



Faculty of Computer Science
Data Science and Business Analytics

Research Project Report

Data Analysis and Visualization Using Yandex DataLens: A Case Study of Consumer Transactions

Presented by

Fishchenko Tamara Alekseevna,
2nd year student,
group 246, DSBA

Project supervisor

Ivanov Andrey Aleksandrovich,
Associate Professor
T-Bank Core Department, HSE University

Moscow 2026



Introduction

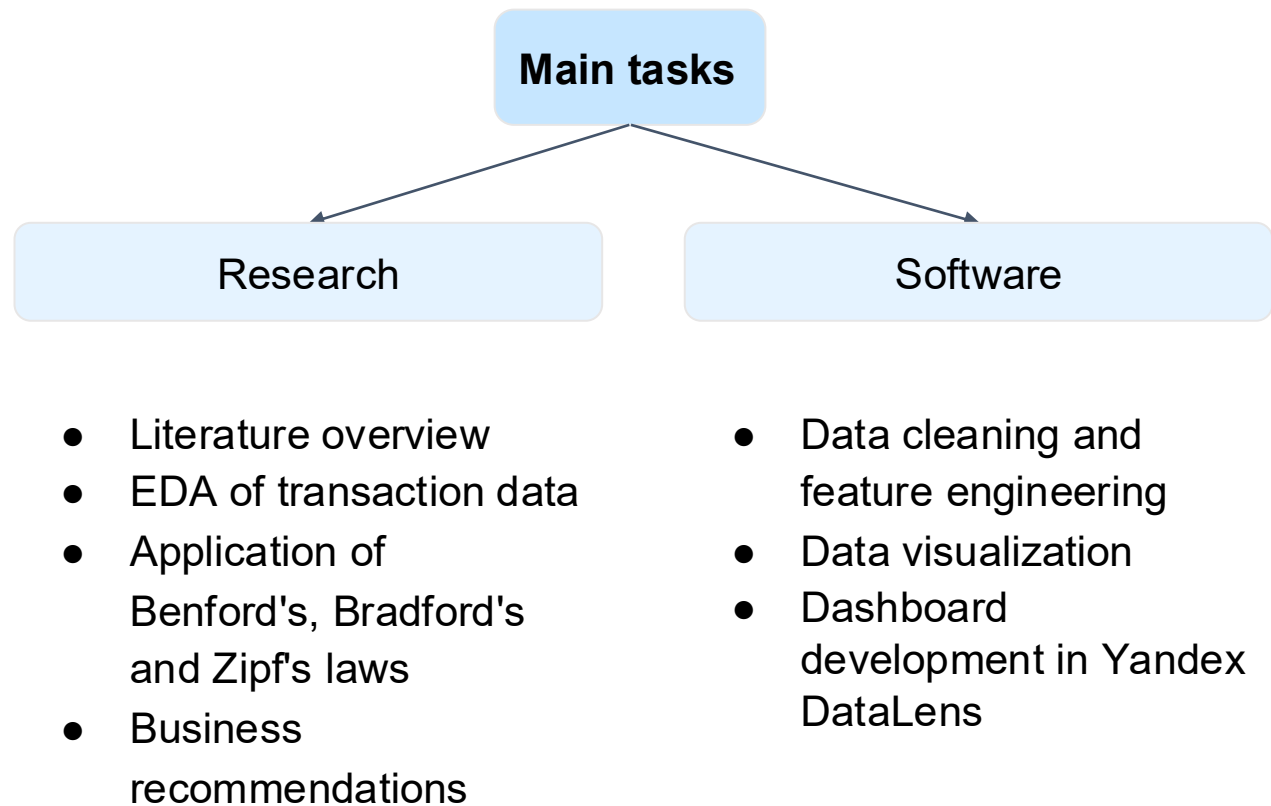
What problem does this project solve?

- The main challenge in modern business is not lack of data but the ability to apply it for strategic choices.
- Yandex DataLens is a cloud-based Business Intelligence (BI) tool for data integration, analysis, and interactive dashboard development.
- The research demonstrates how real raw transactional data can be transformed into actionable business insights of customer behavior patterns and data-driven decision-making.





Objectives, tasks and relevance



Objective

Segment consumers using RFM analysis, study purchasing behavior, develop marketing strategies and verify data reliability using Benford's law.

Relevance

In the digital economy, companies fail to use massive transactional data for strategic decisions, losing their competitive edge.

Target audience

Business analysts and managers, statisticians, students in business informatics.



Dataset description and preparation

Dataset variables

Variable	Description
InvoiceNo	Unique transaction code ('C' prefix = cancellation)
StockCode	Unique product code
Description	Product name
Quantity	Units purchased
InvoiceDate	Date and time of transaction
UnitPrice	Price per unit (GBP)
CustomerID	Unique customer code
Country	Country of transaction

Description of fields

Metric	Before	After
Total records	541,909	531,119
Unique customers	4,372	4,334
Unique StockCodes	4,070	3,811
Countries	38	35

Data preprocessing

- Removed duplicates and invalid transactions.
- Cleaned product and geographic information.
- Corrected product code inconsistencies.
- Filtered non-sales records.
- Created analytical features for revenue, returns and time analysis.



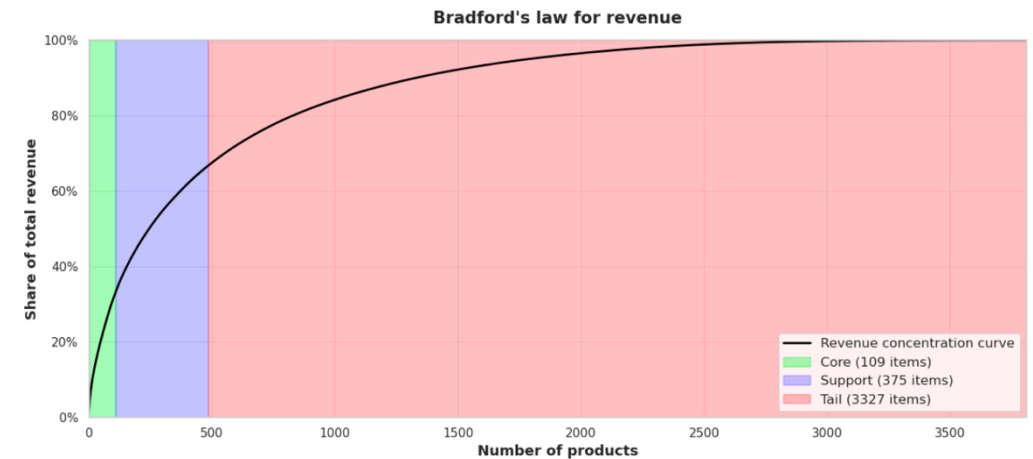
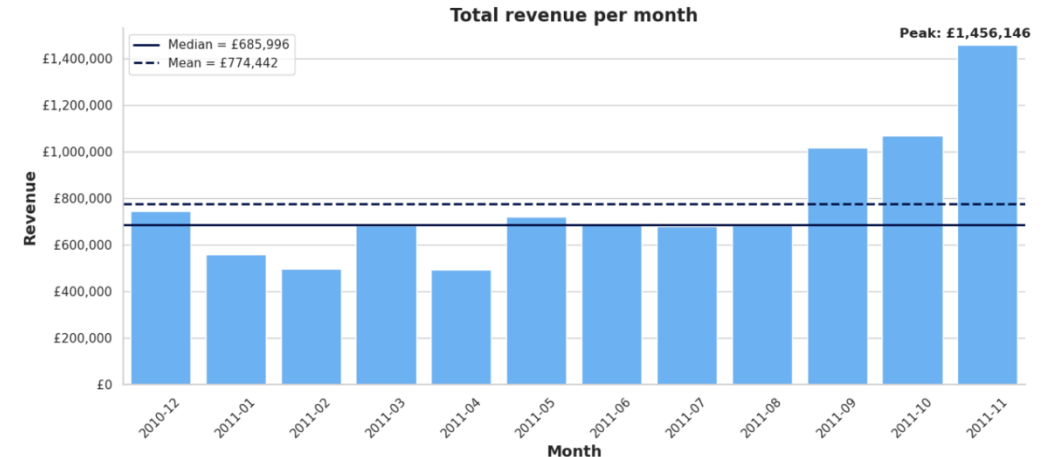
Exploratory Data Analysis (EDA)

Revenue and product concentration

- **Strong seasonality:** Revenue peaked at £1.46M in November and increased during holiday period.
- **Bradford distribution:** 109 core products generated one-third of total revenue, the same as 375 support products and 3,327 long-tail products combined.
- **Volume $\neq$ Revenue:** Low-cost items drove sales volume, while mid-priced items generated most revenue.

Main strategic insights:

- Diversify sales across months.
- Protect main 109 products.
- Promote cross-selling and bundling for mid-priced items.





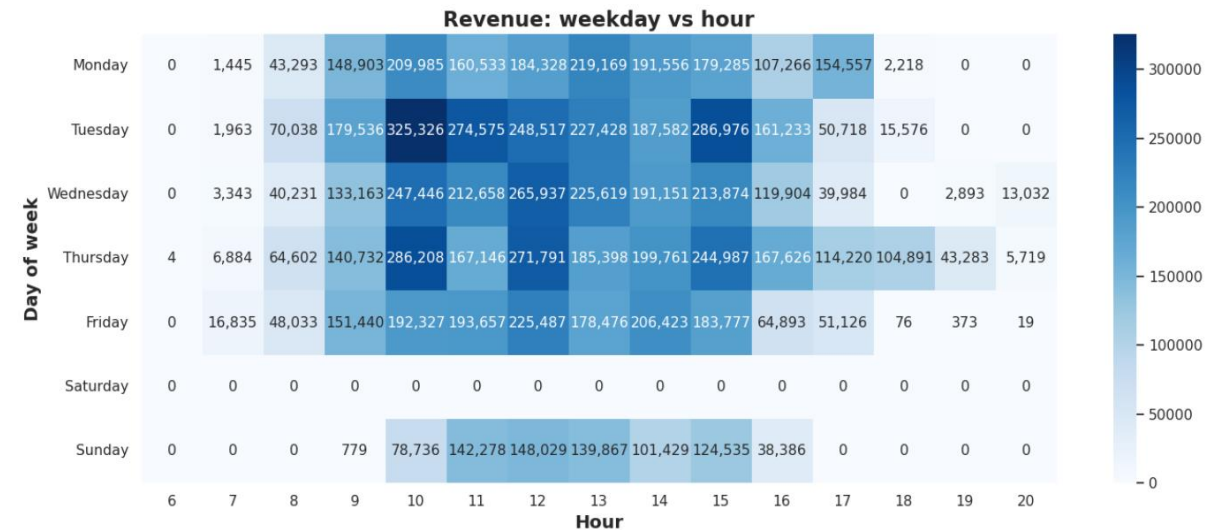
Exploratory Data Analysis (EDA)

Temporal patterns and returns

- **B2B purchasing pattern:** Peak activity occurred on Tuesday at 10 AM, and no transactions were recorded on Saturdays.
- **Rapid returns:** 14.8% of returns occurred within 1 hour (median return time: 8 minutes).
- **Delayed returns:** More than 50% of returns occurred after the purchase day, suggesting inventory adjustments and seasonal stock clearance.

Main strategic insights:

- Use Saturdays for maintenance and operational updates.
- Target customers before peak business hours.
- Implement delayed order processing to reduce rapid cancellations.



Period	Number of returns	Returns (%)
Within 1st hour	425	14.844569
Day 0	641	22.389102
Day 1	84	2.933985
Day 2	164	5.728257
Day 3	116	4.051694
Days 4-7	366	12.783793
Days 8-30	838	29.269997
Over 30 days	654	22.843171



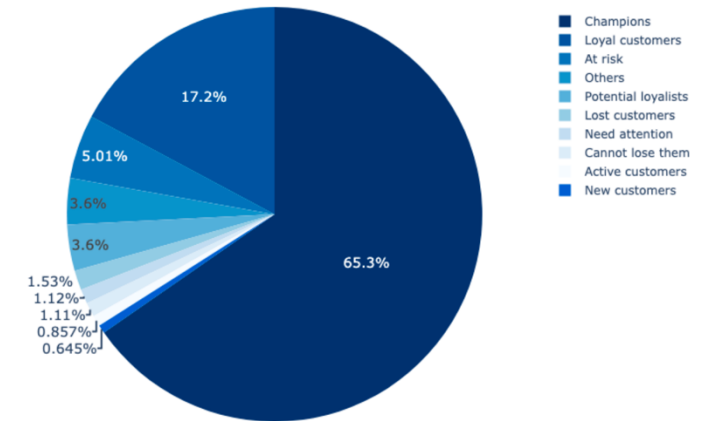
RFM (Recency, Frequency and Monetary) analysis

- Champions (22.1%) generated 65.3% of total revenue.
- Champions and Loyal Customers generated ~80% of revenue from only 26.3% of customers.
- At Risk and Lost Customers represent nearly one-third of the customer base.
- At Risk customers contribute 5.0% of revenue and are prime retention targets.
- Frequency showed the strongest relationship with customer value ($r = 0.55$).

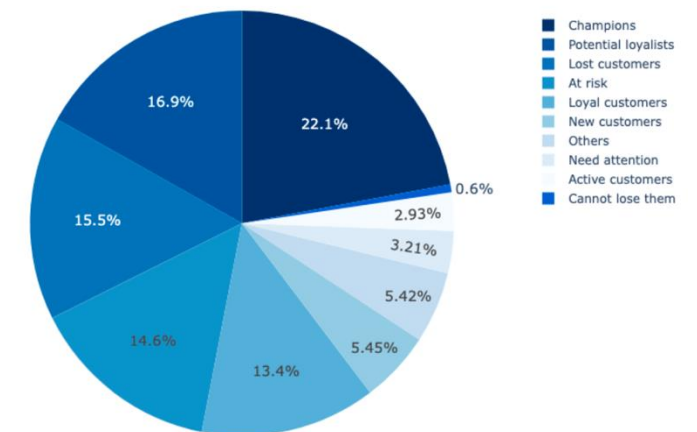
Strategic recommendations:

- Reward Champions through VIP and loyalty programs.
- Improve onboarding and engagement for New Customers.
- Launch reactivation campaigns for At Risk and Lost Customers.
- Provide dedicated support for high-value "Cannot Lose Them" customers.

Revenue share of customer segments



Customer segments



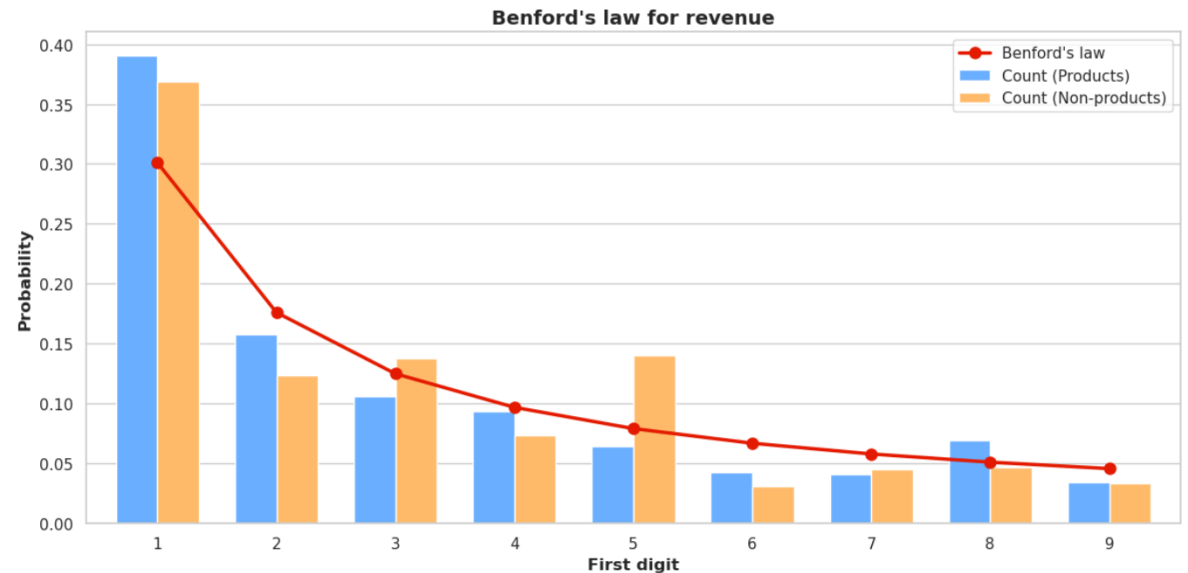


Benford's analysis

- Product and non-product transactions deviated from the theoretical Benford distribution.
- Product prices showed a strong concentration around psychological pricing endings (.95, .25, .65).
- Non-product transactions exhibited fixed-price patterns associated with operational fees and postage costs.
- Observed deviations are consistent with business practices rather than anomalous behavior.

Conclusion

- No evidence of fraud or systematic manipulation was detected, confirming the dataset as a reliable foundation for further analysis and dashboard development.





Final results

- A complete analytical workflow was developed, including data cleaning, exploratory analysis, customer segmentation, anomaly detection, and dashboard implementation.
- Customer purchasing behavior exhibited strong seasonality, revenue concentration, and distinct return patterns
- RFM analysis identified Champions as the most valuable segment, generating 65.3% of revenue while representing only 22.1% of customers
- Benford's, Bradford's, and Zipf's laws provided additional insights into product distributions, pricing behavior, and data reliability.
- No evidence of fraud or systematic manipulation was detected; observed deviations were explained by legitimate business practices.
- • An interactive dashboard was developed in Yandex DataLens to support customer analysis and data-driven decision-making.



Dashboard

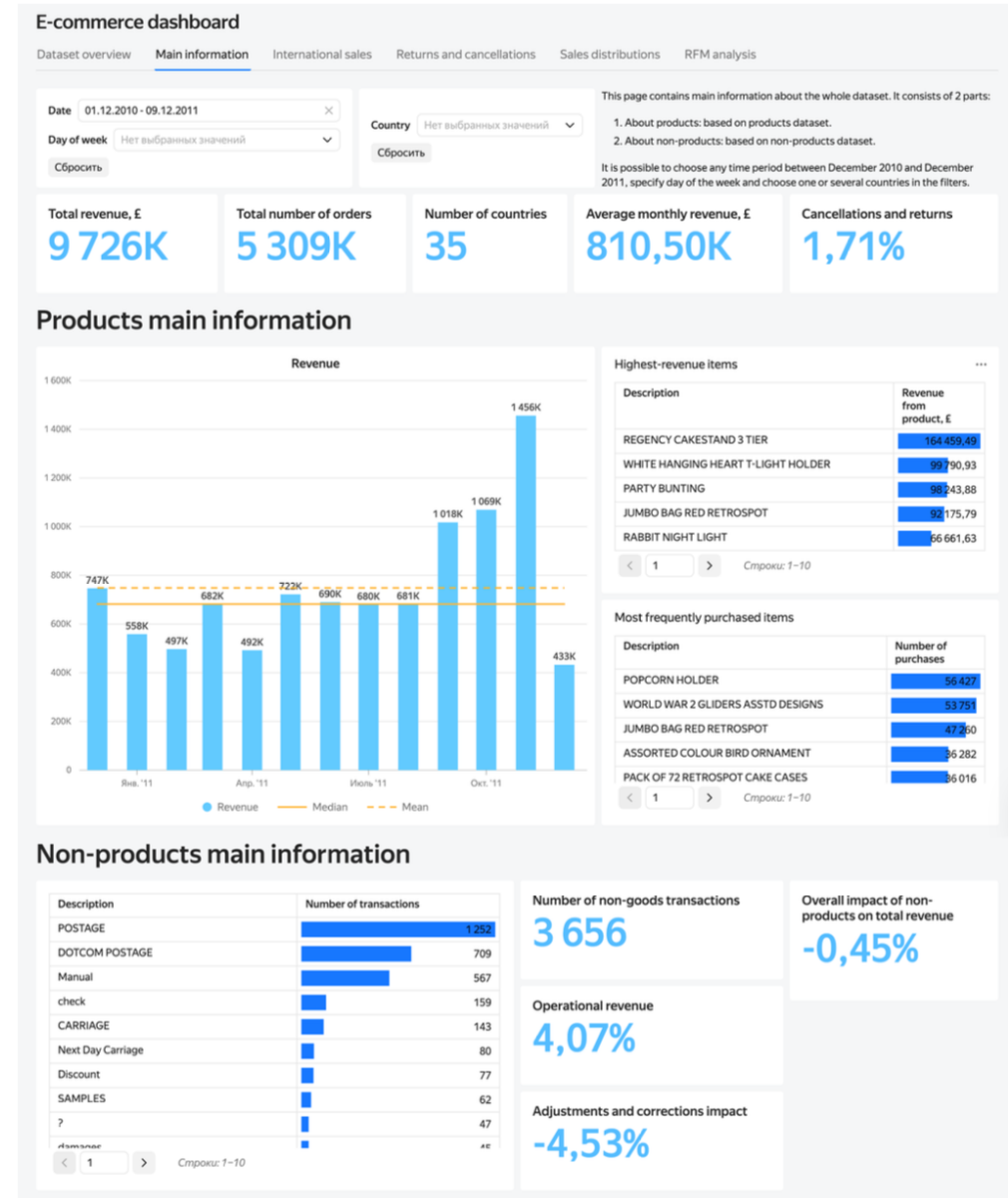


https://datalens.yandex/31qt_b886gkiin

Github



https://github.com/Tamati8/ConsumerTransactions_CaseStudy_Fishchenko

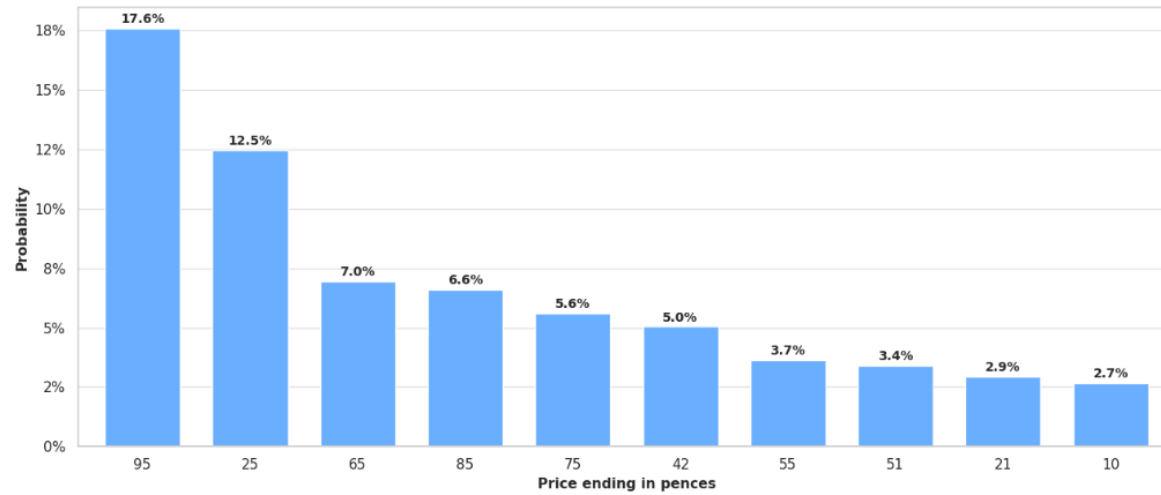




Segment	Count	Description
Champions	956	Highly valuable active customers
Potential loyalists	731	Potential future champions
Lost customers	671	Inactive low-value customers
At risk	633	Previously active customers
Loyal customers	580	Stable high-value customers
New customers	236	Recently acquired customers
Need attention	139	Average customers
Active customers	127	Customers with moderate activity and lower monetary value
Cannot lose them	26	High-value inactive customers
Others	235	Moderate-value rarely active customers



Top 10 price endings (Products)



Top 10 price endings (Non-products)

