Prediction Customer Lifetime Value Using Pareto NBD

Final Project by Michael Jacob E J

Customer Lifetime Value

Brazillian E-commerce Olist

Business Background

- The olist store is a sales model in which your products are displayed and marketed within olist on the largest marketplaces in Brazil and other relevant e-commerces.
- The olist store understand that customers are their most important asset and that it is imperative to estimate the potential value of this asset.
- Currently, there is an escalating demand for web services and web applications, along with continuous growth in worldwide commercial application. the olist store realize that they must be able to measure the value of their web customers, i.e. the expected profit that will be derived from the relationship with web customers from the present until a specified period in the future.
- So to solve this problem issues the olist store need The ability to accurately predict value of a company’s customers that has a large impact to achieve long-term gain.

Customer Lifetime Value

Brazillian E-commerce Olist

Problem Statement

- which customers have the highest spend probability in next 365-days?
- What is your action and recomenddation for customer churn?
- which customers have the lowest spend probability in next 365-days?
- etc

Objective

1.Distinguish active customers from inactive customers.
2.Generate transaction forecasts for individual customers.
3.Predict the purchase volume of the entire customer base.

Proposed Solution

use machine learning to know value of transactions made by a customer with your business over his entire lifetime. Here the lifetime means the time period till your customer purchases with you before moving to your competitors

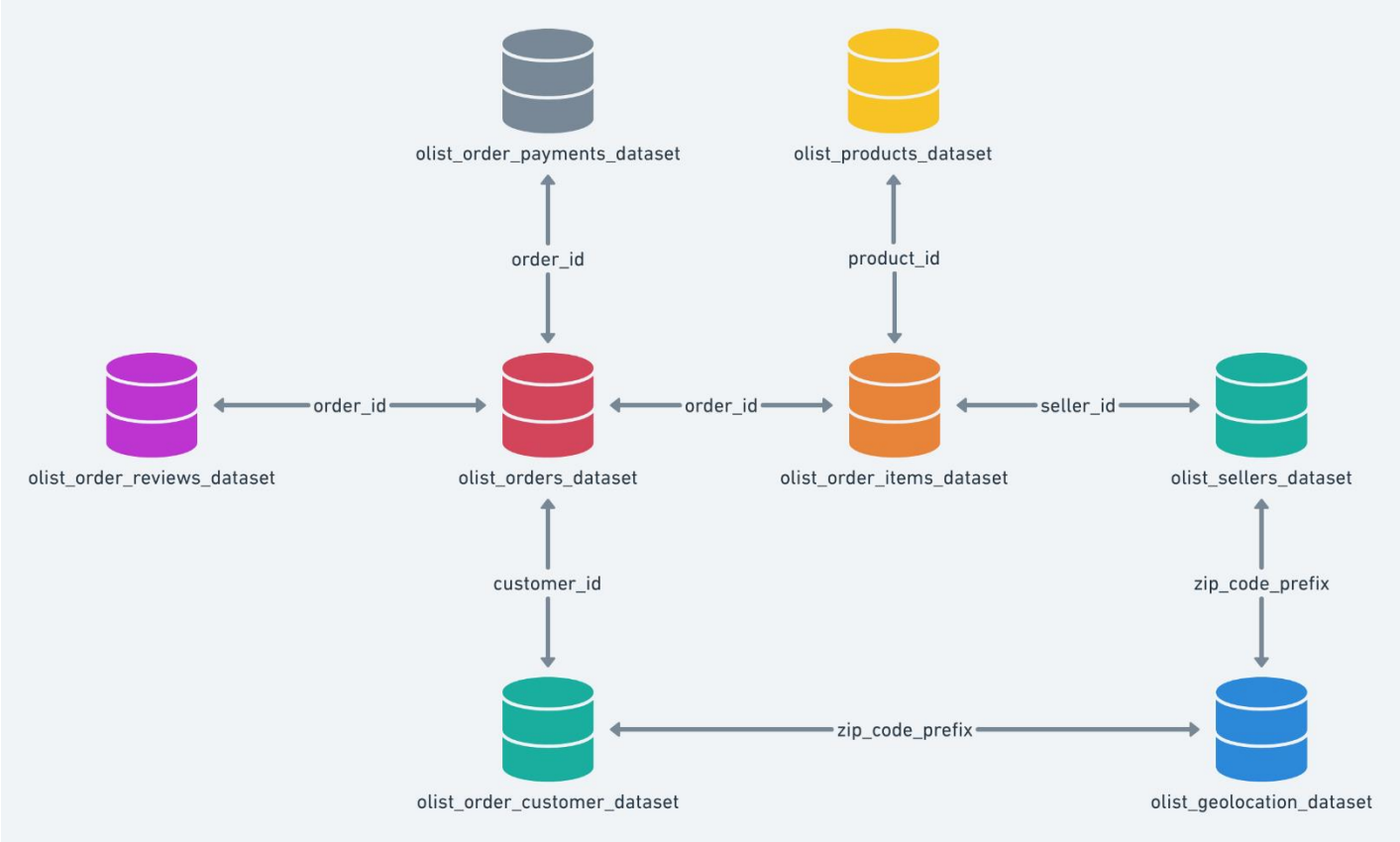
Business Benefit

1. companies can take the short term risks necessary to achieve long-term gain.
2. Marketing managers can develop different offers for different customers, based on the estimated contribution of such customers in the future.
3. Stock exchange analysts can use the company's customer value in order to evaluate its worth to potential investors.
4. Customer value can serve as the basis for decisions regarding new campaigns, allocation of retention vs. acquisitions budgets, etc.

Customer Lifetime Value

Brazillian E-commerce Olist

Data Schema



Customer Lifetime Value

Brazillian E-commerce Olist

Data Understanding

1. The dataset has information of 100k orders from 2016 to 2018
2. Source Data:https://www.kaggle.com/olistbr/brazilian-ecommerce
3. The dataset has 11 column and 98665 rows
4. Just merger 3 csv for predicting customer lifetime value
5. Data dictionary:

- Order id : unique identifier of the order
- Customer id : key to the customer dataset. Each order has a unique customer_id.
- Order purchase time : Shows the purchase timestamp.
- Customer unique id : unique identifier of a customer.
- Customer city : customer city name
- Oder item id : sequential number identifying number of items included in the same order.
- Product id : product unique identifier
- Seller id : seller unique identifier
- Shipping limit date : Shows the seller shipping limit date for handling the order over to the logistic partner.
- Price : item price
- Freight value : item freight value item (if an order has more than one item the freight value is splitted between items)

Customer Lifetime Value

Brazillian E-commerce Olist

Data Preparation

- Python version 3.7.12
- packages : pandas, numpy, lifetimes, seaborns, sklearn, matplotlib

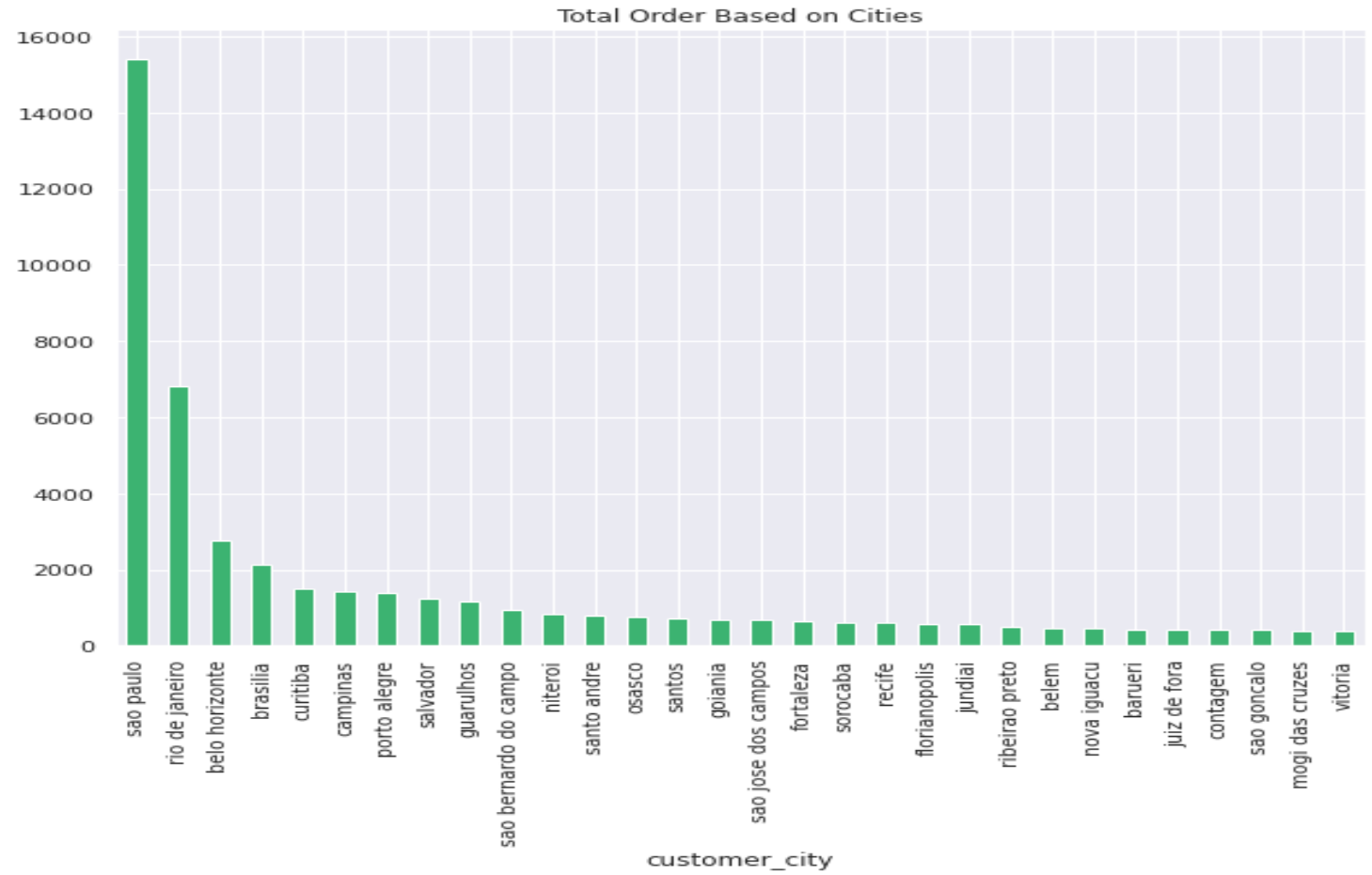
Data Cleansing

1.Remove duplicate data
2.Remove unneeded columns to predict customer lifetime value
3.Change the data type to the appropriate in each column

Customer Lifetime Value

Brazillian E-commerce Olist

Exploratory Data Anaylsis (EDA)

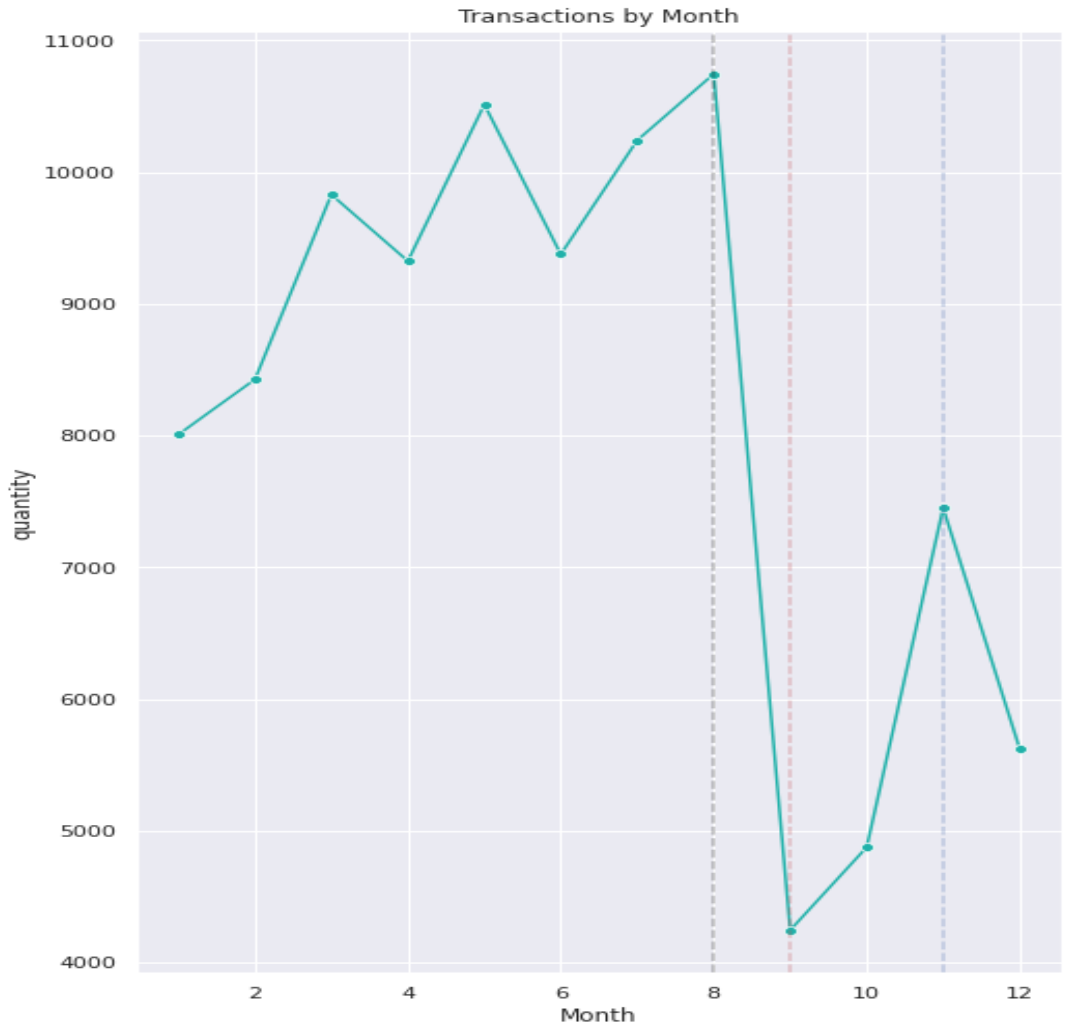


5 city with the best order are sao paulo,rio de janeiro,belo horizonte,brasilia,curitiba.

Customer Lifetime Value

Brazillian E-commerce Olist

Exploratory Data Anaylsis (EDA)

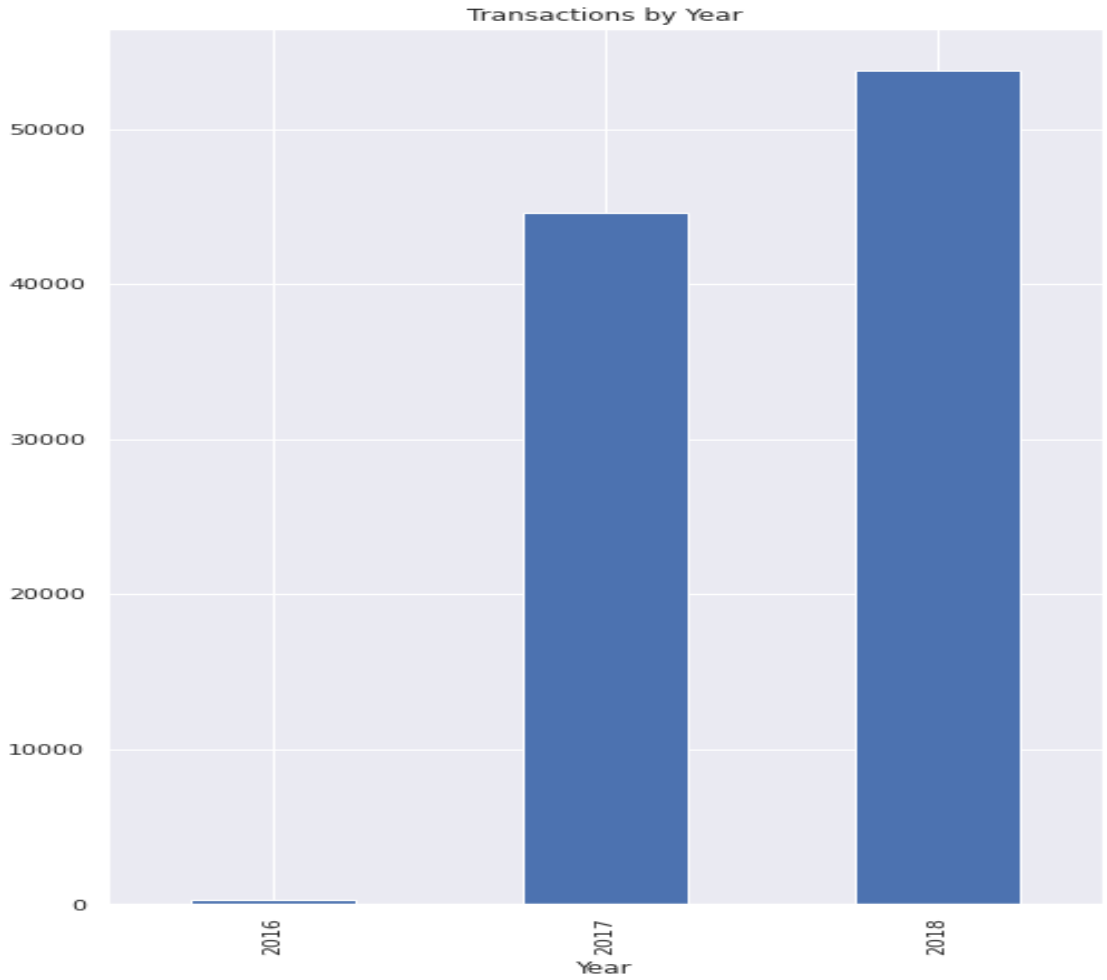


Most Transactions happened in the month of August which is evident due to festive seasons.

Customer Lifetime Value

Brazillian E-commerce Olist

Exploratory Data Anaylsis (EDA)

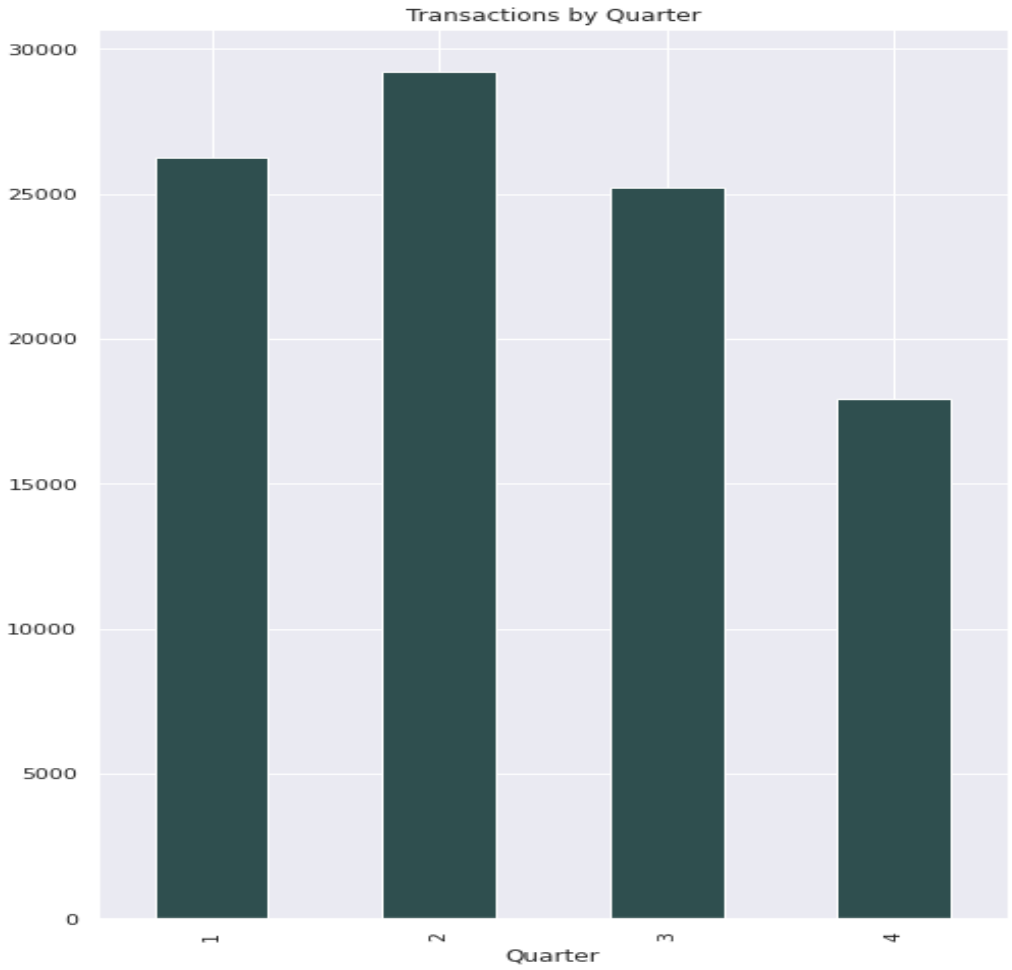


2018 is the year in which we have the most transactions followed by the 2017 and 2016

Customer Lifetime Value

Brazillian E-commerce Olist

Exploratory Data Anaylsis (EDA)

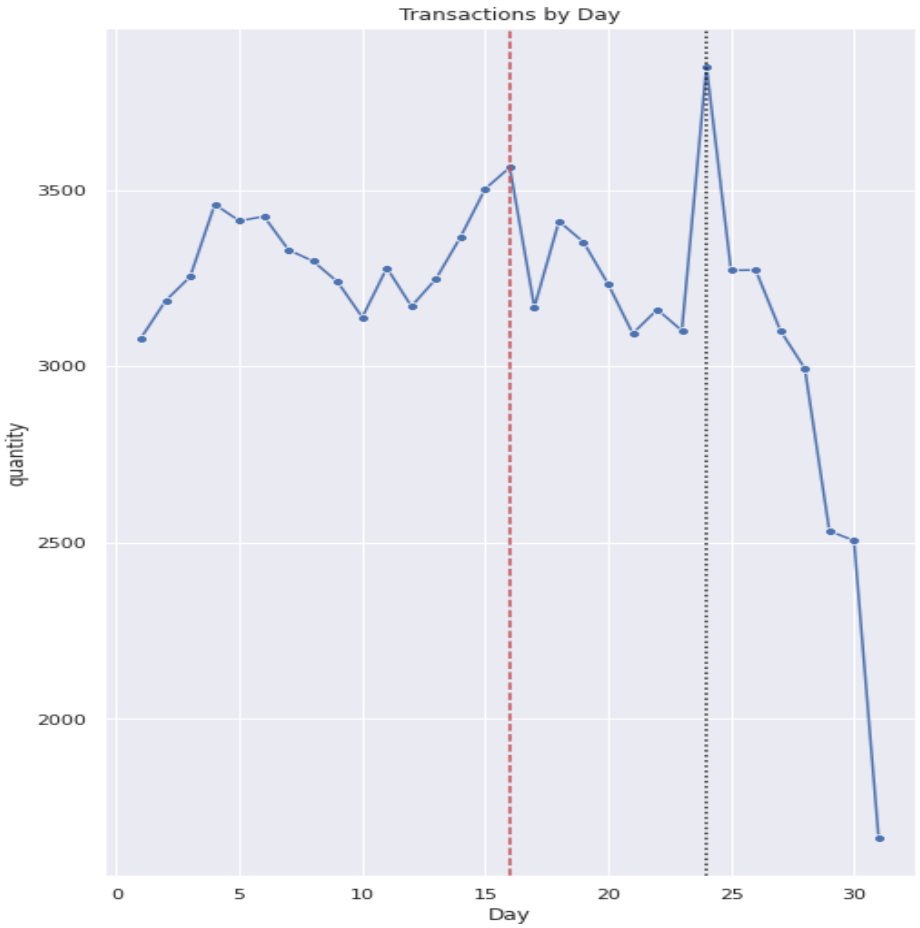


Q2 being the highest when it comes transactions.

Customer Lifetime Value

Brazillian E-commerce Olist

Exploratory Data Anaylsis (EDA)

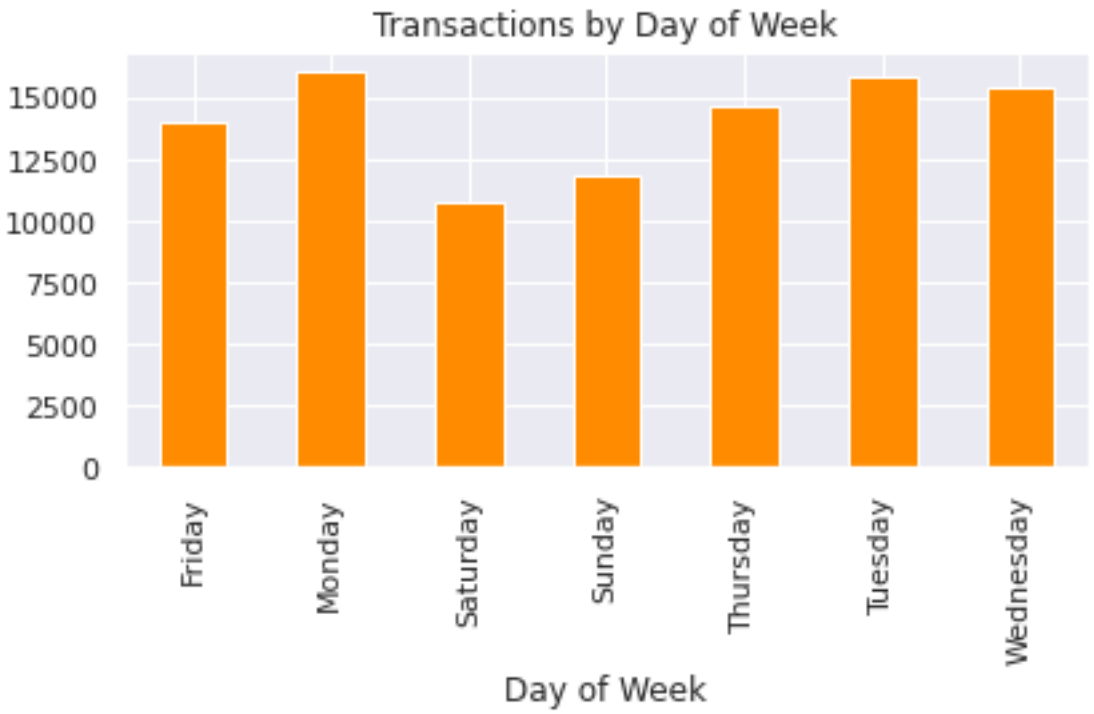


It also observed that in the end of the 1st week and starting of the 3rd week, people tends to buy more.

Customer Lifetime Value

Brazillian E-commerce Olist

Exploratory Data Anaylsis (EDA)

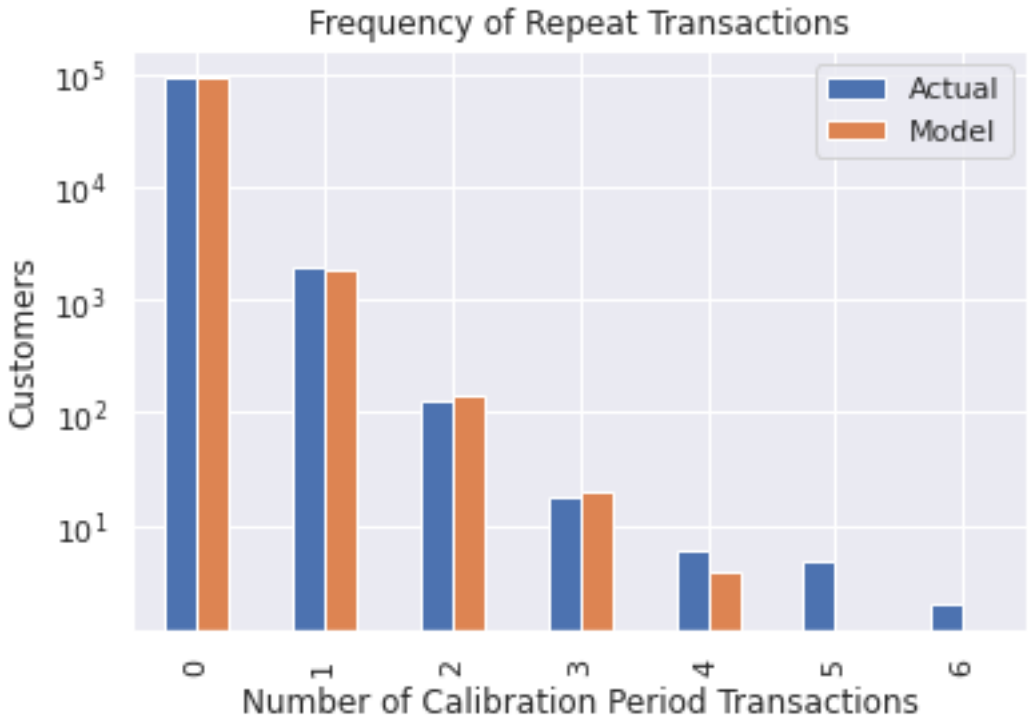


People loves to shop on monday followed by Tuesday and Wednesday.

Customer Lifetime Value

Brazillian E-commerce Olist

Pareto Nbd Modelling

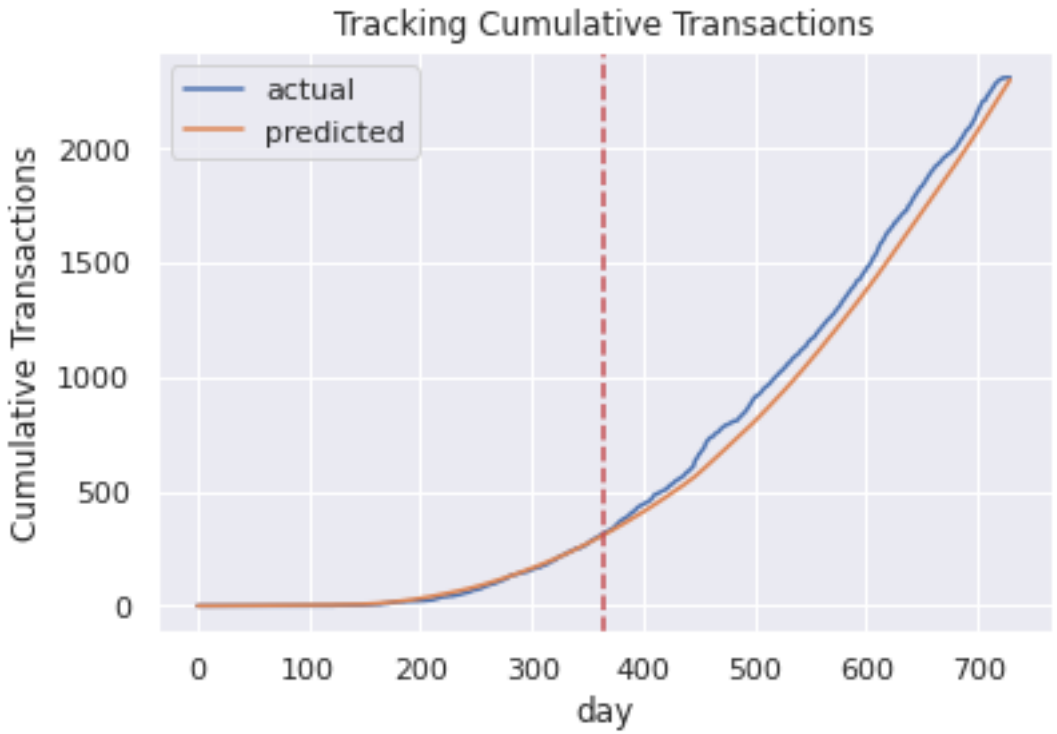


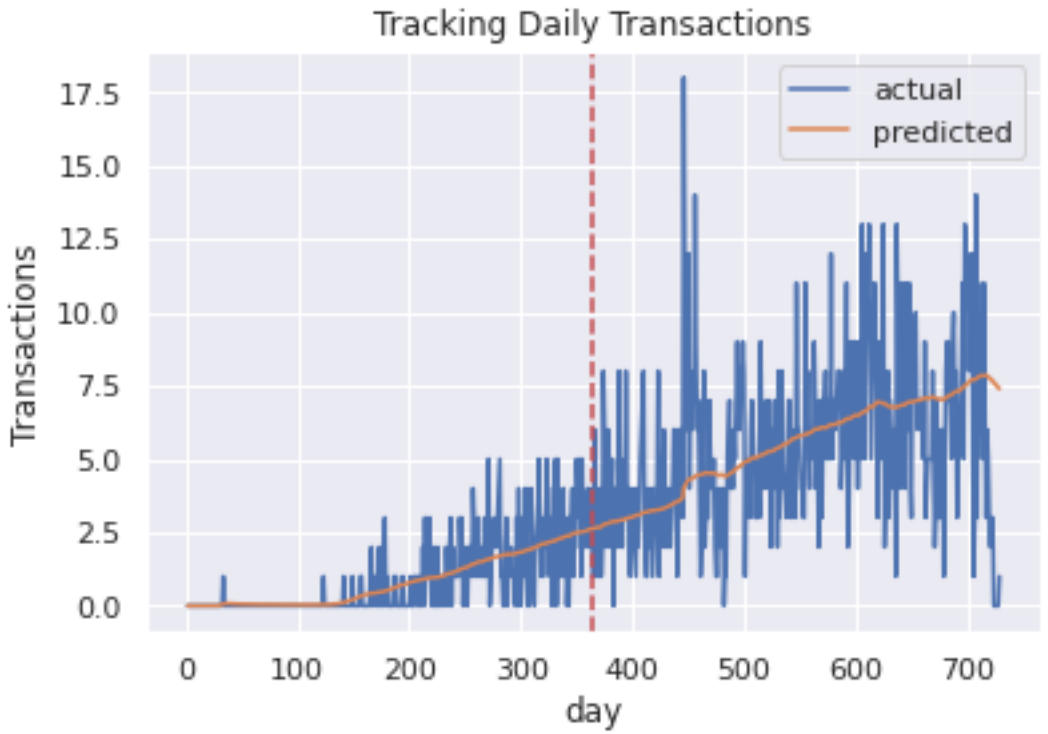
We can see that our actual data and our simulated data line up well. The model is fairly representative of the real data up until seven repeat transactions. There are few customers who make more purchases.

Customer Lifetime Value

Brazillian E-commerce Olist

Pareto Nbd Modelling





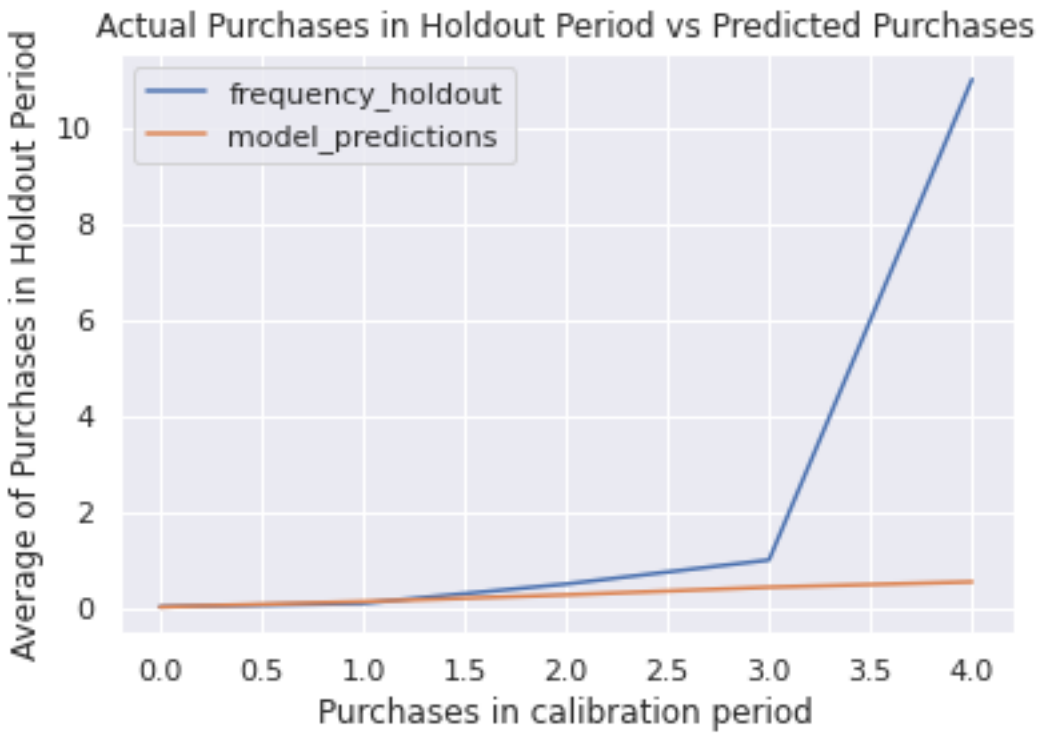
The red line represents the boundary between the calibration period on the left and the holdout period on the right. As you can see, the Pareto/NBD model is good at predicting cumulative transactions.

This plot shows that the model does a decent job capturing general trends in the data.

Customer Lifetime Value

Brazillian E-commerce Olist

Pareto/ Negative Binom Distribution Modelling



The model performs well up to three calibration period purchases, but diverges from the holdout data because of the distribution of the data.

Customer Lifetime Value

Brazillian E-commerce Olist

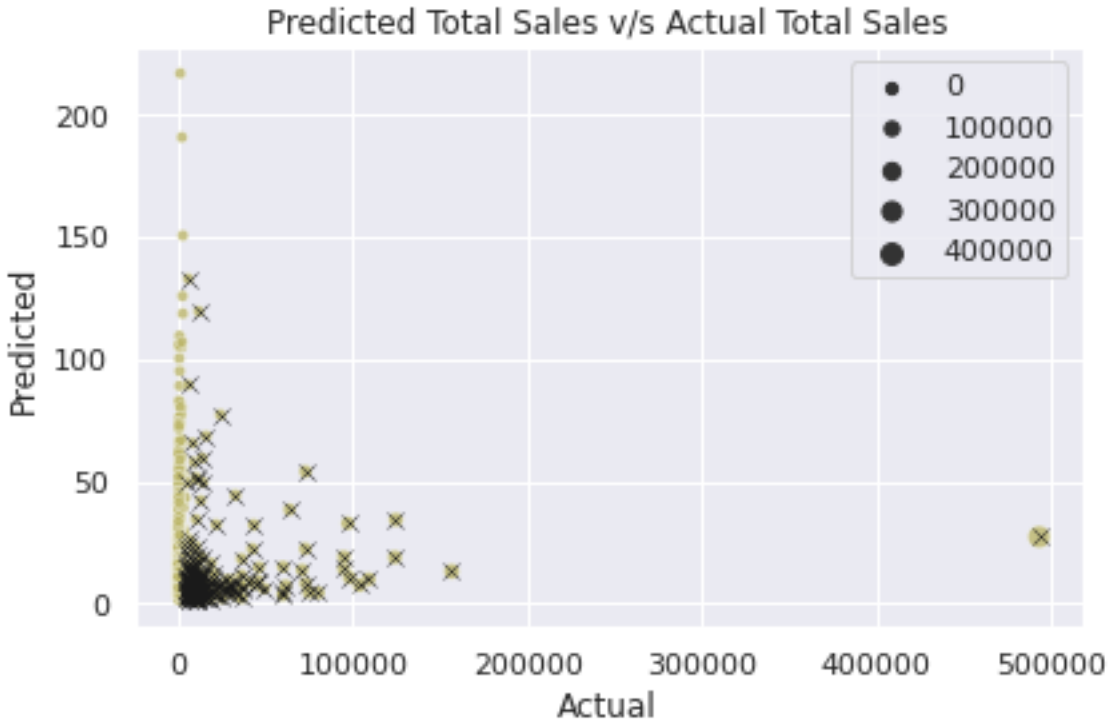
Gamma Distribution Modelling

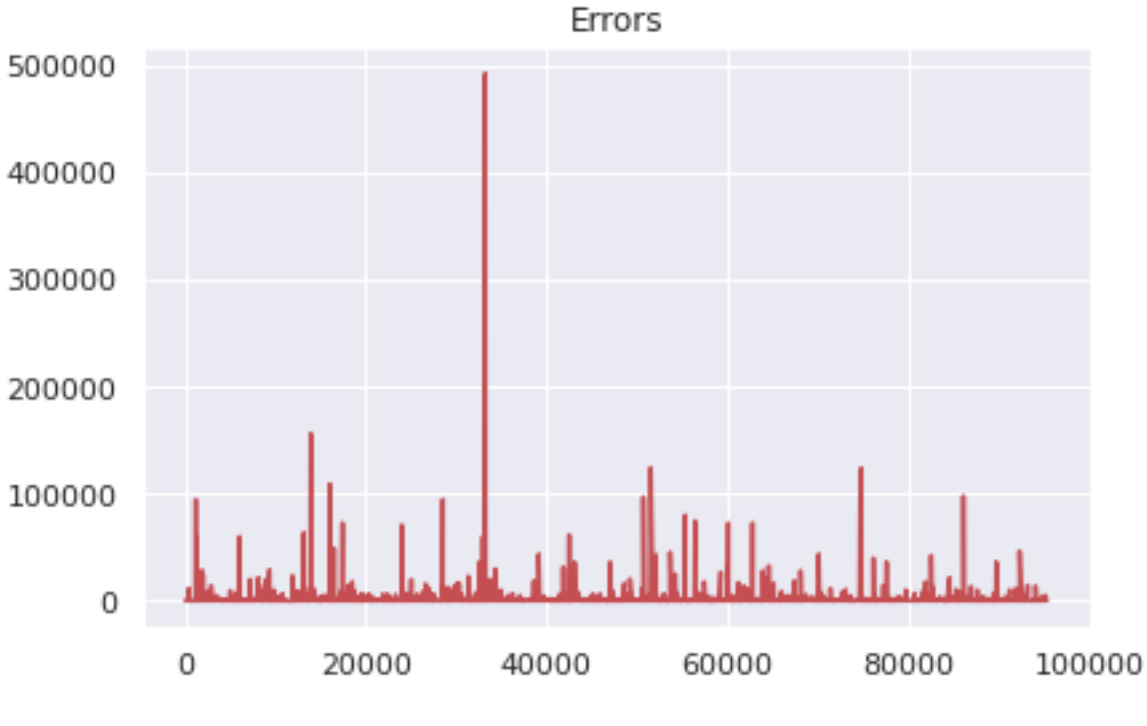
The model we are going to use to estimate the CLV for our data is called the Gamma-Gamma submodel, which relies upon an important assumption.
The Gamma-Gamma submodel assumes that there is no relationship between the monetary value and the purchase frequency. In practice we need to check whether the Pearson correlation between the two vectors is close to 0 in order to use this model.

Customer Lifetime Value

Brazillian E-commerce Olist

Pareto NBD +Gamma Distribution





Actual Avg Sales: 113.26364540367545
Predicted Avg Sales: 113.08382919908112
Mean Sqaured Error: 2940.122843020948

Root Mean Squared Error: 54.222899618343426
Avg Sales Error: 0.1798162045943314 5

Customer Lifetime Value

Brazillian E-commerce Olist

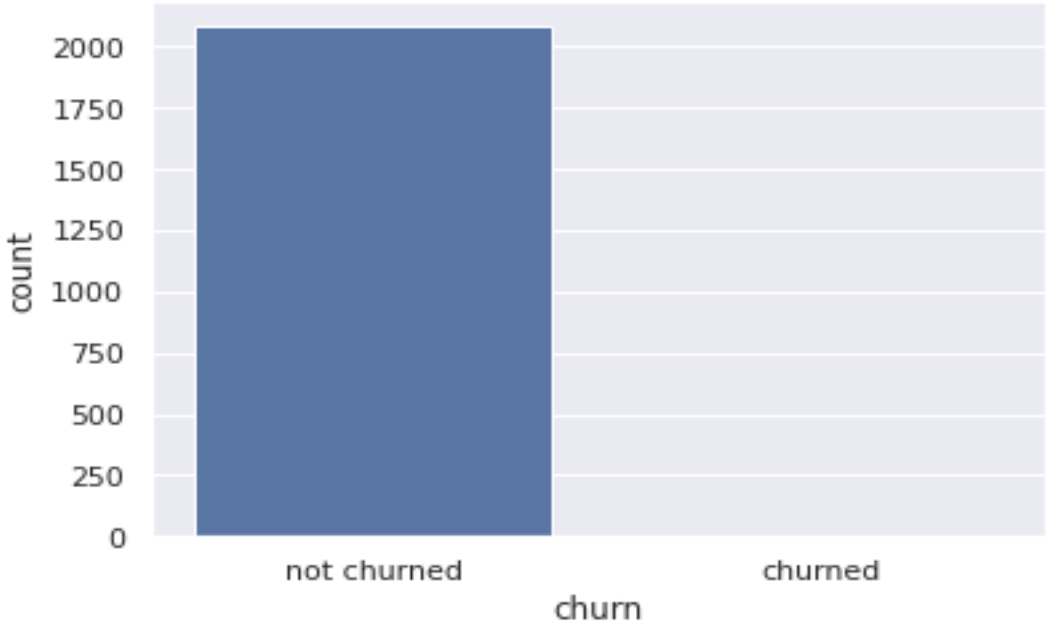
Purchase Prediction Result Summary

	Pareto-NBD
MSE Purchase Error	86.218534
RMSE Purchase Error	9.285394
Avg Purchase Error	0.480401

Customer Lifetime Value

Brazillian E-commerce Olist

Churn Risk From Pareto NBD + Gamma Distribution Modelling



not churned	1951
high risk	132
churned	2

Customer Lifetime Value

Brazillian E-commerce Olist

Action and Recommendation

not churned:
1.Reward Customers
2.Serve them additional benefits to ensure that they remain loyal customers(VIP).

high risk:
1.Identify what happens when a customer defects.
2.Set up alerts to get more details.
3.Know the ideal customer journey.
4.Don’t be afraid of the exit interview.
5.Treat former customers like friends, not enemies.

churned:
1.Give better service
2.Pay attention to complaints