

# Predict Future Sales

## Группа 5



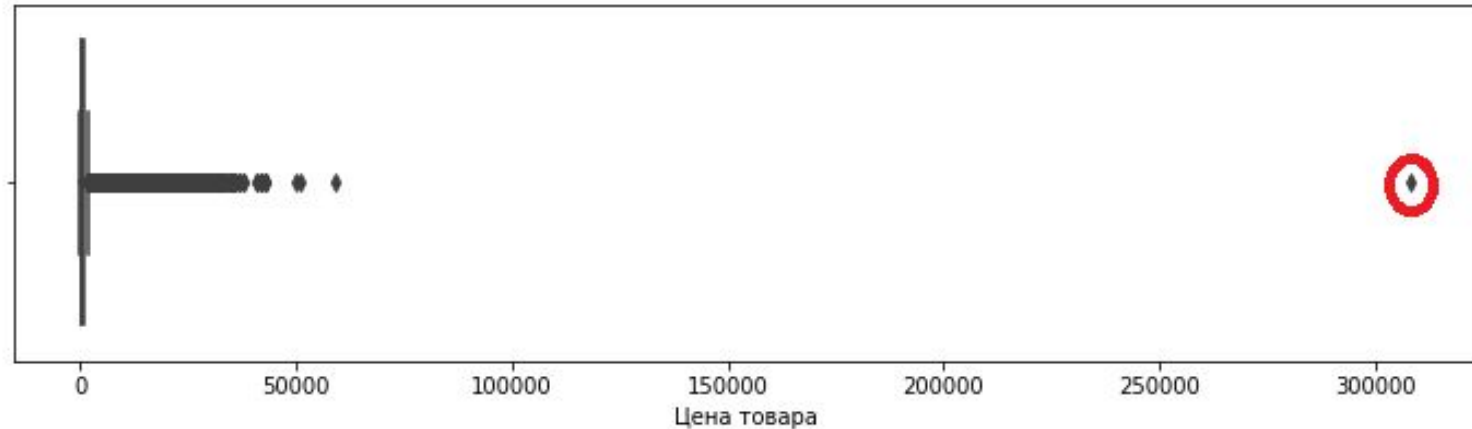
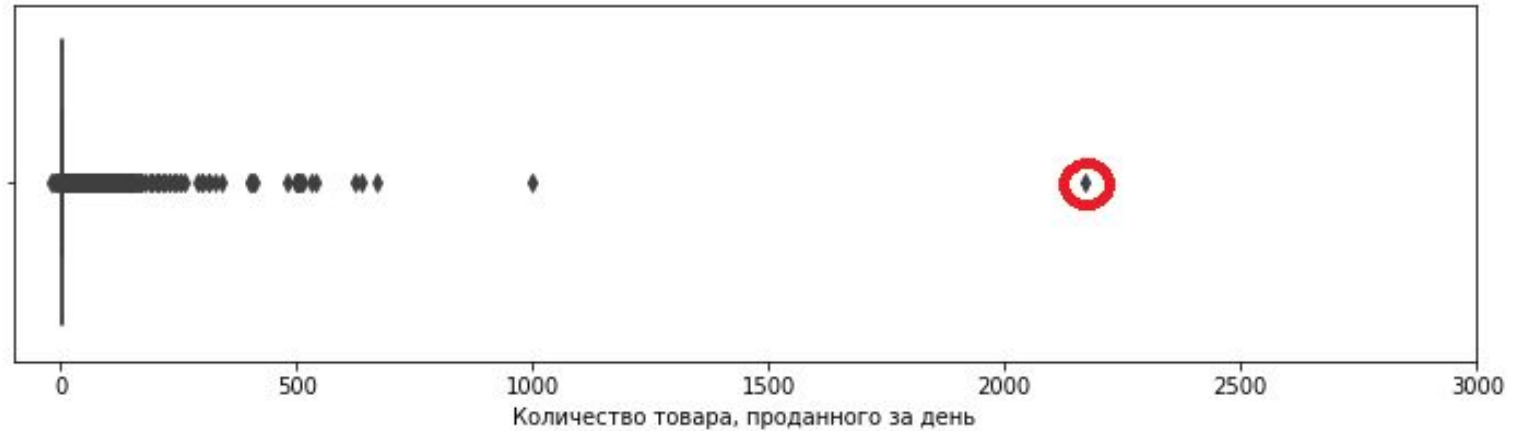
Киселев А. К.  
Морозов Л. Л.  
Захарченко А. В.

# Описание датасета

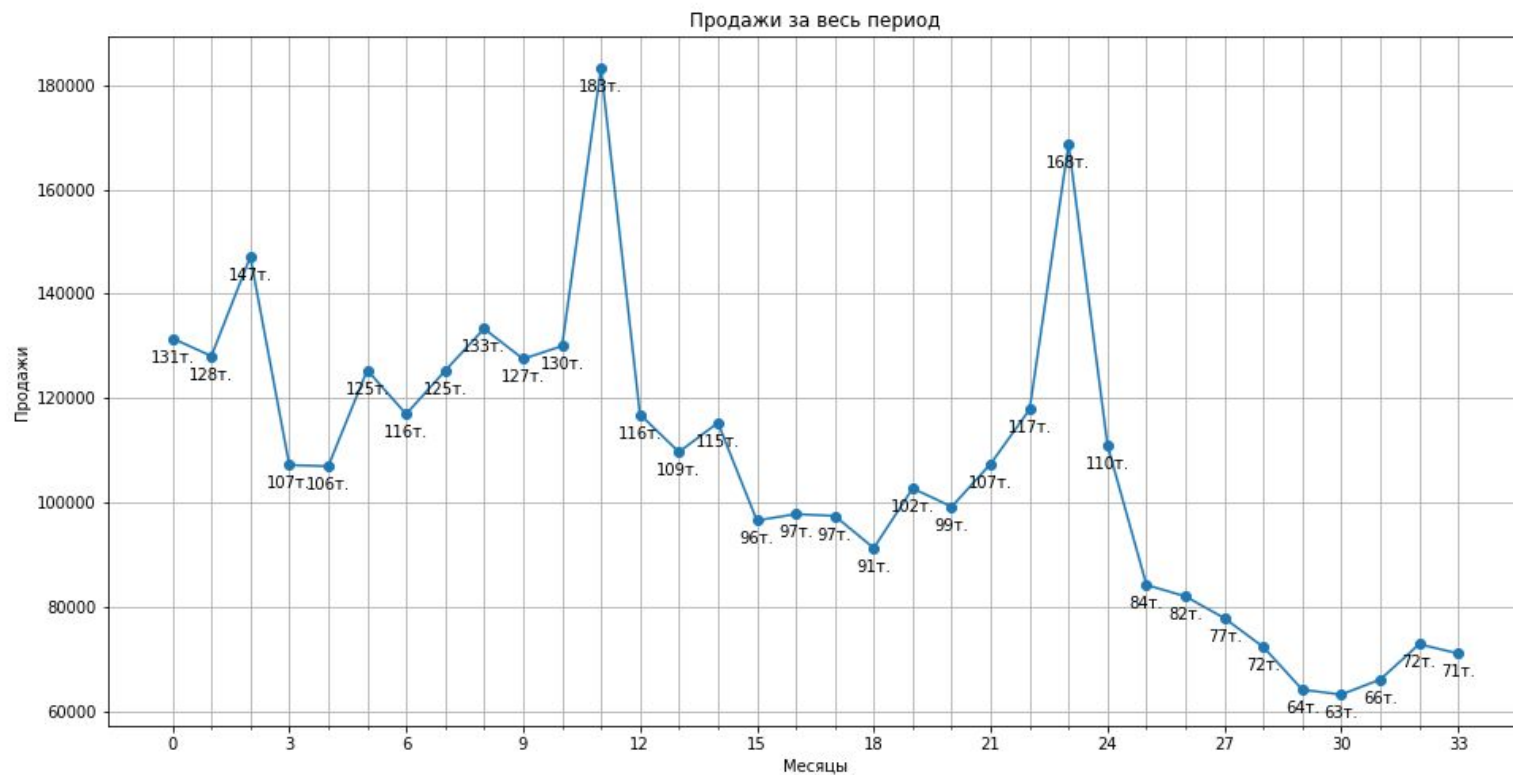
Состав данных: дата (январь 2013 - октябрь 2015), цена товара, количество проданного товара, торговый центр, категория товара.

Предоставляются ежедневные исторические данные о продажах. Задача состоит в том, чтобы спрогнозировать общее количество товаров, проданных в каждом магазине для тестового набора за ноябрь 2015 года.

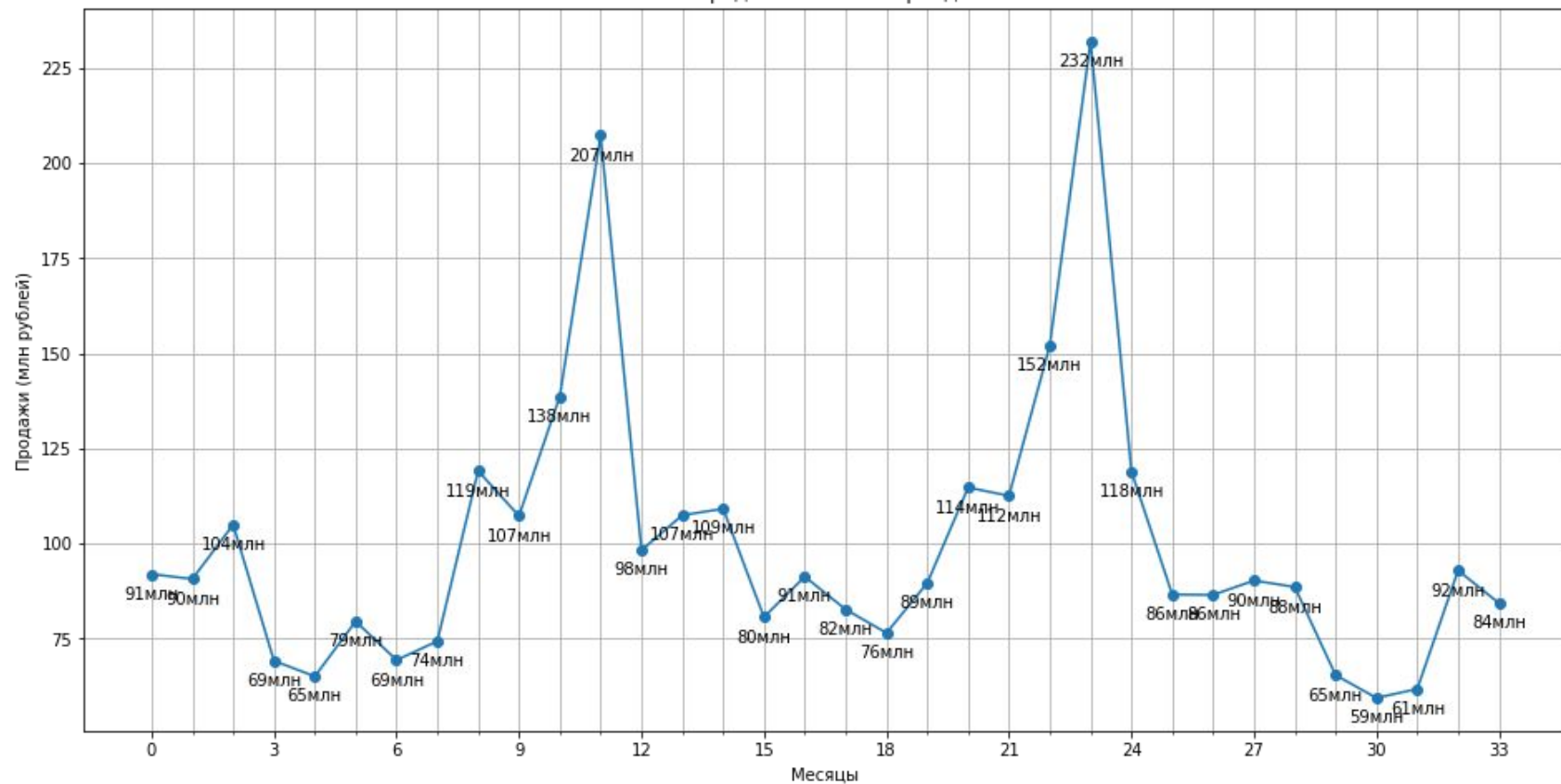
# Удаление выбросов

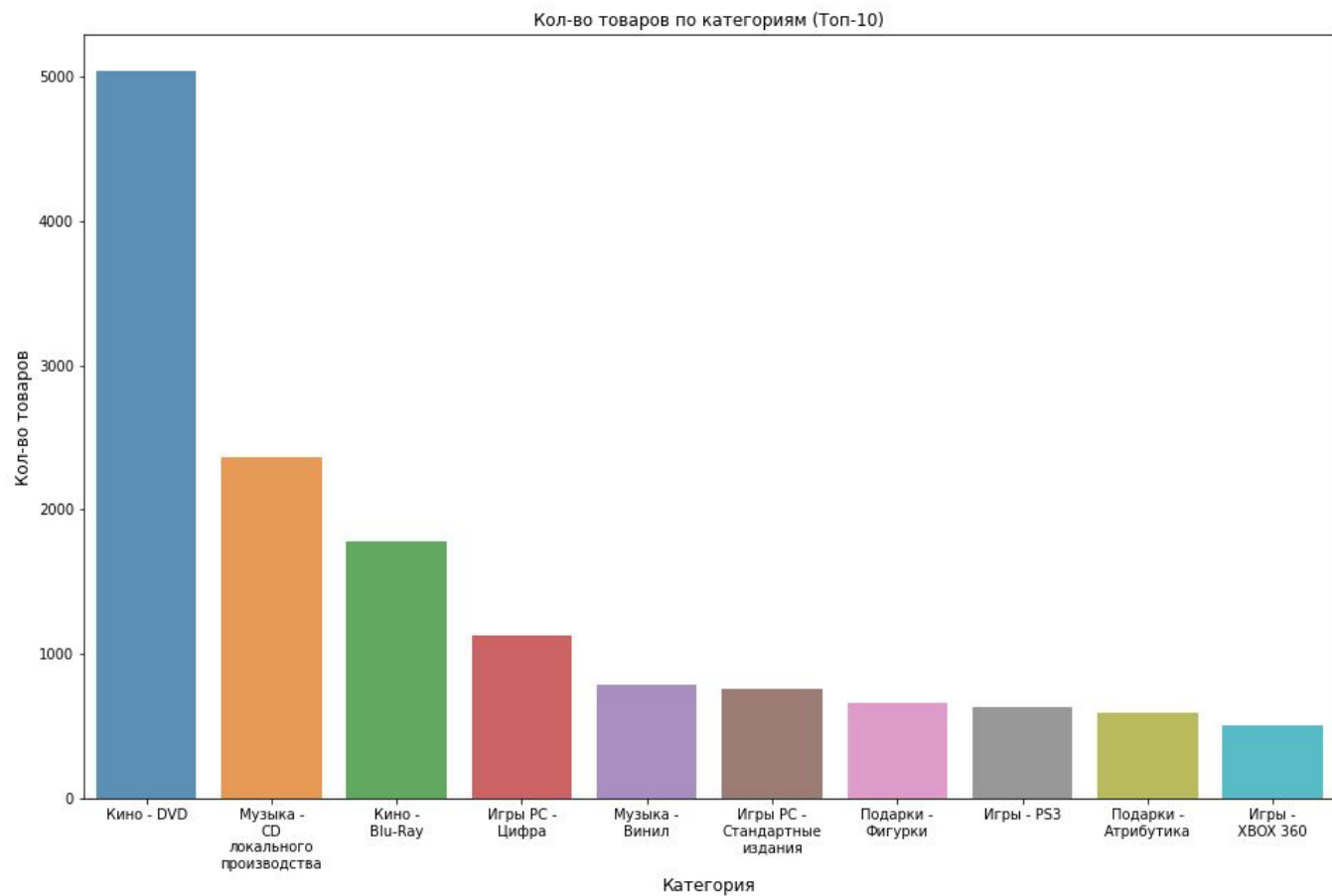


# EDA

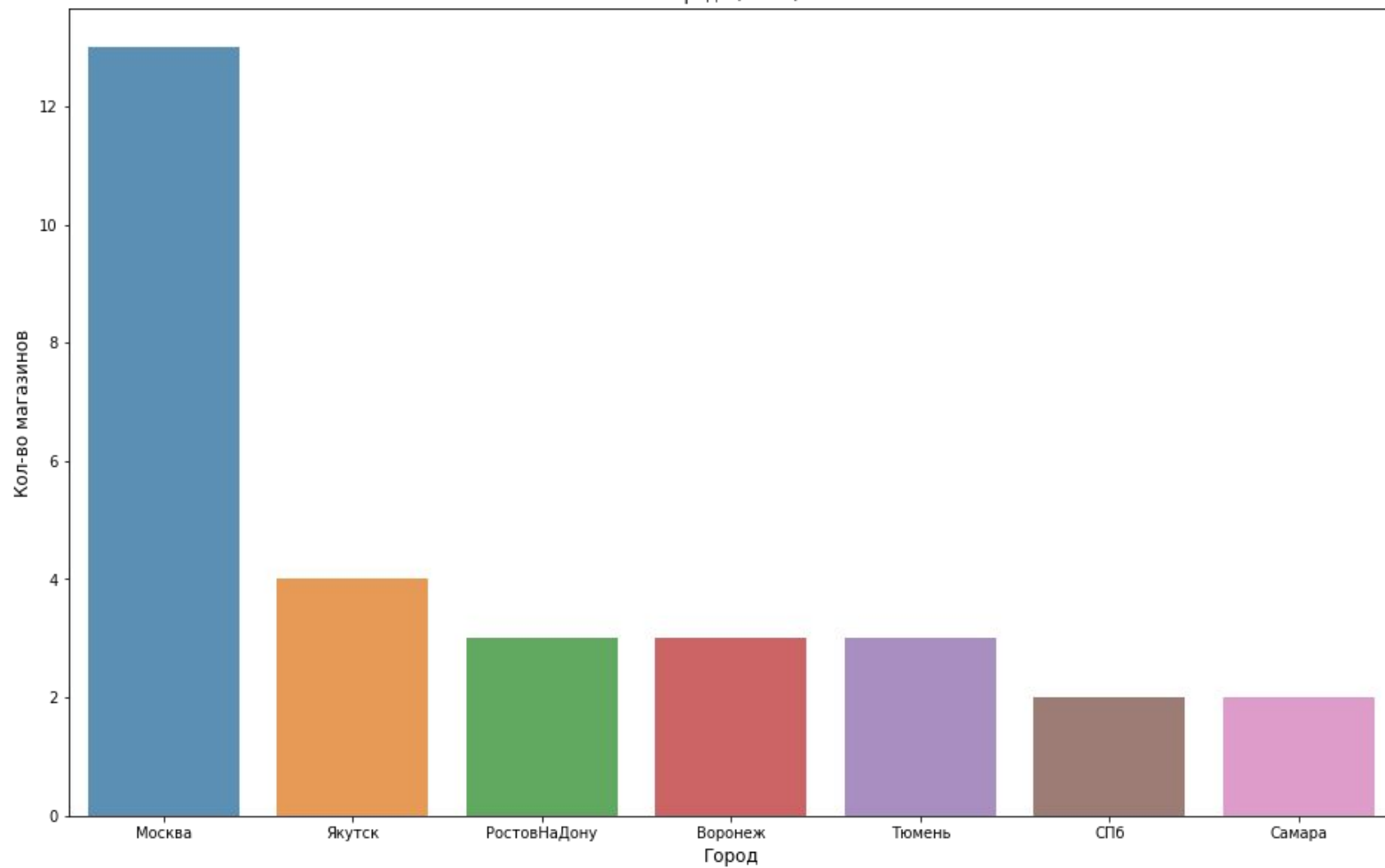


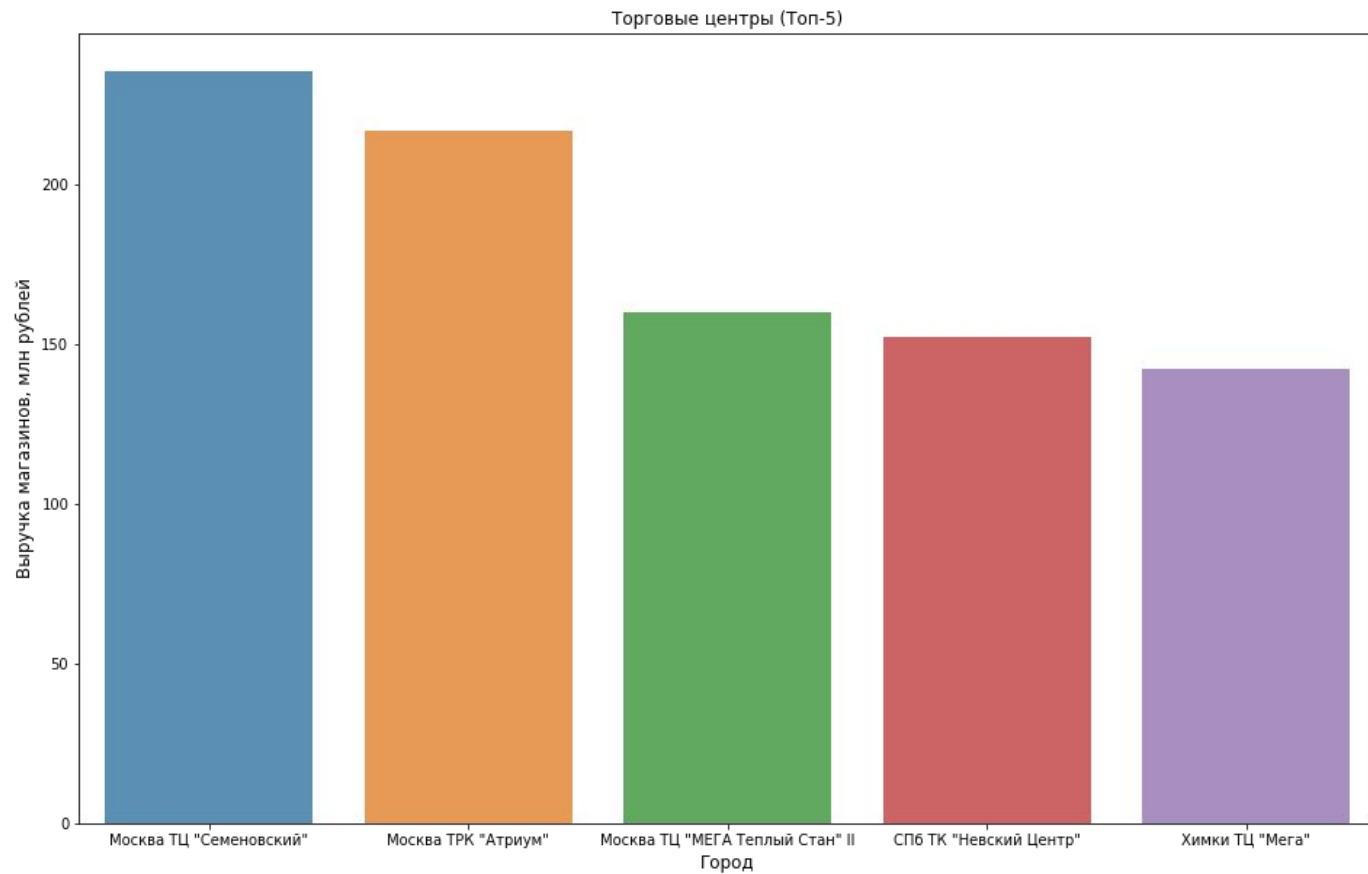
Продажи за весь период





Города (Топ-7)







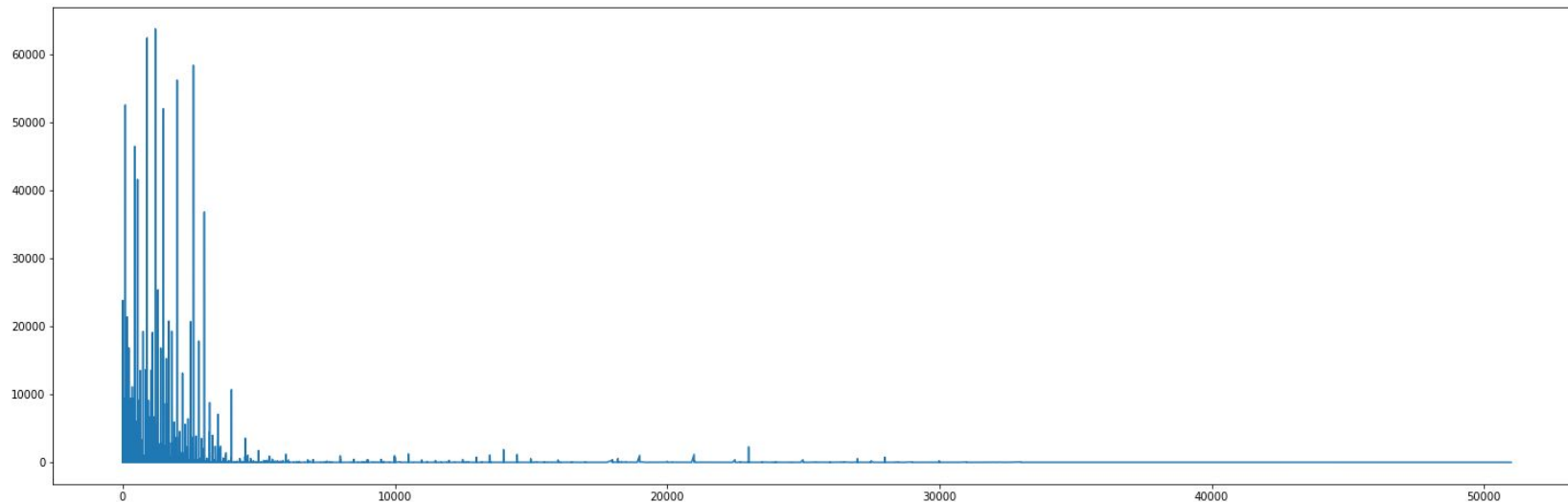
# Feature Engineering

- Доработана дата продажи, разбита на месяц и год.
- Выделен из названия торгового центра город.
- Разбита категория товара на основную категорию и подкатеорию
- Для оценки цены товара и выручки по товару применен метод средних
- Добавлены новые фичи количество продаж в месяце по всем товарам, количество продаж по категориям, количество продаж по подкатегориям, количество продаж в ТЦ, количество продаж в городе.
- Добавлены целевые фичи с задержкой.



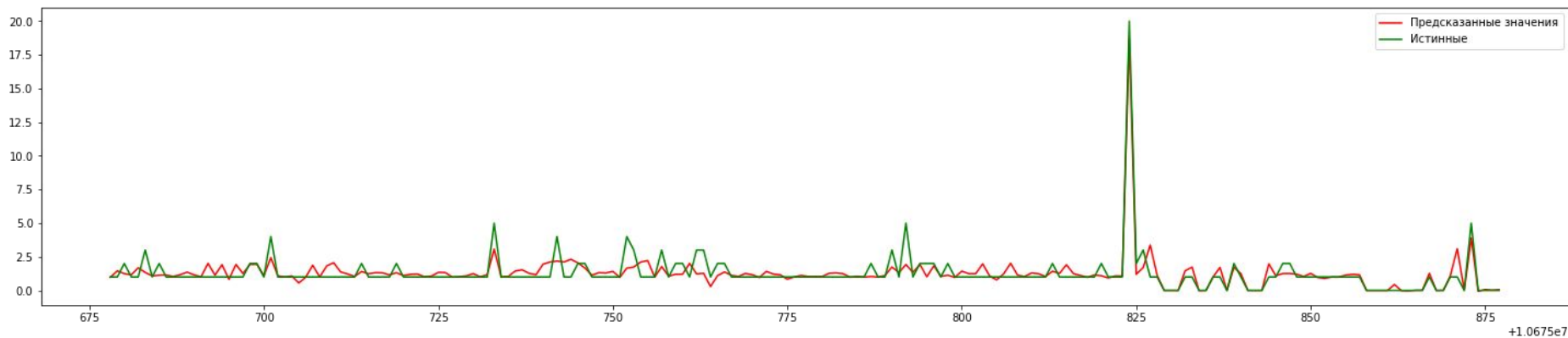
# График зависимости количества товара от цены

Количество проданного товара



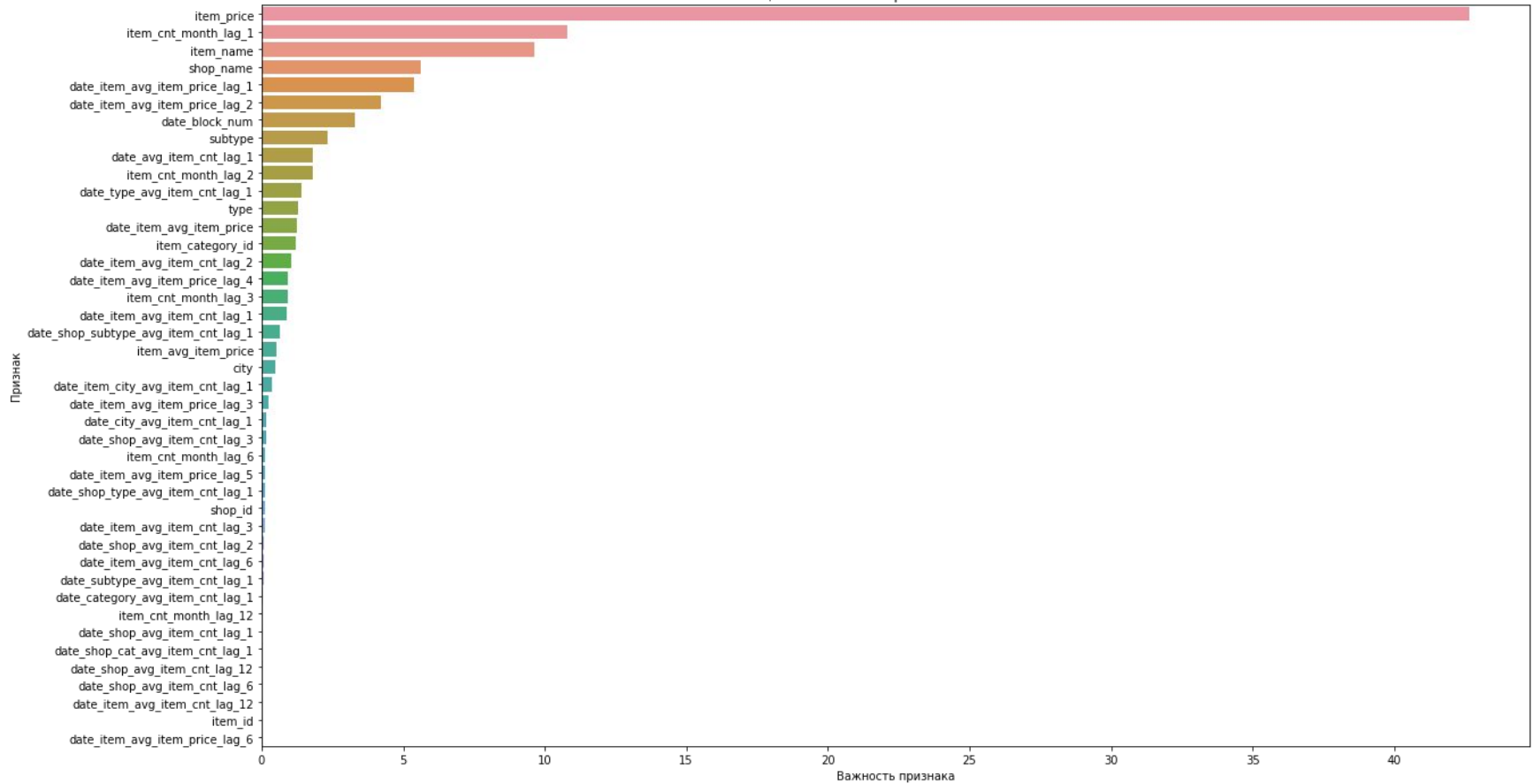
Цена товара

## Сравнение предсказанных значений с истинным



MSE на валидационной выборке: 0.53834  
RMSE на валидационной выборке: 0.73372

Оценка важности признаков CATBOOST



## Место на kaggle

9175	Yamanaka		1.22249	21	1y
------	----------	---	---------	----	----

9176	<b>Project2_Group5</b>	  	1.22249	4	2m
------	------------------------	---	---------	---	----



Your Best Entry!

Your most recent submission scored 1.22249, which is an improvement of your previous score of 1.54437. Great job!

[Tweet this](#)

9177	Jiaxiang Ren		1.22254	5	1y
------	--------------	---	---------	---	----

**презентация закончена**



**Надеюсь вам понравилось!)**

**Спасибо за внимание!)**