Commands + Code + Text

```
[2] import pandas as pd
    import numpy as np
```

```
[3] data = pd.read_csv('/content/drive/MyDrive/Brainwave Internship Projects/amazon_sales_data 2025.csv')
```

```
[4] data.shape
```

```
(250, 11)
```

```
data.head(10)
```

	Order ID	Date	Product	Category	Price	Quantity	Total Sales	Customer Name	Customer Location	Payment Method	Status
0	ORD0001	14-03-25	Running Shoes	Footwear	60	3	180	Emma Clark	New York	Debit Card	Cancelled
1	ORD0002	20-03-25	Headphones	Electronics	100	4	400	Emily Johnson	San Francisco	Debit Card	Pending
2	ORD0003	15-02-25	Running Shoes	Footwear	60	2	120	John Doe	Denver	Amazon Pay	Cancelled
3	ORD0004	19-02-25	Running Shoes	Footwear	60	3	180	Olivia Wilson	Dallas	Credit Card	Pending
4	ORD0005	10-03-25	Smartwatch	Electronics	150	3	450	Emma Clark	New York	Debit Card	Pending
5	ORD0006	14-03-25	T-Shirt	Clothing	20	1	20	John Doe	Dallas	Credit Card	Pending
6	ORD0007	18-03-25	Smartwatch	Electronics	150	4	600	Emma Clark	Houston	PayPal	Completed
7	ORD0008	02-03-25	Smartphone	Electronics	500	1	500	Sophia Miller	Miami	PayPal	Completed
8	ORD0009	08-03-25	T-Shirt	Clothing	20	3	60	Sophia Miller	Boston	PayPal	Completed
9	ORD0010	12-03-25	Smartphone	Electronics	500	1	500	Emily Johnson	San Francisco	Credit Card	Cancelled

Next steps: Generate code with data | View recommended plots | New interactive sheet

```
[7] data.tail()
```

	Order ID	Date	Product	Category	Price	Quantity	Total Sales	Customer Name	Customer Location	Payment Method	Status
245	ORD0246	17-03-25	T-Shirt	Clothing	20	2	40	Daniel Harris	Miami	Debit Card	Cancelled
246	ORD0247	30-03-25	Jeans	Clothing	40	1	40	Sophia Miller	Dallas	Debit Card	Cancelled
247	ORD0248	05-03-25	T-Shirt	Clothing	20	2	40	Chris White	Denver	Debit Card	Cancelled
248	ORD0249	08-03-25	Smartwatch	Electronics	150	3	450	Emily Johnson	New York	Debit Card	Cancelled
249	ORD0250	19-02-25	Smartphone	Electronics	500	4	2000	Emily Johnson	Seattle	Amazon Pay	Completed

```
[19] data.isnull().sum()
```

	0
Order ID	0
Date	0
Product	0
Category	0
Price	0
Quantity	0
Total Sales	0
Customer Name	0
Customer Location	0
Payment Method	0
Status	0
Year	0
Month	0
Day	0
Weekday	0

dtype: int64

```
[22] data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 250 entries, 0 to 249
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Order ID           250 non-null    object
 1   Date               250 non-null    datetime64[ns]
 2   Product            250 non-null    object
 3   Category           250 non-null    object
 4   Price              250 non-null    int64
 5   Quantity           250 non-null    int64
 6   Total Sales        250 non-null    int64
 7   Customer Name      250 non-null    object
 8   Customer Location  250 non-null    object
 9   Payment Method     250 non-null    object
```

```
 9   Payment Method     250 non-null    object
 10  Status             250 non-null    object
 11  Year               250 non-null    int32
 12  Month              250 non-null    object
 13  Day                250 non-null    int32
 14  Weekday            250 non-null    object
dtypes: datetime64[ns](1), int32(2), int64(3), object(9)
memory usage: 27.5+ KB
```

```
[12] data['Date'] = pd.to_datetime(data['Date'], format='%d-%m-%y')
```

```
[15] data['Year'] = data['Date'].dt.year
     data['Month'] = data['Date'].dt.month_name()
     data['Day'] = data['Date'].dt.day
     data['Weekday'] = data['Date'].dt.day_name()
```

```
[16] data.head()
```

	Order ID	Date	Product	Category	Price	Quantity	Total Sales	Customer Name	Customer Location	Payment Method	Status	Year	Month	Day	Weekday
0	ORD0001	2025-03-14	Running Shoes	Footwear	60	3	180	Emma Clark	New York	Debit Card	Cancelled	2025	March	14	Friday
1	ORD0002	2025-03-20	Headphones	Electronics	100	4	400	Emily Johnson	San Francisco	Debit Card	Pending	2025	March	20	Thursday
2	ORD0003	2025-02-15	Running Shoes	Footwear	60	2	120	John Doe	Denver	Amazon Pay	Cancelled	2025	February	15	Saturday
3	ORD0004	2025-02-19	Running Shoes	Footwear	60	3	180	Olivia Wilson	Dallas	Credit Card	Pending	2025	February	19	Wednesday
4	ORD0005	2025-03-10	Smartwatch	Electronics	150	3	450	Emma Clark	New York	Debit Card	Pending	2025	March	10	Monday

Next steps: Generate code with data | View recommended plots | New interactive sheet

```
[11] Order_status = data['Status'].value_counts()
     print(Order_status)
```

```
Status
Completed    88
Pending      85
Cancelled    77
Name: count, dtype: int64
```

We analyzed the Order Status column to categorize and summarize the status of each order as Completed, Pending, or Cancelled.

```
[24] Total_revenue_by_Month = data.groupby('Month')['Total Sales'].sum().sort_values(ascending=False)
     print(Total_revenue_by_Month)
```

```
Month
February    122695
March       117730
April         3420
Name: Total Sales, dtype: int64
```

By grouping the data by month and summing the revenue, we discovered that February recorded the highest sales among the three months—February, March, and April.

```
Total_revenue_by_Category = data.groupby('Category')['Total Sales'].sum().sort_values(ascending=False)
print(Total_revenue_by_Category)
```

```
Category
Electronics        129950
Home Appliances    105000
Footwear             4320
Clothing             3540
Books                1035
Name: Total Sales, dtype: int64
```

While analyzing sales by product category, we found that Electronics and Home Appliances emerged as the top-performing categories, generating the highest revenue.

```
[31] Total_revenue_by_Product = data.groupby('Product')['Total Sales'].sum().sort_values(ascending=False).head(5)
     print(Total_revenue_by_Product)
```

```
Product
Refrigerator       78000
Laptop             58400
Smartphone         48500
Washing Machine    27000
Smartwatch         15750
Name: Total Sales, dtype: int64
```

Top 5 Best-Selling Products:

- Refrigerator led with ₹78,000 in sales, followed by the Laptop (₹58,400), Smartphone (₹48,500), Washing Machine (₹27,000), and Smartwatch (₹15,750).

```
[30] Total_revenue_by_Location = data.groupby('Customer Location')['Total Sales'].sum().sort_values(ascending=False).head(3)
     print(Total_revenue_by_Location)
```

```
Customer Location
Miami      31700
Denver     29785
Houston    28390
Name: Total Sales, dtype: int64
```

In the analysis of total revenue by customer location, Miami recorded the highest sales, leading all other cities in revenue contribution. Denver followed as the second-highest performing location.

```
Total_revenue_by_Payment_method = data.groupby('Payment Method')['Total Sales'].sum().sort_values(ascending=False)
print(Total_revenue_by_Payment_method)
```

```
Payment Method
PayPal         69645
Credit Card    61595
Gift Card      47955
Amazon Pay     32750
Debit Card     31900
Name: Total Sales, dtype: int64
```

During the analysis of total sales by payment method, PayPal emerged as the most preferred option, generating the highest revenue.

```
[39] Week_days_sales_distribution = data.groupby('Weekday')['Total Sales'].sum().sort_values(ascending=False).head(3)
     print(Week_days_sales_distribution)
```

```
Weekday
Monday      42975
Tuesday     42650
Thursday    40815
Name: Total Sales, dtype: int64
```

Using Python scripting, we conducted an in-depth analysis of the dataset to understand key aspects such as sales distribution across various factors including location, payment methods, monthly sales trends, and product categories. The data was clean, with no missing or null values, which ensured accurate and reliable analysis. We will now proceed with the data visualization phase in Power BI to represent these insights effectively and support data-driven decision-making.

AMAZON SALES ANALYSIS 2025

244K	250	256	31%	975
Total revenue	Total orders	Total quantity sold	Cancellation rate	Avg order value

PayPal
69645
Total reve...

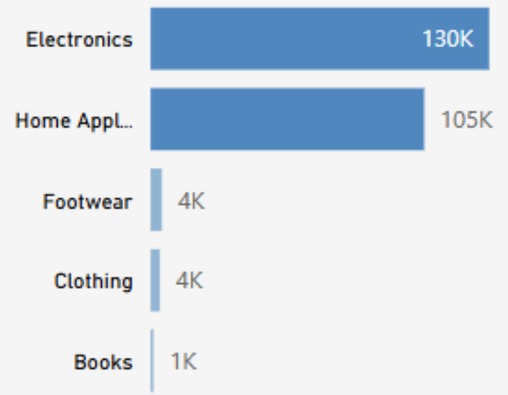
Customer Location
- [ ] Boston
- [ ] Chicago

Category
- [ ] Books
- [ ] Clothing

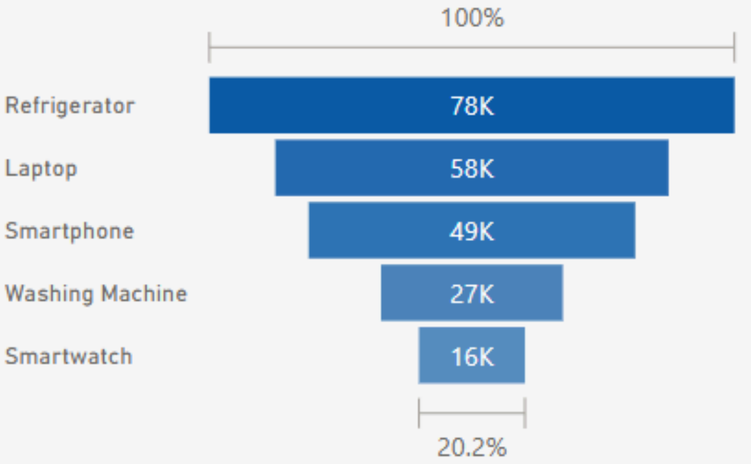
Month: April | February | March

Status: Cancelled | Completed | Pending

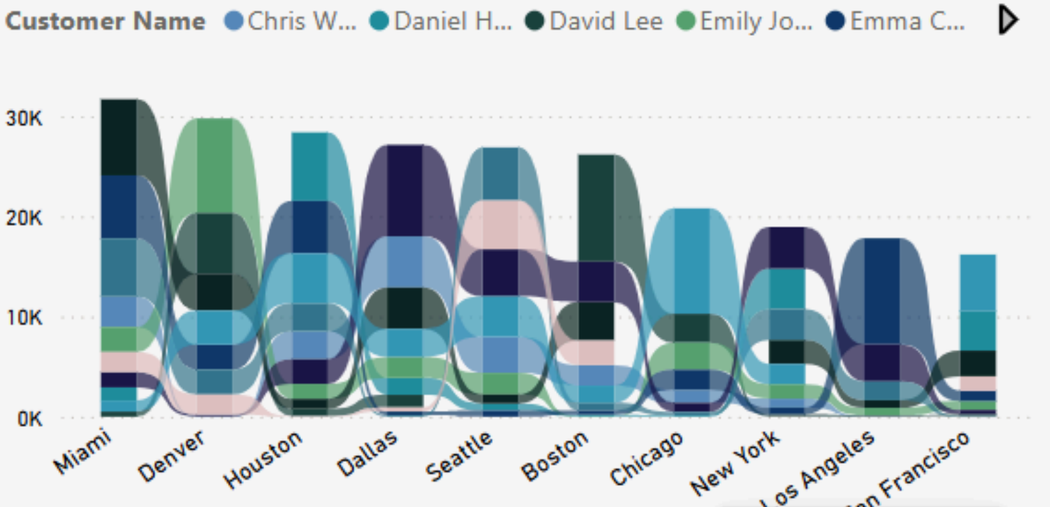
Total revenue by Product Category



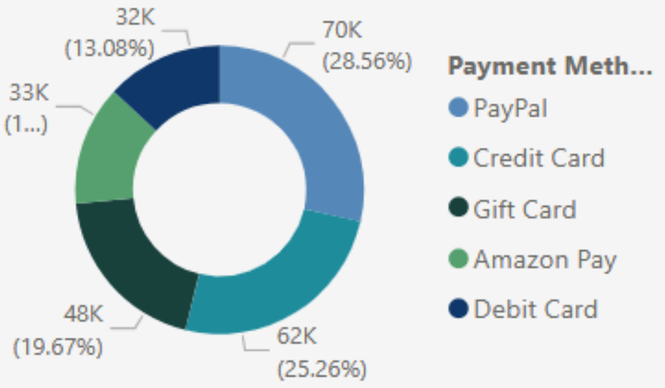
Top 5 Products with highest Revenue



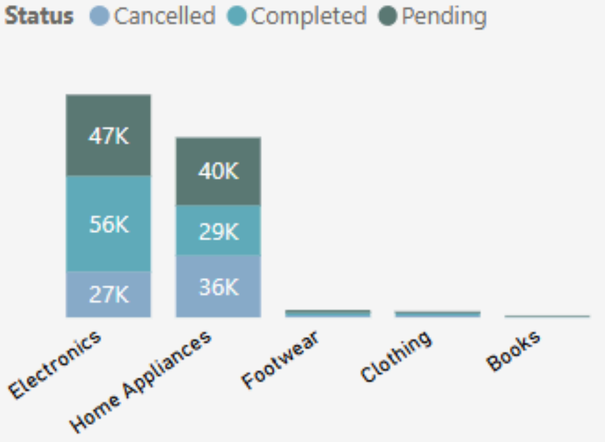
Total revenue by Customer Location and Customer Name



Sum of Total Sales by Payment Method



Sum of Total Sales by Category and Status



Total revenue by Order Month

At the lowest level of data