St. Petersburg State University

Graduate School of Management

Master in Business Analytics and Big Data

SALES PERFORMANCE IMPROVEMENT OF VELODRIVE COMPANY   
USING GEOANALYTICAL METHODS

Master’s Thesis by the 2nd year students:

Polina Abalakova

Irina Pendriak

Research advisor:   
Candidate of Technical Science,

Graduate School of Management

Olga N. Tushkanova

St. Petersburg

2023

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ ВЫПОЛНЕНИЯ

ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Я, Абалакова Полина Юрьевна, студент второго курса магистратуры направления «Бизнес-аналитика», заявляю, что в моей магистерской диссертации на тему «Повышение эффективности продаж компании «Велодрайв» с использованием геоаналитических методов», представленной в службу обеспечения программ магистратуры для последующей передачи в государственную аттестационную комиссию для публичной защиты, не содержится элементов плагиата.

Все прямые заимствования из печатных и электронных источников, а также из защищенных ранее выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Мне известно содержание п. 9.7.1 Правил обучения по основным образовательным программам высшего и среднего профессионального образования в СПбГУ о том, что «ВКР выполняется индивидуально каждым студентом под руководством назначенного ему научного руководителя», и п. 51 Устава федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» о том, что «студент подлежит отчислению из Санкт-Петербургского университета за представление курсовой или выпускной квалификационной работы, выполненной другим лицом (лицами)».

Text

Description automatically generated

(Подпись студента)

01.06.2023 (Дата)

STATEMENT ABOUT THE INDEPENDENT CHARACTER OF

THE MASTER THESIS

I, Polina Abalakova, (second) year master student, program «Business Analytics », state that my master thesis on the topic « Sales Performance Improvement of Velodrive Company Using Geoanalytical Methods», which is presented to the Master Office to be submitted to the Official Defense Committee for the public defense, does not contain any elements of plagiarism.

All direct borrowings from printed and electronic sources, as well as from master theses, PhD and doctorate theses which were defended earlier, have appropriate references.

I am aware that according to paragraph 9.7.1. of  Guidelines for instruction in major curriculum programs of higher and secondary professional education at St. Petersburg University «A master thesis must be completed by each of the degree candidates individually under the supervision of his or her advisor», and according to paragraph 51 of Charter of the Federal State Institution of Higher Education Saint-Petersburg State University «a student can be expelled from St. Petersburg University for submitting of the course or graduation qualification work developed by other person (persons)».

Text

Description automatically generated

(Student’s signature)

01.06.2023 (Date)

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ ВЫПОЛНЕНИЯ

ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Я, Пендряк Ирина Анатольевна, студент второго курса магистратуры направления «Бизнес-аналитика», заявляю, что в моей магистерской диссертации на тему «Повышение эффективности продаж компании «Велодрайв» с использованием геоаналитических методов», представленной в службу обеспечения программ магистратуры для последующей передачи в государственную аттестационную комиссию для публичной защиты, не содержится элементов плагиата.

Все прямые заимствования из печатных и электронных источников, а также из защищенных ранее выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Мне известно содержание п. 9.7.1 Правил обучения по основным образовательным программам высшего и среднего профессионального образования в СПбГУ о том, что «ВКР выполняется индивидуально каждым студентом под руководством назначенного ему научного руководителя», и п. 51 Устава федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» о том, что «студент подлежит отчислению из Санкт-Петербургского университета за представление курсовой или выпускной квалификационной работы, выполненной другим лицом (лицами)».



(Подпись студента)

01.06.2023 (Дата)

STATEMENT ABOUT THE INDEPENDENT CHARACTER OF

THE MASTER THESIS

I, Irina Pendriak, (second) year master student, program «Business Analytics », state that my master thesis on the topic « Sales Performance Improvement of Velodrive Company Using Geoanalytical Methods», which is presented to the Master Office to be submitted to the Official Defense Committee for the public defense, does not contain any elements of plagiarism.

All direct borrowings from printed and electronic sources, as well as from master theses, PhD and doctorate theses which were defended earlier, have appropriate references.

I am aware that according to paragraph 9.7.1. of  Guidelines for instruction in major curriculum programs of higher and secondary professional education at St. Petersburg University «A master thesis must be completed by each of the degree candidates individually under the supervision of his or her advisor», and according to paragraph 51 of Charter of the Federal State Institution of Higher Education Saint-Petersburg State University «a student can be expelled from St. Petersburg University for submitting of the course or graduation qualification work developed by other person (persons)».

 (Student’s signature)

01.06.2023 (Date)

Content

[Introduction 6](#_Toc136548552)

[Chapter 1. Bicycle industry analysis and problem statement 9](#_Toc136548553)

[1.1.Overview of the Velodrive business 9](#_Toc136548554)

[1.2. Overview of sporting goods market in Russia 13](#_Toc136548555)

[1.3. Problem description and expected results 17](#_Toc136548556)

[Summary of Chapter 1 19](#_Toc136548557)

[Chapter 2. Influence of geographic factors on business operations 21](#_Toc136548558)

[2.1. Sales performance factors 21](#_Toc136548559)

[2.2. Application of geomarketing in retail 22](#_Toc136548560)

[2.3. Retail location theory 24](#_Toc136548561)

[2.4. Retail site selection models 27](#_Toc136548562)

[2.5. Current analytical tools 34](#_Toc136548563)

[Summary of Chapter 2 37](#_Toc136548564)

[Chapter 3. Analysis of geofactors influence on Velodrive sales performance 38](#_Toc136548565)

[3.1. Technologies and tools used to extract geodata in the research 38](#_Toc136548566)

[3.2. Data collection and description 40](#_Toc136548567)

[3.3. Data preprocessing 47](#_Toc136548568)

[3.4. Velodrive assortment analysis 47](#_Toc136548569)

[3.5. Building a model to analyze geofactors’ importance and predict annual sales per 1 sq.m. 51](#_Toc136548570)

[3.6. Estimation of business effect 59](#_Toc136548571)

[3.7. Visualization of collected data and research results via Yandex DataLens 61](#_Toc136548572)

[Conclusion 63](#_Toc136548573)

[References 65](#_Toc136548574)

[Appendices](#_Toc136548575)

[Appendix 1. The function to geocode addresses of the stores](#_Toc136548576)

[Appendix 2. The loop for getting a dataset with competitors](#_Toc136548577)

[Appendix 3. Rental efficiency calculation](#_Toc136548578)

[Appendix 4. Spearman's Correlations](#_Toc136548579)

[Appendix 5. Spearman’s rho heatmap](#_Toc136548580)

[Appendix 6. The loop to get bootstrapped coefficients](#_Toc136548581)

[Appendix 7. Getting confidence interval for median bootstrapped coefficients using “bootstrapping” function from scipy.stats](#_Toc136548582)

[Appendix 8. Linear regression reports results from JASP (with assumptions compatibility check)](#_Toc136548583)

[Appendix 9. New locations](#_Toc136548584)

[Appendix 10. P&L statement draft for the new locations](#_Toc136548585)

# **Introduction**

In today's highly competitive business environment companies need to adopt innovative and effective strategies to stay ahead of their competitors. Sales performance is a crucial metric for any business, and companies need to constantly strive to improve this metric to remain profitable and sustainable. The bike industry is no exception, and companies in this field need to continuously analyze their sales data on different scales, including a geographical one, to identify areas of improvement and implement strategies to boost sales performance.

Geoanalytical methods offer a powerful tool for businesses to gain insights into their sales data by analyzing it in the context of geographic location. By incorporating location-based data into sales analysis, businesses can identify patterns and trends that might not be evident in traditional sales analysis. These insights can help companies to optimize their sales strategies, improve customer targeting, and ultimately boost sales performance.

Velodrive company has been operating in the Russian market since 2005 and now has around 30 shops in St. Petersburg, Moscow, Yekaterinburg and Izhevsk. It has a diverse range of products, including bikes and their spare parts, scooters, skateboards, skis and snowboards. Among the additional services, the company provides rental of bicycles and electric scooters, winter equipment, protection and child bike seats, repairs of any complexity in service centers and delivery. Despite the company's success, there is scope for improvement in their sales performance, particularly, in certain locations of Saint-Petersburg.

Velodrive company is facing challenges in improving its sales performance, and it is looking for innovative solutions to overcome them. One possible solution is to use geoanalytical methods to analyze the geographical data and gain insights into customer behavior and market trends. By doing so, the company can identify potential areas of growth, improve its sales strategy, and ultimately increase its profits.

The motivation for this master thesis is to investigate the potential of using geoanalytical methods to improve the sales performance of Velodrive company. This research involves analyzing financial data of the company and substantial amount of geographical data. The results of this analysis provide insights into how the company can improve its sales performance by targeting new areas of Saint-Petersburg and new types of retail locations.

Moreover, this research can contribute to the existing literature on sales performance improvement by applying geoanalytical methods to a real-world case study. The findings of this research can help other businesses facing similar challenges to understand how they can leverage geographic data to improve their sales performance.

The first chapter of thesis provides a comprehensive overview of the Velodrive company’s business, including its product offerings, distribution channels, marketing channels and financial performance. It is then followed by an overview of the industry in Russia, including analysis of trends in sporting goods market and bicycle market, in particular, and review of the government support of sport industry. This analysis sets the stage for understanding Velodrive's current sales performance and identifying opportunities for enhancement. Finally, we formulate the research problem, research questions and tasks.

In the second chapter, we review the existing literature on sales performance improvement strategies, with a specific focus on geoanalytical methods. The review provides a theoretical foundation for the research and highlights the retail location theories, models of optimal site selection and tools used for the solution of the location problem.

In the third chapter we detail the research methodology, which includes data collection, data preparation, data analysis, model building and estimation of business effect. The data is collected from various sources, including the company's sales database, geodata from 2GIS API, and data on target audience collected from DataLens Marketplace. The data preparation phase involves cleaning and transforming the data to make it suitable for analysis. This step implies significant amount of resources, especially, time, due to the nature of the data provided. The data analysis phase implies a combination of statistical and spatial analysis techniques to identify patterns and trends in the data. Then we build a multiple regression model to identify which factors influence the sales the most and make predictions for potential locations in other districts of the city. Finally, we create a 3-year-long financial model draft for evaluating chosen potential locations.

The results of the performed analysis are presented in the form of Jupyter notebooks, Excel spreadsheets, data visualizations and maps to help identify the most important geographic factors in sales performance and their influence. The thesis also mentions possible sales performance improvement methods that were identified through the analysis.

Overall, the goal of this master thesis is to determine the most important geofactors for the success of the point of sale (in terms of revenue), estimate their effect on the business of the company and provide valuable insights and location opportunities to the company, helping it to improve its sales performance and stay competitive in the market.

To achieve this goal, it is necessary to solve the following tasks:

1. To study the existing methodology and approaches to the problem;
2. To collect the required and available data from the company (sales, profits, retail locations, rental fees, corporate presentations on operational performance etc.);
3. To collect geodata from the available sources through APIs on the competition, target audience and other geofactors identified in the previous steps;
4. To process and analyze the received data with the analytical instruments and tools and statistical methods;
5. To build a statistical model to identify the most important geofactors;
6. Suggest potential locations and estimate their performance based on the predictions of the model.

# **Chapter 1. Bicycle industry analysis and problem statement**

## **Overview of the Velodrive business**

Velodrive is a network of specialized stores selling bicycles and sports goods. The network has over 20 stores in St. Petersburg, Moscow, Yekaterinburg, and Izhevsk. The company started its work in 2004 by opening a small tent near the "Chernaya Rechka" metro station in Saint-Petersburg, where they sold and repaired bicycles. In 2005, the company launched its own website, and this year is considered the year of the company's foundation. Thus, the company has been in the market for 17 years already.

The company's business model is diversified both in terms of product range and sales channels. They offer a broad selection of products for both summer and winter seasons, including bicycles, parts, accessories, scooters, electric scooters, balance bikes, hoverboards, skateboards, snowboards, sledges, inflatable sledges, skis, and snow scooters. The company also offers services for repairing bicycles, renting bicycles, scooters, snowboards and skis, seasonal storage of bicycles and snowboards.

Sales channels include:

1. Own retail stores. Velodrive focuses its main activities on retail through its own branded stores. Also, services for bicycle rental and service are provided based on the company's own stores. Part of Velodrive sales is carried out through an online store;
2. Online channel. Most of the online sales come from Velodrive website. However, the company has already expanded into online marketplaces, such as Ozon, Wildberries and Yandex Market, and is striving to boost sales through this channel;
3. Franchise. The company also develops franchising;
4. Wholesale. Velodrive also acts as a wholesale distributor for its franchisees and other market participants. The largest clients are Yandex Lavka, Samokat, Bushe - companies with well-developed courier delivery services.

The company has several store formats: Kids — with an area of 100-150 m2, Bicycle Center — 150-300 m2, Cycling World — over 300 m2. Currently, Velodrive network is represented by 9 self-run stores and 14 points of sale managed by franchisees.

The company has a functional organizational structure which is illustrated in Figure 1. This type of organizational structure implies that the company is divided into smaller departments or units based on the functions or skills required for the job. There are several advantages and disadvantages to this type of organizational structure.

Advantages:

1. Specialization: In a functional organizational structure, employees are grouped by their areas of expertise, which allows them to focus on their specific job responsibilities. This specialization often leads to increased efficiency and productivity, as employees can develop a deep understanding of their roles and responsibilities;
2. Economies of scale: Functional structures can also allow for economies of scale, as similar functions can be grouped together, allowing the company to leverage its resources more efficiently;
3. Efficient use of resources: Because functional structures are designed to group employees based on their areas of expertise, there is less duplication of effort and fewer resources are wasted.

Disadvantages:

1. Silos: The functional organizational structure can sometimes lead to silos, where departments become too focused on their own goals and objectives and lose sight of the larger company mission;
2. Limited communication: Silos can also lead to limited communication between departments, which can make it difficult to share information and ideas across the company;
3. Slow decision-making: Functional structures can sometimes result in slow decision-making, as decisions need to be approved by multiple departments before they can be implemented;
4. Lack of flexibility: Because employees are grouped based on their areas of expertise, it can be difficult to reorganize or shift resources when needed, which can limit the company's ability to respond to changes in the market or new opportunities.

**Figure 1.** Organizational chart of Velodrive company

Velodrive company uses a variety of promotion channels to reach its target audience. One of the primary promotion channels is through sales promotion campaigns, such as "Cyber Month", "Bike Carnival", "Bike Night Sale", and "Velodrive's Birthday", as well as promotions for special occasions like Children's Day and Russia Day.

In addition to sales promotions, Velodrive also utilizes various advertising channels, including SMM and a YouTube channel. The company also advertises on product platforms such as Yandex Market, Google Shopping, and Blizko, as well as through outdoor advertising, such as billboards. Velodrive also engages in cross-promotion with other businesses such as Dodo-Pizza, Fitness House, and Mama Roma.

Velodrive pursues multichannel advertising approach. The company utilizes traditional advertising methods, such as advertising in newspapers and on radio. Apart from that, it also invests in online advertising campaigns through search engines like Yandex and Google. Finally, the company also organizes various events to promote its brand, such as Big Bike Festival which is organized by Velodrive, "ZSD – fest", "To work by bike", "Big bike parade" and "Opening of the cycling season".

****As for the financial performance of Velodrive, the company’s revenue had been growing steadily enough until 2021, when its revenue decreased significantly by 11% to the level of 553 million rubles. The COVID-19 pandemic had a positive impact on the company’s sales as people became more interested in the ways of passing their leisure time, namely in sports activities in this case. The dynamics of the company’s revenues for the last 6 years are shown in Figure 2. According to the company’s internal data, gross profit margin of the Velodrive business was registered at the level of 35% for the studied period, while net profit margin was in the range from 1% to 5%.

**Figure 2.** Basic financial data of Velodrive. Revenues and growth rates for 2017-2022

Source: corporate presentations of Velodrive based on the Financial Statements

One of the limitations that we encounter with, is that Velodrive is not a publicly traded company, and they are not obliged to disclose their financial information. We can only receive data from the company’s representatives or open sources, and not from the official financial statements.

In summary, Velodrive is a sporting goods retailer that stands out in the market due to its strong brand, long-term strategy, and sustainable business model. The company's well-established brand name has been built through a wide retail network and numerous franchisees, making it one of the market leaders in St. Petersburg. Additionally, Velodrive has a clear development strategy for the coming years that is expected to increase its market share and improve business efficiency. Moreover, Velodrive's sustainable business model, which is diversified both in terms of the range of goods and services and sales channels, makes it stand out in the competitive cycling industry. The company's commitment to providing quality bicycles, accessories, winter sports goods, and excellent services through various channels, including retail stores, online stores, wholesale supplies, and franchise, demonstrates its customer-centric approach.

## **. Overview of sporting goods market in Russia**

**1.2.1. Sports goods market overall**

The demand for bicycles, either to buy or to rent, in Russia is driven by several factors, including a growing interest in health and fitness, the desire to reduce congestion and pollution in urban areas, and the availability of new models with advanced features. Moreover, the COVID-19 pandemic has accelerated the demand for bicycles as people looked for alternative modes of transportation and outdoor activities.

The overall sporting goods market in Russia nowadays, just like the total market in the country, is being influenced by the external factors, namely, the exits of numerous international brands from the Russian market. The above trend appeared in spring 2022 and from that moment on the local buying patterns are being transformed and restructured. Before the adaptation of the Russian producers, distributors and businesses to the newly existing business setting, the market most often shows negative dynamics in sales both in absolute and money values. That concerns not only the general tendence in the market, but the sporting goods evolution in Russia in particular.

Sporting goods category include such products as:

1. Sports clothes, shoes and outfit;
2. Sports equipment for various types of activities;
3. Sports accessories;
4. Nutrients and alimentary supplements;
5. Other categories.

According to the research of Tinkoff Data, the sales of sporting goods in Russia decreased by 43% year-on-year in terms of quantity and by 40% in terms of monetary value in 2022. The portal “Shopper’s”, dedicated to the research and news in retail market, suggests that online sales of sports goods declined by 55% quantitatively and by 51% in money value.

As to the online channel specifics of sports goods sales, the research of Data Insight company showed that for the annual period (from July 2020 till June 2021) the sales increased by 57% compared to the analogical previous period and constituted 81,5 billion rubles. In unit terms the sales reached 35,6 million demonstrating a 73% growth rate (Data Insight, 2022).

Data Insight research show that most of the sports goods online consumption (~27%) comes from Moscow and Saint-Petersburg (see Figure 3). Krasnodar and Ekaterinburg account for 5% of sales of this category of retail goods (Data Insight, 2022).



**Figure 3.** Distribution of online sales of sports goods by city in monetary values in 2019

The leaders on the market of sporting goods in 2021 were:

1. Sportmaster (129 billion rubles);
2. Adidas (42 billion rubles);
3. Decathlon (28 billion rubles) (Sobol & Kiseleva, 2022).

In 2022 Adidas and Decathlon left the Russian market which impacted favorably the business of Sportmaster and other local stores, however, we do not have the information yet on how the market shares redistributed.

After the exit of international sports brands the consumers shifted their shopping experience towards online channels. The traffic of people visiting offline stores, shopping centers, has been decreasing since March 2022. Local online marketplaces, such as Wildberries and Ozon, demonstrated a 2,6 times growth in the category of sporting products in the first half of 2022 versus the analogical period of the previous year and reached 26,2 million units (Sobol & Kiseleva, 2022).

In connection with the above-mentioned exit of brands, the question of the location of a retail store becomes more and more relevant. International players left and shopping centers are now looking for new lenders of their facilities. In these circumstances it is essential to understand where the opening of a retail point of sale would be more beneficial. According to director of a sports retail chain Demix Tatiana Ivanova, the most favorable is the co-location with the sporting and sportstyle brands popular among young people, such as Cropp, Sinsay, Urban Vibes, Street Beat etc. (Ivanova, 2022). And frequently a previously taken by “Adidas” place becomes a very attractive point of a store location. Thus, the state of the art on the market presents not only challenges, but also opportunities, at least in terms of store locations opportunities.

**1.2.2. Bicycles market**

According to the research of BusinesStat held in 2022 “The analysis of the bicycle market in Russia”, in 2017-2019 the sales of bicycles in the country were recovering after the fall during the crisis of 2014-2015. For the period of 2017-2019 the sales of bicycles grew from 4,16 million to 4,35 million units. The market recovery was promoted by the governmental aid, popularization of cycling sport and wider range of bicycle assortment presented in retail (BusinesStat, 2022).

During the COVID-19 pandemic the unit sales of bicycles in Russia decreased by 13% versus the level of 2019 – to 3,79 million units (see Figure 4) (BusinesStat, 2022). A bicycle was not a prime necessity, so people decided to postpone the purchase in the conditions of lower incomes. However, the sales could have decreased even more significantly – this was prevented by the decision of people to avoid public transport and their pursuit to spend their free time actively while keeping social distance.



**Figure 4.** Sales of bicycles in Russia in 2017-2021 in millions of units

Source: BusinesStat “The research on the bicycle market in Russia”, 2022

In 2021 the previously deferred demand was exercised. The sales of bicycles grew by 18% compared to 2020 and reached 4,48 million units.

It is important to notice that the bicycle market in Russia is characterized by a high level of dependence on the import supplies. According to the above-mentioned research by BusinesStat, the imported goods account for around 70% of bicycle sales in Russia (BusinesStat, 2022). The bicycles produced in Russia are constructed with the imported spare parts – depending on the degree of production localization – such as group sets and moving parts, brakes, boings, tyres, frames, steering tubes.

The sanctions imposed on Russia do not directly restrict the supplies of bicycles or their spare parts into the country, however, many brands and producers decided to quit the market themselves to support relations with the clients in the countries supporting the sanctions. Thus, the Russian market will, obviously, require some time to restructure the supply chain and build up the deliveries from other countries (e.g., China). According to experts, longer supply chain in bicycle spare parts deliveries will provoke significantly higher prices.

The opportunities indicated for the overall sports products market is also applicable for bicycle sales’ businesses. There are quite a few solid options to locate a new point of sale efficiently or to relocate the existing premises. Yet, a thorough analysis helping to choose such a location is necessary to take the right and data-driven decision.

**1.2.3. Government support for sport and physical activity**

The sports industry receives active support from the government, as evidenced by the adoption of the "Development of Physical Culture and Sports" state program in 2014. This program is intended to run until 2030 and aims to increase the proportion of citizens who regularly participate in physical activities and sports to 70% by that year (Ministry of Sports of the Russian Federation, 2014). For the years 2022 to 2024, the program will be financed with a total of 195 billion rubles (Ministry of Sports of the Russian Federation, 2014).

The City Parking Management Center of St. Petersburg considers the development of bicycle infrastructure to be a crucial task. In 2017, the Bicycle Petersburg project was included in a list of priority programs aimed at providing a balanced transport environment and convenient car-free movement within the city. As a result of the comprehensive work carried out, the length of bicycle lanes in the Northern Capital has tripled.

Currently, cyclists have access to 37 bike routes covering a total distance of 135.4 km (Administration of St. Petersburg, 2023), making it safer and more comfortable for citizens to travel between their homes or workplaces and public transport or metro stops. Moreover, there are now 20 intercept parking and 13 city parking lots with free bike parking, allowing for easy transfer to public transportation.

The city aims to increase the length of bike paths to 300 km by 2030 (TASS, 2021). In addition to creating bike infrastructure, the Transport Committee collaborates actively with public organizations and regularly participates in events to promote cycling and exchange experiences with foreign colleagues.

The popularization of sports and the creation of infrastructure for physical activity can have a positive impact on the sporting goods market. As more people engage in physical activities and sports, the demand for equipment and apparel related to these activities can increase. This can create new opportunities for sporting goods manufacturers and retailers to expand their product lines and increase sales.

## **1.3. Problem description and expected results**

The problem at hand is to determine the optimal retail location for a bicycle company Velodrive based on various geofactors. The goal is to identify a location that maximizes potential sales while minimizing operational costs. Even though we do not look specifically at the profits of the company and focusing our attention on sales, while choosing an optimal location we cannot but take into account the rental expenses and profitability of the rented facilities.

The preliminary research and intuition suggest that the studied problem is more complex than it seems. The company may need to consider various geographic factors such as population density, traffic patterns, demographics, and accessibility to public transportation. It is important to select a location that is easily accessible and visible to potential customers, and that is in close proximity to bike paths or other outdoor recreational areas such as parks, lakes etc.

Additionally, Velodrive needs to consider the competition in the area, as well as the potential logistics expenses and availability of skilled labor. The company must also take into account the zoning and permitting requirements of the local government.

The challenge in this problem is to effectively balance these different factors to identify the best location for the business. An effective analysis of these factors can help Velodrive to identify a location that will optimize its revenue potential and improve its overall business performance.

Apart from the stated problem the company also approached us with the issue of sales’ seasonality. Seasonality is the industry-spread phenomenon as bicycle, and especially their offline sales, are season-dependent. In our research and analyses we try to touch this aspect, however, do not consider it as a target problem to which we are looking for a solution.

To prepare the solution for the above-mentioned problems and based on the research gap, the following **research questions** were formulated:

1. How does the current sales performance look like – overall and by store? What are the peculiarities of the bicycle market sales and Velodrive in particular?
2. Which factors currently impact the management decision on stores’ location?
3. What are the rental fees that the company incur in each retail location?
4. Which geo factors might influence business performance in general? Which factors may be applicable to the bicycle industry in particular? How are they ranged?
5. Apart from rental fees, are there any other expenses that differ based on a store location? Should we take them into consideration?
6. Based on the answers to the previous questions, what are the most suitable locations in Saint-Petersburg for opening new bicycle shops?
7. Does the company have necessary resources to apply the geostrategy recommendations we will come up with? That is, are they feasible?
8. In connection with seasonality issue the below questions are formulated:
9. Which categories are most popular among customers (apart from the core product – bicycles)?
10. What products are in demand all year round?
11. Is it possible to increase the sales of non-dependent on the season articles?

To answer the formulated questions the following research tasks are set:

1. To perform a literature review on the topic;
2. To study the existing methodology and approaches to the problem;
3. To collect the required and available data from the company (sales, profits, retail locations, rental fees, corporate presentations on operational performance etc.);
4. To process the received data, i.e., to transform it into a machine-readable format;
5. To analyze the data with the analytical instruments and tools and statistical methods;
6. To define patterns and peculiarities of the company’s sales overall and by store in terms of different categories’ performance;
7. To study the existing locations’ specifics and visualize them on a map;
8. To collect geodata from the available sources through APIs on the competition, target audience and other geofactors identified in the previous steps;
9. To estimate the performance of each store in terms of the profit/square unit ratio;
10. Depending on the size of the Velodrive chain and available data, either build a statistical or an ML model;
11. Based on the previous steps results, come up with the recommendations on the optimal retail locations.

As a result of our analysis, we are aiming to come up with several most important geographic factors influencing the sales performance of Velodrive and taking into account the above-mentioned metrics of location attractiveness. Apart from that, we would like to suggest the company the optimal location areas and places taking into account the ratio of sales revenue to 1 unit of the rental square.

We would also like to present a dashboard with the current and preferred locations to the company using DataLens tool.

As a side result of our work, we plan to introduce some suggestions on how to tackle the seasonality issue.

## **Summary of Chapter 1**

To sum up, in this chapter we covered the overview of the studied company Velodrive, the essence of its business, fundamental information on their operations, such as organizational structure, basic financials, the size of the network and its business model. We also dived into the bicycle market in general to better understand how the industry overall operates and what is the current state-of-art on the Russian market. We examined how the COVID-19 and the exit of the global companies have impacted the industry. We examined the competition on the bicycle market in Russia, looked at their key performance indicators. To better understand the potential and prospect of development of the market, we looked at the governmental support of the industry and physical activity in the country.

Finally, we formulated the research questions and set the tasks necessary to answer those questions. We envisaged the results that we expect to achieve after the execution of our analysis.

# **Chapter 2. Influence of geographic factors on business operations**

## **2.1. Sales performance factors**

Sales performance is a critical aspect of retail management, and understanding the factors that influence sales can help retailers make informed decisions about marketing, merchandising, and other aspects of their operations. Some of the key factors that can impact sales performance in retail include (Berman & Evans, 2018):

1. Store design and layout: The design and layout of a store can have a significant impact on sales performance. Factors such as the placement of products, the use of signage and displays, and the overall atmosphere of the store can all influence how customers perceive and interact with the space;
2. Product assortment and pricing: The selection of products available for sale, as well as their pricing, is another important factor in sales performance. Retailers must consider factors such as consumer demand, product availability and quality, and competitive pricing when making decisions about their product assortment;
3. Marketing and advertising: The effectiveness of a retailer's marketing and advertising campaigns can have a significant impact on sales performance. This includes factors such as the use of social media, email marketing, and other digital channels, as well as traditional advertising methods such as print ads and radio spots;
4. Staff training and management: The quality of customer service provided by a retailer's staff can also influence sales performance. Training programs that focus on topics such as product knowledge, customer engagement, and problem-solving can help retailers build strong relationships with their customers and drive sales;
5. Location and foot traffic: The location of a retail store can also play a role in sales performance, as stores in high-traffic areas are more likely to attract potential customers. Retailers must consider factors such as the presence of nearby competitors, accessibility, and visibility when selecting a location for their store.

In our research we focus in more details on the last point of discussion - location factor and geoanalytics in retail sales performance.

The topic of the use of the geoanalytical tools in the improvement of business operations has been discussed in several relatively comprehensive books and articles. The coverage of the topic cannot be performed without the knowledge of several key concepts developed in connection with the influence of geoanalytics on the business performance.

## **2.2. Application of geomarketing in retail**

Geomarketing is a relatively new field that combines the principles of geography, marketing, and data analysis to help businesses make better decisions about their marketing strategies. Geomarketing is the application of geographical information systems (GIS) to the field of marketing. According to (Cliquet, 2013), "Geomarketing is the analysis of the spatial distribution of customers, products, and services, and the use of this analysis to design and implement marketing strategies that take into account the geographic context." This means that geomarketing involves analyzing geographic data such as maps, demographics, and customer behavior to develop marketing strategies that are tailored to specific regions.

Geomarketing is a process that uses location-based data to provide insights into market trends and consumer behavior. Geomarketing combines geographic information system (GIS) and marketing tools to analyze customer data and geographic data (Cliquet, 2013). The primary goal of geomarketing is to provide insights into market trends and consumer behavior to help businesses make informed decisions about their marketing and store location strategies.

Connecting location with information is a widely accepted practice for businesses when making decisions. Geographical considerations are involved in selecting a site, targeting an audience, focusing on market segments, scheduling distribution networks, allocating resources, responding to emergencies, and more. Various stakeholders, such as marketers, managers, retailers, real estate professionals, insurers, asset managers, health organizations, urban consulting agencies, travel agencies, and others, are seeking to gain a better understanding of their markets. Geomarketing supports these stakeholders in many ways, including marketing, business decisions, analysis, and research.

Geomarketing can be used to solve the following analytical tasks:

1. Territorial planning that can be viewed at macro and micro levels - within the macro level the selection of the most promising territorial location for business is carried out within the boundaries of a state. At micro level, geomarketing can be used for territorial planning within specific geographic settlement areas;
2. Direct marketing - geomarketing is used to determine the geographic location of the target consumer group;
3. Socio-demographic analysis which includes traditional parameters such as gender, age, income level, family composition, also involves spatial data mapping to identify consumer preferences within specific ‘habitation zones’ (Ramadani, Zendeli, Gerguri-Rashiti, & Dana, 2018);
4. Market analysis - geomarketing enables businesses to assess their competitors and infrastructure with a geographic focus, considering both vehicle and pedestrian traffic. This approach allows for the creation of more precise models and identification of patterns in consumer behavior;
5. Advertising and media planning - geomarketing is used to determine the size, direction, and traffic of a flow, but most importantly to target it. Advertisements should be placed in places where the target audience is concentrated, particularly in places like neighborhoods, transportation routes, and bus stops (Banerjee, 2019);
6. Location analysis - this task involves the best location for a business based on specific criteria like target audience, geographical and temporal concentration of consumers, and competition analysis. The task can be also viewed from another perspective which includes determination of the optimal product range or activity direction within a given price range based on the existing consumer concentration and traffic, as well as competition analysis;
7. Risk analysis includes the calculation of possible risks associated with deviating from the optimal solutions that were identified on the previous steps.

When choosing a store location, businesses use geomarketing to identify the best location based on several criteria such as accessibility, visibility, and market potential. Geomarketing can help businesses to identify potential customers in the area, understand the competition, and assess the potential profitability of the location. For example, geomarketing analysis can help retailers to identify the best locations for their stores based on factors such as the density of the target audience, the number of competitors in the area, and the availability of public transportation (Saini & Bansal, 2023). Research indicates that geomarketing can have a beneficial impact on the growth of businesses, provided that its key factors, including location, industry, socio-demographic variables, and business factors, are thoroughly examined and considered in the decision-making process (Ramadani, Zendeli, Gerguri-Rashiti, & Dana, 2018).

One of the main approaches to geomarketing for store location selection is the use of GIS. GIS is a system that allows users to analyze, store, and manage geographic data. GIS can be used to analyze different types of data such as population density, age groups, income levels, and lifestyle characteristics. By combining these data sets, businesses can identify the best locations for their stores based on the specific needs and preferences of their target customers. For example, a retailer can use GIS to identify areas with high population densities and high-income levels to determine the best locations for their stores.

Another approach to geomarketing for store location selection is the use of customer data. Customer data can provide insights into consumer behavior and preferences, which can be used to identify the best locations for stores. Customer data can be collected through various methods such as surveys, loyalty programs, and social media. By analyzing this data, businesses can understand their target customers' shopping patterns, lifestyle preferences, and brand loyalty, and identify the best locations for their stores based on these factors (Cliquet, 2013).

Despite its many benefits, geomarketing also poses several challenges and limitations that businesses need to consider. One limitation of using geomarketing for store location selection is that it relies on accurate and up-to-date data. The accuracy of the data used in geomarketing analysis can affect the results of the analysis. For example, inaccurate or incomplete data can lead to incorrect conclusions about the best locations for stores. Therefore, businesses must ensure that they have access to accurate and up-to-date data to make informed decisions about their store locations (Saini & Bansal, 2023). Another important challenge of geomarketing is the need for accurate and reliable spatial data, which can be difficult to obtain and maintain. In addition, geomarketing requires specialized skills and expertise in areas such as GIS, spatial analysis and data visualization, which can be costly and time-consuming for businesses to acquire. Furthermore, geomarketing also raises ethical and privacy concerns related to the collection, storage, and use of location-based data, which businesses need to address and comply with relevant laws and regulations.

## **2.3. Retail location theory**

Retail location theory is a branch of spatial economics that seeks to explain the location choices made by retailers. It is concerned with understanding the factors that influence the choice of retail locations, and how retailers can optimize their location strategies to maximize profitability.

**2.3.1. Central place theory**

There are several different theories of retail location, but one of the most well-known is the central place theory, developed by Walter Christaller in the 1930s (Brown, 1993). According to this theory, retail locations are organized in a hierarchical structure, with larger, more specialized retailers located in larger cities or central locations, and smaller, more general retailers located in smaller towns or suburban areas (Berry, 1962). This hierarchy is based on the concept of range and threshold, which refer to the geographic area that a retailer can effectively serve and the minimum level of demand necessary to sustain a business, respectively.

Another important aspect of central place theory in retail location is the principle of minimum differentiation. This principle states that retailers will locate in such a way as to minimize competition and maximize profits. Retailers will avoid locating too close to each other to prevent cannibalizing each other's sales, but they will also avoid locating too far apart to avoid losing potential customers (Brown, 1993).

**2.3.2. Spatial interaction theory**

Another key theory in retail location theory is the spatial interaction theory (Brown, 1993). Developed in the 1950s and 1960s by William Alonso, Edgar Mills, and others, the theory explains the location of retail businesses based on accessibility, market potential, and competition. Accessibility refers to the ease of reaching a particular location, while market potential refers to the size and composition of the local market. Competition refers to the presence of similar businesses in the same location.

Spatial interaction models illustrate how different locations are functionally interdependent. This can be seen by considering a simple interaction model (Rey, 2001) reflected in the following formula (1):

where:

represents the amount of interaction between locations i and j;

and represent the attributes of the two locations (such as population, economic activity, or cultural significance);

represents the distance between the two locations;

β is a parameter that determines the degree to which distance affects interaction.

This equation illustrates the basic principle of spatial interaction theory: the interaction between two locations is a function of the distance between them and the attributes of the locations themselves.

Spatial interaction models can be used to predict the location of retail stores, based on the demographics and behavior of consumers in the surrounding areas (Piovani, Molinero, & Wilson, 2017).

**2.3.3. Bid rent theory**

Bid rent theory (Ladle, Stiller, & Stiller, 2009) is another important theory in retail location. Developed in the 1960s and 1970s by William Wheaton, the theory explains the value of land as a function of its location and the revenue it can generate. According to this theory, retail stores will pay higher rents for prime locations that offer better access to consumers and higher sales potential. This theory also explains why retail stores are often clustered together in areas with high foot traffic, to take advantage of the benefits of agglomeration (Brown, 1993).

**2.3.4. Minimum differentiation principle**

The principle of minimum differentiation is another important concept in retail location theory, which states that firms aim to locate their stores as close as possible to each other while still maintaining a minimum distance to avoid cannibalization of sales (Hotelling, 1929). Eaton and Lipsey present new developments in the theory of spatial competition that challenge the traditional interpretation of the minimum differentiation principle. The authors argue that the traditional interpretation of the minimum differentiation principle assumes that firms compete only on the basis of price (Eaton & Lipsey, 1975). However, in reality, firms may also compete on the basis of quality, location, and other factors. Eaton and Lipsey propose a model that incorporates these additional dimensions of competition and show that the minimum differentiation principle may not always hold true. Specifically, they show that under certain conditions, firms may prefer to locate their stores further apart from each other to differentiate themselves in terms of quality or other factors. This principle explains why retail stores often specialize in particular product categories or offer unique value propositions to differentiate themselves from competitors (Vandell & Carter, 1993).

**2.3.5. Percolation theory**

Recent advances in retail location theory have focused on incorporating new technologies, such as big data, machine learning, and spatial analysis, into traditional models. For example, Piovani et al. (Piovani, Molinero, & Wilson, 2017) use percolation theory and spatial interaction modeling to predict the optimal location and clustering of retail stores in urban areas. Percolation theory is a mathematical framework that has been used to study the connectivity and accessibility of systems, such as electrical networks and porous materials. In the context of urban retail locations, percolation theory can be used to understand the spatial distribution of stores and the connectivity between them. The authors of the article combine percolation theory and spatial interaction modeling to develop a framework for understanding urban retail locations. They use data from a retail complex in Milan, Italy to test their model and find that it accurately predicts the distribution of stores and the connectivity between them.

Other factors that influence retail location decisions include demographics, competition, accessibility, and cost. Retailers may choose locations that are in close proximity to their target customers, that have lower competition, that are easily accessible by car or public transportation, or that offer lower rent or other cost advantages.

In practice, retailers use a variety of tools and techniques to analyze and optimize their location strategies. This may include using GIS software to map and analyze customer demographics and behavior, conducting site surveys and market research, and using predictive analytics to forecast sales and demand.

To sum up, retail location theory provides a useful framework for understanding the factors that influence the location choices made by retailers, and how retailers can use this information to optimize their location strategies and maximize profitability.

## **2.4. Retail site selection models**

It has always been assumed that the key variables of a successful retail distribution company are location, location, location, so it shows the importance of a proper location strategy for retailers (Ghosh & McLafferty, 1982; Cliquet G. , 2020). In retail companies, the opening of a new store or outlet carries an inherent risk because of the high financial costs connected with it. Also, a store that is unsuccessful due to a poor choice of location can have a significant negative impact on the image of the company. Consequently, the analysis of location is vital for commercial enterprises and retail companies in particular.

The study of the problem of optimal site selection has led to the development of several methods and models.

**2.4.1. The analog model**

The analog model, also known as the "comparable store analysis," is a location selection method used in retail to identify potential store locations by comparing them to existing successful stores. This model assumes that the characteristics that made a store successful in one location will also be successful in similar locations.

The Analog Model constituted the first attempt at a formal retail site selection process (Applebaum, 1966). The creator of this model, William Applebaum, focused on the study of existing retail stores to identify potential retail sites. Customers of these existing stores were interviewed to determine where they lived, allowing Applebaum to define primary trade areas for these stores. He used existing store sales levels to project sales potential of future locations. To determine the likely performance of the planned store, he performed a systematic comparison of the characteristics of the proposed store with the characteristics of the existing ‘analogue’ store (Buckner, 1998; Berman & Evans, 2018).

One advantage that the Analog Model has over other methods is its adaptability to assess virtually all types of retail stores. By contrast, the Gravity Model is used primarily with supermarkets and drugstores (Buckner, 1998). However, the Analog Model has some weaknesses:

1. It is highly subjective and, typically, does not work well without an experienced analyst;
2. Developing and maintaining the database through a well-trained staff has a relatively high cost (Buckner, 1998);
3. It assumes that customer behavior is static and does not consider external factors such as changes in consumer preferences or economic conditions;
4. The analog model does not account for the fact that a successful store may have unique characteristics that are not easily replicated in other locations (Berman & Evans, 2018).

**2.4.2. The Huff model**

The Huff model, also known as the Huff gravity model, is a geographic information system (GIS) tool used to estimate the probability that a consumer will visit a retail location based on its proximity and attractiveness compared to other nearby locations. The model is named after its creator, David Huff, who introduced it in a 1963 paper titled ‘A Probabilistic Analysis of Shopping Center Trade Areas’.

The basic idea behind the Huff model is that the probability that a consumer will choose to visit a particular retail location is proportional to the location's attractiveness, and inversely proportional to the distance between the consumer's location and the retail location (Suárez-Vega, Gutiérrez-Acuña, & Rodríguez-Díaz, 2015). The model can be expressed mathematically as follows (Huff, 1963; Sevtsuk & Kalvo, 2018):

where:

is the probability that a consumer at location i will visit location j;

is the attractiveness of location j (e.g., the size of the store or the quality of the products offered);

is the distance between location i and location j;

α is an attractiveness parameter. The α constant is calculated by summing the attractiveness of all locations and then dividing each location's attractiveness by the sum. This ensures that the probabilities for all locations add up to 1;

β is a scaling parameter that adjusts the distance decay function;

n is the total number of stores, including store j.

The output of the Huff model is a map of predicted consumer demand for each retail location, which can be used to make informed decisions about site selection, marketing, and resource allocation.

While the Huff model is a simple and intuitive tool for analyzing consumer behavior, it has some limitations. For example, it assumes that consumers make decisions based solely on distance and attractiveness, and does not account for other factors such as price, quality, or brand reputation (Cliquet G. , Geomarketing: Methods and Strategies in Spatial Marketing, 2013). Additionally, it assumes that consumer preferences and behavior are static over time, which may not be the case in reality. Despite these limitations, the Huff model remains a popular and widely used tool for location analysis and site selection in the retail industry.

**2.4.3. Multiplicative competitive interaction model**

The evolution of retail location theory has led to the development of new models and concepts. The (MCI) model is an econometric model that analyzes market shares and/or market areas in a competitive environment. It was first introduced by Nakanishi and Cooper in 1974 and 1982. The model is designed to help businesses make decisions about market segmentation, targeting, and positioning by providing insights into how customers and suppliers interact with each other in different submarkets (Wieland, 2017).

The MCI model builds on the Huff model, which is a spatial interaction model that predicts the probability of a customer choosing a particular store based on the store's attractiveness and the customer's distance from the store. The MCI model extends the Huff model by allowing for multiple submarkets and multiple suppliers in each submarket. The MCI model can be described as follows (Nakanishi & Cooper, 1974; Bekti, Pratiwi, & Jatipaningrum, 2018):

where:

– probability that a consumer living at i chooses retailer j;

– measure of variable k that describes the pull on consumers of retailer j;

– sensitivity parameter with respect to variable k;

q – total number of variables k considered in the measure of pull on consumers;

– distance between consumer location i and retailer j;

β – sensitivity parameter with respect to distance;

n – number of retail firms considered by the consumer living at i.

This MCI model was used in several studies. One of them was conducted by A. Baviera-Puig, J. Buitrago-Vera and C. Escriba-Perez, they were developing a geomarketing model using the MCI model to help managers design supermarket location strategies based on shop features, competitors, and environment. The model presented combines objective and subjective variables to assess outlets and their trade areas. The subjective component is evaluated by experts, while the objective component describes the characteristics of the outlet and its trade area. The results of the subjective Multiplicative Competitive Interaction (MCI) model reveal that distance (100%) has the greatest impact on the dependent variable, followed by ease of pedestrian access (32%), unemployment rate (30%), separation rate (29%), growth of competitors' market share (27%), brand recognition (21%), aggressiveness of competitors' strategies (20%), years of operation (23%), and number of checkouts (20%) (Baviera-Puig, Buitrago-Vera, & Escriba-Perez, 2016).

**2.4.4. Multiple regression model**

Statistical modelling can also be used in the retail site selection process. One of such instruments is a multiple regression model. It can predict numerous target outcomes such as the volume of sales of a potential store in terms of quantity or revenue or number of customers that will be attracted by the store (Morphet, 1991; Taylor, 2015). With the help of different statistical software or programming languages it can be possible not only to determine the target variable values, but also determine the important features that can influence the level of sales, for example. However, before looking at the statistical significance of different explanatory variables it is important to start with collecting data on them.

The following factors (variables) can be included in the model structure:

1. Demographics: Understanding the demographic characteristics of the target market is crucial for retailers to determine the potential customer base for a location. This includes factors such as age, income, education level, and household size;
2. Traffic and accessibility: The amount of foot and vehicular traffic a location receives is an important consideration for retailers, as it affects the level of visibility and accessibility of the store. Factors such as the presence of nearby highways or public transportation, parking availability, and pedestrian flow are important to consider;
3. Competition: The level of competition in a given area is another important factor in retail site selection. Retailers must consider the number and types of competitors in the area, as well as their pricing and marketing strategies, in order to make informed decisions about where to locate;
4. Site availability and cost: The availability of suitable sites for retail locations, as well as the cost of acquiring or leasing those sites, is a key consideration for retailers. Other factors such as zoning regulations, building codes, and utility access may also come into play.

The above factors are in essence the criteria influencing the choice of the site location. Undoubtedly, there are quite a few other factors that are hard to enumerate. One of such obvious examples are market trends. Keeping abreast of market trends and changes in consumer behavior is crucial for retailers to remain competitive. This includes factors such as shifts in shopping patterns, changes in consumer preferences, and emerging technologies that may affect how customers interact with retailers.

**2.4.5. Spatial-allocation model**

Spatial-allocation models are a type of location selection model used in the retail industry to determine the optimal placement of stores based on customer demand and accessibility. These models take into account various factors such as demographic information, transportation accessibility, and competition in order to identify the most suitable locations for new stores (Ladle, Stiller, & Stiller, 2009) and can be perceived as an advancement of multiple regression models.

The spatial-allocation model is a complex and iterative process that involves a number of steps. The first step is to collect data on customer demographics and behavior, as well as data on the physical and economic characteristics of the potential store locations. This data is typically obtained from a variety of sources, including government census data, market research reports, and customer surveys.

Once the data has been collected, it is used to create a spatial allocation model that takes into account the various factors that influence customer demand and accessibility. The model typically involves the use of statistical techniques such as regression analysis and spatial autocorrelation analysis to identify patterns and relationships in the data.

The spatial-allocation model then generates a set of candidate locations for new stores, which are evaluated based on their predicted sales potential and overall profitability. The model may also consider factors such as the availability of real estate and the cost of construction in order to identify the most economically feasible locations.

One of the key benefits of the spatial-allocation model is that it allows retailers to make informed decisions about where to locate new stores based on data-driven analysis. By considering a wide range of factors, including customer behavior and competition, retailers can identify the most profitable locations for new stores and maximize their return on investment.

However, the spatial-allocation model is not without its limitations. The model relies heavily on data quality and accuracy, and it may be difficult to obtain accurate and comprehensive data in some cases. In addition, the model may be affected by changes in customer behavior or competition over time, which may require periodic updates and revisions.

The spatial-allocation model is a valuable tool for retailers looking to expand their store networks and increase profitability. By considering a range of factors and using data-driven analysis, retailers can make informed decisions about where to locate new stores and optimize their store networks for maximum profitability.

Comparison of above-mentioned models is presented in Table 1.

**Table 1.** Comparison of site selection models.

|  |  |  |
| --- | --- | --- |
| **Model** | **Advantages** | **Disadvantages** |
| Analog model | * Helpful in situations where there is limited data available. * Provides a visual representation of the data, making it easier to interpret and understand. * Useful for predicting future events or outcomes based on past data. * Easy to use. | * Limited by the availability and quality of historical data. * Subjective and rely on the expertise of the user. * May not accurately capture the complexity of the system being modeled. * It does not account for the fact that a successful store may have unique characteristics that are not easily replicated in other locations. |
| Huff Model | * Useful tool for predicting customer behavior and estimating market share. * Сan take into account factors such as the distance between locations and the attractiveness of each location. * Relatively simple and easy to use. * Сan be applied to a wide range of industries and applications. | * Assumes that customers will always choose the closest location, which may not be accurate in all cases. * It does not take into account the impact of advertising or promotions. * Assumes that customers have perfect information about all available locations. * It may not accurately capture the preferences of individual customers. |
| Multiplicative competitive interaction (MCI) model | * Useful for predicting changes in market demand due to changes in the economic environment. * Takes into account a wide range of economic factors, such as interest rates, inflation, and consumer confidence. * Can be used to forecast demand at different levels of aggregation, such as by product or by region. * Widely used and has a strong track record of accuracy. | * Relies on assumptions about the behavior of consumers and the economy, which may not always hold true. * Difficult to incorporate new variables or changes in the economic environment into the model. * Complex and difficult to understand, requiring significant expertise in data analysis and interpretation * Time-consuming and costly to collect and analyze the necessary data. |
| Multiple regression model | * Сan be used to identify the factors that have the greatest impact on a particular outcome. * Сan be used to test hypotheses and determine the strength of relationships between variables. * Can be used to make predictions based on the values of the independent variables. * Widely used and well-established in the field of statistics. | * Assumes that the relationship between the independent and dependent variables is linear, which may not always be the case. * Assumes that there are no other variables that are influencing the outcome, which may not be true. * Requires a large amount of data to be accurate or methods of additional data verification, which may not be available in all cases. |
| Spatial-allocation model | * Useful for predicting the spatial distribution of various phenomena, such as population or traffic flow. * Take into account a wide range of factors, such as the proximity of different locations and the availability of resources. * Can be used to test different scenarios and identify the most effective spatial strategies. * Useful in a variety of fields, including urban planning, transportation, and environmental management. | * Complex and time-consuming to develop. * Require a large amount of data, particularly spatial data, which may not always be available. * Sensitive to changes in the input data and may not accurately predict outcomes in new or changing environments. * May not capture all relevant factors that influence the spatial distribution of the phenomena being studied. * Computationally intensive, which can limit their scalability and applicability to larger datasets or more complex scenarios. |

It is assumed that the combination of different models and approaches lead to a more accurate analysis of the geo retail landscape. Yet, it is rarely possible to conduct several of them as even one method or model requires collection of significant amount of data. Therefore, the choice of a specific approach should be done in accordance with data availability and the agreed limitations in terms of time and costs associated.

## **2.5. Current analytical tools**

**2.5.1. Geographic information systems (GIS)**

With the development of informational technologies and internet network organizations obtained access to different analytical tools allowing them to make better informed decisions. This led to the application of geographic information systems in business.

A geographic information system (GIS) is a computer-based tool that captures, stores, analyzes, and displays geographic data. GIS allows users to view, understand, question, interpret, and visualize data in many ways that reveal relationships, patterns, and trends in the form of maps, globes, reports, and charts (Esri, n.d.).

The development of GIS began in the 1960s, when researchers at Harvard University and the Canada Land Inventory (CLI) recognized the need for a computer-based system to manage land use data (De Smith, Goodchild, & Longley, 2018). The first GIS software, the Canada Geographic Information System (CGIS), was developed by the CLI in 1969 (Foresman, 1998). Since then, GIS technology has continued to advance, with the development of more powerful hardware and software tools that have increased the speed and accuracy of data processing and analysis.

One of the primary advantages of GIS technology is its ability to integrate multiple sources of data. By combining data from different sources, GIS can provide a more complete picture of a particular phenomenon or problem. For example, GIS can be used to integrate data on population density, air quality, and land use to identify areas with high levels of pollution and poor air quality. This information can then be used to inform policy decisions, such as where to locate new parks or where to implement new transportation infrastructure.

GIS is also an important tool for spatial analysis. With GIS, users can perform a wide range of spatial analyses, including proximity analysis, network analysis, and terrain analysis. These analyses can help to identify patterns and relationships between different environmental and social factors. For example, GIS can be used to identify areas that are most vulnerable to natural disasters, such as floods or earthquakes. This information can be used to inform emergency management plans and to allocate resources more effectively.

GIS technology also plays an important role in data visualization. By creating maps and visualizations of spatial data, GIS can help to communicate complex information in a more accessible and understandable format (Calle-Jimenez, Orellana-Alvear, & Prado-Imbacuan, 2019). Maps and visualizations can be used to identify patterns, trends, and relationships that might be difficult to discern from raw data alone. GIS can also be used to create 3D visualizations of landscapes, buildings, and other features, which can help to provide a more immersive and realistic view of the world.

GIS is used in many fields, including environmental management, urban planning, natural resource management, public health, transportation, and many others (De Smith, Goodchild, & Longley, 2018). In the business world, GIS is often used in location analysis and site selection for retail stores and other businesses. GIS can also be used for market analysis, customer profiling, and targeted marketing campaigns (Roig-Tierno, Baviera-Puig, & Buitrago-Vera, 2013).

**2.5.2. Machine learning in retail location analysis**

Machine learning has become increasingly important in the retail industry, particularly in retail location analysis. Retailers are using machine learning algorithms to analyze large amounts of data to gain insights into customer behavior, preferences, and patterns. One way machine learning is used in retail location analysis is through predictive modeling (Ting & Jie, 2022). Retailers can use historical sales data, demographic information, and other relevant data to build models that predict sales performance at potential retail locations. This can help retailers make more informed decisions about where to open new stores or expand existing ones.

The machine learning approach to location problem involves predicting the optimal location for a new retail store or business. This problem is particularly relevant for retailers looking to expand their operations and want to maximize their chances of success by selecting the best location possible.

One of the key challenges in the machine learning retail location problem is identifying the factors that are most important in determining the success of a retail store. Factors such as demographics, population density, traffic flow, and competition are all important considerations when selecting a location for a new retail store. Machine learning algorithms can help identify the key factors that are most important in determining the success of a retail store by analyzing large datasets of demographic and market data.

One of the most common applications of machine learning in retail location analysis is predictive modeling. Predictive models use historical sales data, demographic data, and geographic data to identify patterns and relationships that can be used to predict future sales and revenue (Glaeser, Fisher, & Su, 2019). These models can be used to determine the optimal location for a new store or to analyze the performance of existing stores.

One popular approach to solving the machine learning retail location problem is using supervised learning algorithms (Karamshuk, Noulas, Scellato, Nicosia, & Mascolo, 2013). These algorithms are trained on a labeled dataset that contains information about the success or failure of retail stores in different locations, along with relevant input features such as demographic and market data. The algorithms can then use this dataset to predict the success of a new retail store in a particular location based on the input features.

Another approach to solving the machine learning retail location problem is using unsupervised learning algorithms, such as clustering and dimensionality reduction (Shaytura, et al., 2020). These algorithms can be used to identify patterns in the data that may not be immediately apparent, such as identifying areas with similar population densities or demographics. This approach can be particularly useful in identifying potential locations that may not have been considered otherwise.

Machine learning can also be used to optimize store layouts and product placement (Miao, 2020). Retailers can use machine learning algorithms to analyze customer traffic patterns, identify hotspots, and optimize product placement to increase sales. For example, if a retailer notices that customers frequently purchase coffee and pastries together, they can place these items near each other to increase sales of both items.

In addition to optimizing store locations and layouts, machine learning can also help retailers to personalize their marketing and sales efforts. By analyzing customer data and purchase history, retailers can create targeted marketing campaigns and personalized recommendations that increase customer loyalty and drive sales. For example, a retailer might use machine learning algorithms to recommend specific products to customers based on their purchase history or demographic profile.

While machine learning algorithms have a lot of potential in retail location analysis, it is important to consider their limitations as well. One of the primary limitations of machine learning algorithms is the quality and quantity of data. The algorithms rely heavily on the data inputs, so if the data is inaccurate, incomplete, or biased, the resulting predictions may also be inaccurate or biased. In addition, if the data set is too small, the algorithm may not be able to identify meaningful patterns or relationships.

## **Summary of Chapter 2**

In this section we have explored various factors that impact sales performance in retail, such as store layout, product assortment, pricing, promotion, staff training, location and foot traffic. Additionally, we have examined the role of geomarketing in retail, which involves analyzing spatial data to better understand consumer behavior, improve marketing strategies and find better locations for opening new stores. We have also reviewed the evolution of different retail location theories, including the central place theory, spatial interaction theory, bid rent theory, minimum differentiation principle and percolation theory. These theories provide insights into factors that affect the decisions of retailers in selecting a location, and how they can utilize this knowledge to enhance their location strategies and increase profits.

This section also reviews different models that can be used for store location, they include analog model, Huff model, MCI model, multiple regression model and spatial allocation model. In the last part, the review delves into the use of GIS and machine learning as analytical tools for retail location selection.

While many of the existing studies on retail location selection and sales performance have focused on general retail settings, there is a lack of research that specifically examines these topics in the context of sport goods companies. The sport goods industry is unique in many ways, with factors such as seasonality, product specialization, and brand recognition playing a significant role in sales performance and location selection. Therefore, there is a need for more research that specifically examines the factors that impact sales performance in sport goods stores and the models and tools that can be used to select optimal locations for these stores.

By addressing these gaps and building on the existing literature, we aim to provide valuable insights and practical recommendations that can help Velodrive make data-driven decisions and improve their competitiveness on the market.

# **Chapter 3. Analysis of geofactors influence on Velodrive sales performance**

In the previous chapter we examined the existing approaches and methodologies to the research problem. In order to choose the most appropriate approach for us, we need to first analyze the data that is obtained from a Velodrive company representative from Sales department and geodata that we collect using several tools.

To solve the research tasks that we set out previously, we need to collect information on geofactors. We also need visualize the locations that Velodrive currently has. We realize that a lot of the data we collect will require a visual representation, and most often on maps. That is why we need to find the appropriate and available technologic solutions. Further in our analysis we will consider such tools and solutions as 2GIS, API, Yandex.Audience and Yandex DataLens.

## **Technologies and tools used to extract geodata in the research**

**3.1.1 2GIS API opportunities for geodata extraction**

2GIS is a popular digital map and business directory service that provides detailed information about cities, businesses, and organizations. The name 2GIS stands for "double GIS", as it combines two different types of maps: a detailed street map and an interactive 3D map. 2GIS was first launched in 2003 in Russia and has since expanded to cover over 600 cities in 35 countries, including Ukraine, Kazakhstan, the UAE, Italy, and others. The service is available on various platforms, including desktop and mobile apps for iOS and Android.

One of the main features of 2GIS is its detailed maps, which provide comprehensive information about streets, buildings, and landmarks. Users can easily search for specific addresses or places and get directions, as well as view detailed information about businesses and organizations in their area.

In addition to maps, 2GIS also offers a business directory service, which provides information about local businesses, such as their contact information, hours of operation, and reviews from other users. Businesses can also create their own profiles on 2GIS to promote their services and connect with customers.

Another unique feature of 2GIS is its 3D map, which allows users to view cities and landmarks from different angles and perspectives. This can be especially helpful for getting a better understanding of a city's layout and architecture.

Thus, 2GIS is a useful tool for anyone looking to navigate a city or find information about local businesses and organizations. Its detailed maps and comprehensive business directory make it a popular choice among users in Russia and beyond.

We decided to use 2GIS for obtaining information via API, and Yandex.Maps and Yandex.Audience for general location information and data on human traffic, audience description and the size of the target audience with the interest “Cyclists”.

An API (Application Programming Interface) is a set of protocols, tools, and standards for building software applications. It defines how software components should interact with each other and provides a way for different systems to communicate and exchange data.

An API provides a standard way for developers to access and manipulate data and services from other software applications, databases, and systems. By using an API, developers can integrate different systems and applications without having to understand the underlying implementation details.

APIs can be designed for different purposes, such as retrieving data, updating data, or performing a specific action. There are also different types of APIs, including RESTful APIs, SOAP APIs, and GraphQL APIs, each with their own specific characteristics and use cases.

2GIS provides an API that allows developers or researchers to access its extensive database of geospatial information, including maps, points of interest, and local business information. With the 2GIS API, developers can integrate this information into their own applications and services, enabling users to easily search for and navigate to locations, businesses, and other points of interest, and researchers to retrieve required geoinformation from the platform.

The 2GIS API offers several different endpoints and services, including:

1. Maps API - This endpoint provides access to the 2GIS map tiles and allows developers to embed maps into their applications. The Maps API supports zooming and panning, as well as searching for specific locations and points of interest;
2. Search API - This endpoint allows developers to search the 2GIS database for points of interest and local businesses. Searches can be filtered by category, location, and other criteria;
3. Routing API - This endpoint provides routing and navigation functionality, allowing developers to calculate optimal routes between two or more points, taking into account traffic, road closures, and other factors;
4. Geocoding API - This endpoint allows developers to convert addresses or other location information into geographic coordinates, and vice versa;
5. Reviews API - This endpoint provides access to user reviews and ratings for local businesses listed in the 2GIS database.

In addition to these endpoints, the 2GIS API also offers various tools and resources for developers, including documentation, SDKs, and sample code.

We requested the access to APIs of 2GIS on non-commercial terms.

**3.1.2 Visualization opportunities with Yandex DataLens**

Yandex DataLens is a platform that provides advanced data analysis and visualization capabilities. It is a cloud-based platform that allows users to connect to various data sources, prepare and transform data, and create interactive dashboards and reports.

One of the key features of DataLens is its ability to connect to a wide variety of data sources, including databases, cloud services, and APIs, including Yandex.Maps or 2GIS API. This allows to easily integrate data from different sources and create a source of triangulation for the analysis.

Once the data is connected, users can use DataLens' data preparation tools to clean, transform, and enrich the data. This includes features like filtering, sorting, aggregating, and joining data sets. Users can also create custom formulas and calculations to derive new insights from their data.

DataLens also provides a wide range of visualization options, including charts, graphs, and maps. These visualizations can be customized to match the user's needs and can be used to create interactive dashboards and reports. Users can also create custom widgets and add them to their dashboards for additional functionality.

One of the key advantages of DataLens is its flexibility of use and free access. The platform is designed to make visualization results user-friendly. This allows not only researcher, but also businesses to make data-driven decisions and unlock the full potential of the data.

## **Data collection and description**

**3.2.1. Sales data retrieval**

Our initial data request concerns the basic financial metrics of the company: volume of sales in rubles and units, the cost of goods sold (COGS) and gross profit. The timeframe which the company agreed to provide data for is 3 years. We analyze this period to identify seasonality patterns and to judge about the evolution and dynamics of sales in Velodrive. The data that we request should be presented monthly for each point of sale in Saint-Petersburg for the period from January 2020 up to December 2022.

The dataset that we initially received was in the form of a retrieval of data from the Russian software program “1C”. It contained 487,269 rows and 5 columns.

Figure 5 depicts the hierarchy of the data we obtained from the company. The data is represented by the subsequences of levels. As can be inferred from the above figure, those data cannot be machine-read as they do not have a “flat table” structure. Moreover, the date field is located inside of the sales categories’ structure and is not extracted as an individual axis.

Store

Month and year

Sales manager

Season\*

Sales category

I level

Sales category

II level

Article group

Article model

Article modification

\*Dashed outline signifies an optional layer

**Figure 5**. Levels of hierarchy or the data presented by Velodrive

Apart from the sales data, we also require the key operational information on the points of sale of the company in Saint-Peterburg, including rental square and fees of each point of sale, their address, monthly rental fees and HR costs.

We received sales data from Velodrive store chain in Saint-Petersburg for 14 offline points of sale. They are:

1. Akademicheskaya (prospect Nauki, 19, bld. 2);
2. Baltiyskaya (Naberezhnaya Obvodnogo kanala, 118A);
3. Bogatyrskiy (Bogatirskiy prospect, 13A, hypermarket “Okay”);
4. Zvezdnaya and Zvezdnaya kids (Zvezdnaya st., 1, “Kontinent” mall);
5. Zvezdnaya-2 (Moskovskoe shosse, 36);
6. Kupchino (Balkanskaya square, 5, “Balkaniya Nova 2” mall);
7. Lakhta (Lakhtinskiy prospect, 85, “Garden-City” exhibition center);
8. MaksiSopot (Zheleznovodskaya st., 68, “Maksi Sopot” shopping center;
9. Mercury (Savushkina st., 141);
10. Murino (Murino, Tikhaya st., 12);
11. Ozerki (Engels’ prospect, 109, bld.2);
12. Prosvesheniya (Prosvesheniya prospect, 53, bld.1);
13. RIO (Fuchika st., 2, “RIO” shopping center);
14. City Mall (flowers, presents, fireworks) (Kolomyazhskiy prospect, 17, bld.1).

Velodrive provided us with the rental information on 12 of them, excluding RIO and City Mall.

**3.2.2. Rental data description**

Velodrive provided us with an Excel table on the 12 offline points of sale with 3 descriptive columns which are presented in Table 2.

**Table 2.** Velodrive rental fees data template

|  |  |  |  |
| --- | --- | --- | --- |
| Velodrive store | Actual address | Monthly rental fee | Area, in km squared |

Apart from that, we received a presentation dated from year 2021 on the Velodrive chain and their franchise model. It contains information on:

1. The formats of the company stores;
2. Their audience in terms of gender and age;
3. Factors that a franchisee should pay attention to while choosing a retail location (based on Velodrive experience);
4. Advantages of having a store in a shopping center or as a standalone point of sale;
5. Brands presented in the company chain;
6. Sales, COGS and gross profit structure by categories;
7. Key marketing metrics (customer base, subscribers on social media, mailing list size in VK social network, number of store visitors per day (during the season), number of Internet visits per month (during the season);
8. List of the recommended retail location (based on the Velodrive internal analysis).

Upon receipt of all the data described above our next step is to preprocess data to make it machine-readable, structured and synthesized.

**3.2.3. Extraction of geodata from 2GIS API**

Via 2GIS API we need to retrieve several points:

1. Geocoded coordinates of Velodrive stores;
2. Bicycle stores located in predefined areas;
3. Geocoded coordinates of points of attraction of human traffic.

The areas that serve for our interest are the polygons, specifically hexagons, built with the help of H3 hexagonal indexing system.

Since we work with geodata, we need to geocode the addresses of Velodrive stores for further visualization and spatial analysis. To do it, we used Geocoder API of 2GIS. Appendix 1 presents a function for direct geocoding, it sends a GET request to the API, passing the provided address as a query parameter. The API call requests the latitude and longitude coordinates (items.point) for the specified address. The api\_key variable holds the authentication key required for accessing the API. Then we parse the JSON response and extract coordinates.

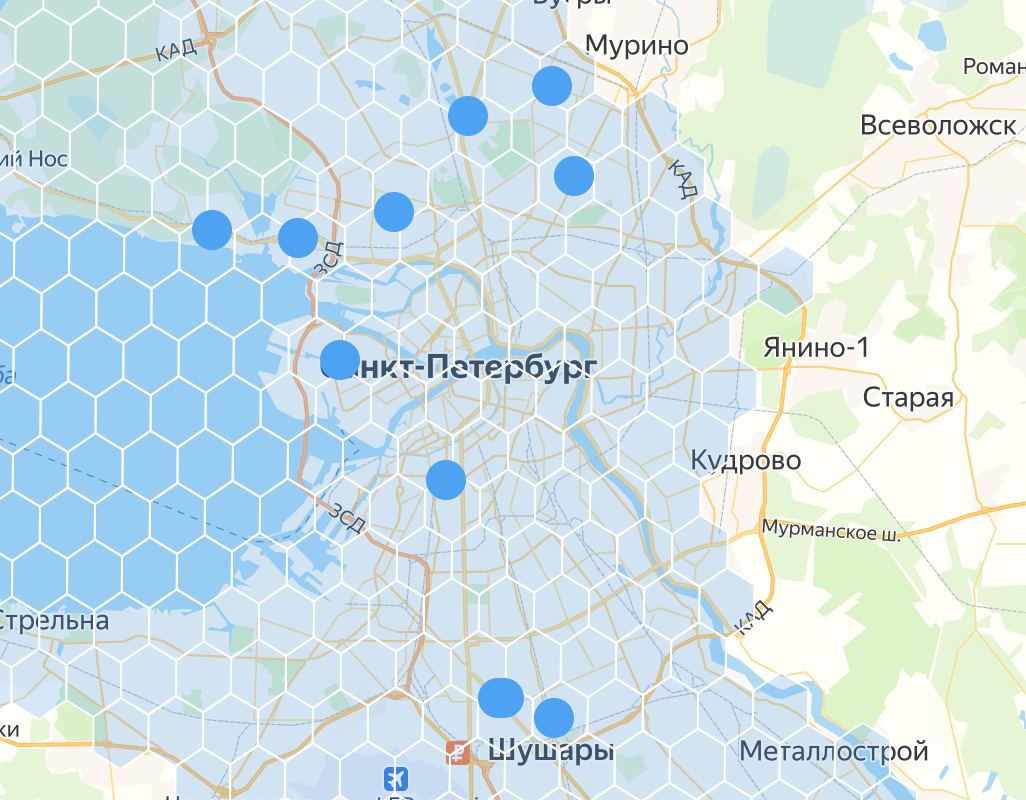
**3.2.4 Introduction to H3 geospatial system and division of the sales area into hexagons**

H3 is a hierarchical hexagonal geospatial indexing system developed by Uber. It is used for partitioning the earth's surface into small, uniformly sized hexagonal cells, allowing for efficient indexing and querying of geospatial data. H3 addresses some of the limitations of traditional geospatial indexing methods, such as rectangular grids, by providing a more flexible and granular approach to representing geospatial data. The hierarchical nature of H3 also allows for spatial queries to be performed at varying levels of resolution, making it a useful tool for applications such as geospatial data analysis, location-based services, and navigation. H3 is an open-source project, available on GitHub, and has been implemented in several programming languages.

H3 provides for 16 resolution types that are differentiated by the average hexagon area, that is choosing a resolution impacts the size of hexagons and, thus, localization of our geoanalytical research (Uber, n.d.).

At first, we convert the coordinates of Saint-Petersburg to the Shapely Polygon format. Shapely is a Python library for computational geometry. It provides a set of geometric operations and data structures that enable you to manipulate and analyze geometric shapes, such as points, lines, and polygons (Gillies, 2023). To create hexagons we use a function that takes a GeoJSON object with coordinates, creates a Folium map, generates a set of hexagonal shapes using H3 library, creates and adds polylines representing the hexagons and the original input polyline to the map, and returns the map, hexagonal polygons, polylines, and hexagon IDs.

In the research we used Resolution 7 which implies the average hexagon area of 5,16 squared kilometers and the average edge length of 1,4 kilometers. That implies the coverage of Saint-Peterburg by hexagons of the same area which can be seen in Figure 6. Blue bubbles indicate the existing locations of Velodrive shops. Each hexagon will be examined by the factors influencing the choice of a retail location.



**Figure 6.** Division of Saint-Petersburg into hexagons of Resolution 7 with Velodrive stores

Figure 6 and other maps are visualized via Yandex DataLens tool. We made sure in advance that the resolution, size and location of polygons in DataLens are the same as we received via 2GIS API.

**3.2.5 Identifying the rivalry level**

The next step we perform is locating the competition on the map to identify the level of rivalry in a hexagon area based on data from 2GIS. The list of the companies that can be considered as Velodrive competitors includes:

1. Velocountry
2. Velograd
3. Velostorm
4. Trial-Sport
5. Kingbike
6. Velomarka
7. RealVelo
8. Birota
9. AlienBike

To identify the competition rate, we created a for loop presented in Appendix 2, which iterates over each competitor in the above-mentioned list. In the loop we send a GET request to 2GIS API specifying:

* Query parameter which is the name of the competitor;
* Fields that we want to get back from the API, we want to receive coordinates of the stores;
* Id of the region where we are searching for, in our case it is Saint-Petersburg with id – 38;
* Key which was provided by 2GIS company.

Then we parse the response content into a JSON format, remove the 'meta' key because this information is not needed and normalize the JSON data into a flat table-like structure using the json\_normalize() function from the pandas library. The 'result.items' column is a list that contains required information so we "explode" it in the current\_competitor DataFrame, normalize it, drop unnecessary columns and concatenates it to the final dataset with competitors.

As a result, we get a dataset with name of the competitor, its address, latitude and longitude of its location. Further we create a GeoDataFrame from the existing pandas DataFrame by adding a geometry column based on the longitude and latitude columns. The resulting GeoDataFrame can be used for spatial analysis and visualization. We perform a spatial join operation between two GeoDataFrames, one with competitors, another with hexagons of Saint-Petersburg, using the "contains" spatial relationship. It's worth noting that for a spatial join to work properly, the GeoDataFrames should have a common coordinate reference system (CRS) and compatible geometry types for the specified spatial relationship. Using this joined GeoDataFrame we calculated how many stores of our competitors are located in each hexagon.

The maximum number of competitors that can be found in a hexagon is two (not accounting for Sportmaster which is not considered as a competitor by Velodrive) which makes Velodrive sphere of business low or moderately competitive in terms of geography.

**3.2.6 Identifying the population of audience with the interest “Cyclists”**

In order to understand the potential volume of the target audience we decided to use DataLens BI Datasets, containing the information on the socio-demographic characteristics of population. The information there is applicable to the hexagons we defined earlier and conform to the H7 resolution. The final dashboard in the format of a map is presented in Figure 7. There, we chose the hexagon with the highest relative interest “Cyclist”.

To be more specific, “Cyclist” interest category reflects the users who conformed to at least one of the following criteria:

1. Visits of the thematic websites for cyclists at least on 2 different days over the last month;
2. Browser search for bicycle accessories or information on the repairing works;
3. Purchase of bicycles or their accessories;
4. Downloads of thematic applications. Map

   Description automatically generated

**Figure 7.** Socio-demographic representation on the map with the interest “Cyclists”

Source: DataLens Marketplace dashboards

As soon as data on human and automobile traffic is not an open data, we decided that the use of this metric – the audience with the interest “Cyclists” – would be a good approximation of the traffic in hexagon areas. Not only it tells us about people who physically pass by a location, but it also takes into account the target audience of Velodrive that is actually interested in the products of this specific company and not nearly located coffeeshops, restaurants, shopping centers and other points of interest that do not guarantee the influx of visitors to Velodrive stores.

## **3.3. Data preprocessing**

Sales data obtained from the company “1C” software contain numerous levels of hierarchy, some of them are not relevant for our analysis. Unfortunately, there is no automated way to restructure the data into the required form, therefore we had to manually transform it to a flat table structure. The target structure has the form presented in Table 3, one instance of Akademicheskaya store is used as an example.

**Table 3.** Final synthesized data structure

|  |  |  |  |
| --- | --- | --- | --- |
| Store | Channel of sale | Sales Category name/Costs category | Sales volume by month (Feb-21), RUB |
| Akademicheskaya | Offline | Bicycles | 1 108 154,33 |
| … | … | … | … |

Considering that we have 14 offline stores and a 3-year period of observation, the size of the table that we created is 815 rows by 39 columns.

It is important to mention that each offline store was also a pick-up point for Internet orders, that is why we included the channel field to be able to see the organic offline sales versus online sales.

In the same table we added operational costs categories from the Velodrive rental fees data: COGS, square, rental fees and calculated ourselves gross and operational profits.

## **3.4. Velodrive assortment analysis**

Based on the summarized data we gathered in our flat table we identified the following sales categories sold in the Velodrive store chain:

1. Bicycles;
2. Bicycle accessories and spare parts;
3. Scooters;
4. Scooters’ spare parts;
5. Balance bikes;
6. Donut tubes;
7. Snowboards;
8. Cross-country skis;
9. Alpine skis;
10. Hoverboards;
11. Motorcycles;
12. Roller skates;
13. Sledges;
14. Sup boards;
15. Snow racer;
16. Skateboards;
17. Markdown articles;
18. Markdown pre-owned articles ;
19. Electric transport;
20. Boats;
21. Services;
22. Ya.Nomenclature;
23. Others (mostly fireworks).

Following our analysis, we can see that, bicycles, their accessories and spare parts account for approximately 80% of the total revenue both for the overall observed period and for each year individually. This is supported by the corporate Velodrive presentation (for the available matching periods).

Figure 8 depicts how the sales dynamics looks like over the observed 3-year period and it validates the conclusion about the dominating sales of bicycles and their spare parts.

**Figure 8**. Velodrive sales dynamics by categories in rubles, 2020-2022

Source: processed data provided by the company

Looking at the chart above, it is easy to notice the seasonal nature of the traditional sales categories – bicycles and their accessories and spare parts. The sales surge in summer when people are looking for the options of an active free time and bicycles become one of the most popular choices. Along with the seasonal peaks in bicycles sales, there is another category that also shows increasing interest of customers in summertime – scooters.

As for the absolute values of sales, in Table 4 we looked at the sales volume and their growth rates of top-2 categories – bicycle and bicycle accessories and spare parts – and compared with the next popular category, scooters.

**Table 4.** Top sales categories of Velodrive in comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Year** | **Bicycles and their accessories and spare parts** | **Growth rate** | **Scooters and their spare parts** | **Growth rate** |
| 2020 | 576 939 302 | - | 50 301 990 | - |
| 2021 | 465 593 419 | -19% | 44 641 848 | -11% |
| 2022 | 507 416 447 | 9% | 41 970 353 | -6% |

First of all, we can see that the sales of scooters and their spare parts are on average 11 times less than that of bicycles and their parts. This is a very large gap between the two categories, and it seems to be very hard to diminish it. Secondly, whereas the bicycles sales grew in 2022 versus 2021 by 9%, the scooters sales decreased by 6%, which does not give reason for expecting the smaller gap either. Both categories showed significantly negative growth rates, presumably, mostly because of the high base in 2020 during COVID-19.

We analyzed the sales in each of the 14 stores, both in offline and online channel. Despite the presence of 19 other sales categories, unfortunately, there is none of them making up for the slump in sales of traditional categories. There are several categories that are seeing growing interest in winter period, especially in some of the shops: donut tubes, sledges, snowboards. Lakhta store specialized on pyrotechnics that accounted for more than 90% in 2020-2021. In 2022 the store started active sale of the traditional category of bicycles, this decreased the share of pyrotechnics to 37% and increased the bicycles share to 44%. The average scooters’ share is kept at the level of 6-8%.

In Zvezdnaya store (adult segment) the share of winter equipment – snowboards – is registered at 7% in both channels. This is a differentiator among all the stores with a share of not a traditional assortment being relatively significant.

An important observation is that from Chapter 1 we know that people postponed their demand for bicycle equipment during COVID-19 active explosion and that demand was realized later in 2021. Velodrive data show the opposite: their sales of bicycles were never as high as in summer 2020 when COVID-19 started spreading intensively in Russia and a lot of relocation restrictions were introduced by local region administrations.

An important observation concerns online channel of sale. Even though online sales are not the focus of our research and seasonality is not the key research object, we think it would be useful to mention that it looks easier to tackle the seasonality issue via online channel, as online sales have a bit different consumer patterns – customers can buy summer items even not during the season. It might be connected, for example, with winter sales of bicycles with lower prices that usually, and with marketing tools (such as contextual ads, internet banners, mailings) it is easier to make a customer become aware of the sale. So, consumers might buy traditional items not during traditional period to save money.

Based on the rental data we received from the company, we set up another table to identify the efficiency of the square used. Our new table structure is presented in Appendix 3.

Gross profit is understood as Revenue volume in rubles deducted by COGS. Operational profit is calculated by deducting from the Gross Profit the sum of Rental fees and HR costs – the predominant costs categories. It is important to note that we do not have the concrete data on HR costs, yet, as it was agreed and validated with the company, HR costs constitute approximately 10% of the revenue of each store. Thus, we took this value as a base for our calculations. For the adequate comparison we expressed both profits and losses in annual terms to better grasp the picture in an annual perspective.

What we found interesting is the example that we used in Table 4 – Murino shop. It is found to be the most expensive place in terms of the cost of 1 sq.m. and at the same time it is the most profitable place expressed by the ratio of operational profit to 1 unit of square. We found that unexpected also because Murino does not belong to Saint-Petersburg itself, it is located in the Leningradskaya oblast. We hope to examine this phenomenon later when building a model of the geo factors.

Apart from Murino, Akademicheskaya shop also seems to be operationally efficient rental location with Operational Income per 1 sq.m. being of 101,014 rubles and cost of 1 rental sq.m. of 13,200 rubles, expressed in 2022 annual terms and based on comparison with intracompany benchmarks. Bogatyrskiy and Ozerki stores also show relatively good performance, but their ratio of operational profit to the rental square is 40,000 rubles lower than that of Murino or Akademicheskaya points of sale. Lakhta and MaxiSopot show the worst operational performance, evaluated by the same criterium. The previously used ratio is 16,300 rubles for the former, and 13,202 rubles for the latter.

We are looking at the operational efficiency of retail locations specifically, that is of 1 unit of rental square, because this is compatible with our research tasks. Therefore, we do not take into account the absolute values of annual Gross revenues deducted by annual rental fees and we realize that the result of analysis would be different if we did so – in most cases the bigger the rental square, the higher the absolute value of rental fees subtracted from gross revenue. But in that case, we are not comparing the efficiency of the location, but sales and profits performance overall.

We also notice that Prosvesheniya store shows negative value for the operational income. Probably, it is connected with the fact that it was opened only in October 2022, and we have an insufficient period of observation for that store.

The conducted analysis allowed us to come to several important conclusions:

1. Velodrive seeks to have a diversified product portfolio structure, however, the data show that bicycles and their accessories and spare parts account for approximately 80% of sales;
2. Velodrive business is characterized by strong seasonality. The traditional sales category – bicycles and accessories – are sold mostly in spring-summer period.
3. There are some categories of goods that are supposed to help to deal with winter slumps in sales. They are pyrotechnics, snowboards, donut tubes, sledges. However, their share in annual sales on average does not exceed 5-7%;
4. Online channel seems to be a good option for the company not only in terms of operational costs reduction (pick-up points are cheaper to have that real offline stores), but also for the sake of seasonality resolution. Yet, the nature of bicycle business does not allow to transform into the e-commerce business, as many customers find important to visit a brick-and-mortar place where they can try on different items.

## **3.5. Building a model to analyze geofactors’ importance and predict annual sales per 1 sq.m.**

Table 1 of Chapter 2 demonstrated different techniques used in site selection process. Based on the available data we decided to determine the most important geofactors with the help of multiple regression model that use ordinary least squares algorithm to determine regression coefficients. We will also use our model also for predicting sales monetary volume for some of new locations in Saint-Peterburg that we consider potentially attractive for Velodrive company.

While building a multiple regression model one should check if the following assumptions associated with OLS-method are met:

1. The model is correctly specified;
2. The size of sample is much greater than the number of estimated parameters of the model;
3. The model is linear;
4. The regressors are fixed in repeated sampling;
5. The variance of regressors is positive and finite;
6. There is no perfect collinearity (perfect linear association) between the regressors;
7. The regressors and the disturbances are not associated;
8. The expected value of disturbances should be zero;
9. The variance of disturbances should be constant and finite (homoskedasticity);
10. There should be no association between the disturbances (no autocorrelation).

In addition to that, a lot of statistical tests imply the normality of disturbances, especially for small samples, like in our case.

Multiple regression model implies checking for all of the above assumptions, apart from that it is also necessary to check the overall pertinence of the model which is usually checked in ANOVA tables of standard statistical packages using F-test. Another necessary check is for the adequate specification of the current model that identifies if any non-linear transformations should be made (unfortunately, without specifying them). We used RESET-test for that introduced by James Ramsey (Ramsey, 1969).

**3.5.1 Geofactors preselected for a multiple regression model**

Based on the information provided by the company and discussions with the company representative on the relevance of different factors we collected data on the following model variables:

1. Annual cost of 1 sq.m. rented;
2. Rental area (sq.m);
3. Floor on which a store is located;
4. Presence on the market, in months;
5. Distance to the nearest metro station, in meters;
6. Parking in the radius of 500 meters (1 – yes, 0 – no);
7. Format of a store: in a shopping center or street retail (1 – shopping center, 0 – street retail);
8. Audience of the Interest “Cyclists”, in users (in the hexagon which a store belongs to);
9. Distance to the nearest Sportmaster, in meters (relevance confirmed by Velodrive);
10. Distance to the nearest hypermarket, in meters (relevance confirmed by Velodrive);
11. Competition level (number of stores-competitors in the hexagon which a store belongs to);
12. Distance to the nearest Velodrive store (including franchise network);
13. Distance to the nearest competitor store (the list of competitors is presented in point 3.2.5);
14. Annual sales per 1 sq.m. – target variable.

While collecting the data we found out that all of the observed Velodrive stores had a parking nearby – either in shopping centers or within the house building yards or at the public free parking, so considering this factor in the model would not make sense.

Before running a model, we made an exploratory data analysis of the variables to look at the distributions, range, standard deviation and correlation between the factors. The EDA results of the dataset are presented in the Jupyter notebook[[1]](#footnote-2).

**3.5.2 Modelling process**

It is important to mention that before assessing the model we made two preliminary data transformations: we standardized all the regressors using StandardScaler function from “sklearn” library since we wanted to estimate the individual contribution of each regressor into the model, and we took a natural logarithm of the target variable because its scale was much larger than the one of the regressors, so the interpretation using percentage change seemed more comprehensible for us, moreover, it favors the compatibility of the model with homoskedasticity and normality of disturbances assumptions.

A usual first step for choosing model variables is looking at the correlation table and determine the variables that are most correlated with the target variable. In our case we used Spearman method for calculating correlation in case there are non-linear associations between the variables. The top- most correlated variables are: Target audience (in the hexagon area which a store belongs to) – 0.543, Annual rental cost per 1 sq.m. – 0.407, Distance to the nearest metro – (-0.406), Rental area – (-0.340), Presence on the market – 0.337 and Floor – (-0.323). It is necessary to mention that p-value for all the correlation coefficients is above the level of significance (0.05). Therefore, we cannot be confident in the existence of the association between the variables, especially considering the size of our dataset. Apart from that, we can see from the correlation matrix (see Appendix 4 and Appendix 5) that the above-mentioned variables are also correlated with other factors including between themselves in some cases which would cause a multicollinearity issue if included in the model. That is why we considered that the most reliable way is to experiment with the model specifications. We tried numerous combinations of factors including pushing them all into the model to see how the results of the model change and if the assumptions of the regression are still met.

Table 5 demonstrates which tests were used for checking different model assumptions and hypotheses. For making a decision on rejecting or not rejecting the null hypothesis we used p-value with the level of significance (alpha) fixed at 0.05.

**Table 5.** Model tests

|  |  |
| --- | --- |
| **Assumption/Hypothesis** | **Test/Method** |
| Overall model pertinence | F-test |
| Coefficients individual significance | t-test, bootstrapping, confidence intervals |
| Adequacy of model specification | Ramsey RESET-test |
| Normality of disturbances | Jarque-Bera test, q-q plot |
| Autocorrelation of disturbances | Durbin-Watson test |
| Multicollinearity | Variance Inflation Factor (VIF) |
| Redundant&Omitted variables | AIC change, F-test |

As our data sample is very small and includes information only on 12 Velodrive stores, we implemented a bootstrapping procedure and built confidence intervals with 95% level of confidence for each of the coefficients and the intercept. We used median values for the regression coefficients after running bootstrapping.

Appendices 6 and 7 show the code with bootstrapping procedure and getting confidence intervals for the median values of the bootstrapped coefficients. None of the coefficients or intercept confidence intervals include zero.

Apart from using Python for running analysis and predicting sales level for new locations we used a statistical software JASP to double-check our results and to quickly see how the model changes when inputs change. The regression report from JASP is also presented in Appendix 8.

Based on our analysis and compatibility with the model assumptions we came to the conclusion that three of the studied variables are statistically significant for the model dependent variable: distance to the nearest metro station, floor and annual rental cost of 1 sq.m. Only the combination of these 3 factors showed the best results in terms of statistical significance (both overall and individual), R-squared, AIC even after running a bootstrapping procedure. This 3-factor model passes successfully all the tests mentioned in Table 6.

We will describe the specification of the model with these variables. The report on running the model with the 3 factors is presented in Appendix 8.

The final model equation with the scaled data is presented in formula (4):

ln (annual sales per 1 sq.m.) = 12.038 + 0.312 annual rental cost of 1 sq.m*.*-

* + - 0.390 distance to the nearest metro – 0.347 floor (4)

**3.5.3 Estimating model errors**

To judge the performance of our final model we used the original “training” data into the equation to calculate absolute errors between the “predicted” and actual values using median values of the bootstrapped regression coefficients. The results are shown in Table 6.

If we were to calculate the error in the annual volume of sales along all the locations taking into account the sign of the error, we would see that our model underpredicts only by ~73.5 thousand rubles. In total and on annual basis this is insignificant. If we look at the absolute values of errors, then the total error for all the locations is 434 207,09 rubles, or 18%.

Model mean errors are presented in Table 7.

**Table 6.** Estimating the model errors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Annual sales per sq.m.** | **Annual sales per sq.m.\_hat** | **Absolute error** | **Absolute percentage error** |
| Akademicheskaya | 387 147,00 | 254 537,28 | 132 609,72 | 34% |
| Baltiyskaya | 199 322,00 | 168 050,74 | 31 271,26 | 16% |
| Bogatyrskiy | 257 479,00 | 189 611,37 | 67 867,63 | 26% |
| Zvezdnaya and Zvezdnaya kids | 209 175,00 | 294 795,29 | 85 620,29 | 41% |
| Zvezdnaya-2 | 99 026,00 | 164 728,51 | 65 702,51 | 66% |
| Kupchino | 194 142,00 | 203 036,48 | 8 894,48 | 5% |
| Lakhta | 99 026,00 | 97 217,49 | 1 808,51 | 2% |
| MaksiSopot | 78 220,00 | 69 326,51 | 8 893,49 | 11% |
| Mercury | 179 249,00 | 191 727,15 | 12 478,15 | 7% |
| Murino | 417 117,00 | 416 030,50 | 1 086,50 | 0% |
| Ozerki | 157 473,00 | 147 181,87 | 10 291,13 | 7% |
| Prosvesheniya | 78 220,00 | 85 903,41 | 7 683,41 | 10% |
| **Total** | **2 355 596,00** | **2 282 146,60** | **434 207,09** | **18%** |

**Table 7.** Modelmean errors

|  |  |
| --- | --- |
| Mean absolute error | 36 183,92 |
| Mean absolute percentage error | 19% |

We see that the model is most erroneous for the two Zvezdnaya retail locations. Speaking about Zvezdnaya-2, we find it necessary to mention that we have sales data for this point only for the last 3 months of 2022. Therefore, we had to choose a proxy location for this store to fill in the annual sales per 1 sq.m. as this is our target variable and it should be comparable among all the retail locations. This proxy location has become Lakhta because it operates not a whole year around and we assumed that even though Zvezdnaya-2 has just launched, its sales should be not lower than for Lakhta. Although, we, of course, admit that this is also a prediction based on the assumption.

We tried to exclude Zvezdnaya-2 from the model, it did not impact significantly the results in terms of errors’ change, so we decided to keep as it is since we already have a small sample.

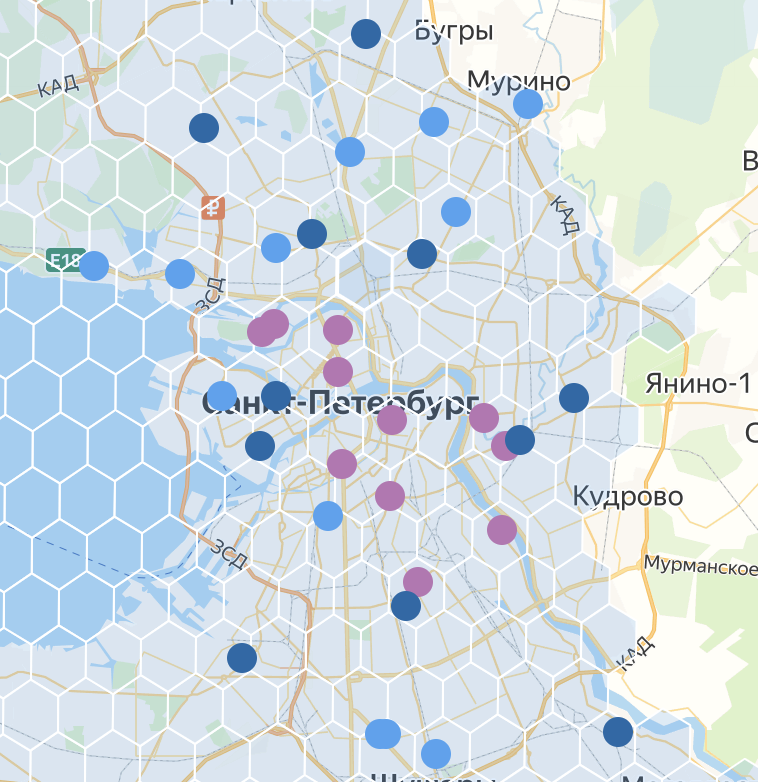
Apart from that, speaking about the two Zvezdnaya stores as a whole, these two locations differ from the others by a very close distance between them (400 meters). Our hypothesis is that this close location is the main reason of a big error. In order to test it, we decided to run a model with the factor “nearest\_store” which corresponds to the nearest Velodrive store including franchisees’ shops. The errors for Zvezdnaya and Zvezdnaya-2 decreased to 20% and 42% correspondingly. Meanwhile, the total percentage error for all the locations increased by 3 p.p. to 21%, and this factor appears not statistically significant because its p-value in the run model is 0.36 which is substantially higher than the level of significance. Yet, this experiment allowed us to confirm that the factor of a very close position of the stores played an important role in the original error. These two locations are significantly different from the other sample representatives and could be considered outliers.

Overall, we consider the results of model implementation on the training data as acceptable.

**3.5.4 Predicting annual sales per 1 sq.m. for the new locations**

For predicting annual sales per 1 sq.m. we selected 12 new locations in Saint-Petersburg that we consider attractive for Velodrive. We based our choices on, first of all, the current map of Velodrive presence on the Saint-Petersburg market including its franchisees’ stores - some areas of the city are still not covered by the company. We also accounted for the size of the target audience with the interest “Cyclists”: even though it does not participate in the model it is better to have a point of sale closer to one’s potential customers. We certainly based our search on available rental commercial locations – for the search we used the local platform “Cian”. Our search was performed on the 17th of May, 2023. We admit that available real estate is a constantly changing factor and we cannot influence it. Yet, for the purpose of methodology introduction to the company we consider it adequate.

Location of the selected sites on the map along with the currently existing stores can be seen in Figure 9, light blue dotes are existing stores of Velodrive, dark blue – franchise stores and purple – new locations.



**Figure** **9.** Newly selected commercial sites on the map

The list of the selected locations with their predicted annual sales per 1 sq.m. are shown in Table 8.

**Table 8.** Sales prediction for the new locations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **floor** | **nearest\_metro, meter** | **annual\_rental\_cost\_of\_sq\_m, RUB** | **rental\_area, sq. m** | **annual\_sales\_per\_sq\_m, RUB** |
| 1 | 1 | 90 | 23 999 | 191 | 691 365,67 |
| 2 | 1 | 650 | 20 400 | 103,4 | 407 926,37 |
| 3 | 1 | 175 | 27 567 | 370 | 922 543,42 |
| 4 | 2 | 90 | 14 112 | 170 | 163 485,21 |
| 5 | 1 | 700 | 30 000 | 339,4 | 946 438,52 |
| 6 | 1 | 250 | 36 000 | 150 | 1 910 378,80 |
| 7 | 1 | 950 | 36 000 | 234 | 1 478 175,92 |
| 8 | 2 | 1000 | 24 000 | 327,8 | 284 013,13 |
| 9 | 1 | 1500 | 12 000 | 100 | 140 777,91 |
| 10 | 1 | 177 | 21 600 | 112 | 540 176,00 |
| 11 | 1 | 760 | 12 000 | 114 | 184 626,16 |
| 12 | 1 | 270 | 18 000 | 205,8 | 378 168,76 |

We can see that predicted sales volume for the new locations are on average higher than for the original, training data. This is explained by the significantly higher rental rates for the sites that we selected which, in turn, is dictated by the location of those sites – some of them are in the city center where the company is not well represented, some are on the intersection of the automobile and human traffic. The information on these 12 sited with their address and corresponding links to Cian is provided in Appendix 9.

**3.5.5 Model limitations**

As any other model, ours has its limitations and simplifications:

1. Annual sales per 1 sq.m. represent a mixed value because original data contained information on both recently opened and long-established stores. We can say that on average in the long run the company can expect the predicted results;
2. More data would improve the quality of the model – not only we would be able to predict sales better, but what is more important in our study we would deeper understand the nature of the observed geo factors, their influence on sales and their associations between each other. We may assume that other factors would become statistically significant in modelling if we had more data. Even though we performed bootstrapping to tackle the issue, it was still done from the same sample. It might be useful to include information on the franchise network, unfortunately, this data cannot be provided by the company due to NDA;
3. We recognize that traffic is an important factor for sales. Yet, there are some notes to be made here. Any traffic is not useful for any type of business. The importance of traffic for a coffeeshop and for a B2B company or a DIY shop would be different. Here, we deal with a bike company, and it is hard to imagine that a lot of people would become interested in the goods of the company while just passing by. It is more probable that they would first search for them on Internet and then go to specifically chosen stores. However, the traffic is evidently substantial for the brand awareness and recognition. More people see and know about the brand, more probable they would make their preliminary search with that brand. This is why we used the target audience with the interest “Cyclists” that was estimated by Yandex Crypta technology as an approximation to the “target traffic”. And for the same reason Velodrive should put efforts in marketing for rising their brand awareness.

Moreover, the traffic is most likely considered in the annual rental cost of 1 sq.m. as this factor impact directly rental price for commercial real estate and this factor proved to be significant and included in the model;

1. Our location suggestions are temporary and may be valid only for limited period of time. The goal of the research is to propose a methodology that can be used in site selection and to estimate the influence of different factors on sales for Velodrive business, specifically. We assume that decisions on the network expansion or relocation are not in the day-to-day agenda for the company, therefore there might little reason to automate this process.

## **. Estimation of business effect**

Based on the prediction of the model we calculated potential sales volumes for 5 selected locations that we consider most attractive for the company. We composed a draft of a 3-year Profit & Loss statement for each of the 5 locations where we tried to consider the most significant cost drivers if decided to launch a Velodrive store there. These cost drivers include:

1. Remodeling locations to fit our requirements;
2. Purchase of sales equipment such as cash desks, equipment for displaying goods, payment terminals, etc.;
3. HR expenses;
4. Recruitment costs associated with hiring new staff;
5. Logistics and delivery expenses;
6. Marketing expenses to promote new stores and attract customers;
7. Income taxes.

We based our costs assumptions on corporate presentations of Velodrive and confirmed our suggestions with the company. For example, we learned that HR costs account for approximately 10% of total sales of Velodrive. These include salaries, taxes and social contributions. Since we acknowledge that first year sales will be lower than the average and lower than the predicted values (for the reasons mentioned previously) we assumed that 1st year sales level would be 70% of the predicted value. This number is based on the growth dynamics of the Velodrive stores that had been launched in 2021 and on average grew by 43% the year after. We also admit that the sales will not be flat and will grow in the third year. Based on the stores, for which we have 3-year-long data, in Year 3 the YoY growth was on average 11%. That is what we included in our assumptions when calculating revenue streams. For the discounting rate we took a risk-free rate for the governmental bonds of 7.5%.

The list of assumptions and high-level results of the P&L are shown in Table 9 and Table 10.

**Table 9.** Financial model assumptions

|  |  |  |  |
| --- | --- | --- | --- |
| Model assumptions | | | |
| HR costs rate | 10% | 1st year sales proportion | 70% |
| Gross profit margin | 40% | 3d year YoY growth | 11% |
| Marketing expenses rate | 5% | risk-free rate | 8% |
| Card payments rate | 50% |  |  |
| Income tax | 20% |  |  |
| Acquiring rate | 2% |  |  |
| Logistics expenses | 1% |  |  |
| Recruitment expenses rate (on HR costs) | 10% |  |  |

**Table 10.** Net income for the 5 selected locations, in RUB

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Location** | **Metric** | **Year 1** | **Year 2** | **Year 3** |
| Ligov (shopping center) | Discounted Net Income | 11 098 062 | 24 196 017 | 24 659 348 |
| Ligov (shopping center) | Net income margin | 13% | 21% | 21% |
| Krestovskii (general purpose) | Discounted Net Income | 2 591 000 | 5 253 154 | 5 573 862 |
| Krestovskii (general purpose) | Net income margin | 9% | 14% | 15% |
| Petrogradskaya (general purpose) | Discounted Net Income | 27 315 967 | 44 094 085 | 46 251 219 |
| Petrogradskaya (general purpose) | Net income margin | 13% | 16% | 16% |
| Admiralteiskii (shopping center) | Discounted Net Income | 28 081 302 | 41 885 192 | 43 631 668 |
| Admiralteiskii (shopping center) | Net income margin | 15% | 17% | 17% |
| Chernyshevskaya (general purpose) | Discounted Net Income | 32 008 668 | 49 239 383 | 51 439 508 |
| Chernyshevskaya (general purpose) | Net income margin | 14% | 16% | 17% |

As discussed above, substantial difference in sales amount between predicted values and Velodrive original, training data are explained by the selection of places with significantly higher annual rental cost per 1 sq.m. for the new locations which is in turn dictated by the location. The weighted average profitability among the new locations is close to 17% which if, certainly, much higher than currently (that latest figures we have are for Year 2019 – 1,1%), yet Velodrive plans to increase it till 11,2% by 2024 within the current strategy taking into account their existing locations and with shutting down some of the least profitable and opening new sites. Considering the much higher sales volume and not that high costs, as well as the fact that we calculate margins only for the new locations, we consider the predicted profit margins realistic.

The detailed P&L can be found in Appendix 10.

Based on the calculations described above we measured the annual profit per 1 sq.m. for each of the five new locations. The results are presented in Table 11.

**Table 11.** Net annual profits per 1 sq.m. for the five new locations, in RUB

|  |  |  |  |
| --- | --- | --- | --- |
| **New location** | **Year 1** | **Year 2** | **Year 3** |
| Ligov (shopping center) | 60 369 | 107 997 | 121 990 |
| Krestovskii (general purpose) | 25 003 | 58 711 | 66 967 |
| Petrogradskaya (general purpose) | 85 341 | 150 136 | 169 292 |
| Admiralteiskii (shopping center) | 198 583 | 322 690 | 361 357 |
| Chernyshevskaya (general purpose) | 145 339 | 243 172 | 273 090 |

Judging by the results above, locations in a shopping center Admiralteiskii and near metro station Chernyshevskaya are preferable in terms of the balance between revenues and costs, including initial investments.

## **3.7. Visualization of collected data and research results via Yandex DataLens**

Based on collected data and predictions of the model we have created a dashboard using Yandex DataLens[[2]](#footnote-3). It allows Velodrive to track their monthly and annual sales performance, analyze sales by categories and channels and analyze store performance in terms of its location.

We used 6 datasets with information on existing stores of Velodrive, their franchise stores, new potential locations, competitors, target audience and Velodrive sales.

The dashboard consists of the following parts:

* Selectors of year, store and sales channels connected with bar chart and stacked are chart;
* Bar chart on monthly sales: it displays monthly sales of 12 Velodrive stores, allowing to track sales performance over a 3-year period from 2020 to 2022. The chart presents sales figures for each month, providing a visual comparison between different periods and highlighting patterns in sales;
* Stacked area chart on monthly sales by categories: it visualizes sales data categorized by different product and service categories. The chart shows the cumulative sales volume for each category over time, with each category represented by a colored area. The chart allows to observe the contribution of each category to the overall sales trend and identify any shifts or changes in category performance;
* Selectors of gender, age and income level of target audience connected with map;
* Map of Saint-Petersburg with hexagons for target audience: it represents the locations of existing stores, franchise stores and new locations. Each current store of Velodrive contains information on name, address, type of location (shopping center or street rerail), floor, rental area, presence on the market, distance to the nearest metro station, annual rental costs of 1 sq. m and annual sales per 1 sq. m. Potential locations contains same date except presence on the market. For franchise store we only have data on name and address since other data is confidential. The map is divided into hexagons that were determined by the H3 geo-indexing system, the average dimensions of their faces is 1.22 km. Each hexagon contains information on number of people interested in cycling. This visualization helps to understand which districts of the city are the most promising in terms of target audience location;
* Heat map on competitors: the map contains the same information on the stores as the previous one, however, instead of hexagons with target audience data it represents a heat map that highlights areas with high competitor activity. This information helps to identify potential challenges and opportunities in specific districts of the city and make strategic decisions accordingly.

Overall, this dashboard combines various visual elements, including bar charts, stacked area charts and maps to provide a comprehensive view of sales, target audience, store locations, and competitor level. By presenting data in an intuitive and interactive format, this dashboard enables company to extract meaningful insights and make data-driven decisions for their business.

# **Conclusion**

In our master thesis, we have developed a statistical model of linear regression to understand the most important geofactors influencing sales of Velodrive company in the 12 locations for which we were provided with data. Using this model, we also predicted the potential sales volumes for 12 preselected locations where the company could allocate their stores. We calculated potential business effect on the company assuming several key costs drivers and growth dynamics and came up with the possible profits and profit margins for the 5 locations. We calculated net annual profits per 1 sq.m. for these sites and identified two most efficient in terms of the balance between potential revenues and costs, including initial investments.

We visualized the results of our research on a dashboard where Velodrive can easily see the dynamics of sales, in total and by categories, their locations on the map along with the competition and their target audience. We also added on the map the newly suggested sites with the data on potential sales volume, so that the company can judge on these locations easily.

To achieve our results, we followed several methodological steps:

* 1. We deep dived into the specifics of bicycle industry, its dynamics, revenue and cost drivers and peculiarities of Velodrive business with its position on the market, sales structure and seasonality.
  2. We studied literature on using geoanalytics in business, in particular, in retail. We looked at the similar cases of retail companies and which methods and tools they used to determine the most appropriate retail locations. We chose the most suitable approach for us based on the data we have and can get in open sources.
  3. We divided Saint-Peterburg into hexagons of Resolution 7 to calculate competition rate and the level of target audience of the company. We collected data on the potentially important factors influencing the sales of the company via 2GIS API, Yandex Maps and Yandex DataLens, combined them with the company internal data on the current rental cost and revenues per 1 sq.m. We assigned annual rental sales per 1 sq.m. as a target variable for which we studied influential factors and made predictions on the new locations. We standardized data and ran several statistical tests for different model specifications to identify the model that would respect OLS estimation rules and simultaneously be pertinent. After that we were able to judge which factors play the most important role in the target variable.
  4. We based our selection of new locations on the current presence map of the company including its franchise network and those factors that turned out significant in the modelling process. Those are: distance to the nearest metro station, floor on which a store is located and annual cost of 1 sq.m. of the rented space. Based on the data we had, we checked that there is no multicollinearity between the model variables, as it might reasonably seem. We suggest that annual rental cost may be more influenced by other factors, such as area of the city (center or residential districts), human and automobile traffic, time of construction, type of real estate (economy, business, premium etc.) that as a result became included indirectly into our model. We also discussed why traffic could be of not very high importance for this type of business.
  5. We measured the economic effect for the 5 locations, based on several assumptions confirmed with Velodrive. We calculated potential revenues, costs (investments and operational) and profits for the new locations.
  6. Finally, we created a dashboard with all collected data and research results via Yandex DataLens that would allow Velodrive to make data-driven decision on opening of new locations.

The main limitations for our results lie in the lack of data, a small training dataset within a limited timeframe that includes the impact of COVID-19, and the inaccessibility of certain information such as the company's operational expenses, franchise financials, and traffic. The small dataset itself represents a major flaw in our analysis since it has a limited statistical power to detect meaningful relationships or patterns. We assume that other factors for which we had gathered information could also become significant in the model. In addition, access to the operational costs could make profits predictions more realistic.

Concerning the seasonality issue, current sales structure does not seem to allow for significant changes that could tackle this issue. As mentioned previously, managing seasonality seems more accessible via online channel and by introducing range of winter products. However, the company already pursues this tactic. It seems to us that even though Velodrive seeks to diversify its assortment to increase winter sales, customers are simply not aware that they can find something other than bicycles and their accessories in the stores named Velodrive. Probably, more targeted and clear marketing messages would increase brand awareness in this case.

Further research could include implementation of Machine Learning algorithms on larger datasets. These algorithms can handle complex relationships between input variables and sales outcomes, capturing non-linear patterns that may not be captured by traditional regression models. Also, further research could include a wider range of geofactors that could be collected from third-party providers. These steps would allow to enrich the dataset, identify more important factors and enhance the model's predictive capabilities.

There is also an issue with the constantly changing commercial estate market – that would require regular filtered search and detection of available retail sites for store allocation. Yet, from business perspective and withing the framework of one city it does not seem viable, since usually companies do not consider relocations or expansions that often.

# **References**

1. Administration of St. Petersburg. (2023, March 28). Transport Committee - About bike paths. Retrieved April 20, 2023, from https://www.gov.spb.ru/gov/otrasl/c\_transport/razvitie-velosipednoj-infrastruktury/o-velodorozhkah/
2. Applebaum, W. (1966). Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales. Journal of Marketing Research, Vol. 3, No. 2 , 127-141.
3. Banerjee, S. (2019). Geomarketing and Situated Consumers: Opportunities and Challenges. rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising (pp. 1–4). New York: Association for Computing Machinery.
4. Baviera-Puig, A., Buitrago-Vera, J., & Escriba-Perez, C. (2016). Geomarketing Models in Supermarket Location Strategies. Journal of Business Economics and Management 17 (6), 1205-1221.
5. Bekti, R., Pratiwi, N., & Jatipaningrum, M. (2018). Multiplicative Competition Interaction Model to obtained Retail Consumer Choice based on Spatial Analysis. IOP Conference Series: Earth and Environmental Science, Volume 187. Bogor, Indonesia: The 4th International Seminar on Sciences.
6. Berman, B. R., & Evans, J. R. (2018). Retail Management: A Strategic Approach, 13th edition. Pearson.
7. Berry, B. B. (1962). Retail location and consumer behavior. Papers of the Regional Science Association 9, 65–106.
8. Brown, S. (1993). Retail location theory: evolution and evaluation. International Review of Retail, Distribution & Consumer Research, Vol. 3, No. 2, 185-229.
9. Buckner, R. W. (1998). Site Selection: New Advancements in Methods and Technology. Lebhar-Friedman Books.
10. BusinesStat. (2022). "The research on the bicycle market in Russia" by BusinesStat. Retrieved April 17, 2023, from https://businesstat.ru/news/bicycles/
11. Calle-Jimenez, T., Orellana-Alvear, B., & Prado-Imbacuan, R. (2019). GIS and User Experience in Decision Support for Retail Type Organizations. International Conference on Information Systems and Software Technologies (ICI2ST) (pp. 156-161). Quito: IEEE.
12. Cliquet, G. (2013). Geomarketing: Methods and Strategies in Spatial Marketing. John Wiley & Sons.
13. Cliquet, G. (2020). From Geomarketing to Spatial Marketing. In Spatial Economics Volume II (pp. 277-305). Colombo: Palgrave Macmillan, Cham.
14. Cliquet, G., & Baray, J. (2020). Location-Based Marketing: Geomarketing and Geolocation. Wiley-ISTE.
15. Data Insight. (2022, February). Marketing research of the online market of sporting goods in 2021 by "Data Insight". Retrieved April 17, 2023, from https://datainsight.ru/DI-Sport2021
16. De Smith, M. J., Goodchild, M. F., & Longley, P. (2018). Geospatial Analysis: A Comprehensive Guide to Principles, Techniques, and Software Tools. 6th ed.
17. Eaton, B. C., & Lipsey, R. G. (1975). The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition. The Review of Economic Studies 42, no. 1 , 27–49.
18. Esri. (n.d.). What is GIS? Retrieved March 14, 2023, from https://www.esri.com/en-us/what-is-gis/overview
19. Field, A. P., Miles, J., & Field, Z. (2012). Discovering statistics using R. London: Sage.
20. Foresman, T. W. (1998). The history of geographic information systems : perspectives from the pioneers.
21. Ghosh, A., & McLafferty, S. (1982). Locating Stores in Uncertain Environments: A Scenario Planning Approach. Journal of Retailing, 58, 5-22.
22. Gillies, S. (2023, April 7). The Shapely User Manual. Retrieved from Shapely 2.0.1 Documentation: https://shapely.readthedocs.io/en/stable/manual.html
23. Glaeser, C. K., Fisher, M., & Su, X. (2019). Optimal Retail Location: Empirical Methodology and Application to Practice. Manufacturing & Service Operations Management 21(1), 86-102.
24. Hotelling, H. (1929). Stability in Competition. Economic Journal, 39 , 41-57.
25. Huff, D. L. (1963). Probabilistic Analysis of Shopping Center Trade Areas. Land Economics 39, no. 1, 81-90.
26. Ivanova, T. (2022, November 24). New Retail. Retrieved April 17, 2023, from https://new-retail.ru/marketing/sportivnoe\_importozameshchenie\_novye\_lidery\_na\_rossiyskom\_rynke/
27. Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., & Mascolo, C. (2013). Geo-Spotting: Mining Online Location-Based Services for Optimal Retail Store Placement. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 793–801). New York: Association for Computing Machinery.
28. Ladle, J. K., Stiller, D., & Stiller, D. (2009). Retail Site Selection: A New, Innovative Model for Retail Development. Cornell Real Estate Review Vol. 7, 17.
29. Miao, Y. (2020). A Machine-Learning Based Store Layout Strategy in Shopping Mall. The 2020 International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy. SPIOT 2020. Advances in Intelligent Systems and Computing, vol 1282. (pp. 170-176). Springer.
30. Ministry of Sports of the Russian Federation. (2014). Development of Physical Culture and Sports. Retrieved April 20, 2023, from Portal of state programs of the Russian Federation: https://programs.gov.ru/Portal/programs/passport/13
31. Morphet, C. S. (1991). Applying multiple regression analysis to the forecasting of grocery store sales: an application and critical appraisal. International Review of Retail, Distribution and Consumer Research, 329-351.
32. Nakanishi, M., & Cooper, L. G. (1974). Parameter Estimation for a Multiplicative Competitive Interaction Model: Least Squares Approach. Journal of Marketing Research, Vol. 11, No. 3, 303-311.
33. Piovani, D., Molinero, C., & Wilson, A. (2017, October 4). Urban retail location: Insights from percolation theory and spatial interaction modeling. Retrieved March 12, 2023, from PLoS One: https://doi.org/10.1371/journal.pone.0185787
34. Ramadani, V., Zendeli, D., Gerguri-Rashiti, S., & Dana, L.-P. (2018). Impact of geomarketing and location determinants on business development and decision making. Competitiveness Review, Vol. 28 No. 1, 98-120.
35. Ramsey, J. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. Journal of the Royal Statistical Society. Series B (Methodological), 31(2), 350–371.
36. Rey, S. J. (2001). Mathematical Models in Geography. In International Encyclopedia of the Social & Behavioral Sciences.
37. Roig-Tierno, N., Baviera-Puig, A., & Buitrago-Vera, J. (2013). Business opportunities analysis using GIS: the retail distribution sector. Glob Bus Perspect 1, 226–238.
38. Saini, S., & Bansal, R. (2023). Geo-Marketing: A New Tool for Marketers. In Enhancing Customer Engagement through Location Based Marketing (pp. 102-112).
39. Sevtsuk, A., & Kalvo, R. (2018). Patronage of urban commercial clusters: A network-based extension of the Huff model for balancing location and size. Environment and Planning B:Urban Analytics and City Science Vol.45(3), 508–528.
40. Shaytura, S., Feoktistova, V., Minitaeva, A., Olenev, L., Chulkov, V., & Kozhaev, Y. (2020). Spatial geomarketing powered by Big Data. Turismo: Estudos &Práticas No. 5, 1-13.
41. Sobol, M., & Kiseleva, A. (2022, July 8). Vedomosti. Retrieved April 17, 2023, from https://www.vedomosti.ru/business/articles/2022/07/07/930408-prodazhi-tovarov-marketpleisah
42. Suárez-Vega, R., Gutiérrez-Acuña, J. L., & Rodríguez-Díaz, M. (2015). Locating a supermarket using a locally calibrated Huff model. International Journal of Geographical Information Science, 29:2, 217-233.
43. TASS. (2021, July 14). In St. Petersburg, the length of bike paths will more than double by 2030. Retrieved April 20, 2023, from https://tass.ru/obschestvo/11904289
44. Taylor, R. D. (2015). Multiple Regression Analysis as a Retail Site Selection Method: An Empirical Review. In The 1980’s: A Decade of Marketing Challenges (pp. 184–187). Springer, Cham.
45. Ting, C.-Y., & Jie, M. Y. (2022). Location Profiling for Retail-Site Recommendation Using Machine Learning Approach. Proceedings of the International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022) (pp. 48–67). Atlantis Press.
46. Uber. (n.d.). Tables of Cell Stats. Retrieved from H3: https://h3geo.org/docs/core-library/restable/
47. Vandell, K. D., & Carter, C. C. (1993). Retail Store Location and Market Analysis: A Review of the Research. Journal of Real Estate Literature 1, no. 1 , 13-45.
48. Wieland, T. (2017). Market Area Analysis for Retail and Service Locations with MCI. The R Journal, Vol. 9/1, 298-323.

# **Appendices**

## **Appendix 1. The function to geocode addresses of the stores**

# Define a function to get the coordinates of an address using the 2GIS API

**def** get\_coordinates(address):

# Send a GET request to the 2GIS API for the address

response = requests.get(f'https://catalog.api.2gis.com/3.0/items/geocode? ’

‘q={address}&fields=items.point&key={api\_key}')

# Parse the JSON response to extract the latitude and longitude coordinates

result = response.json()['result']

**if** result['items']:

latitude = result['items'][0]['point']['lat']

longitude = result['items'][0]['point']['lon']

**return** pd.Series({'latitude': latitude, 'longitude': longitude})

**else**:

**return** pd.Series({'latitude': None, 'longitude': None})

## **Appendix 2. The loop for getting a dataset with competitors**

**for** competitor **in** comp\_stores:

response = requests.get(f'https://catalog.api.2gis.com/3.0/items?q={competitor}’

‘&fields=items.point&region\_id=38&key={api\_key}')

parsed\_response = response.json()

parsed\_response.pop('meta')

current\_competitor = pd.json\_normalize(parsed\_response)

current\_competitor = current\_competitor.explode('result.items')

current\_competitor = pd.json\_normalize(current\_competitor['result.items'].tolist())

current\_competitor = current\_competitor.drop(['id', 'type'], axis=1)

competitors\_df = pd.concat([competitors\_df, current\_competitor],ignore\_index=True)

## **Appendix 3. Rental efficiency calculation**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | **Store** | **Rental square** | **Revenue** | **Gross profit** | **Operational profit** | **Monthly rental fees** | **Annual rental fees** | **Operational profit per 1 sq.m.** | **Gross profit per 1 sq.m.** | **Annual price per 1 sq.m.** |
| 2021 | Akademicheskaya | 172 | 46 645 661 | 17 835 570 | 10 900 604 | 189 200 | 2 270 400 | 63 375,60 | 103 695,17 | 13 200,00 |
| 2022 | Akademicheskaya | 172 | 66 589 360 | 26 303 821 | 17 374 484 | 189 200 | 2 270 400 | 101 014,44 | 152 929,19 | 13 200,00 |
| 2022 | Akademicheskaya | 117 | 21 377 307 | 8 181 367 | 5 014 037 | 93 600 | 1 029 600 | 42 855,01 | 69 926,22 | 8 800,00 |
| 2020 | Bogatyrskiy | 183 | 67 486 044 | 26 652 103 | 17 143 498 | 230 000 | 2 760 000 | 93 664,96 | 145 616,03 | 15 079,50 |
| 2021 | Bogatyrskiy | 183 | 35 146 192 | 14 132 196 | 7 857 577 | 230 000 | 2 760 000 | 42 930,54 | 77 212,46 | 15 079,50 |
| 2022 | Bogatyrskiy | 183 | 47 126 440 | 19 582 775 | 12 110 131 | 230 000 | 2 760 000 | 66 164,73 | 106 992,16 | 15 079,50 |
| 2020 | Zvezdnaya | 423 | 115 551 114 | 44 125 560 | 25 838 448 | 561 000 | 6 732 000 | 61 083,80 | 104 315,74 | 15 914,89 |
| 2021 | Zvezdnaya | 423 | 92 042 979 | 37 277 476 | 21 341 178 | 561 000 | 6 732 000 | 50 451,96 | 88 126,42 | 15 914,89 |
| 2022 | Zvezdnaya | 423 | 88 481 213 | 36 331 918 | 20 751 797 | 561 000 | 6 732 000 | 49 058,62 | 85 891,06 | 15 914,89 |
| 2022 | Zvezdnaya-2 | 395 | 1 788 039 | 956 370 | -152 433 | 310 000 | 930 000 | -385,91 | 2 421,19 | 2 354,43 |
| 2020 | Kupchino | 194 | 48 871 394 | 17 749 700 | 9 262 561 | 300 000 | 3 600 000 | 47 794,43 | 91 587,72 | 18 575,85 |
| 2021 | Kupchino | 194 | 31 221 687 | 12 223 949 | 5 501 781 | 300 000 | 3 600 000 | 28 388,96 | 63 075,07 | 18 575,85 |
| 2022 | Kupchino | 194 | 37 624 734 | 14 413 834 | 7 051 361 | 300 000 | 3 600 000 | 36 384,73 | 74 374,79 | 18 575,85 |
| 2020 | Lakhta | 70 | 3 868 009 | 1 540 808 | 1 020 626 | 96 000 | 192 000 | 14 580,37 | 22 011,54 | 2 742,86 |
| 2021 | Lakhta | 70 | 3 482 640 | 1 994 382 | 1 203 499 | 96 000 | 384 000 | 17 192,84 | 28 491,17 | 5 485,71 |
| 2022 | Lakhta | 70 | 6 931 848 | 2 602 196 | 1 141 011 | 96 000 | 1 152 000 | 16 300,16 | 37 174,23 | 16 457,14 |
| 2020 | MaksiSopot | 320 | 934 256 | 142 048 | -238 993 | 287 616 | 287 616 | -746,85 | 443,90 | 898,80 |
| 2021 | MaksiSopot | 320 | 29 696 685 | 12 353 374 | 5 932 313 | 287 616 | 3 451 392 | 18 538,48 | 38 604,29 | 10 785,60 |
| 2022 | MaksiSopot | 320 | 25 030 245 | 10 178 965 | 4 224 549 | 287 616 | 3 451 392 | 13 201,71 | 31 809,27 | 10 785,60 |
| 2020 | Mercury | 842 | 181 479 543 | 67 056 138 | 33 733 424 | 1 264 563 | 15 174 760 | 40 053,94 | 79 620,21 | 18 018,00 |
| 2021 | Mercury | 842 | 138 413 246 | 50 397 126 | 21 381 042 | 1 264 563 | 15 174 760 | 25 387,13 | 59 839,86 | 18 018,00 |
| 2022 | Mercury | 842 | 150 963 756 | 58 007 713 | 27 736 578 | 1 264 563 | 15 174 760 | 32 933,48 | 68 876,41 | 18 018,00 |
| 2020 | Murino | 72 | 17 180 347 | 6 451 415 | 3 963 381 | 110 000 | 770 000 | 55 046,96 | 89 602,99 | 10 694,44 |
| 2021 | Murino | 72 | 25 846 988 | 11 366 678 | 7 461 980 | 110 000 | 1 320 000 | 103 638,61 | 157 870,53 | 18 333,33 |
| 2022 | Murino | 72 | 30 032 449 | 12 006 988 | 7 683 743 | 110 000 | 1 320 000 | 106 718,65 | 166 763,72 | 18 333,33 |
| 2020 | Ozerki | 697 | 144 338 435 | 82 707 525 | 55 136 600 | 540 496 | 6 485 952 | 79 105,60 | 118 662,16 | 9 305,53 |
| 2021 | Ozerki | 697 | 94 404 211 | 53 826 670 | 34 083 231 | 540 496 | 6 485 952 | 48 899,90 | 77 226,21 | 9 305,53 |
| 2022 | Ozerki | 697 | 109 758 551 | 62 419 374 | 40 844 867 | 540 496 | 6 485 952 | 58 600,96 | 89 554,34 | 9 305,53 |
| 2022 | Prosvesheniya | 298 | 1 291 264 | 737 963 | -305 594 | 286 000 | 858 000 | -1 025,48 | 2 476,39 | 2 879,19 |

## **Appendix 4. Spearman's Correlations**

|  | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | |  | | **Spearman's rho** | | **p** | | **Lower 95% CI** | | **Upper 95% CI** | |
| rental\_area |  | - |  | annual\_rental\_cost\_of\_sq\_m |  | -0.280 |  | 0.379 |  | -0.826 |  | 0.333 |  |
| rental\_area |  | - |  | floor |  | 0.350 |  | 0.265 |  | -0.193 |  | 0.760 |  |
| rental\_area |  | - |  | target\_audience |  | 0.039 |  | 0.905 |  | -0.635 |  | 0.637 |  |
| rental\_area |  | - |  | nearest\_metro |  | 0.109 |  | 0.737 |  | -0.580 |  | 0.788 |  |
| rental\_area |  | - |  | nearest\_hypermarket |  | -0.469 |  | 0.124 |  | -0.886 |  | 0.195 |  |
| rental\_area |  | - |  | presence\_on\_market |  | 0.602 |  | 0.038 |  | -0.011 |  | 0.956 |  |
| rental\_area |  | - |  | nearest\_sportmaster |  | -0.443 |  | 0.149 |  | -0.869 |  | 0.402 |  |
| rental\_area |  | - |  | competition\_rate |  | -0.185 |  | 0.564 |  | -0.714 |  | 0.571 |  |
| rental\_area |  | - |  | nearest\_store |  | -0.354 |  | 0.258 |  | -0.834 |  | 0.348 |  |
| rental\_area |  | - |  | nearest\_competitor |  | 0.347 |  | 0.269 |  | -0.329 |  | 0.886 |  |
| rental\_area |  | - |  | annual\_sales\_per\_sq\_m |  | -0.340 |  | 0.279 |  | -0.863 |  | 0.295 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | floor |  | 0.295 |  | 0.353 |  | -0.391 |  | 0.826 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | target\_audience |  | 0.000 |  | 1.000 |  | -0.575 |  | 0.548 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | nearest\_metro |  | 0.021 |  | 0.948 |  | -0.547 |  | 0.586 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | nearest\_hypermarket |  | 0.166 |  | 0.607 |  | -0.435 |  | 0.678 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | presence\_on\_market |  | 0.200 |  | 0.534 |  | -0.448 |  | 0.859 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | nearest\_sportmaster |  | -0.200 |  | 0.532 |  | -0.859 |  | 0.600 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | competition\_rate |  | 0.051 |  | 0.874 |  | -0.496 |  | 0.560 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | nearest\_store |  | 0.104 |  | 0.748 |  | -0.538 |  | 0.705 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | nearest\_competitor |  | -0.035 |  | 0.914 |  | -0.661 |  | 0.612 |  |
| annual\_rental\_cost\_of\_sq\_m |  | - |  | annual\_sales\_per\_sq\_m |  | 0.407 |  | 0.189 |  | -0.113 |  | 0.750 |  |
| floor |  | - |  | target\_audience |  | -0.009 |  | 0.977 |  | -0.532 |  | 0.479 |  |
| floor |  | - |  | nearest\_metro |  | -0.189 |  | 0.556 |  | -0.690 |  | 0.414 |  |
| floor |  | - |  | nearest\_hypermarket |  | 0.060 |  | 0.852 |  | -0.432 |  | 0.570 |  |
| floor |  | - |  | presence\_on\_market |  | 0.323 |  | 0.306 |  | -0.147 |  | 0.720 |  |
| floor |  | - |  | nearest\_sportmaster |  | -0.639 |  | 0.025 |  | -0.839 |  | -0.266 |  |
| floor |  | - |  | competition\_rate |  | -0.270 |  | 0.396 |  | -0.730 |  | 0.258 |  |
| floor |  | - |  | nearest\_store |  | -0.014 |  | 0.965 |  | -0.556 |  | 0.445 |  |
| floor |  | - |  | nearest\_competitor |  | 0.247 |  | 0.439 |  | -0.421 |  | 0.749 |  |
| floor |  | - |  | annual\_sales\_per\_sq\_m |  | -0.323 |  | 0.305 |  | -0.768 |  | 0.258 |  |
| target\_audience |  | - |  | nearest\_metro |  | -0.067 |  | 0.837 |  | -0.674 |  | 0.627 |  |
| target\_audience |  | - |  | nearest\_hypermarket |  | -0.279 |  | 0.380 |  | -0.908 |  | 0.496 |  |
| target\_audience |  | - |  | presence\_on\_market |  | 0.461 |  | 0.131 |  | -0.234 |  | 0.867 |  |
| target\_audience |  | - |  | nearest\_sportmaster |  | -0.320 |  | 0.310 |  | -0.789 |  | 0.297 |  |
| target\_audience |  | - |  | competition\_rate |  | 0.103 |  | 0.751 |  | -0.599 |  | 0.762 |  |
| target\_audience |  | - |  | nearest\_store |  | 0.147 |  | 0.648 |  | -0.568 |  | 0.777 |  |
| target\_audience |  | - |  | nearest\_competitor |  | 0.211 |  | 0.511 |  | -0.487 |  | 0.716 |  |
| target\_audience |  | - |  | annual\_sales\_per\_sq\_m |  | 0.543 |  | 0.068 |  | -0.011 |  | 0.818 |  |
| nearest\_metro |  | - |  | nearest\_hypermarket |  | 0.067 |  | 0.836 |  | -0.776 |  | 0.766 |  |
| nearest\_metro |  | - |  | presence\_on\_market |  | -0.068 |  | 0.833 |  | -0.715 |  | 0.642 |  |
| nearest\_metro |  | - |  | nearest\_sportmaster |  | -0.053 |  | 0.871 |  | -0.713 |  | 0.607 |  |
| nearest\_metro |  | - |  | competition\_rate |  | 0.288 |  | 0.364 |  | -0.587 |  | 0.885 |  |
| nearest\_metro |  | - |  | nearest\_store |  | -0.011 |  | 0.974 |  | -0.646 |  | 0.773 |  |
| nearest\_metro |  | - |  | nearest\_competitor |  | -0.260 |  | 0.414 |  | -0.813 |  | 0.635 |  |
| nearest\_metro |  | - |  | annual\_sales\_per\_sq\_m |  | -0.406 |  | 0.190 |  | -0.930 |  | 0.313 |  |
| nearest\_hypermarket |  | - |  | presence\_on\_market |  | -0.322 |  | 0.308 |  | -0.809 |  | 0.423 |  |
| nearest\_hypermarket |  | - |  | nearest\_sportmaster |  | 0.335 |  | 0.287 |  | -0.335 |  | 0.796 |  |
| nearest\_hypermarket |  | - |  | competition\_rate |  | -0.257 |  | 0.421 |  | -0.877 |  | 0.501 |  |
| nearest\_hypermarket |  | - |  | nearest\_store |  | 0.677 |  | 0.016 |  | 0.068 |  | 0.952 |  |
| nearest\_hypermarket |  | - |  | nearest\_competitor |  | 0.216 |  | 0.500 |  | -0.490 |  | 0.842 |  |
| nearest\_hypermarket |  | - |  | annual\_sales\_per\_sq\_m |  | -0.297 |  | 0.348 |  | -0.871 |  | 0.421 |  |
| presence\_on\_market |  | - |  | nearest\_sportmaster |  | -0.254 |  | 0.427 |  | -0.783 |  | 0.395 |  |
| presence\_on\_market |  | - |  | competition\_rate |  | -0.091 |  | 0.779 |  | -0.649 |  | 0.639 |  |
| presence\_on\_market |  | - |  | nearest\_store |  | -0.072 |  | 0.825 |  | -0.659 |  | 0.490 |  |
| presence\_on\_market |  | - |  | nearest\_competitor |  | 0.351 |  | 0.263 |  | -0.327 |  | 0.869 |  |
| presence\_on\_market |  | - |  | annual\_sales\_per\_sq\_m |  | 0.337 |  | 0.283 |  | -0.405 |  | 0.680 |  |
| nearest\_sportmaster |  | - |  | competition\_rate |  | -0.024 |  | 0.942 |  | -0.629 |  | 0.686 |  |
| nearest\_sportmaster |  | - |  | nearest\_store |  | 0.356 |  | 0.256 |  | -0.391 |  | 0.782 |  |
| nearest\_sportmaster |  | - |  | nearest\_competitor |  | -0.106 |  | 0.743 |  | -0.777 |  | 0.599 |  |
| nearest\_sportmaster |  | - |  | annual\_sales\_per\_sq\_m |  | 0.136 |  | 0.674 |  | -0.507 |  | 0.761 |  |
| competition\_rate |  | - |  | nearest\_store |  | -0.260 |  | 0.414 |  | -0.677 |  | 0.247 |  |
| competition\_rate |  | - |  | nearest\_competitor |  | -0.783 |  | 0.003 |  | -0.944 |  | -0.359 |  |
| competition\_rate |  | - |  | annual\_sales\_per\_sq\_m |  | 0.231 |  | 0.469 |  | -0.607 |  | 0.898 |  |
| nearest\_store |  | - |  | nearest\_competitor |  | 0.158 |  | 0.624 |  | -0.500 |  | 0.789 |  |
| nearest\_store |  | - |  | annual\_sales\_per\_sq\_m |  | -0.011 |  | 0.973 |  | -0.640 |  | 0.637 |  |
| nearest\_competitor |  | - |  | annual\_sales\_per\_sq\_m |  | -0.069 |  | 0.832 |  | -0.771 |  | 0.675 |  |
|  | | | | | | | | | | | | | |
| Note.  Confidence intervals based on 1000 bootstrap replicates. | | | | | | | | | | | | | |

## **Appendix 5. Spearman’s rho heatmap**

A picture containing text, screenshot

Description automatically generated

## **Appendix 6. The loop to get bootstrapped coefficients**

# resample with replacement each row

boot\_slope\_rental\_cost = []

boot\_slope\_metro = []

boot\_slope\_floor = []

boot\_interc = []

n\_boots = 1000

n\_points = 12

**for** \_ **in** range(n\_boots):

# sample the rows, same size, with replacement

sample\_df = scaled\_df.sample(n=n\_points, replace=True)

# fit a linear regression

ols\_model\_temp = smFrmApi.ols(formula = 'log\_annual\_sales\_per\_sq\_m ~ annual\_rental\_cost\_of\_sq\_m + floor + nearest\_metro', data=sample\_df)

results\_temp = ols\_model\_temp.fit()

# append coefficients

boot\_interc.append(results\_temp.params[0])

boot\_slope\_rental\_cost.append(results\_temp.params[1])

boot\_slope\_floor.append(results\_temp.params[2])

boot\_slope\_metro.append(results\_temp.params[3])

## **Appendix 7. Getting confidence interval for median bootstrapped coefficients using “bootstrapping” function from scipy.stats**

boot\_slope\_rental\_cost = (boot\_slope\_rental\_cost,)

boot\_interc = ( boot\_interc,)

boot\_slope\_floor = (boot\_slope\_floor,)

boot\_slope\_metro = (boot\_slope\_metro,)

interc\_ci = bootstrap(boot\_interc, np.median, confidence\_level=0.95,

random\_state=1, method='BCa')

rental\_cost\_ci = bootstrap(boot\_slope\_rental\_cost, np.median, confidence\_level=0.95,

random\_state=1, method='BCa')

metro\_ci = bootstrap(boot\_slope\_metro, np.median, confidence\_level=0.95,

random\_state=1, method='BCa')

floor\_ci = bootstrap(boot\_slope\_floor, np.median, confidence\_level=0.95,

random\_state=1, method='BCa')

## **Appendix 8. Linear regression reports results from JASP (with assumptions compatibility check)**

**Table 1.** Model Summary - log(annual\_sales)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | R | R2 | Adjusted R2 | RMSE | Durbin-Watson | | |
| Autocorrelation | Statistic | p |
| H0 | 0.000 | 0.000 | 0.000 | 0.567 | 0.125 | 1.387 | 0.262 |
| H1 | 0.895 | 0.800 | 0.725 | 0.297 | 0.302 | 1.128 | 0.110 |

Table 2. ANOVA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model |  | Sum of Squares | df | Mean Square | F | p |
| H₁ | Regression | 2.832 | 3 | 0.944 | 10.681 | 0.004 |
|  | Residual | 0.707 | 8 | 0.088 | - | - |
|  | Total | 3.539 | 11 | - | - | - |

Note. The intercept model is omitted, as no meaningful information can be shown.

Table 3. Coeﬃcients

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model |  | Unstandardized | **Standard Error** | **Standardized** | **t** | p | 95% CI | | Collinearity Statistics | |
| Lower | Upper | **Tolerance** | **VIF** |
| H₀ | (Intercept) | 12.042 | 0.164 | - | 73.540 | < .001 | 11.681 | 12.402 |  |  |
| H₁ | (Intercept) | 12.042 | 0.086 | - | 140.311 | < .001 | 11.844 | 12.240 |  |  |
|  | nearest\_metro | −0.387 | 0.091 | −0.713 | −4.266 | 0.003 | −0.596 | −0.178 | 0.895 | 1.117 |
|  | ﬂoor | −0.331 | 0.091 | −0.609 | −3.641 | 0.007 | −0.540 | −0.121 | 0.892 | 1.121 |
|  | annual\_rental\_  cost\_of\_sq\_m | 0.311 | 0.087 | 0.573 | 3.578 | 0.007 | 0.111 | 0.512 | 0.973 | 1.028 |

Table 4. Bootstrap Coeﬃcients

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model |  | Unstandardized | **Bias** | **Standard Error** | p\* | 95% CI | |
| Lower | Upper |
| H₀ | (Intercept) | 12.041 | 0.003 | 0.155 | < .001 | 11.749 | 12.348 |
| H₁ | (Intercept) | 12.050 | 0.005 | 0.132 | < .001 | 11.784 | 12.225 |
|  | nearest\_metro | −0.391 | −0.010 | 0.210 | 0.026 | −0.611 | −0.110 |
|  | ﬂoor | −0.343 | −0.032 | 0.122 | 0.031 | −0.526 | −0.089 |
|  | annual\_rental\_cost\_of\_sq\_m | 0.310 | 0.009 | 0.112 | 0.002 | 0.160 | 0.650 |

Note. Bootstrapping based on 1000 replicates.

Note. Coeﬃcient estimate is based on the median of the bootstrap distribution.

\* Bias corrected accelerated.

Table 5. Descriptives

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **N** | **Mean** | **SD** | **SE** |
| log(annual\_sales) | 12 | 12.042 | 0.567 | 0.164 |
| nearest\_metro | 12 | −5.551×10−17 | 1.044 | 0.302 |
| ﬂoor | 12 | 1.480×10−16 | 1.044 | 0.302 |
| annual\_rental\_cost\_of\_sq\_m | 12 | 1.850×10−16 | 1.044 | 0.302 |

**Table 6.** Collinearity Diagnostics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Dimension** | **Eigenvalue** | **Condition Index** | Variance Proportions | | | |
| **(Intercept)** | **nearest\_metro** | **ﬂoor** | **annual\_rental\_cost\_of\_sq\_m** |
| H₁ | 1 | 1.301 | 1.000 | 0.000 | 0.336 | 0.349 | 0.001 |
|  | 2 | 1.053 | 1.111 | 0.000 | 0.071 | 0.043 | 0.800 |
|  | 3 | 1.000 | 1.140 | 1.000 | 0.000 | 0.000 | 0.000 |
|  | 4 | 0.646 | 1.419 | 0.000 | 0.593 | 0.608 | 0.198 |

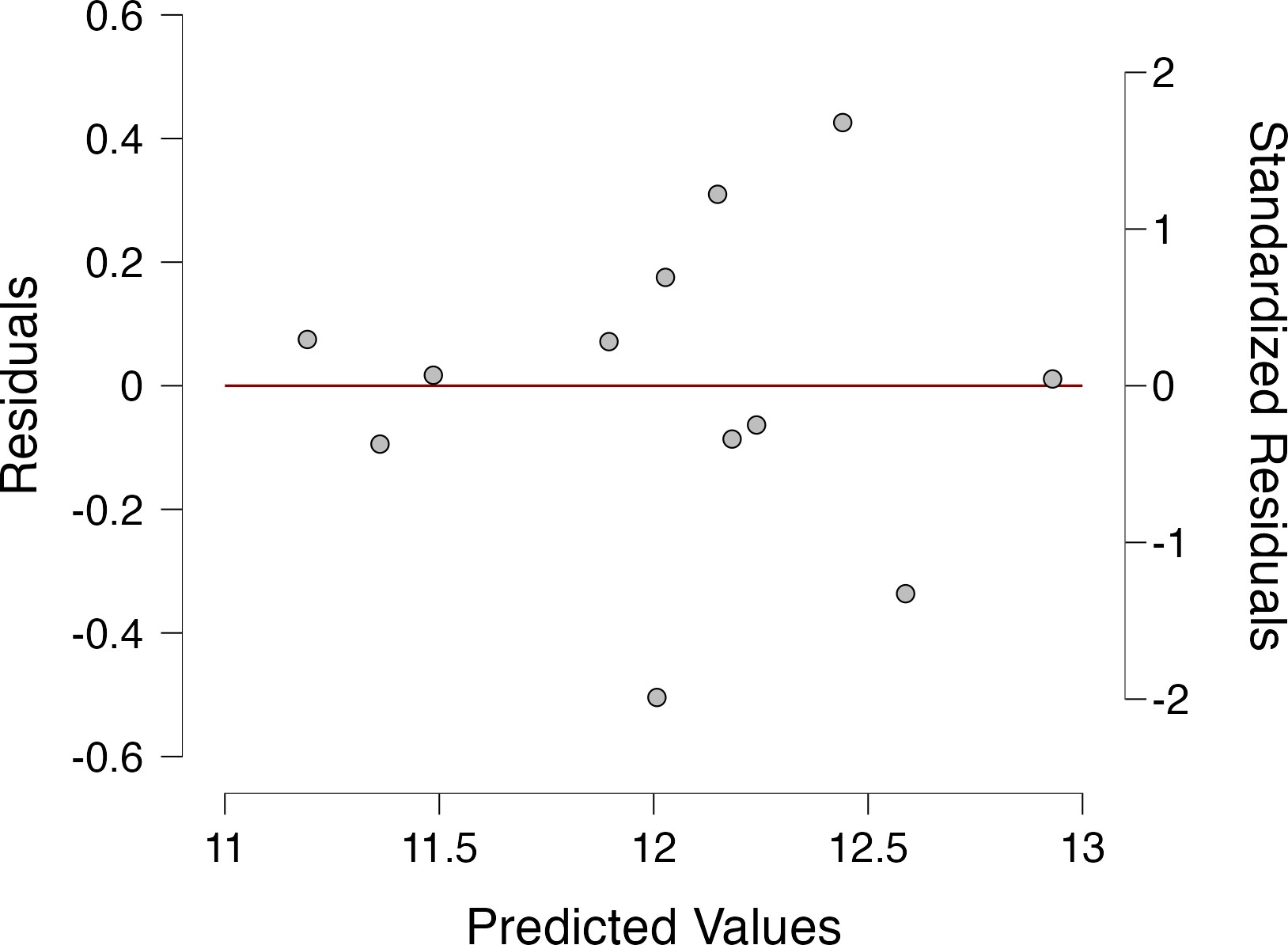


Figure 1. Residuals vs. Predicted

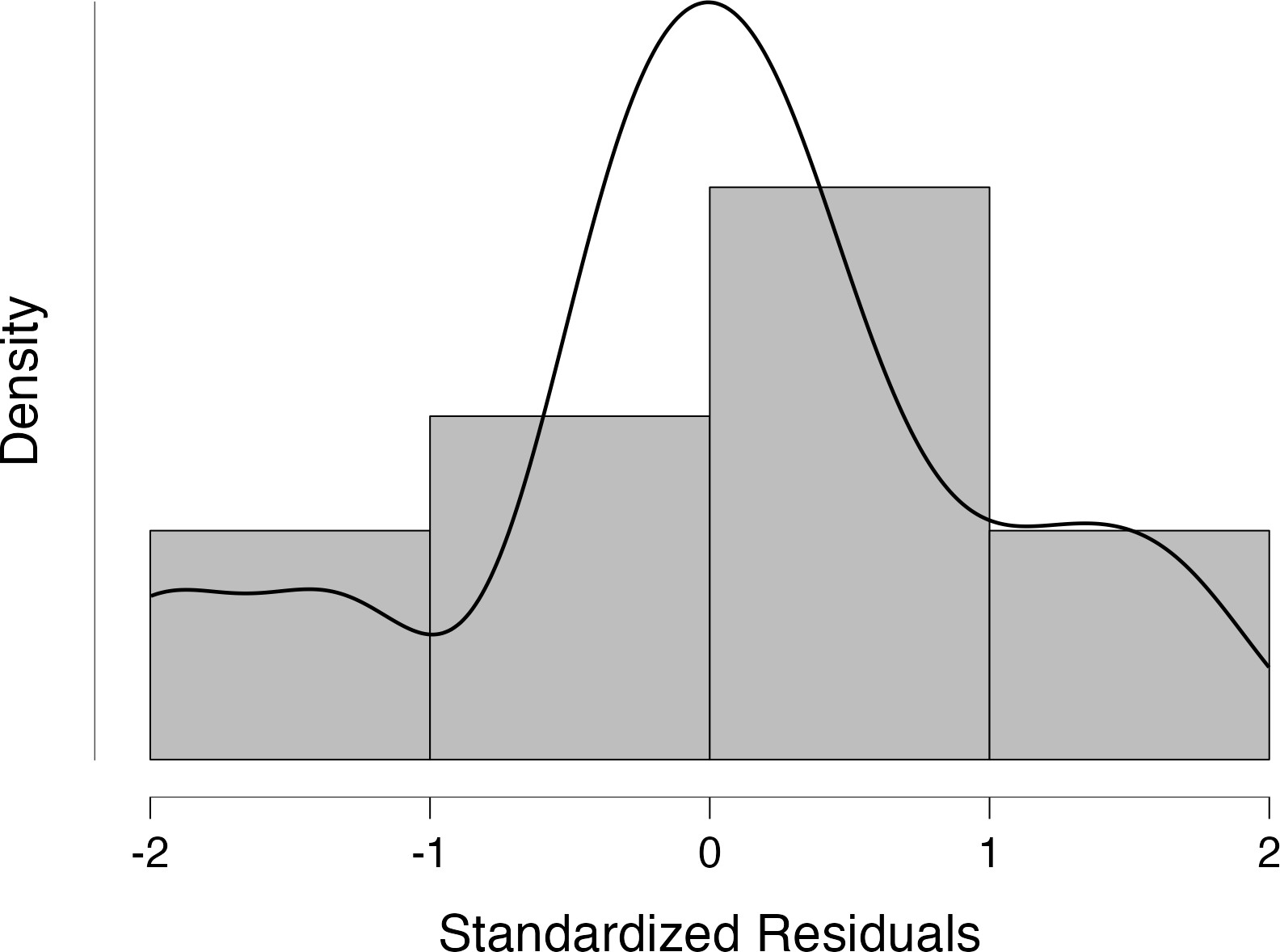


Figure 2. Standardized Residuals Histogram

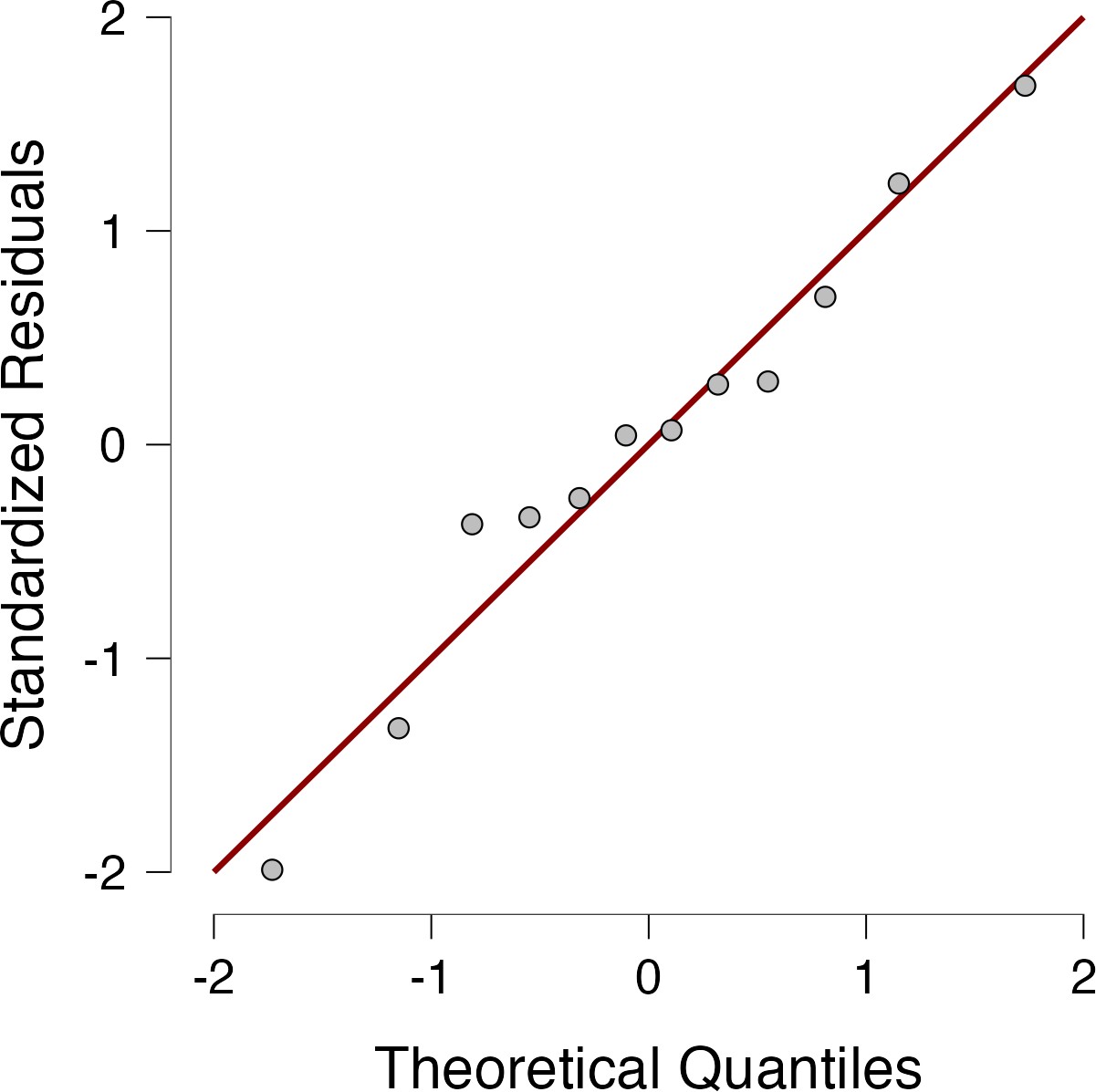


Figure 3. Q-Q Plot Standardized Residuals

## **Appendix 9. New locations**

|  |  |  |
| --- | --- | --- |
| **Store** | **Link** | **Address** |
| Ligov | <https://spb.cian.ru/rent/commercial/280900177/> | Saint-Petersburg, Ligovskii prosp., 153 |
| Krestovskii | <https://spb.cian.ru/rent/commercial/269688270/> | Saint-Petersburg, Esperova st., 10 |
| Krestovskii – 2 | <https://spb.cian.ru/rent/commercial/286072567/> | Saint-Petersburg, Morskoj prosp., 28 |
| Ligov – 2 | <https://spb.cian.ru/rent/commercial/286877792/> | Saint-Petersburg, Ligovskii prosp., 153 |
| Petrogradskaya | <https://spb.cian.ru/rent/commercial/286166797/> | Saint-Petersburg, Professora Popova st., 23 |
| Admiralteiskii | <https://spb.cian.ru/rent/commercial/280494110/> | Saint-Petersburg, Moskovskii prosp., 3aB |
| Chernyshevskaya | <https://spb.cian.ru/rent/commercial/283393506/> | Saint-Petersburg, per. Baskov, 2 |
| Zanevskii | <https://spb.cian.ru/rent/commercial/272561084> | Saint-Petersburg, Zanevskii prosp., 65K2 |
| Novocherskasskaya | <https://spb.cian.ru/rent/commercial/281103385/> | Saint-Petersburg, prosp. Shaumyana, 2 |
| Buharestskaya | <https://spb.cian.ru/rent/commercial/271174694/> | Saint-Petersburg, Salova st., 61 |
| Elizarovskaya | https://spb.cian.ru/rent/commercial/278410451 | Saint-Petersburg, Obshchestvennyi per., 5 |
| Gorkovskaya | <https://spb.cian.ru/rent/commercial/280000498> | Saint-Petersburg, Kronverkskii prosp., 35 |

## **Appendix 10. P&L statement draft for the new locations**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Location** | **Metric** | **Year 1** | **Year 2** | **Year 3** |
| **Ligov (shopping center)** | **Total sales** | **92 435 590** | **132 050 843** | **146 576 435** |
| Ligov (shopping center) | Sales per 1 sq.m. | 483 956 | 691 366 | 767 416 |
| **Ligov (shopping center)** | **Operational Costs** | **- 22 561 215** | **- 27 036 052** | **- 29 505 403** |
| Ligov (shopping center) | Annual cost of 1 sq.m. | - 23 999 | - 23 999 | - 23 999 |
| Ligov (shopping center) | Total rental costs | - 4 583 809 | - 4 583 809 | - 4 583 809 |
| Ligov (shopping center) | Location remodelling | - 668 500 | - | - |
| Ligov (shopping center) | Works per 1 sq.m. | - 2 000 | - | - |
| Ligov (shopping center) | Materials per 1 sq.m. | - 1 500 | - | - |
| Ligov (shopping center) | Housing&Utility | **-** | **-** | **-** |
| Ligov (shopping center) | Sales equipment (cash desks, payment terminals, Acquiring (2%)) | - 1 594 856 | - 1 324 108 | - 1 469 364 |
| Ligov (shopping center) | Director computer | - 25 000 | - | - |
| Ligov (shopping center) | Cash desk computer | - 25 000 | - | - |
| Ligov (shopping center) | Printer | - 4 000 | - | - |
| Ligov (shopping center) | Barcode scanner | - 3 500 | - | - |
| Ligov (shopping center) | Check-printing machine | - 23 000 | - | - |
| Ligov (shopping center) | Label printer | - 33 000 | - | - |
| Ligov (shopping center) | Cash box | - 2 400 | - | - |
| Ligov (shopping center) | Spare materials | - 1 000 | - | - |
| Ligov (shopping center) | Software installation | - 5 000 | - | - |
| Ligov (shopping center) | Fiscal memory device (3Y) | - 14 000 | - | - |
| Ligov (shopping center) | Wires and network equipment | - 1 000 | - | - |
| Ligov (shopping center) | Payment terminal | - 30 000 | - | - |
| Ligov (shopping center) | Internet connection | - 3 600 | - 3 600 | - 3 600 |
| Ligov (shopping center) | Acquiring | - 924 356 | - 1 320 508 | - 1 465 764 |
| Ligov (shopping center) | Demonstration equipment | - 500 000 | - | - |
| Ligov (shopping center) | HR costs (incl. taxes, insurance and other expenditures) | - 9 243 559 | - 13 205 084 | - 14 657 644 |
| Ligov (shopping center) | Personnel recruitment | - 924 356 | - | - |
| Ligov (shopping center) | Logistics and delivery expenses | - 924 356 | - 1 320 508 | - 1 465 764 |
| Ligov (shopping center) | Marketing | - 4 621 779 | - 6 602 542 | - 7 328 822 |
| Ligov (shopping center) | Average gross profit margin | 40% | 40% | 40% |
| Ligov (shopping center) | Gross Income | 36 974 236 | 52 820 337 | 58 630 574 |
| Ligov (shopping center) | Income before taxes | 14 413 021 | 25 784 285 | 29 125 171 |
| Ligov (shopping center) | Income tax | - 2 882 604 | - 5 156 857 | - 5 825 034 |
| **Ligov (shopping center)** | **Net Income** | **11 530 417** | **20 627 428** | **23 300 137** |
| Ligov (shopping center) | Net income margin | 12% | 16% | 16% |
| **Ligov (shopping center)** | **Discounted Net Income** | **10 725 969** | **17 849 586** | **18 755 691** |
| **Krestovskii (general purpose)** | **Total sales** | **29 525 711** | **42 179 587** | **46 819 342** |
| Krestovskii (general purpose) | Sales per 1 sq.m. | 285 548 | 407 926 | 452 798 |
| **Krestovskii (general purpose)** | **Operational Costs** | **- 8 578 628** | **- 9 283 490** | **- 10 072 248** |
| Krestovskii (general purpose) | Annual cost of 1 sq.m. | - 20 400 | - 20 400 | - 20 400 |
| Krestovskii (general purpose) | Total rental costs | - 2 109 360 | - 2 109 360 | - 2 109 360 |
| Krestovskii (general purpose) | Location remodelling | - 734 140 | - | - |
| Krestovskii (general purpose) | Works per 1 sq.m. | - 3 500 | - | - |
| Krestovskii (general purpose) | Materials per 1 sq.m. | - 3 600 | - | - |
| Krestovskii (general purpose) | Housing&Utility | **-** | **-** | **-** |
| Krestovskii (general purpose) | Sales equipment (cash desks, payment terminals, Acquiring (2%)) | - 715 757 | - 425 396 | - 471 793 |
| Krestovskii (general purpose) | Director computer | - 25 000 | - | - |
| Krestovskii (general purpose) | Cash desk computer | - 25 000 | - | - |
| Krestovskii (general purpose) | Printer | - 4 000 | - | - |
| Krestovskii (general purpose) | Barcode scanner | - 3 500 | - | - |
| Krestovskii (general purpose) | Check-printing machine | - 23 000 | - | - |
| Krestovskii (general purpose) | Label printer | - 33 000 | - | - |
| Krestovskii (general purpose) | Cash box | - 2 400 | - | - |
| Krestovskii (general purpose) | Spare materials | - 1 000 | - | - |
| Krestovskii (general purpose) | Software installation | - 5 000 | - | - |
| Krestovskii (general purpose) | Fiscal memory device (3Y) | - 14 000 | - | - |
| Krestovskii (general purpose) | Wires and network equipment | - 1 000 | - | - |
| Krestovskii (general purpose) | Payment terminal | - 30 000 | - | - |
| Krestovskii (general purpose) | Internet connection | - 3 600 | - 3 600 | - 3 600 |
| Krestovskii (general purpose) | Acquiring | - 295 257 | - 421 796 | - 468 193 |
| Krestovskii (general purpose) | Demonstration equipment | - 250 000 | - | - |
| Krestovskii (general purpose) | HR costs (incl. taxes, insurance and other expenditures) | - 2 952 571 | - 4 217 959 | - 4 681 934 |
| Krestovskii (general purpose) | Personnel recruitment | - 295 257 | - | - |
| Krestovskii (general purpose) | Logistics and delivery expenses | - 295 257 | - 421 796 | - 468 193 |
| Krestovskii (general purpose) | Marketing | - 1 476 286 | - 2 108 979 | - 2 340 967 |
| Krestovskii (general purpose) | Average gross profit margin | 40% | 40% | 40% |
| Krestovskii (general purpose) | Gross Income | 11 810 284 | 16 871 835 | 18 727 737 |
| Krestovskii (general purpose) | Income before taxes | 3 231 656 | 7 588 345 | 8 655 489 |
| Krestovskii (general purpose) | Income tax | - 646 331 | - 1 517 669 | - 1 731 098 |
| **Krestovskii (general purpose)** | **Net Income** | **2 585 325** | **6 070 676** | **6 924 391** |
| Krestovskii (general purpose) | Net income margin | 9% | 14% | 15% |
| **Krestovskii (general purpose)** | **Discounted Net Income** | **2 404 954** | **5 253 154** | **5 573 862** |
| **Petrogradskaya (general purpose)** | **Total sales** | **224 854 864** | **321 221 235** | **356 555 571** |
| Petrogradskaya (general purpose) | Sales per 1 sq.m. | 662 507 | 946 439 | 1 050 547 |
| **Petrogradskaya (general purpose)** | **Operational Costs** | **- 53 736 116** | **- 64 793 210** | **- 70 800 047** |
| Petrogradskaya (general purpose) | Annual cost of 1 sq.m. | -30 000 | -30 000 | -30 000 |
| Petrogradskaya (general purpose) | Total rental costs | - 10 182 000 | - 10 182 000 | - 10 182 000 |
| Petrogradskaya (general purpose) | Location remodelling | - 2 409 740 | - | - |
| Petrogradskaya (general purpose) | Works per 1 sq.m. | - 3 500 | - | - |
| Petrogradskaya (general purpose) | Materials per 1 sq.m. | - 3 600 | - | - |
| Petrogradskaya (general purpose) | Housing&Utility | **-** | **-** | **-** |
| Petrogradskaya (general purpose) | Sales equipment (cash desks, payment terminals, Acquiring (2%)) | - 2 919 049 | - 3 215 812 | - 3 569 156 |
| Petrogradskaya (general purpose) | Director computer | - 25 000 | - | - |
| Petrogradskaya (general purpose) | Cash desk computer | - 25 000 | - | - |
| Petrogradskaya (general purpose) | Printer | - 4 000 | - | - |
| Petrogradskaya (general purpose) | Barcode scanner | - 3 500 | - | - |
| Petrogradskaya (general purpose) | Check-printing machine | - 23 000 | - | - |
| Petrogradskaya (general purpose) | Label printer | - 33 000 | - | - |
| Petrogradskaya (general purpose) | Cash box | - 2 400 | - | - |
| Petrogradskaya (general purpose) | Spare materials | - 1 000 | - | - |
| Petrogradskaya (general purpose) | Software installation | - 5 000 | - | - |
| Petrogradskaya (general purpose) | Fiscal memory device (3Y) | - 14 000 | - | - |
| Petrogradskaya (general purpose) | Wires and network equipment | - 1 000 | - | - |
| Petrogradskaya (general purpose) | Payment terminal | - 30 000 | - | - |
| Petrogradskaya (general purpose) | Internet connection | - 3 600 | - 3 600 | - 3 600 |
| Petrogradskaya (general purpose) | Acquiring | - 2 248 549 | - 3 212 212 | - 3 565 556 |
| Petrogradskaya (general purpose) | Demonstration equipment | - 500 000 | - | - |
| Petrogradskaya (general purpose) | HR costs (incl. taxes, insurance and other expenditures) | - 22 485 486 | - 32 122 123 | - 35 655 557 |
| Petrogradskaya (general purpose) | Personnel recruitment | - 2 248 549 | - | - |
| Petrogradskaya (general purpose) | Logistics and delivery expenses | - 2 248 549 | - 3 212 212 | - 3 565 556 |
| Petrogradskaya (general purpose) | Marketing | - 11 242 743 | - 16 061 062 | - 17 827 779 |
| Petrogradskaya (general purpose) | Average profit margin | 40% | 40% | 40% |
| Petrogradskaya (general purpose) | Gross Income | 89 941 946 | 128 488 494 | 142 622 228 |
| Petrogradskaya (general purpose) | Income before taxes | 36 205 830 | 63 695 284 | 71 822 181 |
| Petrogradskaya (general purpose) | Income tax | - 7 241 166 | - 12 739 057 | - 14 364 436 |
| **Petrogradskaya (general purpose)** | **Net Income** | **28 964 664** | **50 956 227** | **57 457 745** |
| Petrogradskaya (general purpose) | Net income margin | 13% | 16% | 16% |
| **Petrogradskaya (general purpose)** | **Discounted Net Income** | **26 943 874** | **44 094 085** | **46 251 219** |
| **Admiralteiskii (shopping center)** | **Total sales** | **200 589 774** | **286 556 820** | **318 078 070** |
| Admiralteiskii (shopping center) | Sales per 1 sq.m. | 1 337 265 | 1 910 379 | 2 120 520 |
| **Admiralteiskii (shopping center)** | **Operational Costs** | **- 43 001 659** | **- 54 118 259** | **- 59 476 872** |
| Admiralteiskii (shopping center) | Annual cost of 1 sq.m. | -36 000 | -36 000 | -36 000 |
| Admiralteiskii (shopping center) | Total rental costs | - 5 400 000 | - 5 400 000 | - 5 400 000 |
| Admiralteiskii (shopping center) | Location remodelling | - 825 000 | - | - |
| Admiralteiskii (shopping center) | Works per 1 sq.m. | - 3 000 | - | - |
| Admiralteiskii (shopping center) | Materials per 1 sq.m. | - 2 500 | - | - |
| Admiralteiskii (shopping center) | Housing&Utility | **-** | **-** | **-** |
| Admiralteiskii (shopping center) | Sales equipment (cash desks, payment terminals, Acquiring (2%)) | - 2 676 398 | - 2 869 168 | - 3 184 381 |
| Admiralteiskii (shopping center) | Director computer | - 25 000 | - | - |
| Admiralteiskii (shopping center) | Cash desk computer | - 25 000 | - | - |
| Admiralteiskii (shopping center) | Printer | - 4 000 | - | - |
| Admiralteiskii (shopping center) | Barcode scanner | - 3 500 | - | - |
| Admiralteiskii (shopping center) | Check-printing machine | - 23 000 | - | - |
| Admiralteiskii (shopping center) | Label printer | - 33 000 | - | - |
| Admiralteiskii (shopping center) | Cash box | - 2 400 | - | - |
| Admiralteiskii (shopping center) | Spare materials | - 1 000 | - | - |
| Admiralteiskii (shopping center) | Software installation | - 5 000 | - | - |
| Admiralteiskii (shopping center) | Fiscal memory device (3Y) | - 14 000 | - | - |
| Admiralteiskii (shopping center) | Wires and network equipment | - 1 000 | - | - |
| Admiralteiskii (shopping center) | Payment terminal | - 30 000 | - | - |
| Admiralteiskii (shopping center) | Internet connection | - 3 600 | - 3 600 | - 3 600 |
| Admiralteiskii (shopping center) | Acquiring | - 2 005 898 | - 2 865 568 | - 3 180 781 |
| Admiralteiskii (shopping center) | Demonstration equipment | - 500 000 | - | - |
| Admiralteiskii (shopping center) | HR costs (incl. taxes, insurance and other expenditures) | - 20 058 977 | - 28 655 682 | - 31 807 807 |
| Admiralteiskii (shopping center) | Personnel recruitment | - 2 005 898 | - | - |
| Admiralteiskii (shopping center) | Logistics and delivery expenses | - 2 005 898 | - 2 865 568 | - 3 180 781 |
| Admiralteiskii (shopping center) | Marketing | - 10 029 489 | - 14 327 841 | - 15 903 903 |
| Admiralteiskii (shopping center) | Average gross profit margin | 40% | 40% | 40% |
| Admiralteiskii (shopping center) | Gross Income | 80 235 909 | 114 622 728 | 127 231 228 |
| Admiralteiskii (shopping center) | Income before taxes | 37 234 250 | 60 504 469 | 67 754 356 |
| Admiralteiskii (shopping center) | Income tax | - 7 446 850 | - 12 100 894 | - 13 550 871 |
| **Admiralteiskii (shopping center)** | **Net Income** | **29 787 400** | **48 403 575** | **54 203 485** |
| Admiralteiskii (shopping center) | Net income margin | 15% | 17% | 17% |
| **Admiralteiskii (shopping center)** | **Discounted Net Income** | **27 709 209** | **41 885 192** | **43 631 668** |
| **Chernyshevskaya (general purpose)** | **Total sales** | **242 125 216** | **345 893 165** | **383 941 414** |
| Chernyshevskaya (general purpose) | Sales per 1 sq.m. | 1 034 723 | 1 478 176 | 1 640 775 |
| **Chernyshevskaya (general purpose)** | **Operational Costs** | **- 54 338 439** | **- 67 229 438** | **- 73 697 640** |
| Chernyshevskaya (general purpose) | Annual cost of 1 sq.m. | -36 000 | -36 000 | -36 000 |
| Chernyshevskaya (general purpose) | Total rental costs | - 8 424 000 | - 8 424 000 | - 8 424 000 |
| Chernyshevskaya (general purpose) | Location remodelling | - 1 661 400 | **-** | **-** |
| Chernyshevskaya (general purpose) | Works per 1 sq.m. | - 3 500 | - | - |
| Chernyshevskaya (general purpose) | Materials per 1 sq.m. | - 3 600 | - | - |
| Chernyshevskaya (general purpose) | Housing&Utility | - | - | - |
| Chernyshevskaya (general purpose) | Sales equipment (cash desks, payment terminals, Acquiring (2%)) | - 3 091 752 | - 3 462 532 | - 3 843 014 |
| Chernyshevskaya (general purpose) | Director computer | - 25 000 | - | - |
| Chernyshevskaya (general purpose) | Cash desk computer | - 25 000 | - | - |
| Chernyshevskaya (general purpose) | Printer | - 4 000 | - | - |
| Chernyshevskaya (general purpose) | Barcode scanner | - 3 500 | - | - |
| Chernyshevskaya (general purpose) | Check-printing machine | - 23 000 | - | - |
| Chernyshevskaya (general purpose) | Label printer | - 33 000 | - | - |
| Chernyshevskaya (general purpose) | Cash box | - 2 400 | - | - |
| Chernyshevskaya (general purpose) | Spare materials | - 1 000 | - | - |
| Chernyshevskaya (general purpose) | Software installation | - 5 000 | - | - |
| Chernyshevskaya (general purpose) | Fiscal memory device (3Y) | - 14 000 | - | - |
| Chernyshevskaya (general purpose) | Wires and network equipment | - 1 000 | - | - |
| Chernyshevskaya (general purpose) | Payment terminal | - 30 000 | - | - |
| Chernyshevskaya (general purpose) | Internet connection | - 3 600 | - 3 600 | - 3 600 |
| Chernyshevskaya (general purpose) | Acquiring | - 2 421 252 | - 3 458 932 | - 3 839 414 |
| Chernyshevskaya (general purpose) | Demonstration equipment | - 500 000 | - | - |
| Chernyshevskaya (general purpose) | HR costs (incl. taxes, insurance and other expenditures) | - 24 212 522 | - 34 589 317 | - 38 394 141 |
| Chernyshevskaya (general purpose) | Personnel recruitment | - 2 421 252 | - | - |
| Chernyshevskaya (general purpose) | Logistics and delivery expenses | - 2 421 252 | - 3 458 932 | - 3 839 414 |
| Chernyshevskaya (general purpose) | Marketing | - 12 106 261 | - 17 294 658 | - 19 197 071 |
| Chernyshevskaya (general purpose) | Average profit margin | 40% | 40% | 40% |
| Chernyshevskaya (general purpose) | Gross Income | 96 850 086 | 138 357 266 | 153 576 565 |
| Chernyshevskaya (general purpose) | Income before taxes | 42 511 647 | 71 127 828 | 79 878 925 |
| Chernyshevskaya (general purpose) | Income tax | - 8 502 329 | - 14 225 566 | - 15 975 785 |
| **Chernyshevskaya (general purpose)** | **Net Income** | **34 009 318** | **56 902 262** | **63 903 140** |
| Chernyshevskaya (general purpose) | Net income margin | 14% | 16% | 17% |
| **Chernyshevskaya (general purpose)** | **Discounted Net Income** | **31 636 575** | **49 239 383** | **51 439 508** |

1. https://github.com/irinapendryak/master\_thesis [↑](#footnote-ref-2)
2. https://datalens.yandex/dyizpxgcxigi3 [↑](#footnote-ref-3)