For my project, I began by downloading a dataset from Kaggle. Upon opening the dataset, I noticed that it contained a significant number of empty cells, which could potentially cause issues when working with the data in Python. To address this, I opted to replace these empty slots with the string 'NA' using Excel's find and replace feature. After making this adjustment, I imported the cleaned data into a Jupyter Notebook for better data visualization and analysis.

**Amazon Sales Project**

Subsequently, I encountered an issue with the date formats within the dataset. The dates were recorded in two different formats: 'MM-DD-YY' and 'MM/DD/YYYY.' Unfortunately, neither of these formats was in the desired 'DD/MM/YY' format. To standardize the date format, I employed pivot tables and the INDEX-MATCH formula in Excel to ensure that all dates conformed to the correct format, enabling proper data processing.

Once the date formats were harmonized, I proceeded to check for missing values in the dataset using the df.isnull().sum() function in Python. This inspection revealed a substantial number of missing values, it turned out to be from the NA values I previously replaced . To ensure the data's integrity and quality, I performed further data cleaning and preparation in Excel. With the data fully cleaned and preprocessed, I was able to commence my data analysis in Python.

**Results based of the Analysis ran in python:**

The max amount paid by an Indian customer is equivalent to (5584 INR) and the average amount spent is (609.363662 INR).

I Performed som exploratory data analysis, I first checked the date span of the data :

From:

Timestamp('2022-03-31 00:00:00')

Until:

Timestamp('2022-06-29 00:00:00')

Questions: SALES Analysis:

-What is the total sales revenue ? :

78592678.30 INR from 2022-03-31 until 22-06-29.  -What are the top 10 product by sale?

|  |  |
| --- | --- |
| JNE3797 | 2933482 |
| J0230 | 1944948 |
| SET268 | 1303924 |
| J0341 | 1275323 |
| J0003 | 981973.8 |
| JNE3405 | 846822.4 |
| J0008 | 819170 |
| SET345 | 723402.8 |
| SET324 | 720096.8 |
| SET278 | 719791.48 |

-What are the top 10 products by QTY?

|  |  |
| --- | --- |
| JNE3797 | 3692 |
| JNE3405 | 2067 |
| SET268 | 1731 |
| J0230 | 1710 |
| J0341 | 1471 |
| J0003 | 1390 |
| SET324 | 1152 |
| SET345 | 1134 |
| JNE3373 | 1070 |
| JNE3399 | 971 |

-What is the total number of orders?  77804

-What is the AOV?

TR/TO :

78592678.30 / 77804= 1010.14 INR  -What are the peak sales Periods?

|  |  |
| --- | --- |
| 30/04/2022 | 969634.4 |
| 01/05/2022 | 1079958 |
| 02/05/2022 | 1172327 |
| 03/05/2022 | 1190673 |
| 04/05/2022 | 1209364 |

CUSTOMER Analysis:  -What are the top cities/states by sales ?

|  |  |  |
| --- | --- | --- |
| **ship-city** | **ship-state** | **Amount** |
| BENGALURU | KARNATAKA | 4723666 |
| HYDERABAD | TELANGANA | 3250847 |
| MUMBAI | MAHARASHTRA | 2437985 |
| NEW DELHI | DELHI | 2345059 |
| CHENNAI | TAMIL NADU | 2074972 |
| PUNE | MAHARASHTRA | 1535676 |
| KOLKATA | WEST BENGAL | 943375 |
| GURUGRAM | HARYANA | 850907 |
| THANE | MAHARASHTRA | 671973 |
| NOIDA | UTTAR PRADESH | 605745 |

-What are the top cities/states by number of orders?

|  |  |  |
| --- | --- | --- |
| **ship-city** | **ship-state** | **Qty** |
| BENGALURU | KARNATAKA | 7339 |
| HYDERABAD | TELANGANA | 5020 |
| MUMBAI | MAHARASHTRA | 3839 |
| NEW DELHI | DELHI | 3532 |
| CHENNAI | TAMIL NADU | 3469 |
| PUNE | MAHARASHTRA | 2405 |
| KOLKATA | WEST BENGAL | 1501 |
| GURUGRAM | HARYANA | 1237 |
| THANE | MAHARASHTRA | 1063 |
| NOIDA | UTTAR PRADESH | 883 |

-Are there any order cancelation trends?

|  |  |  |
| --- | --- | --- |
|  | **DateN** | Cancelled orders |
|  | 2022-04-30 | 240 |
|  | 2022-05-01 | 234 |
|  | 2022-05-02 | 318 |
|  | 2022-05-03 | 334 |
|  | 2022-05-04 | 304 |

In the beginning of May until the middle of May there is an increase in the number of cancelled orders.

-What is the distribution of B2B vs B2C customers?

\*B2B: COUNT: 871

PERCENTAGE: 0.67%

\*B2C: COUNT: 128104

PERCENTAGE: 99.33%

SHIPPING ANALYSIS:

-What percentage of orders where Fulfilled by Amazon vs Merchants?

Amazon 69.546811 %

Merchant 30.453189 %  -What is the Preferred shipment method?

Expedited 68.707114 %

Standard 31.292886 %  PRODUCT ANALYSIS

-Which categories of products are most popular?

|  |  |
| --- | --- |
| **Category** | Count |
| Set | 50284 |
| kurta | 49877 |
| Western Dress | 15500 |
| Top | 10622 |
| Ethnic Dress | 1159 |
| Blouse | 926 |
| Bottom | 440 |
| Saree | 164 |
| Dupatta | 3 |

The 2 most popular categories of products by far are Set and Kurta.\

-Are there any specific styles that are more popular than others?

|  |  |
| --- | --- |
| Style | Count |
| JNE3797 | 4224 |
| JNE3405 | 2263 |
| J0230 | 1868 |
| SET268 | 1860 |
| J0341 | 1630 |
| J0003 | 1627 |
| SET324 | 1284 |
| SET345 | 1250 |
| JNE3373 | 1173 |
| JNE3440 | 1054 |

The JNE3797 and the JNE3405 seem to be the 2 mos popular styles but not by far from the rest.

-What is the average quantity of products ordered?

0.904431

Total quantity ordered : 116649  PROMOTIONAL ANALYSIS

-what is the most used promotional id ?

IN Core Free Shipping 2015/04/08 23-48-5-108 :

Used : 46100