Classification of sellers with high probability to stop selling with Olist and **cluster analysis**

Context

Olist is inserted in the marketplaces context, giving to the small sellers the opportunity of selling its products in the great brazilian e-commerces.

Considering its revenue comes from a percentage of transactions made in case that Olist inserted the seller in the e-commerce, we understand that the variables who has more impact on revenue are:

- Volume of sellers;
- Sells frequency of the sellers;
- Value of each transaction (average ticket).

- **Business problem**
  High number of sellers who stop selling with Olist in the first months of operation.

- **Motivation**
  Decreasing the number of sellers who stop selling with Olist, the company can make more revenue in the long-therm, because Olist already has a significant entrance of sellers through time.

- **Method**
  Identifying sellers with high probability to stop selling and make clusters, for possible priorizations and different action plans for each cluster.

Solution's architecture

AZURE DATA LAKE STORAGE
-Original and processed datasets
-Classification model
-Clustering model

AZURE DATABRICKS
-Python scripts for data processing

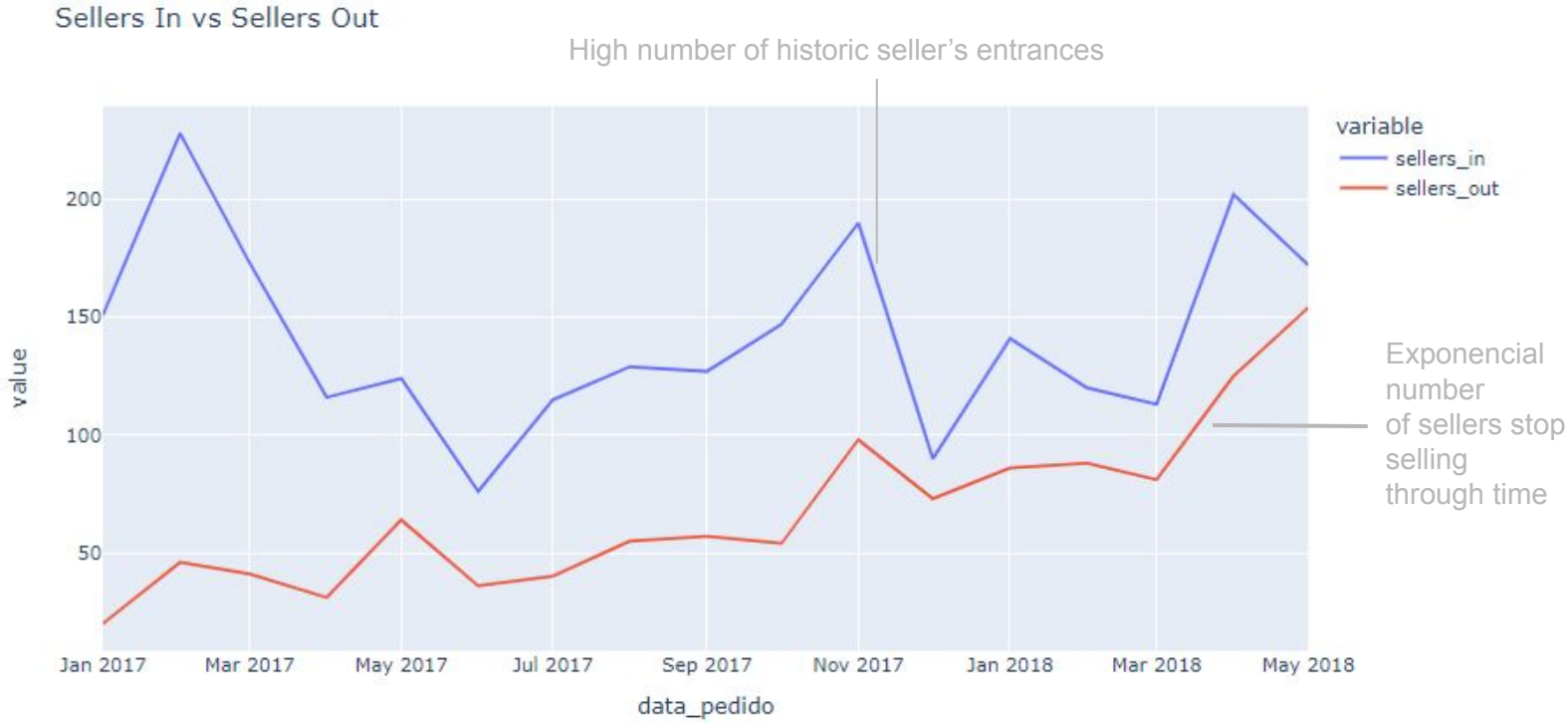
PYCARET

python™

AZURE DATA FACTORY
-Cloud ETL service to scale-out serverless data integration and transformation using python scripts located in Databricks

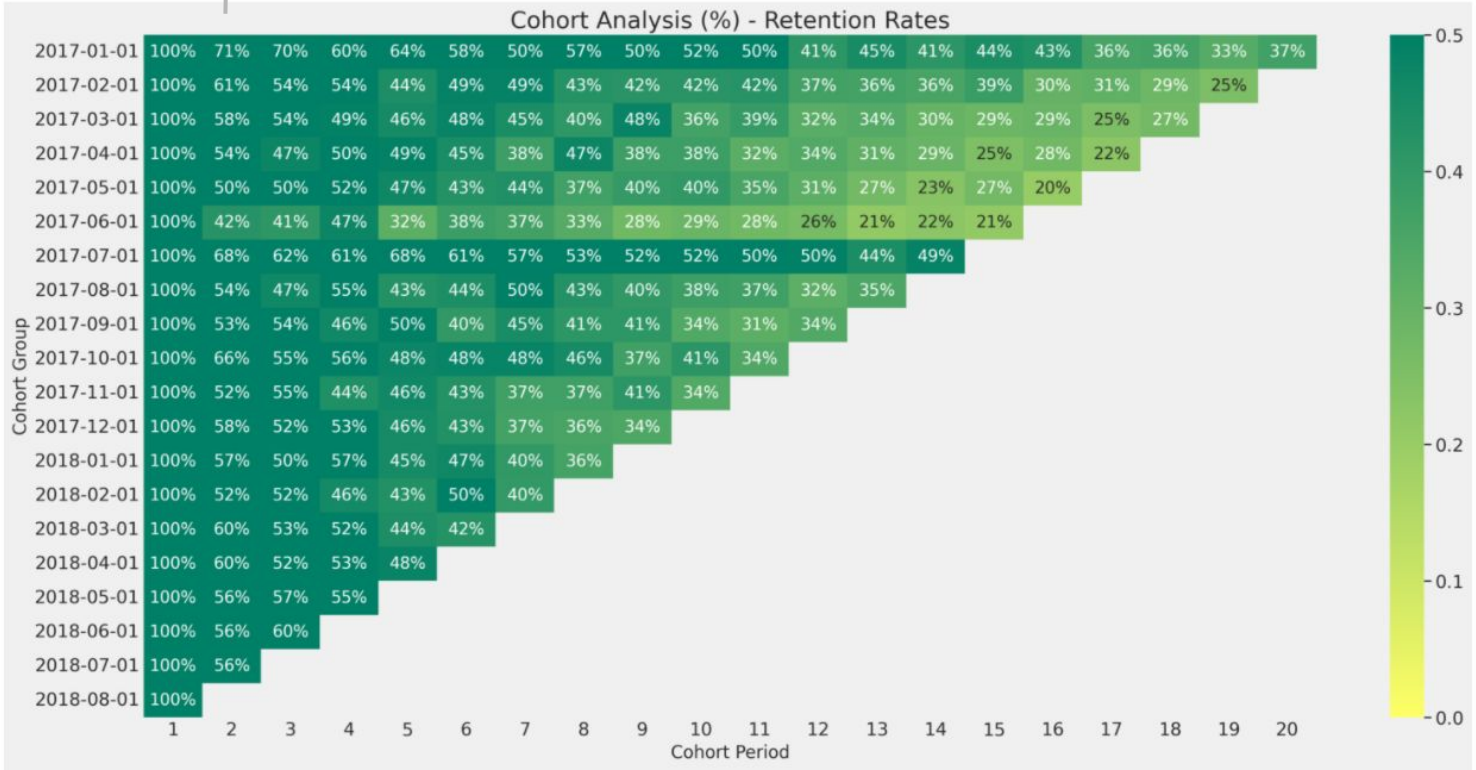
Streamlit

Exploratory analysis



Exploratory analysis

Sellers who joined Olist in Jan/2017, Feb/2017 and sold in the second, third month and so on…

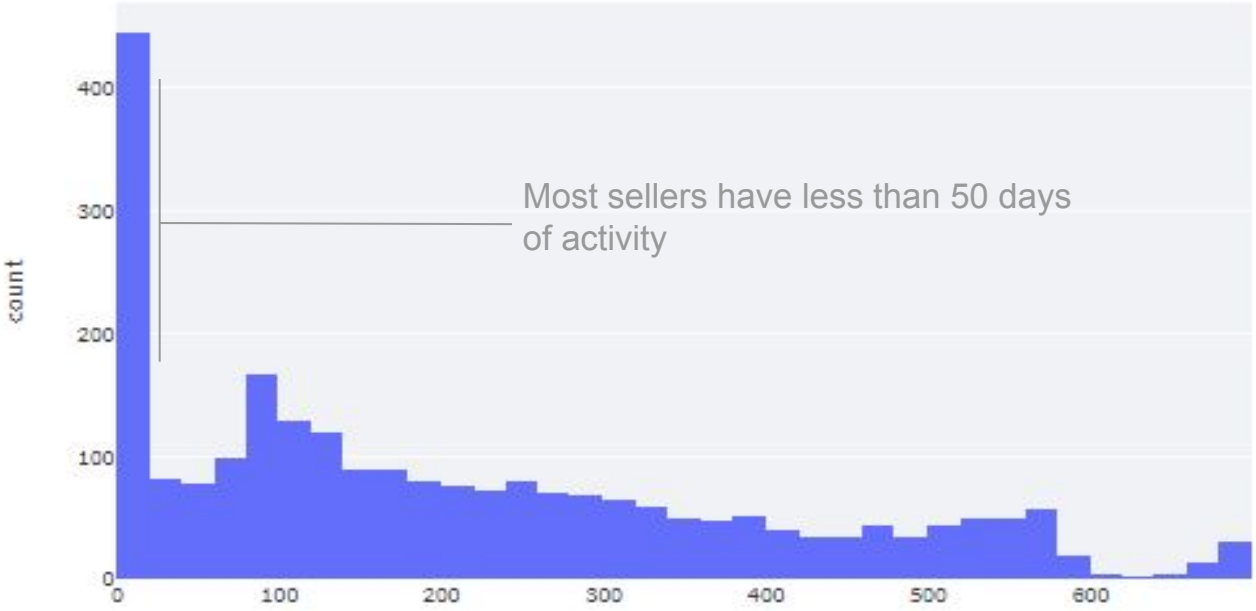


Average of **46 p.p* of the sellers doesn't sell products in the next month of entrance.** The percentage keeps decreasing through time

*p.p: percentual points

Exploratory analysis

Distribution of number of days between first sell date and last sell date



Looking through these insights, **Olist should spent some effort to retain sellers on its base**. The strategy to make this possible, could be analyzed between the business and data teams.

A way to make it easier is creating a **Data App that allow the CRM team to get sellers with high possibility to stop selling**, understanding its characteristics.

Here is the url to the Data App where you'll find the classification and cluster analysis:
https://share.streamlit.io/streamlitstack/olist_stack/main/app.py