

# **Разработка аналитического приложения для прогнозирования оттока клиентов и анализа розничных продаж**

**Партнер: Ростелеком**



# Наша команда: Скрытый пул



**Артемий:**  
Full-stack dev / аналитик



**Сергей:**  
Аналитик: проверка  
гипотез



**Дмитрий:**  
Аналитик / капитан



**Марк:**  
ML-инженер:  
Разработка модели



**Илья:**  
ML-инженер:  
оптимизация модели

# Решение

## Задача:

- Создание аналитического приложения прогнозной модели оттока клиентов и интерактивных дашбордов, позволяющие провести анализ розничных продаж товаров и клиентскую аналитику.

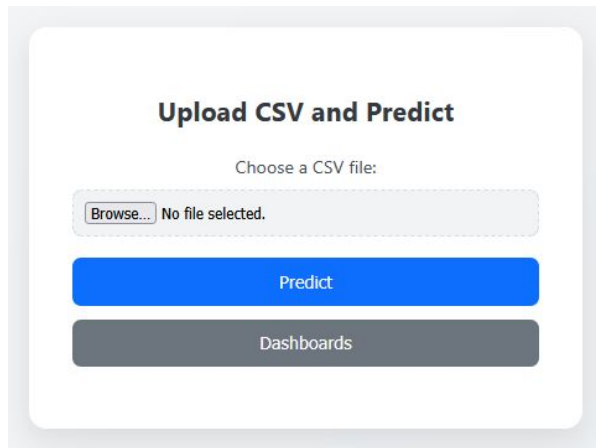
## Решение данной задачи было реализовано следующим образом:

- Проведение аналитического исследования (RFM, ABCDX, ABC-XYZ - аналитика, выстраивание гипотез)
- Разработка веб-приложения для прогнозирования оттока клиентов с поддержкой интерактивных дашбордов

# Техническая реализация

Для решения поставленной задачи были использованы следующие инструменты:

- **Аналитика:** python (pandas, numpy, scipy), SQL
- **Построение модели:** scikit-learn, LightGBM, SHAP
- **Визуализация:** matplotlib, seaborn, plotly, Grafana



**Архитектура приложения:**

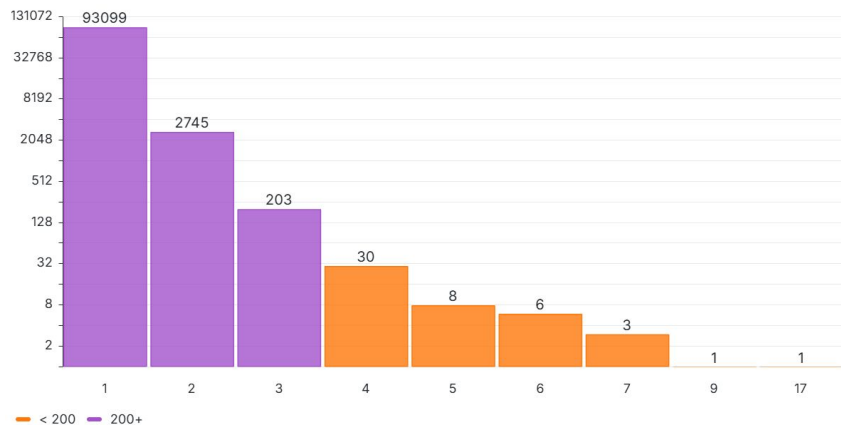
Python(FastAPI), PostgreSQL,  
Grafana, Nginx

# **Аналитическое исследование**

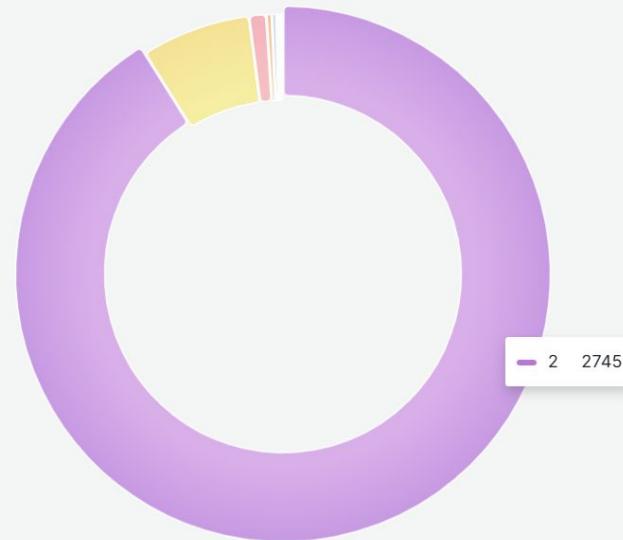


# Сильный дисбаланс

Распределение количества заказов по всем пользователям



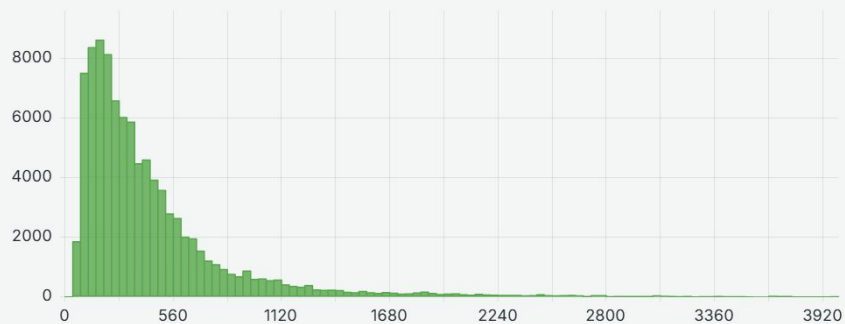
Статистика по количеству заказов среди пользователей, сделавших более 1 заказа



## Анализ метрик

Метрики

LTV-distribution

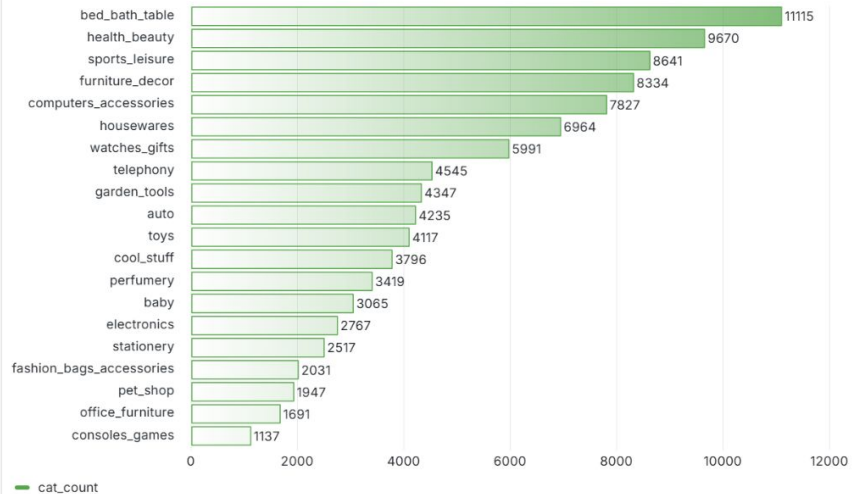


MAU

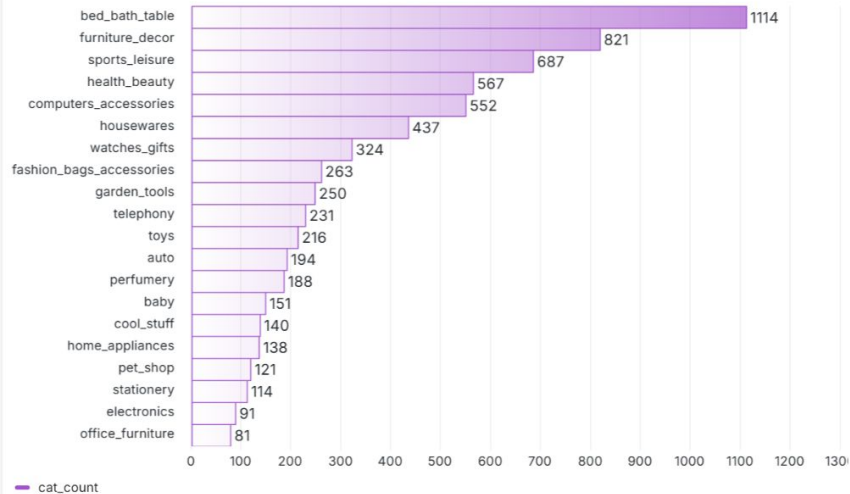


## Анализ популярности категорий

Популярность категорий товаров среди всех пользователей



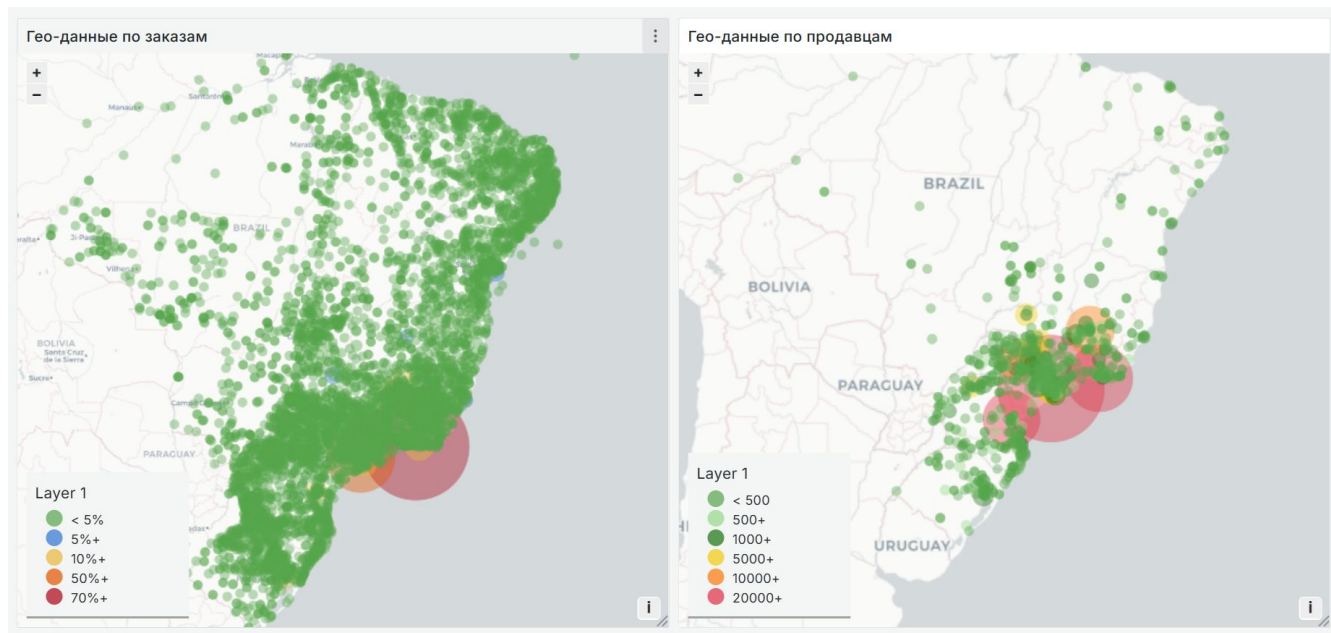
Популярность категорий товаров среди пользователей, сделавших более одного заказа



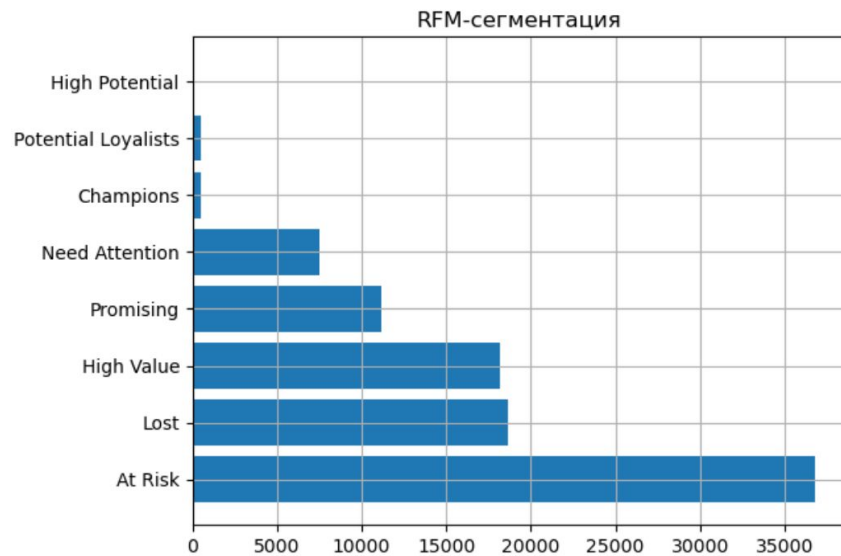
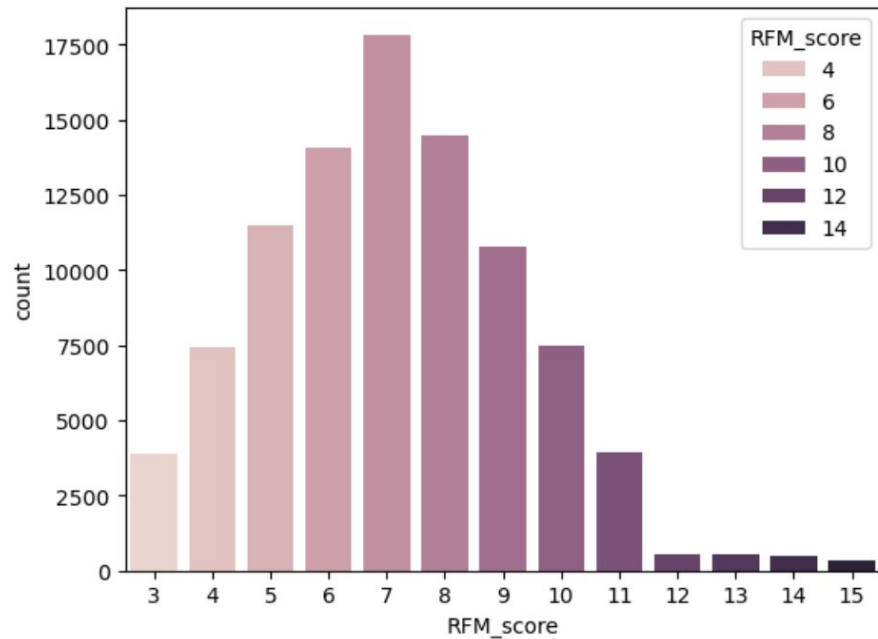


# EDA

## Изучение геоданных

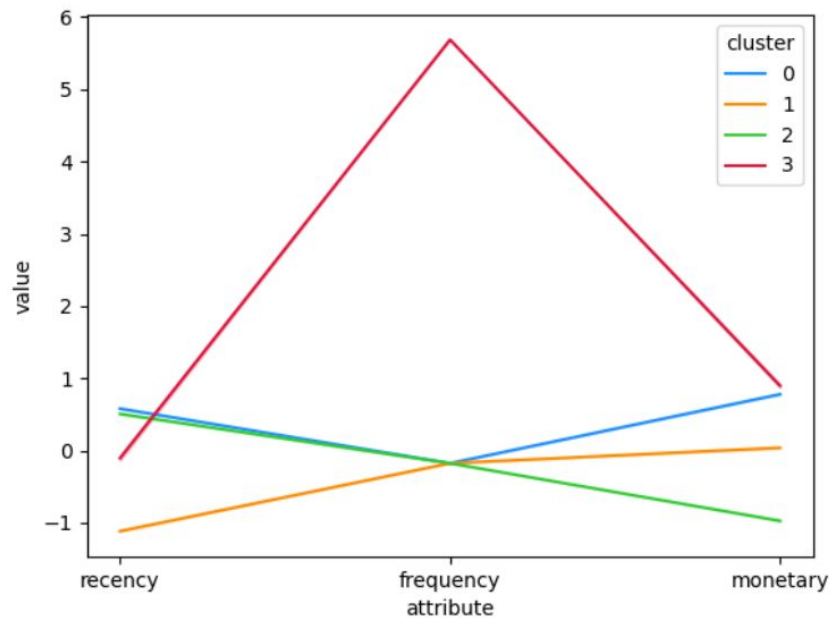
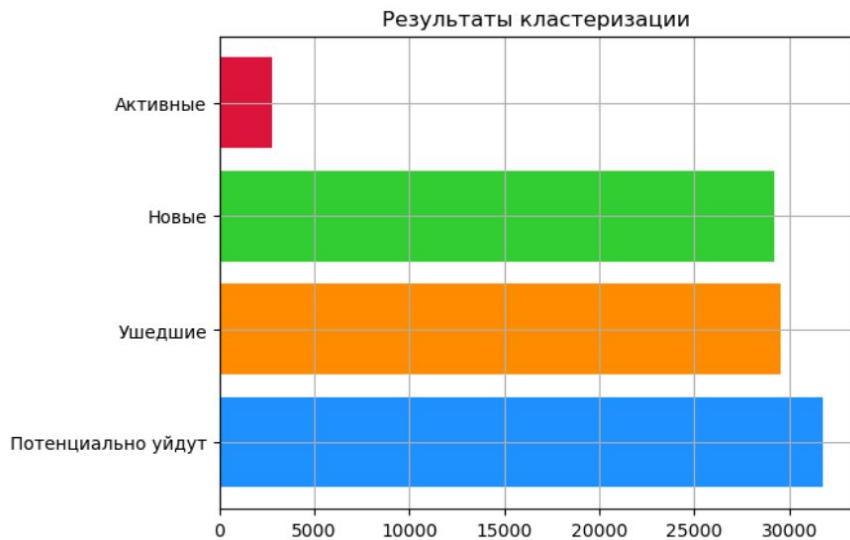


# RFM-анализ



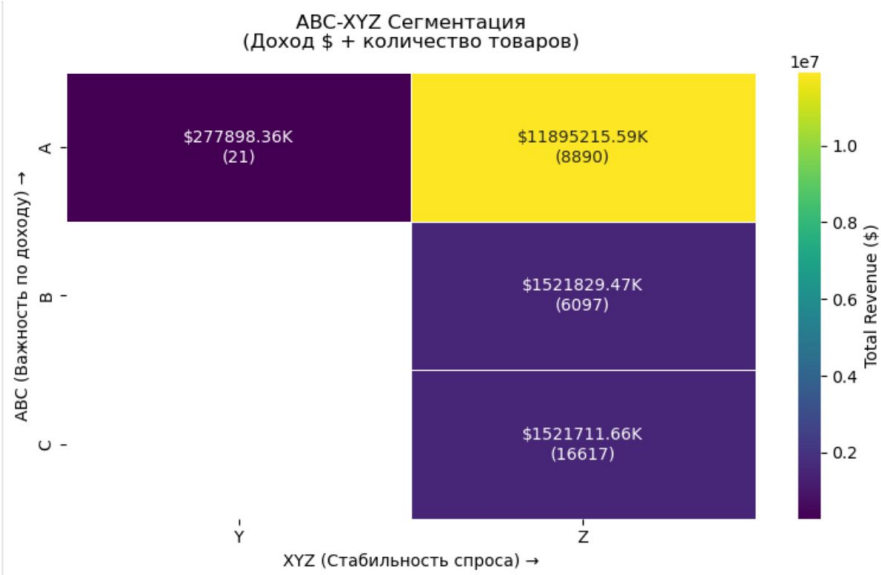
# Кластеризация клиентской аудитории

Кластеризация на базе результатов FRM-анализа



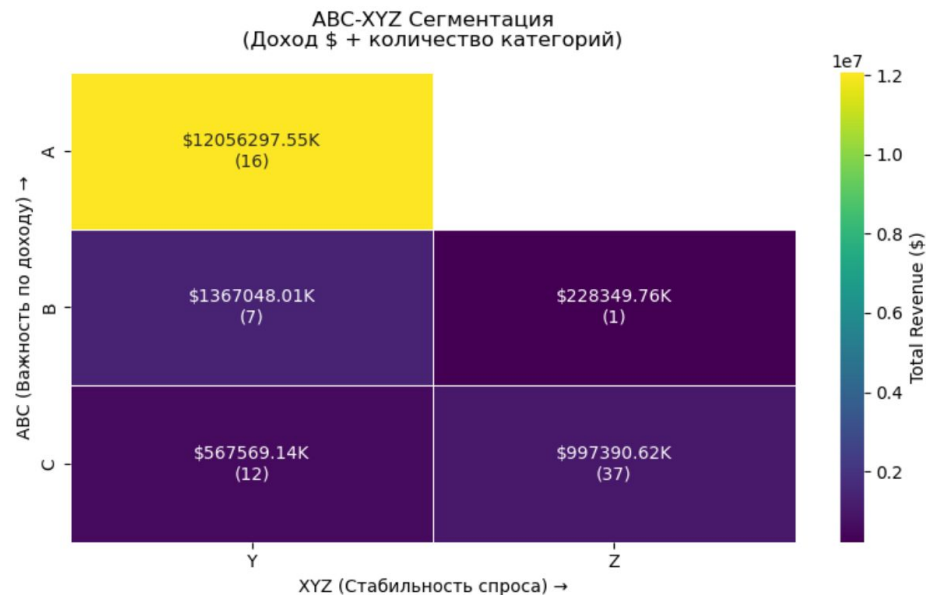
# ABC/XYZ-анализ по товарам

	ABC_XYZ	total_sales	average_sales
1	AZ	11895215.59	517183.286522
2	BZ	1521829.47	66166.498696
3	CZ	1521711.66	66161.376522
0	AY	277898.36	12082.537391



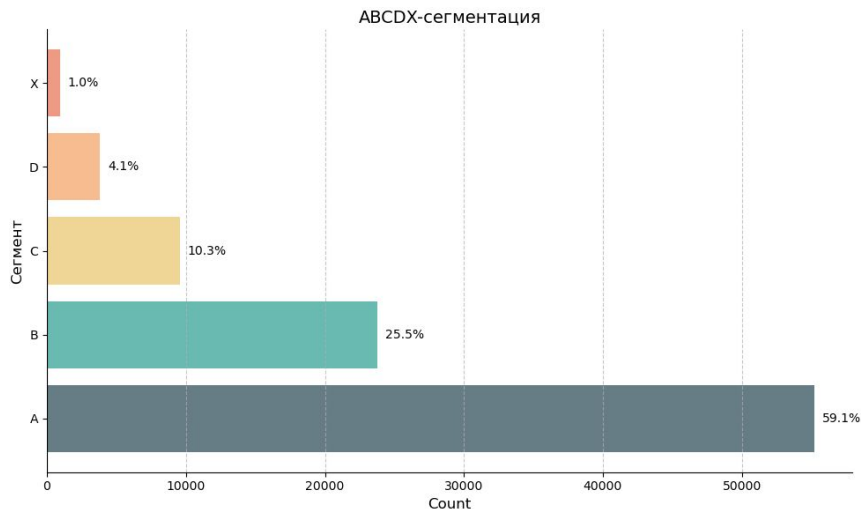
# ABC/XYZ-анализ по категориям товаров

	ABC_XYZ	total_sales	average_sales
0	AY	12056297.55	524186.850000
1	BY	1367048.01	59436.870000
4	CZ	997390.62	43364.809565
3	CY	567569.14	24676.919130
2	BZ	228349.76	9928.250435

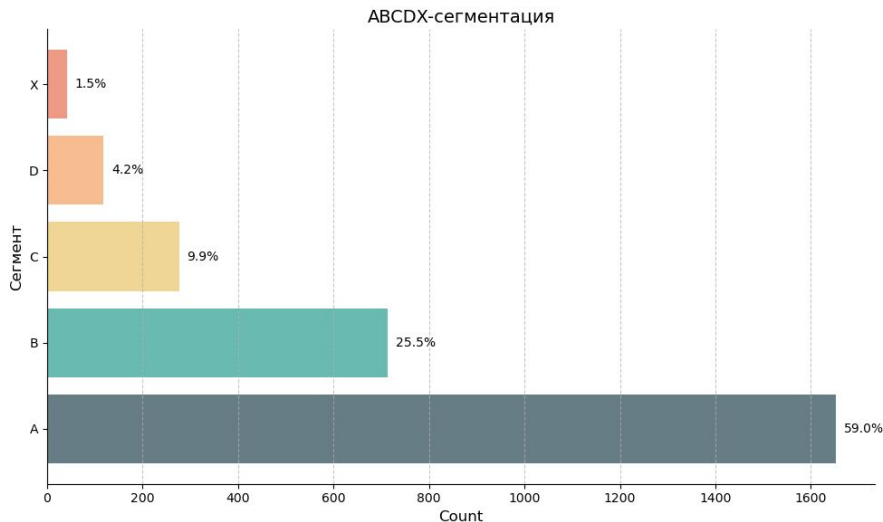


# ABCDX-сегментация

По всем пользователям



По пользователям, совершившим более одного заказа

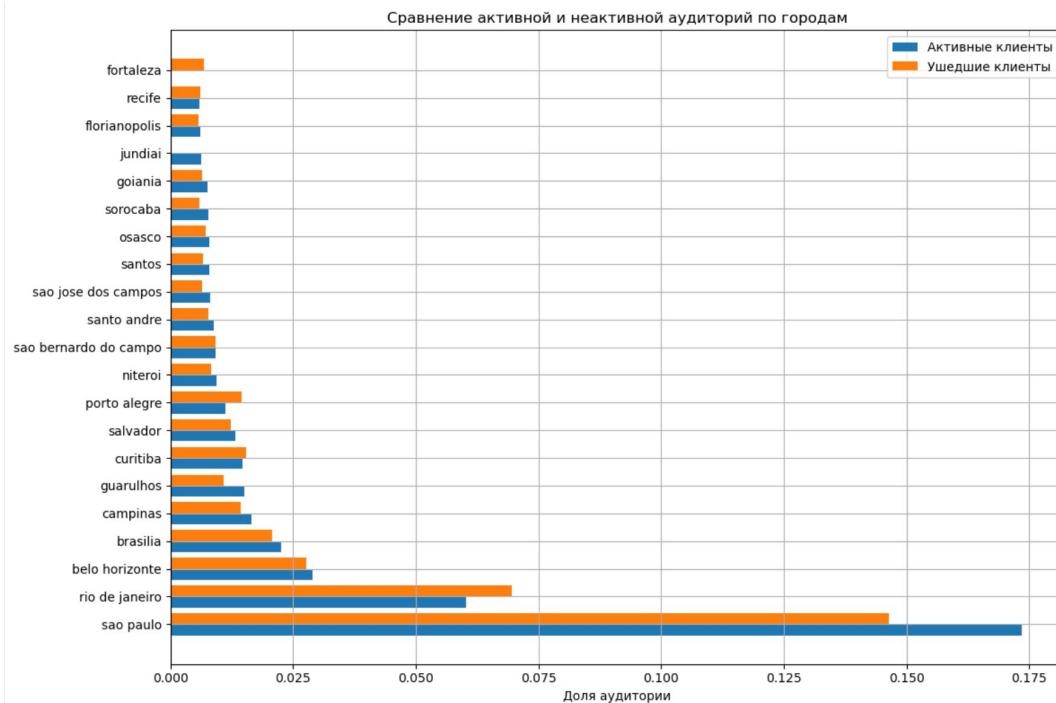


# **Формирование гипотез**



# Формирование гипотез

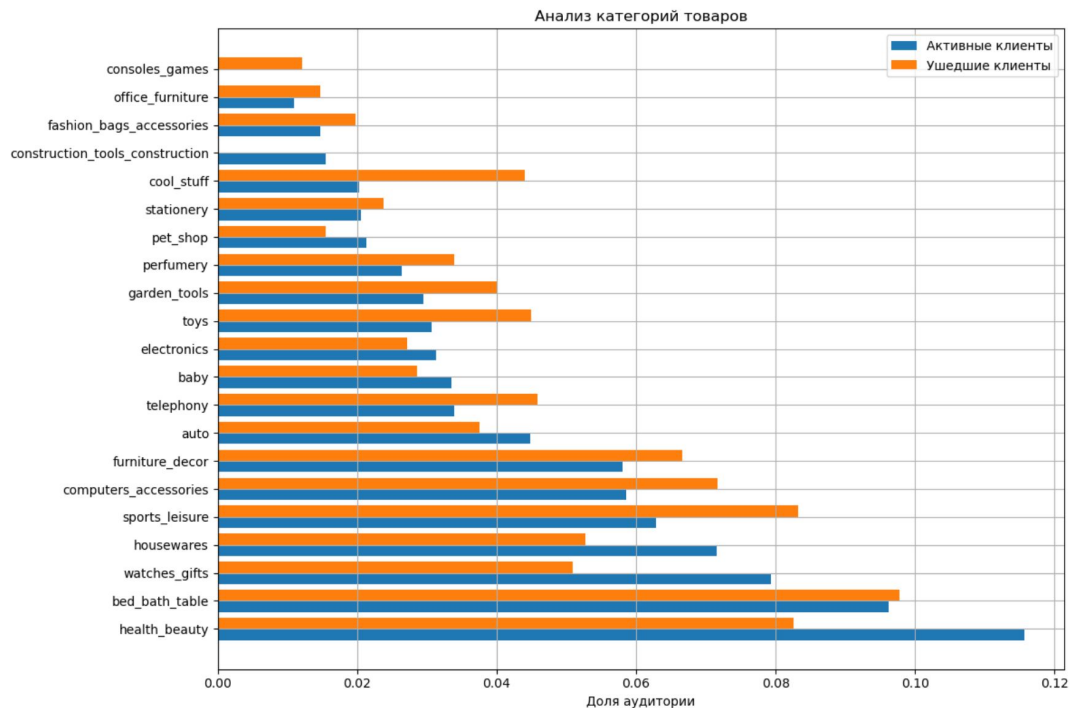
## Зависимость показателей активности от города клиента





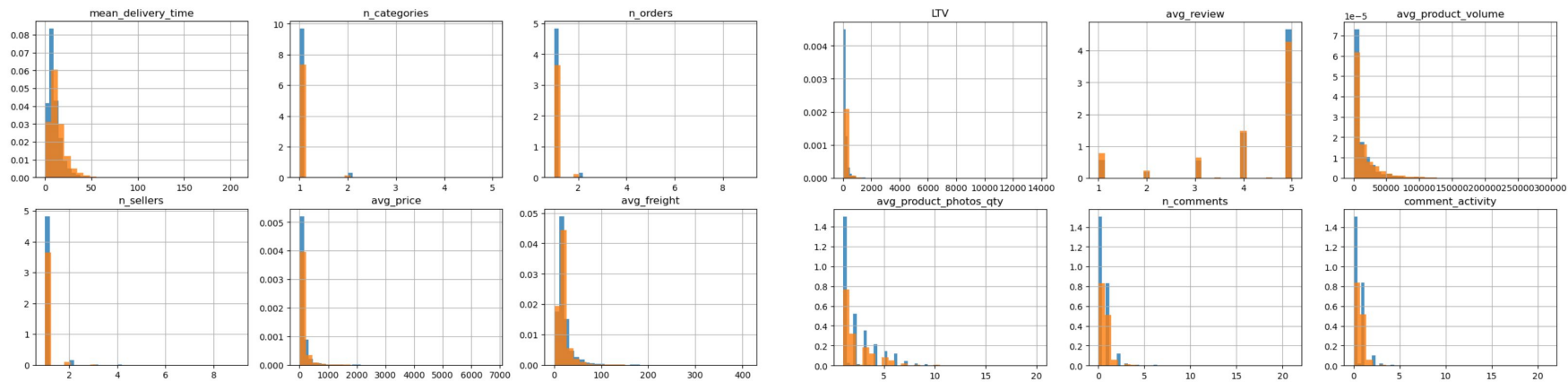
# Формирование гипотез

## Зависимость показателей активности от категорий товаров



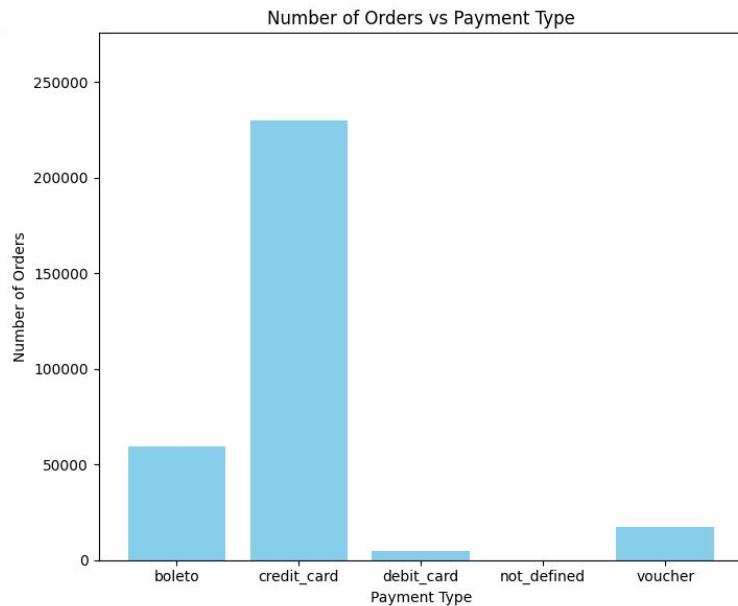
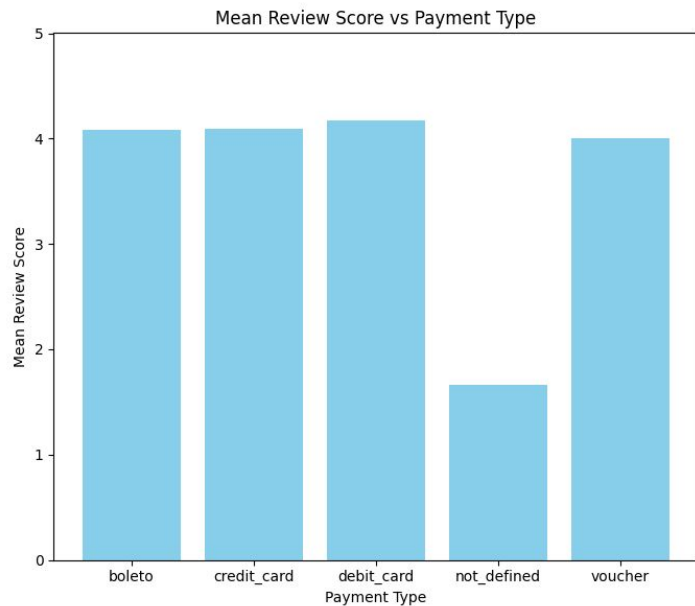
# Формирование гипотез

Сравнение распределений по выборкам активных и ушедших пользователей



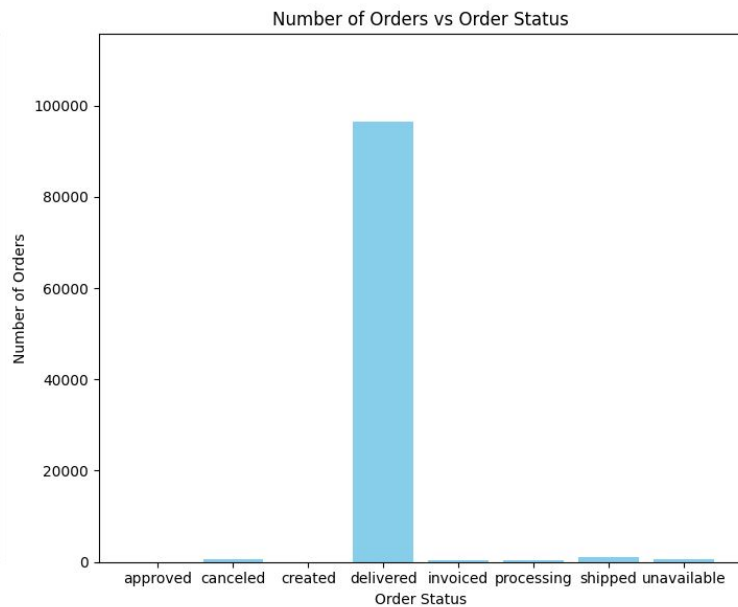
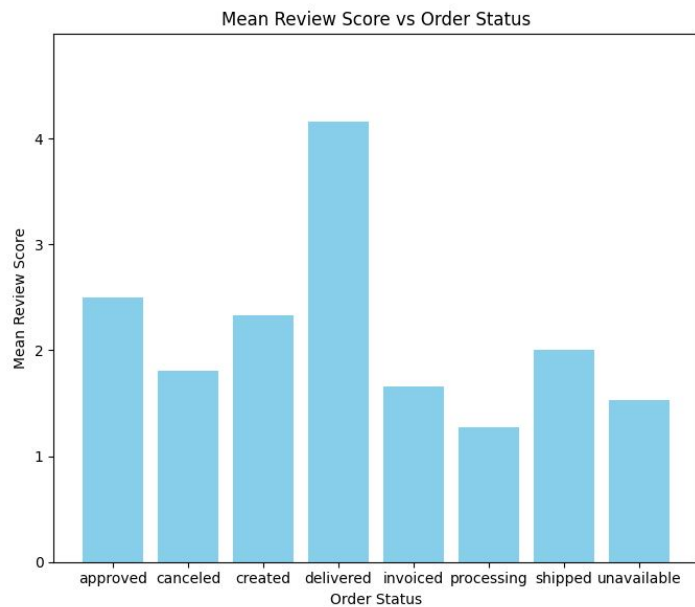
# Формирование гипотез

## Статистика по методам оплаты



# Формирование гипотез

## Статистика по статусам заказов

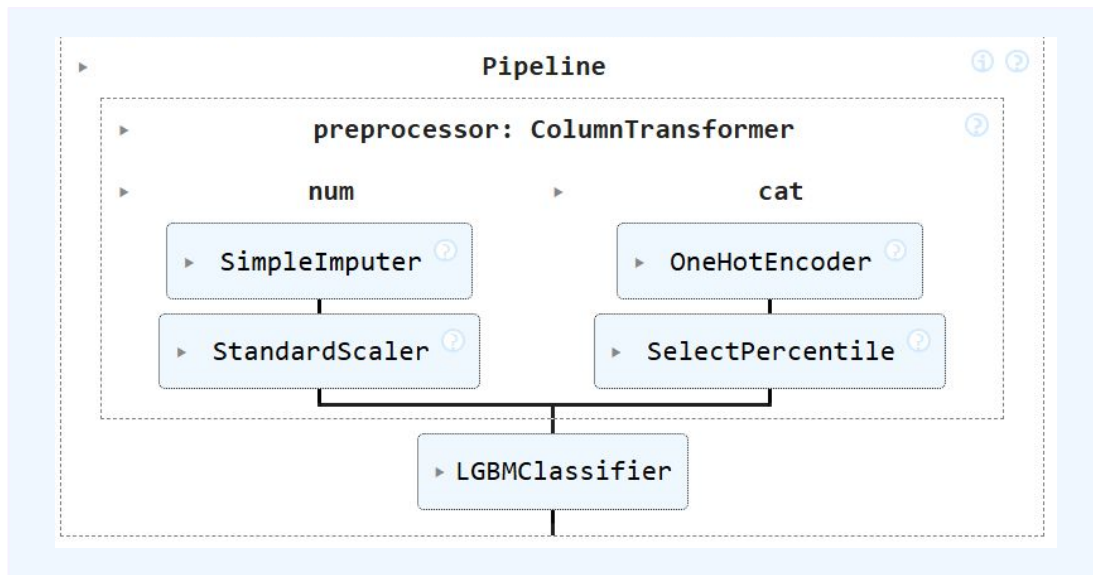


# **Разработка модели**



# Разработка модели

- **Итоговая модель:** LightGBM
- **Рассмотренные варианты:** RandomForest, LinReg, SVM



# Разработка модели

## • Признаки

- customer\_unique\_id
- price
- freight\_value
- order\_purchase\_timestamp
- order\_delivered\_customer\_date
- product\_category\_name
- review\_score
- order\_id
- review\_comment\_message
- customer\_city

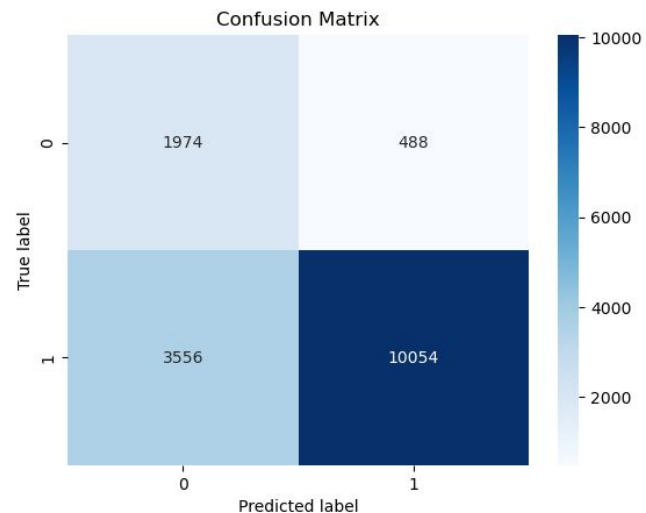
	mean_delivery_time	n_categories	n_orders	avg_price	avg_freight	avg_review	most_common_category	most_common_city
customer_unique_id								
0000366f-3b9a-7992-bf8c-76cfd3221e2	6.0	1	1	129.90	12.00	5.0	cama_mesa_banho	cajamar
0000b849-f77a-49e4-a4ce-2b2a4ca5be3f	3.0	1	1	18.90	8.29	4.0	beleza_saude	osasco
0000f46a-3911-fa3c-0805-444483337064	25.0	1	1	69.00	17.22	3.0	papelaria	sao jose
0000f6cc-b074-5a6a-4b88-665a16c9f078	20.0	1	1	25.99	17.63	4.0	telefonica	belem
0004aac8-4e0d-f4da-2b14-7fca70cf8255	13.0	1	1	180.00	16.89	5.0	telefonica	sorocaba

# Оценка качества модели

Classification report:

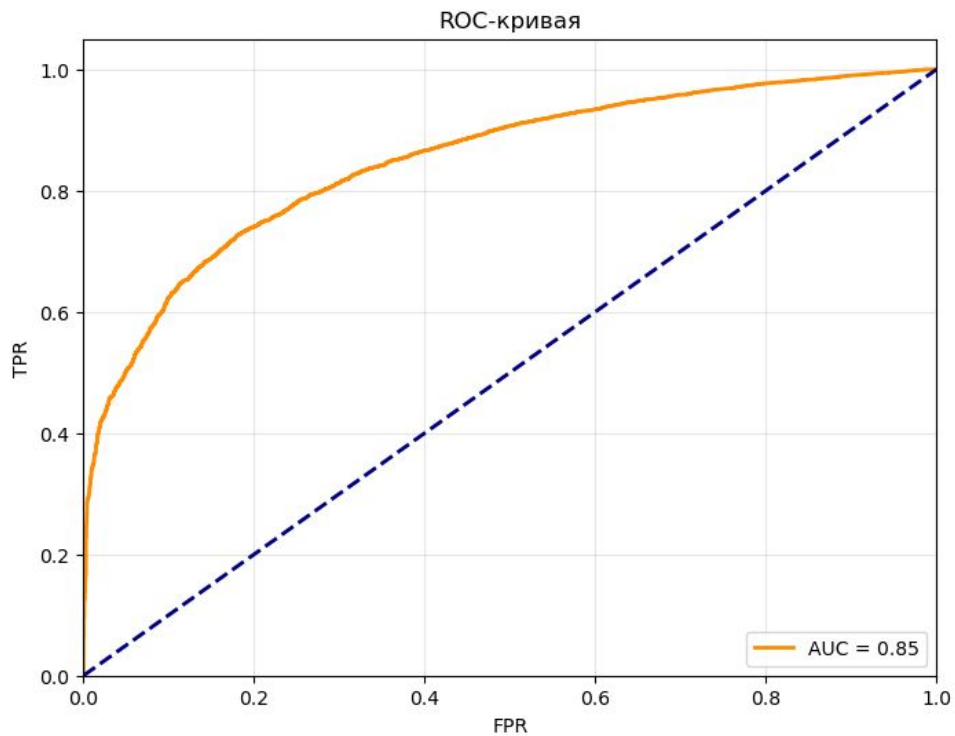
	precision	recall	f1-score	support
0	0.36	0.80	0.49	2462
1	0.95	0.74	0.83	13610
accuracy			0.75	16072
macro avg	0.66	0.77	0.66	16072
weighted avg	0.86	0.75	0.78	16072

AUC-ROC: 0.8489660473286534



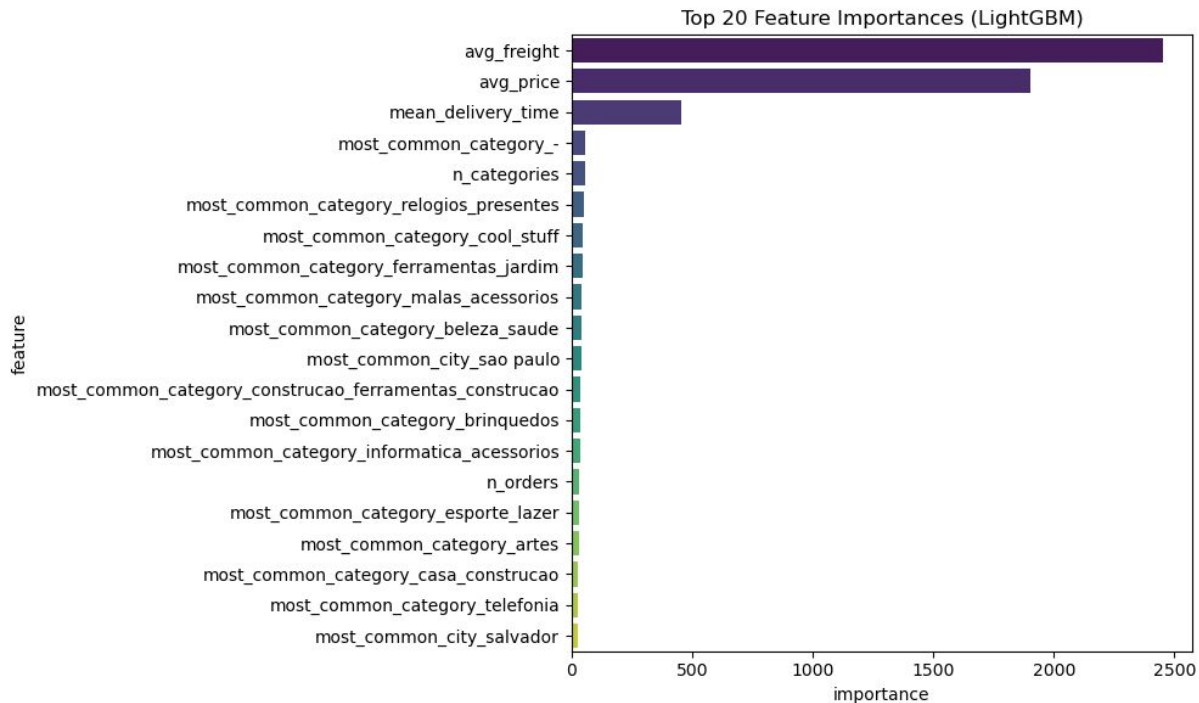


# Оценка качества модели



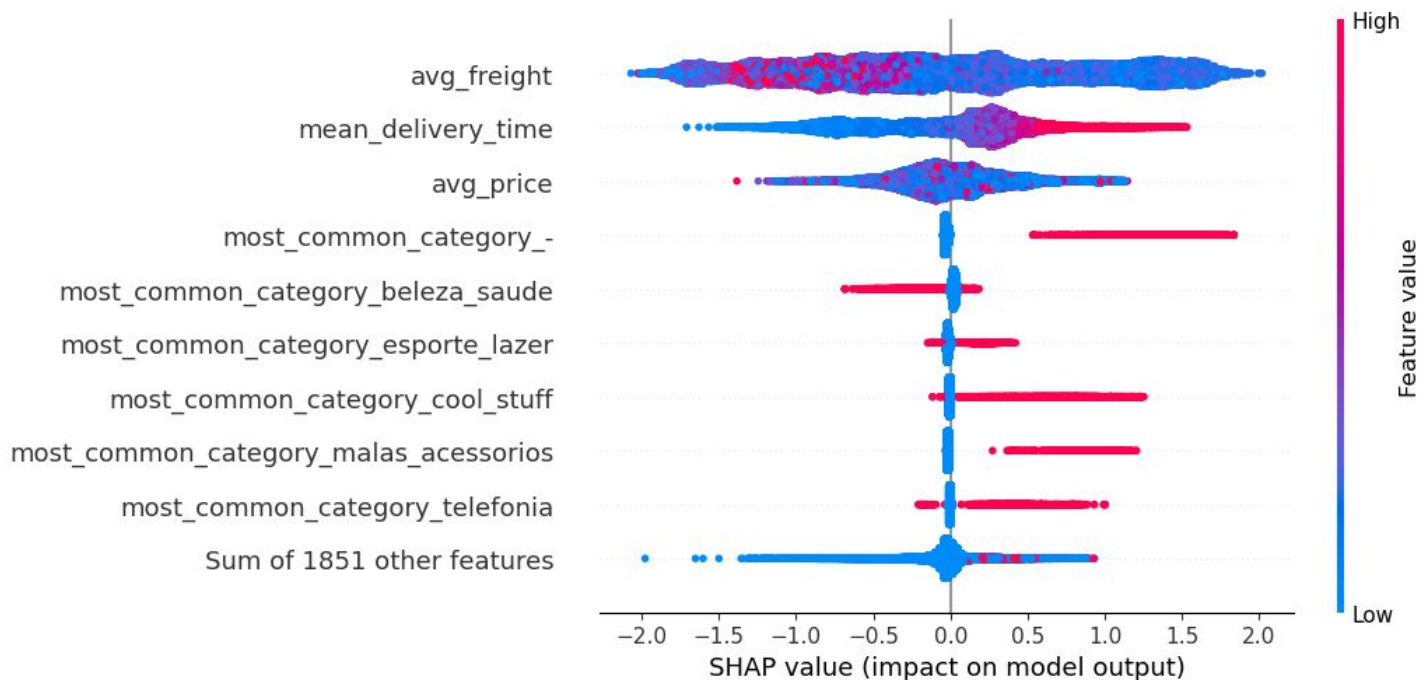
# Интерпретация результатов обучения модели

Оценка значимости признаков средствами LightGBM



# Интерпретация результатов обучения модели

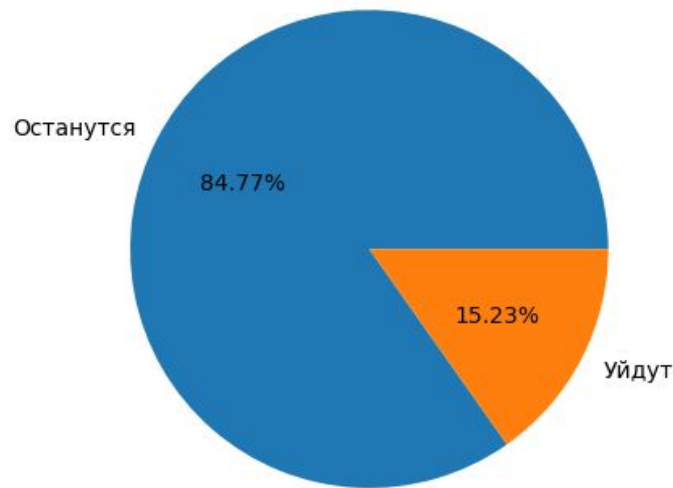
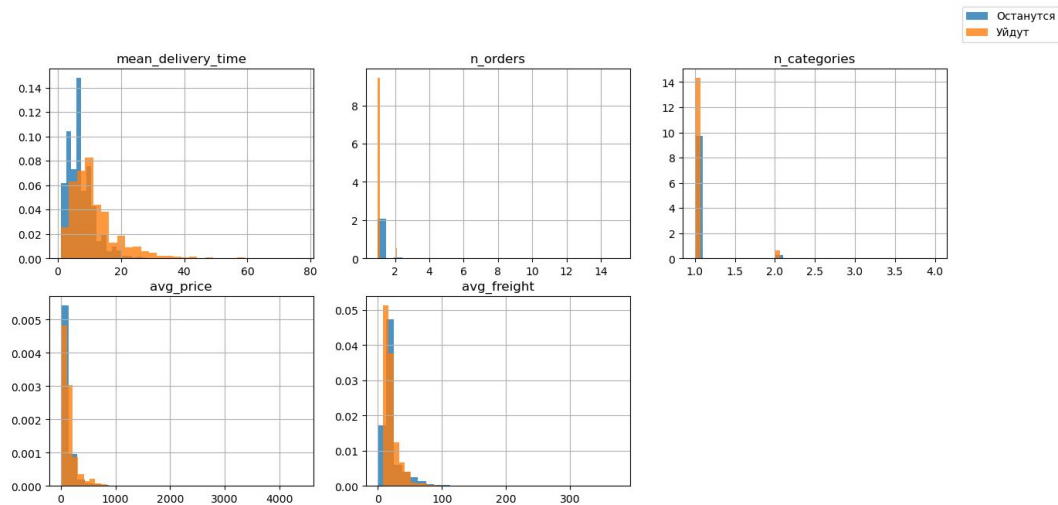
Оценка значимости признаков средствами SHAP



# **Предложения по снижению оттока клиентов**



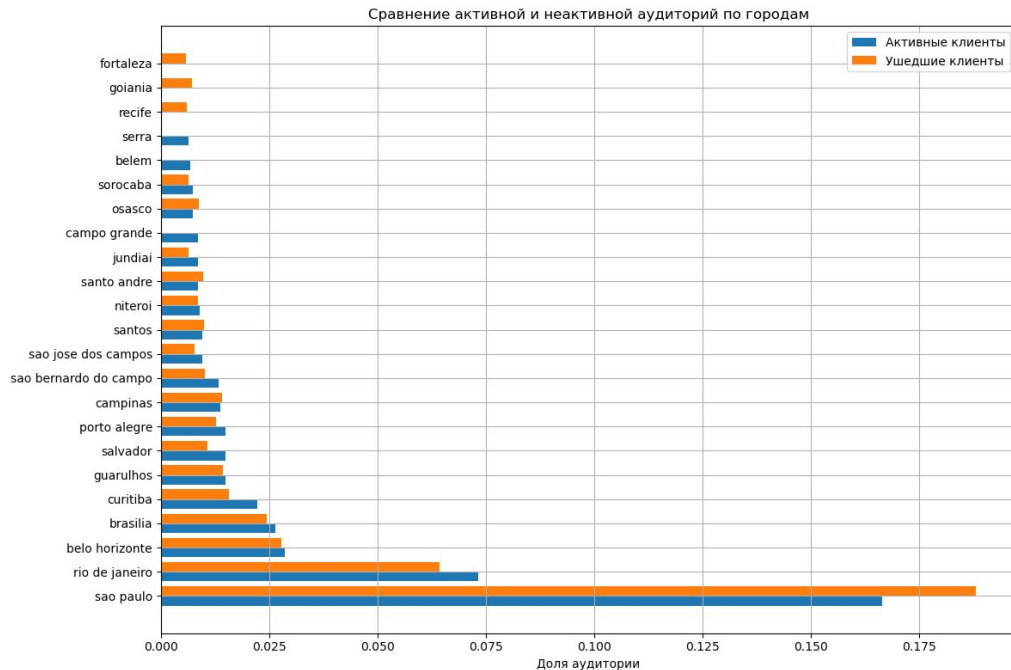
# Результаты предсказаний модели



- Основное отличие в распределениях наблюдается у среднего времени доставки

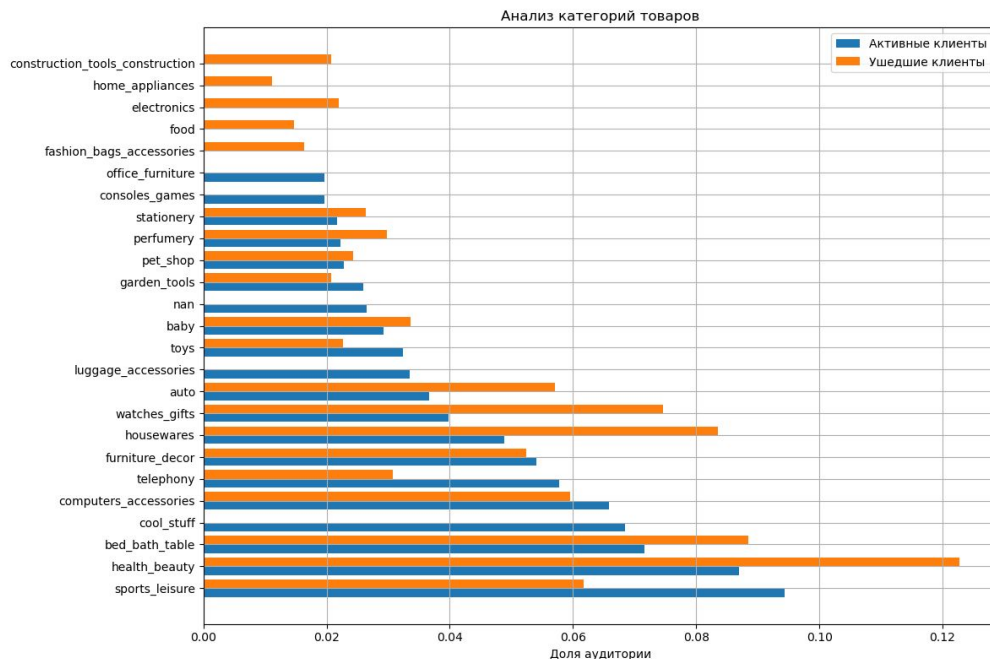
# Результаты предсказаний модели

- Анализ геоданных



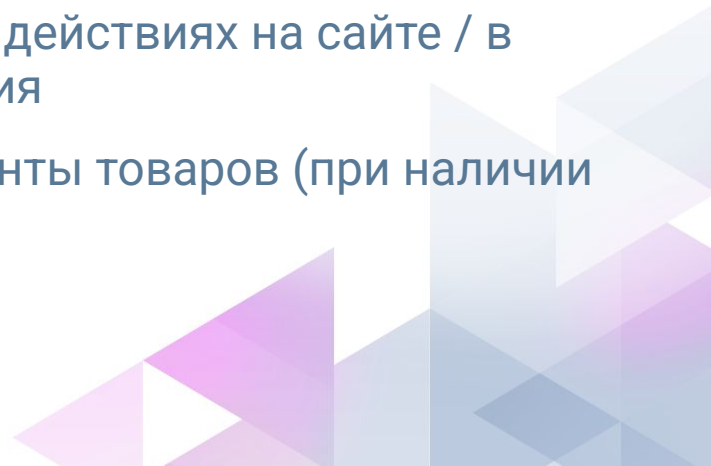
# Результаты предсказаний модели

- Анализ категорий товаров



# Предложения по снижению оттока клиентов

- Оптимизация времени доставки
- Дальнейшее исследование геоданных (в совокупности с оптимизацией доставки, например, открытие логистических центров)
- Следует обратить внимание на категории товаров, с сильными различиями распределений пользователей
- Анализ действий пользователя (аналитика о действиях на сайте / в приложении), выявление паттернов поведения
- Оптимизация рекомендательной системы ленты товаров (при наличии таковой)

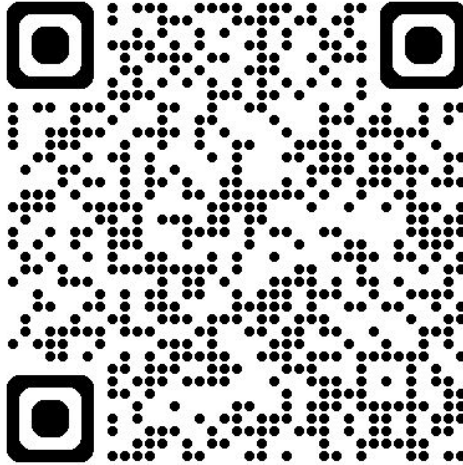




**One more thing ...**



## One more thing ...



Login: admin

Password: kingkonghahaha

# Выводы и результаты

Итоги работы: как и кому решение может быть полезно?  
В ходе работы над проектом было проведено

- Аналитическое исследование
- Разработка веб-приложения с поддержкой дашбордов
- Разработка и оптимизация ML-модели
- Формирование практических рекомендаций на основе данных, полученных на предыдущих этапах

