

1. Introduction

Olist is a marketplace that allows different companies to sell their products online. Many of them are small producers or traders with little experience in business administration. My goal is to help them identify potential ways to improve their sales and success on this specific marketplace. The client of this project are product sellers on olist and olist itself. The goal of my analysis is therefore to give some recommendations to the sellers on how to increase their sales on the platform.

2. Questions - Dataset

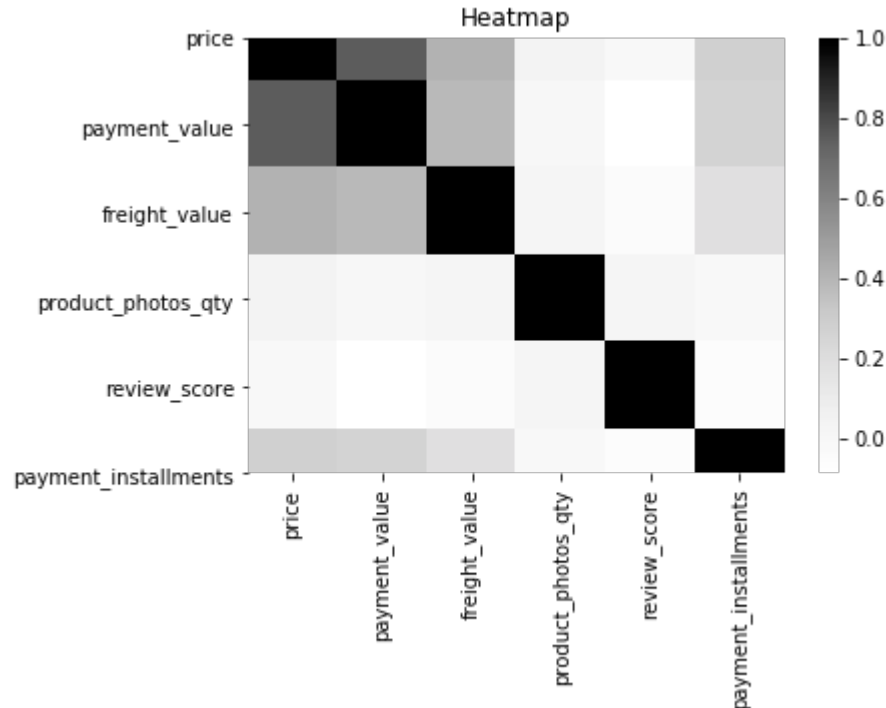
| Category | Question |
|----------|---|
| Products | <ul style="list-style-type: none">• What are the most popular product categories?• What is the most popular product?• Are products with higher ratings more popular?• Are some products more popular in some states?• Seasonality in Sales? |
| Payments | <ul style="list-style-type: none">• What is the preferred payment method?• Does the payment method depend on the state?• Do people use more than one payment option at the same time?• Do people use installments to pay for their orders?• Correlation btw. purchase value, and installments |
| Delivery | <ul style="list-style-type: none">• What is the average delivery time? Does it depend on States?• Is fast delivery correlated with higher reviews?• Are reviews higher if the package arrives before the anticipated time?• What is the average freight value? |
| Sellers | <ul style="list-style-type: none">• Revenue of the most successful seller?• Revenue distribution of sellers?• What do they sell on average?• Do positive reviews lead to more sales?• Where are sellers located?• Do sellers sell more locally or over the entire country? |

3. Information about the Dataset

3.1. Description of the Dataset

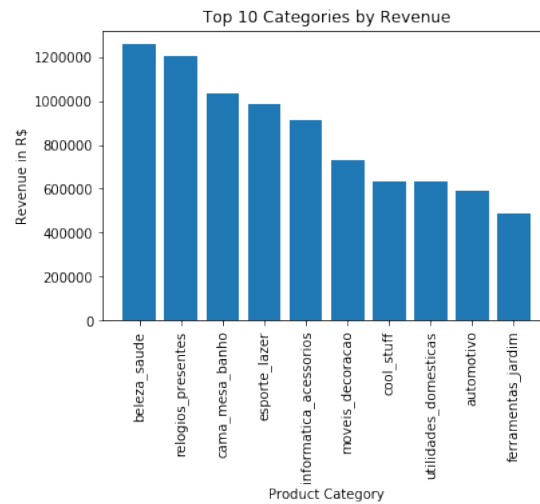
The dataset I'm using for my capstone project is available [here](#). It is a dataset about an e-commerce marketplace in Brazil. It contains csv-files with information about around 100'000 purchase orders on the platform. The files come from a relational database and can be merge in the following way:

4. Exploratory Analysis



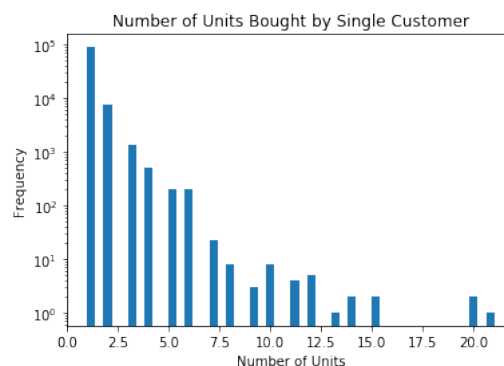
- The strongest correlation can be examined between freight value and the price of the product, a potential explanation of that is that bigger products generally have a higher value.
- There is almost no correlation between the price and the review score - probably the expectations of the user are in line with the price paid and if the price is lower the customer already expects a lower quality.
- There is a low correlation between the number of installments and the price, a potential explanation would be that there are several customer groups: higher income customers that generally do not use installments, lower income customer that use installments more frequently - it would be interesting to investigate whether within this subgroup the relationship is different. Another possibility is that paying everything at once gives the user discounts and thus people are more incentivized to pay all up front.

4.1. Products



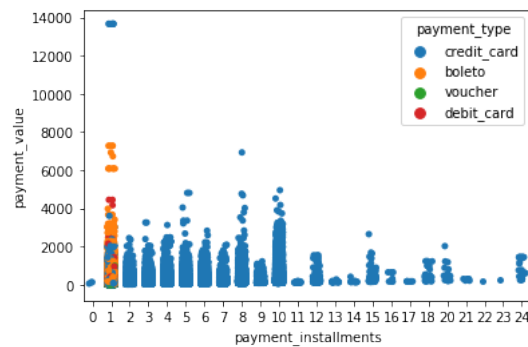
- Looking at the answers above shows, that the best categories for e-commerce are things that people use often but it is generally a specific size. One of the difficulties of e-commerce stores is still that it is hard to choose sizes for clothes for instance, therefore people are more likely to buy these things in a physical store and buy online things that they know what they are buying.
- The median revenue of a product is BRL 136.68 and the mean is 412.47 this shows not only that the average product does not generate a lot of revenue, but also confirms the hypothesis that there are many product that almost do not sell at all and some products that generate a lot of revenue.
- The best selling product generates BRL 63'885, as we only have the product_id we cannot investigate what product it is.
- There seems to be no correlation between the review score and the revenue of the product.

4.2. Customers



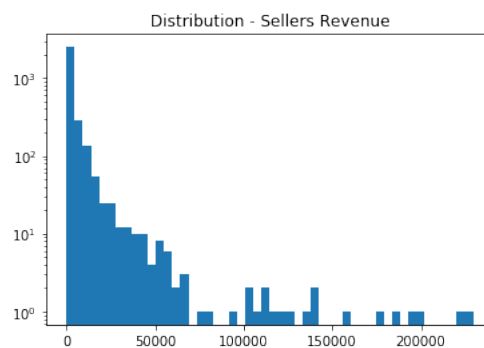
- Similar to the sellers, the majority of buyers conduct very little purchases, there are some exceptions that buy a lot

4.3. Installments



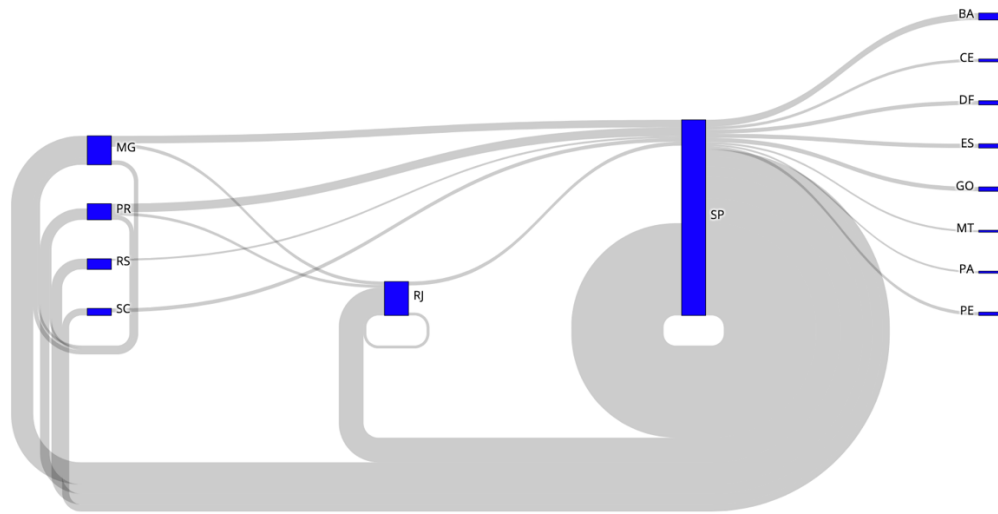
- It seems like there is no clear trend between the number of installments and the payment value of the order. It would be interesting to test whether the income has an impact on the number of installments chosen, but unfortunately due to the data available this is not possible.

4.4. Seller



- There are many sellers on the platform, but a few of them are able to create a significant revenue, this can be seen in the distribution of the seller's revenue, the mean revenue and the median revenue. The median revenue is with BRL 821.48 substantially lower than the mean BRL 4'391.48. This means that there are some outliers that highly influence the mean, but do not have an impact on the median.
- It seems like there is no relationship between the order value and the total revenue of the seller.
- The largest seller created a revenue of BRL 229'472.63 in the examined period.

4.5. Geographical Relationships of Sales



- The Sankey Diagram confirms that São Paulo is the economic powerhouse of Brazil, most of the revenue is generated within this state (followed by Rio de Janeiro and Minas Gerais). The origin of the products are also mainly these states, but there is a significant part of revenue also coming from other states.