

Customer Satisfaction Analysis

Objective: We shall be analysing customer satisfaction by using their review score as a metric of satisfaction.

Dataset: [https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?resource=download)
[resource=download](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?resource=download)

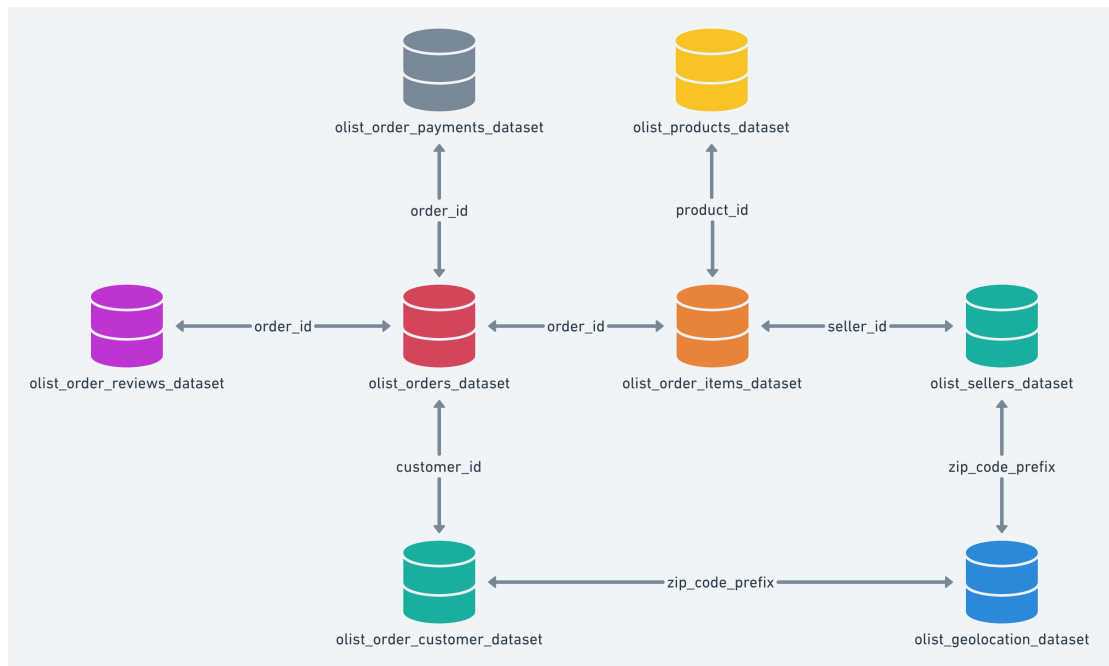
Brazilian E-Commerce Public Dataset by Olist Welcome! This is a Brazilian ecommerce public dataset of orders made at Olist Store. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates. This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

1. Importing the files

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [ ]: customers = pd.read_csv("datasets/olist_customers_dataset.csv")
geoloc = pd.read_csv("datasets/olist_geolocation_dataset.csv")
order_reviews = pd.read_csv("datasets/olist_order_reviews_dataset.csv")
order_items = pd.read_csv("datasets/olist_order_items_dataset.csv")
order_payments = pd.read_csv("datasets/olist_order_payments_dataset.csv")
orders = pd.read_csv("datasets/olist_orders_dataset.csv")
products = pd.read_csv("datasets/olist_products_dataset.csv")
sellers = pd.read_csv("datasets/olist_sellers_dataset.csv")
product_category_name_translation = pd.read_csv("datasets/product_category_name_
```

2. Relationship Between The Tables



```
In [ ]: customers.head()
```

```
Out[ ]:
```

	customer_id	customer_unique_id	customer_zi
0	06b8999e2fba1a1fbc88172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0	
1	18955e83d337fd6b2def6b18a428ac77	290c77bc529b7ac935b93aa66c333dc3	
2	4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e	
3	b2b6027bc5c5109e529d4dc6358b12c3	259dac757896d24d7702b9acbbff3f3c	
4	4f2d8ab171c80ec8364f7c12e35b23ad	345ecd01c38d18a9036ed96c73b8d066	

```
In [ ]: geoloc.head()
```

```
Out[ ]:
```

	geolocation_zip_code_prefix	geolocation_lat	geolocation_lng	geolocation_city	geolo
0	1037	-23.545621	-46.639292	sao paulo	
1	1046	-23.546081	-46.644820	sao paulo	
2	1046	-23.546129	-46.642951	sao paulo	
3	1041	-23.544392	-46.639499	sao paulo	
4	1035	-23.541578	-46.641607	sao paulo	

```
In [ ]: order_reviews.head()
```

Out[]:

	review_id	order_id	review_score
0	7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb	4
1	80e641a11e56f04c1ad469d5645fdfde	a548910a1c6147796b98fdf73dbeba33	5
2	228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b	5
3	e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b	5
4	f7c4243c7fe1938f181bec41a392bdeb	8e6bfb81e283fa7e4f11123a3fb894f1	5

In []:

order_items.head()

Out[]:

	order_id	order_item_id	product_id
0	00010242fe8c5a6d1ba2dd792cb16214	1	4244733e06e7ecb4970a6e2683c13e61
1	00018f77f2f0320c557190d7a144bdd3	1	e5f2d52b802189ee658865ca93d83a8
2	000229ec398224ef6ca0657da4fc703e	1	c777355d18b72b67abbef9df44fd0fc
3	00024acbcd0a6daa1e931b038114c75	1	7634da152a4610f1595efa32f14722fc
4	00042b26cf59d7ce69dfabb4e55b4fd9	1	ac6c3623068f30de03045865e4e1008c

In []:

order_payments.head()

Out[]:

	order_id	payment_sequential	payment_type	payment_ins
0	b81ef226f3fe1789b1e8b2acac839d17	1	credit_card	
1	a9810da82917af2d9aefd1278f1dcfa0	1	credit_card	
2	25e8ea4e93396b6fa0d3dd708e76c1bd	1	credit_card	
3	ba78997921bbcdc1373bb41e913ab953	1	credit_card	
4	42fdf880ba16b47b59251dd489d4441a	1	credit_card	

In []:

orders.head()

Out[]:

	order_id	customer_id	order_status
0	e481f51cbdc54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered
1	53cdb2fc8bc7dce0b6741e2150273451	b0830fb4747a6c6d20dea0b8c802d7ef	delivered
2	47770eb9100c2d0c44946d9cf07ec65d	41ce2a54c0b03bf3443c3d931a367089	delivered
3	949d5b44dbf5de918fe9c16f97b45f8a	f88197465ea7920adcdbec7375364d82	delivered
4	ad21c59c0840e6cb83a9ceb5573f8159	8ab97904e6daea8866dbdbbc4fb7aad2c	delivered

In []: `products.head()`

Out[]:

	product_id	product_category_name	product_name_lenght	p
0	1e9e8ef04dbcff4541ed26657ea517e5	perfumaria	40.0	
1	3aa071139cb16b67ca9e5dea641aaa2f	artes	44.0	
2	96bd76ec8810374ed1b65e291975717f	esporte_lazer	46.0	
3	cef67bcfe19066a932b7673e239eb23d	bebes	27.0	
4	9dc1a7de274444849c219cff195d0b71	utilidades_domesticas	37.0	

In []: `sellers.head()`

Out[]:

	seller_id	seller_zip_code_prefix	seller_city	seller_state
0	3442f8959a84dea7ee197c632cb2df15	13023	campinas	SP
1	d1b65fc7debc3361ea86b5f14c68d2e2	13844	mogi guacu	SP
2	ce3ad9de960102d0677a81f5d0bb7b2d	20031	rio de janeiro	RJ
3	c0f3eea2e14555b6faeea3dd58c1b1c3	4195	sao paulo	SP
4	51a04a8a6bdcdb23deccc82b0b80742cf	12914	braganca paulista	SP

In []: `product_category_name_translation.head()`

```
Out[ ]: product_category_name product_category_name_english
```

0	beleza_saude	health_beauty
1	informatica_acessorios	computers_accessories
2	automotivo	auto
3	cama_mesa_banho	bed_bath_table
4	moveis_decoracao	furniture_decor

3. Data Cleaning

3.1 Handling Missing Value

```
In [ ]: customers.isnull().sum()
```

```
Out[ ]: customer_id          0
customer_unique_id        0
customer_zip_code_prefix  0
customer_city             0
customer_state            0
dtype: int64
```

```
In [ ]: geoloc.isnull().sum()
```

```
Out[ ]: geolocation_zip_code_prefix  0
geolocation_lat                    0
geolocation_lng                    0
geolocation_city                   0
geolocation_state                  0
dtype: int64
```

```
In [ ]: order_reviews.isnull().sum()
```

```
Out[ ]: review_id          0
order_id          0
review_score      0
review_comment_title    87656
review_comment_message  58247
review_creation_date    0
review_answer_timestamp  0
dtype: int64
```

```
In [ ]: # Since these two have missing values
del order_reviews['review_comment_title']
del order_reviews['review_comment_message']
```

```
In [ ]: order_reviews.isnull().sum()
```

```
Out[ ]: review_id          0
order_id          0
review_score      0
review_creation_date    0
review_answer_timestamp  0
dtype: int64
```

```
In [ ]: order_items.isnull().sum()
```

```
Out[ ]: order_id          0
order_item_id         0
product_id            0
seller_id             0
shipping_limit_date   0
price                 0
freight_value         0
dtype: int64
```

```
In [ ]: order_payments.isnull().sum()
```

```
Out[ ]: order_id          0
payment_sequential     0
payment_type           0
payment_installments   0
payment_value          0
dtype: int64
```

```
In [ ]: orders.isnull().sum()
```

```
Out[ ]: order_id          0
customer_id             0
order_status            0
order_purchase_timestamp 0
order_approved_at       160
order_delivered_carrier_date 1783
order_delivered_customer_date 2965
order_estimated_delivery_date 0
dtype: int64
```

```
In [ ]: del orders['order_approved_at']
del orders['order_delivered_carrier_date']
del orders['order_delivered_customer_date']
orders.isnull().sum()
```

```
Out[ ]: order_id          0
customer_id             0
order_status            0
order_purchase_timestamp 0
order_estimated_delivery_date 0
dtype: int64
```

```
In [ ]: products.isnull().sum()
```

```
Out[ ]: product_id          0
product_category_name     610
product_name_lenght       610
product_description_lenght 610
product_photos_qty        610
product_weight_g           2
product_length_cm          2
product_height_cm          2
product_width_cm           2
dtype: int64
```

```
In [ ]: # Replacng missing categories with "Unknown"
products['product_category_name'].replace(np.nan, 'Unknown', inplace=True)
```

```
In [ ]: # To only keep product_id and product_category_name columns
products=products[['product_id','product_category_name']].copy()
```

```
In [ ]: products.isnull().sum()
```

```
Out[ ]: product_id          0
product_category_name      0
dtype: int64
```

```
In [ ]: sellers.isnull().sum()
```

```
Out[ ]: seller_id          0
seller_zip_code_prefix     0
seller_city                0
seller_state               0
dtype: int64
```

```
In [ ]: product_category_name_translation.isnull().sum()
```

```
Out[ ]: product_category_name          0
product_category_name_english         0
dtype: int64
```

3.2 Managing Inconsistent Data

```
In [ ]: order_payments['payment_value'] = order_payments['payment_value'].astype(float)
order_payments.head()
```

```
Out[ ]:
```

	order_id	payment_sequential	payment_type	payment_ins
0	b81ef226f3fe1789b1e8b2acac839d17	1	credit_card	
1	a9810da82917af2d9aefd1278f1dcfa0	1	credit_card	
2	25e8ea4e93396b6fa0d3dd708e76c1bd	1	credit_card	
3	ba78997921bbcdc1373bb41e913ab953	1	credit_card	
4	42fdf880ba16b47b59251dd489d4441a	1	credit_card	



4. Data Preprocessing

4.1 String Manipulation

```
In [ ]: customers['customer_city'] = customers['customer_city'].str.title()
customers.head()
```

	customer_id	customer_unique_id	customer_zi
0	06b8999e2fba1a1fbc88172c00ba8bc7	861eff4711a542e4b93843c6dd7febb0	
1	18955e83d337fd6b2def6b18a428ac77	290c77bc529b7ac935b93aa66c333dc3	
2	4e7b3e00288586ebd08712fdd0374a03	060e732b5b29e8181a18229c7b0b2b5e	
3	b2b6027bc5c5109e529d4dc6358b12c3	259dac757896d24d7702b9acbbff3f3c	
4	4f2d8ab171c80ec8364f7c12e35b23ad	345ecd01c38d18a9036ed96c73b8d066	

```
In [ ]: geoloc['geolocation_city'] = geoloc['geolocation_city'].str.title()
geoloc.head()
```

	geolocation_zip_code_prefix	geolocation_lat	geolocation_lng	geolocation_city	geolo
0	1037	-23.545621	-46.639292	Sao Paulo	
1	1046	-23.546081	-46.644820	Sao Paulo	
2	1046	-23.546129	-46.642951	Sao Paulo	
3	1041	-23.544392	-46.639499	Sao Paulo	
4	1035	-23.541578	-46.641607	Sao Paulo	

```
In [ ]: sellers['seller_city'] = sellers['seller_city'].str.title()
sellers.head()
```

	seller_id	seller_zip_code_prefix	seller_city	seller_state
0	3442f8959a84dea7ee197c632cb2df15	13023	Campinas	SP
1	d1b65fc7debc3361ea86b5f14c68d2e2	13844	Mogi Guacu	SP
2	ce3ad9de960102d0677a81f5d0bb7b2d	20031	Rio De Janeiro	RJ
3	c0f3eea2e14555b6faeea3dd58c1b1c3	4195	Sao Paulo	SP
4	51a04a8a6bdcb23deccc82b0b80742cf	12914	Braganca Paulista	SP

```
In [ ]: product_category_name_translation['product_category_name_english'] = product_cat
product_category_name_translation.head()
```


Out []: **product_category_name product_category_name_english**


0	beleza_saude	Health Beauty
1	informatica_acessorios	Computers Accessories
2	automotivo	Auto
3	cama_mesa_banho	Bed Bath Table
4	moveis_decoracao	Furniture Decor

4.2 Discretisation

```
In [ ]: order_reviews_bin = [1,2,3,6]
order_reviews_label = ['Poor', 'Average', 'Good']
order_reviews['review_score_group'] = pd.cut(order_reviews['review_score'], bins=
order_reviews.head()
```

Out []:

	review_id	order_id	review_score
0	7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb	4
1	80e641a11e56f04c1ad469d5645fdfe	a548910a1c6147796b98fdf73dbeba33	5
2	228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b	5
3	e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b	5
4	f7c4243c7fe1938f181bec41a392bdeb	8e6bfb81e283fa7e4f11123a3fb894f1	5



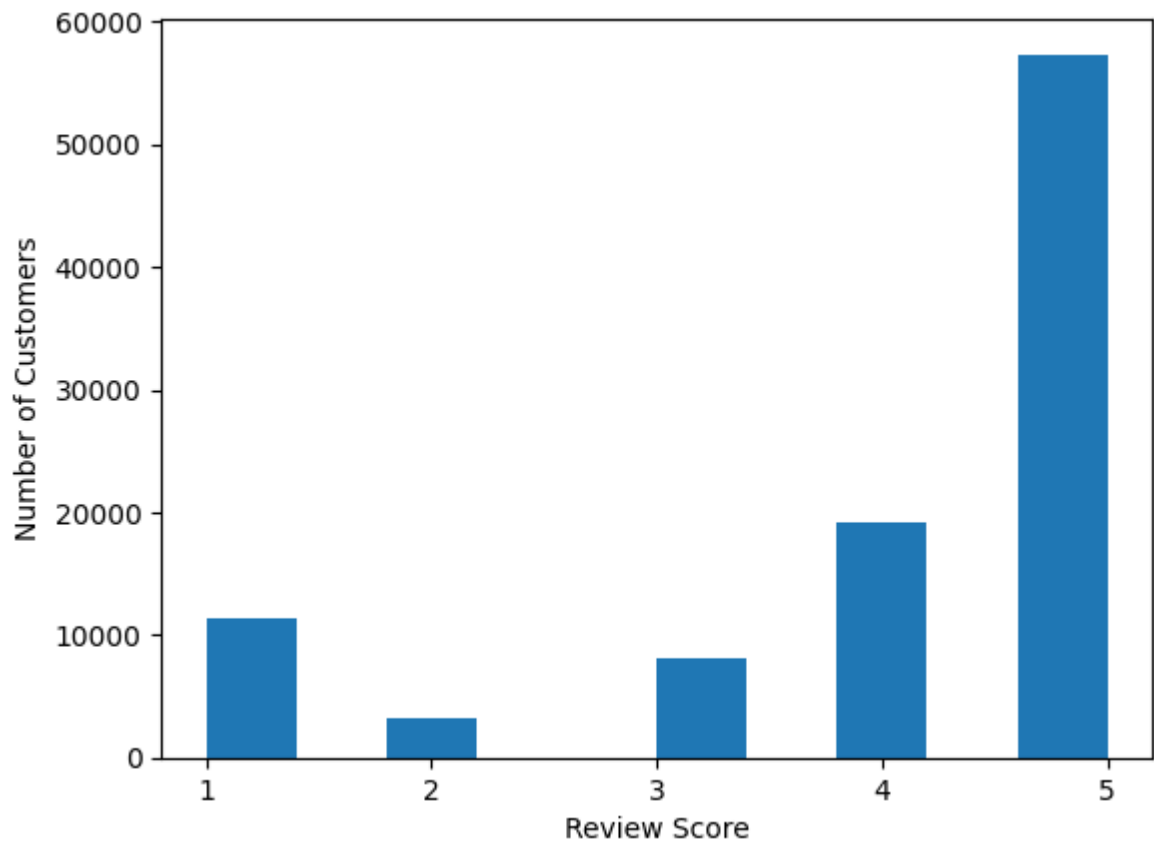
5. Data Analysis And Visualization

1. Histogram for customer review
2. Customer reviews grouped by state
3. Best 5 reviewed product category analysis
4. Worst 5 reviewed product category analysis
5. Top 10 Seller with highest product reviews
6. Bottom 10 Seller with lowest product reviews
7. Payment method used by most satisfied customers
8. Freight value paid by most satisfied customers
9. Percentage of seller who satisfied most customers (Sellers whose products have 4 or above stars rating)
10. Percentage of customers most satisfied (Satisfaction = Review score of 4 or more)

5.1 Histogram for customer review

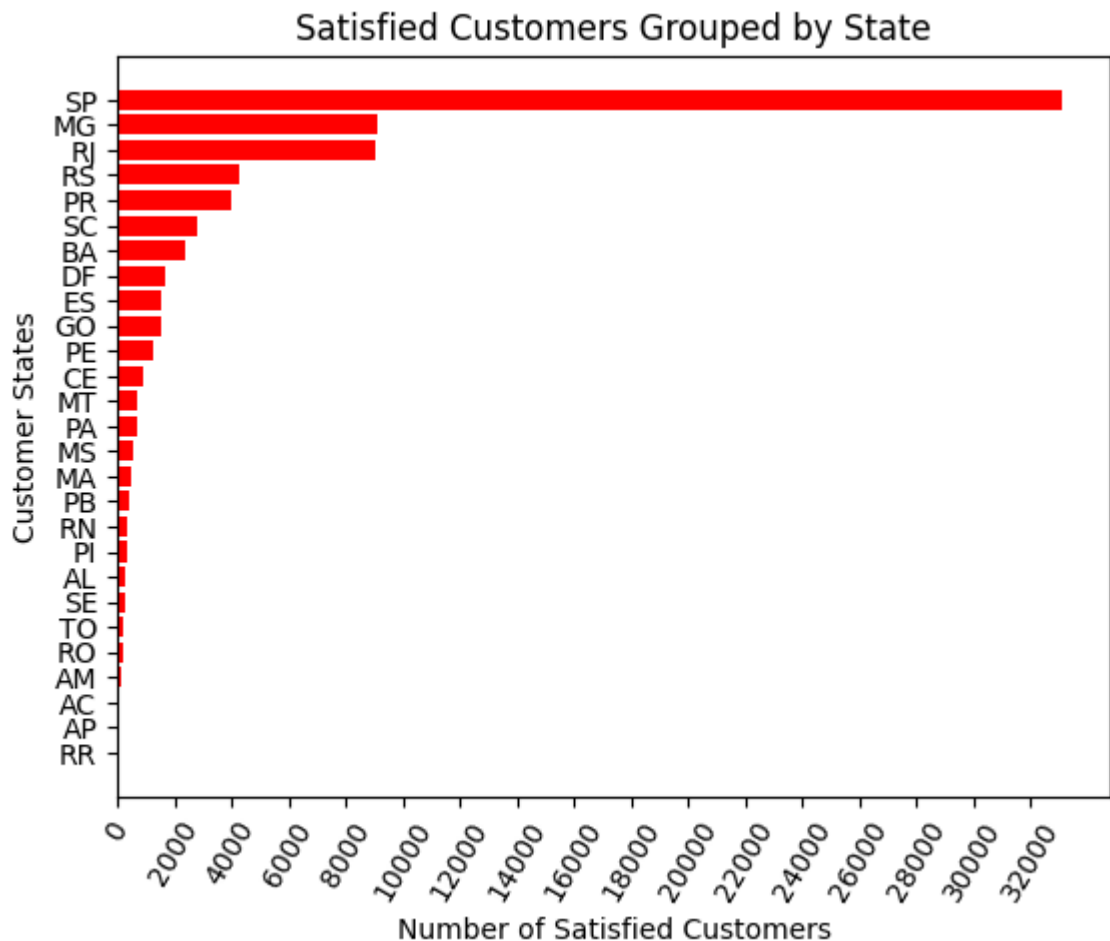
```
In [ ]: plt.hist(order_reviews['review_score'])
plt.xticks([1,2,3,4,5])
plt.xlabel('Review Score')
```

```
plt.ylabel('Number of Customers')
plt.show()
```



5.2 Customer reviews grouped by state

```
In [ ]: grp2 = pd.merge(customers, orders, on='customer_id')
grp2 = pd.merge(grp2, order_reviews, on='order_id')
grp2 = grp2.drop(grp2[(grp2['review_score'] < 4)].index)
grp2 = grp2[['customer_state', 'customer_id']]
grp2 = grp2.groupby('customer_state').count().reset_index()
grp2 = grp2.sort_values(by='customer_id')
x = grp2['customer_id']
y = grp2['customer_state']
plt.barh(
    y = y,
    width= x,
    color='r',
)
plt.xticks(range(0,x.max(),2000), rotation = 60)
plt.title('Satisfied Customers Grouped by State')
plt.ylabel('Customer States')
plt.xlabel('Number of Satisfied Customers')
plt.show()
```



5.3 Best 5 reviewed product categories analysis

```
In [ ]: product_category_review = pd.merge(products,order_items,on='product_id')
product_category_review = pd.merge(product_category_review ,orders,on='order_id')
product_category_review = pd.merge(product_category_review ,order_reviews,on='or
product_category_review = product_category_review[['product_category_name','revi
product_category_review = product_category_review.groupby('product_category_name
product_category_review = pd.merge(product_category_review,product_category_name
product_category_review = product_category_review[['product_category_name_englis
product_category_review = product_category_review.sort_values(by='review_score',
product_category_review.columns = ['Product Category', 'Review Score']
product_category_review.head()
```

```
Out[ ]:
```

	Product Category	Review Score
17	Cds Dvds Musicals	5.0
45	Musical Instruments	4.0
51	Market Place	4.0
50	Luggage Accessories	4.0
49	Books Technical	4.0

5.4 Worst 5 reviewed product category analysis

```
In [ ]: product_category_review = product_category_review.sort_values(by='Review Score')
product_category_review.head()
```

```
Out[ ]:
```

	Product Category	Review Score
65	Security And Services	2.0
55	Office Furniture	3.0
42	Diapers And Hygiene	3.0
70	Housewares	4.0
15	Home Comfort 2	4.0

5.5 Top 10 Seller with highest product reviews

```
In [ ]: grp3 = pd.merge(order_reviews,orders,on='order_id')
grp3 = pd.merge(grp3,order_items,on='order_id')
grp3 = pd.merge(grp3,sellers,on='seller_id')
grp3 = grp3[['review_score','seller_id']]
grp3 = grp3.groupby('seller_id').mean().round().reset_index()
grp3 = grp3.sort_values(by = 'review_score', ascending = False)
grp3.head(10)
```

```
Out[ ]:
```

	seller_id	review_score
2605	d8b8f2cf9ff6ba0389072541cb42498c	5.0
606	33a17d60c64393351ebf1ef860f4e0f2	5.0
595	32e5635e63cb374eb63afdd242fb6134	5.0
1356	6f835fd4be26989b1b064399da346143	5.0
597	32f83ffe11cd40f7adcf4eef171f52d9	5.0
2575	d566c37fa119d5e66c4e9052e83ee4ea	5.0
2574	d558ebe531605a1285ab2b1bc3256dfb	5.0
600	333c4210e76a1aa2ab817b99437e3ff1	5.0
2573	d52cbce9845184537284a23c3bc3da0e	5.0
602	334cab711dee080b079fa5779b584783	5.0

5.6 Bottom 10 Seller with lowest product reviews

```
In [ ]: grp3 = pd.merge(order_reviews,orders,on='order_id')
grp3 = pd.merge(grp3,order_items,on='order_id')
grp3 = pd.merge(grp3,sellers,on='seller_id')
grp3 = grp3[['review_score','seller_id']]
grp3 = grp3.groupby('seller_id').mean().round().reset_index()
grp3 = grp3.sort_values(by = 'review_score')
grp3.head(10)
```

	seller_id	review_score
1967	a247197e2e9c19a6a53a6888cb8b660f	1.0
2740	e46bc031f2c5bae4ccb40bb90712e9b4	1.0
128	0aa124728afc1131dff5655f4c6f487b	1.0
2439	c97aa4ee7420f937da13b7f9e2228b99	1.0
301	1992f8fb6b19fccccd97ca819811e7267	1.0
1265	67e43d802fde8cfd3f9580124f8167d1	1.0
1823	9599519be538b98748162a2b48248960	1.0
983	5206cc4bc2297c833e6061c49bf9c43f	1.0
297	1967a9e2ad6f51802b093147d861df58	1.0
1661	87f3e35268860433e13d577825aada95	1.0

5.7 Payment method used by most satisfied customers

```
In [ ]: grp = pd.merge(order_reviews,orders,on='order_id')
grp = pd.merge(grp,customers,on='customer_id')
grp = pd.merge(grp, order_payments,on='order_id')
grp = grp[['review_score','customer_id', 'payment_type']]
grp = grp.drop(grp[(grp['review_score']<4)].index)
grp = grp[['customer_id', 'payment_type']]
grp = grp.groupby('payment_type').count().reset_index()
grp.head()
```

	payment_type	customer_id
0	boleto	15217
1	credit_card	59098
2	debit_card	1212
3	voucher	4303

5.8 Freight value paid by most satisfied customers

```
In [ ]: grp = pd.merge(order_reviews,orders,on='order_id')
grp = pd.merge(grp,customers,on='customer_id')
grp = pd.merge(grp, order_items,on='order_id')
grp = grp[['review_score','customer_id', 'freight_value']]
grp = grp.drop(grp[(grp['review_score']<4)].index)
grp = grp[['freight_value']]
avg_frieght = grp.mean().round(2)
print("Average frieght value paid by most satisfied customers =",avg_frieght.val
```

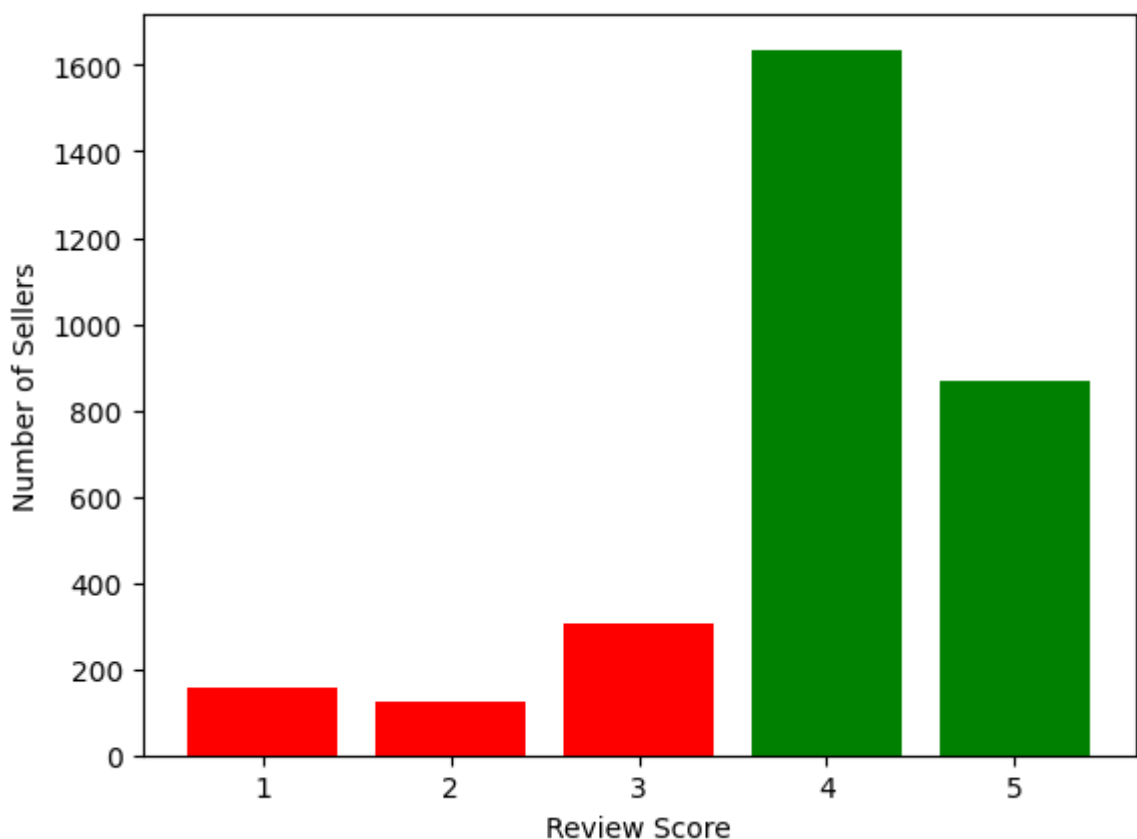
Average frieght value paid by most satisfied customers = 19.69

5.9. Percentage of seller who satisfied most customers (Sellers whose products have 4 or above stars rating)

```
In [ ]: grp = pd.merge(order_reviews, orders, on='order_id')
grp = pd.merge(grp, order_items, on='order_id')
grp = pd.merge(grp, sellers, on='seller_id')
grp = grp[['review_score', 'seller_id']]
grp = grp.groupby('seller_id').mean().round().reset_index()
grp = grp.groupby('review_score').count().reset_index()
total_sellers = sellers['seller_id'].unique().size
satisfactory_sellers = grp['seller_id'].where(grp['review_score'] > 3).sum()
percentage_satisfactory_sellers = (satisfactory_sellers / total_sellers) * 100
print(f"Hence, {round(percentage_satisfactory_sellers, 2)}% are satisfactory sell
```

Hence, 80.94% are satisfactory sellers receiving 4 or more as review score.

```
In [ ]: x = grp.review_score
y = grp.seller_id
plt.bar(x, y, color=[('g' if i > 3 else 'r') for i in x])
plt.xlabel('Review Score')
plt.ylabel('Number of Sellers')
plt.show()
```



5.10 Percentage of customers satisfied (Satisfaction = Review score of 4 or more)

```
In [ ]: customer_reviews = pd.merge(orders, order_reviews, on='order_id')
customer_reviews['review_score']
total_reviews = order_reviews.review_score.count()
satisfied_reviews = order_reviews.review_score.where(lambda s: s >= 4).count()
```

```
percentage_satisfied_customers = (satisfied_reviews/total_reviews)*100  
print(f"Hence, {round(percentage_satisfied_customers,2)}% of customers are satis
```

Hence, 77.07% of customers are satisfied with there orders.

6. Conclusion

- **High Customer Ratings:** A significant number of customers (55,000) have given a 5-star rating, indicating a generally positive sentiment towards the products and services.
- **Regional Satisfaction:** The state of São Paulo (SP) stands out with the highest number of satisfied customers, approximately 32,000, suggesting potential regional preferences or effective marketing strategies in that area.
- **Best-Reviewed Product Categories:** Certain product categories, such as "Cds Dvds Musicals," "Musical Instruments," and "Market Place," received top-notch ratings, reflecting the success and popularity of these items among customers.
- **Least Favorable Product Categories:** Conversely, product categories like "Security And Services" and "Office Furniture" received lower ratings, signaling potential areas for improvement in product quality or customer experience.
- **Top-Rated Sellers:** The top 10 sellers with the highest product reviews consistently achieved a perfect 5.0 score, showcasing their excellence in customer satisfaction and service.
- **Lowest-Rated Sellers:** On the other hand, the bottom 10 sellers with the lowest product reviews received a 1.0 score, indicating areas that may require attention to enhance customer satisfaction and trust.
- **Preferred Payment Methods:** The majority of satisfied customers used credit cards for transactions (59,098), followed by boleto (15,217), debit cards (1,212), and vouchers (4,303), offering insights into popular payment preferences.
- **Freight Value:** The most common freight value paid by satisfied customers is \$19.69, providing a benchmark for shipping cost expectations among the content customer base.
- **Satisfactory Sellers:** A substantial 80.94% of sellers received a review score of 4 or above, indicating that the majority of sellers are meeting or exceeding customer expectations.
- **Overall Customer Satisfaction:** A commendable 77.07% of customers expressed satisfaction with their orders, reinforcing the positive outlook on the overall customer experience within the analyzed data set.

In summary, this customer satisfaction analysis unveils valuable information about product performance, seller reliability, and customer preferences, offering actionable insights to enhance the overall satisfaction and loyalty of the customer base.