# Problem Statement

Olist is a Brazilian e-commerce platform that connects small and medium-sized businesses to customers across Brazil. The platform operates as a marketplace, where merchants can list their products and services and customers can browse and purchase them online.

The Olist sales dataset is a collection of anonymized data about orders placed on the Olist from September 2016 to September 2018. It contains a wide range of information about each order, including the order date, product details, payment and shipping information, customer and seller IDs, and customer reviews. The dataset also includes information about the sellers who list their products on Olist, as well as data on customer behavior and demographics. The dataset is designed to help analysts and researchers better understand the e-commerce landscape in Brazil and identify opportunities for growth and optimization.

1. Revenue and Profit Analysis:

What are the monthly and yearly revenue and profit trends?

Which product categories contribute the most to overall profit?

What are the profit margins for different product categories and sub-categories?

How do seasonal sales patterns affect revenue and profit?

What are the top-performing products in terms of revenue and profit?

1. Customer Analysis:

Who are the top customers by revenue and profit?

What is the average order value and frequency for different customer segments?

How does customer lifetime value vary across different segments?

What factors drive higher customer lifetime value?

How do discounts and promotions impact customer purchasing behavior?

1. Inventory Management:

What are the inventory turnover rates for different product categories?

How does inventory level correlate with salles performance?

Identify slow-moving and fast-moving products.

How effective are current inventory management practices?

What recommendations can be made to optimize inventory levels?

1. Sales Channel Performance:

How do different sales channels (e.g., online, in-store) compare in terms of revenue and profit?

Which channels have the highest customer acquisition and retention rates?

How does marketing spend impact salles across different channels?

What is the ROI of marketing campaigns for each channel?

How do customer preferences differ across sales channels?

Data Dictionary

1/olist_customers_dataset.csv

```
customer_id: unique identifier for each customer

customer_unique_id: unique identifier for each customer (anonymized)

customer_zip_code_prefix: zip code prefix of the customer's address

customer_city: city where the customer is located

customer_state: state where the customer is located
```

2/ olist_geolocation_dataset.csv

```
geolocation_zip_code_prefix: zip code prefix for the location

geolocation_lat: latitude of the location

geolocation_lng: longitude of the location

geolocation_city: city of the location

geolocation_state: state of the location
```

3/ olist_orders_dataset.csv

```
order_id: unique identifier for each order

customer_id: unique identifier for the customer who placed the order

order_status: current status of the order (e.g. delivered, shipped,
canceled)

order_purchase_timestamp: date and time when the order was placed

order_approved_at: date and time when the payment for the order was
approved

order_delivered_carrier_date: date and time when the order was handed
over to the carrier
```

```
order_delivered_customer_date: date and time when the order was delivered to the customer

order_estimated_delivery_date: estimated date when the order is expected to be delivered
```

4/ olist_order_items_dataset.csv

```
order_id: unique identifier for the order

order_item_id: unique identifier for each item within an order

product_id: unique identifier for the product being ordered

seller_id: unique identifier for the seller who listed the

product shipping_limit_date: date and time when the seller has to ship the product

price: price of the product

freight_value: shipping fee for the product
```

5/ olist_order_payments_dataset.csv

```
order_id: unique identifier for the order

payment_sequential: index number for each payment made for an order

payment_type: type of payment used for the order (e.g. credit card, debit card, voucher)

payment_installments: number of installments in which the payment was made

payment_value: value of the payment made
```

6/ olist_products_dataset.csv

```
product_id: unique identifier for each product

product_category_name: name of the category that the product belongs to

product_name_lenght: number of characters in the product name

product_description_lenght: number of characters in the product description

product_photos_qty: number of photos for the product
```

```
product_weight_g: weight of the product in grams

product_length_cm: length of the product in centimeters

product_height_cm: height of the product in centimeters

product_width_cm: width of the product in centimeters
```

7/ olist_sellers_dataset.csv

```
seller_id: unique identifier for each seller

seller_zip_code_prefix: zip code prefix for the seller's location

seller_city: city where the seller is located

seller_state: state where the seller is located
```

8/ product_category_name_translation.csv

```
product_category_name: name of the product category in Portuguese

product_category_name_english: name of the product category in English
```

9/ olist_order_reviews_dataset.csv

```
review_id: unique identifier for each review

order_id: unique identifier for the order that the review is associated with

review_score: numerical score (1-5) given by the customer for the product

review_comment_title: title of the review comment

review_comment_message: text of the review comment

review_creation_date: date and time when the review was created

review_answer_timestamp: date and time when the seller responded to the review (if applicable)
```

Note: The review comment fields (i.e. review_comment_title and review_comment_message) are optional, and may not be present in all reviews.

```
# read all csv file from the data csv file by pandas

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#
```

# Read data and check data

```
customers_dataset = pd.read_csv("olist_customers_dataset.csv")
geolocation_dataset = pd.read_csv('olist_geolocation_dataset.csv')
order_items_dataset = pd.read_csv('olist_order_items_dataset.csv')
order_reviews_dataset = pd.read_csv('olist_order_reviews_dataset.csv')
order_dateset = pd.read_csv('olist_orders_dataset.csv')
products_dataset = pd.read_csv('olist_products_dataset.csv')
sellers_dataset = pd.read_csv('olist_sellers_dataset.csv')
product_category_name_translation =
pd.read_csv('product_category_name_translation.csv')
order_payments_dataset =
pd.read_csv('olist_order_payments_dataset.csv')

order_payments_dataset.head()
```


```
{"type":"dataframe","variable_name":"order_payments_dataset"}
```


```
customers_dataset.head()
```


```
{"summary":"{\n  \"name\": \"customers_dataset\",\n  \"rows\": 99441,\
n  \"fields\": [\n    {\n      \"column\": \"customer_id\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 99441,\n        \"samples\": [\n
\"c7432c6d237ffd6aa36a007b4237ec38\",\n
\"7f399d641e2e2064470145178c9e8778\",\n
\"ba5642b730704dc0f74b7cf715b41ed5\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"customer_unique_id\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 96096,\n        \"samples\": [\n
\"b5d6fa3d2213927296ac893f14f4461c\",\n
\"0520a11c7af8a73b703f1d2e722c7c8a\",\n
\"7a19f3fff09616cbb1cf8fdaa05ae032\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"customer_zip_code_prefix\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
29797,\n        \"min\": 1003,\n        \"max\": 99990,\n
\"num_unique_values\": 14994,\n        \"samples\": [\n
4343,\n          96820,\n          8598\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

```
n    },\n    {\n      \"column\": \"customer_city\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 4119,\n        \"samples\": [\n          \"luis
antonio\",\n          \"canitar\",\n          \"lagoinha do piaui\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"customer_state\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 27,\n        \"samples\": [\n          \"ES\",\
n          \"DF\",\n          \"BA\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\
n}","type":"dataframe","variable_name":"customers_dataset"}
```


```
geolocation_dataset.head()
```


```
{"type":"dataframe","variable_name":"geolocation_dataset"}
```


```
order_items_dataset.head()
```


```
{"type":"dataframe","variable_name":"order_items_dataset"}
```


```
order_reviews_dataset.head()
```


```
{"summary":"{\n  \"name\": \"order_reviews_dataset\",\n  \"rows\":
99224,\n  \"fields\": [\n    {\n      \"column\": \"review_id\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 98410,\n        \"samples\": [\n
\"ce4db7bc68bf7fa8d618d214203b15a3\",\n
\"7c93bc469d3617c77df6aed62ee09477\",\n
\"f39646c4059692442fe5734784cd9d50\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"order_id\",\n      \"properties\":
{\n        \"dtype\": \"string\",\n        \"num_unique_values\":
98673,\n        \"samples\": [\n
\"bbfb6afc6b09bb994e0093a008ec7675\",\n
\"182e6edb8cb190cada2690d7c7d78e76\",\n
\"99a3c8e51539b3ccb221a27cc29367f2\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"review_score\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
1,\n        \"min\": 1,\n        \"max\": 5,\n
\"num_unique_values\": 5,\n        \"samples\": [\n          5,\n
2,\n          1\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"review_comment_title\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 4527,\n
\"samples\": [\n          \"O PRODUTO N\\u00c3O CHEGOU .\",\n
\"N\\u00e3o recebi retorno\",\n          \"Pedido com 10 toalhas \"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"review_comment_message\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
```

\"num_unique_values\": 36159,\n        \"samples\": [\n
\"Paguei dois produtos com 60 e s\\u00f3 veio um na nota fiscal consta
duas unidades por\\u00e9m quando eu abri a caixa s\\u00f3 tinha uma
unidade desde j\\u00e1 pe\\u00e7o v\\u00eania \",\n          \"Foi ate
entregue antes do prazo... Parab\\u00e9ns.\",\n          \"prateleira
de tamanho ideal ao que eu esperava,\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"review_creation_date\",\n
\"properties\": {\n        \"dtype\": \"object\",\n
\"num_unique_values\": 636,\n        \"samples\": [\n          \"2017-
05-03 00:00:00\",\n          \"2017-03-13 00:00:00\",\n
\"2017-08-03 00:00:00\"\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"review_answer_timestamp\",\n      \"properties\": {\n
\"dtype\": \"object\",\n        \"num_unique_values\": 98248,\n
\"samples\": [\n          \"2018-08-23 13:09:13\",\n          \"2017-
11-01 21:19:58\",\n          \"2017-07-20 11:07:29\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\
n}","type":"dataframe","variable_name":"order_reviews_dataset"}

```
order_dateset.head()
```

{"summary":"{\n  \"name\": \"order_dateset\",\n  \"rows\": 99441,\n
\"fields\": [\n    {\n      \"column\": \"order_id\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 99441,\n        \"samples\": [\n
\"b9a6c5f5df52c7226ac85aee7524c27f\",\n
\"261e71d2349c713eafa9f3df5972b95d\",\n
\"67b50899f52995848c427e361e10dde3\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"customer_id\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 99441,\n        \"samples\": [\n
\"f160aaf480efdfa7268f0fa535f73e76\",\n
\"d6708bbbd2d419475869a84e41f620a1\",\n
\"1b353c00c71689afba44554e43cc5a76\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"order_status\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 8,\n        \"samples\": [\n
\"invoiced\",\n          \"canceled\",\n          \"delivered\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"order_purchase_timestamp\",\n
\"properties\": {\n        \"dtype\": \"object\",\n
\"num_unique_values\": 98875,\n        \"samples\": [\n
\"2017-12-03 02:01:18\",\n          \"2018-05-15 22:52:30\",\n
\"2018-08-22 21:56:07\"\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"order_approved_at\",\n      \"properties\": {\n

\"dtype\": \"object\",\n        \"num_unique_values\": 90733,\n
\"samples\": [\n          \"2017-05-17 03:55:26\",\n          \"2018-
03-04 10:55:33\",\n          \"2017-10-05 02:49:39\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"order_delivered_carrier_date\",\n
\"properties\": {\n        \"dtype\": \"object\",\n
\"num_unique_values\": 81018,\n        \"samples\": [\n
\"2018-04-24 14:38:40\",\n          \"2017-03-18 10:37:00\",\n
\"2017-12-29 17:22:42\"\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"order_delivered_customer_date\",\n      \"properties\":
{\n        \"dtype\": \"object\",\n        \"num_unique_values\":
95664,\n        \"samples\": [\n          \"2017-03-27 12:43:40\",\n
\"2018-02-23 16:39:23\",\n          \"2017-10-09 22:16:39\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"order_estimated_delivery_date\",\n      \"properties\": {\n
\"dtype\": \"object\",\n        \"num_unique_values\": 459,\n
\"samples\": [\n          \"2018-07-27 00:00:00\",\n          \"2018-
08-06 00:00:00\",\n          \"2017-09-13 00:00:00\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"order_dateset"}

```
products_dataset.head()
```

{"summary":"{\n  \"name\": \"products_dataset\",\n  \"rows\": 32951,\n
\"fields\": [\n    {\n      \"column\": \"product_id\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 32951,\n        \"samples\": [\n
\"f819f0c84a64f02d3a5606ca95edd272\",\n
\"b9de40e43fccb6ba53b7eadbd5c49077\",\n
\"26afe4ed5890d941fcba14205863eec7\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"product_category_name\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 73,\n        \"samples\": [\n
\"utilidades_domesticas\",\n          \"pc_gamer\",\n
\"eletronicos\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"product_name_lenght\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 10.245740725237239,\n        \"min\":
5.0,\n        \"max\": 76.0,\n        \"num_unique_values\": 66,\n
\"samples\": [\n          13.0,\n          72.0,\n          40.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"product_description_lenght\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
635.1152246349538,\n        \"min\": 4.0,\n        \"max\": 3992.0,\n
\"num_unique_values\": 2960,\n        \"samples\": [\n
1314.0,\n          40.0,\n          123.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\

```
n    },\n    {\n      \"column\": \"product_photos_qty\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
1.7367656379315444,\n        \"min\": 1.0,\n        \"max\": 20.0,\n
\"num_unique_values\": 19,\n        \"samples\": [\n          1.0,\n
9.0,\n          17.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"product_weight_g\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 4282.038730977007,\n        \"min\":
0.0,\n        \"max\": 40425.0,\n        \"num_unique_values\": 2204,\
n        \"samples\": [\n          567.0,\n          241.0,\n
2646.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"product_length_cm\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 16.914458054065967,\n        \"min\":
7.0,\n        \"max\": 105.0,\n        \"num_unique_values\": 99,\n
\"samples\": [\n          88.0,\n          75.0,\n          10.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"product_height_cm\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
13.637554061749494,\n        \"min\": 2.0,\n        \"max\": 105.0,\n
\"num_unique_values\": 102,\n        \"samples\": [\n          6.0,\n
38.0,\n          54.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"product_width_cm\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 12.079047453227835,\n        \"min\":
6.0,\n        \"max\": 118.0,\n        \"num_unique_values\": 95,\n
\"samples\": [\n          80.0,\n          38.0,\n          9.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    }\n  ]\
n}","type":"dataframe","variable_name":"products_dataset"}
```


```
sellers_dataset.head()
```


```
{"summary":"{\n  \"name\": \"sellers_dataset\",\n  \"rows\": 3095,\n
\"fields\": [\n    {\n      \"column\": \"seller_id\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 3095,\n        \"samples\": [\n
\"1f7fd2a6fcd5a6fa5d8a4dabc72aaae0\",\n
\"6e1862e15f33d9994bc25922a85e1efc\",\n
\"e628d4a53c109f09ca88098338b3a3f5\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"seller_zip_code_prefix\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
32713,\n        \"min\": 1001,\n        \"max\": 99730,\n
\"num_unique_values\": 2246,\n        \"samples\": [\n
87160,\n          6787,\n          36325\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"seller_city\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 611,\n        \"samples\": [\n
```