AMAZON DATSET SALES ANALYSIS

The project revolves around analyzing a dataset containing order information from an online platform, specifically focusing on order fulfillment statuses across various geographical regions. The primary goal is to extract meaningful insights regarding shipment operations, category popularity, and customer behavior trends.

Objectives:

- Identify Delivery and Cancellation Trends: Determine which cities have the highest delivery and cancellation rates. This helps in understanding logistical strengths and weaknesses in specific locations.
- Ascertain Popular Categories by Region: Utilize SQL queries to find the most popular product categories in each state, providing insights into regional consumer preferences which can inform stock and marketing strategies.
- Assess Shipping Service Levels: Analyze how frequently different shipping speeds (like expedited shipping) are chosen across various categories, aiding in optimizing shipping operations and customer satisfaction.

Methods:

- Data Querying and Manipulation: Using SQL, data is queried to perform groupings, aggregations, and rankings to discern patterns in order fulfillment and cancellations. This involves:
- Calculating the number of orders by status (e.g., delivered, canceled) per city.
- Identifying the most common categories per state and evaluating how shipment methods correlate with product categories.
- Ranking and partitioning data to focus analysis on specific geographical or categorical dimensions.

Tools Used:

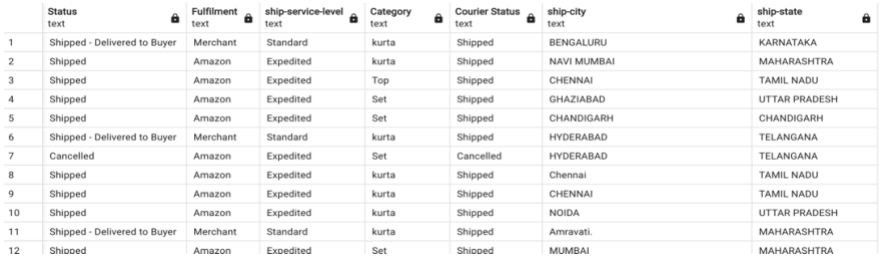
- SQL: For data extraction and manipulation. SQL queries are crafted to handle different aspects of the data, such as filtering, counting, and ranking, tailored to extract specific insights like the most popular categories or cities with the highest order fulfillment.
- pgAdmin (PostgreSQL): The database management tool used for handling SQL queries, facilitating data analysis through a graphical interface and SQL command utility.

Outcomes:

The analysis will reveal:

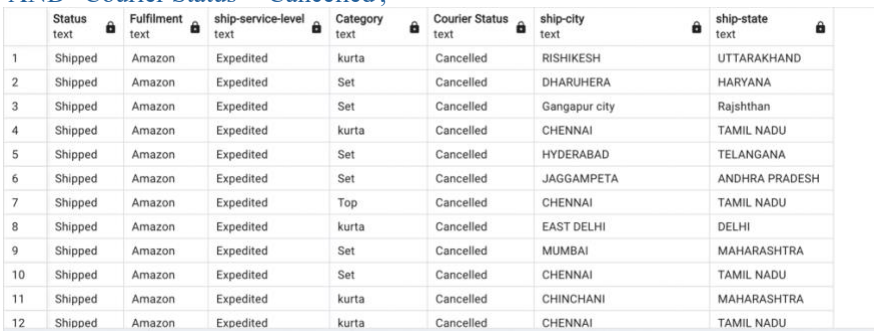
- Key cities that excel or falter in order delivery and cancellations, suggesting areas for logistical improvements or targeted customer service enhancements.
- Categories that dominate in certain states, which could guide inventory management and promotional campaigns.
- Insights into how different shipping options are preferred across categories, potentially influencing how shipping policies are structured to enhance customer satisfaction.
- This project empowers the business to make data-driven decisions, enhancing operational efficiency and aligning product offerings with consumer demands. It highlights the value of SQL in performing complex analyses and the strategic importance of understanding market trends at a granular level.

1. Selecting all columns from the table amazon
```
Select * from amazon;
```



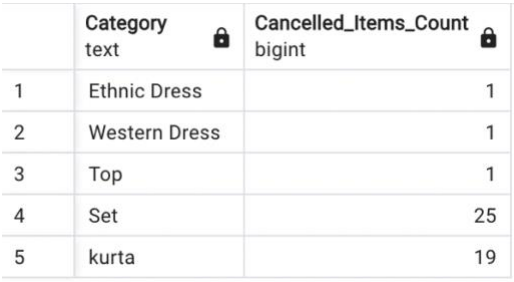
	Status text	Fulfilment text	ship-service-level text	Category text	Courier Status text	ship-city text	ship-state text
1	Shipped - Delivered to Buyer	Merchant	Standard	kurta	Shipped	BENGALURU	KARNATAKA
2	Shipped	Amazon	Expedited	kurta	Shipped	NAVI MUMBAI	MAHARASHTRA
3	Shipped	Amazon	Expedited	Top	Shipped	CHENNAI	TAMIL NADU
4	Shipped	Amazon	Expedited	Set	Shipped	GHAZIABAD	UTTAR PRADESH
5	Shipped	Amazon	Expedited	Set	Shipped	CHANDIGARH	CHANDIGARH
6	Shipped - Delivered to Buyer	Merchant	Standard	kurta	Shipped	HYDERABAD	TELANGANA
7	Cancelled	Amazon	Expedited	Set	Cancelled	HYDERABAD	TELANGANA
8	Shipped	Amazon	Expedited	kurta	Shipped	Chennai	TAMIL NADU
9	Shipped	Amazon	Expedited	kurta	Shipped	CHENNAI	TAMIL NADU
10	Shipped	Amazon	Expedited	kurta	Shipped	NOIDA	UTTAR PRADESH
11	Shipped - Delivered to Buyer	Merchant	Standard	kurta	Shipped	Amravati.	MAHARASHTRA
12	Shipped	Amazon	Expedited	Set	Shipped	MUMBAI	MAHARASHTRA

2. SQL query to identify goods that have been marked as shipped but were later cancelled by the customer, based on the courier status variable
```
 select * from amazon
 where "Status"='Shipped' AND "Courier Status"='Cancelled';
```



	Status text	Fulfilment text	ship-service-level text	Category text	Courier Status text	ship-city text	ship-state text
1	Shipped	Amazon	Expedited	kurta	Cancelled	RISHIKESH	UTTARAKHAND
2	Shipped	Amazon	Expedited	Set	Cancelled	DHARUHERA	HARYANA
3	Shipped	Amazon	Expedited	Set	Cancelled	Gangapur city	Rajshthan
4	Shipped	Amazon	Expedited	kurta	Cancelled	CHENNAI	TAMIL NADU
5	Shipped	Amazon	Expedited	Set	Cancelled	HYDERABAD	TELANGANA
6	Shipped	Amazon	Expedited	Set	Cancelled	JAGGAMPETA	ANDHRA PRADESH
7	Shipped	Amazon	Expedited	Top	Cancelled	CHENNAI	TAMIL NADU
8	Shipped	Amazon	Expedited	kurta	Cancelled	EAST DELHI	DELHI
9	Shipped	Amazon	Expedited	Set	Cancelled	MUMBAI	MAHARASHTRA
10	Shipped	Amazon	Expedited	Set	Cancelled	CHENNAI	TAMIL NADU
11	Shipped	Amazon	Expedited	kurta	Cancelled	CHINCHANI	MAHARASHTRA
12	Shipped	Amazon	Expedited	kurta	Cancelled	CHENNAI	TAMIL NADU

3. SQL query that counts cancelled items grouped by category, where the items were initially marked as shipped but later cancelled
```
SELECT "Category", COUNT(*) AS "Cancelled_Items_Count"
FROM amazon
WHERE "Status" = 'Shipped' AND "Courier Status" = 'Cancelled'
GROUP BY "Category";
```



	Category text	Cancelled_Items_Count bigint
1	Ethnic Dress	1
2	Western Dress	1
3	Top	1
4	Set	25
5	kurta	19

4. Query will give you a complete list of categories, each with the number of expedited orders, sorted from the most to the least
```
SELECT "Category", COUNT(*) AS "Expedited_Order_Count"
FROM amazon
WHERE "ship-service-level" = 'Expedited'
GROUP BY "Category"
ORDER BY "Expedited_Order_Count" DESC;
```

	Category text	Expedited_Order_Count bigint
1	Set	17972
2	kurta	16786
3	Top	4418
4	Western Dress	4323
5	Blouse	391
6	Ethnic Dress	366
7	Bottom	151
8	Saree	71

5. A list showing how many orders were fulfilled by the merchant and Amazon in each category.

```
SELECT "Category", "Fulfilment", COUNT("Category") AS "Order_Count"
FROM amazon."Orders"
WHERE "Fulfilment" IN ('Merchant', 'Amazon')
GROUP BY "Category", "Fulfilment"
ORDER BY "Category", "Fulfilment";
```

	Category text	Fulfilment text	Order_Count bigint
1	Blouse	Amazon	400
2	Blouse	Merchant	117
3	Bottom	Amazon	154
4	Bottom	Merchant	79
5	Ethnic Dress	Amazon	372
6	Ethnic Dress	Merchant	134
7	Saree	Amazon	71
8	Saree	Merchant	15
9	Set	Amazon	18215
10	Set	Merchant	6423
11	Top	Amazon	4455
12	Top	Merchant	1084

6. Cities with maximum cancellations

```
SELECT "ship-city", COUNT("ship-city") AS "Cancel_Count"
FROM amazon
WHERE "Status" = 'Cancelled'
GROUP BY "ship-city"
ORDER BY "Cancel_Count" DESC
```

	ship-city text	Cancel_Count bigint
1	BENGALURU	471
2	HYDERABAD	364
3	CHENNAI	229
4	MUMBAI	216
5	NEW DELHI	213
6	PUNE	165
7	KOLKATA	101
8	THANE	88
9	GURUGRAM	85
10	LUCKNOW	61
11	PATNA	51
12	NOIDA	50

7. Cities with Maximum deliveries

```
SELECT "ship-city", COUNT(*) AS "Delivered_Count"
FROM amazon
WHERE "Status" = 'Shipped - Delivered to Buyer'
GROUP BY "ship-city"
ORDER BY "Delivered_Count" DESC;
```

	ship-city text	Delivered_Count bigint
1	BENGALURU	1222
2	HYDERABAD	877
3	MUMBAI	783
4	NEW DELHI	760
5	CHENNAI	564
6	PUNE	474
7	KOLKATA	307
8	GURUGRAM	217
9	THANE	203
10	LUCKNOW	198
11	NOIDA	165
12	AHMEDABAD	165

8. Cities with Maximum cancellations

```
SELECT "ship-city", COUNT(*) AS "Cancellation_Count"
FROM amazon
WHERE "Status" = 'Cancelled'
GROUP BY "ship-city"
ORDER BY "Cancellation_Count" DESC
```

	ship-city text	Cancellation_Count bigint
1	BENGALURU	471
2	HYDERABAD	364
3	CHENNAI	229
4	MUMBAI	216
5	NEW DELHI	213
6	PUNE	165
7	KOLKATA	101
8	THANE	88
9	GURUGRAM	85
10	LUCKNOW	61
11	PATNA	51
12	NOIDA	50

9. SQL query to find out category with maximum occurrences

```
WITH CategoryCounts AS (
SELECT "ship-state", "Category", COUNT(*) AS "Category_Count"
FROM amazon
GROUP BY "ship-state", "Category"
),
MaxCategories AS (
SELECT "ship-state", "Category", "Category_Count",
  RANK() OVER (PARTITION BY "ship-state" ORDER BY "Category_Count" DESC) AS "Rank"
FROM CategoryCounts
)
SELECT "ship-state", "Category", "Category_Count"
FROM MaxCategories
WHERE "Rank" = 1;
```

	ship-state text	Category text	Category_Count bigint
1	ANDAMAN & NICOBAR	Set	48
2	ANDHRA PRADESH	kurta	1155
3	APO	Set	1
4	AR	Top	1
5	ARUNACHAL PRADESH	Set	30
6	ASSAM	Set	375
7	Arunachal Pradesh	Set	2
8	BIHAR	Set	601
9	Bihar	Set	10
10	CHANDIGARH	Set	74
11	CHHATTISGARH	Set	222
12	Chandigarh	kurta	3
13	DADRA AND NAGAR	Set	12
14	DELHI	Set	1562
15	Delhi	kurta	32
16	GOA	kurta	238
17	Goa	kurta	7
18	Gujarat	Set	810
19	HARYANA	Set	1047
20	HIMACHAL PRADESH	Set	196
21	JAMMU & KASHMIR	Set	180