How Much Can You Earn On Olist?

Predicting Seller Performance on Our E-Commerce Platform

Presentation By Alexis Lim

Agenda

Company Background

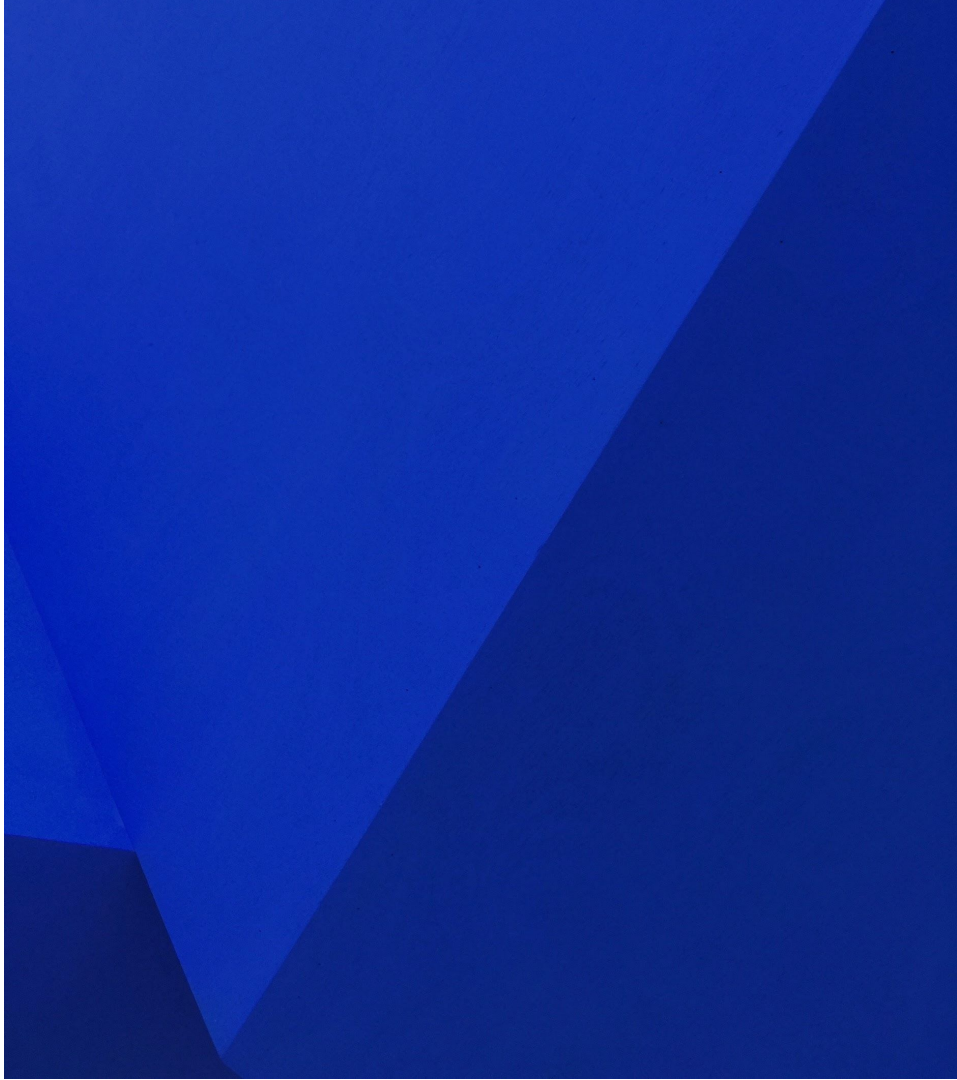
Problem Statement

Data Summary & Exploration

Modeling Approach

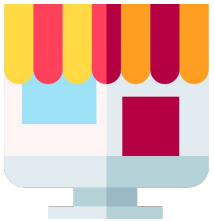
Model Evaluation

Conclusion and Insights



Olist is a marketplace of marketplaces

Our Business Model



We connect with small businesses from all over Brazil.

And help them to list their products on major marketplaces under the Olist Store.

Customers choose us for our wide variety of products from different sellers.

Sellers choose us for ease of use and wider access to more customers.

Our problem statement

I'm interested, but **how much can I earn**?

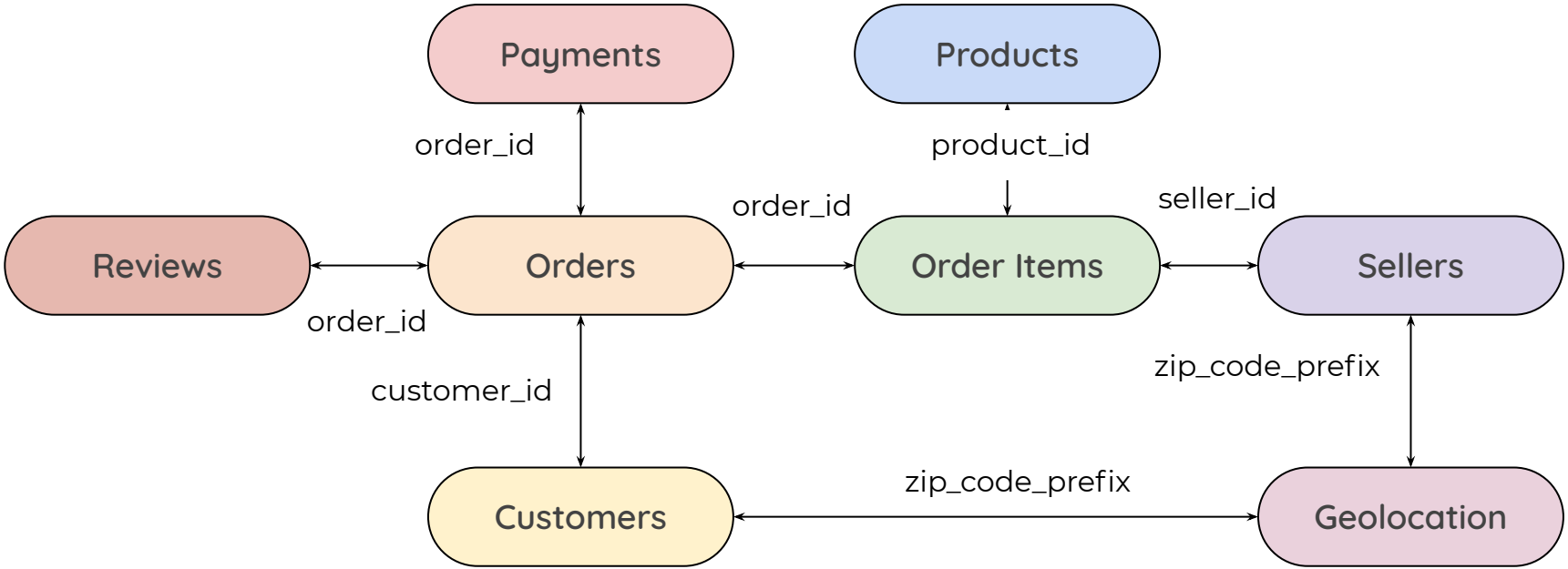
As part of the data science team in Olist, we have been given a sample of sales data (~100K sales orders) to perform analysis.

Guiding Questions:

1. How much can a seller expect to earn on the platform?
2. What steps can a seller take to increase their sales?

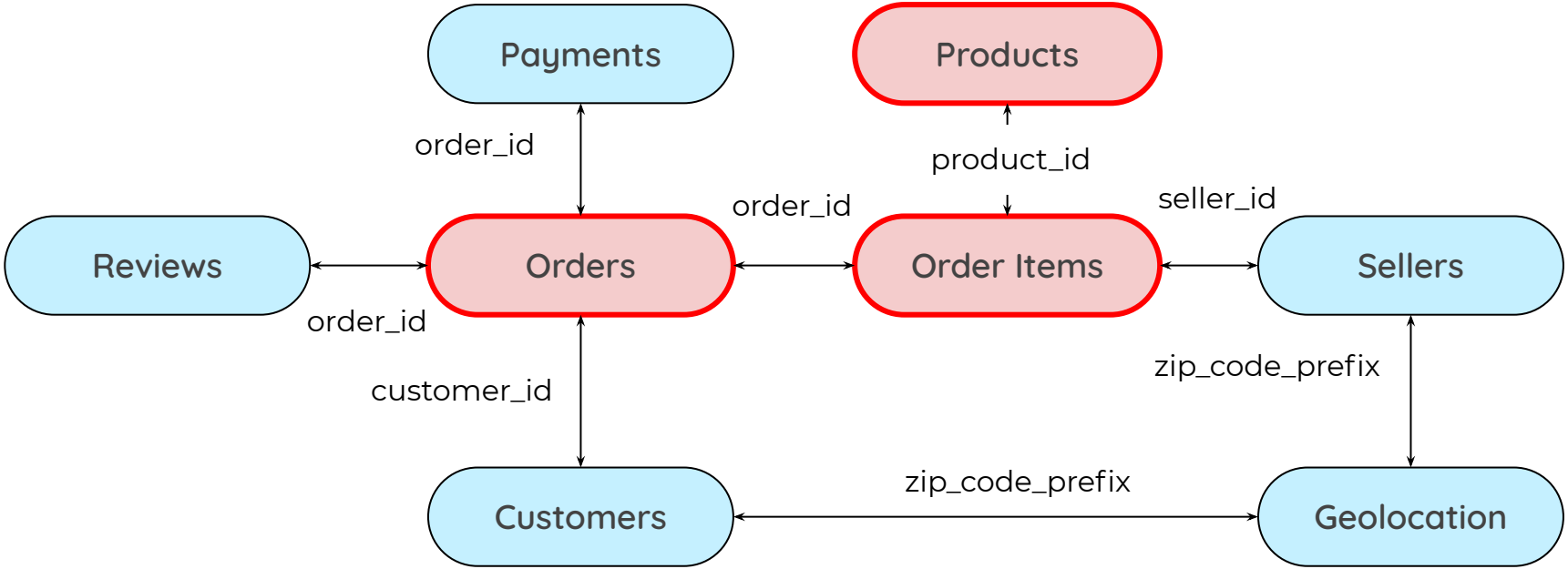
Data summary

We have a total of **8 datasets** given to us which gives us different columns of data across 100K orders, from 2016 to 2018.



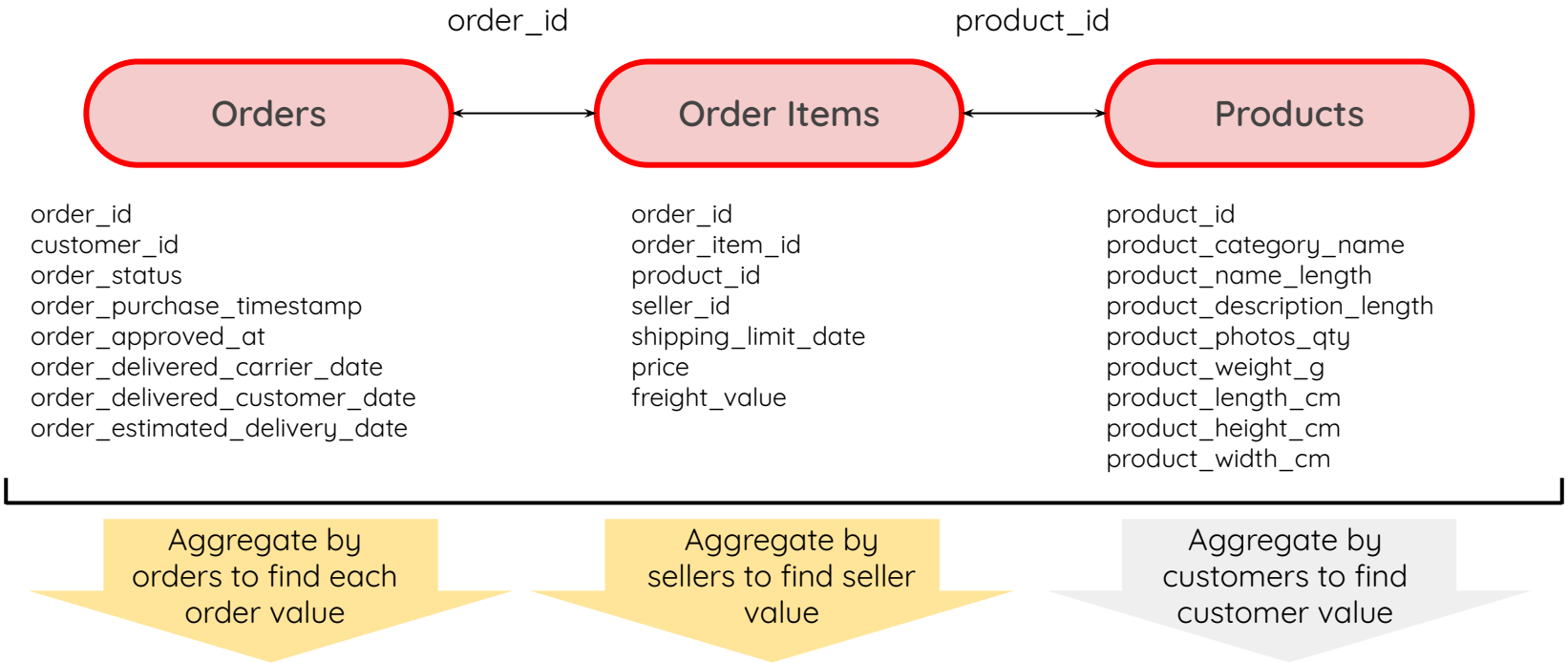
Data summary

We have a total of **8 datasets** given to us which gives us different columns of data across 100K orders, from 2016 to 2018.

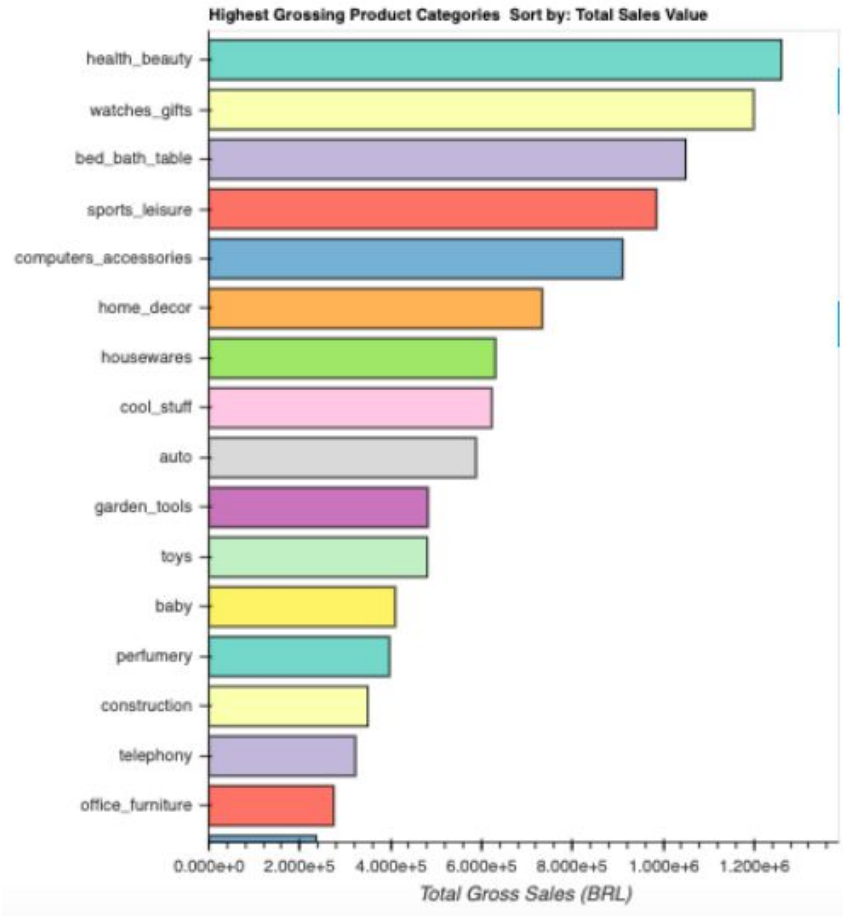


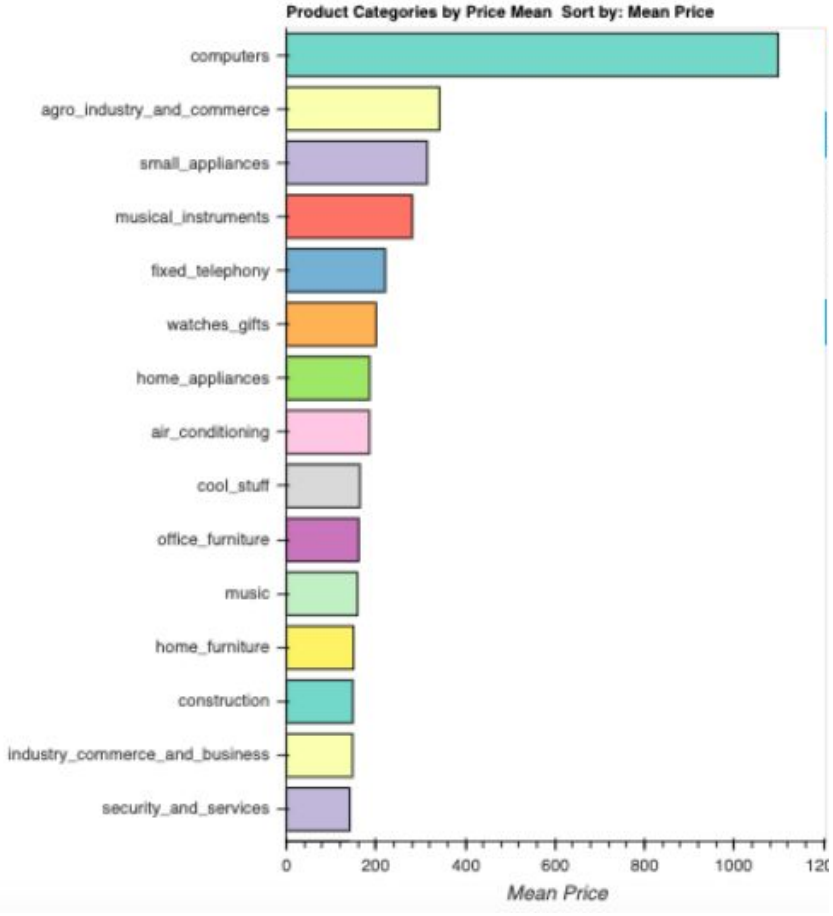
Creating a master dataset

To perform analysis on our data, we join the orders, order items and product datasets to create a master order items dataset.

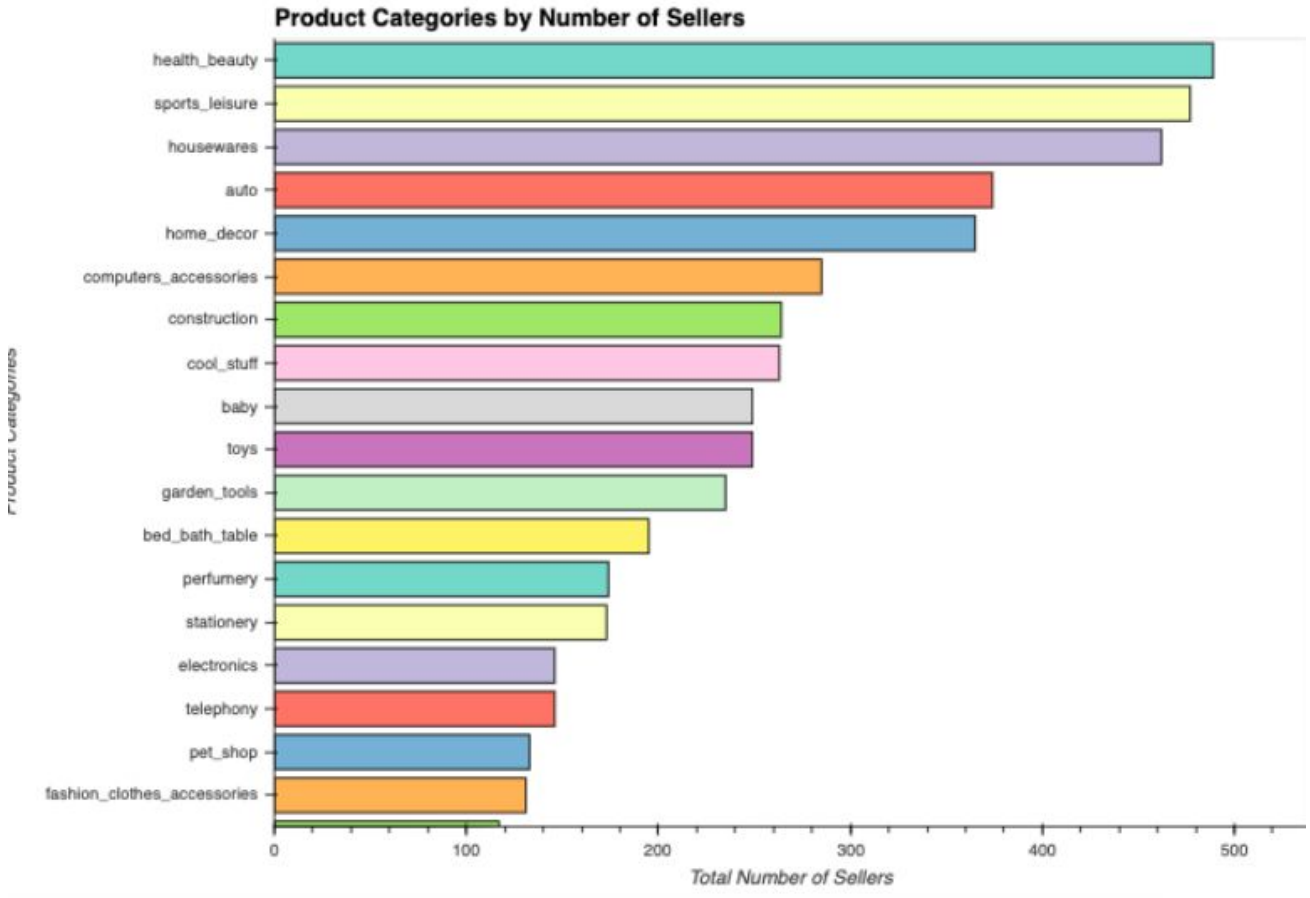


Top product categories

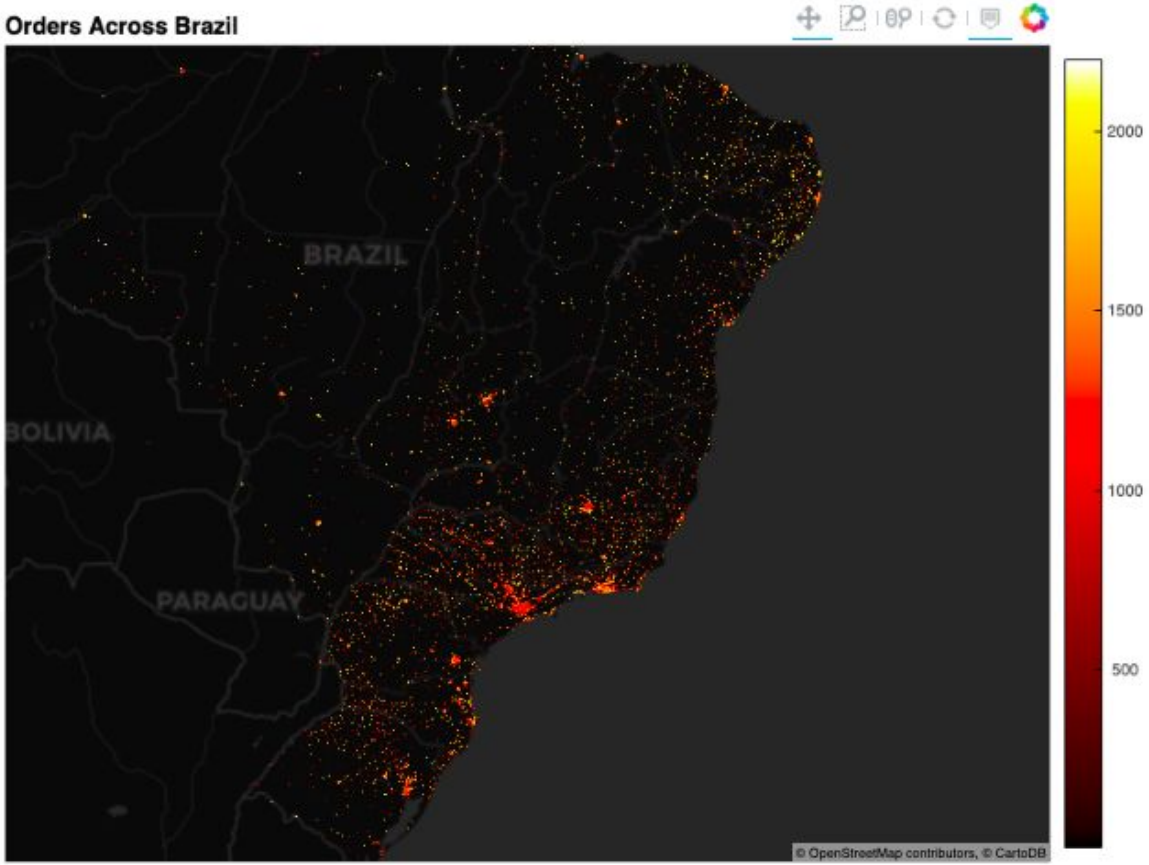




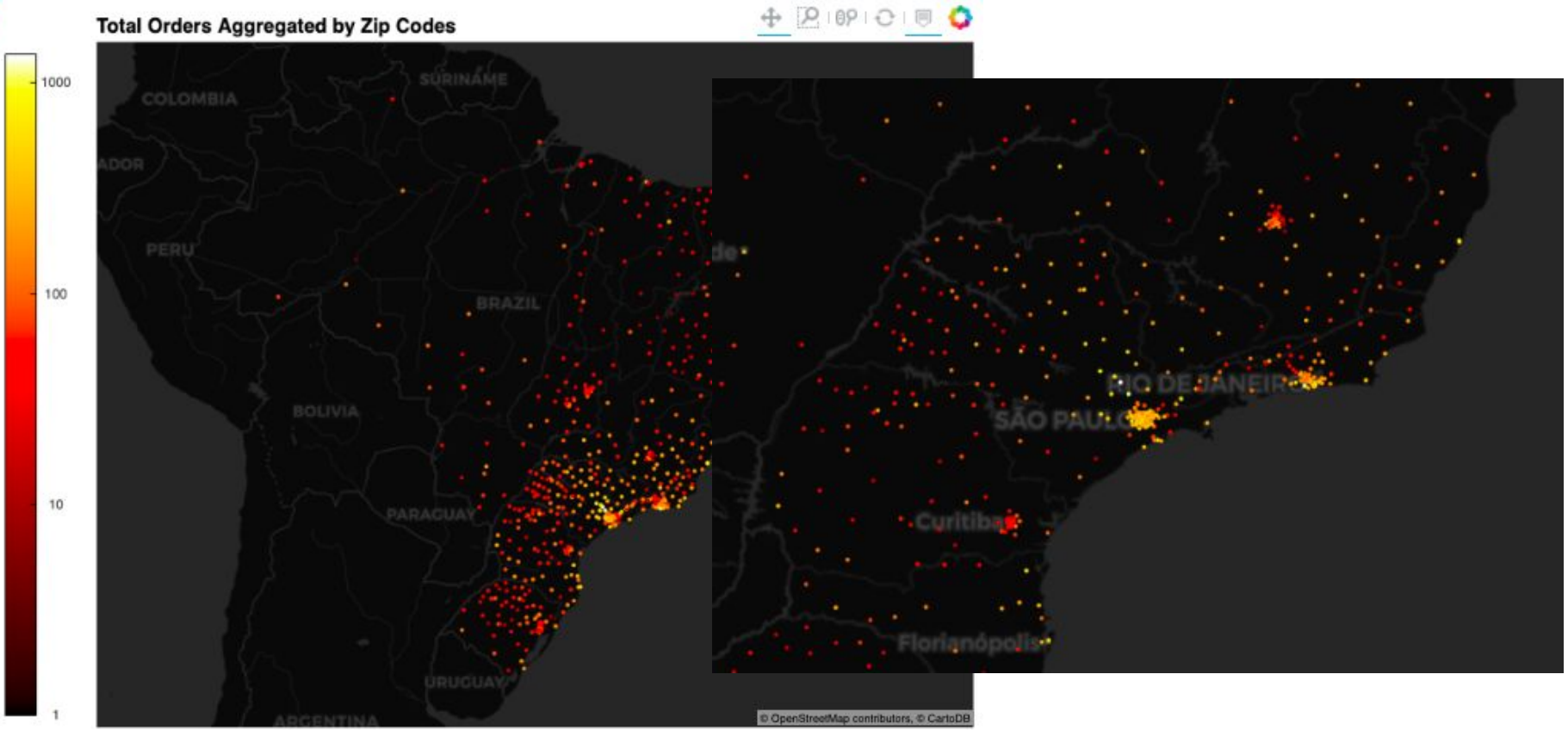
Top product categories



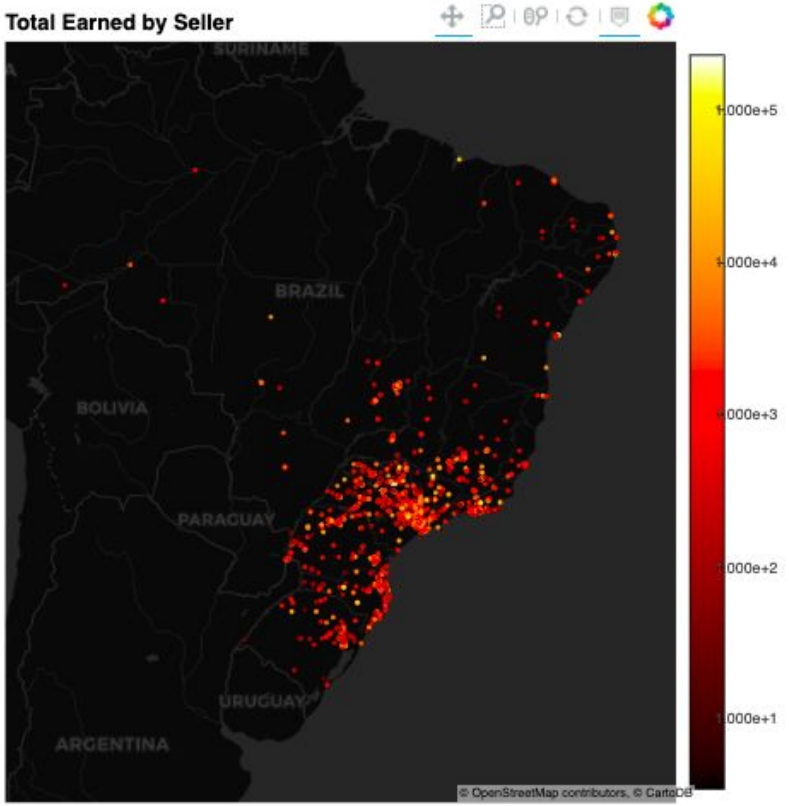
Orders are primarily from the southeast

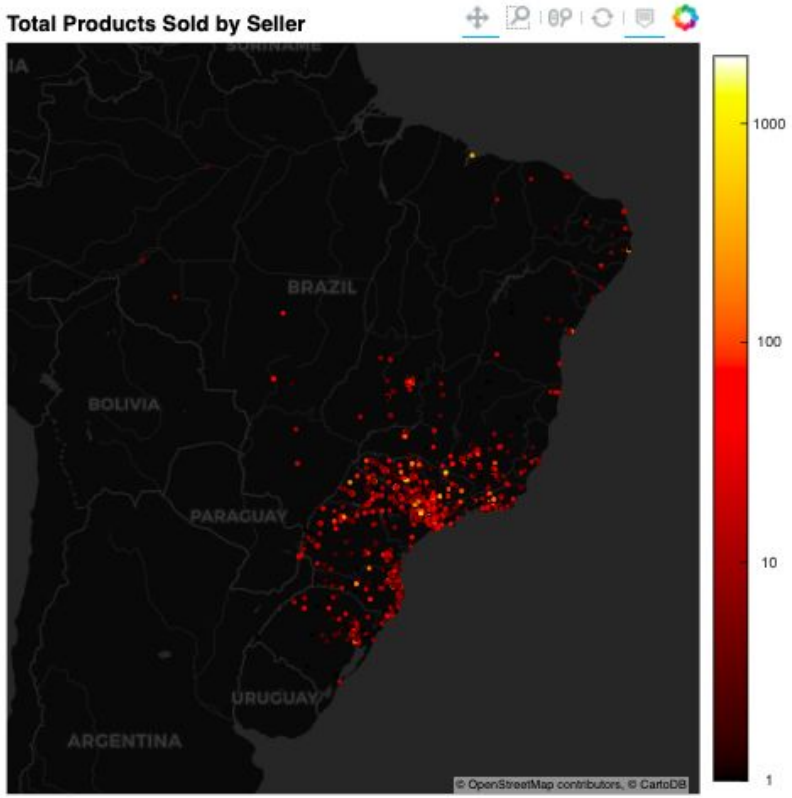


Orders are primarily in the southeast

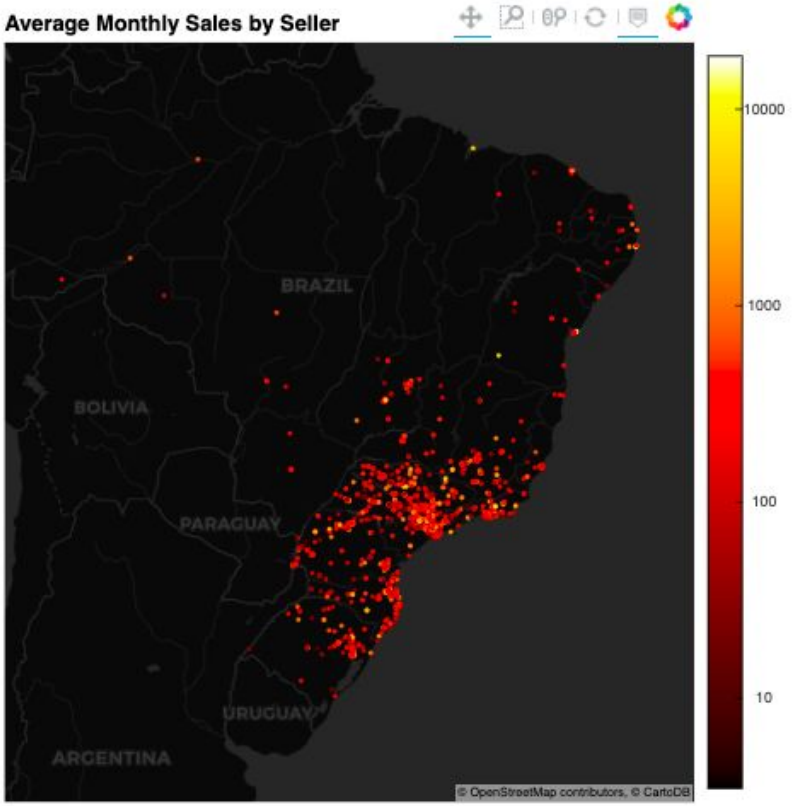


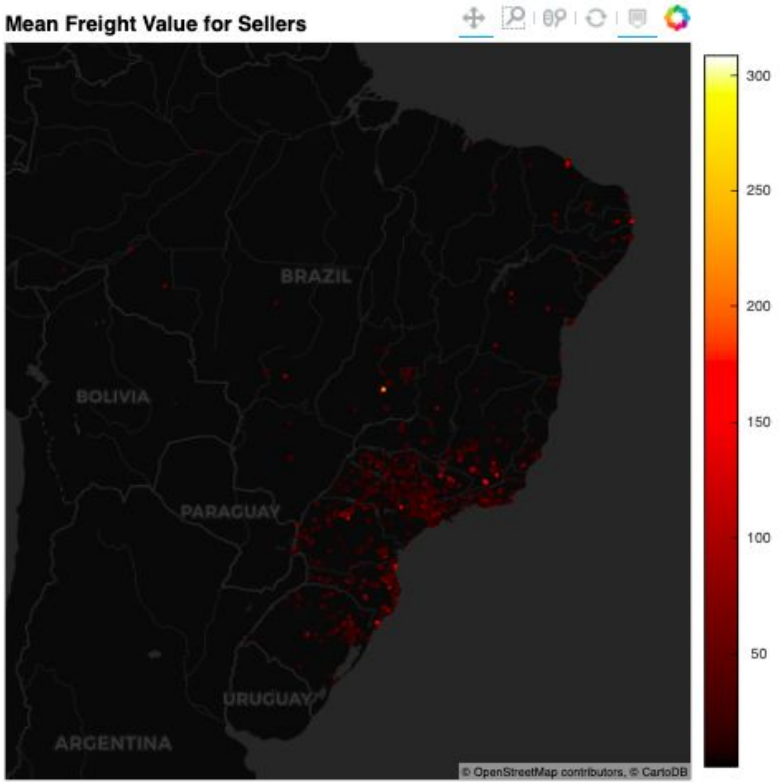
Sellers tend to sell more in the southeast





Sellers tend to sell more in the southeast





How to predict seller performance?

I'm interested, but how much can I earn?

I sell **books** online to earn part-time income, and the price range is from **R$40 to R$100, averaging R$60**. I also sell some **CDs and DVDs.** I can sell up to **50 products** a month from my inventory and have around **10 unique listings** at one time.

I am situated in **Sao Paulo**.

???

How to predict seller performance?

I'm interested, but how much can I earn?

I sell **books** online to earn part-time income, and the price range is from **R$40 to R$100**. I also sell some **CDs and DVDs.** I can sell up to **50 products** a month from my inventory and have around **10 unique listings** at one time.

I am situated in **Sao Paulo**.

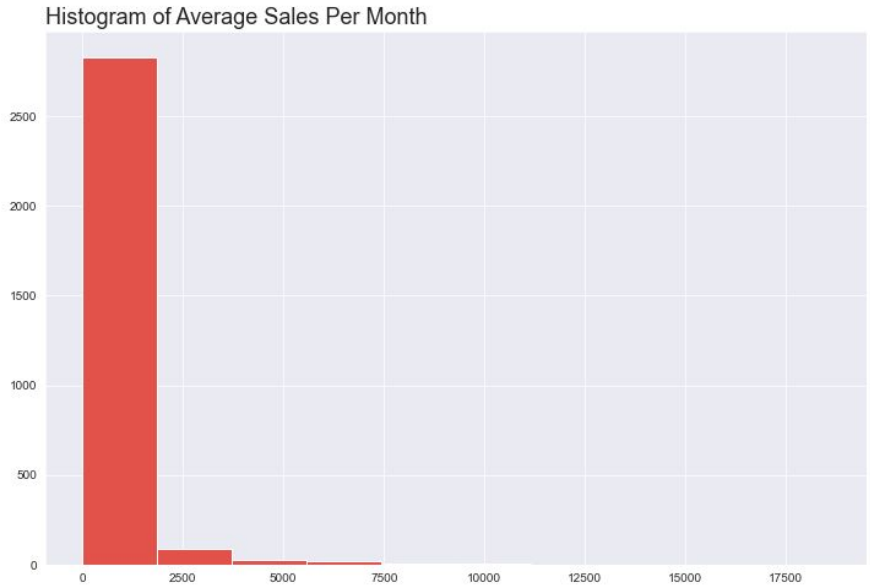
Formulating seller characteristics from our data

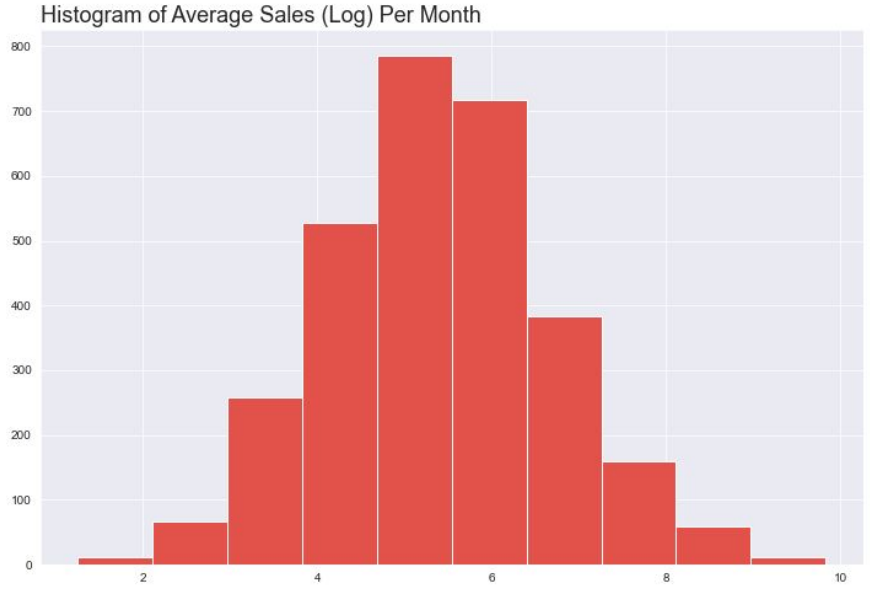
Product Category	Number of Unique Products	Product inventory	Price range
Get the most common category and number of categories	Count unique product ids	Average product sold / month	Maximum, minimum, average price

Sales Value — Calculating the active sale months in the data and average sales value across the period

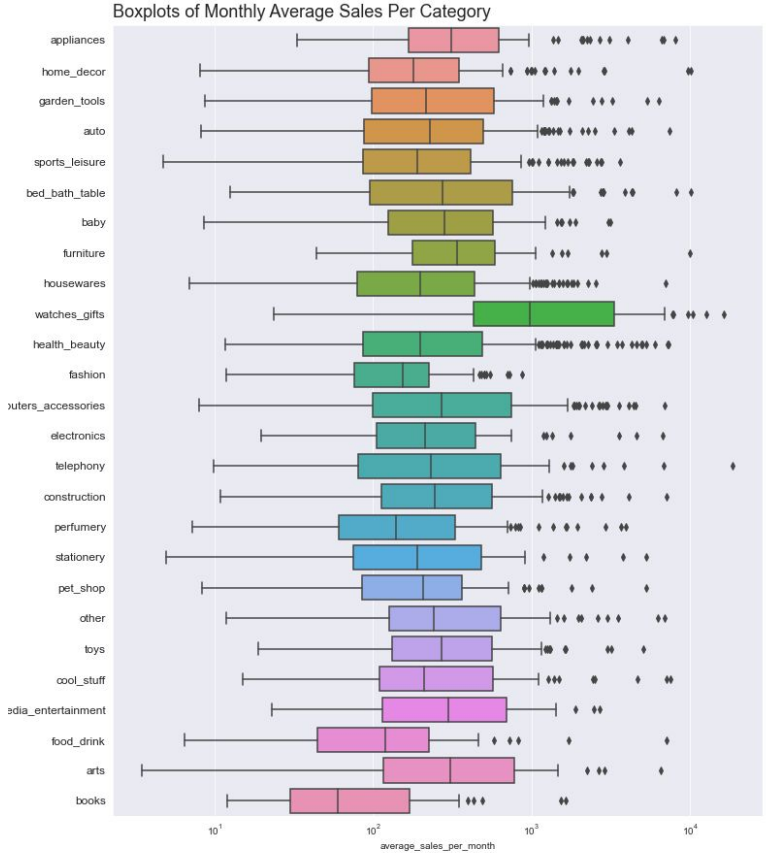
Note: As this is sample data, we do not get the real-life situation of seller performance. The best case scenario is to perform these steps on the entire seller performance across a fixed period of time.

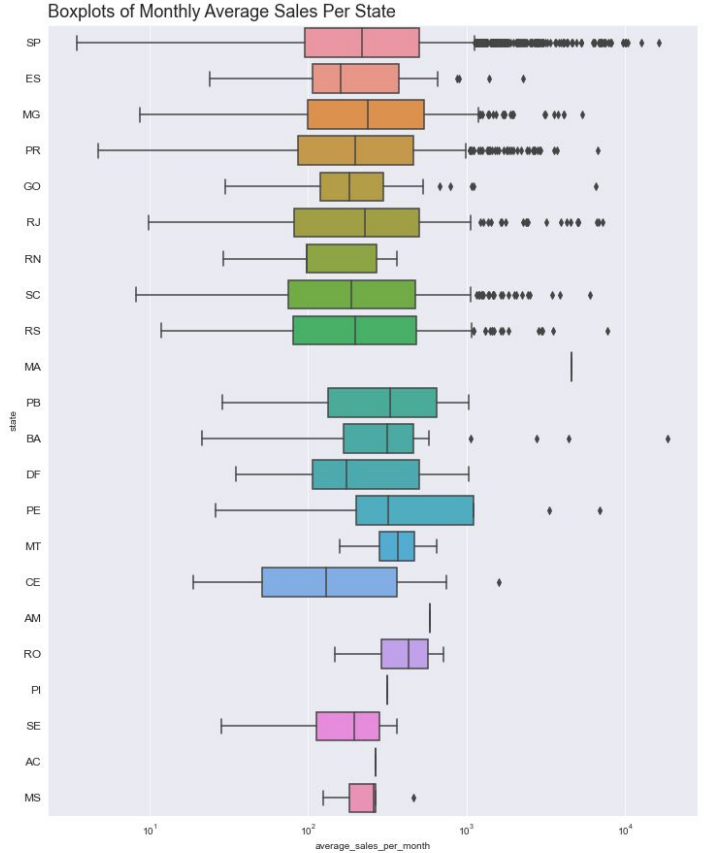
Seller average sales per month





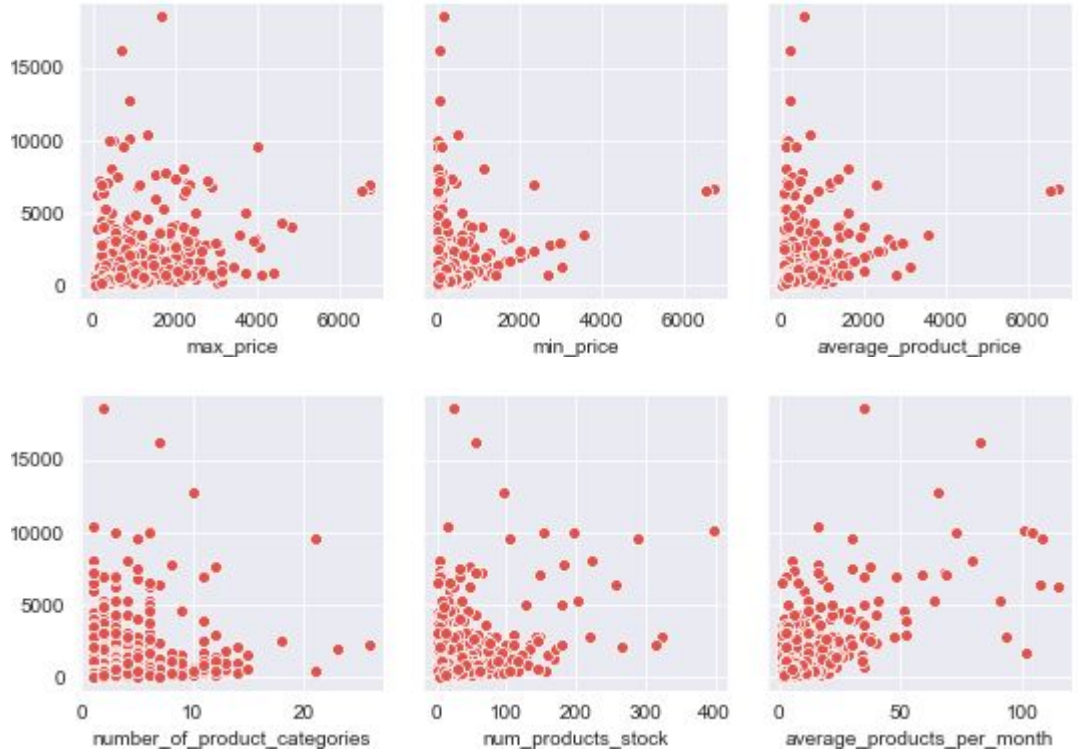
Seller average sales per month





Seller average sales per month

Scatterplot with continuous input features



Solving our regression problem

As we are predicting a continuous variable, this is a regression problem. Two popular methods for regression are linear regression methods and decision trees.

Input Variables

Product volume / month
Min, max, average product price
No. of product categories
No. of unique products
Dominant product category (One hot encoding)

Tested:
State (One hot encoding)
Zip codes (One hot encoding)

Target Variable

Average Sales / Month

Average Sales / Month (Log)

Evaluation Metrics

Root Mean Squared Error

Baseline: 950

R2 Score

Baseline: -0.0

Final model

Input Variables	Target Variable	Evaluation Metrics
Top predictors: average_products_per_month max, min, average price number of products in stock number of product categories watches_gifts	Average Sales / Month	**Test Data** **RMSE:** 209 **R2 Adjusted:** 0.94 **Train Data** **RMSE:** 153 **R2 Adjusted:** 0.98

The final model selected was a **Random Forest model** which leverages decision tree algorithms. The top features in deciding the variable were product volume / month and the minimum, maximum and average product prices.

How much can he earn?

I'm interested, but how much can I earn?

I sell **books** online to earn part-time income, and the price range is from **R$40 to R$100**. I also sell some **CDs and DVDs.** I can sell up to **50 products** a month from my inventory and have around **10 unique listings** at one time.

I am situated in **Sao Paulo**.

Model Prediction:
R$13,634 a month

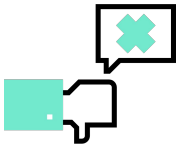
How can sellers increase their sales?

Judging our model

Pros

Accurate on Test Data

Simple to Implement



Cons

Reproducible?

Business Concerns

Moving forward

Better Data & More Features

- More comprehensive data
- Predict optimal seller product volume
- Get more seller and product characteristics (size of company frequency of catalog update, product image and descriptions etc.)

Evaluation

- Existing sellers
- Incoming sellers

Use Cases

- Market Olist to potential sellers
- Gauge potential of new sellers for our marketing team and client relations team

Thank You!

Please let me know if you have any questions.