Master's degree thesis

LOG950 Logistics

Demand forecasting, distribution, replenishment and in-store redistribution of goods in shoe retail

Ivan Kaipau

Alena Navitskaya

Number of pages including this page: 83

Molde, 24.05.2019

Molde University College
Specialized University in Logistics

Mandatory statement

Each student is responsible for complying with rules and regulations that relate to examinations and to academic work in general. The purpose of the mandatory statement is to make students aware of their responsibility and the consequences of cheating. Failure to complete the statement does not excuse students from their responsibility.

*Please complete the mandatory statement by placing a mark **in each box** for statements 1-6 below.*		
1.	**I/we hereby declare that my/our paper/assignment is my/our own work, and that I/we have not used other sources or received other help than mentioned in the paper/assignment.**	☒
2.	**I/we hereby declare that this paper** 1. Has not been used in any other exam at another department/university/university college 2. Is not referring to the work of others without acknowledgement 3. Is not referring to my/our previous work without acknowledgement 4. Has acknowledged all sources of literature in the text and in the list of references 5. Is not a copy, duplicate or transcript of other work	Mark each box: 1. ☒ 2. ☒ 3. ☒ 4. ☒ 5. ☒
3.	I am/we are aware **that any breach of the above will be considered as cheating, and may result in annulment of the examination and exclusion from all universities and university colleges in Norway for up to one year, according to the** Act relating to Norwegian Universities and University Colleges, section 4-7 and 4-8 **and** Examination regulations section 14 and 15.	☒
4.	**I am/we are aware that all papers/assignments may be checked for plagiarism by a software assisted plagiarism check**	☒
5.	**I am/we are aware that Molde University College will handle all cases of suspected cheating according to prevailing guidelines.**	☒
6.	**I/we are aware of the University College's** rules and regulation for using sources	☒

Personal protection

Personal Data Act

Research projects that processes personal data according to Personal Data Act, should be notified to Data Protection Services (NSD) for consideration.

Have the research project been considered by NSD? ☐yes ☒no

- If yes:

Reference number:

- If no:

I/we hereby declare that the thesis does not contain personal data according to Personal Data Act.: ☒

Act on Medical and Health Research

If the research project is effected by the regulations decided in Act on Medical and Health Research (the Health Research Act), it must be approved in advance by the Regional Committee for Medical and Health Research Ethic (REK) in your region.

Has the research project been considered by REK? ☐yes ☒no

- If yes:

Reference number:

Publication agreement

ECTS credits: 30

Supervisor: Irina Gribkovskaia

Agreement on electronic publication of master thesis

Author(s) have copyright to the thesis, including the exclusive right to publish the document (The Copyright Act §2).

All theses fulfilling the requirements will be registered and published in Brage HiM, with the approval of the author(s).

Theses with a confidentiality agreement will not be published.

I/we hereby give Molde University College the right to, free of charge, make the thesis available for electronic publication: ☒yes ☐no

Is there an agreement of confidentiality? ☒yes ☐no

(A supplementary confidentiality agreement must be filled in)

- If yes:

Can the thesis be online published when the period of confidentiality is expired? ☒yes ☐no

Date: 11.03.2019

Høgskolen i Molde

STANDARD AGREEMENT

This agreement is between

.........Ivan Kaipau, Alena Navitskaya......................(Student(s)).

...Irina Gribkovskaia......... (Faculty Advisor at Molde University College),

.......BELWEST.. (Company/Institution),

And Molde University College/MSc Logistics Program Coordinator, concerning the use of specifications and results reported in the Master's degree thesis in accordance with the study plan for the Master's degree program in Logistics at Molde University College.

1. The student will complete the work assigned for the Master's degree thesis in cooperation with the company/institution
(Optional):..

The title of the thesis is: Demand forecasting, distribution, replenishment and in-store redistribution of goods in shoe retail

2. The student has copyrights to the thesis. Those copies of the thesis submitted for evaluation along with descriptions and models, such as computer software that is included as part of or as an attachment to the thesis, belongs to Molde University College. The thesis and its attachments can be used by the College for teaching and research purposes without charge. The thesis and its attachments must not be used for other purposes.

3. The student has the right to publish the thesis, or parts of it, as an independent study or as part of a larger work, or in popularized form in any publication.

4. The company/institution has the right to receive a copy of the thesis with attachments, and the College's evaluation of it. The company/institution will have three (3) months from the time the thesis is submitted to the College for censoring to determine whether a patent is possible and to apply for a patent for all or part of the results in the thesis. The specifications and results in the thesis can be used by the company/institution in its own activities.

5. An additional confidentiality agreement may be entered into between the parties, if the company/institution sees this as necessary.

6. Each part of the agreement should have one copy of the agreement. The final signature should be from the Program Coordinator/Dean validating the agreement.

Place...Molde.................. Date of final signature.....23.05.2019.........

Student(s) — Faculty Advisor: Irina Gribkovskaia

Company/Institution — Program Coordinator / Dean

Høgskolen i Molde

SUPPLEMENTARY AGREEMENT/CONFIDENTIALITY AGREEMENT

(Supplement to Point 5 of the Standard Agreement)

This agreement is between

Ivan Koipau, Alena Menitskaya (Student(s)),

Irina Gribkovskaia (Faculty Advisor at Molde University College),

BELWEST (Company/Institution),

and Molde University College/MSc logistics Program Coordinator. This is a supplement to the Standard Agreement, which defines the use of specifications and results reported in a Master's degree thesis in accordance with the study plan for the Master's degree program in Logistics at Molde University College.

1. As stated in Point 5 of the Standard Agreement the parties named above agree to limit access to the Master's degree thesis for 1 Year(s) from the date of this agreement. The limit should be 1 to 5 years, for longer confidentiality periods, contact the program coordinator for a special agreement.
2. The reason for keeping the thesis results confidential is so that the company/institution can avoid giving away information that would give others a competitive advantage.
3. The thesis must be submitted to Molde University College. It will be kept in a secure room in the main archive at the College during the period agreed upon. The thesis can be borrowed during this period if a written request is submitted and the company/institution gives permission. Academic staff members can access the thesis after signing a declaration of secrecy. After the period of confidentiality is ended the thesis will be placed in the library for general use.
4. Four copies of this agreement must be signed so that each party will have a copy. The agreement is valid when it has been approved and signed by the program coordinator or Dean at Molde University College.
5. The confidentiality agreement does not exclude members of the academic staff from being present at the final oral presentation of the thesis.

Place: Molde

Date of last signature: 13.05.2019

Student(s)

Irina Gribkovskaia

Faculty Advisor

Company/Institution

Agreement approved by:

Program Coordinator/Dean

Preface

We want to express gratitude and appreciation to people and organizations, without whom the writing of this diploma would be impossible.

We want to say thanks to the Norwegian government, which in the framework of the project provided us with the opportunity and means to study at Molde University College. Thanks to Sergey Markov, the program coordinator on the Belarusian side, for believing in us and letters of recommendation. The preparation of a diploma would have been impossible without the administration and teachers of Molde University College. Without their professional approach and leadership, we would not have received the knowledge that is needed to write a thesis.

We express our sincere appreciation to our supervisor Irina Gribkovskaia for assistance in the preparation of the thesis. Her serious scientific approach and knowledge of the research topic helped us to avoid errors in the study. She always gave us empathetic guidance and tireless support, from choosing a topic to the final completion of work. She helped a lot in research, helped to correct mistakes, gave valuable advice. We would like to take this opportunity to express our gratitude and deepest respect.

We also want to say thanks to the people from the company who helped us during the work on the thesis. We express our deep gratitude to the owner of the company, for the opportunity to delve into all the processes of the company, work with any data, conduct real research. Also, thanks for taking the time during numerous conversations.

The general director of the company explained to us how the business works in the modern world and what challenges we face. Thanks to him for the advice in which direction we need to move and explaining what tasks for business are priorities.

Thanks to CIO of the company, for the information on how and where you can get the data we need to conduct the study. We are also deeply grateful for additional courses on the use of SAP BI, organized by them. This helped us save a lot of time and acquire new knowledge.

The head of the logistics department also provided invaluable assistance. He told us how the company's transport system works and provided the necessary data and explanations. An invaluable contribution to the understanding of warehouse logistics and the processes taking place in the warehouse was made the head of the warehouse of finished products. He explained complex things to us in simple language and allowed us to evaluate the work of the warehouse from different angles.

The head of the trade department made a great contribution to understanding the principles of distribution of goods. She showed us the principles of work that her department uses now and what shortcomings she sees in this. She also told us about the limitations and problems that we will encounter when conducting research and experiments. Thanks also to her for organizing the data collection. Without her help, we would collect data for half a year.

We express our gratitude to product development specialist, for many hours of conversations and useful criticism.

For a huge amount of data collected we thank programmer. He helped us organize the data collection process in a short time and always took our wishes into account.

We apologize if you forgot to thank someone. And we want to say thanks to all the people whom we did not mention, but who helped us in writing a thesis directly or indirectly.

Summary

Offline sales in the fashion industry are currently facing intense competition from online sales. In order to successfully exist, offline retailers need to optimize their processes.

Demand forecasting is a critical factor for retailers in the shoe industry and in particular for company B. The demand forecast close to real values allows the company to prevent sales losses due to a shortage of goods or overcrowded stores with goods. This thesis paper discusses methods and approaches for demand forecasting based on machine learning and time series forecasting.

Also in the thesis work reviewed methods for the primary distribution of goods in the fashion industry. The adaptation of these methods to the needs of the company has been made. The replenishment of goods in the store is considered depending on the demand for a certain period and the current balance of goods in the store to minimize sales losses. The same approach is used at the end of the collection's life cycle to transfer goods between stores. In this real case, the relevant demand forecasting methods are selected. The most suitable and rarely used forecasting method has been implemented. Also considered are the methods of primary distribution, replenishment and redistribution of goods between stores, which are used in the company at the moment. Own heuristics were also proposed for these processes and an experiment was conducted on their effectiveness.

To perform all these tasks, it was necessary to collect a large array of data. Data was collected and formatted for convenient use for analysis.

Contents

1.0 Introduction

The first subchapter will present an analysis of the shoe market and competitors in Belarus and Russia. After that, the background and motivation of the research, the problem of the research, the organization of the rest of the thesis are presented.

1.1 The situation in the fashion industry

Online commerce with the development of the Internet has received a tremendous impetus to development. This is well seen in companies like Amazon, Alibaba, Ebay, etc. Consequently, there is an outflow of buyers from offline retail to online, which leads to a drop in revenues of companies that only work online. The decline of offline retail is especially noticeable on the example of the US market. Many have noticed this in the light of recent news about the closure of Macy's large retail chain. Macy's is a 160-year-old company that has lost $ 34 billion in market capitalization to participating companies in the retail sector.

Many retailers did not see the benefits of digital opportunities. Over the past few years, consumers have realized that going to the store is a much less productive process compared to online shopping. And it does not depend on which online stores are discussed in comparison. Below are the indicators that the analytical group PYMNTS received. They compared data from the Census Bureau and other sources and built their own model based on them. Analysts say that stores selling clothing, sports goods and electronics cannot confirm that 90% of retail sales still occur in their premises, if they occur at all.

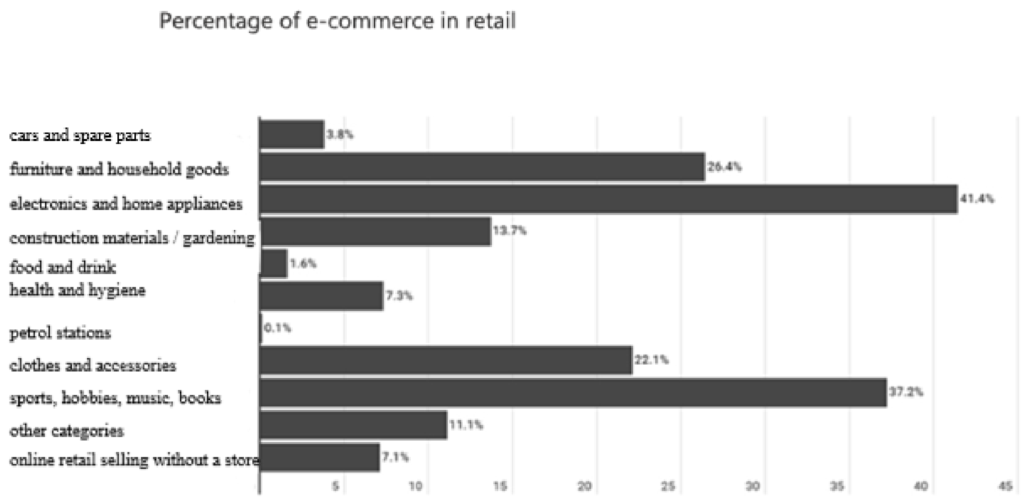


Figure 1.1: E-commerce share in retail (Wirex 2019)

You do not need to be an analyst and build complex information models in order to understand that people began to buy less in physical stores than ever before. It was enough for the last four years to visit physical outlets and communicate with their staff, who would say that the flow of customers is steadily decreasing.

Mobile applications and logistics innovations have significantly improved online shopping practices. But the in-store experience has become less reliable. Consumers, who call time the most valuable resource, want the buying process to be convenient and free from any uncertainties. Trips to the physical store do not always meet these two requirements. Therefore, people prefer to buy online those things that were once more convenient to buy only in physical retail - clothes, jewelry, sports goods, electronics. Increasingly, this also includes home goods and car parts

A PYMNTS research last fall, which was attended by 2,600 people, confirmed this thesis in two product categories. These are clothes and cosmetics. After all, it is assumed that the opportunity to personally touch and try out samples of goods gives real stores a tangible advantage. However, only 42% and 34% of consumers who bought clothes or cosmetics, respectively, reported that they purchased the goods in a physical store.

Over the past few years, retailers have spent hundreds of millions of dollars trying to lure consumers into their stores. They organized demonstrations of fashionable clothes and shoes, called experts for the presentations of new products, agreed with the stars and famous athletes to represent brands, install special mirrors to make fitting of clothes or shoes more comfortable. They gave the sales staff tablets with recommendations, combinations of products or tips on the use of certain products. They offered discounts and promotions, valid only when visiting physical outlets.

Next in line are robots who greet customers at the doorstep of the store and help them find the right products. Perhaps there is an application of AR and VR technologies. A few years ago, analysts expressed the opinion that the future of physical retail as a category and physical stores as points of interaction with consumers will be similar to the evolution in the media sphere: only the most large-scale or highly specialized projects will survive. The largest retailers will be able to take advantage of their branching to provide a range and efficient logistics. This will satisfy the needs of consumers across all channels, including digital.

Physical retail of the future will pass the road of modern media: there will be only players who have adopted new technologies, and business models offering a new look at familiar processes (Wirex 2019).

1.2 Market situation and competitor analysis

Since our company has a network in Russia and Belarus, it is necessary to divide the analysis of the market and competitors into two parts. The capacity of the markets is significantly different and the competitors are also different. In Belarus, the main producers and retailers of footwear are Marco, Kari, Megatop and the company, which is referred to in the diploma. Less than a year ago, the company Inditex came to Belarus, with its subsidiaries Zara, Bershka, Pull & Bear, Massimo Dutti, etc. In the future, Inditex and its subsidiaries may become a major player in the Belarusian market, but now their footwear sales are insignificant and occupy about 3% of the market.

In 2018, footwear production in Belarus decreased by 4% compared with 2017, to 9.6 million pairs, and consolidated the downward trend observed in recent years (Kollontai 2018).

Another reduction in output occurred against the background of an increase in 2018 in retail sales of footwear in the whole country by 11.2% to $ 486 million, including large and medium-sized trade organizations - by 12.6% to $ 278.6 million.

Consistently increasing the number of stores, some shoe chains continued to strengthen their presence in the market. In November 2018, the share of shoe sales in the markets and shopping centers increased to 8.5% (against 8% a year earlier), which strengthened the position of this retail segment. Growth in sales could benefit mainly sellers of imported products.

In 2018, the import-oriented sales model was formed on the Belarusian footwear market: foreign footwear was widely represented in trade, while the share of domestic footwear in sales volumes decreased to 47.9% against 52.8% in 2017. Our company also influenced the decrease in the share of footwear produced in Belarus on the market, which began selling imported footwear in stores.

The largest share in the Belarusian shoe market is occupied by the manufacturer and retailer of footwear “Marco”. This company owns 37% of the market, the annual production of about 5 million pairs. Retail chain has 81 stores throughout Belarus. Half of the shoes produced are sold wholesale to Russia and sold in stores of other shoe retailers. Some shoes are made in China and India.

Also one of the largest shoe retailers on the Belarusian market is Megatop. This company does not have its own production, the company buys shoes from small local manufacturers, and also imports shoes from China. This company has 69 stores in Belarus. Kari is also represented in Belarus. This is an international company that sells shoes, accessories and clothing. In Belarus, has 39 stores.

Our company is represented in Belarus 51 store. About 25% of the company's annual output goes to Belarusian stores. 18% of imported footwear goes to Belarusian shops.

Based on the above, we can say that our company is in the top 3 shoe retailers in Belarus. We also have a big advantage over Megatop, which is currently superior to us in the number of stores. They do not have their own production. Therefore they are less dynamic in replenishing stores with goods. This leads to large sales in the stores and, consequently, to sales losses. Since we can more quickly respond to customer demand and meet their needs through our own production in Belarus. We have the opportunity to compete for the second line of the rating of the largest shoe retailers in Belarus.

Experts predict a growing footwear market in Russia. At the end of the year, it will grow by 4 percent with the improvement of the economic situation (Popova 2018).

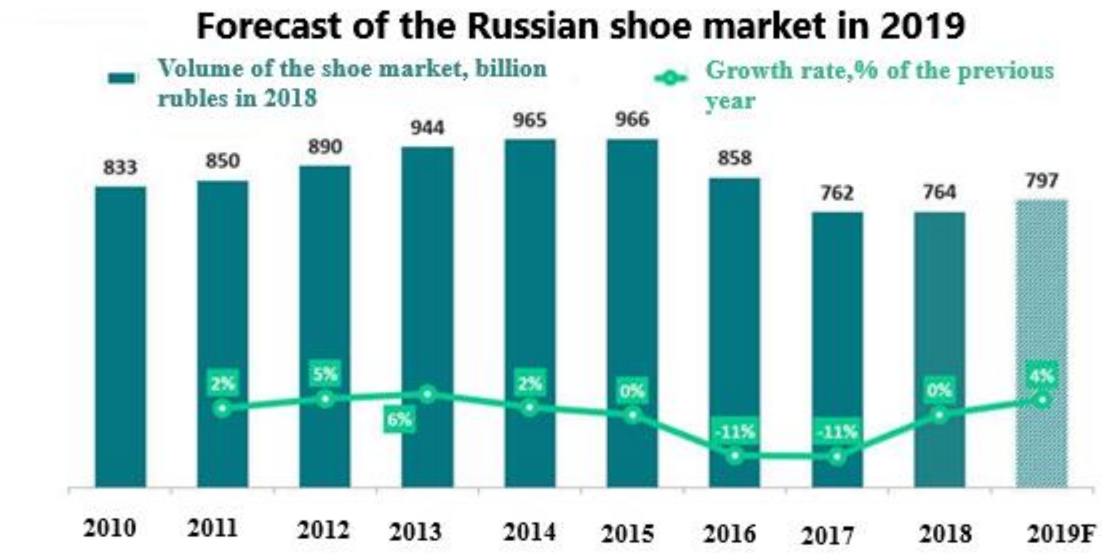


Figure 1.2: Forecast of the Russian shoe market in 2019 (Popova 2018)

The size of the shoe market in 2018 was 18 percent lower than the 2014 crisis year. Care continues for weak players and the closure of inefficient stores. Retailers continuously optimize business processes. However, a quick recovery of the shoe market to the level of 2014 is not expected. At the same time, buyers continue to strive for discounts and special offers. 33 percent of Russians in 2018 tried to buy shoes that are sold with discounts. 11 percent specifically waited for discounts and promotions for buying shoes.

Figure 1.3: Leaders among shoe retailers by number of stores on February 2018 (Popova 2018)

The data in the table for the year is a bit outdated. Now our company has 342 stores in Russia, 37 stores will be opened during 2019. In general, experts believe that our company is in the top 5 shoe retailers in Russia based on a combination of factors such as the number of stores, turnover, revenue, etc.

1.3 Research problem

In this thesis, we consider the case of a shoe company, which has an extensive network of stores in Belarus and Russia. Based on the above, it is a great luxury to lose new or already loyal to the brand customers because of situations when the store did not have the right shoe model or was not the right size. The purpose of this thesis is to propose methods for forecasting the demand for goods in stores to minimize the situation when a customer chose our store, perhaps even a specific shoe model, but could not buy it because it was not available in the store. Proper demand forecasting is critical for a company for a number of reasons, including the growing share of online trading. Also, near-to-reality demand forecasting allows the company to increase revenue, reduce operating expenses and working capital.

Company B wants to use the forecast of demand for products in stores with a time horizon from 14 to 275 days to distribute products to stores. An error in forecasting demand up to 90 days should not exceed 25%, while in forecasting demand with a time horizon of 275 days, the prediction error should not exceed 40%. The companies are also interested in developing heuristics for the distribution, replenishment and redistribution of goods between stores. An effective way is needed to increase sales, reduce sales losses, increase revenues and reduce working capital.

In the framework of this thesis, we set ourselves the tasks:

- Offer a demand forecasting method and test its performance in real life;
- Offer heuristics for the initial placement, replenishment and transfer of goods between stores.

1.4 Organization of thesis

The second chapter presents a description of the problem, explains the company's model of work in the shoe business, describes the company, and describes the processes that are currently used by the company to forecast demand, distribution, replenishment and transfer of goods. The third chapter is defined problem (research tasks) and the formulation of the problem. In the fourth chapter, we made a review of the literature on the topics of forecasting the demand in retail, the optimal range of the store during the initial placement, replenishment of goods in stores, the transfer of goods to stores to increase sales. The fifth chapter describes the methodology for solving the tasks assigned to us. The sixth chapter describes the data that was available to us and how to obtain this data. The seventh chapter discusses the selection and development of a prediction method. The eighth chapter presents heuristics for initial placement, replenishment and transfer. The ninth chapter is the last part of the diploma, consisting of conclusions, recommendations and directions for future research.

2.0 Problem description

The first subchapter describes the general principles of business in the shoe industry. In the second subchapter, the company's case is described. The third subchapter describes the demand forecasting method that was used in the company. The methods used by the company to distribute and replenish goods are described in the fourth and fifth sub-chapters. In the sixth sub-chapter, existing approaches are criticized.

2.1 Shoe industry

Shoe retail is part of fashion retail, which is characterized by seasonality of the product. In the shoe business there are four main seasons:

- Winter;
- Summer;
- Demi-season (insulated shoes for the off-season during spring and autumn);
- All-season or year-round.

The average duration of the seasons varies from region to region. For example, in St. Petersburg (59 ° 56.3457 '0 "N, 30 ° 18.9521' 0" E) the duration of the winter season is very different from the duration of the winter season in Sochi (43 ° 35′57 ″ N, 39 ° 43′32 ″ E). In order to simplify this part of the diploma averaged over all regions, the duration of the seasons. We assume that the summer and winter seasons last 4 months, the demi-season lasts for two months from mid-September to mid-November and from mid-March to mid-May. For each season, designers are developing a new collection of shoes, focusing on new trends in the fashion world, or simply rethink the classics. After that, a new collection begins to be produced.

Obviously, it is silly, for example, to sell winter shoes in the summer. Of course, almost no one will buy it, and therefore the company will not receive income. Consequently, shoes of the right season need to be delivered to stores at the right time. Suppose we know the time when you need to deliver shoes. The next step is to understand how many shoes you need to get into stores. This step will be considered in this thesis in fact, this is demand forecasting. Suppose we know how many shoes customers want to buy during the entire season.

Should we immediately bring all the shoes to the store and wait for them to buy? The probability of selling on store shelves is higher than in stock. However, this is physically impossible for several reasons. The most significant reasons are the limited capacity of the stores and not always the right amount of shoes is available at the very beginning of the season. In reality, production cannot provide all the shoes for distribution to the stores at once, the reproduction takes place during the season.

Restrictions of real life lead to the fact that shoes for the season are delivered to stores in several stages. The first batch, usually the largest, is called initial placement. Subsequent batches from the warehouse during the season are called replenishment.

Towards the end of the season, stocks in warehouses are running out. Sales and revenues are also falling. In some stores from the seasonal collection, almost nothing remains or the dimension of the articles is incomplete. For example, the article should be presented in the amount of 36-40, but only 36 and 40 sizes remained in the store. In this case, the goods are redistributed between stores. This operation is necessary for the company to stimulate sales and maximize the speed of sale of the collection. This stage in the life of the collection is called the transfer between stores.

Next, the company needs to minimize the number of pairs of shoes of this collection in the store. The sales stage begins, which includes big discounts on seasonal shoes.

In general, the life cycle of the collection can be described by the schedule:

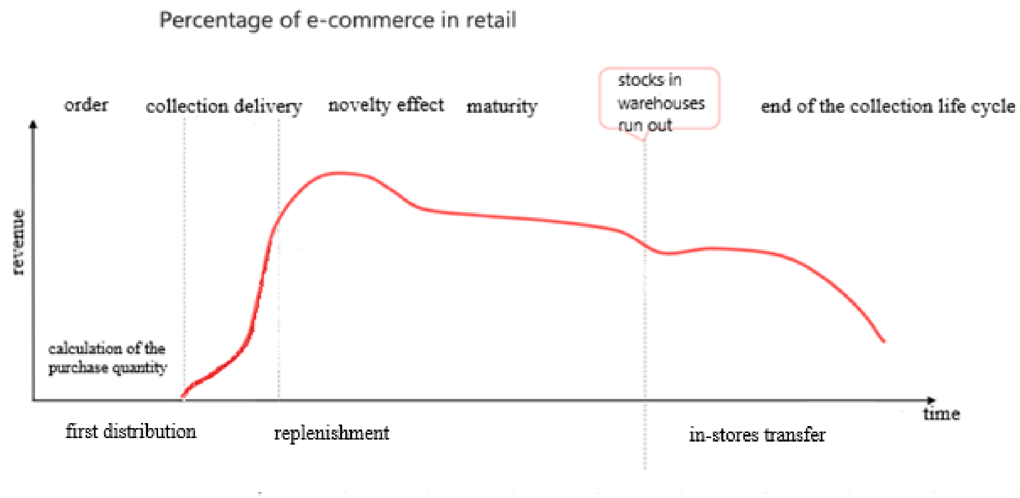


Figure 2.1: Shoe collection life cycle

When the season ends, comes a new collection of shoes for the new season. The remains of the old collection are returned to the warehouse of the company for storage.

2.2 About company

In our example, we will work with a real problem for the Belarusian shoe company B. The company produces various types of shoes for all seasons and bags. Company B was founded in the late 80s in the city of Vitebsk. It was the time of the decline of the USSR, the country's economy was in a disastrous state. Nevertheless, people had money to buy food, clothes, shoes and other goods, but there was an incredible shortage in the stores. At that time there was a situation in which it was possible to buy something only with coupons issued at the place of work. However, even with a special coupon, it was necessary to wait until goods were brought to the stores. But the production is almost not working because of the terrible economic condition of the country.

Company B was opened with the participation of foreign capital, so the problems of the Soviet economy did not affect the company directly. At that time, company B only had shoe manufacturing and did not have its own stores. State-owned companies or structures bought shoes from company B and distributed shoes to the population by coupons through government stores. Thus, the company gained fame and respect in the USSR, and later in Belarus and Russia.

Picture 2.1: Illustration of shoes made by company B

Company B won the hearts of the people not only because of the economic condition of the country. The way of development of the factory was revolutionary in the shoe industry. Since its inception, the company has used advanced world technologies and unique knowledge necessary for production. comfortable and high-quality shoes. Company B today is a high-tech, full-cycle enterprise. This is one of the largest shoe companies in Eastern Europe, successfully developing for 28 years. The company works in tandem with science, introduces innovative technologies in production, designers collaborate with Italian design studios. The company is aware of global fashion trends and gives everyone the joy of good shopping.

Due to the expansion of staff, IT department of company B was reorganized into a separate company. Employees are involved in programming, IT support, data analysis, robotics, etc.

This decision was made due to the need to keep up with the times and comply with the technological level of the leading companies in the shoe industry.

Picture 2.2: Illustration of shoes made by company B

Company B has its own production in Vitebsk, which produces more than 2 million pairs of men's and women's shoes per year. The company also buys more than 1.5 million pairs of shoes and about 400 thousand from Italian, Spanish and Portuguese manufacturers and delivers them to warehouses in Vitebsk for distribution to stores for sale. The company also has its own network of stores selling goods in Belarus and Russia. New stores are opening, some of the old ones are closing. At the moment, the company has 342 stores in Russia, and 52 stores in Belarus. In 2019, the store chain is expected to increase by 10 percent. The geography of stores in Russia from Kaliningrad to Ulan-Ude, from Sochi to Novosibirsk. The company employs about 4,000 people. The company's annual turnover in 2018 was $ 97 million.

Picture 2.3: Illustration of shoes made by company B

For the production, storage and delivery of goods to the stores, the company has several production workshops, two warehouses and a vehicle fleet. The main production facilities are located in the city of Vitebsk in the north of Belarus. Production facilities are located on the same territory. The main industrial building consists of four floors. On average, about 200 technological operations are carried out for the production of a pair of shoes, therefore, one workshop is located on one floor. On the ground floor there is a workshop for laying out and cutting pre-processed skins. The workshop receives a task for which articles you need to get the details from the skin. In the workshop there are both modern complexes for laying and cutting skins, as well as morally and physically outdated installations. The task for layout and cutting comes in electronic form for modern complexes. In the form of physical samples that need to be laid out on the skin, and then served on the cutting.

Nevertheless, even modern equipment cannot yet completely replace a person. For example, the skin of an animal on a ridge is thicker than on a belly. Therefore, in accordance with the requirements of the production of various models of shoes, the skin of an animal must be zoned into several parts. This is necessary because some parts can only be put in certain areas. Modern complexes do not zone the skin, so this technological operation is performed by people.

As an experiment, one of the authors of this thesis modernized the software of the automated complex for automatic zoning. As a result, when the skin is on the desktop of the complex, using computer vision, it is determined on the table and its orientation in space. As a result, the operator of the complex sees the below image on the monitor in front of him.

This is followed by the process of troubleshooting, which also does not do without a person. A worker with a laser pointer marks areas with defects and assigns them a certain degree of damage. This is necessary because some parts can be applied even to those places where there are defects. Troubleshooting requires a lot of experience, exceptional attentiveness, sharp eyesight and sharp tactile sensations from the operator. This job is one of the highest paid in the workplace. It is also considered that only women can save the required level of care in the process of defending for a long time.

After the operation of troubleshooting, there are two possible modes of operation on the automated complex. The first mode is fully automatic layout and cutting. The complexes used for the layout of the genetic algorithm, which is iterative and computational. Because of this, several iterations of the layout algorithm take several minutes, which is quite often prohibitively long for production.

Picture 2.4: The result of the algorithm of automatic skin zoning

Therefore, operators often stop the algorithm after the first or second iteration. Because of this, the percentage of use of the useful skin area is far from the maximum possible. The second mode of operation is the layout of parts on the skin by the operator. On the screen of the complex, the operator, using a computer mouse, places the parts into accessible places over the entire skin area. Real-time complex lasers project the location of parts on the skin. This is done for the convenience of the operator of the complex and to monitor possible errors. When working with skins of light color, it is possible to work only in the second mode of operation.

Picture 2.5: Cutting complex

After that, the parts are conveyed to the second and third floors, where sewing shops operate. Seamstresses make the upper part of the shoes from details, after which the upper of the shoe is sent to the next floor to be joined to the sole. Soles for shoes are either purchased from supplier companies, or manufactured by the company B. The workshop where the soles are made is called injection molding. The main production building on the fourth floor connects the top of the shoe to the sole. A special glue is applied to the sole and top of the shoe, after which it is processed in special equipment under high pressure and temperature.

After the last technological operation, shoes undergo quality control. If the shoe meets the quality requirements, it is marked and packed in boxes to the warehouse. If the shoe does not meet quality standards, then it is sent for revision.

After labeling and packaging, the shoes are immediately sent to the finished goods warehouse, because there is no storage space in the production building. The warehouse of finished products is located 1.5 kilometers from the production building. The products are delivered by the transport of company B. The building of the finished product warehouse is a two-story building with two freight elevators with an area of about 8,000 square meters. Up to 600,000 pairs of shoes can be stored in the warehouse of finished products at a time At the moment, the work of the warehouse of finished products is organized as follows. After the arrival of the car with shoes, it is unloaded on carts and lifted to the second floor, where all the boxes from the lot are placed on the storage sites. After that, part of the shoe is packaged in a box, in which six pairs of shoes are usually placed, the box is marked and put in the address storage box. The box can be stored in the address box on both the second and the first floor. Some shoes are not packaged in boxes, but placed in boxes in address cells. As a rule, this is done to quickly select the right pair of shoes to the store.

Picture 2.6: Product warehouse

Zone A is called shelving where shoes are stored “placers”, Zone B is called shelving where shoes are stored in boxes. The process of placing boxes or “placers” by workers is called the replenishment of zone A or zone B.

After the task comes to the warehouse to send a certain batch of shoes to the stores, in the SAP warehouse system, a task is issued to the workers. This operation is called the selection of shoes. Each worker has a portable scanner, and it is up to him to take a certain box or box with shoes in a certain address cell into the trolley.

After the boxes are selected, they are marked and put on pallets for shipment. For the “placer” there is another operation: packaging. Packaging takes place in boxes that are pre-assembled and labeled. This was done by other workers during the placer pick. After packing "placer" in the box, they are also put on pallets to the boxes already standing. After that, a batch of shoes ready for shipment to the store.

There is also a second finished goods warehouse. It is a one-story building with an area of about 8,000 square meters. It processes and stores shoes that are returned from stores at the end of the sales season. Before putting the shoes into address storage, the operation of receiving shoes is performed on this hoard.This operation is necessary because there are cases when the shoes are returned from stores in non-commodity form. At the acceptance stage, the shoes are checked for compliance with the presentation. If a shoe passes the test, it goes to address storage in zone A or B. If the shoe does not pass control, then it goes through the procedure of purchasing the presentation before being stored.

Also imported shoes arrive at the second warehouse. For her, the procedure for placing in storage is no different from the one when the shoes returned from stores. It is not uncommon for a box to have only one shoe or a brick instead of shoes. Naturally, in such cases, the company imposes penalties for the importer for non-compliance with the contract for the supply of shoes.

In general, the transfer of goods in company B can be illustrated by the scheme.

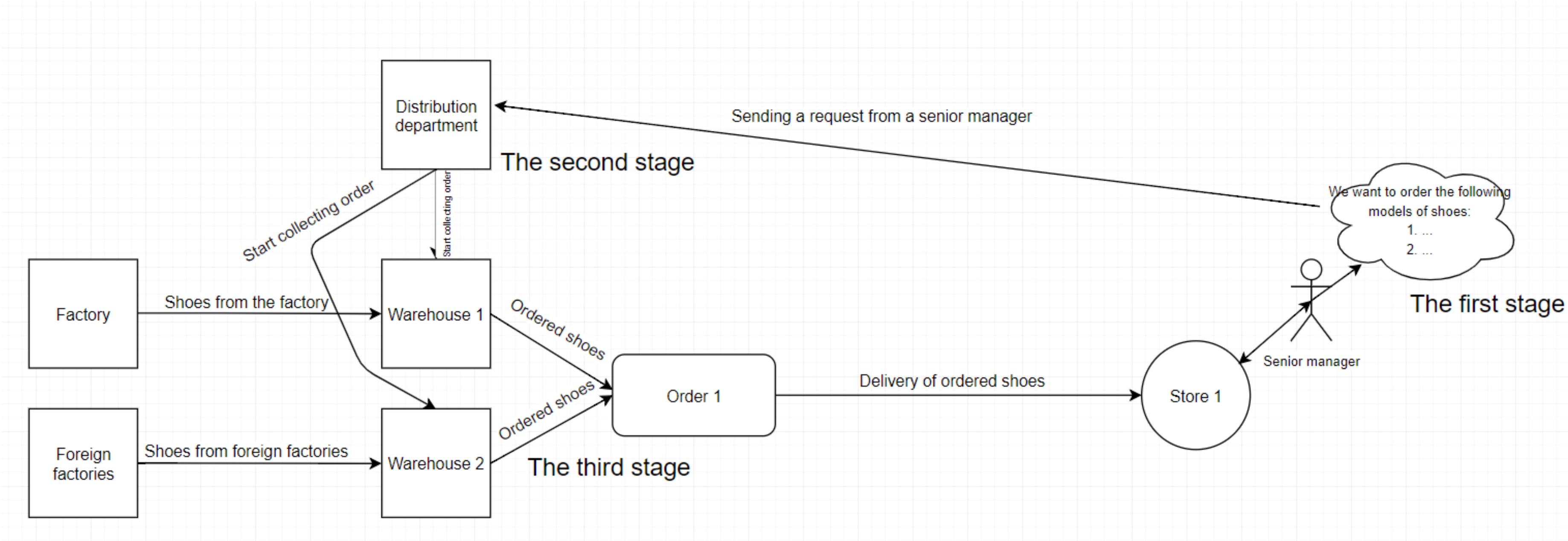
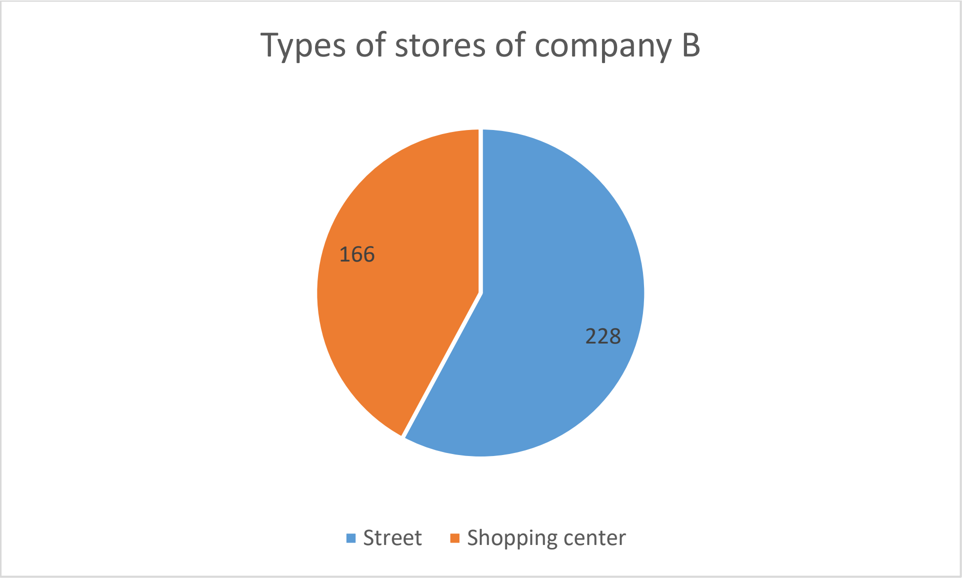


Figure 2.2: Current shoe distribution scheme in company B

At the same time, company B is not only one of the largest shoe manufacturers, but also one of the largest shoe retailers in Russia and Belarus. The chain of stores currently consists of 342 stores in Russia and 52 stores in Belarus. By the end of 2019, it is planned to open about thirty new stores, mainly in Russia.

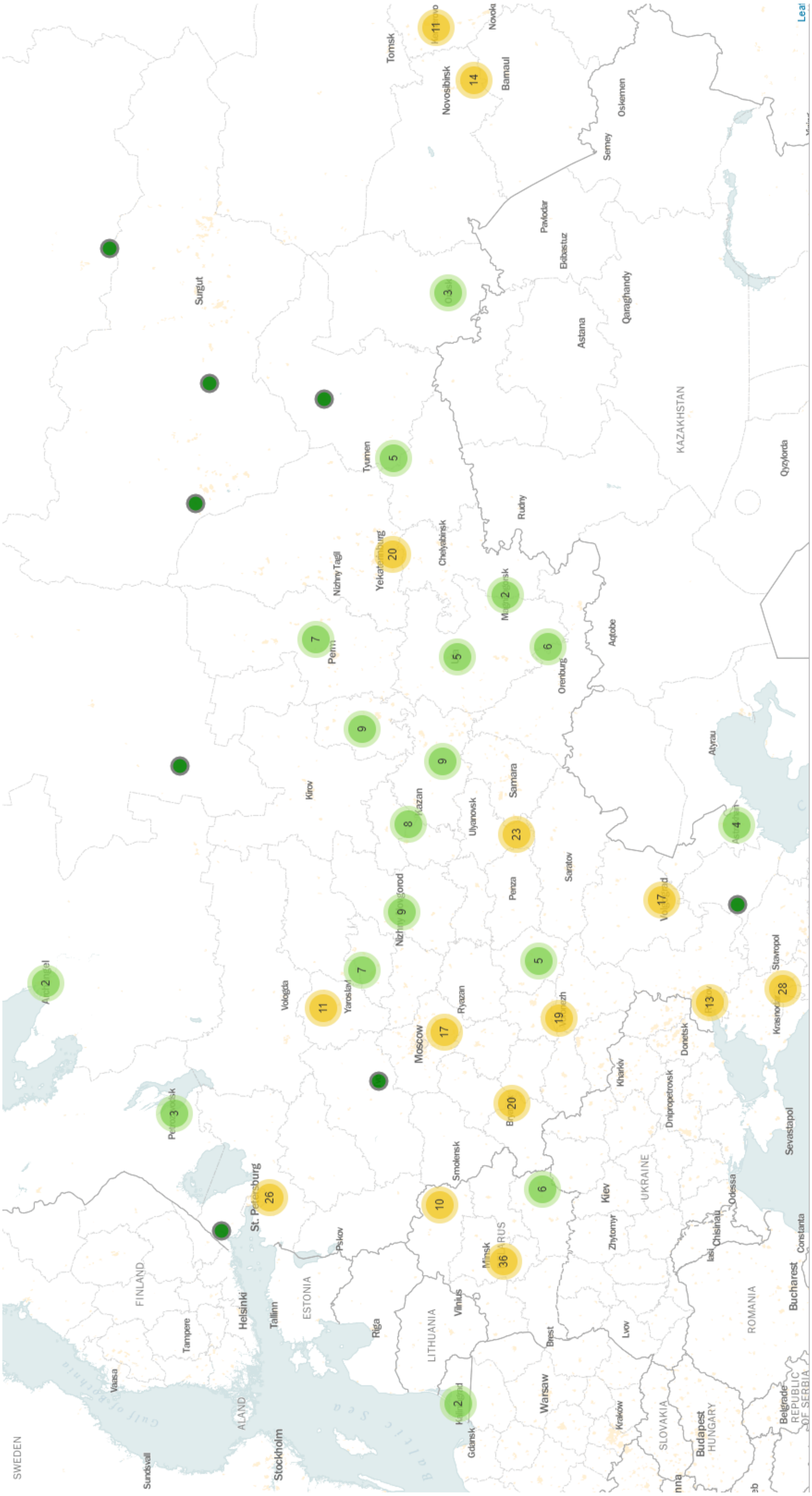
Picture 2.7: Showcase store company B

Types of stores can be divided into two types: street and located in shopping centers. The distribution of stores by type is presented below. Currently, street-type stores are still more, but in recent years there has been a trend that the share of stores located in shopping centers has steadily increased. At the moment, the share of stores located in shopping centers is 42.1%.



Figures 2.3: Types of stores company B

For convenience, we have visualized the company's network of stores. On the interactive map, you can zoom the map, select a store and view information on it. The map was developed in the Python programming language using the Folium visualization library. Circles mean the location of the stores in the region and their number. The color of the circle means the density of the stores in the region.


SWEDEN
Sundsvall
Stockholm
ALAND
Vaasa
Tampere
Helsinki
Tallinn
FINLAND
ESTONIA
Riga
LITHUANIA
Vilnius
Minsk
Pskov
St. Petersburg
Smolensk
Moscow
Ryazan
Yaroslavl
Vologda
Kirov
Perm
Kazan
Ulyanovsk
Samara
Penza
Saratov
Orenburg
Chelyabinsk
Tyumen
Nizhny Tagil
Surgut
Tomsk
Novosibirsk
Barnaul
Semey
Oskemen
Pavlodar
Ekibastuz
Astana
Qaraghandy
KAZAKHSTAN
Qyzylorda
Rudny
Aqtobe
Atyrau
Stavropol
Donetsk
Kharkiv
Dnipropetrovsk
Sevastopol
Kiev
UKRAINE
Zhytomyr
Odessa
Chisinau
Iasi
Constanta
Bucharest
ROMANIA
Lvov
Brest
Warsaw
Kraków
SLOVAKIA
Budapest
HUNGARY
Belgrade
REPUBLIC OF SERBIA
Gdansk

2.3 Company demand forecasting

Forecasting demand for products in stores is crucial for any retail company. The problem lies in the shortage or surplus of products. At the very beginning of the company's life cycle, the excess products do not significantly worry the owners. The curve of growth of margin and sales initially goes up. This suggests that all goods not sold this month can be sold in the following. If the surpluses can hide behind optimistic forecasts for some time, then it is easier to see deficits: there is nothing to ship to stores, and all that could be earned during shipment is lost profit.

The solution to the problem seems obvious - you can simply increase the safety stock. But as soon as we begin to do this, the value of the stock begins to grow and does not solve the problem of deficiencies. Over time, we wonder: why with the growth of the company and the volume of production, profits stop growing? And sometimes even there is not enough working capital to pay off receivables or to purchase the necessary components for real orders. Just at this moment, attention switches to money frozen in stocks. What to do in this case? One way to find a balance between shortages and surpluses is accurate forecasting of demand in stores.

At the moment, in company B, the global forecast of sales in the markets of Belarus and Russia in the preparation of a business plan is carried out by specialists of the economic analysis and controlling department. Forecasts in this department are tactical, with a time horizon of one month, or strategic, with a time horizon of 1-3 years. A more accurate definition of the activities of this department is the word planning, not forecasting. The result of the work of this department is informing other departments about how much we have to sell products in the next month or year. Further, in the marketing and advertising department it is predicted how much a certain article should be produced. Production forecast algorithm:

- Displays the average weekly sales rate. Taken to calculate the periods when the value of the indicator was not zero. To calculate the average value of the weekly sales rate of the article, in the reports of the SAP BI system, select the indicator “Rate of sale of the article”. After that weeks of active sales are selected. For shoes of each season, there are weeks of active sales. For all stores, the weeks of active sales of each season are the same, although many stores are located in different climatic zones. For the summer season, active weeks are 10-41 weeks. For spring all-season shoes 4-23 weeks, for autumn all-season shoes 29-49 weeks, for spring insulated shoes 4-19 weeks, for autumn insulated shoes 33-49 weeks, for winter shoes 36-13 weeks;

- Predicted sales are calculated as the average sales rate per item per week multiplied by the number of active sales weeks multiplied by the number of stores and multiplied by a coefficient of 0.7. The number of stores is calculated as the sum of existing stores and stores that will be opened. The coefficient of 0.7 is the accepted threshold value of the implementation coefficient;
- The need to start is determined by the estimated sales volume of the article minus the current balance of the article. Under the current balance refers to the number of pairs of shoes of a certain article in the finished goods warehouse, in stores at the moment or in transport;
- After receiving the value of the parameter of the need to run, the decision is made to launch the article into production. If the article should be presented with full size, dimensions 36-41 or 35-41 for women's shoes, 40-45 for men's shoes, then the article will go into production with the “need to start” parameter set to more than 500.

This algorithm is applied only to articles that are issued several seasons in a row, that is, they are transferred from the old collection to the new one. For absolutely new articles, this algorithm cannot be applied due to the lack of statistics. When forecasting sales of new articles, or rather, forecasting the volume of production, decisions are made based on the expert opinion of the most experienced employees. The approaches they apply are not documented. In general terms, their methodology was able to identify during the interview. The approach is that the new article X is very similar to the article Y from the previous collection. Article Y we produced in the amount of 1000 pairs, the implementation ratio was 0.7. Therefore, we will produce 1000 pairs of the article Y. There may be some fluctuations in the parameter of the forecast sales volume associated with new trends in the fashion world. Table 1 shows the calculation of the sales forecast for a specific article. The estimated number of pairs sold based on sales statistics for 2017, 2018, as well as the beginning of 2019 is given. As can be seen from the table, sales are planned at the level of 2180 pairs of this article. The current balance is 409 pairs, therefore the calculated launch volume is equal to 1771 pairs.

количество действующих магазинов, РФ	342	количество недель активных продаж	21 всесезонный ассортимент, 29-49 нед. 2019
количество действующих магазинов, РБ	52	коэффициент реализации	70
количество новых магазинов	3	процент полной ростовки	10
количество пар на артикул в резерв на новые магазины	5	процент компенсационной ростовки	90

Артикул	Размер	Прогнозное значение ТО, пар			Расчетный объем запуска, пар	Предлагаемый объем запуска, пар	уд.вес, %	в т.ч.		Выпуск, пар						Продажи 2017, пар			Продажи 2018, пар			Продажи 2019, пар				Текущий остаток, пар					
		РФ	РБ	итого				компенсационная ростовка	полная ростовка	2016	2017	2018	2019	итого	уд.вес, %	РФ	РБ	итого	РФ	РБ	итого	РФ	РБ	итого	уд.вес, %	РФ	РБ	склад	в пути	итого	уд.вес, %
1637948/1	Результат	1 265	242	2 180	1 771	1 000		900	100	309	2 525	2 924	0	5 758		1 649	543	2 192	2 260	600	2 860	94	28	122		339	41	17	12	409	
	F350					40	4,00	33	13	0	80	70	0	150	2,61	37	30	67	63	10	73	3	1	4	2,78	2	1	0	0	3	0,73
	F360					120	12,00	104	13	29	230	310	0	569	9,88	165	46	211	240	65	305	8	0	8	10,13	23	3	1	2	29	7,09
	F370					270	27,00	257	13	82	518	669	0	1 269	22,04	381	137	518	508	145	653	17	3	20	23,02	37	2	4	1	44	10,76
	F380					270	27,00	258	13	77	528	738	0	1 343	23,32	402	115	517	542	162	704	19	7	26	24,10	52	3	2	0	57	13,94
	F390					190	19,00	175	13	74	550	560	0	1 184	20,56	351	120	471	449	112	561	19	7	26	20,45	72	12	5	4	93	22,74
	F400					40	4,00	29	13	30	230	418	0	678	11,77	139	39	178	291	52	343	20	6	26	10,57	92	11	3	3	109	26,65
	F410					30	3,00	19	13	17	199	159	0	375	6,51	99	27	126	121	42	163	6	4	10	5,78	45	8	2	2	57	13,94
	F420					40	4,00	24	13	0	190	0	0	190	3,30	75	29	104	46	12	58	2	0	2	3,17	16	1	0	0	17	4,16

2.4 Distribution of goods on the shops in the company

To describe the distribution problem, we need to describe the distribution process that exists now. Now the number of stores is about 400. In each store there is a senior manager. At the moment, the senior manager of each store creates a list of items that he wants to get from the factory. There is a motivation system for senior store managers. The more goods sold in the store, the more cash bonuses the senior manager. Now we see the personal interest of people who make decisions about placing an order in a store.

What negative consequences can be for a company if you follow the above strategy? Senior managers are trying to order the shoes that they think will sell better. It often happens that a senior manager orders products more than the store can sell. The goal of any private company is to maximize profits. But if the goods are not sold in the store, the company incurs losses. This may be due to the fact that the purchasing power in the area for this type of product has been exhausted.

Secondly, each shoe model has its own level of sales. In other words, this is the popularity of the shoe. Senior managers order such models mainly in large quantities, so the demand for such models from senior managers exceeds supply from production. In this case, those stores whose senior managers have the best ability to convince will get the necessary shoe model.

As a result of this approach, some stores get more goods than they can sell. Other stores may sell more merchandise, but cannot, because they have no goods to sell. The existing supply chain is based on the experience and personal interests of employees who compete with each other for goods and do not apply scientific methods.

Recently, the company is beginning to realize the negative aspects of this approach. As an experiment, the distribution of goods across the shops in the company created focus groups of employees who are trying to systematize the approach to the distribution of goods. Experimental approaches have been applied for the past year and a half. One of the authors of this thesis got into such a focus group, so we can tell you about the heuristics that were used to distribute the winter collection to stores.

After arriving at the warehouses of finished products for the winter range, footwear was selected from production and unsold residues from the stores, which had a salable appearance. Also, for the distribution of the winter assortment, the store rating for sales of the winter assortment was made from 08/01/2018 through 03/01/2019. At least 3 sizes of one article must be received in each store, with at least 1 size being the most common. These sizes for women's shoes are sizes 37-39, for men 42-43. The distribution of articles starts with the collection in 2018 in descending order. During the year, the collection is distributed according to the articles with the most residues.

The algorithm includes several stages. The first stage of the algorithm is the highest priority, the second stage is the second in priority, and so on in descending order:

- If a store has 5 or more sales of a certain article, the full size range of this article is delivered to this store. If it is not possible to supply a full size range, at least 3 sizes are supplied. At the same time at least 1 size is the most popular;
- If a store has 4 sales of a certain article for a period of a taken store rating and the rest of this article is in store 0 or 1, at least 3 sizes are supplied, with at least 1 size being the most popular;
- If in the store for the period of the taken rating of stores there were 3 sales of a certain article and the remainder of this article in the store is zero, at least 3 sizes are supplied, with at least 1 size being the most popular;
- In stores where a specific item has not been presented during the past winter seasons, the full size range of this item is supplied. If it is impossible to supply a full size range, at least 3 sizes are supplied, with at least 1 size being the most popular;
- If two or less sales of a certain article occurred in the store for the period of the taken store rating, the remaining sizes are delivered, but not a fuller size range;
- Replenish the remnants of articles with purchased sizes to the full size range. Usually the full size range for women's shoes is 35-40 sizes, for men 40-45 sizes.

This heuristic managed to distribute about 250000 pairs of shoes of past winter collection.

2.5 Replenishment of stores in the company

Replenishment of stores is engaged in the sales department of the company. The work process of the department is to track sales of goods in the store and respond to it. The department employs about 20 people. The authors of the thesis for a long time monitored the work of the staff of this department and compiled the average work algorithm. The algorithm is very simple. If there is a sale of a certain article in a certain size, then the employee adds to the next delivery to this store the exact same pair of shoes. They must also take into account the opinion of senior managers who have their own preferences for the assortment and its quantity presented in their store.

2.6 Criticism of acting methods

Company B's main problem in demand forecasting, distribution, replenishment and transfer between merchandise stores is not using scientific approaches. Let's look at each part separately and determine what is going wrong. We believe that the company does not forecast the demand for goods in stores, because a sales forecast does not mean a forecast for demand. For example, a buyer entered X store and purchased a 35 size article Y. We considered it as a sale and were taken into account in the statistics. However, after half an hour, another potential buyer came to the same store, which saw article Y, but did not find his size 35 and left without buying anything. At the moment, the company does not take into account lost sales in forecasting. We believe that sales forecast at the item level are in fact the planning of the production volume of this article.

The overall sales forecast, developed by the department of economic analysis and controlling, is essentially a sales plan, which in percentage terms correlates strongly with the expansion of the chain of stores as a percentage. The above criticism concerns the forecasting of demand for articles that were already on sale and there are sales statistics for The method of forecasting the demand for articles from the new collection, for which there are no statistics, is very dangerous for the company due to the fact that a person is not a computer and may be mistaken or under the influence of cognitive distortions, also known as life experience. In addition, an employee who makes decisions only on the basis of his thoughts puts the company at a disadvantage. If he wants to leave the company, it is not possible to find a replacement for him. Therefore, the company offers such an employee a large salary and various social bonuses. Sooner or later, the amount of dividend payments to an employee becomes greater than the benefits brought by this employee.

The distribution of goods in stores is also a problem for the company. The system of motivation of senior managers is built in such a way that the personal interest of employees is not aimed at achieving the common good for the company, but to increase their income even to the detriment of the company. Cases where a senior manager orders products more than the store can accommodate are not rare. Because of this, the work of the staff slows down, customers have an unpleasant feeling after visiting the store. There were cases that, due to the congestion of the store, fines were exhibited by the fire inspectorate due to the fact that the evacuation routes were blocked in the event of a product fire.

We did not see a scientific approach to solving problems when replenishing stores. The task of replenishment works on the human factor. As practice shows, employees at this stage make a lot of mistakes that lead to decisions that do not bear the company's benefit.

Summing up, we consider it necessary to bring a scientific approach to solving the above tasks.

3.0 Research tasks

We set ourselves the task of developing a method for forecasting the demand for goods in stores, conducting experiments on the effectiveness of the method of forecasting proposed by us. We also want to implement and conduct experiments of already known methods of forecasting based on machine learning. We also set ourselves the task of creating heuristics for distributing goods to stores, replenishing stores with merchandise and transferring goods between stores.

To perform the tasks for each of the problems identified, we conducted literature studies. Since the problems that we need to solve are crucial for each retailer, and the retail business is very competitive, most of the scientific articles on our problems cannot be found in the public domain. Many authors of scientific articles on demand forecasting in retail collaborate or work in research departments of the largest retail networks in the world, such as Nike, Adidas, and so on.

Therefore, we made a decision to hold negotiations with companies that deal with the problems of forecasting demand, optimizing the distribution of goods to the stores of the chain, and so on. Thus, we received information about what methods and approaches are currently used in the most technological retail companies in the world.

Also, in order to accomplish the tasks set by us, it was necessary to carry out a huge work on the date of mining and data analysis. Without data, it is impossible to solve the problems posed by us, therefore the task of obtaining data and analyzing them is one of the main tasks of this thesis. Therefore, we needed to determine what data we need, in what way it is possible to obtain it, in what form it should be presented for analysis and with the help of which approaches and tools to work with them.

We also have the task of analyzing and studying existing forecasting methods, choosing the most relevant forecasting methods for us.

After developing their own method of forecasting and analyzing existing methods of forecasting, the task is to develop a methodology for distributing goods to stores. If there is an opportunity, we will use it to conduct an experiment on the effectiveness of our methodology in comparison with the method that is used in company B at the moment.

The next task is to develop a methodology for replenishing goods in stores during the season.

The next step is the development of a methodology for the transfer of goods between stores, given the legal restrictions in force in Russia.

4.0 Literature review

The review of literature is divided into two subchapters. The first subchapter will provide an overview of the literature on demand forecasting. The second subchapter contains a review of the literature on the topic of optimizing product placement.

4.1 Demand forecasting

In (Li and Lim 2018), the authors suggested using greedy heuristics to predict demand in fashion retailing. They work with a real case for one fashion retailer in Singapore. The essence of heuristics is to predict sales of a certain sku (stock keeping unit) without taking into account other products that are sold in the store. The demand forecast is calculated for each sku in each store for each day.

The complexity of the algorithm is O (n3), which makes it inapplicable to our case because of the large number of stores and sku. The Mean Absolute Percentage Error when using the heuristics proposed by the authors was 20% with a planning horizon of 15 days. With a time, horizon of 275 days, MAPE was 77.2%. This work did not take into account the possibility that the goods may not remain in the store on one of the days. The article stipulated that the entire range of stores is always available for purchase. In the real world, this is an impossible condition.

The last few years it has become a trend to use neural networks and artificial intelligence to forecast demand in retail. One of the landmark works in this direction is the study given in (Veiga et al. 2016). This paper shows the results of WNN (Wavelets Neural Network) and TS (Takagi-Sugeno Fuzzy System). The authors of the article use these neural networks to predict the demand in food retail. Three categories of products were selected: yogurt, milk and dairy dessert. This was done not only to simplify the presentation of the technical part of the article, but also because the prediction methods proposed in the article are computationally complex. Although they show minimal error with a short forecast horizon. As a result of the analysis of the article and the results presented in it, we came to the conclusion that these methods are difficult to apply in our case. This is due to the extremely difficult implementation of the neural network architecture and the operating time of the neural networks. In addition, the prediction error increases greatly with the increase in the forecast horizon.

An interesting case is considered in article by (Ma, Fildes, and Huang 2016). This is important because the methods that are considered in it make it possible to forecast demand during the validity of promotions. Just take into account how the demand for products of one category will affect the demand of another category. The advantages of this technique include the simplicity of the data that is needed for the model to work.

Number	Category	Number of SKUs	Mean units sold per week	Median units sold per week	Percentages of weeks concerning promotional activities		
					Price reductions	Displays	Features
1	Beer	98	12.80	7	0.30	0.27	0.13
2	Carbonated beverages	76	38.25	16	0.42	0.09	0.18
3	Coffee	46	5.90	5	0.34	0.02	0.10
4	Cold cereal	119	15.60	9	0.20	0.05	0.13
5	Frozen dinners	79	18.50	13	0.43	0.04	0.17
6	Frozen pizza	62	21.05	14	0.47	0.10	0.31
7	Frankfurters	21	22.95	10	0.35	0.08	0.16
8	Margarine/Butter	21	29.20	13	0.37	0.05	0.13
9	Mayonnaise	17	15.70	12	0.21	0.03	0.08
10	Milk	40	59.60	24	0.19	0.01	0.06
11	Peanut butter	16	14.30	10	0.22	0.01	0.07
12	Salty snacks	80	17.95	11	0.31	0.12	0.12
13	Soup	129	15.05	9	0.23	0.03	0.10
14	Spaghetti sauce	70	9.40	7	0.38	0.03	0.11
15	Yogurt	52	49.45	37	0.29	0.01	0.08

Table 4.1: Used data in (Ma, Fildes, and Huang 2016)

Data that is used to work the model is not difficult to obtain in company B. The second advantage was the clear algorithm scheme. To forecast demand, the Multistage LASSO process is used, presented in the diagram below.

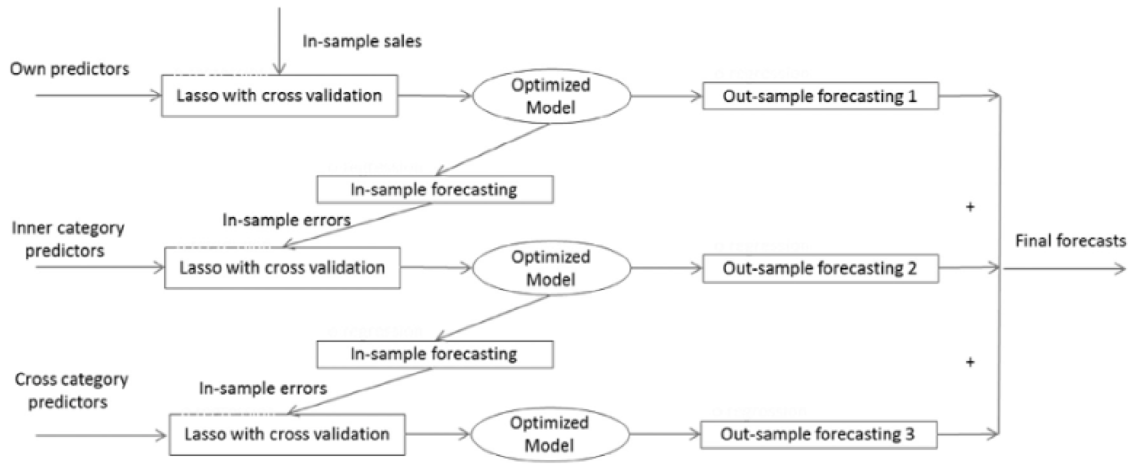


Figure 4.1: Model used in (Ma, Fildes, and Huang 2016)

However, a prediction error with a time horizon of more than 4 days is unacceptable for our task.

Forecasting true demand includes your entire supply chain. Everyone needs to share information that increases demand. This is a tough task. Most likely, you will have to develop trust covering your supply chain, that is, with other organizations. Once you do this, the operations for everyone will become smoother.

Effective demand forecasting is in fact part of integrated business planning or sales, inventory, planning operations. Inventory accuracy is one of three key performance indicators (KPIs) that you will use to measure your planning effectiveness. The outcome of your business depends on high prediction accuracy. Unfortunately, consumer demand is incredibly unstable, which can make forecasting difficult. In addition to historical sales data, you need to include:

- Market research;
- Consumer research;
- Resource and Capacity Planning;
- Possible economic, political or social factors.

Many businesses are trying to keep accurate historical data. A flexible inventory management system will help you keep numbers, as well as facilitate daily operations.

The purpose of demand forecasting is consumer demand forecasting. Demand forecasting models are grouped into two categories: qualitative and quantitative. Quantitative methods rely on data, and qualitative methods rely on expert opinion.

Methods of demand forecasting are divided into estimated and experimental. They are usually used when historical data is not available for your products. As in the case of new products. A frequently used evaluation method is based on the opinions of available experts. You will receive subjective evaluations from suppliers, marketing, sales and purchasing specialists.

There are quite a few different qualitative methods, from basic to complex. All of them can be very valuable in developing an accurate forecast.

Causal methods use a number of internal and external factors, such as price, legislation, and even weather, as well as another qualitative method to accurately draw a picture of future inventory requirements.

Projective forecasting methods are based on mathematical formulas that can provide an easier answer to your forecasting needs. These methods use past data to estimate future demand. Often software is used to forecast demand.

Life cycle prediction can be useful for technology products or any inventory with a high probability of obsolescence. This type of forecasting takes into account the fact that demand changes during the product life cycle. Accurate forecasting of the life cycle of your products can help you save even less and take into account future releases of new versions of your product.

When considering trends in inventory, it is important to consider seasonality. Seasonality is the change in demand on a periodic basis. Trends - this change in demand, which occurs for various reasons, for example, when changing consumer behavior. If the product is sold at a constant rate for many years, and then demand increases, this is a trend.

Companies have always tried to get ahead of sales trends.

The correlation between accurate forecast and cost reduction in the supply chain has long been clear. Progress in technologies such as artificial intelligence and data analysis has allowed companies to see accuracy rates like never before.

The potential savings in the supply chain exceed the cost of tools such as software for demand forecasting. This means that you cannot afford to ignore this technological development of the supply chain. As large and small businesses are struggling for forecast accuracy, expect to see all kinds of investment in this technology in the coming months and years. Higher forecast accuracy does not depend only on the methods that you use quarterly or annually to plan your reserves. What you and your employees do every day in the warehouse, receiving inventory and completing orders, is more important. Every day is based on the foregoing. If your daily numbers do not match, you will not be able to disentangle it and effectively create accurate prediction models. Focus on the accuracy and precision of the entire supply chain, and you will get the advantage of more accurate demand forecasting (Gasparian et al. 2018).

Marketing professionals consider forecasting to be an important part of their work.

We discuss methods for forecasting demand. People often use the terms “demand” and “sales” in the same meaning. It is reasonable to do this because these two terms are equal when sales are not limited to the offer.

Sometimes it is advisable to forecast demand directly. When direct prediction is not feasible or where uncertainty is and changes are expected to be significant, marketing managers may have to predict market size or Product Category. In addition, they will have to predict the actions and reactions of key decision makers, such as competitors, suppliers, distributors, employees, governments and themselves - especially when questions are related. These actions can help predict market share. The resulting predictions allow calculate the demand forecast.

Methods Based on Judgment

There are methods based on judgment. In this case, you need to ask the experts what will happen. This is a good method if the experts are unbiased, big changes are unlikely, the relations are well understood by the experts, the experts have accurate information, the experts receive accurate and generalized reviews about their predictions.

Forecasting markets, also known as betting markets, information markets and futures markets, have a long history. Some commercial organizations provide Internet markets and software that allows participants to bid on trade contracts.

The Delphi method was designed to help gather the knowledge of various experts, while avoiding the shortcomings of traditional group meetings. The latter are a waste of time. The Delphi Forecast is the median or form of the final expert forecasts.

The outcomes of similar situations from the past can help the marketer predict the outcome of a new situation. People often use analogies to make predictions, but they do not do it in a structured way. The method of structured analogies uses a formal process to overcome the biased and inefficient use of information from similar situations.

Game theory has been advertised in textbooks and research as a way to get more accurate predictions in situations involving negotiations or other conflicts. Despite the tremendous research efforts, there are no studies that directly test the ability of game theory to predict.

The idea of judicial decomposition, is to divide the problem of forecasting into parts that are easier to predict than the whole. Details you can predict separately using methods that correspond to each part. Parts are combined to produce a forecast. One approach is to break the problem down into components. For example, to predict sales for a brand, you can predict sales by industry, market share, and selling price per unit. Then collect the predictions by multiplying the components together.

Subjective evaluation turns subjective judgments into structured procedures. Experts ask. They are then asked to make predictions for various cases, which may be real or hypothetical. For example, it may be new products. The data is then converted into a regression equation.

As the name implies, expert systems are structured rule representations that experts use to predict or diagnose. Rules are often created from opinions, according to which weather forecasters talk about what they do when making forecasts. Expert opinion, collaborative analysis and bootstrapping can also help in developing expert systems.

Simulated interaction is a form of role-playing game for predicting the decisions of people who interact with others. This is especially useful when the situation is related to a conflict. To use simulated interaction, the administrator prepares a description of the target situation, describes the main roles of the main characters and provides a list of possible solutions. Role players learn the situation. Then they improvise realistic interactions with other players until they decide. The decisions of the role participants are used to make a forecast.

In surveys, people are asked about their intentions, how they intend to behave in certain situations. Similarly, in a survey of expectations, people are asked how they expect to behave. Expectations are different from intentions, because people understand that unexpected things happen.

By interviewing consumers about their preferences for alternative product design in a structured way, it can be concluded how different functions will affect demand. Thus, a potential client is forced to compromise between the various functions, choosing one of each pair of offers in such a way that it represents how they will choose in the market. The obtained data can be analyzed by comparing the selection with the characteristics of the product. This method is called “conjoint analysis,” since respondents are jointly reviewing product features.

Methods requiring quantitative data

Extrapolation methods use historical data about what they want to predict. Exponential smoothing is the most popular and cost-effective statistical extrapolation method. It implements the principle that the latest data should be weighed more intensively and “smooth out” cyclical fluctuations to predict trends. To use exponential smoothing to extrapolate, the administrator must first clear and cancel the seasonality of the data and select reasonable smoothing factors. The administrator then calculates the mean and trend based on the data and uses them to get a forecast.

Retail scanning technology provides reliable and up-to-date data to extrapolate sales of existing products. As a result, the forecast accuracy should improve, especially because the error in assessing the current situation is reduced. Ignorance of what you start with is often the main source of errors in predicting future values.

Experts can identify situations that are similar to this situation. They can be used to extrapolate the results of a situation. For forecasting using quantitative analogies, situations are identified that are similar to the target situation and for which data are available.

Rule based prediction (RBF) is a type of expert system that allows you to integrate the knowledge of managers in this field with time series data in a structured and inexpensive way.

Neural networks are computer-intensive methods that use decision-making processes similar to those of the human brain. They have the ability to learn when changing patterns and updating their parameter estimates. A lot of data is required to evaluate neural network models and reduce the risk of data iteration.

Data mining uses sophisticated statistical analysis to identify relationships. There are ambitious statements and great research efforts, but we do not know which mining methods are useful for forecasting. Most of this was due to the fact that only a few studies used the

correct data evaluation scheme.
Causal models are based on prior knowledge and theory. These models allow you to explore the impact of marketing activities, such as price changes, as well as key aspects of the market, thus providing information for contingency planning.

Segmentation involves splitting the problem into independent parts, using data for each part to make a forecast, and then combining the par (Armstrong and Green 2005).

4.2 Assortment planning

Assortment planning is difficult problem for company. There is a scientific approach such as product assortment planning (PAP). When companies planning the assortment they determine variety of merchandise (number of categories), the depth of merchandise (number of stock-keeping units within a categories) and service level or the amount of inventory to allocate to each stock-keeping unit (SKU) (number of individual items of a particular stock-keeping unit). These indicators should be in balance. It is also necessary to consider amount of money that can invest in inventory and physical space. However, there are limitations in these categories.

Considering all these factors, companies hope to satisfy customers. They want to provide the right merchandise in the right store at the right time.

Product assortment planning consider consumer perceptions and preferences, their supply-side constrained and the external environmental factors, such as economic conditions and competitor's strategies. Customers benefit from these costly investments by finding and buying what they want and it brings company revenue.

Retailers must identify how many and which stock-keeping units, number of distinct brands or product types to offer, the number of variants of each brands, number of units of each variants of each brand or product or type to carry in inventory.

Consumers prefer flexibility because buying cases are often separated in time from consumption and often change their preferences. Another factor related to the size of the assortment is associated with the assumption about the stability of consumer preferences over time. In many cases, there is no single most preferred element, since the most preferred element is created at the time of selection, depending on the circumstances of the decision. Consumers may not be able to select a favorite element because they are trying to maximize global utility through a sequence of choices rather than local utility during a choice and can negatively perceive a large range if they cause frustration or a feeling of depression. In order to complicate the seller's task when assembling an assortment, the actual variety of the assortment may not match the intended diversity that the consumer faces. When a consumer finds an acceptable product in one retail store, he may still be unsure whether there are similar products in other stores, and will want to go to another store to explore other alternatives hoping to find the best product, although this includes the cost of search.

Even if the retailer is able to determine the optimal assortment for transportation, he may be unjustified to store such an assortment. Therefore, stock-out situations are key factors for retailers who need to predict how consumers will react to these events.

Despite progress in understanding how consumers influence PAP solutions, much remains unknown. PAP problems from the point of view of the consumer are associated with the complex problem of working with a range of products that are attractive but difficult to choose.

The most obvious restriction on the size and composition of the range is the space available in the store. Excluding costs, the size of an ideal store is equal to the sum of all ideal assortment categories. Since the expansion of the physical size of stores is very expensive and often impossible, the total floor area remains substantially fixed. The variety of categories and the depth of the SKU within the categories that determine the type of seller and its position in the market.

In addition to consumer responses and retailer constraints, external factors affecting the organization make important contributions to the retailer's PAP solutions. The same product category can be purchased in stores that use very different retail formats.

Macroeconomic and environmental trends also influence the diversity and depth of PAP solutions.

Retailers must adapt the range and depth of their range to the changing tastes and profile of consumers. Most retailers segment their markets or shopping areas primarily based on customer-specific, market-specific or store-specific factors, and then modify the variety and depth of their product range based on these factors (Mantrala et al. 2009).

Visual merchandising is the practice of displaying goods to help customers find what they want and encourage large purchases. Visual merchandising is much more than just installing shelves. It helps you find the optimal store layout and determine exactly where to place the goods.

Visual merchandising is the process of designing your layouts, racks and products to maximize sales and give customers an exciting shopping experience. Visual merchandising as a science explains how shoppers behave in a store, and thus, how your product placement can provide them with an optimal shopping experience. Make sure that when customers go to your store, they know that they are in the right place. Influencing the shopping habits of buyers - the correct layout of the shelf and the placement of products lead buyers in all areas of your store where income items live. With these two things in mind, you can come up with a floor plan and a merchandising plan for the relevant product to maximize the full sales potential of your store.

A good floor layout ensures that you can place all your products in the places that give the most change for sale. Buyers should notice how easy it is to move and shop in a store without taking their eyes off the store shelves.

Since you are planning where you want your goods to appear in your store, it is important to know the type of store. The store layout refers to how your racks are placed all over your store, so your goods are easily bought and the flow of customers goes to products that bring you the most profit. There are three basic layouts and designs that you can use, and it is important to know a little about each one, as they will determine how you plan your product placement. It is important to spend a lot of time planning your goods placement, because when everything is done right, your customers will always take extra things on the way to get what they really need. They will also find it easier to find the products they want to buy. Today's consumers are bombarded with goods and offers. The question is how to design and deliver the offers that stand out. Understanding this can help retailers decide how, when and where to display goods. Manufacturers also recognize the importance of ensuring that consumers pay attention to their goods and offers, to look for ways to make their goods stand out among competitors on the shelf (Reed 2012).

5.0 Methodology

Systematizing the information obtained in the process of studying the literature, negotiations with companies that are engaged in solving problems, announced in this thesis, interviews with employees of company B, it was possible to systematize the methods and approaches used to predict demand in shoe retail:

- Econometric approaches;
- Methods of machine learning;
- Deep learning, using artificial intelligence.

Machine learning methods include probabilistic and statistical approaches such as regressions (linear, nonlinear, autoregression), the support vector method and its variations, random forest regression, model ensemble combining, and so on.

Deep learning includes prediction using recurrent and convolutional neural networks of various architectures.

Having conducted a comprehensive analysis, it is clear that in the modern world with the current level of technology development, econometric approaches are outdated. They were very popular in the 80-90s, but now the mood in business circles and the scientific literature is that machine learning and artificial intelligence have a future of forecasting.

In-depth training methods require a huge amount of data and computational resources. This direction for research seems very promising. One of the indicators of the effectiveness of deep learning methods is the success of companies that use them in demand forecasting. We are talking about companies such as Amazon and Zalando. Last but not least, their explosive growth is due to lower logistics costs due to effective demand forecasting.

Unfortunately, within the framework of the thesis, we cannot choose methods associated with deep learning. In January 2019, company B purchased the NVIDIA DGX-Workstation, which is focused on working with deep learning algorithms. Delivery and tuning of the station to Belarus was expected at the end of March 2019. However, legal difficulties have made adjustments. The station arrived at company B on 4 May. Naturally, the research could not wait so long, therefore, in this thesis work, machine learning methods are used to predict. Since forecasting is long-term, you cannot rely on methods that use weather data to forecast demand. These methods can work well with a time horizon of about a week, then the quality of the forecast deteriorates dramatically. In world practice for quite a long time there is a trend to use time series to forecast demand with a time horizon of up to a year.

There are many approaches to predict time series, such as ARIMA, ARCH, regression models, neural networks. Implementation from scratch of the above methods (except for regression models that make a prediction at the level of “what happened yesterday, it will be tomorrow”) requires a team of highly qualified mathematicians and at least six months.

Comparative analysis of forecasting methods shows that the mathematical apparatus used in the model implemented in this thesis is more accurate in predicting almost the entire time interval than other most popular forecasting methods. More information about the method will be described in Chapter 7.

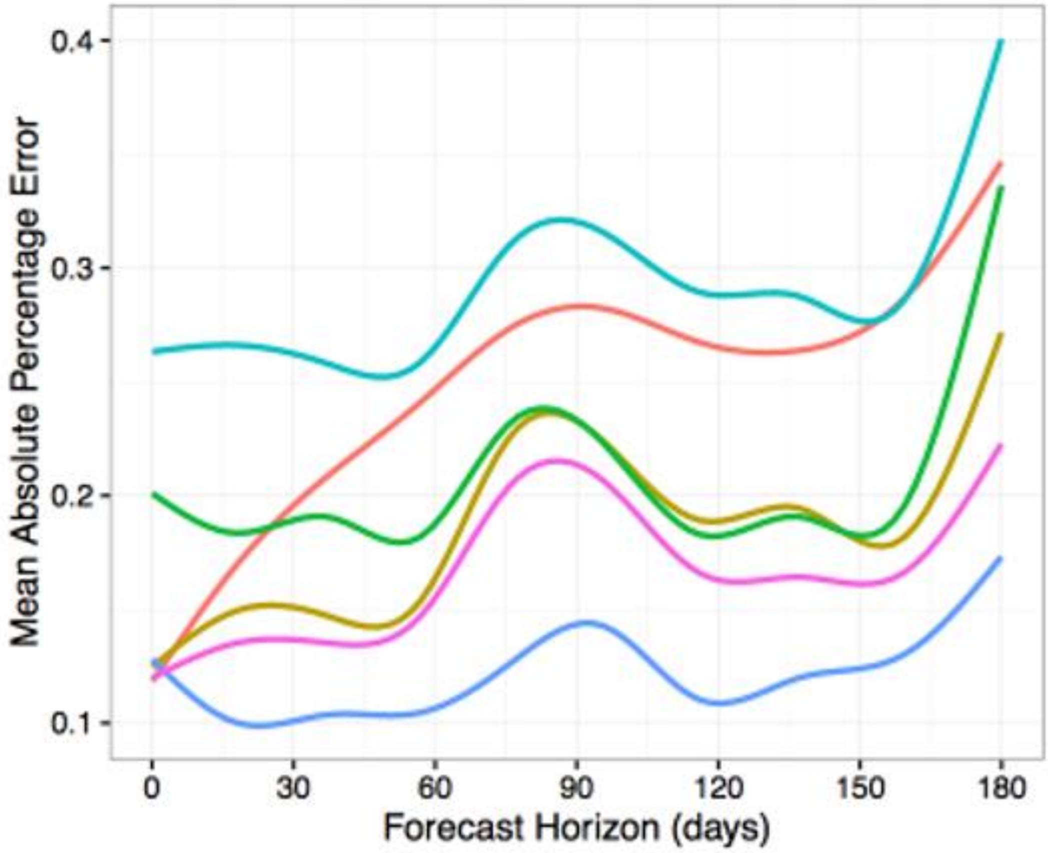


Figure 5.1: Comparison of the accuracy of various prediction algorithms

Having conducted an extensive study, including in closed sources, no retail companies were found that would use the method used in this thesis to forecast demand.

To solve the problems of initial placement, replenishment and transfer, the main criterion for choosing a method was to reduce the human influence factor and to make the processes systematic and orderly.

6.0 Data mining and data transformation

To develop methodologies for solving the problems posed, a large amount of data is needed. Different techniques may be required for different methods. When using random forest regression, weather data may be needed in the regions where the stores are located to increase the accuracy of demand forecast. On the other hand, for forecasting methods based on time series, in the simplest case, it is enough for each store to know the daily sales and the daily supply of goods in the store.

We proceed from the concept that there is not much data, therefore we tried to get as much data as possible. Software from SAP and POS (point-of-sale) stores a huge amount of data on the company's activities, so before we start data mining, we compiled a list of the data that we need to obtain first. Here is a list of data that we wanted to get:

1. Directory of goods
    - Product ID
    - Name of product
    - Description of product groups (gender, season, stylistic line, etc.)
    - Date of entry of goods
    - A set of attributes of the product (color, size, style, etc.)
2. Directory of shops
    - Name
    - Store Address
    - Store ID
    - Region
    - Store characteristics
        - o Location (street / shopping center)
        - o Area
        - o Open / closed showcase
3. Sales history (by day for 3 years)
    - Product Identifier (SKU)
    - Store ID
    - Date
    - Sales, units
4. History of balances (by day for 3 years)
    - Product Identifier (SKU)

- Store ID
- Date
- The rest of the goods at the end of the day

5. The history of changes in prices for goods
   - Product Identifier (SKU)
   - Store ID
   - Starting date of the price
   - The price of the product
6. Promotional History
   - Product Identifier (SKU)
   - Store ID
   - Type of promotion / mechanics
   - Amount of discount
   - Name of promotion
   - Starting date of the promotion.
   - End date of promotion

Data with a list of products and their characteristics could be obtained through SAP BI. Uploading data was done to Excel. The result of the work of unloading the directory of goods was the following table:

	A	B	C	D	E	F	G	H	I
1	vendor_code	size	name	sex	season	style_line	date_start	lining_material	heel_height
2	1638267Р	37	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
3	1638267Р	38	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
4	1638267Р	42	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
5	1638267Р	36	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
6	1638267Р	41	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
7	1638267Р	35	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
8	1638267Р	40	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
9	1638267Р	39	Туфли Летние	Женские	Летняя	модельная классика	2016-01-01	Кожа Натуральн	Н/К
10	1638270	38	Туфли	Женские	Всесезонная	модельная классика	2015-01-01	Кожа Натуральн	В/К
11	1638270	40	Туфли	Женские	Всесезонная	модельная классика	2015-01-01	Кожа Натуральн	В/К
12	1638270	35	Туфли	Женские	Всесезонная	модельная классика	2015-01-01	Кожа Натуральн	В/К
13	1638270	39	Туфли	Женские	Всесезонная	модельная классика	2015-01-01	Кожа Натуральн	В/К
14	1638270	41	Туфли	Женские	Всесезонная	модельная классика	2015-01-01	Кожа Натуральн	В/К
15	1638270	36	Туфли	Женские	Всесезонная	модельная классика	2015-01-01	Кожа Натуральн	В/К
16	1638270	37	Туфли	Женские	Всесезонная	модельная классика	2015-01-01	Кожа Натуральн	В/К
17	1638270/О	35	Туфли Летние	Женские	Летняя	модельная классика	2015-01-01	Кожа Натуральн	В/К
18	1638270/О	41	Туфли Летние	Женские	Летняя	модельная классика	2015-01-01	Кожа Натуральн	В/К
19	1638270/О	37	Туфли Летние	Женские	Летняя	модельная классика	2015-01-01	Кожа Натуральн	В/К
20	1638270Р	41	Туфли	Женские	Всесезонная			Кожа Натуральн	В/К
21	1638270Р	36	Туфли	Женские	Всесезонная			Кожа Натуральн	В/К
22	1638270Р	37	Туфли	Женские	Всесезонная			Кожа Натуральн	В/К
23	1638270Р	35	Туфли	Женские	Всесезонная			Кожа Натуральн	В/К
24	1638270Р	38	Туфли	Женские	Всесезонная			Кожа Натуральн	В/К
25	1638270Р	40	Туфли	Женские	Всесезонная			Кожа Натуральн	В/К
26	1638270Р	39	Туфли	Женские	Всесезонная			Кожа Натуральн	В/К
27	1638271	40	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
28	1638271	36	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
29	1638271	37	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
30	1638271	35	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
31	1638271	41	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
32	1638271	38	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
33	1638271	39	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
34	1638271Р	41	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
35	1638271Р	39	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К
36	1638271Р	35	Туфли	Женские	Всесезонная	модельная классика	2016-01-01	Кожа Натуральн	В/К

Table 6.1: Item Characteristics Data

The table consists of 9 columns and 140062 rows. Product ID is represented by a combination of the first two columns (vendor_code) and size. Since in company B the same shoe model of different colors is considered to be different article numbers, then by adding size to the article, we get SKU. The sex attribute shows whether the SKU is women's or men's shoes.

Season shows the seasonality of shoes. Style_line means to which stylistic line the shoe belongs. This division is rather conditional, and often there is no stylistic line for shoes. Date_start - the date when the SKU was taken into development. Lining_material indicates the material from which the shoes are made. Heel_height characterizes the height of the heel of shoes. Heel height is not indicated in centimeters, but in categories: low heel, medium heel, high heel.

The next step was to get the data for the stores. The identifier, address, region and location (street / shopping center) of the shops were also obtained from SAP BI and uploaded to Excel. However, in SAP BI there was no data on the area of stores. These data, although not complete, were obtained thanks to the employees of the marketing and advertising department. Then we manually attached to our Excel file data on the total area of the store (total_area) and the area of the sales area (sale_area) in square meters. As a result, the following table was obtained:

	A	B	C	D	E	F
1	id_store	address	region	location	total_area	sale_area
2	24110	г. Красногорск, ул. Знаменская, д.5	ЦЕНТР	ТЦ		
3	24109	рп. Ржавки, мкрн. 2, стр. 20	ЦЕНТР	ТЦ		
4	24107	г. Одинцово, ул. Неделина, д.9а	ЦЕНТР	стрит		
5	24104	г. Рославль, ул. Пролетарская, 82	ЦЕНТР	ТЦ	137	60
6	24101	г. Мытищи, пр-т Шараповский, вл.2, стр.3, п.1	ЦЕНТР	ТЦ	121	93
7	26104	г. Москва, ул. Красного Маяка, д.2Б	МОСКВА	ТЦ		
8	26103	г. Москва, пр-д Чечерский, д.51	МОСКВА	ТЦ		
9	26102	г. Москва, ул. Большая Тульская, д.13	МОСКВА	ТЦ		
10	02110	г. Воронеж, пр-т Ленинский, д.174П	ЦЕНТР	ТЦ	93	59
11	02124	г. Воронеж, б-р. Победы, д. 23Б	ЦЕНТР	ТЦ	193	150
12	02122	г. Липецк, ул. Титова, д.10	ЦЕНТР	ТЦ	261	145
13	02120	г. Воронеж, пр-т Московский, д.129/1	ЦЕНТР	ТЦ	98	65
14	02119	г. Липецк, ул. А.Г.Стаханова, д.36	ЦЕНТР	ТЦ	192	149
15	02113	г. Воронеж, пр-т Ленинский, д.116а	ЦЕНТР	ТЦ	140	56
16	02111	г. Липецк, пр-т Мира, д.1	ЦЕНТР	стрит	87	40
17	02109	г. Воронеж, ул. Плехановская, д.43	ЦЕНТР	стрит	124	79
18	02108	г. Воронеж, ул. Маршака, д.7Б	ЦЕНТР	стрит	109	47
19	02107	г. Воронеж, пр-т Московский, д.78	ЦЕНТР	стрит	153	79
20	02102	г. Липецк, пл. Заводская, д.2	ЦЕНТР	стрит	127	66
21	02101	г. Липецк, ул. Водопьянова, д.70	ЦЕНТР	стрит	525	96
22	02103	г. Липецк, ул. Советская, д.77	ЦЕНТР	стрит	88	41
23	02115	г. Воронеж, ул. Кирова, д.9	ЦЕНТР	стрит	122	85
24	06146	г. Санкт-Петербург, ул. Савушкина, д.141, лит.А	СЕВЕРО-ЗАПАД	ТЦ		
25	06147	г. Санкт-Петербург, пр-т Гражданский, д.41 литер А	СЕВЕРО-ЗАПАД	ТЦ		
26	06140	г. Санкт-Петербург, пл. Комендантская, д.1, лит.А	СЕВЕРО-ЗАПАД	ТЦ	253	
27	06133	г. Санкт-Петербург, ш. Пулковское, д.25	СЕВЕРО-ЗАПАД	ТЦ	187	119
28	06144	г. Санкт-Петербург, пр-т Заневский, д.71, корп.2, лит.А	СЕВЕРО-ЗАПАД	ТЦ		
29	06141	г. Санкт-Петербург, Брантовская дорога, д.3, пом. 2.49	СЕВЕРО-ЗАПАД	ТЦ		
30	06145	г. Санкт-Петербург, пр-т Стачек, д. 99, лит. А	СЕВЕРО-ЗАПАД	ТЦ		
31	06130	г. Санкт-Петербург, ул. Пражская, д.48/50	СЕВЕРО-ЗАПАД	ТЦ	153	109
32	06137	г. Санкт-Петербург, пр-т Большевиков, д.18, лит.А	СЕВЕРО-ЗАПАД	ТЦ	102	70
33	06120	г. Санкт-Петербург, пр-т Художников, д.14, лит.А	СЕВЕРО-ЗАПАД	стрит	104	64
34	06101	г. Санкт-Петербург, пр-т Заневский, д.10	СЕВЕРО-ЗАПАД	стрит	168	89

Table 6.2: Store Characteristics Data

The third step in the date-mining process was the process of obtaining sales data. Sales history is in the PIC system (point-of-sale) and is stored in the Oracle database. Therefore, to obtain this data, we wrote SQL queries to the database. It is possible to upload data from Excel database to Excel file, but more than two million lines cannot be imported to Excel file. It was decided to upload the sales history for each store to a separate file. Total sales history is a 392 Excel file of the following form:

	A	B	C	D	E
1	vendor_code	size	id_store	date	count
2	531030/1	39	23010	2016-01-03	1
3	531030/1	38	23010	2016-01-03	1
4	432050	37	23010	2016-01-04	1
5	1638040/1	39	23010	2016-01-05	1
6	430066	40	23010	2016-01-05	1
7	234197	38	23010	2016-01-05	1
8	314176	42	23010	2016-01-05	1
9	430330	38	23010	2016-01-05	1
10	413040	43	23010	2016-01-05	1
11	214106	42	23010	2016-01-06	1
12	430201	40	23010	2016-01-06	1
13	430380	41	23010	2016-01-06	1
14	534001	38	23010	2016-01-09	1
15	435106	40	23010	2016-01-09	1
16	413010	42	23010	2016-01-10	1
17	421000	44	23010	2016-01-10	2
18	314030	42	23010	2016-01-11	1

Table 6.3: Sales history data

However, another problem arose: since not only shoes are sold in stores, but also related products such as socks, stockings, tights, shoe care products, etc. When uploading data from the database, they also fell into the sales data. We only need to forecast demand for shoes, so we needed to get rid of the "noise in the data." Benefit related products are of zero size and easy to remove using the programming language Python.

```
for file in files:
    data = pd.read_csv(file, encoding='utf-16')
    data = data.drop(data[data.SIZE == 0.0].index)
```

Figure 6.1: Fragment of program code for converting Excel files into csv files

It is worth considering that there are cases when the buyer returns the purchased goods. We track returns in sales history as a regular sale with the count parameter set to -1. Of course, this method of tracking returns has one drawback: usually returns occur on a different day than buying shoes. Therefore, it happens that on one particular day 17 pairs of shoes were bought, 2 of which were returned to the store the next day. We still consider the number of sales on that day equal to 17, but the next day we consider the number of sales made minus the returns made on that day.

Also at the stage of unloading sales history, we faced the problem of Big data. The data weighs several gigabytes, and downloading from the database took several days due to the load on the system by other processes. Therefore, it was decided to write the SQL queries as simple as possible and to transform and clean the data without accessing the database using the Python programming language and the Pandas library.

The next stage was the unloading of the history of the remnants of the goods in the shops for every day. The look of the tables is the same as in the files containing information about the sales history. However, even the history of residuals for 3 years for one store does not fit into one Excel file. The history of balances for one store takes from 3 to 5 Excel files. Total data on the history of the remnants occupy 48 gigabytes. For the convenience of working with leftover data, it was decided to use the csv extension files. This file format has no limit on the number of lines. Transformation of Excel files to csv format was done using Python.

```
def stocks(path):
    """Convert .xlsx files to .csv format"""

    files = glob.glob(path)
    i = 1

    for file in files:
        data_xls = pd.read_excel(file, 'Sheet0', index_col=None)
        data_xls.to_csv('d:\Anaconda3\Scripts\Future_prediction\Stock\история остатков_' + str(i) + '.csv')
        i = i + 1

stocks('d:\Work\Data_SAS\Stock\*.xlsx')
```

Figure 6.2: Fragment of program code for converting Excel files into csv files

The same situation exists with the history of changes in prices for goods. The solution to the problem was the same as in the previous paragraph. The view of the table with the data on price changes for goods in stores is as follows. Price for the markets of Russia and Belarus is always in Russian rubles.

	A	B	C	D	E
1	vendor_code	size	id_store	start_date	price
2	437438/1	37	24104	2015-09-09	3997
3	437438/1	37	24101	2015-09-09	3997
4	437438/1	37	23022	2015-09-09	3997
5	437438/1	37	23021	2015-09-09	3997
6	437438/1	37	17124	2015-09-09	3997
7	437438/1	37	17122	2015-09-09	3997
8	437438/1	37	17121	2015-09-09	3997
9	437438/1	37	17118	2015-09-09	3997
10	437438/1	37	17116	2015-09-09	3997
11	437438/1	37	06147	2015-09-09	3997
12	437438/1	37	06146	2015-09-09	3997
13	437438/1	37	06145	2015-09-09	3997
14	437438/1	37	06144	2015-09-09	3997
15	437438/1	37	06141	2015-09-09	3997
16	437438/1	37	06140	2015-09-09	3997
17	437438/1	37	02124	2015-09-09	3997
18	437438/1	37	02122	2015-09-09	3997

Table 6.4: Data on prices of goods in the stores

A similar situation exists with data on promotions. In the csv format, promotion data takes up 120 gigabytes of memory. Unloading was slow, and errors were found in the database: in one and the same SKU in 2016, two actions acted simultaneously and it was unclear which of them actually worked. However, the most interesting thing was that this SKU in 2016 was in the store, which began to work only in 2018! An error report was sent to the software development department, but later it was decided not to use the promo history to forecast demand.

Many people advised us to use weather data in forecasting demand, so we did not try. However, the problem was how to assemble it for more than 200 cities in a short time? There were two possible solutions to the problem:

- Use paid services that contain historical data about the weather in different regions;
- Develop your own software that can download weather data from weather forecast sites. Such sites store weather data in the regions for a long time.

We checked both ways to solve the problem. In the first case, we collaborated with company A to obtain historical weather data. The cost of their services was $ 800 for one-year weather history for one city. We selected the 10 smallest cities in Russia where there are stores of company B and received incomplete data from company A. Of the ten cities selected, only four cities had weather data. It was decided to refuse the services of company A. Therefore, it was decided to write a program for web scraping of the weather forecast site. Python and the BeatiufulSoup library were used. The use of weather data is necessary as one of the attributes for machine learning algorithms, for example, for random forest regression.

For the problem of distribution of goods to the shops of critical information is the volume of the store. Knowing the maximum capacity of the store, you can forget about such problems as store overload and penalties associated with this. Direct data on the capacity of stores in the company was not. The option to go to each store with a tape measure and measure its volume is extremely time-consuming. Our technique is as follows:

- Determine the size of the boxes in which shoes are delivered to stores. This information can be found in SAP BI;
- For each day, sum up the amount of residues in the store. Knowing the size of his box for each article, calculate the total volume in cubic meters;
- Find the day for each store when the volume occupied by the shoes was maximum;
- Interview senior store managers to clarify the results and get an expert opinion.

In the first step, we obtained the types and sizes of boxes that are now used to store shoes. Dimensions are in millimetres.

Name of box	Length	Width	Height	Cover height
D15	294	120	100	35
D17	291	158	98	35

D30	320	295	108	45
D34	365	310	108	45
H17	325	182	110	40
H18	350	250	133	45
H-22	349	220	131	45
H18-H	349	189	131	45
H18B	365	275	135	45
H24	329	230	113	35
S45	445	300	108	50
S55	550	305	112	50
S49	500	300	110	50
S56	650	310	120	45

Table 6.5: Shoe box specifications

After that, a table was created with a melon about the workload of stores. There are three large columns in the table:

- The number of unique articles;
- Total stock of shoes;
- Volume of total stock in cubic decimeters.

In each of the three large columns, there are 12 columns defining the value of the large columns for each month in the past 12 months. Each store is represented by five lines. The first four lines show the value of the columns for all-season, winter, summer and warm shoes for a specific month. The fifth line is the sum of the previous four cells. After each large column there is a column called “Maximum Value”. Thus, for each store, we learned the maximum number of unique articles that were in the store at one time, the maximum value of the stock of shoes in the store and the maximum value of the occupied volume in cubic decimeters.

After that, the resulting table was transferred to senior store managers to get their opinion on maximum volume figures. Back we got a table with the maximum allowable amount of shoes in stores during the summer and winter season. Column F shows the maximum capacity of the store during the summer season, column G shows the maximum capacity during the winter season. The columns highlighted in yellow show the month in which the stores were reset. In the future, we focused on the figures obtained for the maximum capacity of stores, expressed in pairs of shoes when solving problems of distribution, replenishment and transfer.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
1						IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	IC2: Кол-в	Макс. кол-во упак.ед.	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	IC2: Общи	Макс. Запас кор. (мес)	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	IC2: Объём	Макс. Объём в дм3
2	Предприят	Магазин	Город	Климатич	Мат.:Сезон	ЯНВ 2018	ФЕВ 2018	МАР 2018	АПР 2018	МАЙ 2018	ИЮН 2018	ИЮЛ 2018	АВГ 2018	СЕН 2018	ОКТ 2018	НОЯ 2018	ДЕК 2018	ЯНВ 2019		ЯНВ 2018	ФЕВ 2018	МАР 2018	АПР 2018	МАЙ 2018	ИЮН 2018	ИЮЛ 2018	АВГ 2018	СЕН 2018	ОКТ 2018	НОЯ 2018	ДЕК 2018	ЯНВ 2019		ЯНВ 2018	ФЕВ 2018	МАР 2018	АПР 2018	МАЙ 2018	ИЮН 2018	ИЮЛ 2018	АВГ 2018	СЕН 2018	ОКТ 2018	НОЯ 2018	ДЕК 2018	ЯНВ 2019	
3	2708	07108 Альметьевск, ул. Ле	Альметье	Центр	Всесезонна	536	479	558	614	657	678	580	606	612	530	516	506	505	678	1 588	1 321	1 684	1 820	1 872	1 918	1 721	1 749	1 738	1 469	1 418	1 378	1 373	1 918	12 892	9 905	12 439	13 993	14 232	14 587	12 938	13 330	13 162	11 313	11 035	10 795	10 750	14 587
4	2708	07108 Альметьевск, ул. Ле	Альметье	Центр	Зимняя	257	239	52	11	11	11	11	10	128	260	304	336	335	336	674	584	92	12	11	11	11	10	500	1 085	1 114	1 168	1 145	1 168	9 398	8 054	1 262	576	169	169	169	157	6 857	14 221	14 405	15 105	14 823	15 105
5	2708	07108 Альметьевск, ул. Ле	Альметье	Центр	Летняя	27	27	295	440	504	521	452	438	194	17	17	16	16	521	36	36	1 019	1 534	1 610	1 522	1 185	1 074	410	18	18	17	17	1 610	244	244	6 364	9 387	10 146	9 554	7 401	6 722	2 591	116	114	110	110	10 146
6	2708	07108 Альметьевск, ул. Ле	Альметье	Центр	Утепленна	266	258	271	254	114	113	113	165	225	239	243	240	240	271	819	794	817	712	338	337	336	583	747	748	737	728	724	819	10 399	10 184	10 567	9 359	4 327	4 316	4 305	6 900	8 531	8 420	8 268	8 167	8 129	10 567
7	2708	07108 Альметьевск, ул. Ле	Альметье	Центр	Всего	1 086	1 003	1 176	1 319	1 286	1 323	1 156	1 219	1 159	1 046	1 080	1 098	1 096	1 323	3 117	2 735	3 612	4 078	3 831	3 788	3 253	3 416	3 395	3 320	3 287	3 291	3 259	4 078	32 932	28 386	30 632	33 314	28 874	28 626	24 813	27 108	31 140	34 070	33 823	34 177	33 812	34 177
8	2425	04125 Арзамас, ул. Калини	Арзамас	Центр	Всесезонна	547	514	536	589	724	694	683	683	694	375	361	366	366	724	1 714	1 590	1 642	1 780	2 229	2 073	2 056	1 979	1 999	1 243	1 125	1 110	1 104	2 229	13 419	11 356	11 749	13 044	17 281	16 209	16 228	15 610	15 039	9 648	8 830	8 677	8 626	17 281
9	2425	04125 Арзамас, ул. Калини	Арзамас	Центр	Зимняя	192	176	113	39	38	32	32	32	126	255	307	323	320	323	552	481	259	47	46	39	39	39	536	900	1 004	980	944	1 004	7 970	6 799	3 630	1 005	708	602	602	602	7 185	11 371	12 701	12 375	11 872	12 701
10	2425	04125 Арзамас, ул. Калини	Арзамас	Центр	Летняя	48	54	132	264	427	403	363	336	173	44	44	37	37	427	81	122	528	1 033	1 486	1 254	1 027	901	404	59	59	52	52	1 486	471	733	3 063	6 106	9 829	8 425	7 065	6 260	3 096	339	335	294	294	9 829
11	2425	04125 Арзамас, ул. Калини	Арзамас	Центр	Утепленна	283	329	328	238	135	134	134	192	254	274	273	269	267	329	877	1 014	965	604	282	281	281	553	779	777	745	726	722	1 014	11 773	13 025	12 447	8 022	3 460	3 451	3 451	6 712	8 934	8 789	8 396	8 171	8 123	13 025
12	2425	04125 Арзамас, ул. Калини	Арзамас	Центр	Всего	1 070	1 073	1 109	1 130	1 324	1 263	1 212	1 243	1 247	948	985	995	990	1 324	3 224	3 207	3 394	3 464	4 043	3 647	3 403	3 472	3 718	2 979	2 933	2 868	2 822	4 043	33 633	31 914	30 889	28 177	31 279	28 688	27 346	29 184	34 254	30 147	30 261	29 517	28 915	34 254
13	3214	12114 Армавир, ул. Мира,	Армавир	Юг	Всесезонна	535	459	504	535	565	612	619	636	632	611	579	571	569	636	1 658	1 400	1 512	1 580	1 693	1 855	1 873	1 883	1 787	1 657	1 541	1 521	1 505	1 883	12 646	10 327	10 997	11 958	12 294	13 479	13 815	13 975	13 302	12 382	11 534	11 403	11 286	13 975
14	3214	12114 Армавир, ул. Мира,	Армавир	Юг	Зимняя	205	136	18	17	17	16	16	15	68	213	262	319	318	319	564	368	26	25	25	23	23	22	246	818	842	978	954	978	7 537	5 006	359	675	329	302	302	290	3 107	10 072	10 354	12 135	11 803	12 135
15	3214	12114 Армавир, ул. Мира,	Армавир	Юг	Летняя	46	45	241	312	368	490	487	434	147	142	134	133	133	490	69	72	877	1 139	1 225	1 424	1 360	1 146	268	253	235	233	233	1 424	441	449	5 367	6 850	7 440	8 562	8 215	6 948	1 740	1 643	1 528	1 518	1 518	8 562
16	3214	12114 Армавир, ул. Мира,	Армавир	Юг	Утепленна	264	279	274	265	173	171	169	185	212	239	239	235	231	279	775	823	788	756	488	483	466	539	613	685	662	659	650	823	9 372	10 002	9 612	9 219	5 288	5 228	5 022	5 839	6 553	7 462	7 206	7 142	7 047	10 002
17	3214	12114 Армавир, ул. Мира,	Армавир	Юг	Всего	1 050	919	1 037	1 129	1 123	1 289	1 291	1 270	1 059	1 205	1 214	1 258	1 251	1 291	3 066	2 663	3 203	3 500	3 431	3 785	3 722	3 590	2 914	3 413	3 280	3 391	3 342	3 785	29 995	25 784	26 335	28 701	25 352	27 571	27 354	27 052	24 702	31 559	30 622	32 198	31 654	32 198
18	3234	12134 Армавир, ул. Ворон	Армавир	Юг	Всесезонна	609	561	589	633	675	639	692	722	773	747	692	674	669	773	2 117	1 847	1 893	1 974	2 119	1 949	2 120	2 166	2 172	1 947	1 828	1 764	1 740	2 172	16 690	13 412	13 800	14 595	15 224	14 103	15 362	15 730	15 751	14 092	13 205	12 804	12 623	16 690
19	3234	12134 Армавир, ул. Ворон	Армавир	Юг	Зимняя	275	193	33	29	29	29	27	25	125	245	339	415	409	415	871	529	42	35	35	35	32	28	546	1 008	1 118	1 397	1 342	1 397	11 894	7 258	529	831	432	432	400	353	6 488	11 759	12 948	16 556	15 870	16 556
20	3234	12134 Армавир, ул. Ворон	Армавир	Юг	Летняя	165	165	313	412	479	511	517	471	159	153	151	149	149	517	445	447	1 452	1 739	1 739	1 653	1 505	1 252	349	331	329	326	326	1 739	2 733	2 730	8 996	10 690	10 754	10 268	9 415	7 853	2 195	2 079	2 052	2 029	2 029	10 754
21	3234	12134 Армавир, ул. Ворон	Армавир	Юг	Утепленна	293	301	295	293	132	130	130	201	258	290	285	285	285	301	1 005	964	915	877	360	359	346	707	868	939	868	882	864	1 005	12 324	11 635	11 109	10 637	4 184	4 082	3 916	7 322	9 094	10 099	9 098	9 116	8 923	12 324
22	3234	12134 Армавир, ул. Ворон	Армавир	Юг	Всего	1 342	1 220	1 230	1 367	1 315	1 309	1 366	1 419	1 315	1 435	1 467	1 523	1 512	1 523	4 438	3 787	4 302	4 625	4 262	3 996	4 003	4 153	3 935	4 225	4 143	4 369	4 272	4 625	43 642	35 035	34 434	36 753	30 593	28 885	29 093	31 258	33 527	38 028	37 302	40 505	39 444	43 642
23	4208	22108 Архангельск, пр-т Т	Архангел	Север	Всесезонна	597	494	542	617	687	699	600	686	662	671	653	655	652	699	1 752	1 374	1 547	1 653	1 717	1 764	1 520	1 816	1 683	1 742	1 678	1 651	1 642	1 816	13 967	9 836	11 286	12 277	12 896	13 418	11 583	13 822	12 458	12 916	12 525	12 311	12 246	13 967
24	4208	22108 Архангельск, пр-т Т	Архангел	Север	Зимняя	300	275	218	209	151	151	151	143	190	358	399	435	431	435	738	623	435	418	277	278	278	259	631	1 279	1 301	1 434	1 383	1 434	10 555	8 789	6 016	6 074	3 492	3 504	3 504	3 252	8 054	16 282	16 369	17 968	17 353	17 968
25	4208	22108 Архангельск, пр-т Т	Архангел	Север	Летняя	56	54	134	318	541	550	483	453	185	184	184	173	173	550	118	108	480	1 245	1 675	1 572	1 271	1 149	363	359	357	342	342	1 675	811	718	3 101	7 748	10 781	10 194	8 413	7 642	2 579	2 558	2 536	2 437	2 437	10 781
26	4208	22108 Архангельск, пр-т Т	Архангел	Север	Утепленна	312	334	329	258	255	252	147	217	217	264	260	253	252	334	929	1 000	872	604	586	573	298	808	670	783	777	775	773	1 000	12 187	12 823	11 292	8 036	7 393	7 265	3 663	9 809	8 018	9 115	9 010	8 929	8 891	12 823
27	4208	22108 Архангельск, пр-т Т	Архангел	Север	Всего	1 265	1 157	1 223	1 402	1 634	1 652	1 381	1 499	1 254	1 477	1 496	1 516	1 508	1 652	3 537	3 105	3 334	3 920	4 255	4 187	3 367	4 032	3 347	4 163	4 113	4 202	4 140	4 255	37 520	32 166	31 694	34 134	34 562	34 381	27 164	34 524	31 110	40 871	40 439	41 645	40 927	41 645
28	4214	22114 Архангельск, ул. Со	Архангел	Север	Всесезонна												183	183	183												924	914	924												6 798	6 736	6 798
29	4214	22114 Архангельск, ул. Со	Архангел	Север	Зимняя												303	301	303												1 231	1 177	1 231												15 219	14 546	15 219
30	4214	22114 Архангельск, ул. Со	Архангел	Север	Утепленна												130	130	130												601	600	601												6 538	6 525	6 538
31	4214	22114 Архангельск, ул. Со	Архангел	Север	Всего	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	616,00	614,00	616	0	0	0	0	0	0	0	0	0	0	0	2 756	2 691	2 756	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	28 554,72	27 806,56	28 555
32	3447	14147 Асбест, ул. Ленингр	Асбест	Север	Всесезонна	603	530	560	612	649	553	588	648	666	714	700	683	682	714	1 881	1 633	1 719	1 850	1 916	1 555	1 677	1 860	1 887	2 149	2 063	1 982	1 979	2 149	14 532	11 845	12 451	13 741	14 437	11 938	12 972	14 438	13 868	15 932	15 389	14 875	14 858	15 932
33	3447	14147 Асбест, ул. Ленингр	Асбест	Север	Зимняя	241	242	97	11	11	11	11	11	11	238	336	403	401	403	740	689	225	11	11	11	11	11	11	781	993	1 056	1 041	1 056	10 597	9 823	3 179	673	146	146	146	146	146	9 444	12 040	12 816	12 622	12 816
34	3447	14147 Асбест, ул. Ленингр	Асбест	Север	Летняя	16	98	232	339	461	460	431	412	51	52	52	48	48	461	23	197	849	1 273	1 618	1 402	1 227	1 147	79	83	83	77	77	1 618	156	1 189	5 385	7 661	10 382	9 141	8 210	7 720	449	465	482	450	450	10 382
35	3447	14147 Асбест, ул. Ленингр	Асбест	Север	Утепленна	245	236	245	233	84	84	84	133	95	193	199	197	197	245	745	703	696	573	194	193	191	367	228	580	599	592	591	745	9 340	8 904	8 849	7 579	2 468	2 458	2 425	4 519	2 769	6 833	6 987	6 913	6 902	9 340
36	3447	14147 Асбест, ул. Ленингр	Асбест	Север	Всего	1 105	1 106	1 134	1 195	1 205	1 108	1 114	1 204	823	1 197	1 287	1 331	1 328	1 331	3 389	3 222	3 489	3 707	3 739	3 161	3 106	3 385	2 205	3 593	3 738	3 707	3 688	3 739	34 625	31 762	29 864	29 654	27 433	23 682	23 753	26 823	17 233	32 674	34 897	35 052	34 831	35 052
37	4508	25108 Астрахань, ул. Мину	Астрахан	Юг	Всесезонна	468	456	493	550	635	631	703	780	767	810	686	673	671	810	1 616	1 501	1 599	1 746	2 047	1 981	2 520	2 773	2 568	2 601	2 019	1 950	1 940	2 773	11 893	11 090	11 797	13 078	15 303	14 824	19 207	21 169	19 122	19 449	15 542	15 013	14 939	21 169
38	4508	25108 Астрахань, ул. Мину	Астрахан	Юг	Зимняя	169	162	8	8	8	8	8	8	56	232	269	298	295	298	431	414	9	9	9	9	9	9	208	873	891	908	880	908	5 898	5 606	129	458	125	125	125	125	2 459	11 042	11 142	11 238	10 887	11 238
39	4508	25108 Астрахань, ул. Мину	Астрахан	Юг	Летняя	66	56	99	335	472	478	503	479	197	196	44	44	44	503	148	127	337	1 255	1 497	1 327	1 443	1 295	462	460	57	57	57	1 497	984	824	2 134	7 523	9 575	8 745	9 450	8 470	3 177	3 159	371	371	371	9 575
40	4508	25108 Астрахань, ул. Мину	Астрахан	Юг	Утепленна	236	231	231	230	99	99	99	132	163	236	230	227	226	236	829	735	696	682	307	307	307	447	545	720	639	642	637	829	10 106	8 942	8 528	8 407	3 438	3 438	3 438	4 928	5 874	7 937	6 859	6 800	6 745	10 106
41	4508	25108 Астрахань, ул. Мину	Астрахан	Юг	Всего	939	905	831	1 123	1 214	1 216	1 313	1 399	1 183	1 474	1 229	1 242	1 236	1 474	3 024	2 777	2 641	3 692	3 860	3 624	4 279	4 524	3 783	4 654	3 606	3 557	3 514	4 654	28 881	26 462	22 587	29 467	28 441	27 132	32 220	34 692	30 633	41 587	33 915	33 421	32 943	41 587
42	4509	25109 Астрахань, ул. Анри	Астрахан	Юг	Всесезонна	441	454	482	541	632	628	608	770	757	792	611	601	601	792	1 370	1 591	1 682	1 827	2 175	2 153	2 069	2 770	2 574	2 530	1 904	1 842	1 835	2 770	9 933	11 603	12 442	13 649	16 357	16 089	15 578	21 130	19 079	18 865	14 709	14 249	14 199	21 130
43	4509	25109 Астрахань, ул. Анри	Астрахан	Юг	Зимняя	147	146	5	5	5	5	5	6	31	195	230	264	260	264	338	348	6	6	6	6	6	7	116	670	695	738	717	738	4 533	4 619	86	388	83	83	83	92	1 369	8 171	8 377	8 713	8 497	8 713
44	4509	25109 Астрахань, ул. Анри	Астрахан	Юг	Летняя	9	30	67	301	441	450	421	441	152	146	18	18	18	450	14	91	271	1 120	1 405	1 283	1 077	1 158	344	332	22	22	22	1 405	91	579	1 645	6 667	9 042	8 371	7 078	7 595	2 339	2 275	129	129	129	9 042
45	4509	25109 Астрахань, ул. Анри	Астрахан	Юг	Утепленна	209	257	259	256	131	131	131	169	186	230	226	225	225	259	656	851	781	760	398	398	398	581	615	699	665	659	656	851	8 101	10 141	9 356	9 106	4 333	4 333	4 333	6 215	6 434	7 334	6 892	6 789	6 766	10 141
46	4509	25109 Астрахань, ул. Анри	Астрахан	Юг	Всего	806	887	813	1 103	1 209	1 214	1 165	1 386	1 126	1 363	1 085	1 108	1 104	1 386	2 378	2 881	2 740	3 713	3 984	3 840	3 550	4 516	3 649	4 231	3 286	3 261	3 230	4 516	22 657	26 942	23 529	29 810	29 815	28 876	27 072	35 033	29 221	36 644	30 107	29 881	29 592	36 644
47	4510	25110 Астрахань, ул. Боев	Астрахан	Юг	Всесезонна	524	558	597	651	734	763	747	840	809	828	676	664	660	840	1 661	2 085	2 101	2 093	2 515	2 679	2 587	2 920	2 529	2 414	1 835	1 748	1 735	2 920	12 789	15 183	15 505	15 659	19 021	20 160	19 512	21 882	18 600	17 822	13 841	13 146	13 050	21 882
48	4510	25110 Астрахань, ул. Боев	Астрахан	Юг	Зимняя	206	200	16	16	16	16	16	16	81	359	420	428	422	428	476	520	17	17	17	17	17	17	365	1 436	1 460	1 329	1 276	1 460	6 493	7 026	261	591	250	250	250	250	3 920	17 212	17 522	15 846	15 204	17 522
49	4510	25110 Астрахань, ул. Боев	Астрахан	Юг	Летняя	103	97	148	404	556	575	580	511	145	145	34	34	34	580	257	252	578	1 461	1 775	1 668	1 628	1 334	327	327	55	55	55	1 775	1 602	1 596	3 463	8 869	11 274	10 452	10 558	8 722	2 230	2 230	293	295	295	11 274
50	4510	25110 Астрахань, ул. Боев	Астрахан	Юг	Утепленна	284	333	326	326	225	225	225	301	319	366	340	330	327	366	984	1 205	1 122	1 082	710	710	710	1 152	1 099	1 116	965	938	926	1 205	12 272	15 059	14 145	13 522	8 707	8 707	8 707	12 987	12 268	12 505	10 536	10 076	9 922	15 059
51	4510	25110 Астрахань, ул. Боев	Астрахан	Юг	Всего	1 117	1 188	1 087	1 397	1 531	1 579	1 568	1 668	1 354	1 698	1 470	1 456	1 443	1 698	3 378	4 062	3 818	4 653	5 017	5 074	4 942	5 423	4 320	5 293	4 315	4 070	3 992	5 423	33 157	38 865	33 375	38 640	39 252	39 569	39 027	43 841	37 018	49 770	42 192	39 363	38 472	49 770
52	4513	25113 Астрахань, пл. Бока	Астрахан	Юг	Всесезонна	441	464	501	566	644	687	688	798	783	811	646	642	638	811	1 419	1 575	1 667	1 845	2 193	2 378	2 398	2 721	2 454	2 341	1 774	1 732	1 711	2 721	11 199	11 571	12 300	13 393	16 401	17 665	17 845	20 269	18 020	17 215	13 456	13 152	12 993	20 269
53	4513	25113 Астрахань, пл. Бока	Астрахан	Юг	Зимняя	148	147	12	12	12	12	12	12	83	316	350	370	362	370	348	388	14	14	14	14	14	14	345	1 192	1 086	1 078	1 016	1 192	4 761	5 233	195	187	187	187	187	187	3 785	14 218	12 786	12 673	11 904	14 218
54	4513	25113 Астрахань, пл. Бока	Астрахан	Юг	Летняя	103	104	155	428	542	541	553	492	156	142	30	30	30	553	279	277	617	1 552	1 637	1 470	1 509	1 206	322	295	39	39	39	1 637	1 765	1 739	3 664	9 549	10 648	9 743	10 014	8 060	2 258	2 109	257	257	257	10 648
55	4513	25113 Астрахань, пл. Бока	Астрахан	Юг	Утепленна	281	305	295	290	164	164	165	232	251	293	284	274	273	305	961	1 032	925	888	487	487	489	861	867	904	819	799	789	1 032	11 759	12 457	11 178	10 346	5 556	5 556	5 574	9 007	9 082	9 630	8 590	8 284	8 176	12 457
56	4513	25113 Астрахань, пл. Бока	Астрахан	Юг	Всего	973	1 020	963	1 296	1 362	1 404	1 418	1 534	1 273	1 562	1 310	1 316	1 303	1 562	3 007	3 272	3 223	4 299	4 331	4 349	4 410	4 802	3 988	4 732	3 718	3 648	3 555	4 802	29 483	31 000	27 337	33 476	32 792	33 150	33 621	37 523	33 145	43 172	35 089	34 365	33 330	43 172
57	4002	20102 Ачинск, 7 микрорайо	Ачинск	Север	Всесезонна	799	817	834	897	945	1 001	841	834	834	841	823	814	813	1 001	2 645	2 799	2 790	2 976	3 002	3 118	2 445	2 423	2 403	2 405	2 317	2 277	2 269	3 118	20 429	20 571	20 390	21 879	21 970	23 274	18 589	18 559	17 727	17 796	17 178	16 901	16 847	23 274
58	4002	20102 Ачинск, 7 микрорайо	Ачинск	Север	Зимняя	342	242	21	21	21	21	21	152	208	337	344	368	363	368	955	596	23	23	23	23	23	340	686	1 192	1 040	1 073	1 025	1 192	13 138	8 251	287	853	276	276	276	4 719	9 490	15 478	13 252	13 664	13 038	15 478
59	4002	20102 Ачинск, 7 микрорайо	Ачинск	Север	Летняя	31	31	284	402	468	422	392	355	2	2	2	2	2	468	64	64	899	1 392	1 581	1 225	1 092	954	4	4	4	4	4	1 581	562	562	5 583	8 179	9 808	7 636	6 899	6 005	24	24	23	23	23	9 808
60	4002	20102 Ачинск, 7 микрорайо	Ачинск	Север	Утепленна	305	331	307	291	290	205	205	234	236	286	283	282	282	331	975	1 084	862	785	784	519	518	740	681	838	831	825	820	1 084	12 311	13 387	10 736	10 013	9 501	6 299	6 286	9 149	8 224	9 905	9 809	9 739	9 685	13 387

	B	C	D	E	F	G	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AW
1					Оптимальный запас пар лето (структура лето+всес 80%; утепл - 20%)	Оптимальный запас пар зима (структура зима+утепл-60%; всес 40%)	Макс. кол-во уникальных артикулов	Общий остаток в магазине в парах													Макс. Запас кор. (мес)	Макс. Объем в дм3
2	Магазин	Город	Климат	Мат.:Се				ЯНВ 20	ФЕВ 20	МАР 20	АПР 20	МАЙ 20	ИЮН 20	ИЮЛ 20	АВГ 20	СЕН 20	ОКТ 20	НОЯ 20	ДЕК 20	ЯНВ 20		
3	07108 Альметьевс	Альметье	Центр	Всесезонная			678	1 588	1 321	1 684	1 820	1 872	1 918	1 721	1 749	1 738	1 469	1 418	1 378	1 373	1 918	14 587
4	07108 Альметьевс	Альметье	Центр	Зимняя			336	674	584	92	12	11	11	11	10	500	1 085	1 114	1 168	1 145	1 168	15 105
5	07108 Альметьевс	Альметье	Центр	Летняя			521	36	36	1 019	1 534	1 610	1 522	1 185	1 074	410	18	18	17	17	1 610	10 146
6	07108 Альметьевс	Альметье	Центр	Утепленна	например		271	819	794	817	712	338	337	336	583	747	748	737	728	724	819	10 567
7	**07108 Альметьев**	**Альметье**	**Центр**	**Всего**	**3700**	**3300**	**1 323**	**3 117**	**2 735**	**3 612**	**4 078**	**3 831**	**3 788**	**3 253**	**3 416**	**3 395**	**3 320**	**3 287**	**3 291**	**3 259**	**4 078**	**34 177**
8	07102 Казань, ул.	Казань	Центр	Всесезонная			790	1 830	1 976	2 251	2 238	2 288	2 216	1 930	2 271	2 193	1 975	1 850	1 764	1 759	2 288	16 967
9	07102 Казань, ул.	Казань	Центр	Зимняя			396	852	768	49	47	47	47	47	46	713	1 224	1 231	1 287	1 247	1 287	16 723
10	07102 Казань, ул.	Казань	Центр	Летняя			583	148	145	1 336	2 066	1 896	1 621	1 338	951	278	247	243	239	238	2 066	12 679
11	07102 Казань, ул.	Казань	Центр	Утепленная			344	1 008	1 015	1 049	878	505	502	489	787	932	823	804	789	785	1 049	13 367
12	**07102 Казань, ул**	**Казань**	**Центр**	**Всего**	**4500**	**4000**	**1 659**	**3 838**	**3 904**	**4 685**	**5 229**	**4 736**	**4 386**	**3 804**	**4 055**	**4 116**	**4 269**	**4 128**	**4 079**	**4 029**	**5 229**	**41 698**
13	07130 Казань, ул.	Казань	Центр	Всесезонная			647				1 739	1 899	1 881	1 939	2 266	2 308	1 955	1 659	1 355	1 339	2 308	17 084
14	07130 Казань, ул.	Казань	Центр	Зимняя			356									694	929	1 105	1 319	1 276	1 319	16 180
15	07130 Казань, ул.	Казань	Центр	Летняя			499				1 942	1 903	1 786	1 567	1 305	132	16	16	16	16	1 942	12 152
16	07130 Казань, ул.	Казань	Центр	Утепленная			243				300	292	290	281	681	947	779	755	735	729	947	10 136
17	**07130 Казань, ул**	**Казань**	**Центр**	**Всего**	**4000**	**3500**	**1 220**	**0,00**	**0,00**	**0,00**	**3 981**	**4 094**	**3 957**	**3 787**	**4 252**	**4 081**	**3 679**	**3 535**	**3 425**	**3 360**	**4 252**	**36 417**
18	07133 Казань, ул.	Казань	Центр	Всесезонная			533								2 008	1 970	1 822	1 887	1 801	1 793	2 008	14 398
19	07133 Казань, ул.	Казань	Центр	Зимняя			352								391	935	1 249	1 189	1 141	1 098	1 249	15 880
20	07133 Казань, ул.	Казань	Центр	Летняя			92								317	295	291	290	290	290	317	2 151
21	07133 Казань, ул.	Казань	Центр	Утепленная			223								768	929	834	813	795	791	929	10 205
22	**07133 Казань, ул**	**Казань**	**Центр**	**Всего**	**4800**	**4200**	**1 193**	**0,00**	**0,00**	**0,00**	**0,00**	**0,00**	**0,00**	**0,00**	**3 484**	**4 129**	**4 196**	**4 179**	**4 027**	**3 972**	**4 196**	**40 148**
23	07109 Набережны	Набережн	Центр	Всесезонная			789	1 605	1 519	1 943	2 036	2 130	2 395	2 130	2 133	2 031	1 752	1 565	1 503	1 495	2 395	17 982
24	07109 Набережны	Набережн	Центр	Зимняя			357	641	593	17	17	17	17	17	17	600	1 072	1 117	1 220	1 198	1 220	15 825
25	07109 Набережны	Набережн	Центр	Летняя			493	92	92	1 045	1 511	1 478	1 470	1 075	837	230	21	21	22	22	1 511	9 073
26	07109 Набережны	Набережн	Центр	Утепленная			312	858	886	921	796	368	368	367	567	717	651	648	643	640	921	11 636
27	**07109 Набережн**	**Набережн**	**Центр**	**Всего**	**4100**	**3200**	**1 437**	**3 196**	**3 090**	**3 926**	**4 360**	**3 993**	**4 250**	**3 589**	**3 554**	**3 578**	**3 496**	**3 351**	**3 388**	**3 355**	**4 360**	**35 455**
28	07114 Набережны	Набережн	Центр	Всесезонная			774	2 116	1 707	1 959	2 030	2 206	1 951	1 890	1 896	1 955	1 868	1 710	1 674	1 667	2 206	16 456
29	07114 Набережны	Набережн	Центр	Зимняя			364	564	549	12	13	13	13	13	13	780	1 283	1 269	1 301	1 265	1 301	16 588
30	07114 Набережны	Набережн	Центр	Летняя			544	210	209	1 316	1 833	1 811	1 567	1 155	1 011	446	92	91	91	91	1 833	11 085
31	07114 Набережны	Набережн	Центр	Утепленная			285	780	854	847	703	309	309	309	616	737	711	728	717	716	854	10 784
32	**07114 Набережн**	**Набережн**	**Центр**	**Всего**	**5200**	**4700**	**1 476**	**3 670**	**3 319**	**4 134**	**4 579**	**4 339**	**3 840**	**3 367**	**3 536**	**3 918**	**3 954**	**3 798**	**3 783**	**3 739**	**4 579**	**38 802**

Also, the department of transport logistics provided us with data on lead time for each store, if distribution occurs from finished goods warehouses. Of course, we collected and worked with even more data, but we are not sure that the data will be needed in this thesis, so we did not describe them in this chapter. If any additional data is needed to solve the problems we are faced with, then we will talk about this data in other chapters.

7.0 Selection and development of a forecasting method

By systematizing what is written in the literature review and research methodology, we concluded that we will use time-based methods to forecast demand in stores. We will test on our data using the Random Forest Regression algorithm. To begin with, we must prepare our data for working with time series and add lost sales to the data.

To estimate lost sales, we used SAP BI features. Now in SAP BI you can calculate the average value of the weekly sales rate. We also have the remnants of the goods in the shops every day. We tracked down the daily balances when the item is not in the store. For example, consider the article X, it refers to the spring collection, the active sales of which fall on 4-23 weeks. Only 20 weeks of active sales. Moreover, a report from SAP BI showed us that the average sales rate per week for one store of article X is 2 pairs of shoes per week. From the history of balances, we see that in store Y from 20 active sales weeks, 8 weeks of item X were not in store Y. Therefore, we estimate lost sales of item X in store Y in the amount of 16 pairs. These 16 pairs of lost sales were randomly added to a copy of the sales history data. This was done in order to predict not sales, but demand. Therefore, we made this assumption in order to imagine that at any time in the store can satisfy customer demand 100%. So, let's proceed to the preparation of data for forecasting.

7.1 Data preparation

Let us remind you that we have sales data for each store over the past three years. The estimated value of lost sales has been added to this data. So now we have the right to call our data historical data on demand in stores.

For our algorithm, based on predicting the time series for each store, we need to bring the data to the form shown in Table N.

Date	y
2016-02-16	20
2016-02-17	29
…	…
2019-03-04	17
…	…

Table 7.1: Sample data for prediction

The values in the y column indicate the number of sales (including lost sales) that occurred on February 16, 2016. Let me remind you what our demand data now looks like after transferring them to csv format. First of all, we need to remove products that are not shoes and remove columns that are redundant in time-series forecasting of demand. Figure N shows the code that performs the operations we need. We used the Python programming language, the Pandas library, the Jupyter Notebook development environment.

	A	B	C	D	E	F	G	H	I	J	K	L
1	,ID_SHOP,SHOPNUM,SHOPNAME,ART,SEASON,GROUPMW,SIZE,KOL,SALE_DATE											
2	0,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",BW-1028, , ,0.0,1,2016-02-16											
3	1,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",GR17, ,Женские,0.0,1,2016-02-16											
4	2,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",KS1001-01, , ,0.0,1,2016-02-16											
5	3,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",141001,Зимняя,Женские,35.0,1,2016-02-16											
6	4,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",15С1456 2/40/БРОН3, ,Женские,0.0,1,2016-02-16											
7	5,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",1635000,Утепленная,Женские,39.0,1,2016-02-16											
8	6,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",1635041,Утепленная,Женские,39.0,1,2016-02-16											
9	7,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",1637162/1,Всесезонная,Женские,38.0,1,2016-02-16											
10	8,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",1637511,Всесезонная,Женские,38.0,1,2016-02-16											
11	9,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",216072,Летняя,Мужские,44.0,1,2016-02-16											
12	10,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",238100,Летняя,Женские,38.0,1,2016-02-16											
13	11,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",238165,Летняя,Женские,38.0,1,2016-02-16											
14	12,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",238375,Летняя,Женские,39.0,1,2016-02-16											
15	13,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",330605,Зимняя,Женские,37.0,1,2016-02-16											
16	14,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",410051,Зимняя,Мужские,41.0,1,2016-02-16											
17	15,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",410095,Зимняя,Мужские,40.0,1,2016-02-16											
18	16,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",413096,Зимняя,Мужские,44.0,1,2016-02-16											
19	17,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",437172,Всесезонная,Женские,40.0,1,2016-02-16											
20	18,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",445001/Б,Утепленная,Женские,43.0,1,2016-02-16											
21	19,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",445010,Утепленная,Женские,37.0,1,2016-02-16											
22	20,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",445010,Утепленная,Женские,38.0,1,2016-02-16											
23	21,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",5С3134 Р27, ,Мужские,0.0,1,2016-02-16											
24	22,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",526071,Всесезонная,Мужские,41.0,1,2016-02-16											
25	23,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",547037/1,Всесезонная,Женские,38.0,1,2016-02-16											
26	24,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",925000,Утепленная,Мужские,44.0,1,2016-02-16											
27	25,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",BW-1097, , ,0.0,1,2016-02-17											
28	26,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",КАДЕМИ 40СМ, , ,0.0,1,2016-02-17											
29	27,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",01502/01,Всесезонная,Мужские,46.0,1,2016-02-17											
30	28,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",016200,Всесезонная,Мужские,44.0,1,2016-02-17											
31	29,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",037501,Всесезонная,Женские,38.0,1,2016-02-17											
32	30,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",138300,Летняя,Женские,40.0,1,2016-02-17											
33	31,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",15С1456 3/40/БРОН3, ,Женские,0.0,1,2016-02-17											
34	32,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",1632002/1,Утепленная,Женские,40.0,1,2016-02-17											
35	33,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",1635005,Утепленная,Женские,39.0,1,2016-02-17											
36	34,1,1,"ПОЛОЦК ФИРМ. МАГАЗИН №1 ул. Коммунистическая, д.13а ",195101, , ,0.0,1,2016-02-17											

Table 7.2: Historical demand data before processing

```python
def delete_rows_in_csv(path):
    """Removes lines in a .csv file that contain sales of related products"""

    files = glob.glob(path)

    for file in files:
        data = pd.read_csv(file, encoding='utf-16')
        data = data.drop(data[data.SIZE == 0.0].index)
        del data['Unnamed: 0']
        del data['ID_SHOP']
        del data['SHOPNUM']
        del data['SHOPNAME']
        del data['ART']
        del data['SEASON']
        del data['GROUPMW']
        del data['SIZE']
        data.to_csv('Data_for_predict/'+file[10:14]+'.csv', encoding='utf-16')

delete_rows_in_csv('Full_data/*.csv')
```

Figure 7.1: A piece of code to remove unnecessary data

After this operation, we received data containing the purchase ID, quantitatively purchased shoes and the date of purchase.

	A	B
1	,KOL,SALE_DATE	
2	3,1,2016-02-16	
3	5,1,2016-02-16	
4	6,1,2016-02-16	
5	7,1,2016-02-16	
6	8,1,2016-02-16	
7	9,1,2016-02-16	
8	10,1,2016-02-16	
9	11,1,2016-02-16	
10	12,1,2016-02-16	

After that we need to summarize the sales of the day and change the columns in some places.

```python
def prepare_for_predict(path):
    """Permutation of KOL and SALE_DATE columns. Daily sales summation"""

    files = glob.glob(path)

    for file in files:
        data = pd.read_csv(file, encoding='utf-16')
        del data['Unnamed: 0']
        data['ds'] = pd.to_datetime(data['SALE_DATE'])
        data = data.set_index('ds')
        data.drop(['SALE_DATE'], axis=1, inplace=True)
        data.columns = ['y']
        data = data.resample('D').sum()
        data.to_csv('Ready_data/'+file[17:21]+'.csv', encoding='utf-16')

prepare_for_predict('Data_for_predict/*.csv')
```

After that we get an almost ready set of data for all the stores in the network. It remains to delete only the purchase ID. As a result, we get what we wanted.

	A	B
1	ds,y	
2	2016-02-16,20	
3	2016-02-17,29	
4	2016-02-18,30	
5	2016-02-19,15	
6	2016-02-20,30	
7	2016-02-21,42	
8	2016-02-22,18	
9	2016-02-23,19	
10	2016-02-24,15	

7.2 Forecast Method Description

We use the time series model with three main components of the model: trend, seasonality and holidays. They are combined in the following equation (Harvey and Peters 1990):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

Here, g (t) is a trend function that models non-periodic changes in the value of a time series, s (t) represents periodic changes (for example, weekly and annual seasonality), and h (t) represents the effect of holidays that occur on potentially irregular graphs one or more days. The term error represents any specific changes that are not related to the model.

This specification is similar to the generalized additive model (GAM) class of regression models with potentially non-linear smoothers applied to regressors. Here we only use time as a regressor, but perhaps several linear and non-linear functions of time as components (Hastie and Tibshirani 1987). Modeling seasonality as an additive component is the same approach used for exponential smoothing (Gardner 1985). Multiplicative seasonality, where the seasonal effect is a factor that multiplies g (t).

In essence, we formulate the problem of forecasting as an exercise on a curve, which by its nature differs from time series models, which explicitly take into account the structure of time dependence in the data. Although we discard some important output advantages of using a generative model, such as ARIMA, this formulation provides a number of practical advantages (J. Taylor and Letham 2017):

- Flexibility: we can easily adapt the seasonality to several periods and allow us to analyze various assumptions about trends;
- Unlike ARIMA models, measurements should not be distributed regularly, and we do not need to interpolate missing values, for example, from removing outliers;
- The training of the model is carried out very quickly, which allows us to immediately explore the result of training with different parameters of the model;

The prediction model has easily interpretable parameters that the analyst can modify to impose assumptions on the forecast. Moreover, the model can easily be expanded to include new components (J. Taylor and Letham 2017).

7.3 Model implementation

To use the above model in the research department of the company Facebook, a library was developed containing the implementations of mathematical functions. We were faced with the task of understanding how this library works, in order to understand how to apply it to our demand forecasting. This library has greatly simplified and accelerated our work, because we did not have to implement the mathematical methods described in (Harvey and Peters 1990) from scratch.

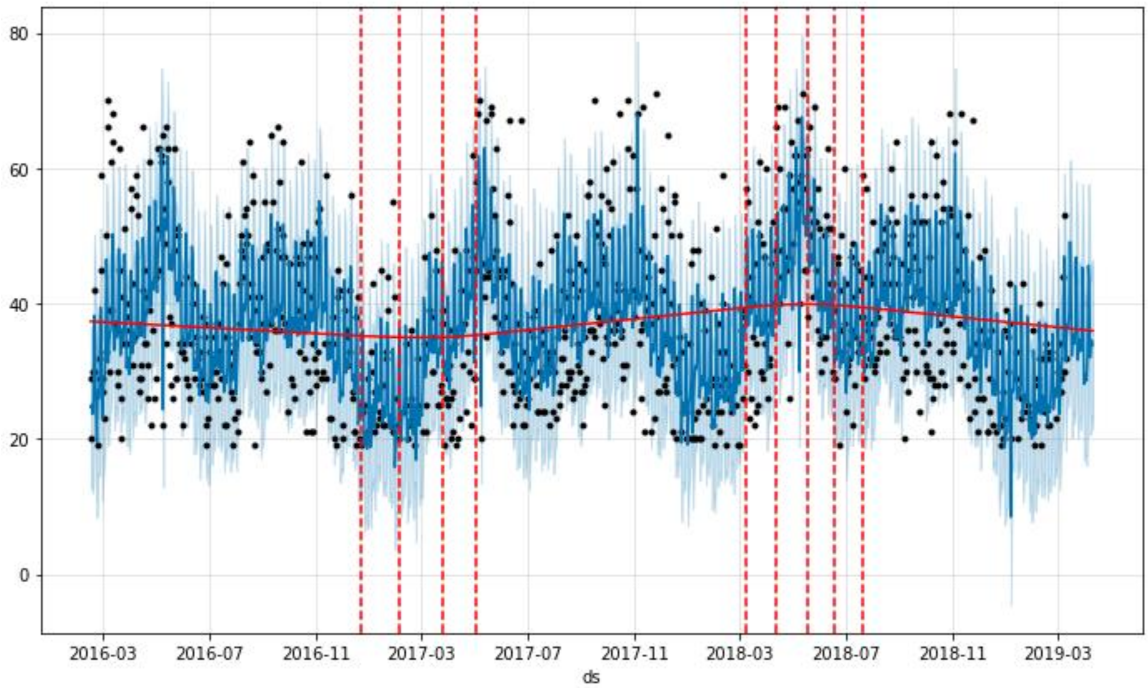
So, first we need to download a file with historical data on demand in the store. We also decided to remove anomalous days from the very beginning. We considered the day abnormal if the demand on that day exceeded twice the value of the arithmetic average for all days of observations. Next, we created an empty demand forecasting model by activating the parameter of weekly seasonality. After that they added to the model data on holidays and vacations in Russia and Belarus. After that added monthly seasonality, lasting 30.5 days. After that, you need to choose how often the forecast we need. You can adjust any period of time, but we, of course, are interested in the daily forecast. Planning horizon can be configured from one day to one year. After that, you can train the model on historical data. The output is the following Table N, which can be downloaded to the Excel file.

	ds	yhat	yhat_lower	yhat_upper
1146	2019-04-07	42.945359	31.045040	54.298432
1147	2019-04-08	32.012209	20.560192	43.272301
1148	2019-04-09	34.246160	21.932629	45.579514
1149	2019-04-10	34.541470	23.502869	46.532006
1150	2019-04-11	33.852194	21.417881	45.633005

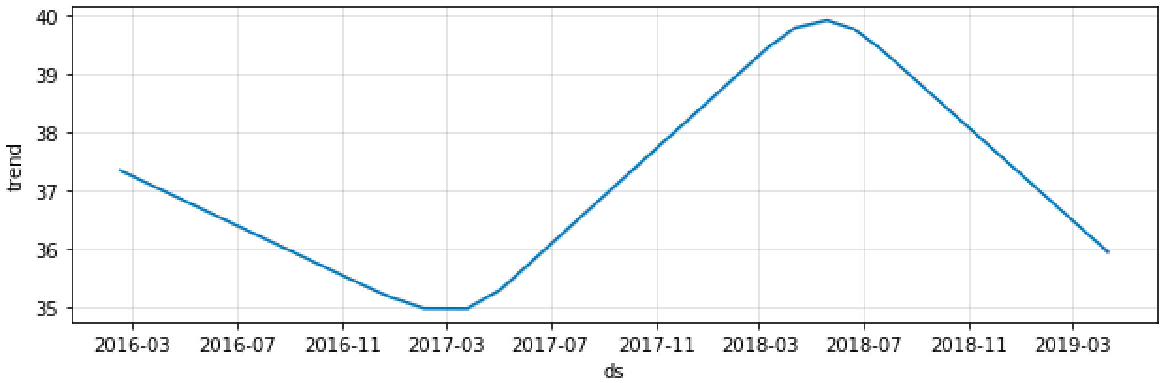
The yhat column predicts demand for the date indicated in the ds column. The remaining two columns are the lower and upper intervals, with which, with a 95% probability, the real value of the demand will fall.

As an added bonus, if you wish, we can display additional graphs for analytics.

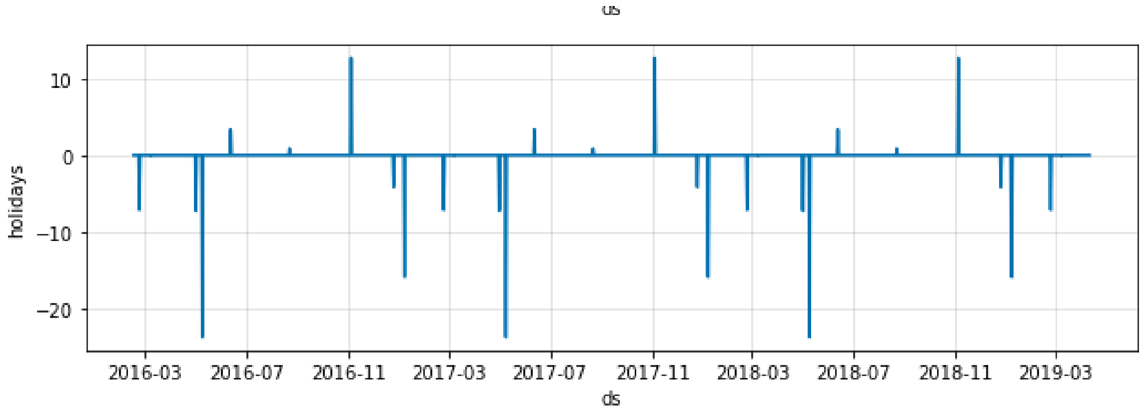
For example, the schedule of changes in demand.



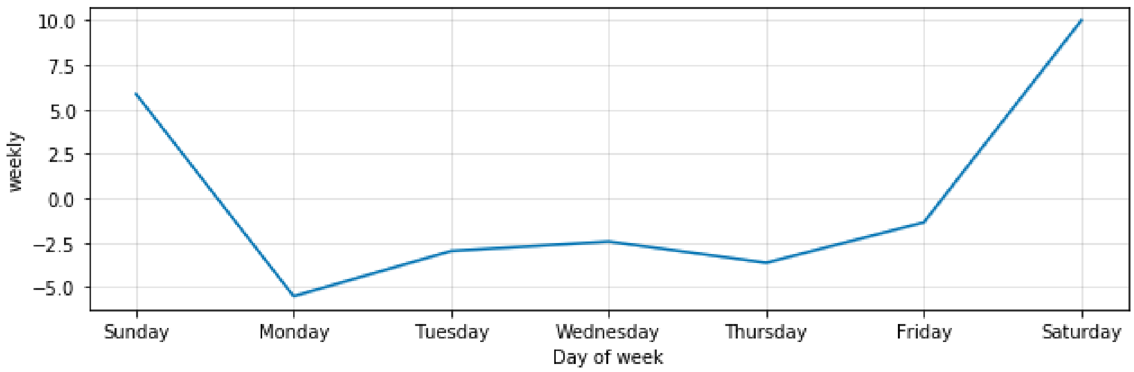
Changing the trend of daily demand in the store during the collection of statistics.



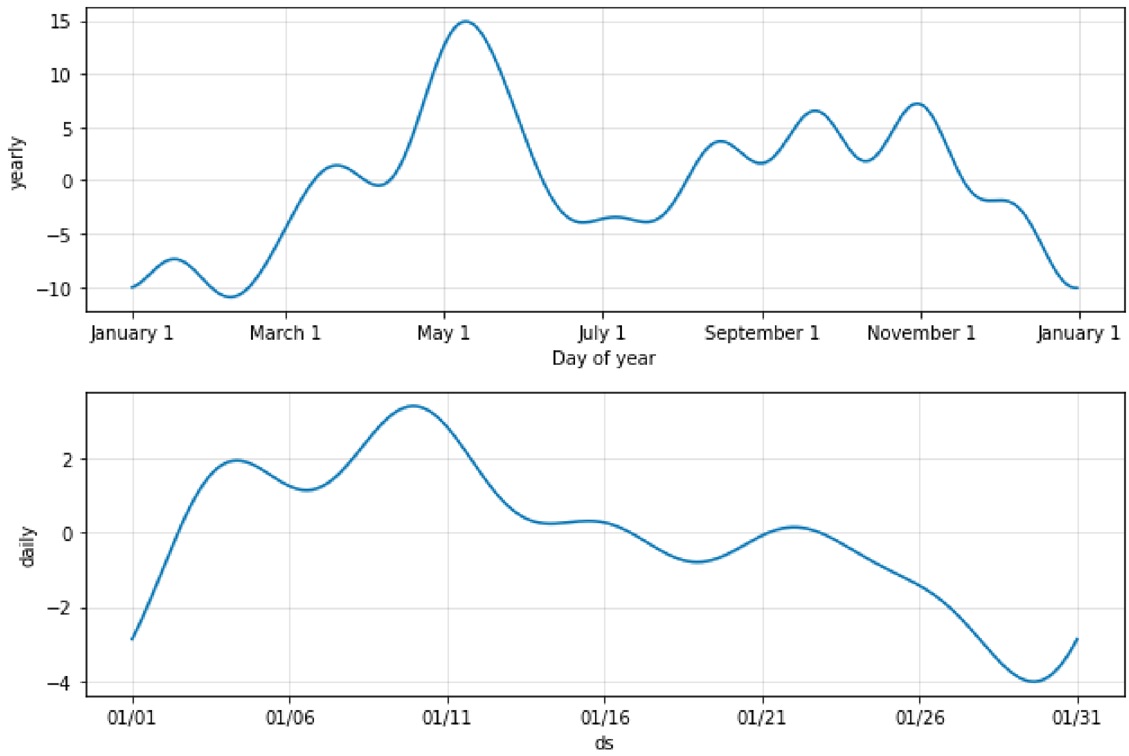
It is possible to provide analytics on sales changes during the holidays.



Weekly changes in demand are available for analysis.



Monthly and annual changes in demand are also available for analysis.



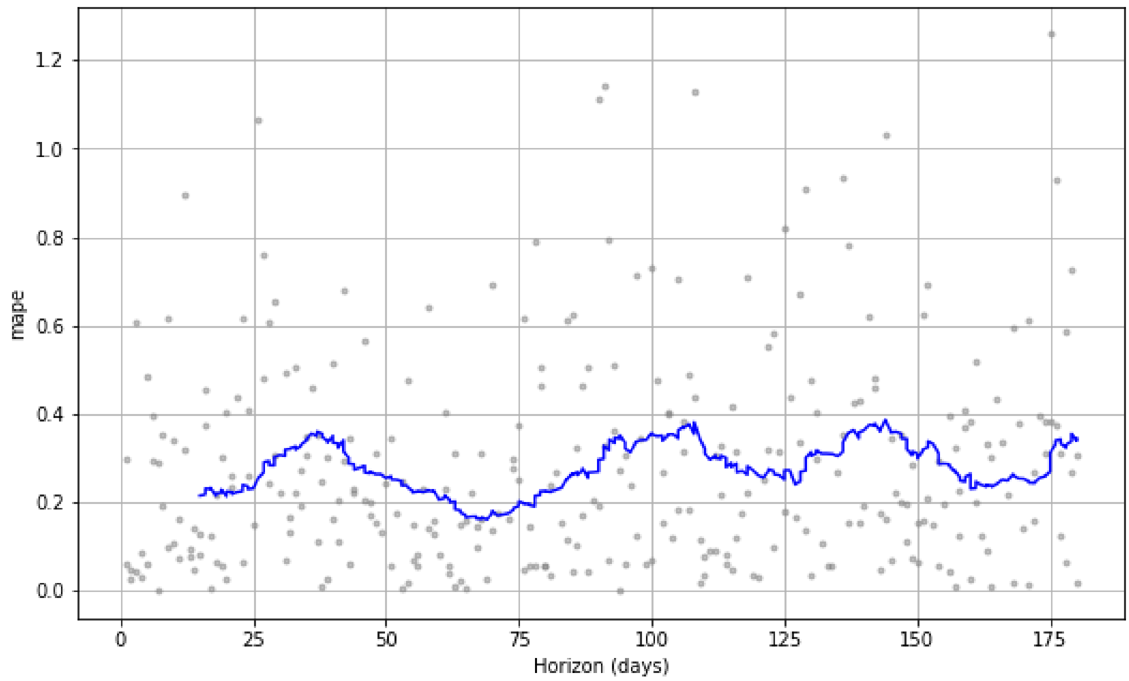
7.4 Experimental results

The prediction error was considered in two ways: cross-validation and demand results in real life. The cross-validation method is to break down historical demand data into 10 parts. Training models to produce on 9 parts of the data, and the model must predict a tenth of the data. After that, the real data and the results of the forecast are compared.

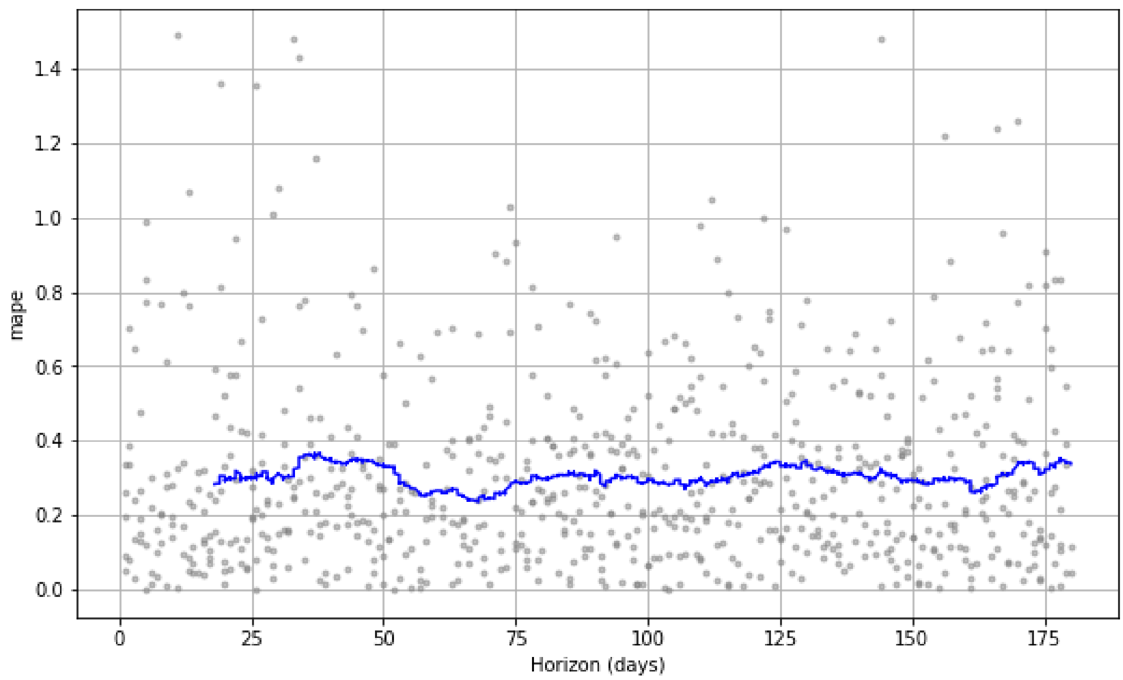
How is a mistake counted? MAPE (mean absolute percentage error) - the average absolute error of our forecast. Let y_i be real demand, $\bar{y}_i$ be forecasted demand. Then $e_i = | y_i - \bar{y}_i |$ - forecast error. Then: $p_i = \frac{e_i}{y_i}$

This is the absolute error of the forecast, expressed as a percentage.

Different stores have different error values. For example, for store X, the error at the forecast horizon of 25 days was 23.1%. When forecasting for 180 days ahead, the error was 33.9%.



For Store Y, the prediction error ranged from 28.3% to 34%. The growth of the error is due to the fact that for the Y data store for training the model was 30% less than for the X store.



8.0 Description of heuristics for distribution, replenishment and redistribution

8.1 Description of a heuristic for distribution

Before describing the method of distribution of goods, it is necessary to introduce the concept of branch. The network of stores of company B are not objects of one company. Due to Russian tax legislation, the chain of stores is divided into several separate companies. This was done in order to reduce the tax burden on the company. The main limitation associated with this way of organizing a chain of stores is that the company should not employ more than 100 people.

In Belarus, the situation is different. For the convenience of the company, stores located in Belarus are also divided into separate companies. The criterion of separation was the geographical location of stores in Belarus. There are 32 such small companies in total.

Company name	Number of stores
Company B1	11
Company B2	19
Company B3	22
Company B4	12
Company B5	23
Company B6	16
Company B7	15
Company B8	15
Company B9	13
Company B10	13
Company B11	13
Company B12	30
Company B13	20
Company B14	16
Company B15	9
Company B16	14
Company B17	14
Company B18	6

Company B19	8
Company B20	7
Company B21	13
Company B22	20
Company B23	3
Company B24	4
Company B25	2
Company B26	5
Company B27	6
Company B28	6
Company B29	12
Company B30	8
Company B31	6
Company B32	10

The exchange of products between companies is not possible directly. The only way to transfer products between companies is to send goods from one company to the finished goods warehouse. After that, it is possible to send arriving shoes from one company to another. Inside the company, the exchange of goods between stores is free. This clarification is very important for the in-store transfer.

In previous chapters, a method was described of how to obtain historical data from sales history data. For each season, data on the demand of stores within one company were combined. The total demand for the whole network for all seasons was also calculated; active sales weeks for the seasons were taken into account. For each company, the specific share of demand was calculated on the basis of total demand for active sales weeks for each season. For example, the production has produced 1,500 pairs of articles for the new seasonal collection. There is a certain branch X, the specific share of demand of which is 5%. Consequently, 5% of the total number of article pairs available for distribution are distributed to the branch. Inside the branch, a store rating is also compiled based on the following factors:

- Demand data last year during the active sales weeks;
- Store format. Priority for distribution is considered to be shops located in shopping centers due to the higher permeability of potential customers;

- Maximum store capacity.

The growth box (full size line) receive the first 150 stores on the top of the rating. The following 100-120 stores in the ranking receive the most popular sizes. So, if the number of available pairs of the article was 1500, then at the initial distribution at the beginning of the season 1050-1110 pairs of shoes would be distributed. 26-30% of the remaining shoes remain in stock for replenishment during the season.

When the maximum capacity of the store is reached, it is eliminated from the distribution and is not taken into account in the rating.

Thus, when using this heuristic, the company reduces the factor of influence of the personal interest of employees in the distribution of goods. This heuristic caused extreme disapproval of senior store managers due to the fact that they no longer have the ability to influence the range of the store. Nevertheless, the company approved the use of this heuristic for distributing shoes throughout the network, starting in the autumn season of 2019.

8.2 Description of heuristics for replenishing stores with goods

There are two possible strategies for replenishment:

- Use (s, s) strategy for inventory management, where S is the maximum capacity of the store, expressed in pairs of shoes. s (safety stock) - the occupancy rate of the store is 80%. Replenishment occurs in all stores of the branch. The main condition is that at least one store from the branch reaches the level of safety stock. Also, replenishment occurs at least once a week;
- Use a fixed frequency of deliveries, using demand forecasting, calculate how many shoes you need to fill up in the store.

For both strategies, we need to know the time it takes to make a decision to replenish the store until you receive the goods in the store (Lead time). The data were obtained in cooperation with the logistics department.

Company name	Lead time, days
Company B1	3
Company B2	2
Company B3	4
Company B4	5
Company B5	2
Company B6	4

Company B7	6
Company B8	6
Company B9	6
Company B10	8
Company B11	9
Company B12	7
Company B13	7
Company B14	7
Company B15	8
Company B16	11
Company B17	6
Company B18	13
Company B19	11
Company B20	12
Company B21	9
Company B22	2
Company B23	2
Company B24	1
Company B25	7
Company B26	1
Company B27	1
Company B28	1
Company B29	1
Company B30	1
Company B31	1
Company B32	1

Our approach is to use a fixed frequency of deliveries for branches, for which the lead time takes no more than 3 days. Transportation costs for the replenishment of such stores are minimal. Frequency of supply ensures prompt replenishment of the product range and the absence of empty shelves. For branches, to which the lead time is more than three days, it makes sense to use (s, S) strategy for replenishment, since a sufficient number and variety of assortment is still present in the store. At the same time, funds are not spent on the work of transport, which at the same time can serve stores. They are located not so far from the warehouses of finished products.

So, when using the first strategy, you need to predict the date when stocks at least in one store of the branch will reach 80% of the maximum. Consider this case on a specific example. Take Store X, owned by affiliate B12. Lead time for the stores of this branch is 7 days. The maximum capacity in the summer season of store X is 4000 pairs. 80% of the store's workload is 3200 pairs. It remains to use the implemented method of forecasting demand in order to find out when the date will come to replenish stocks in the store. Also, using the data on balances, it is possible to calculate the number of pairs of shoes in the store on any date of interest within three years.

On March 12, the number of pairs of shoes in store X was equal to 3,560 pairs. It is necessary to find out after how many days, starting from March 13, 340 pairs of shoes will be sold out. We use the software to predict how long the sale of 340 pairs of shoes will occur. To solve this problem, the program has been slightly modified. The sales forecast occurred every day until the sum of the quantity of goods sold per day did not become 340. The table shows the projected sales for 25 days, during this period 341 pairs of shoes will be sold out. Consequently, the decision to replenish the branch should also be made on March 31, provided that shop X at branch B12 reached the safety stock the fastest.

```
mean_val = pd.to_numeric(forecast['yhat'][1716:]).sum()
print (mean_val)
```

```
341.0612799080272
```

In the company, this strategy was approved by the management and accepted for use. The approach described in the thesis allows us to obtain a general forecast of the demand for shops and the delivery dates. This information is transmitted to the department of corporate trade, which determine which articles will go to the store. Which determine which articles will go to the store.

	ds	yhat	yhat_lower	yhat_upper
1716	2019-03-13	12.796391	7.979354	17.928155
1717	2019-03-14	12.732164	7.998729	17.938799
1718	2019-03-15	13.567699	8.155737	18.721673
1719	2019-03-16	14.753316	9.606795	19.611157
1720	2019-03-17	14.768291	9.628375	19.729461
1721	2019-03-18	12.712524	7.648052	17.865795
1722	2019-03-19	13.733212	8.625042	18.596588
1723	2019-03-20	13.505159	8.149829	18.167506
1724	2019-03-21	13.371005	8.574074	18.816210
1725	2019-03-22	14.063589	8.940134	19.124524
1726	2019-03-23	15.007222	9.853050	20.372099
1727	2019-03-24	14.759494	9.561665	19.487292
1728	2019-03-25	12.538397	7.629061	17.541120
1729	2019-03-26	13.545509	8.556165	18.247536
1730	2019-03-27	13.393038	8.319782	18.669712
1731	2019-03-28	13.290974	7.937443	18.310007
1732	2019-03-29	13.880993	8.923718	19.055542
1733	2019-03-30	14.620042	9.493968	19.666546
1734	2019-03-31	14.202814	9.372212	18.925977
1735	2019-04-01	11.979973	7.192059	17.024434
1736	2019-04-02	13.170502	8.216934	17.913722
1737	2019-04-03	13.265897	8.328814	18.421449
1738	2019-04-04	13.300798	8.832119	18.544416
1739	2019-04-05	13.811259	9.021832	19.074847
1740	2019-04-06	14.291017	8.841376	19.141150

The implementation of the second strategy is easier than the first. For each store in the branch you need to predict sales in a time horizon of 3 and 7 days. It depends on the frequency of delivery, which can be 1 or 2 times a week. Also for each store you need to calculate the current balance of shoes, which is not difficult. This strategy is also accepted for use by company B.

8.3 Description of the method for the transfer of goods between stores

By the end of the season, seasonal shoes run out in finished goods warehouses. The replenishment stage in the collection life cycle comes to an end. It is time to transfer between stores. Company B has two types of transfers:

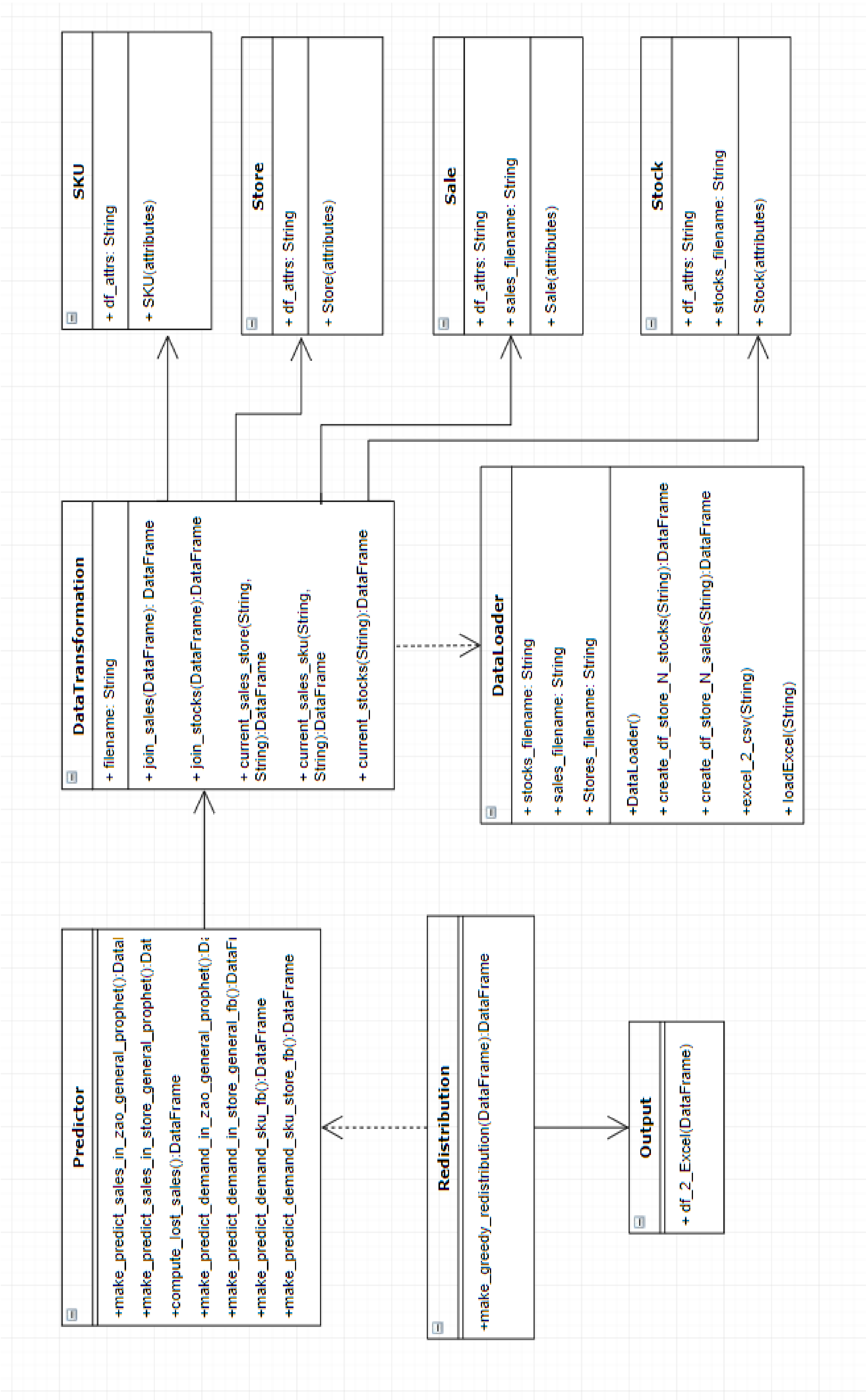
- The transfer of goods between stores within the same branch. We introduce for this type of transfer the name "small" transfer;
- The transfer of goods between stores not belonging to one branch using finished goods warehouses. For this type of transfer, we introduce the name “big” transfers.

The complexity of the transfer phase in comparison with the stages of initial placement and replenishment lies in the fact that now it is necessary to operate with demand forecasting not at the “shop-demand” level, but “shop-SKU-demand”. Due to the fact that at the end of the season in stores you cannot find articles with full sizes, you have to operate with goods at the level of SKU. Due to the number of stores and SKUs, the software must quickly produce a result due to the high load. Therefore, it was decided to implement heuristics, which is not computationally complex. For this reason, it was decided to abandon the calculation of the probability of selling SKU in all stores of the branch and transferring SKU to the store with the highest probability of sales. When calculating the sale probability, Bayesian methods are used that are computationally laborious.

The essence of this heuristics is estimated for each SKU. In which store this SKU will sell faster. If a store is found inside the branch, where the potential sale rate of a given SKU is higher than in the store where it is located, then an edge is created between the stores, to which other SKU are subsequently added. The work of heuristics is a graph where nodes are stores, and edges are those SKUs that need to be transferred from store to store. After that, for each edge is considered the potential additional income derived from the transfer of a set of SKU between the two stores. The potential additional income is compared to the cost of transporting this transfer. If the cost of transportation is higher than the potential additional income, then the transfer is not made. If the potential income is higher than the cost of transportation, the transfer is made. Non-redistributed shoes are shipped back to the finished goods warehouses. For this footwear the mechanism of replenishment of shops works.

It was decided to test this heuristic within one branch with the number of stores equal to 12. Also, 100 SKUs were selected for redistribution, which are sold in the stores of this branch.

Unfortunately, the moment of the beginning of the transfer comes May 25, 2019, so the thesis cannot describe the results of the experiment.


SKU
+ df_attrs: String
+ SKU(attributes)
Store
+ df_attrs: String
+ Store(attributes)
Sale
+ df_attrs: String
+ sales_filename: String
+ Sale(attributes)
Stock
+ df_attrs: String
+ stocks_filename: String
+ Stock(attributes)
DataTransformation
+ filename: String
+ join_sales(DataFrame): DataFrame
+ join_stocks(DataFrame):DataFrame
+ current_sales_store(String, String):DataFrame
+ current_sales_sku(String, String):DataFrame
+ current_stocks(String):DataFrame
DataLoader
+ stocks_filename: String
+ sales_filename: String
+ Stores_filename: String
+DataLoader()
+ create_df_store_N_stocks(String):DataFrame
+ create_df_store_N_sales(String):DataFrame
+excel_2_csv(String)
+ loadExcel(String)
Predictor
+make_predict_sales_in_zao_general_prophet():Data
+make_predict_sales_in_store_general_prophet():Dat
+compute_lost_sales():DataFrame
+make_predict_demand_in_zao_general_prophet():D
+make_predict_demand_in_store_general_fb():DataF
+make_predict_demand_sku_fb():DataFrame
+make_predict_demand_sku_store_fb():DataFrame
Redistribution
+make_greedy_redistribution(DataFrame):DataFrame
Output
+ df_2_Excel(DataFrame)

9.0 Conclusions and future researches

In this thesis, we managed to implement a demand forecasting method based on time series, which was not previously used in shoe retail. It also developed software that is used by company B daily. A side effect of solving the forecasting problem was systematized data of the company, which allowed obtaining new analytical data for company employee.

A method for distributing goods to stores was proposed, based on the demand rating in stores. The influence of the human factor on the distribution of goods has been reduced. This method was used in the distribution of shoes from the summer collection of 2019.

Also developed and implemented strategies for replenishing the range in stores, based on the frequency of deliveries and the use of knowledge about inventory management. Made integration with software to predict demand in stores.

Heuristics were proposed and software was developed for the transfer of goods between stores. To test the heuristics in the near future an experiment is planned at one of the branches of company B.

The economic department has calculated that the economic profit from the introduction of methods for forecasting demand and replenishment was about 570,000 dollars. The decrease in lost sales was 2.1%.

Further studies plan to build infrastructure for working with large amounts of data and more accurate data collection. One of the goals is to reduce the demand forecast error for a large time horizon. It is also planned to introduce a new methodology for estimating lost sales. One of the priorities for the future is to optimize the work of production and warehouses. In the future, forecasting will be made at the “shop-article” level. This will further reduce the human influence on the distribution of goods. It is planned to introduce article clustering methods based on analogues based on computer vision algorithms.

In the near term for the implementation of all planned research, it is planned to establish a data science department, including a group of specialists in analytical logistics.

10.0 References

Armstrong, J., and Kesten Green. 2005. *Demand Forecasting: Evidence-Based Methods*.

Gardner, Everette S. 1985. "Exponential smoothing: The state of the art." *Journal of Forecasting $V 4* (1):1-28.

Gasparian, M. S., M. V. Karmanov, I. A. Kiseleva, V. I. Kuznetsov, and N. A. Sadovnikova. 2018. *Modeling of the demand forecasting*. Vol. 9.

Harvey, A. C., and S. Peters. 1990. "Estimation procedures for structural time series models." *Journal of Forecasting $V 9* (2):89-108.

Hastie, Trevor, and Robert Tibshirani. 1987. *Generalized Additive Models: Some Applications*. Vol. 82.

J. Taylor, Sean, and Benjamin Letham. 2017. *Forecasting at Scale*. Vol. 72.

Kollontai, Oksana. 2018. "Footwear production in Belarus continues to decline, despite rising consumption." accessed 24-04-2019. http://www.belmarket.by/proizvodstvo-obuvi-v-belarusi-prodolzhaet-sokrashchatsya-nesmotrya-na-rost-potrebleniya.

Li, Chongshou, and Andrew Lim. 2018. "A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing." *European Journal of Operational Research $V 269* (3):860-869.

Ma, Shaohui, Robert Fildes, and Tao Huang. 2016. "Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information." *European Journal of Operational Research $V 249* (1):245-257.

Mantrala, Murali, Michael Levy, Barbara Kahn, Edward Fox, Peter Gaidarev, Bill Dankworth, and Denish Shah. 2009. *Why is Assortment Planning so Difficult for Retailers? A Framework and Research Agenda*. Vol. 85.

Popova, Natalia. 2018. "What awaits the clothing and footwear market - forecast." accessed 24-04-2019. https://fashionunited.ru/novostee/beeznyes/chto-zhdet-rynok-odezhdy-i-obuvi-prognoz/2018061522170.

Reed, Travis J. 2012. "Visual Merchandising Magic." *Eyecare Business $V 26* (3).

Veiga, Claudimar Pereira da, Cássia Rita Pereira da Veiga, Weslly Puchalski, Leandro dos Santos Coelho, and Ubiratã Tortato. 2016. "Demand forecasting based on natural computing approaches applied to the foodstuff retail segment." *Journal of Retailing and Consumer Services $V 31*:174-181.

Wirex. 2019. "Why the traditional retail model is already dead." https://habr.com/ru/company/wirex/blog/437292/.