Final Project - Coder School

PROJECT PROPOSAL

WHY IS WATCHES_GIFTS A FASCINATING CATEGORY TO START YOUR BUSINESS ON OLIST?

Tran Viet Thang Long

olist

Overview

Olist is a company that provides an online e-commerce site aggregation platform designed to facilitate direct sales on large e-commerce sites. Its platform connects entrepreneurs with major online retailers and allows shopkeepers to advertise and sell in the marketplaces.

Sep 2016 - Oct 2018

Overview

Target Audience

New sellers looking for a category to start their business on Olist.

Metric

Category's Performance
Review Score

R$ 13.25M
total seller's revenue

71 categories

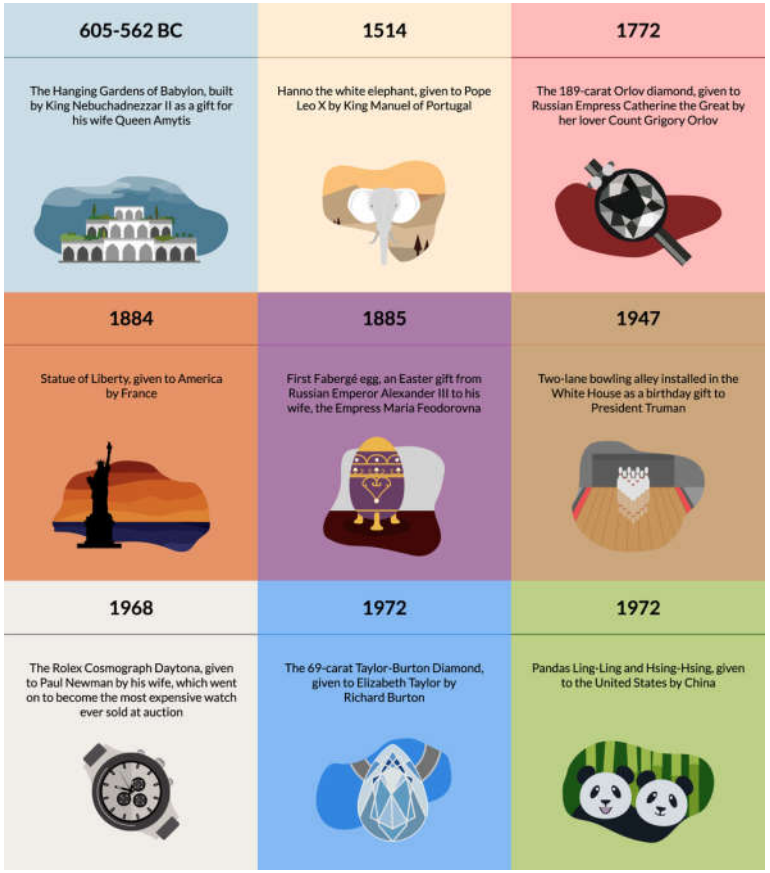
99441 orders

3095 sellers

Content

Analysis Flow

Research



Why do we give gifts?

As gifting plays such an important role in our social fabric, we give gifts for many, sometimes conflicting, reasons. At times our culture requires it, for example, Christmas or birthday presents. At other times, it builds and reinforces relationships with family members and potential mates, and can be done for a variety of reasons.

- **The Psychology of Gift Giving** -

Part 1:

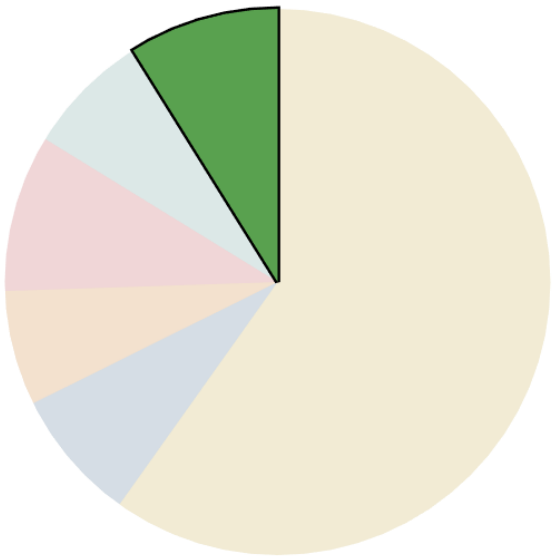
Performance

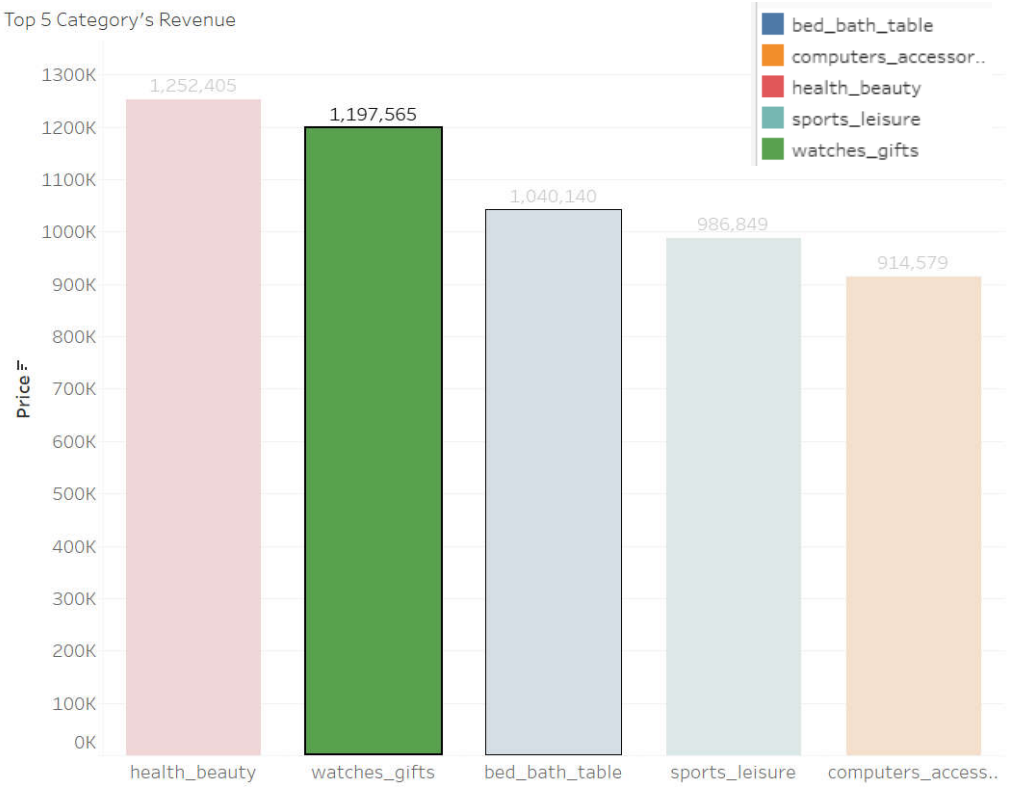
Category's Overall

Top 5 category cover 40.5% total seller's revenue

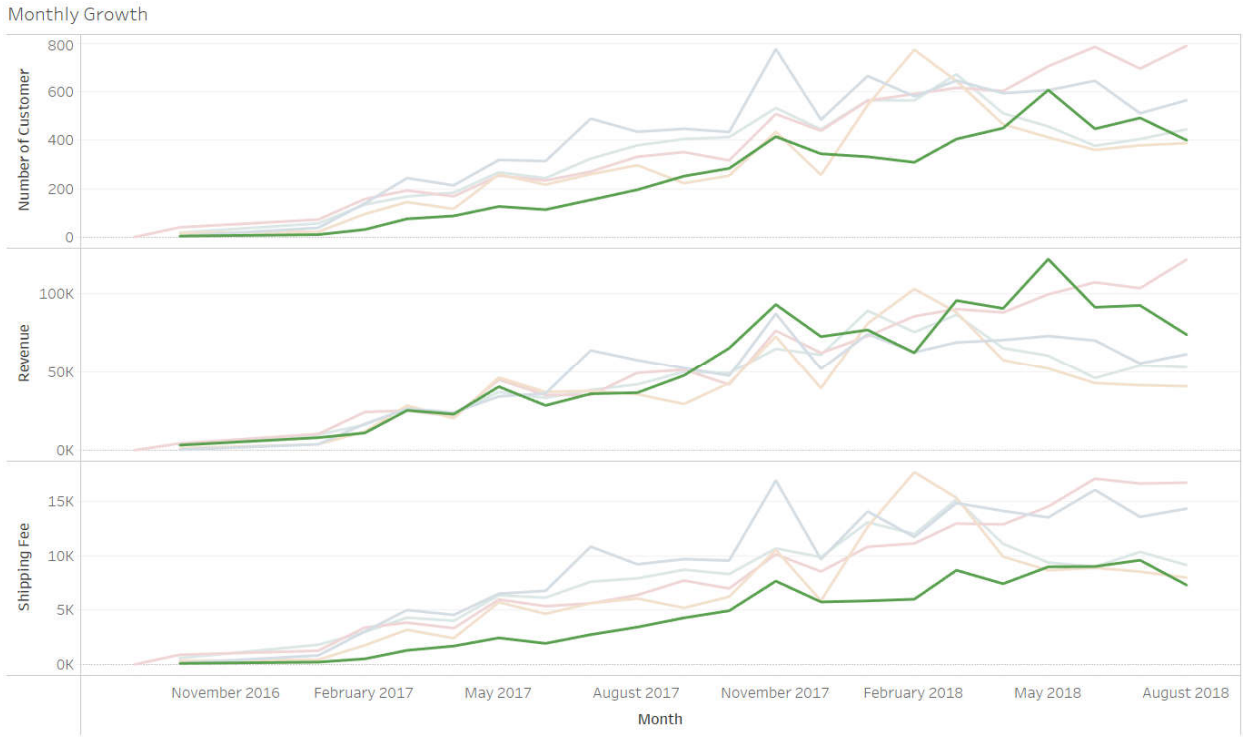
watches_gifts reaches R$ 1197K, top 2/71 categories

9% total seller's revenue





Part 1.1: Monthly Growth



Monthly Number of Unique Customer Growth Rate: **33%**

Monthly Revenue Growth Rate: **33.1%**

Lowest Shipping Fee Growth in the group

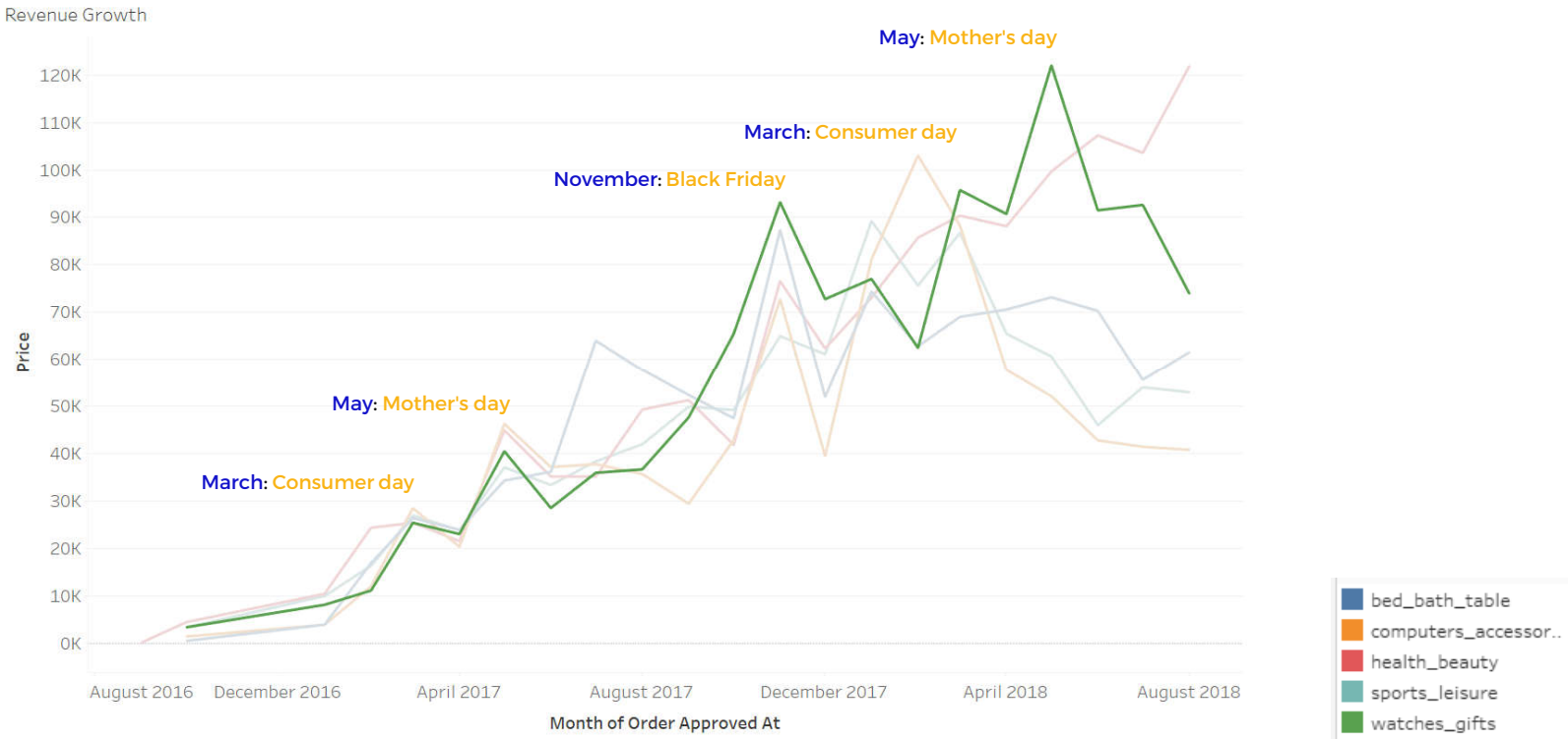
Performance

Part 1.1: Monthly Growth

In Brazil, 58 million consumers made at least one virtual purchase in 2018, representing 27 percent of the country's population and an increase of 6 percent compared to 2017 - making eCommerce a viable sales channel worth exploring.

- **Brazil eCommerce** (The International Trade Administration (ITA), U.S. Department of Commerce) -

Part 1.1: Monthly Growth



3 peak revenue months: **March**, **May**, and **November**

Performance

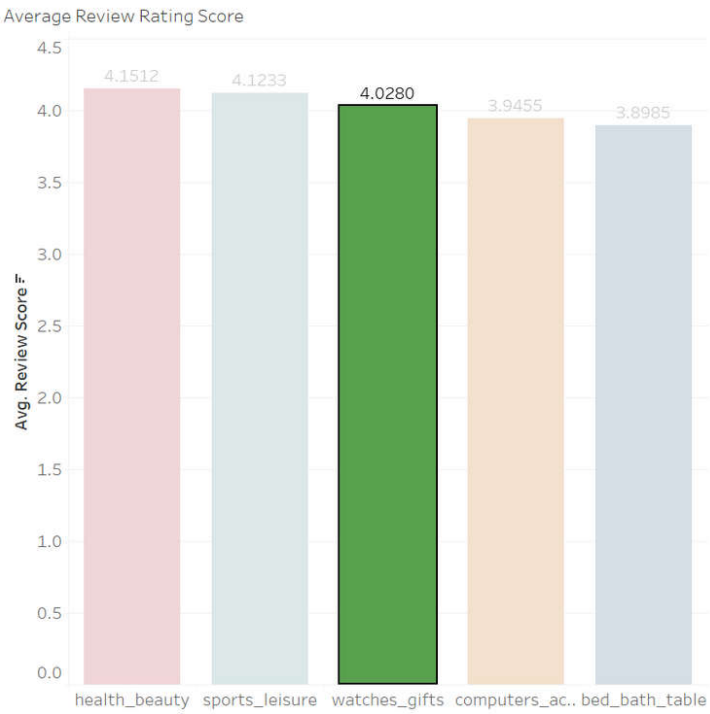
Part 1.2: Seller

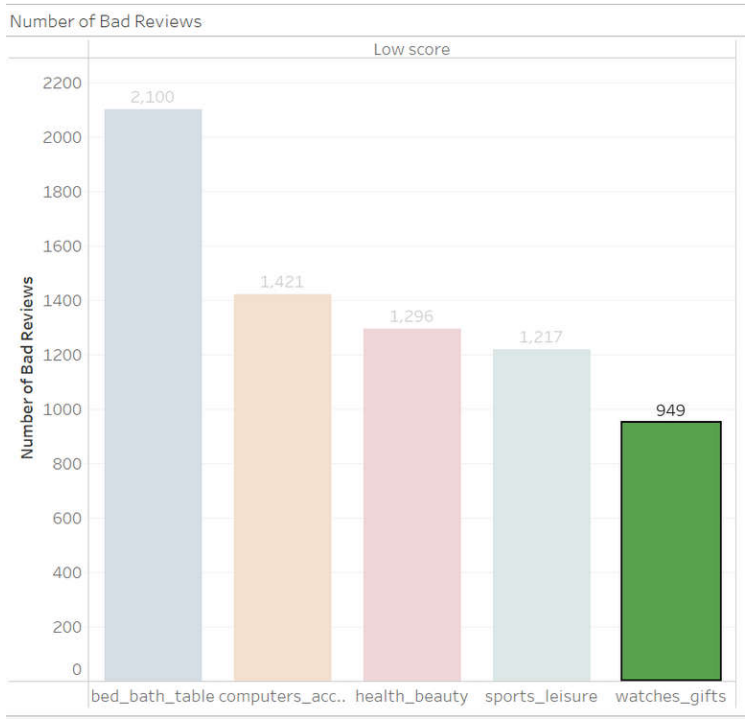
Higher Revenue per seller:
at least **127%**

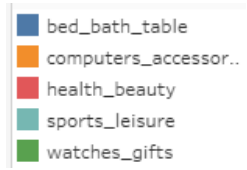
Lower Number of Sellers:
at least **49.5%**

Performance

Part 1.3: Review







Lowest Number of Bad Reviews

37%

lower than the other 4's average score.

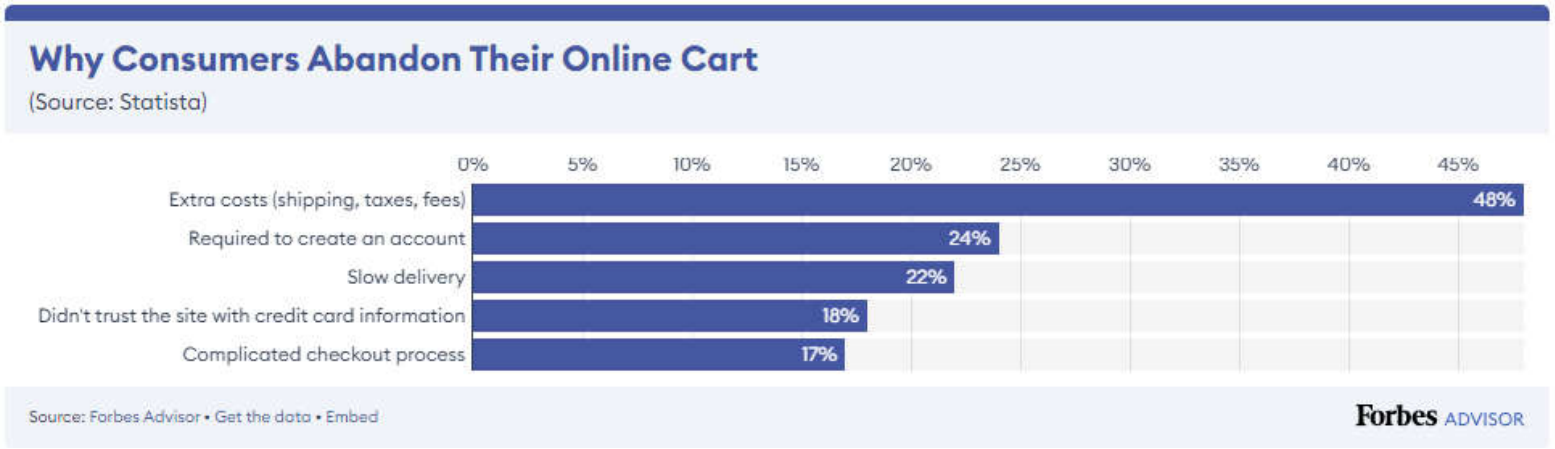
Performance

Part 2:

Shipping

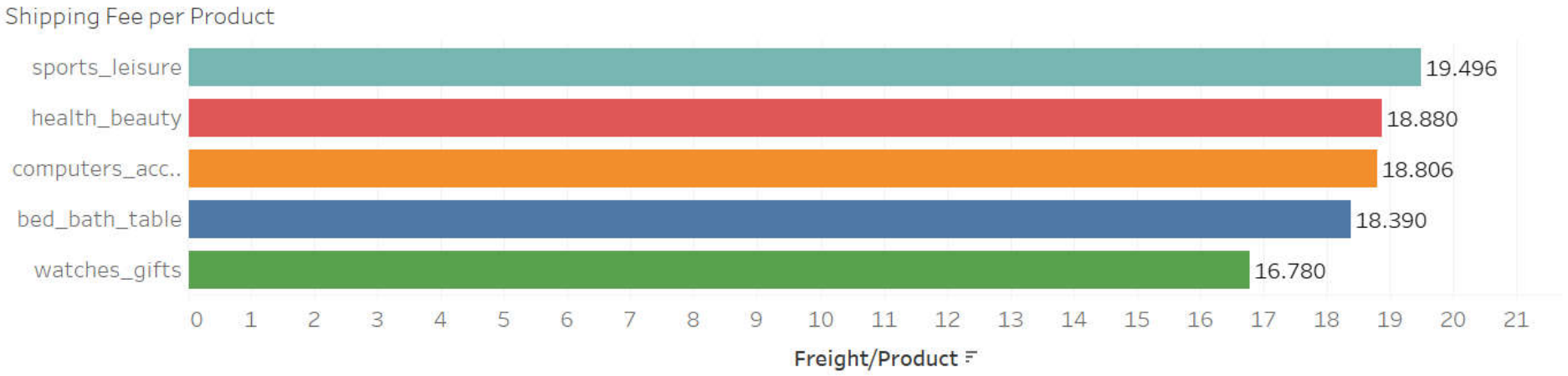
Part 2.1: Shipping Fee

Shipping



Part 2.1: Shipping Fee

Shipping



The average shipping fee per product is **11.3%** lower than the other 4's average

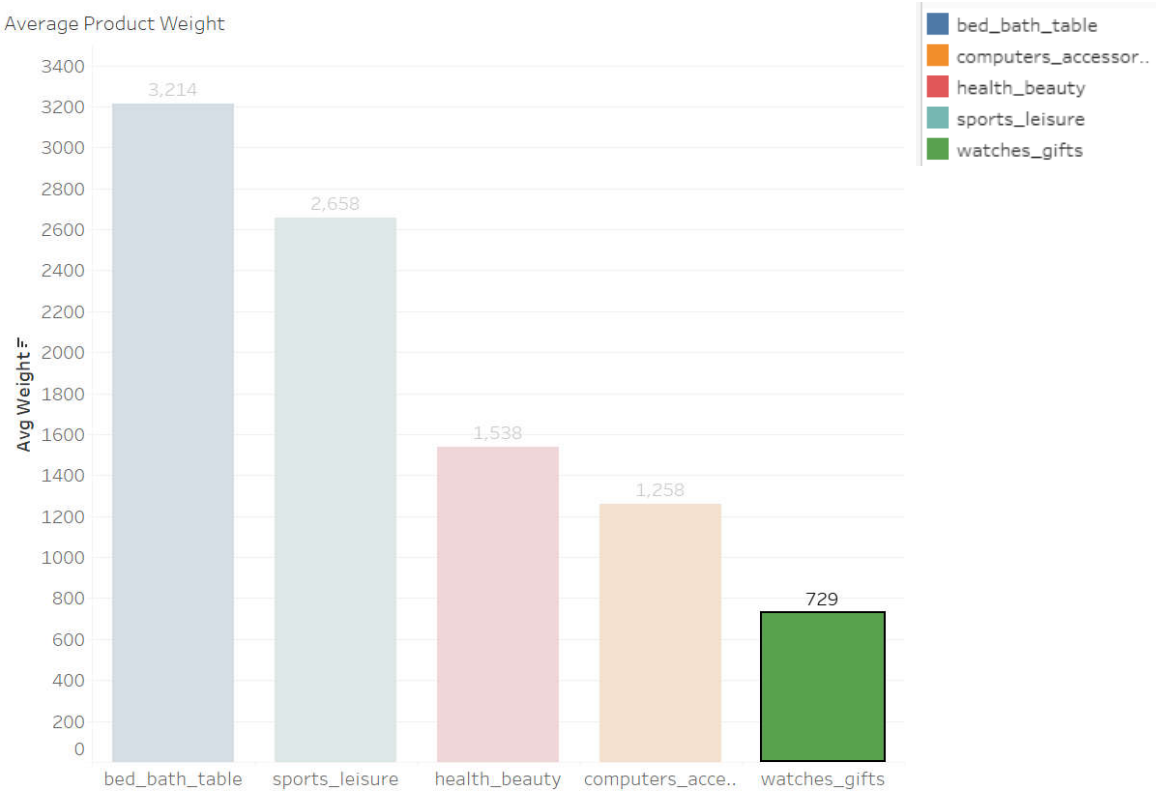
To get the price, measure the weight and dimensions of your package. The shipping charge is calculated based on these measurements and the customer's location.

- Shipping Cost Calculation: Determine National & International **Shipping Rates** (Sku Vault) -

Part 2.1: Shipping Fee

Shipping

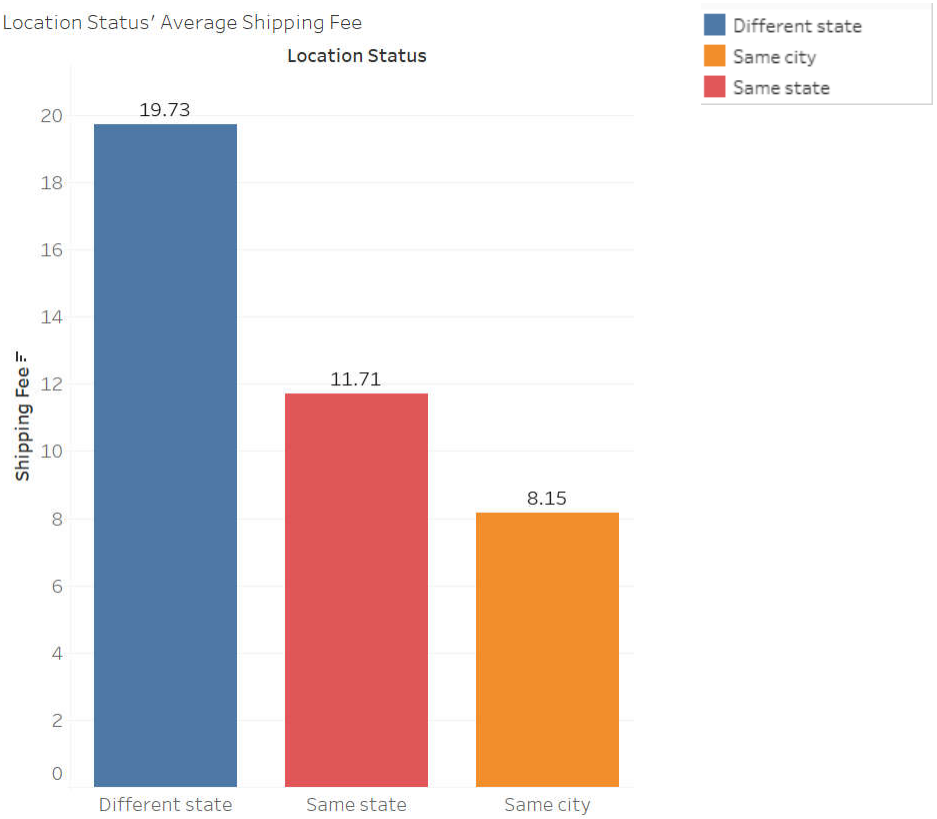
The average product's weight is at least lower than other **42%** categories in the top 5



Part 2.1: Shipping Fee

Shipping

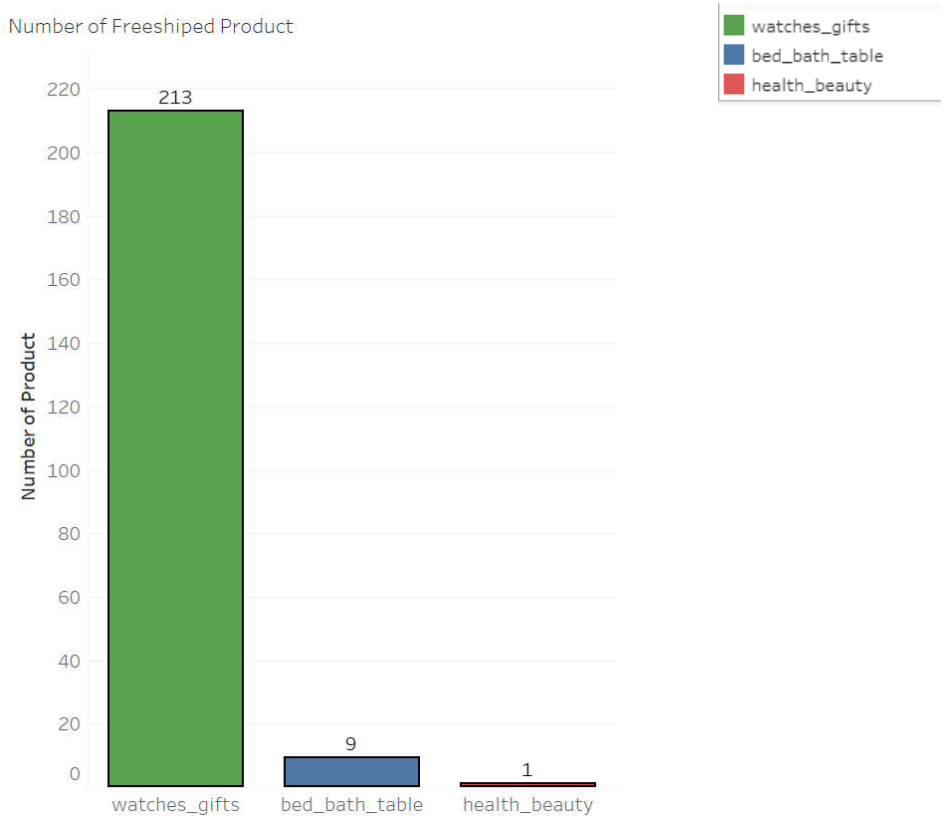
Differences between the seller and customer's state increase shipping fees **68.5%** compared to customer and seller in the same state and **142%** compare to customer and seller in the same city



Part 2.1: Shipping Fee

Shipping

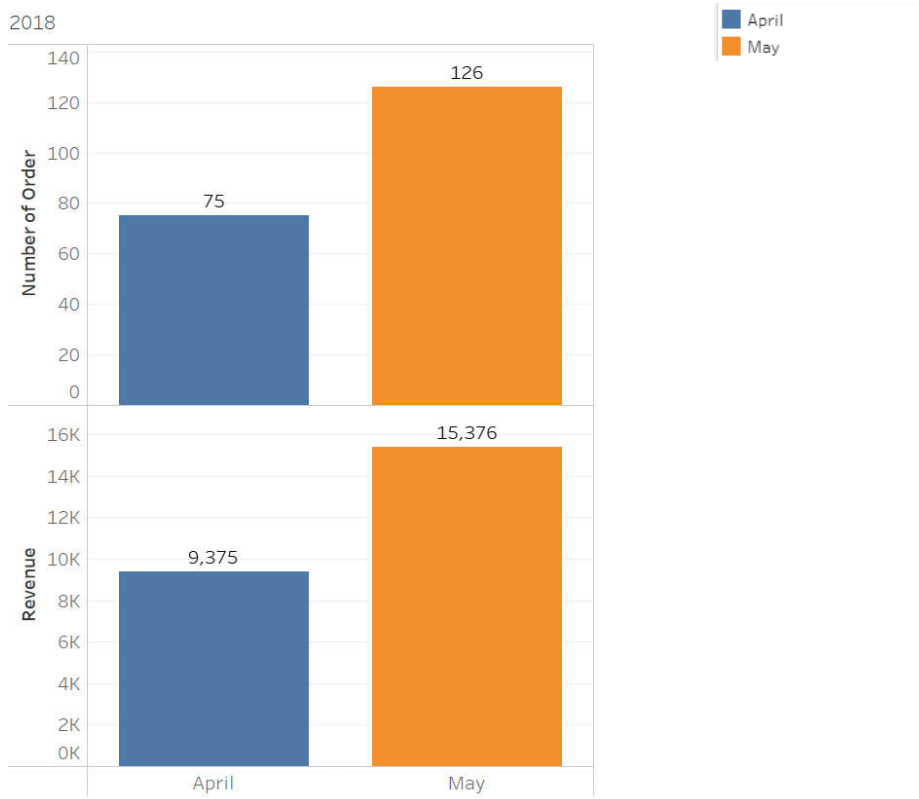
Only 2 watches_gifts' sellers properly apply freeship to 2 products in April, and May 2018



Part 2.1: Shipping Fee

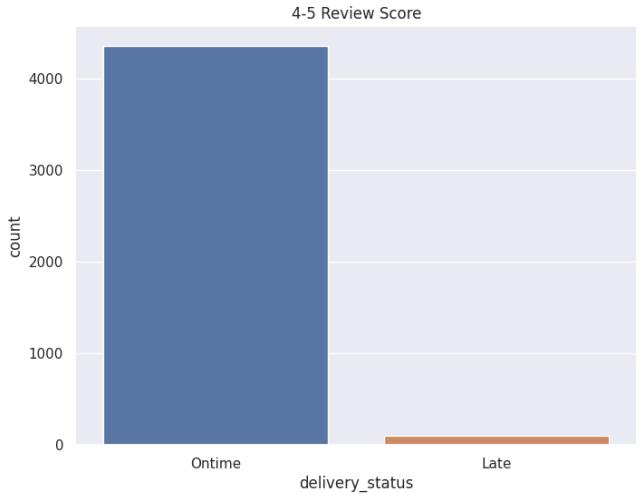
Shipping

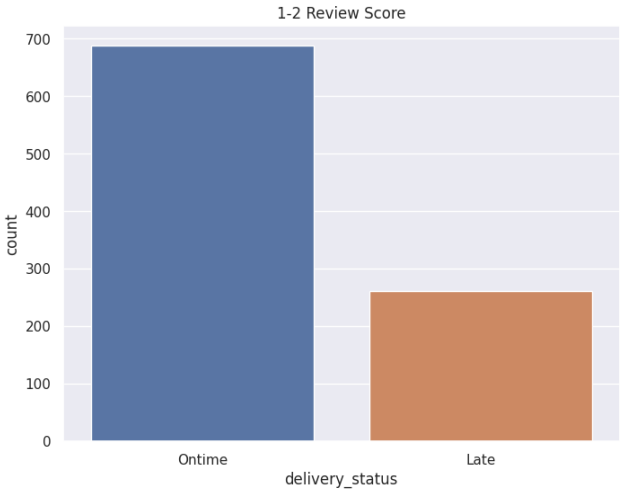
Contribute **20.6%** to the Number of Products and **12.6%** to Revenue in All-time Peak Revenue month: May 2018



Part 2.2: Late Shipping

Shipping



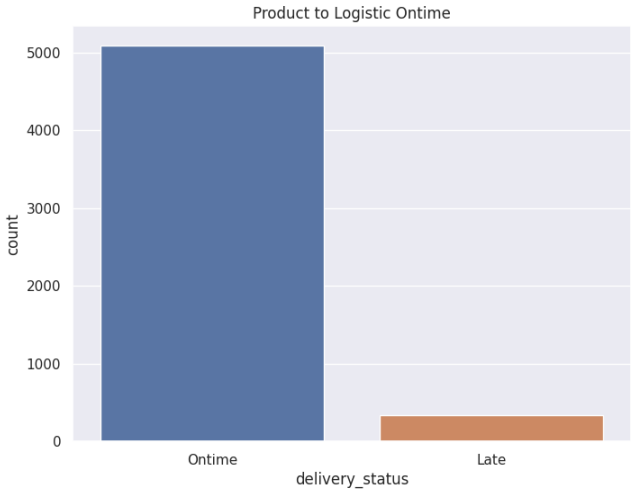


2.11% Late Shipping Rate

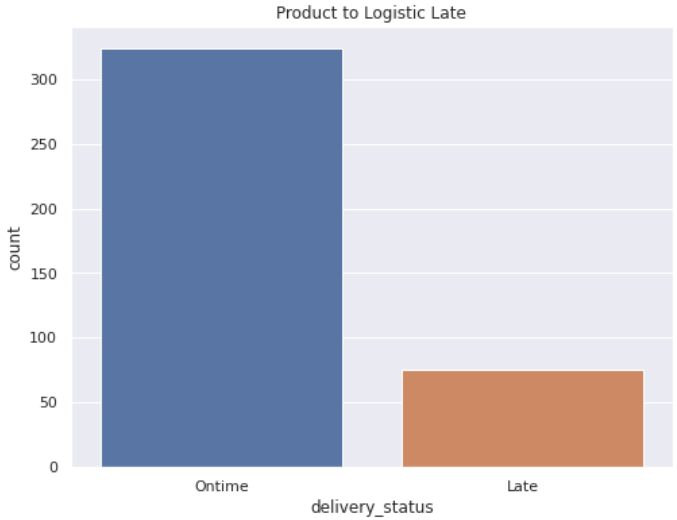
27.5% Late Shipping Rate

The late shipping rate in the 1-2 Review Score group is **13** times higher than the late shipping rate in the 4-5 Review Score

Part 2.2: Late Shipping



6.2% Late Shipping Rate

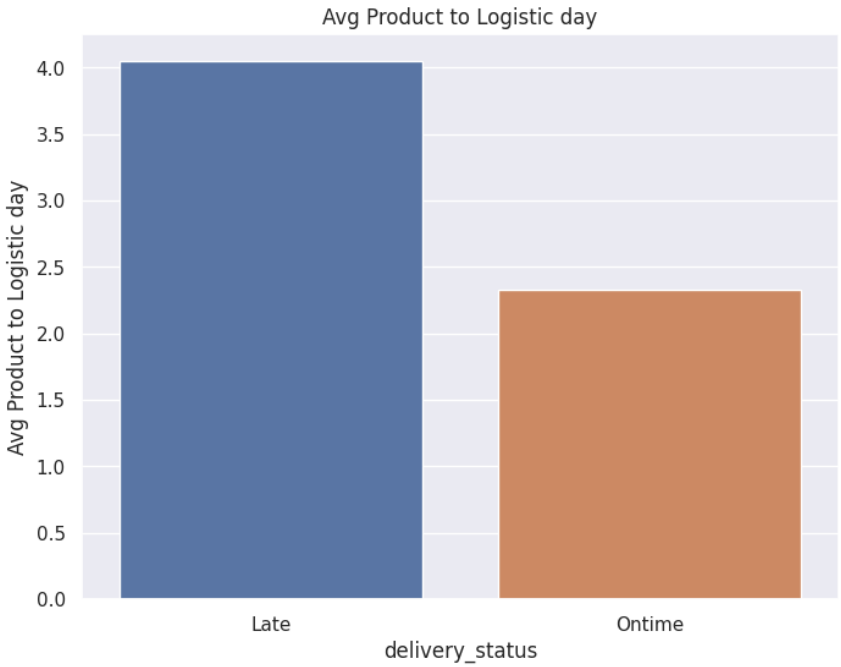


18.8% Late Shipping Rate

The late shipping rate when the seller delivers products to the logistic partner late is **3 times** higher than the late shipping rate when the seller delivers products on time

Part 2.2: Late Shipping

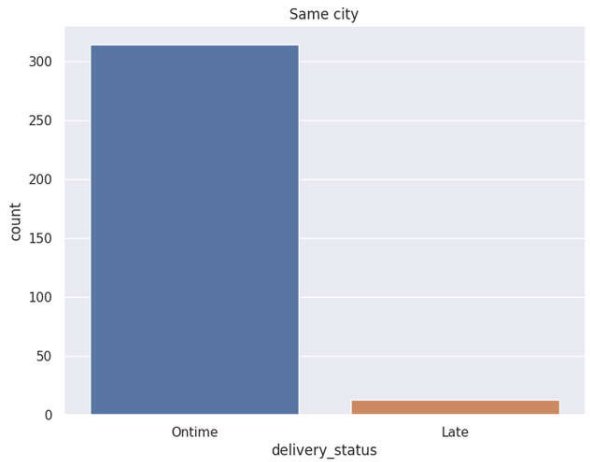
Shipping



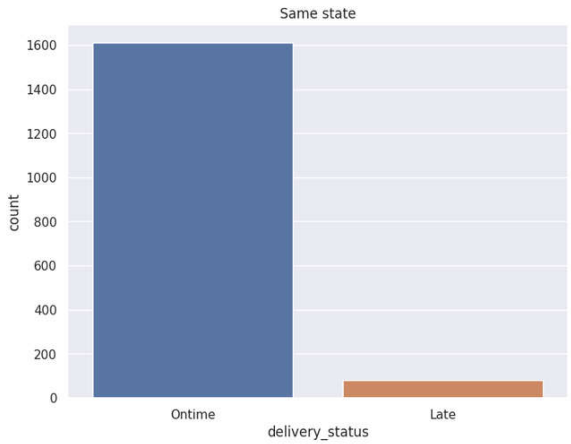
Late orders need **4.05 days** on average to deliver products to the logistic partner when Ontime orders only need **2.33 days**

Part 2.2: Late Shipping

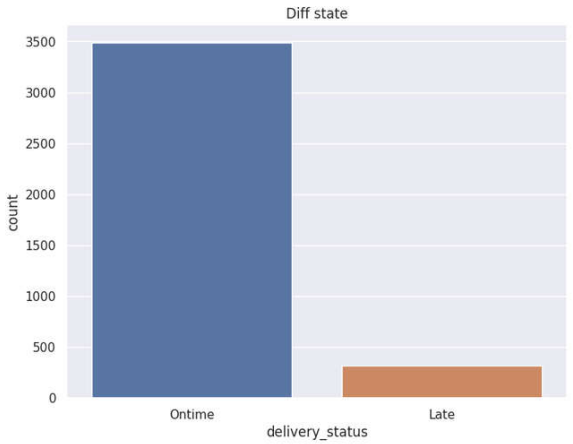
Shipping



4% Late Shipping Rate



4.8% Late Shipping Rate



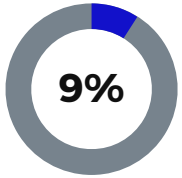
8.35% Late Shipping Rate

The late shipping rate when the seller is in a different state from the customer is **~2 times** higher than the late shipping rate when the seller is in the same city or state as the customer

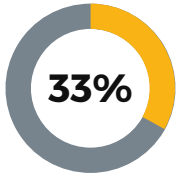
Part 3:

Summary & Recommendation

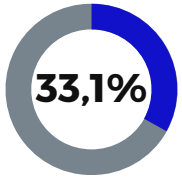
Part 3.1: Summary



Total Seller's Revenue

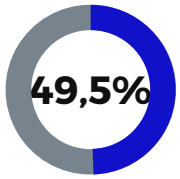


Monthly Number of Unique Customer Growth Rate

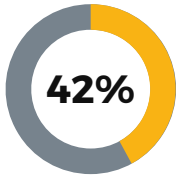


Monthly Revenue Growth Rate

127% **Higher in Revenue per Seller**



Lower in Number of Sellers



Lower in Average Product's Weight

Summary & Recommendation

Part 3.2: Recommendation

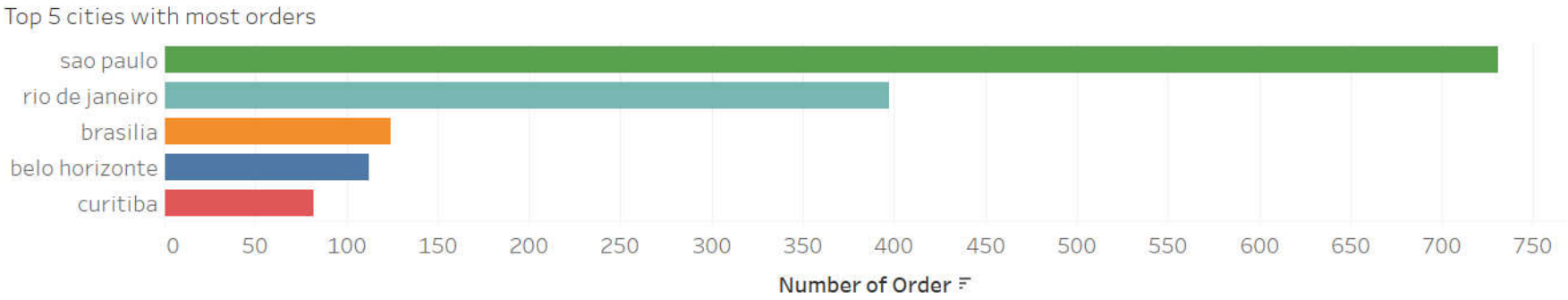
Build **a Promotion Campaign** and apply **Freeship** on products in **Peak Revenue Month** and **Pre - Peak Revenue Month**

March: Consumer Day
May: Mother's Day
November: Black Friday

Summary & Recommendation

Part 3.2: Recommendation

Set up **storage** in **hot cities** to minimize **Late Shipping Rates** and **Shipping Fees** for customer

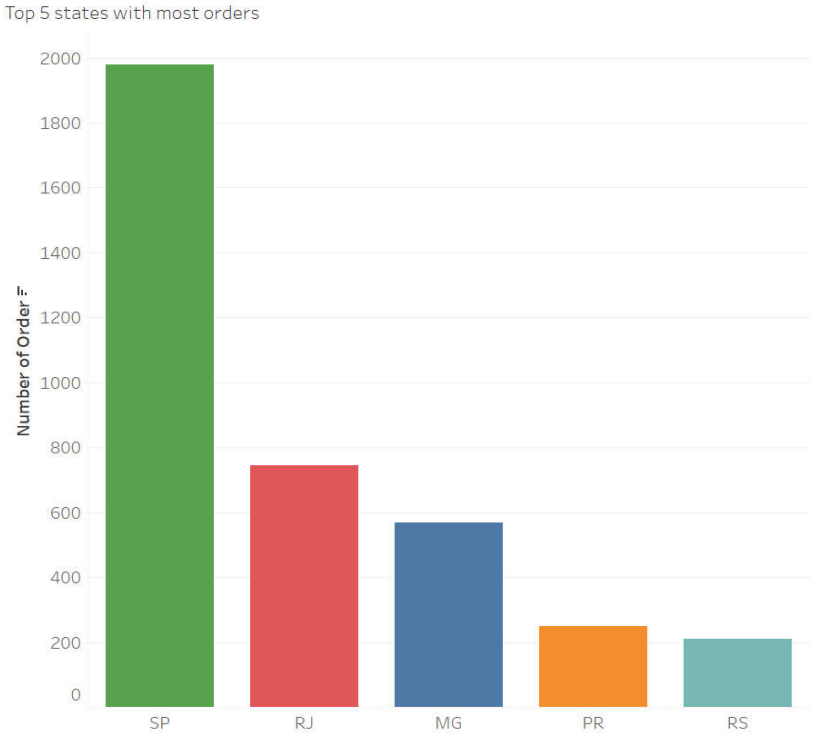


1. Sao Paulo
2. Rio de Janeiro
3. Brasillia
4. Belo Horizonte
5. Curitiba

Summary & Recommendation

Part 3.2: Recommmendation

Or **hot states** to minimize **Late Shipping Rates** and **Shipping Fees** for customer



SP: Sao Paulo

RJ: Rio de Janeiro

MG: Minas Gerais

PR: Paraná

RS: Rio Grande do Sul

Summary & Recommendation

Ideal delivery time to the Logistic Partner:

2.33 days

to avoid late delivery

Summary & Recommendation

Part 4:

Machine Learning Report

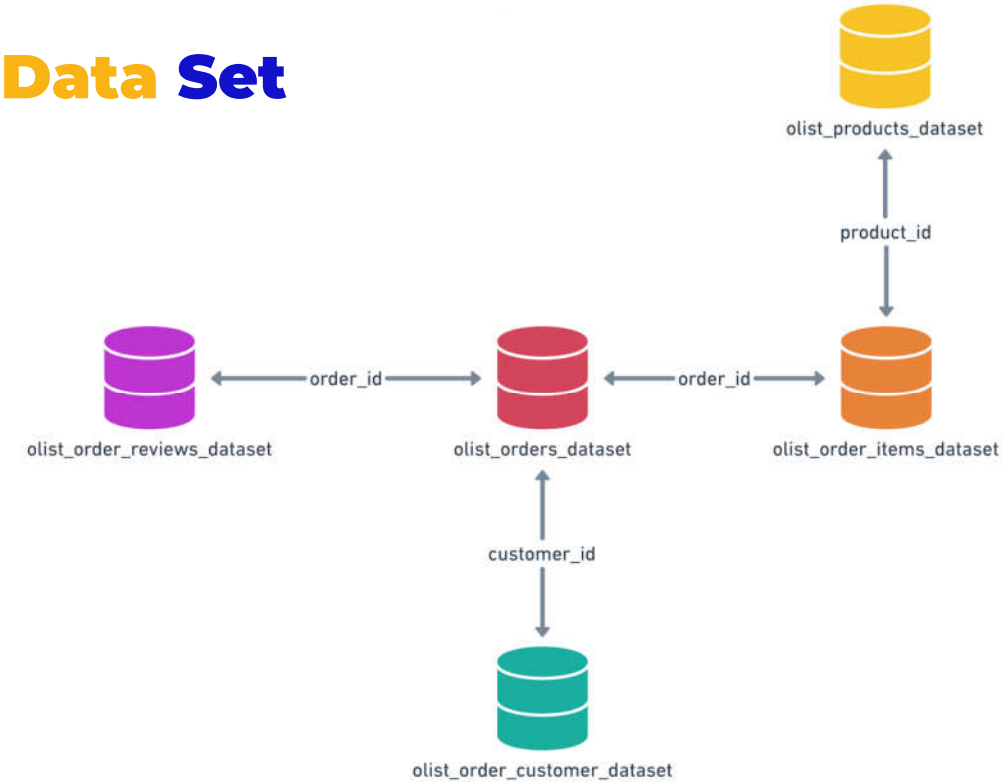
Part 4.1: Preprocessing

Machine Learning Report

Predict Review Rating

Metric: f1 - score

Data Set



Part 4.1: Preprocessing

Machine Learning Report

Convert feature columns into labeled 0/1 columns and categorical columns

Drop Null values

Columns list: price, freight_value, delivery_status, rate, product_category_name, logisstatus, locationstatus, product_weight_g, fee_status, order_status, volume, seller_state, customer_state

Machine Learning Report

Train test split: ratio 8/2
X: All columns except rate
y: rate column values

rate column: 0 if review score <= 3
1 if review score > 3

Feature Engineering:
MinMaxScaler: Convert Numerical columns
LabelEncoder: Convert Categorical columns

Model training: **Logistic Regression**

Report 1

Train

	precision	recall	f1-score	support
0	0.75	0.25	0.37	21313
1	0.80	0.97	0.88	66870
accuracy			0.80	88183
macro avg	0.78	0.61	0.63	88183
weighted avg	0.79	0.80	0.76	88183

Test

	precision	recall	f1-score	support
0	0.78	0.10	0.17	5290
1	0.78	0.99	0.87	16756
accuracy			0.78	22046
macro avg	0.78	0.54	0.52	22046
weighted avg	0.78	0.78	0.70	22046

Model training: **Logistic Regression**

Report 2: Oversampling with SMOTE

Train

	precision	recall	f1-score	support
0	0.84	0.27	0.41	53496
1	0.62	0.96	0.75	66870
accuracy			0.65	120366
macro avg	0.73	0.61	0.58	120366
weighted avg	0.72	0.65	0.60	120366

Test

	precision	recall	f1-score	support
0	0.76	0.24	0.36	5290
1	0.80	0.98	0.88	16756
accuracy			0.80	22046
macro avg	0.78	0.61	0.62	22046
weighted avg	0.79	0.80	0.76	22046

Machine Learning Report

Model training: **Logistic Regression**

Report 3: Undersampling with NearMiss

Train

	precision	recall	f1-score	support
0	0.77	0.62	0.69	21313
1	0.68	0.82	0.74	21313
accuracy			0.72	42626
macro avg	0.73	0.72	0.72	42626
weighted avg	0.73	0.72	0.72	42626

Test

	precision	recall	f1-score	support
0	0.59	0.22	0.32	5290
1	0.79	0.95	0.87	16756
accuracy			0.78	22046
macro avg	0.69	0.59	0.59	22046
weighted avg	0.75	0.78	0.74	22046

Model training: **Random Forest Classifier**

Report 4

Train

	precision	recall	f1-score	support
0	0.88	0.30	0.44	53496
1	0.63	0.97	0.76	66870
accuracy			0.67	120366
macro avg	0.75	0.63	0.60	120366
weighted avg	0.74	0.67	0.62	120366

Test

	precision	recall	f1-score	support
0	0.62	0.29	0.40	5290
1	0.81	0.94	0.87	16756
accuracy			0.79	22046
macro avg	0.72	0.62	0.63	22046
weighted avg	0.76	0.79	0.76	22046

Part 4.3: Summary

Machine Learning Report

Most accurate Logistic Regression report:
Report 2: Oversampling with SMOTE

Most accurate model:
Random Forest Classifier

Machine Learning Report

Preprocessing Step by Step for better accuracy with Logistic Regression model:

1. Convert feature columns to 0/1 label columns and categorical columns

2. Drop Null Value

3. Train Test Split: ratio 8/2

4. Feature Engineering
MinMaxScale
LabelEncoder

5. Oversampling using SMOTE

Final Project - CoderSchool

THANK YOU

Tran Viet Thang Long