

Data Science Career Track

# Predict Customer Satisfaction



By: Simantini Patil



# Target audience

## E-commerce companies



# Project Overview

**Traditional ways to measure customer satisfaction:**

1. Collecting the feedback through reviews
2. Customer survey

Many companies today measure customer satisfaction because a major factor in customer loyalty is customer happiness.

Instead of waiting until the customer's feedback/ review, companies can predict customer satisfaction while they are still in contact with the customer.

# Project Objective

The main hypothesis of this project:

**“the product and order fulfillment influences customer review rating”**

In this project, machine learning techniques were applied to the dataset to predict customer review ratings.

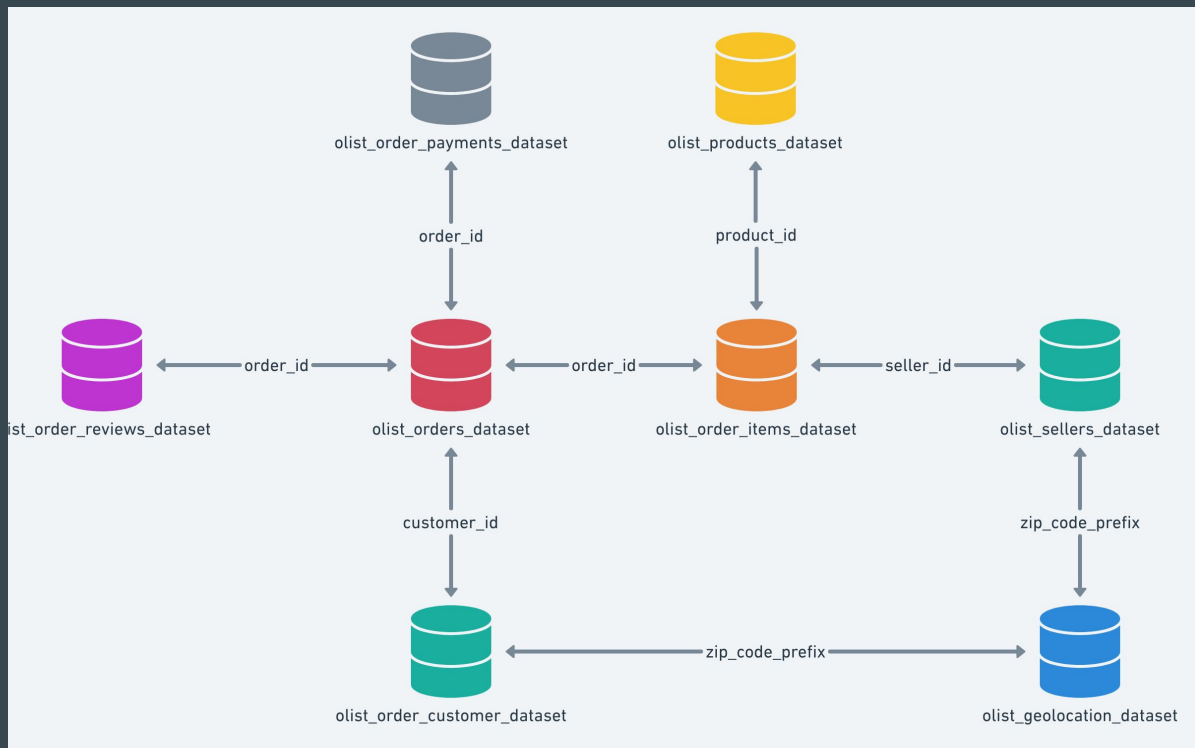
## Data set

Brazilian E-Commerce Public Dataset by Olist - obtained from kaggle

<https://www.kaggle.com/olistbr/brazilian-ecommerce>

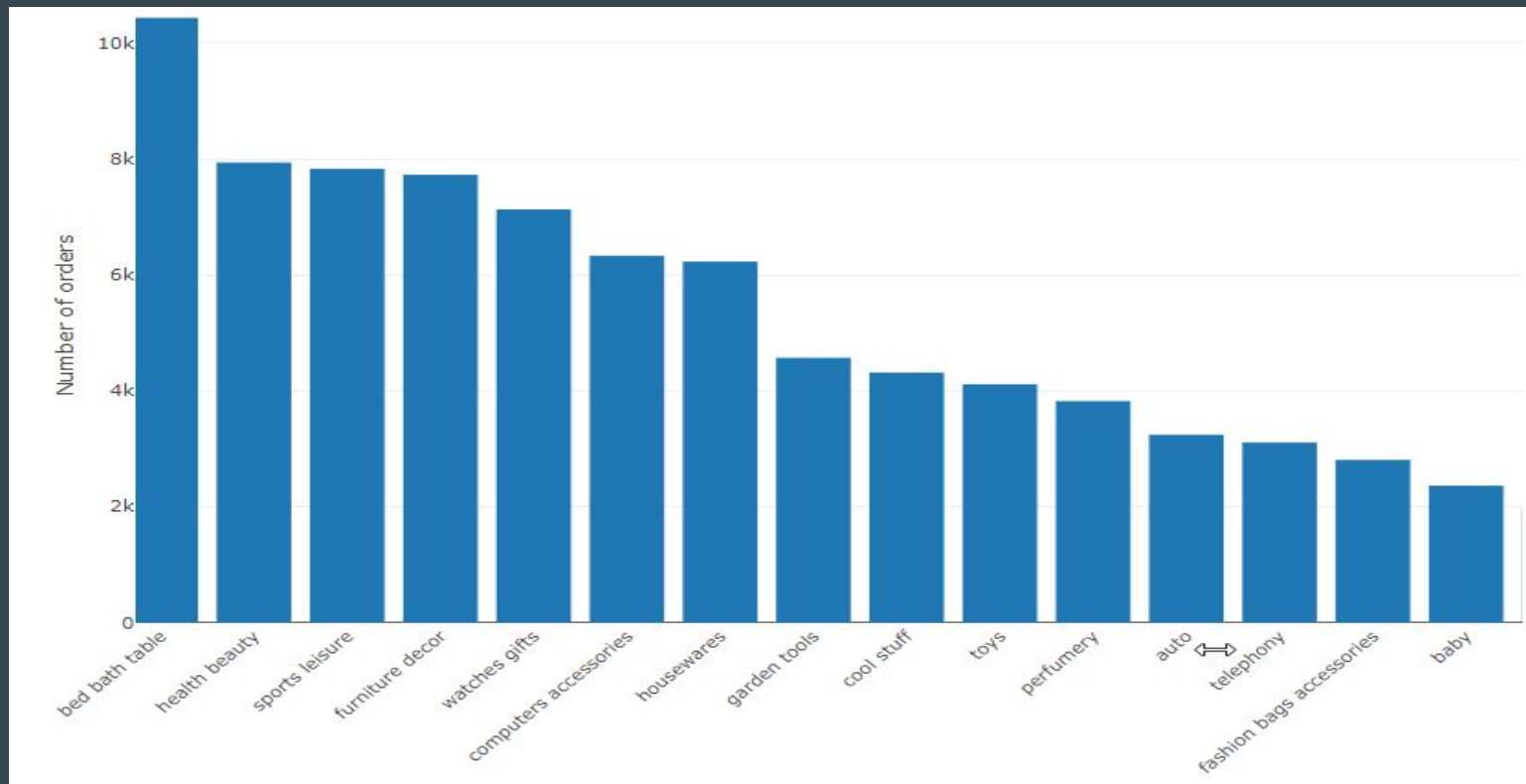
# Understanding the data

Dataset contains 10000 records and 25 columns



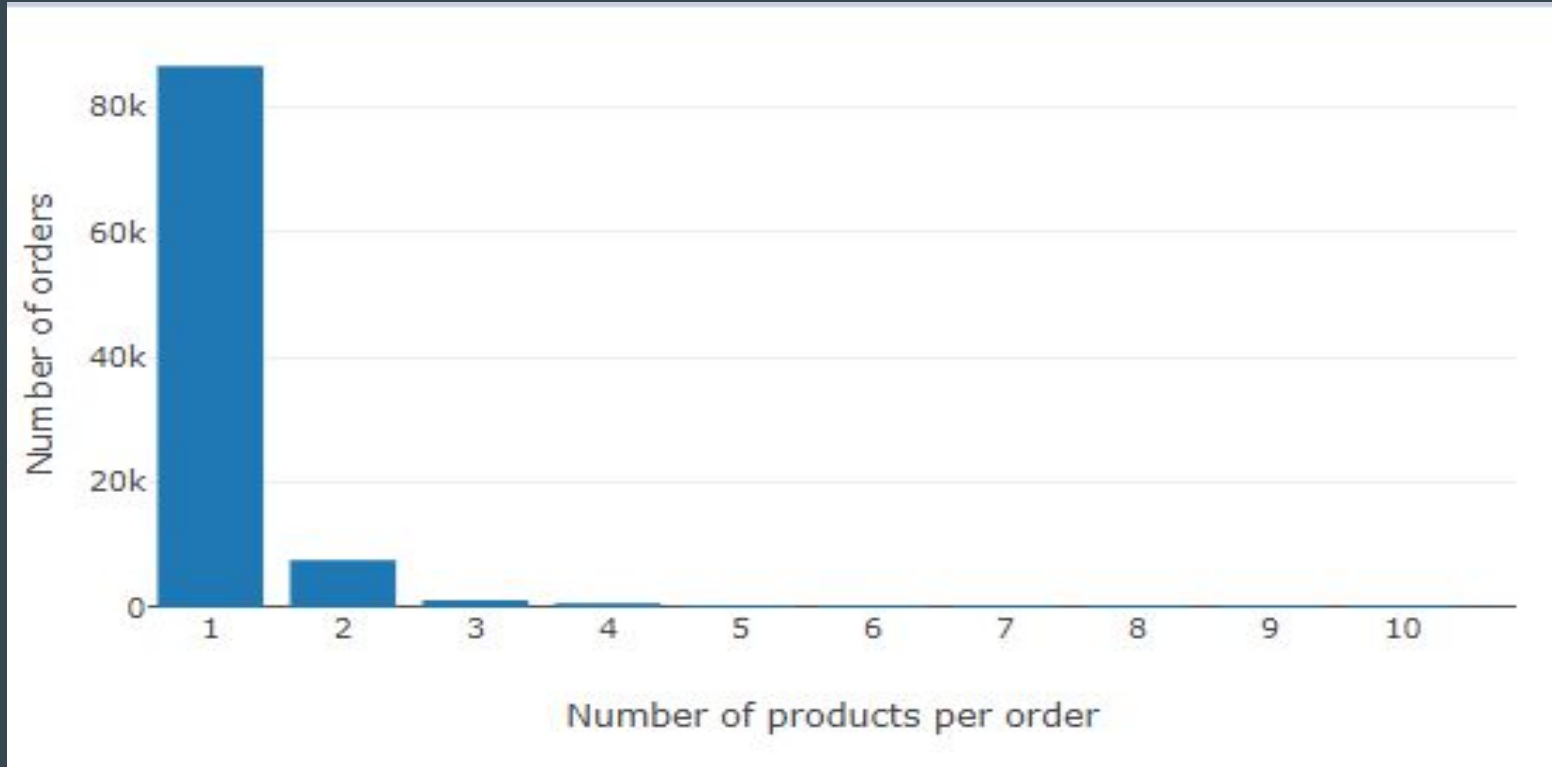
# Exploratory Data Analysis

Most bought products on Olist:



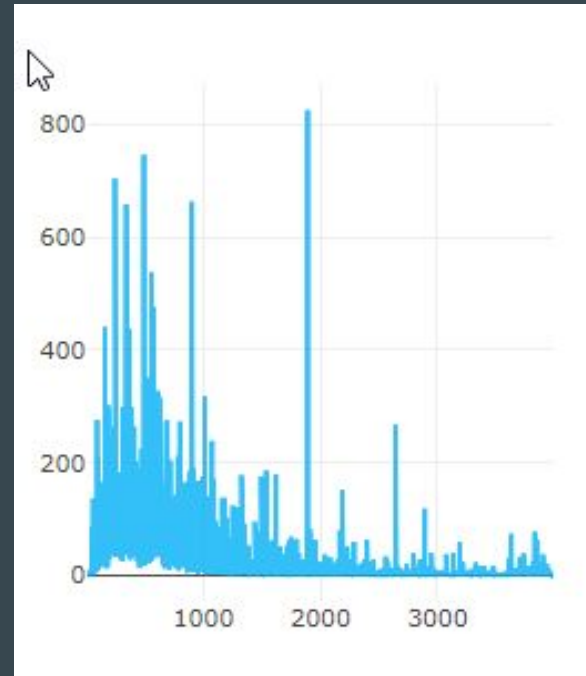
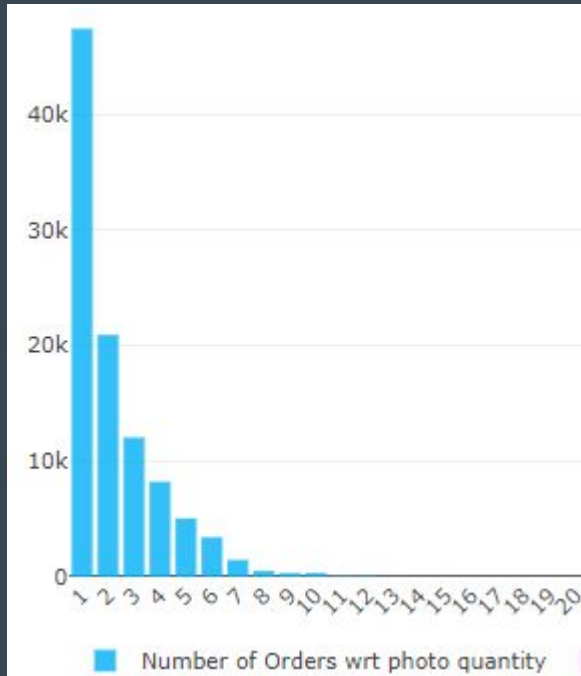
# Exploratory Data Analysis

Number of products customers usually order



# Exploratory Data Analysis

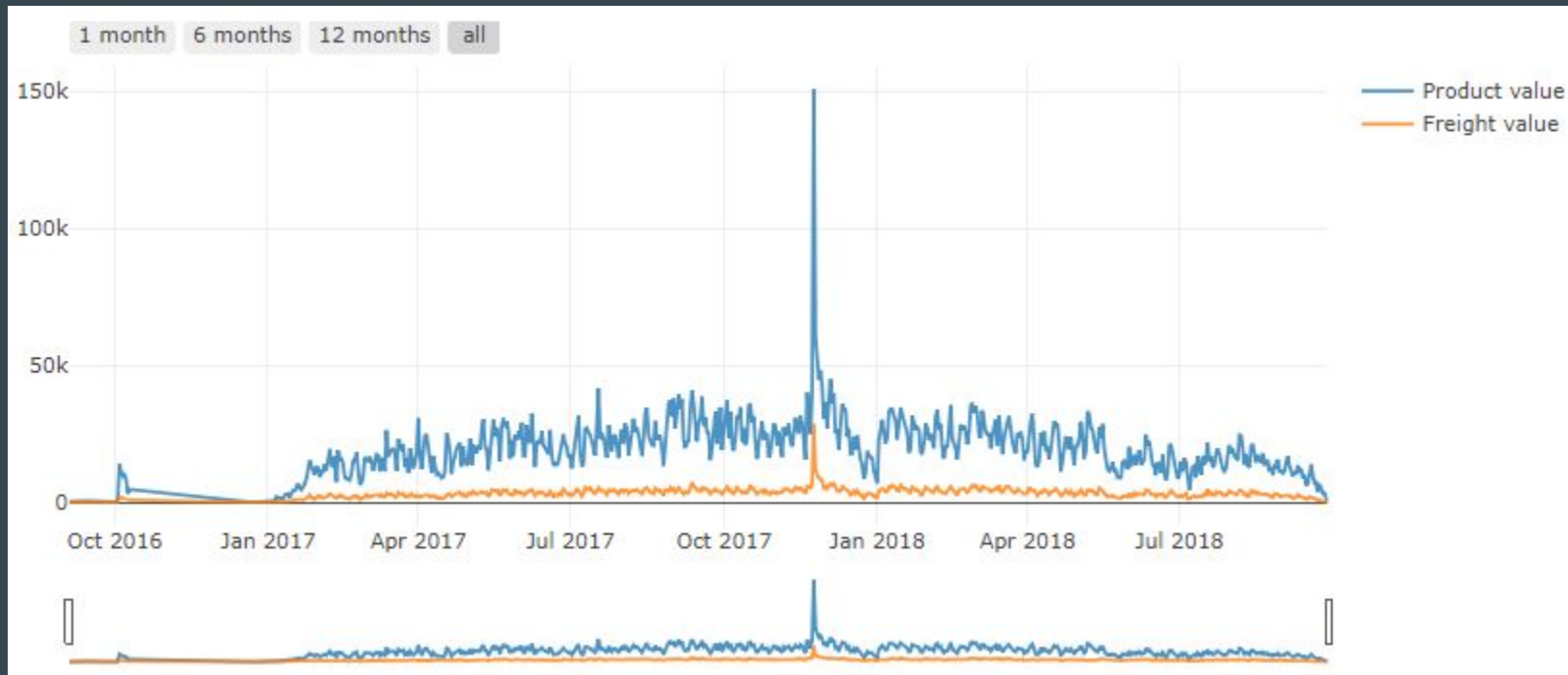
Number of photos per order and Product description length and order count





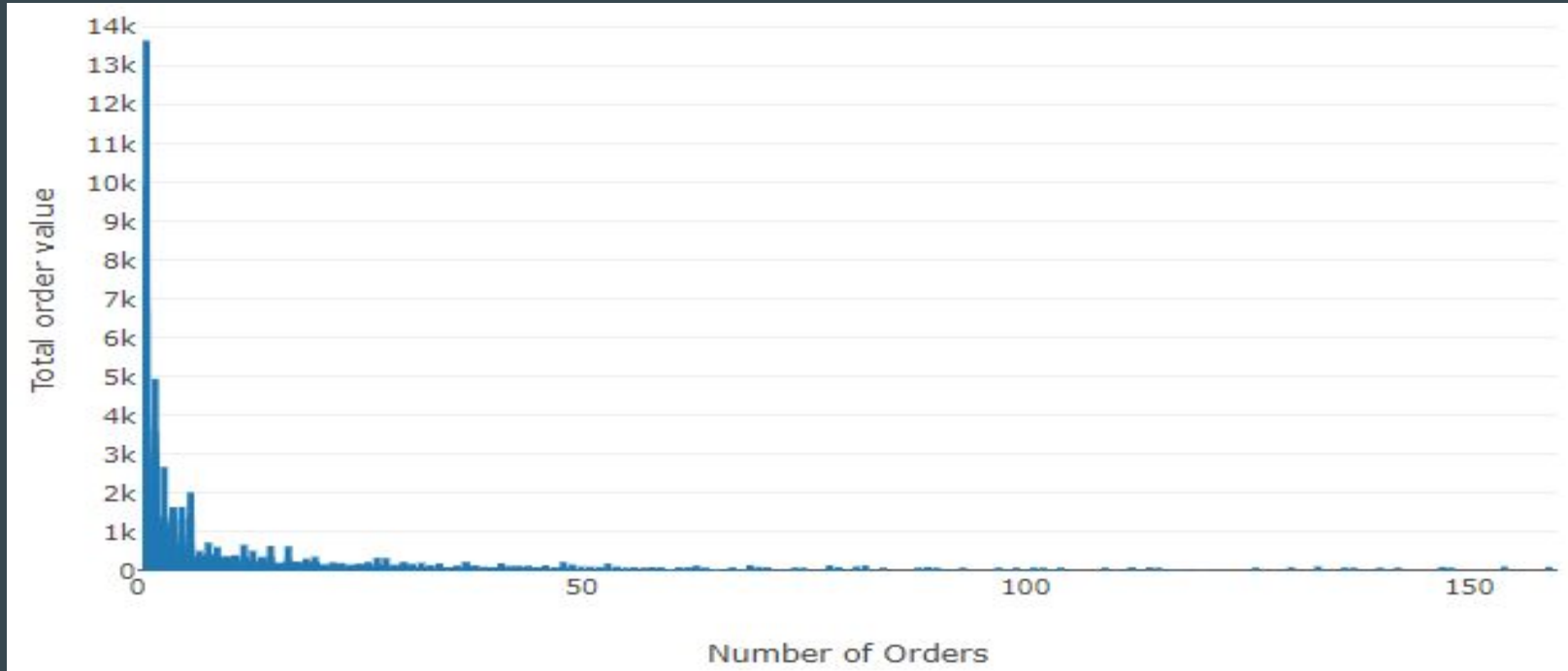
# Exploratory Data Analysis

Product and freight values by date



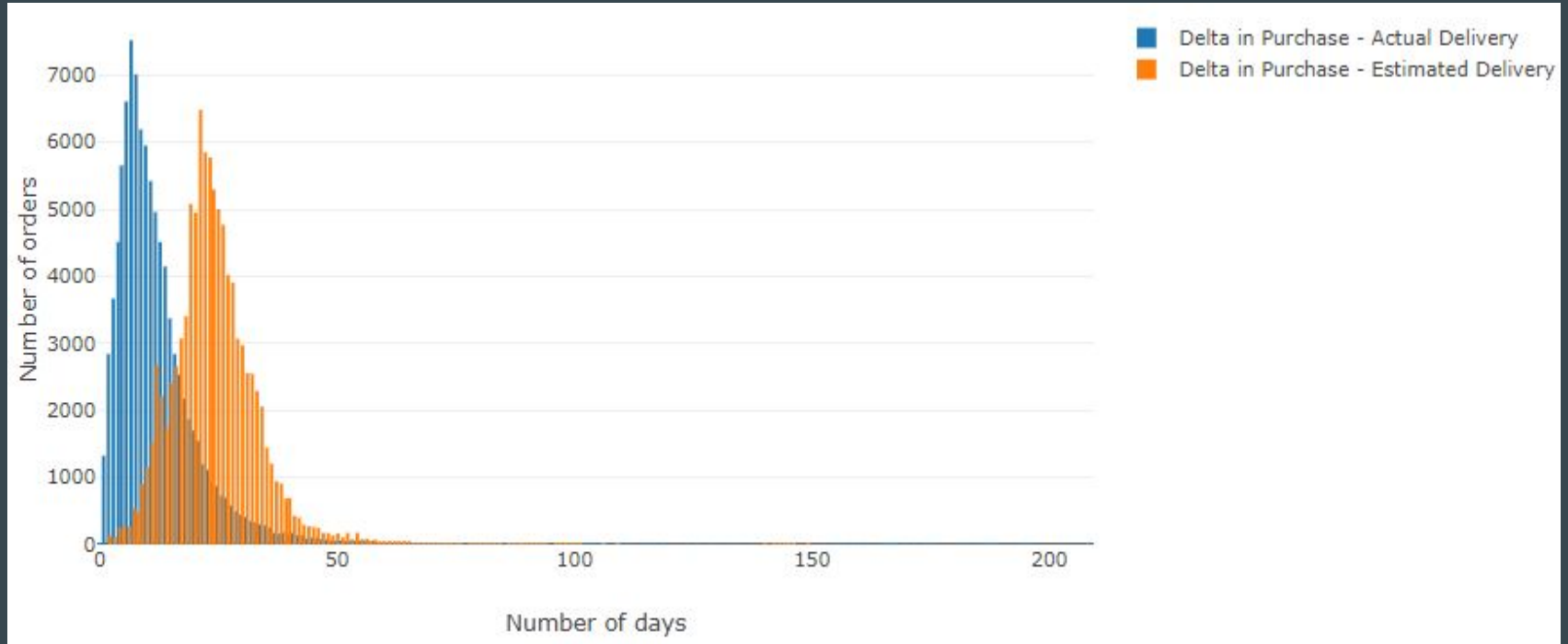
# Exploratory Data Analysis

Money spent on each order/ Transaction value per order



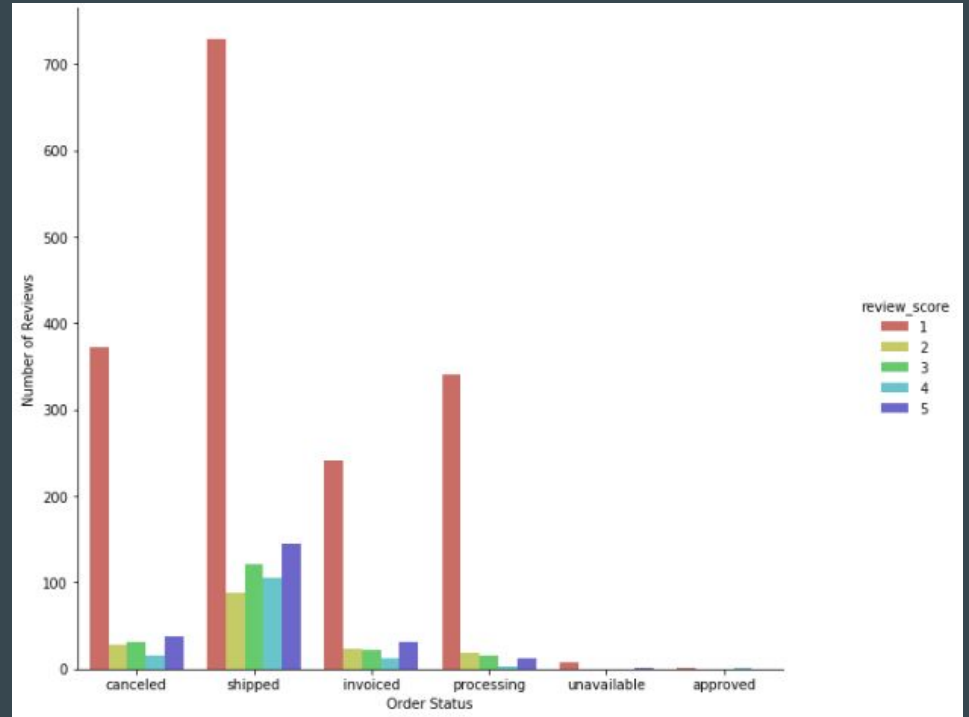
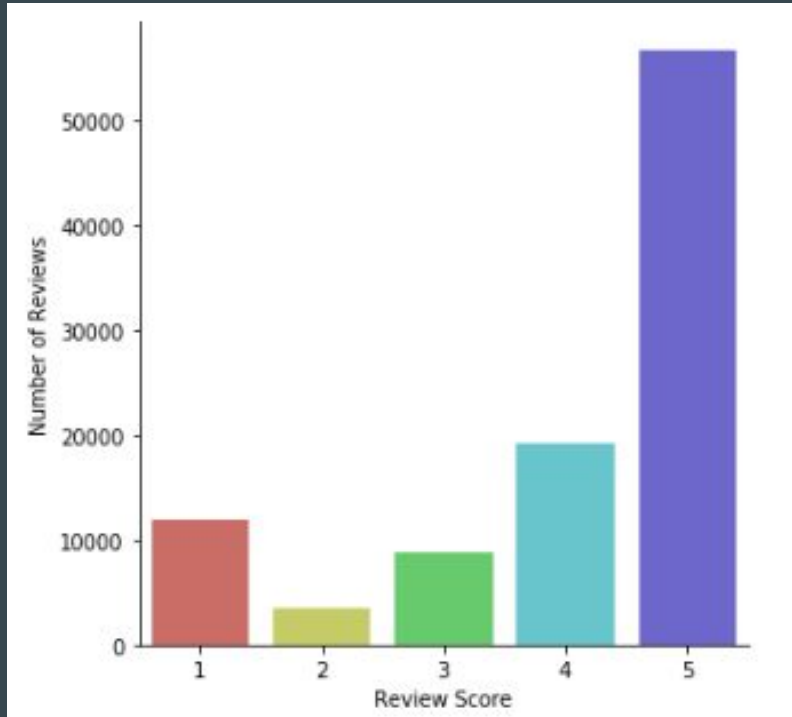
# Exploratory Data Analysis

Number of days from day of purchase till estimated and actual delivery date



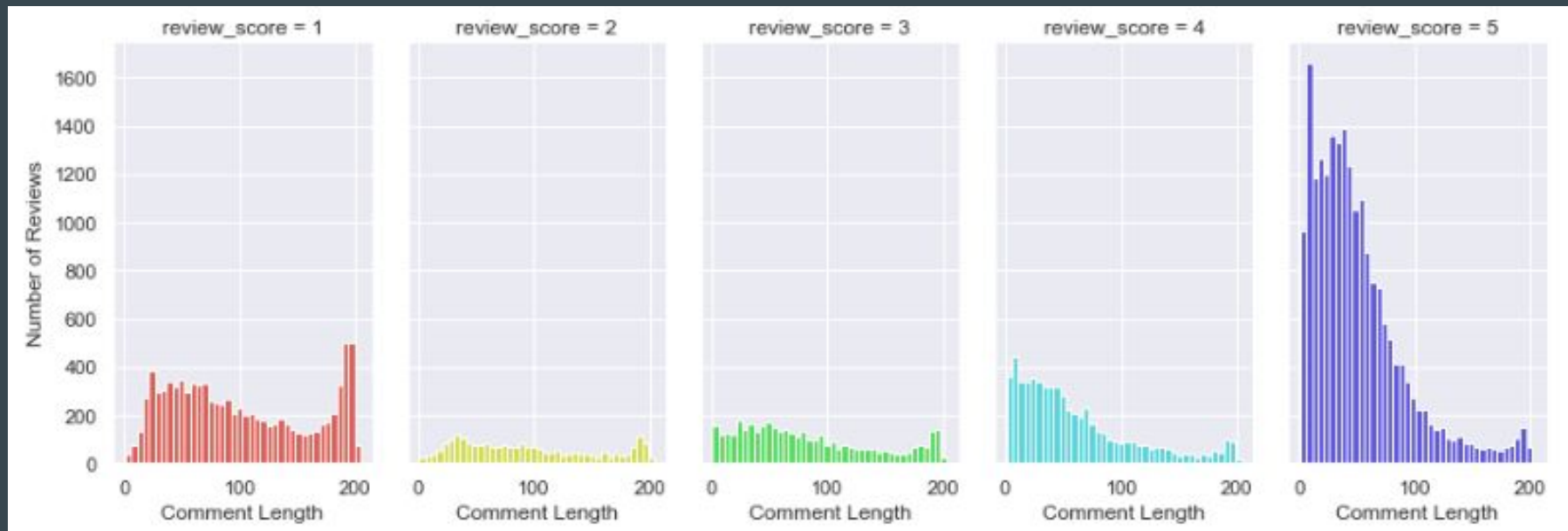
# Exploratory Data Analysis

Number of reviews per score and Review counts per order status



# Exploratory Data Analysis

Review length per review score



# Feature Hypothesis

order\_items\_qty

Increase in item quantity should increase review score

product\_description\_lenght

More information about buying product should have positive relationship with review score

product\_photos\_qty

More photos about buying product should have positive relationship with review score

product\_name\_lenght

Shorter names can be ambiguous or easy to understand,so relationship with review score is unknown

# Feature Hypothesis

delivery\_accuracy

item coming on time or earlier than it was described  
should have positive relationship with review score

order\_products\_value

more expensive perhaps means better quality or  
higher expectation towards order fulfilment

order\_freight\_value

more freight means more expectation towards delivery  
accuracy

To check above hypothesis, an OLS model was used on above mentioned features.

# Feature Engineering

It's clear that we have to use more informative features to model this problem so below new features were created:

1. Estimated Delivery Time in working days
2. Estimated Delivery Time in working days
3. Delivery Time Delta in working days
4. Delayed
5. Average Product Value
6. Order Freight Ratio
7. Purchase Day of Week



# Hyperparameter tuning

The data set had values in different ranges.

To avoid potential negative effect on the performance of ML algorithms:

- Normalized data using scikit-learn's Normalizer
- categorical columns were encoded using Labelencoder
- Data was split into train and test datasets

# Modeling

- Random Forest Classifier: 0.5526
- Logistic Regression: 0.5932
- GradientBoostingClassifier: 0.6074
- GradientBoostingClassifier with GSCV: 0.6088

GradientBoostingClassifier performed best with 60% accuracy.

# Conclusion

The customers usually are satisfied (score 5) or not satisfied(1) so the number of reviews with score 1 and 5 are comparatively more than 2,3 and 4.

It can be said that delivery accuracy is one of the most important factors in keeping customers happy.

# Limitations

Customer satisfaction can not be predicted solely based on objective facts without taking the vast amount of subjective sensor and service process data into account. Many subjective factors influence review rating such as :

- Incorrect descriptions of the product
- Poor client service
- Poor response to information request
- Lack of communication
- Rude or Uninformed customer care staff
- Needs not accurately defined
- Promises not carried out
- Repeated complaints from the same customer

# Further Analysis

Analyze Customer reviews using text mining to predict their behaviour