Customer Analytics: Segmentation of the customer base and behavioral profiling

Completed by
Daria Lapko
Konstantin Shashkov
Timofey Sluev

Supervised by
Artem Glazkov

Moscow, 2023

Rationale

Goal

of the Project

To identify the groups of consumers with similar behavioural patterns and develop a strategy of personified behaviour-based customer interaction for an online store.

Problems

to solve

- Investigate a database of orders placed to an online store and understand its data model.
- Hypothesise the potential clusters and propose a strategy to approach them.
- Verify the quality of the data and prepare it for future analysis.
- Statistically identify the optimal number of customer groups.
- Develop and test a series of clustering models to determine the customer groups.
- Outline and describe the detected customer profiles for the successful model.
- Propose a strategy to address the customers based on their belonging to one of the determined groups.
- Assess the proposed strategy from the financial perspective.

Known Cases

of customer profiling

Technosila case

The once large Russian technology retailer performed a recency-frequency (RF) segmentation of their customer base. They identified 7 groups of consumers and developed 8 scenarios to approach them. The conducted marketing campaign resulted in 15.09% higher click rate and 348.18% higher conversion than achieved with the “Reactivation” scenario increasing the revenue by 71.51% with 43.46% less sent emails.

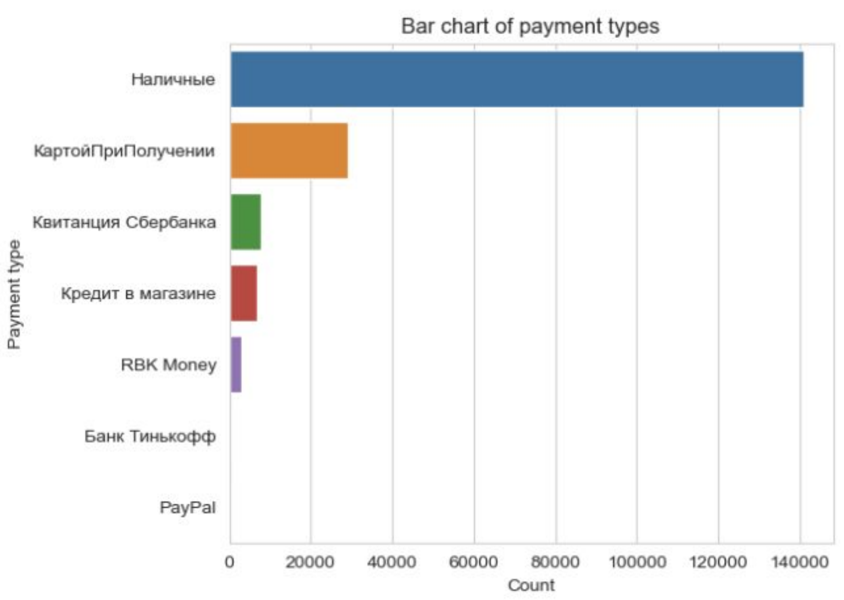
Bausch+Lomb case

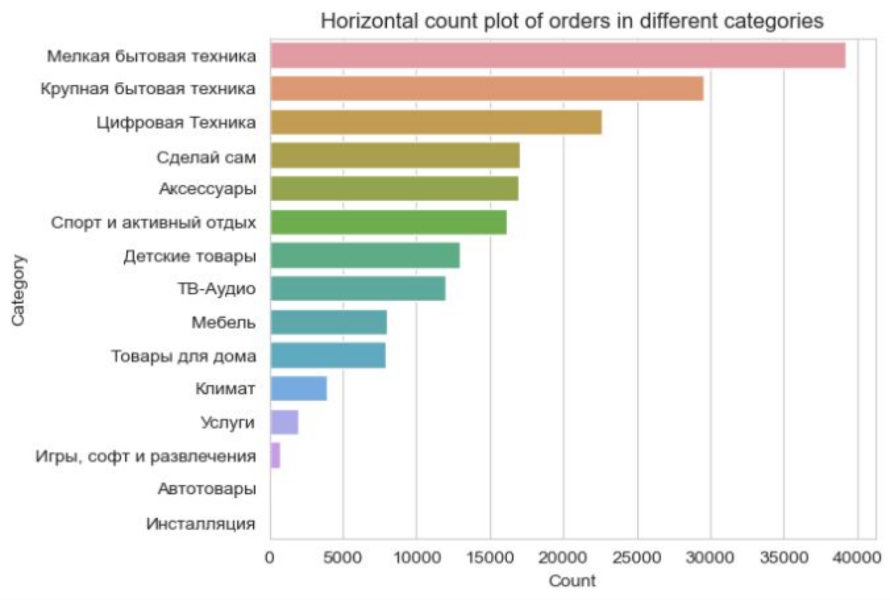
In 2020, the American eye health products manufacturer launched a loyalty programme in the Russian market and performed customer segmentation to personalise it. They identified 9 types of customers through RF analysis and 7 groups using the recency-frequency-monetary (RFM) approach. This way they determined a strategy to improve the performance of the loyalty programme in the following year.

Stages of the Project

Exploratory Data Analysis

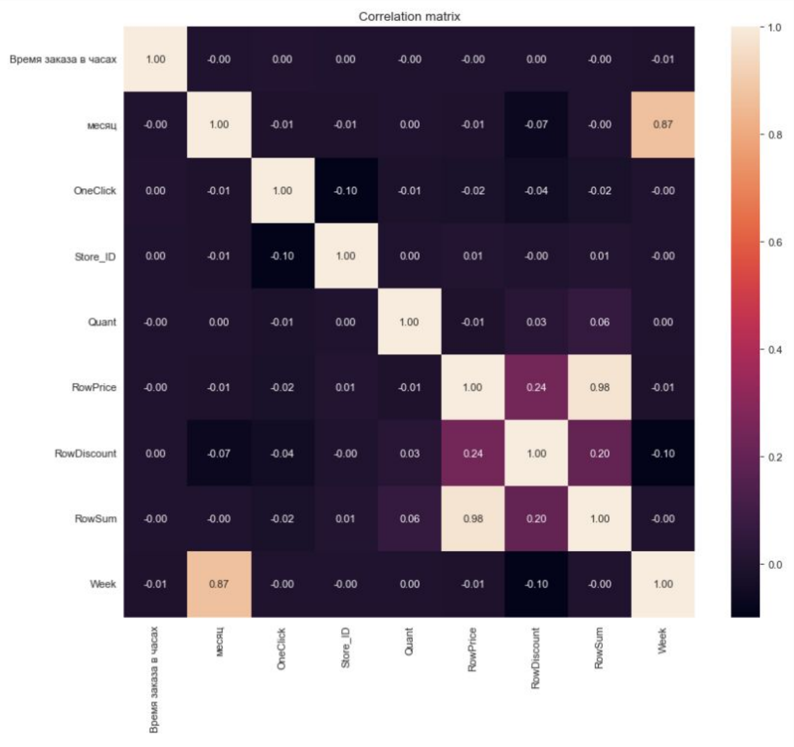
We used graphical presentation to explore the relationship between the number of orders and various characteristics of those. For instance, the graphs below suggest that cash and credit cards are the most popular payment methods, while everyday-use technology and digital goods are the most ordered product categories.





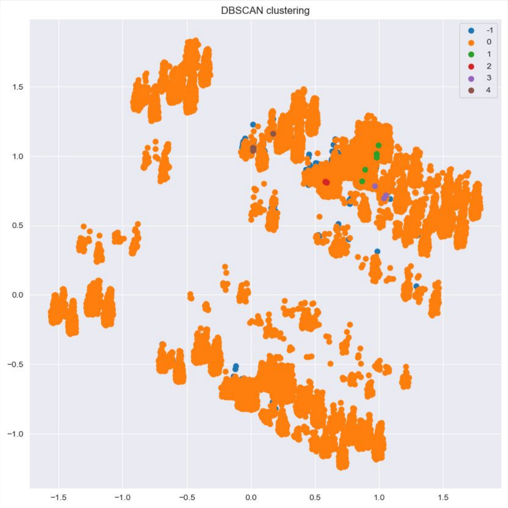
Exploratory Data Analysis

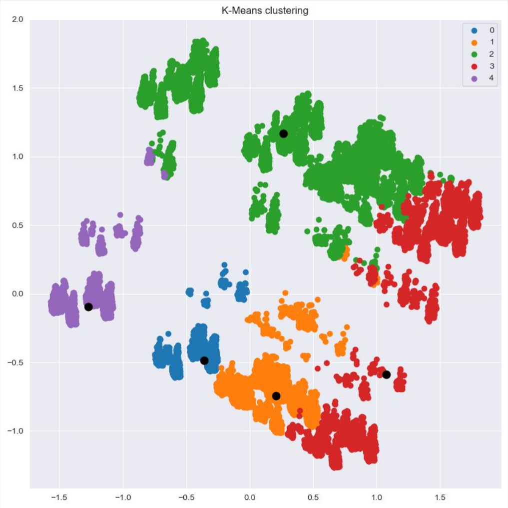
- Dropped missing values and duplicated records
- Categorical data:
    - Merged similar values in one column
    - Renamed and swapped columns for convenience
    - Dropped rows with contradictions in data
    - Dropped columns which do not allow for correct clusterization
    - Divided date into day, month and weekday
    - Encoded data for future use in models
- Numerical data:
    - Constructed and analysed correlation matrix
    - Dropped columns, highly correlated with other features

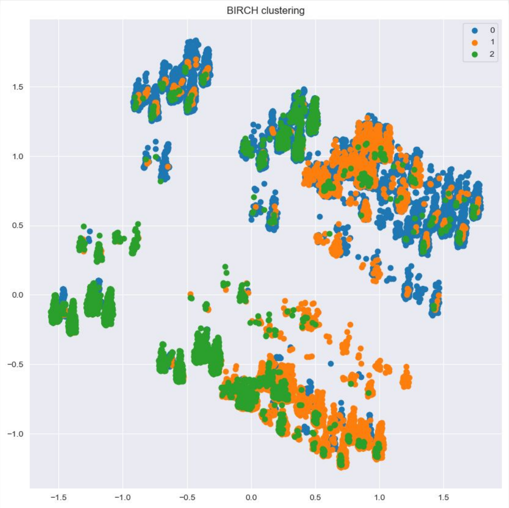


Correlation heatmap of the original dataset

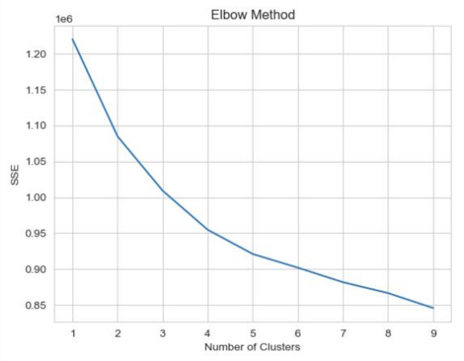
Model comparison

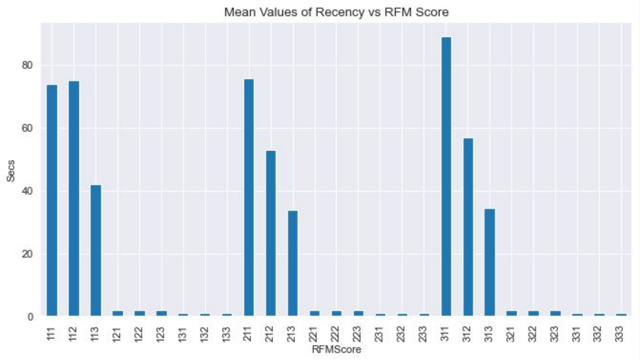






We examined 4 various clustering models. BIRCH and DBSCAN failed to detect any meaningful clusters in the provided data, so we proceeded with the Recency Frequency Monetary (RFM) method and 5-cluster k-Means algorithm.





Model comparison

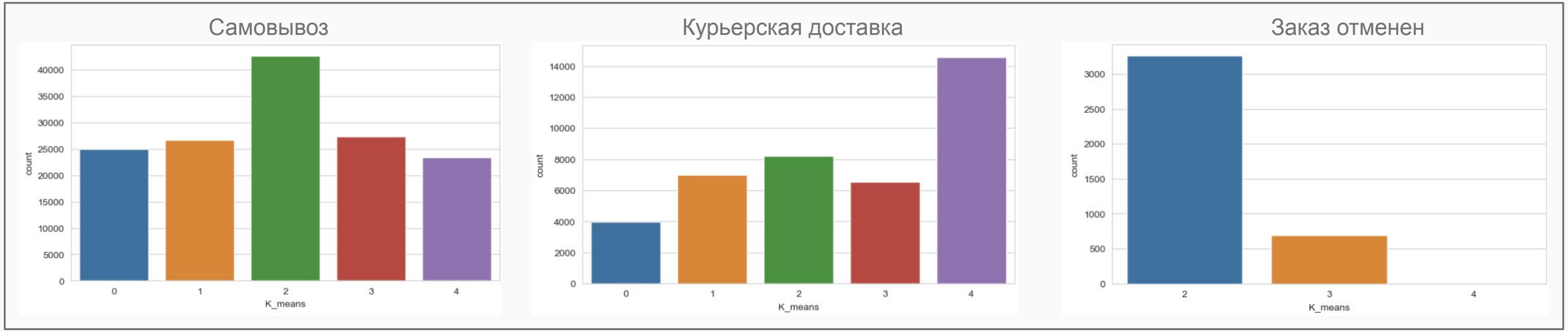
K-means has shown the largest out-of-cluster deviation and the best interpretability among all the models, so we decided to use it for our final segmentation.

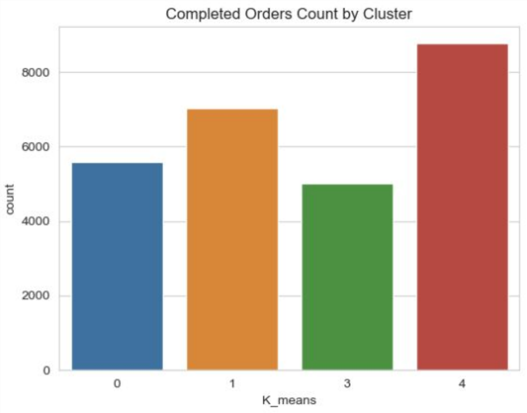
RFMScore	Время заказа в часах	OneClick	Quant	RowPrice	RowDiscount	RowSum	Week	dayNum	monthNum	weekdayNum
132	14.58	0.11	1.02	5371.11	234.33	5180.21	9.89	9.36	3.00	3.09
232	14.59	0.10	1.03	5247.84	64.12	5230.67	12.90	15.93	3.48	3.36
332	14.59	0.10	1.03	5111.23	22.18	5141.42	15.91	20.27	4.00	2.87

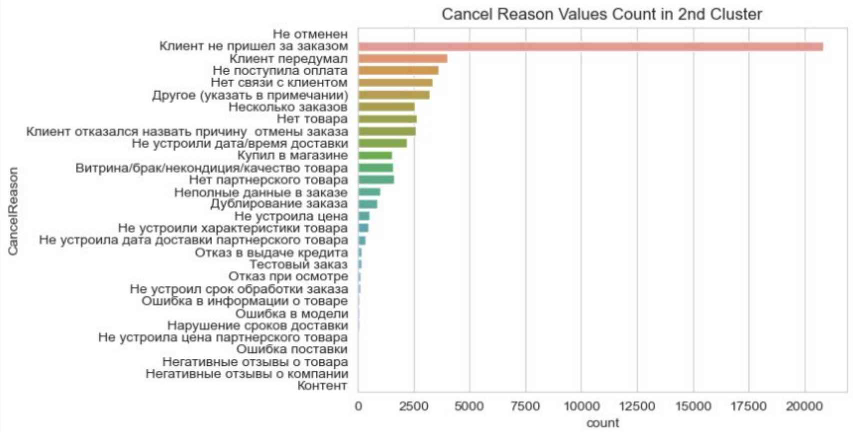
K_means	Время заказа в часах	OneClick	Quant	RowPrice	RowDiscount	RowSum	Week	dayNum	monthNum	weekdayNum
0	14.78	0.06	1.04	8754.86	169.76	8687.58	12.85	15.18	3.48	3.17
1	14.04	0.06	1.05	8987.20	202.36	8904.68	12.89	15.22	3.49	3.17
2	14.29	0.12	1.10	11824.06	136.60	12072.54	12.83	15.26	3.47	3.07
3	14.76	0.08	1.05	9199.56	144.72	9175.07	12.95	15.29	3.50	3.11
4	14.67	0.05	1.06	7511.27	125.45	7557.35	12.92	15.23	3.49	3.06

Customer segmentation with K-means

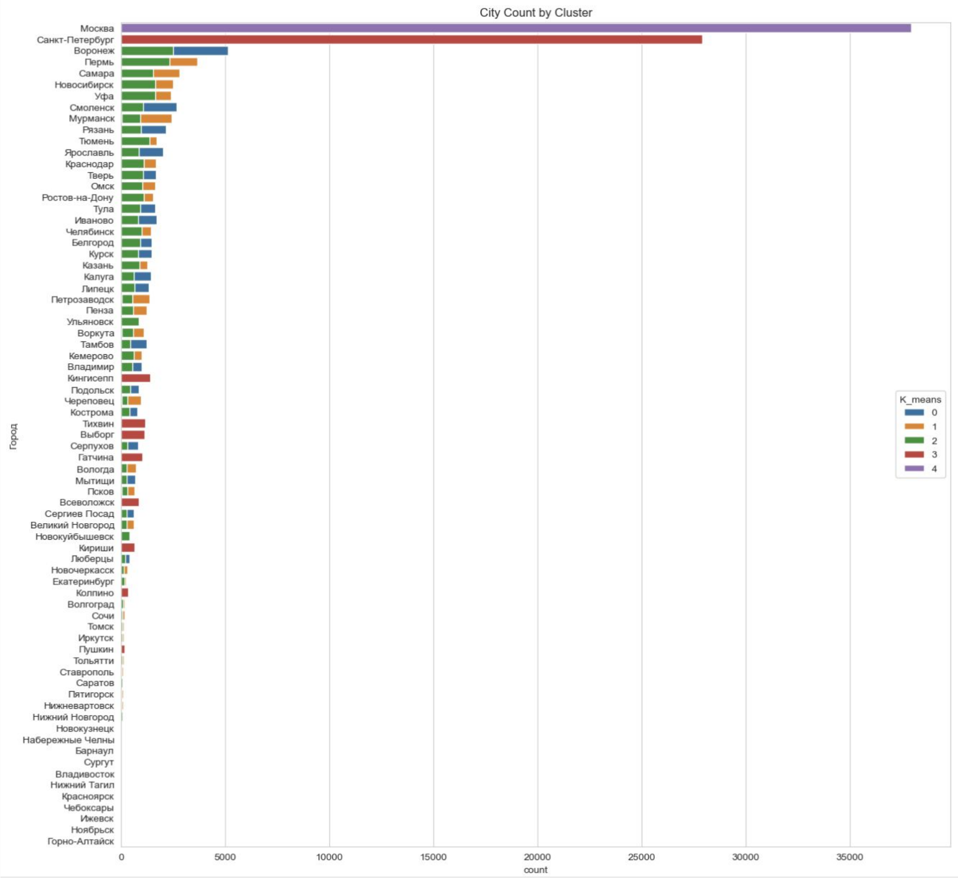
Delivery Type by Cluster







Customer segmentation with K-means



1. **Cluster 0:** The smallest cluster. The customers are mainly located in large non-capital cities of Central Russia. Their average order sum is neither high nor low, and they prefer self-delivery. Interestingly, they have the latest order time (about 15:00).
2. **Cluster 1:** Customers from this cluster live in province. Their order sum is the median among the clusters, and they tend to have the highest discounts, using the delivery service relatively often and never cancelling their orders.
3. **Cluster 2:** The largest cluster. Located in province. Has the highest average order price. Only cancels orders, while the most frequent reasons is not coming to collect them, deciding not to buy or failing to pay.
4. **Cluster 3:** Mostly occupied by customers from Saint-Petersburg and the Leningrad region. Has a fairly low share of completed orders. It also has the second-largest average order price among all clusters (besides cluster 2).
5. **Cluster 4:** Has the lowest average price and discount. The preferred delivery type is courier, and the orders are cancelled the least among all clusters. Almost completely occupied by Moscow city customers.

Strategy to Approach the Customer Groups

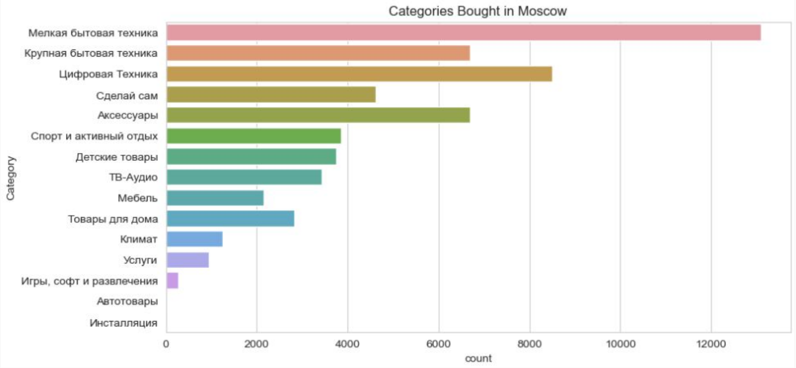
1. **Cluster 0:** This cluster seems to refer to the middle class of large cities in province. One should probably target them with averagely priced goods, and discounts might be efficient for them. Meanwhile, they are not interested in services such as delivery. Communication with this group may be held by both the traditional means like phone calls or SMS, and modern analogues, such as E-mail, messengers, and social media.

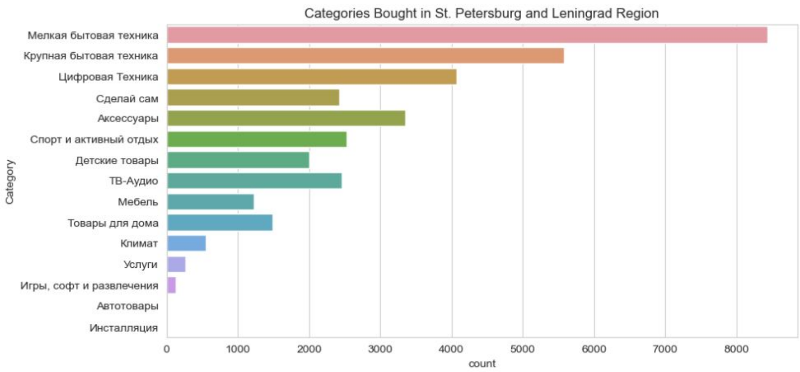
2. **Cluster 1:** We should focus on making appealing discounts and advertising them properly to attract customers from this cluster. We should also develop the delivery services for them. This group is likely to be used to traditional means of communication such as SMS, phone calls, or even mail.

3. **Cluster 2:** A marketing campaign addressing this group of customers is not likely to be efficient since they do not appear to be serious about their orders. Note that this cluster consists fully of people who dismiss their orders, so it might be reasonable not to target this audience.

4. **Cluster 3:** We should target this cluster keeping in mind that it contains primarily citizen from Saint Petersburg and the Northwestern region. Hence, it might be more sensible to suggest them northern region-specific goods. It may also make sense to launch marketing campaigns related to the local events and celebrations. People in this cluster are likely to use modern means of communication like E-mail, messengers and social media.

5. **Cluster 4:** The customers in this cluster live in Moscow and make small orders with lower total price often using delivery service. Discounts do not seem to be efficient for them, while benefits on delivery might be. Once again, region- and event-specific goods might catch their attention. Most likely, they prefer modern ways of communication, such as E-mail, messengers, and social media.





Conclusion & Recommendations
for Future Work

Conclusion

- Having analysed the online store data, we identified 5 groups of customers.
- We noticed a considerable correlation between geographical regions and customer behaviour. In particular, the behavioural patterns of those living in Moscow and St. Petersburg significantly differ from each-other and the rest of the country.
- Self-delivery appears to be quite popular in province, but seemingly badly available in smaller cities. Meanwhile, in capital cities, courier shipping is more popular. The company may want to further develop their shipping methods specifically for each region.
- Certain customers seem not to be serious about their orders. They fail to pay or collect their delivery all the time. It might be a fairly complicated problem to deal with. The store cannot simply stop serving these clients, but loses money on shipping and storage. The wisest approach in this case would be to stop targeting this audience to cut costs since marketing investment is not likely to attract any revenue in their case.
- The means of communication might differ depending on location. While popular methods like E-mail and SMS may be universally applicable, people in large cities might be more responsive to modern means, such as messengers and social media.

References

- Rakhmanova, Vladislava. “Kejs ‘Tekhnosily’: Rost Povtornyh Zakazov v 4,5 Raza.” *Vc.Ru*, 7 June 2017, vc.ru/flood/24340-rfm-segmentation?ysclid=liygzqvj24960921270.
- “Kejs Bausch+Lomb: Segmentaciya Klientov v Programme Loyal’nosti FMCG-Brenda s Pomoshch’yu RFM-Analiza.” *Inbox Marketing*, inbox-marketing.ru/cases/rfm-segmentaciya-klientov. Accessed 16 June 2023.
- ProductStar. “RFM-Analiz: Segmentiruem Klientov Dlya Optimizacii Reklamnoj Kampanii.” *Vc.Ru*, 21 July 2020, vc.ru/s/productstar/143804-rfm-analiz-segmentiruem-klientov-dlya-optimizacii-reklamnoy-kampanii.

Thank you for your attention!

Timofey Sluev
tvsluev@edu.hse.ru

Konstantin Shashkov
kashashkov@edu.hse.ru

Daria Lapko
dalapko@edu.hse.ru