Amazon Delivery Data Analysis Report

1. Introduction

Overview:
This report provides an analysis of the Amazon Delivery dataset. Our goal is to understand how various factors—like traffic, weather, and time of day—affect delivery performance. We also look into patterns behind delivery times, areas of operation, and customer order patterns.

Key Questions Addressed:

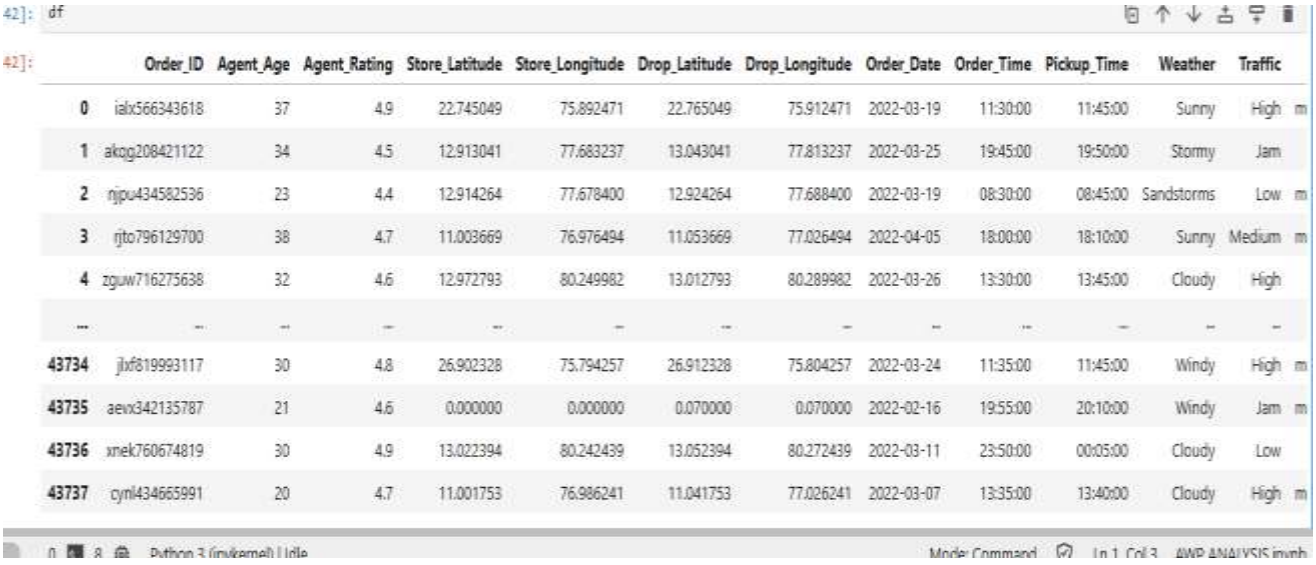
- How fast are deliveries happening?
- What impact do external factors such as traffic or weather have on delivery times?
- Are there any common trends in order timing or geographic locations?

---

2. Data Overview

Records and Columns: The dataset includes delivery orders with details such as order ID, delivery time, agent information (age and rating), store and drop-off locations, order timings, weather, traffic, delivery vehicle, service area, and product category.

- **Understanding the Data:** Each column helps us explore different aspects of the delivery process, from timing and location to conditions that may affect performance.



42]: df

42]:

	Order_ID	Agent_Age	Agent_Rating	Store_Latitude	Store_Longitude	Drop_Latitude	Drop_Longitude	Order_Date	Order_Time	Pickup_Time	Weather	Traffic	
0	ialx566343618	37	4.9	22.745049	75.892471	22.765049	75.912471	2022-03-19	11:30:00	11:45:00	Sunny	High	m
1	akqg208421122	34	4.5	12.913041	77.683237	13.043041	77.813237	2022-03-25	19:45:00	19:50:00	Stormy	Jam	
2	njpu434582536	23	4.4	12.914264	77.678400	12.924264	77.688400	2022-03-19	08:30:00	08:45:00	Sandstorms	Low	m
3	rjto796129700	38	4.7	11.003669	76.976494	11.053669	77.026494	2022-04-05	18:00:00	18:10:00	Sunny	Medium	m
4	zguw716275638	32	4.6	12.972793	80.249982	13.012793	80.289982	2022-03-26	13:30:00	13:45:00	Cloudy	High	
...	...	...	...	...	...	...	...	...	...	...	...	...	
43734	jlxf819993117	30	4.8	26.902328	75.794257	26.912328	75.804257	2022-03-24	11:35:00	11:45:00	Windy	High	m
43735	aevx342135787	21	4.6	0.000000	0.000000	0.070000	0.070000	2022-02-16	19:55:00	20:10:00	Windy	Jam	m
43736	xnek760674819	30	4.9	13.022394	80.242439	13.052394	80.272439	2022-03-11	23:50:00	00:05:00	Cloudy	Low	
43737	cynl434665991	20	4.7	11.001753	76.986241	11.041753	77.026241	2022-03-07	13:35:00	13:40:00	Cloudy	High	m

Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 3 AWP ANALYSIS.ipynb

3. Data Cleaning

What We Did:

- **Remove Incomplete or Duplicate Data:** We carefully checked the information to ensure there were no repeated or missing entries that could affect the analysis.
- **Handling Extreme Values:**
  To ensure the analysis reflects the usual delivery experience, we removed unusually high or low delivery times (extreme cases) so that our charts and insights represent typical performance.

Why This Matters:
Cleaning the data means our insights are based on accurate, reliable information. It helps us better understand normal delivery patterns without distraction from anomalies.

```
2]: df_filtered
```

2]:

	Order_ID	Agent_Age	Agent_Rating	Store_Latitude	Store_Longitude	Drop_Latitude	Drop_Longitude	Order_Date	Order_Time	Pickup_Time	Weather	Traffic	
0	ialx566343618	37	4.9	22.745049	75.892471	22.765049	75.912471	2022-03-19	11:30:00	11:45:00	Sunny	High	m
1	akqg208421122	34	4.5	12.913041	77.683237	13.043041	77.813237	2022-03-25	19:45:00	19:50:00	Stormy	Jam	
2	rjpu434582536	23	4.4	12.914264	77.678400	12.924264	77.688400	2022-03-19	08:30:00	08:45:00	Sandstorms	Low	m
3	rjto796129700	38	4.7	11.003669	76.976494	11.053669	77.026494	2022-04-05	18:00:00	18:10:00	Sunny	Medium	m
4	zguw716275638	32	4.6	12.972793	80.249982	13.012793	80.289982	2022-03-26	13:30:00	13:45:00	Cloudy	High	
...	...	...	...	...	...	...	...	...	...	...	...	...	
43734	jlof819993117	30	4.8	26.902328	75.794257	26.912328	75.804257	2022-03-24	11:35:00	11:45:00	Windy	High	m
43735	aevx342135787	21	4.6	0.000000	0.000000	0.070000	0.070000	2022-02-16	19:55:00	20:10:00	Windy	Jam	m
43736	xnek760674819	30	4.9	13.022394	80.242439	13.052394	80.272439	2022-03-11	23:50:00	00:05:00	Cloudy	Low	
43737	cynl434665991	20	4.7	11.001753	76.986241	11.041753	77.026241	2022-03-07	13:35:00	13:40:00	Cloudy	High	m
43738	nsyz997960170	23	4.9	23.351058	85.325731	23.431058	85.405731	2022-03-02	17:10:00	17:15:00	Fog	Medium	

43660 rows × 16 columns

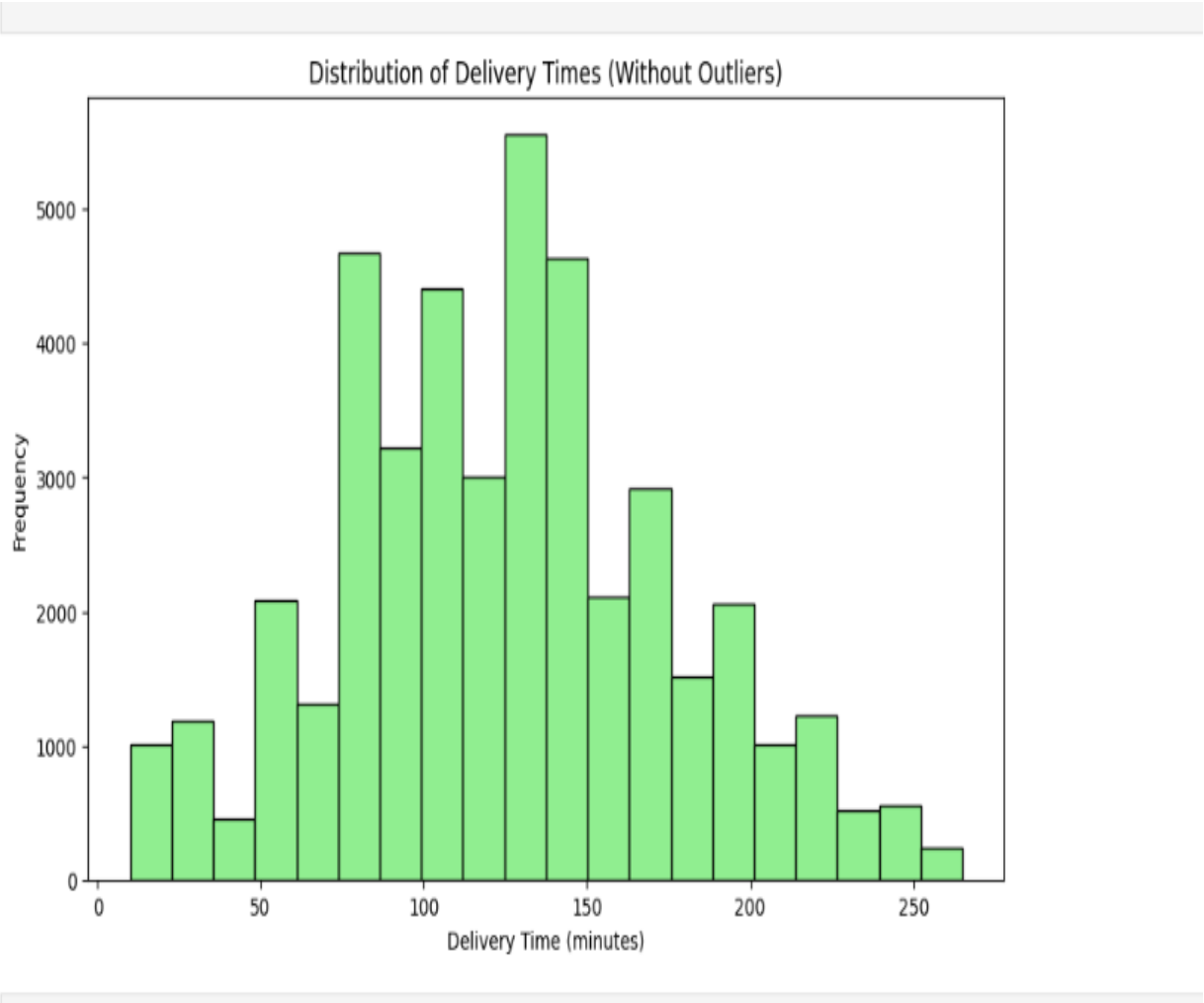
4. Visual Insights

4.1. Delivery Time Distribution

What We Explored:
We looked at how long deliveries take in general. The focus was on understanding the typical range of delivery times while omitting extreme cases.

Insight:
Most deliveries fall within a consistent time range, with only a few taking significantly longer. This tells us that, under normal conditions, the delivery process is fairly predictable.

Histogram of Delivery Times:

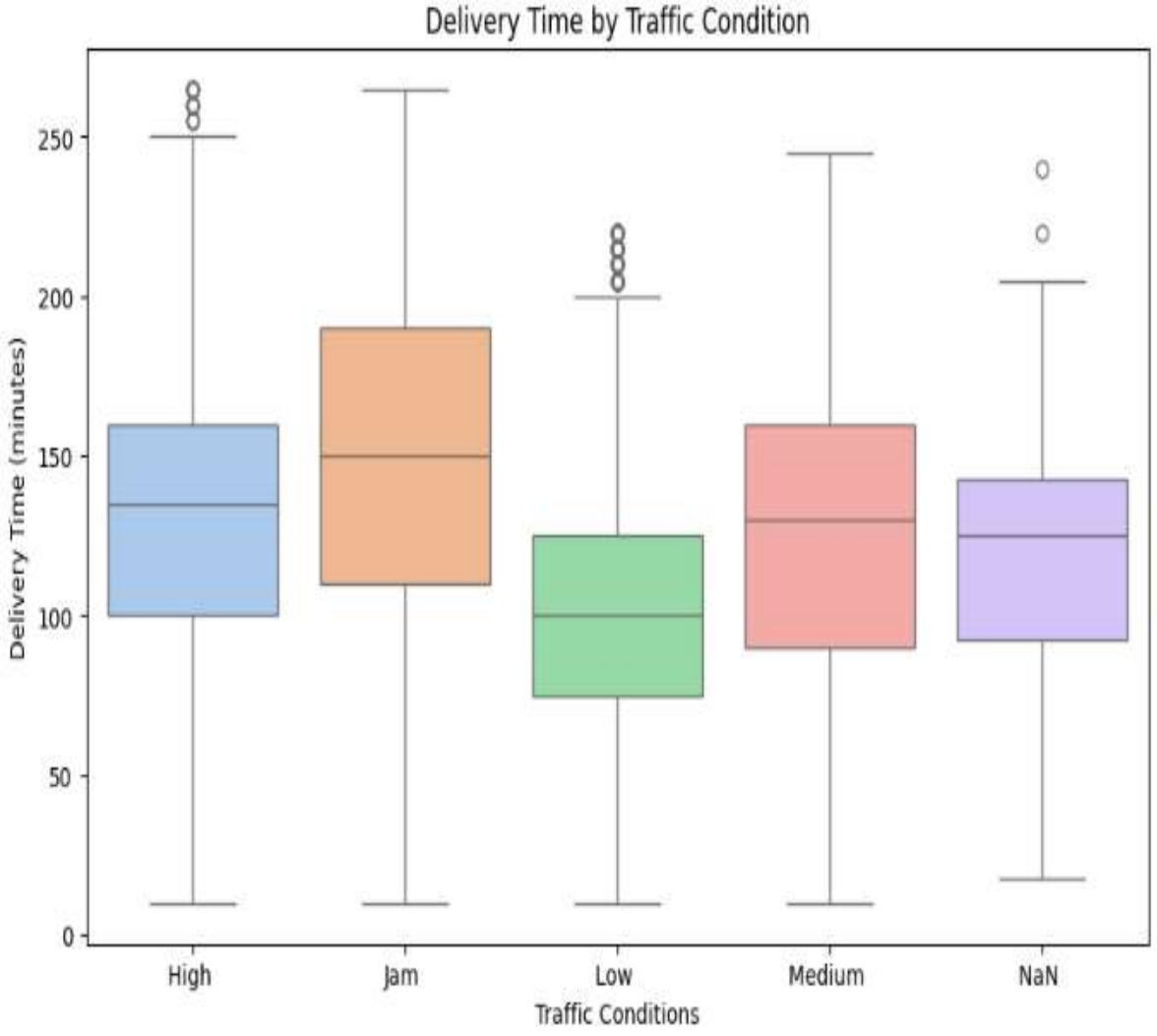


4.2. Impact of Traffic on Delivery Time

What We Explored:
We compared delivery times under different traffic conditions, such as light, moderate, and heavy traffic.

Insight:
The analysis shows that deliveries in heavy traffic tend to take longer, while those in lighter traffic run quicker. This insight can help in planning delivery routes and setting realistic expectations during rush hours.

Boxplot by Traffic Conditions:

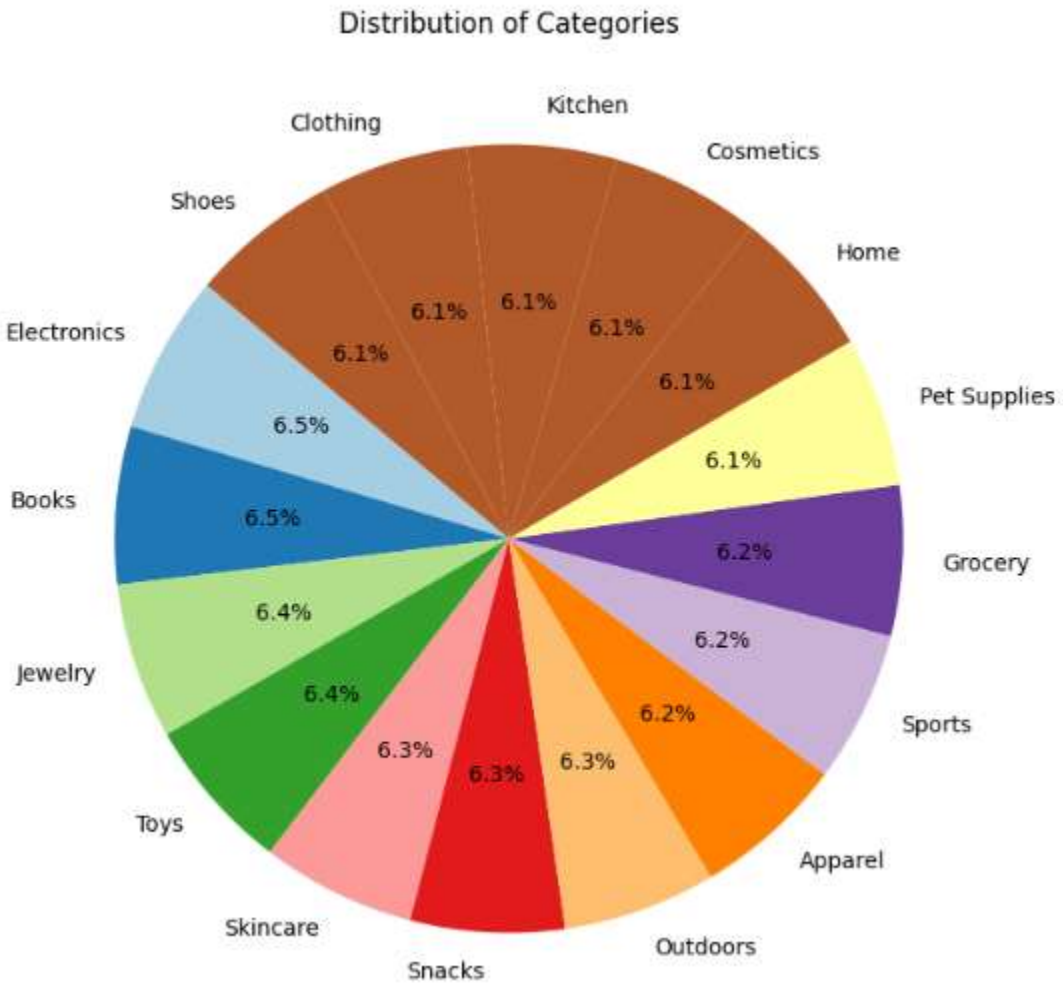


4.3. Category Distribution

What We Explored:
We examined how the different product categories are distributed across orders. This gives us an understanding of which types of products are most frequently delivered.

Insight:
A significant portion of deliveries falls into a few key categories, highlighting the areas of high demand. Knowing this helps prioritize efforts and allocate resources effectively.

Pie Chart of Delivery Categories:

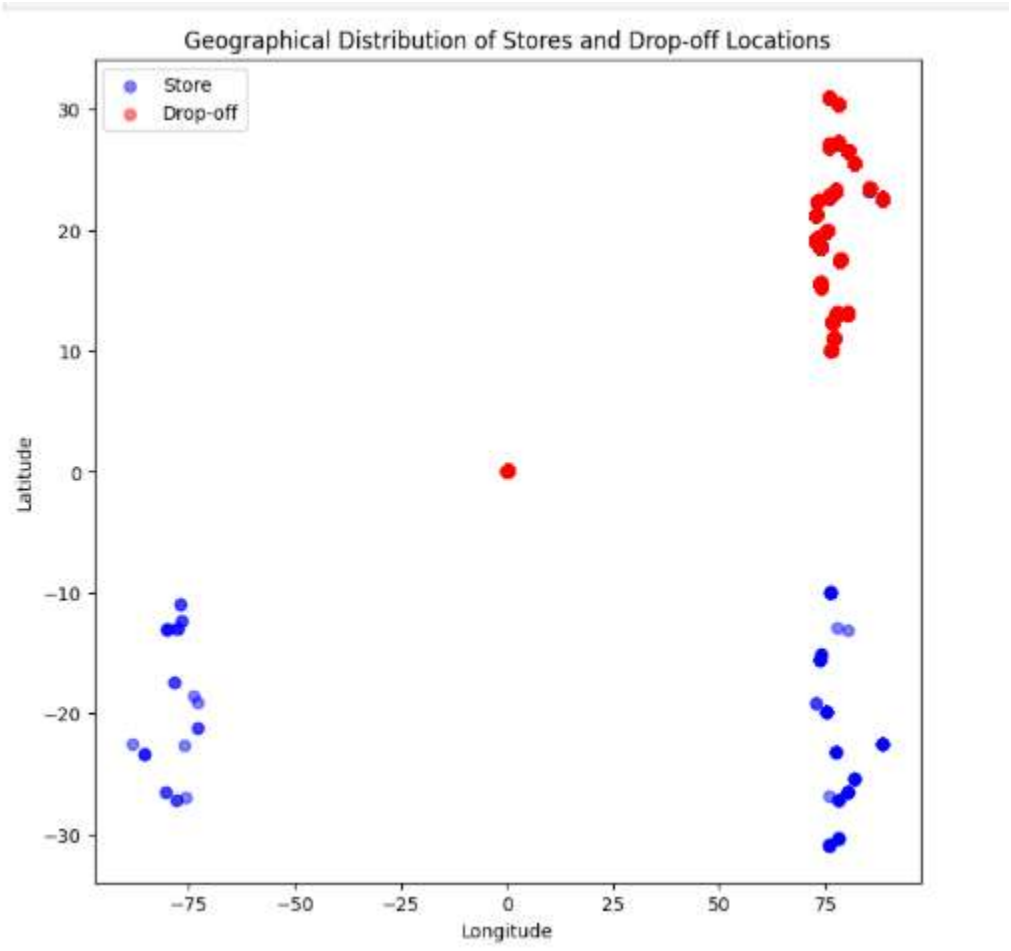


4.4. Geographical Distribution

What We Explored:
I mapped the locations of the stores and drop-off points to see geographic patterns that might impact delivery performance.

Insight:
The spatial analysis reveals clusters where distances between the store and the drop-off locations may be contributing to longer delivery times. This points to areas where route optimization could be most beneficial.

Scatter Plot or Map of Locations:



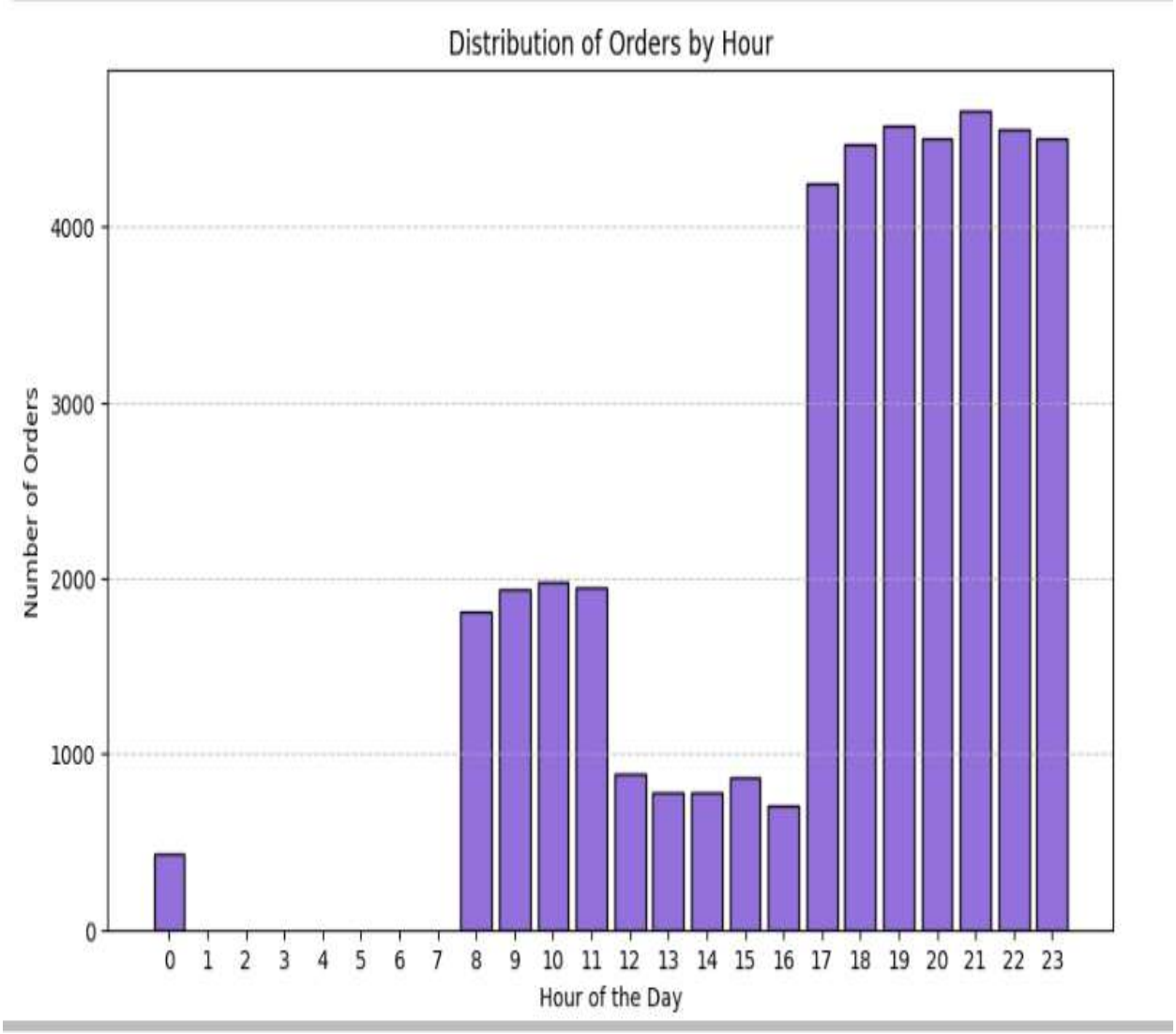
4.5. Order Timing Patterns

What We Explored:

I analyzed at what times of day orders are placed to uncover peak periods during the day.

Insight:
The analysis indicates specific hours where orders spike, suggesting periods of high demand. Recognizing these peak hours can lead to better staffing and scheduling to handle the rush.

Bar/Line Chart Showing Orders by Hour:



5. Conclusions & Recommendations

Key Insights:

- **Consistent Delivery Performance:**
  Most orders are delivered within a predictable timeframe under normal conditions.
- **Traffic Effects:**
  High traffic is closely related to longer delivery times. Adjusting schedules during these periods could help reduce delays.
- **Popular Product Categories:**
  A few categories dominate the delivery orders, suggesting where to focus service improvements.
- **Geographical Trends:**
  Specific regions show a pattern of longer distances between stores and drop-off points, which might benefit from route adjustments.
- **Peak Order Times:**
  Knowing the busiest times of day helps in planning workforce allocation and improving customer service during high-demand periods.

Recommendations:

1. **Route Optimization:**
   Consider analyzing specific geographic clusters to design more efficient routes.
2. **Traffic-Aware Scheduling:**
   Adapt delivery schedules based on traffic predictions to minimize delays during peak hours.
3. **Operational Focus on Key Categories:**
   Prioritize resource allocation for the most frequently delivered product categories.
4. **Resource Planning:**
   Enhance staffing during peak order times to ensure timely deliveries and maintain high service quality.

---

6. Final Remarks

This analysis provides a clear picture of the current state of Amazon deliveries by highlighting significant factors that affect performance. These insights pave the way for targeted improvements in route planning, scheduling, and resource management to ultimately enhance customer satisfaction.