Amazon delivery dataset

A2: Collaborate with team members and Develop a Shiny Dashboard Application

Aleksas

Introduction

There are many industries and many areas where a systemized data analysis and visualization can provide key insights for business improvement. In this case, the focus of analysis will be the business problem that could be described as “understanding potential improvements in delivery logistics”, with the company being Amazon and its deliveries. This short report will briefly discuss the data and its cleaning processes, as well as visualizations, insights and recommendations for improvement.

Data Cleaning

The data cleaning process consisted of multiple steps to ensure consistent outputs by the dashboards. Initially, missing values were addressed by removing incomplete rows, including 91 instances where "nan" appeared in the Traffic column. Additionally, outlying values within Delivery_Time were identified through the IQR method, leading to the removal of 79 extreme values, which had values above 170 minutes, or below 2. Further, filtering was applied to eliminate unrealistic agent ages, inconsistent ratings, and invalid order times. To maintain uniformity, data types were standardized, ensuring proper formatting of date/time fields, preserving numerical integrity in numerical variables, and refining categorical non-numerical columns for consistency. Additionally, duplicate records were removed, and categorical inconsistencies were resolved. These systematic steps resulted in a clean, structured dataset, free from missing values, anomalies, and formatting discrepancies, ensuring reliability for subsequent analysis.

Insights and Recommendations for Amazon according to the Amazon Delivery Time Analysis.

The Amazon Delivery Time Analysis Dashboard includes filters for selecting a date range and category (Weather, Traffic, Vehicle, Area, Category). The Bar Chart Tab visualizes the average delivery time by the selected category with percentage labels, using a bar chart for comparison. The Delivery Delay Prediction Dashboard features a Prediction Tab, where users can input traffic, weather, vehicle type, agent rating, delivery area, and order hour to predict delivery time using Linear Regression and Random Forest models. The Model Performance Tab includes a scatter plot comparing both models' actual vs. predicted delivery times.

Impact of Weather on Delivery Time

Cloudy (136.2 min) and foggy (134.1 min) weather contribute to the most significant delivery delays, perhaps due to low visibility and cautious driving. Good weather (101.9 min) results in the quickest deliveries that benefit from optimal road conditions. Stormy (121.7 min) and windy (121.9 min) weather also result in moderate delays. As weather affects efficiency in delivery, it can be minimized by optimizing for adverse conditions through enhanced vehicle performance and scheduling. Incentives such as Weather Surge Pay can offer extra payouts to drivers working in adverse conditions, Other incentives can also help. They could introduce a Flexible Delivery Discount to encourage customers to opt for delayed deliveries during storms in exchange for Amazon credits. Finally, Priority Routes can be implemented using AI-driven routes to make sure that urgent deliveries are prioritized in case of intense weather conditions.

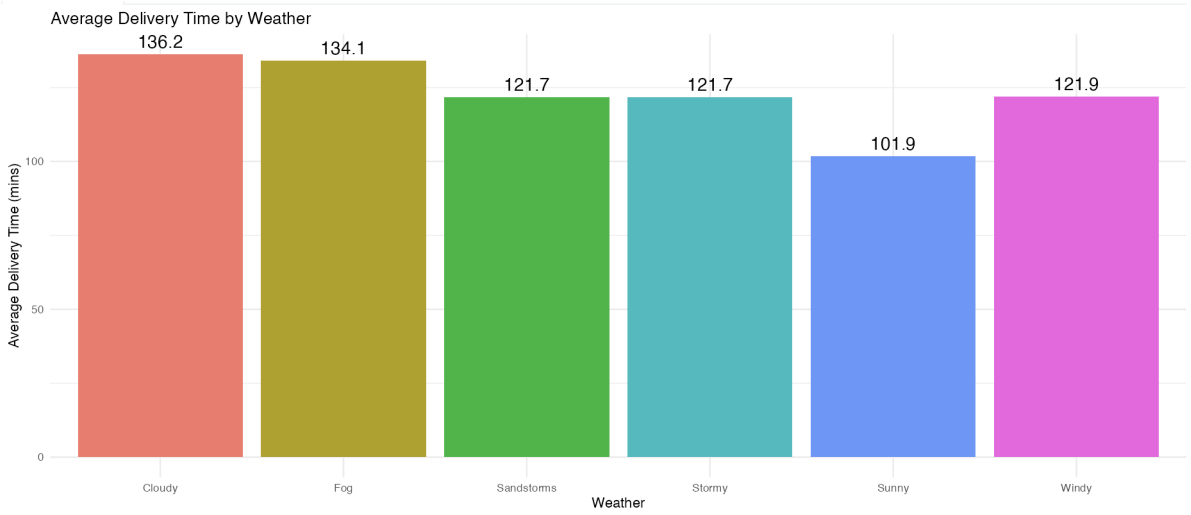

Figure 1 - Average Delivery Time by Weather

Traffic Conditions & Delivery Time

Traffic congestion (143.5 min) creates the largest delays, followed by heavy-traffic (127.8 min) and medium-traffic (126.3 min) conditions, both of which lead to high levels of slowdown. Low-traffic conditions (101.4 min), on the other hand, facilitate the fastest deliveries. This underlines the necessity for real-time traffic surveillance and AI route optimization to alleviate congestion-created delays. In order to minimize delays, AI-driven route planning and real-time traffic monitoring can enhance delivery routes and improve efficiency. Amazon can also adopt incentives such as Driver Bonuses that can reward drivers for completing deliveries faster in high-traffic areas, while Surge Pricing Discounts can encourage customers to opt for off-peak delivery slots. Finally, Smart Routing Rewards can incentivize drivers to consistently follow AI-optimized routes, which will ensure faster and more reliable deliveries.

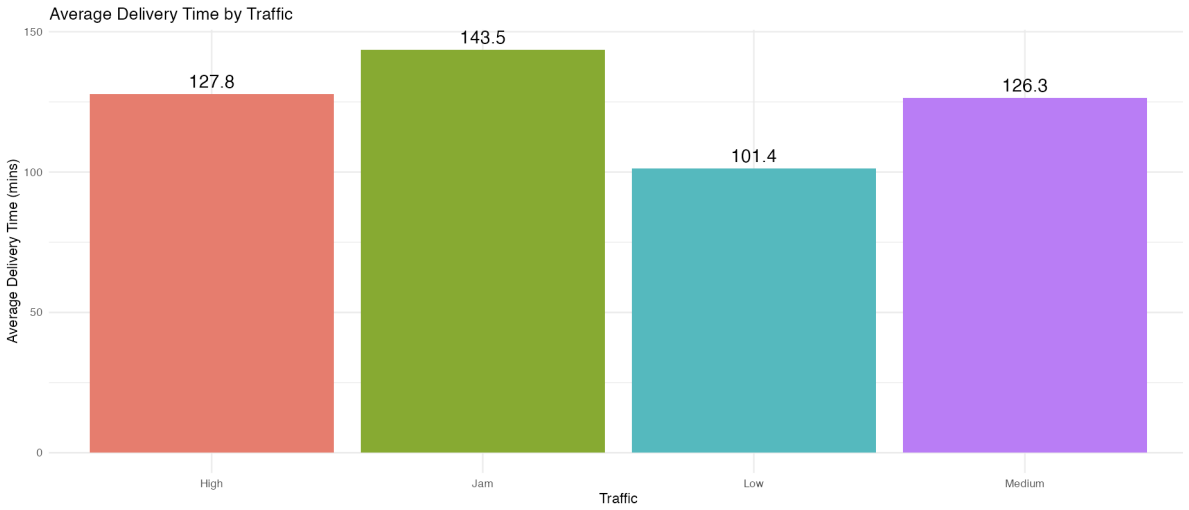

Figure 2 - Average Delivery Time by Traffic

Vehicle Allocation Strategy

Motorcycles have the longest delivery time (128.2 min), most likely due to limited capacity and longer routes. Scooters and vans perform better (115.9 min each), thus they are suitable for faster deliveries. Strategic vehicle deployment is essential, with vans being more suitable for bulk deliveries and scooters for last-mile city delivery. To optimize vehicle allocation, Amazon can introduce incentives such as a Green Fleet Bonus, offering higher payouts for EV and bicycle deliveries, Fuel Reimbursement to support drivers in high-traffic zones, and High-Efficiency Driver Rewards for those consistently maintaining fast deliveries across different vehicle types.

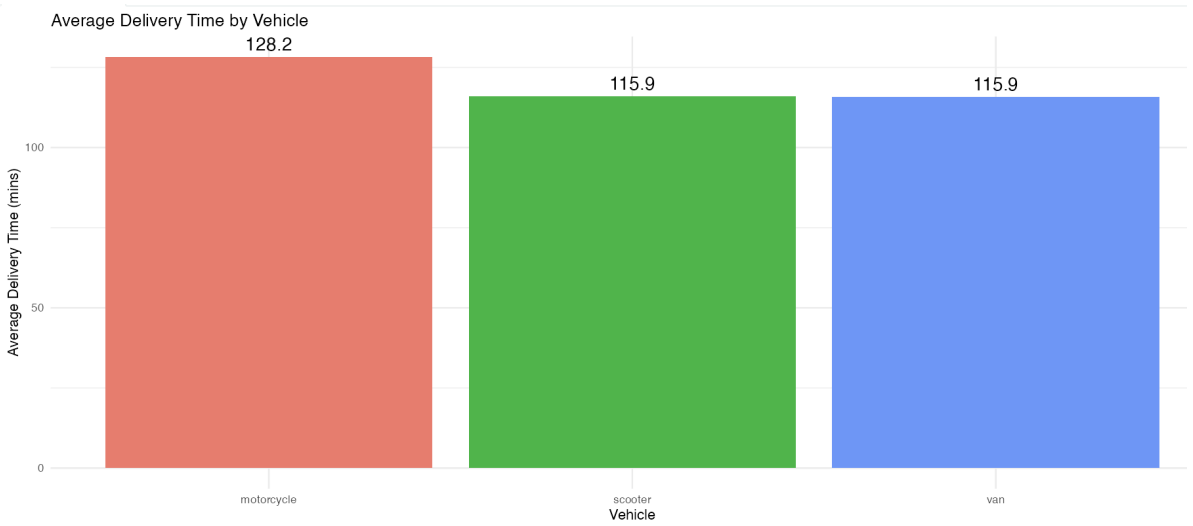

Figure 3 - Average delivery time by vehicle

Regional Delivery Performance

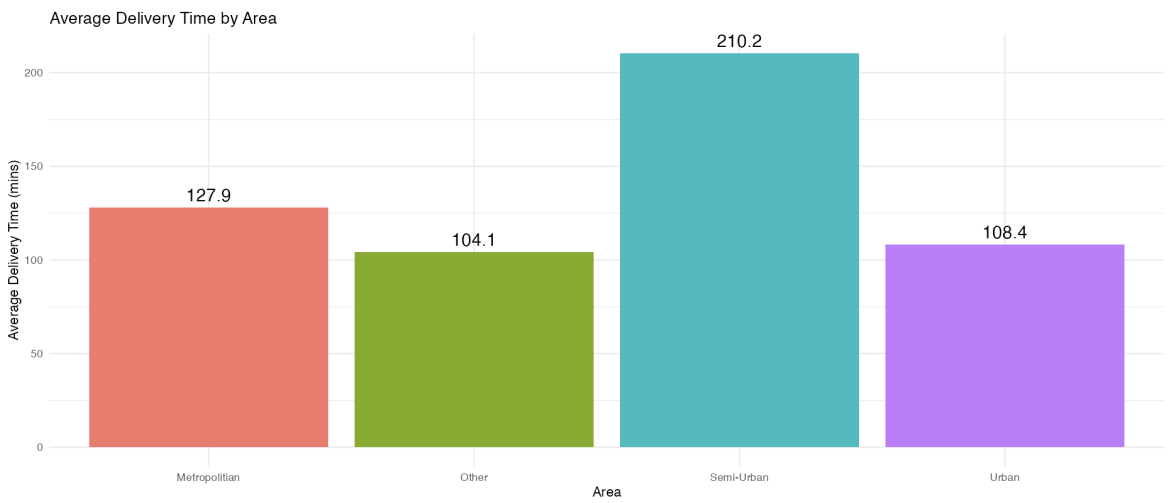

Figure 4 - Average Delivery Time by Area

Metropolitan areas experience the highest delivery delays (74.8%), and urban areas also show significant delays (22.4%), which shows the need for localized distribution solutions. To speed up urban deliveries, Amazon can expand micro-fulfillment hubs, which will reduce last-mile delivery times and increase efficiency. Incentives such as Local Partner Commissions can encourage businesses to offer space for micro-hubs in exchange for a percentage of order revenue. In the same manner, Fast Pickup Discounts can incentivize customers to retrieve orders from Amazon lockers, reducing delivery strain. Finally, a Warehouse Performance Bonus can reward fulfillment centers that consistently meet delivery benchmarks, ensuring faster and more reliable service.

Delivery Time by Product Category

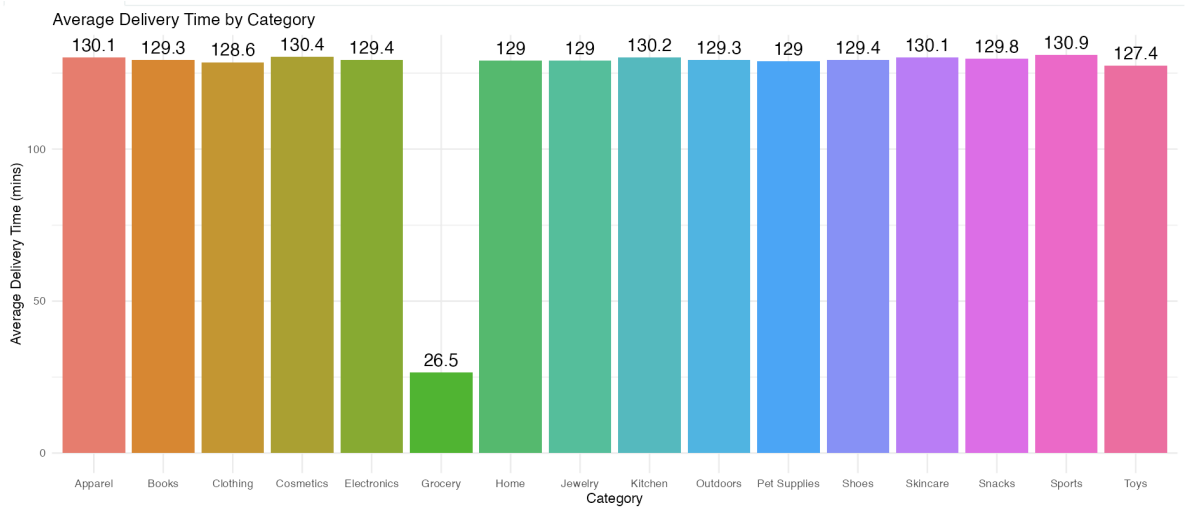

Figure 5 - Average delivery time by category

Semi-urban locations have the highest delivery delays (210.2 min), indicating operational issues in these areas. Metropolitan locations (127.9 min) also suffer from high delays, possibly due to the high volume of deliveries and traffic. Urban (108.4 min) and "Other" locations (104.1 min) have comparatively shorter delivery times. Micro-fulfillment hubs expanded in semi-urban and metropolitan locations can mitigate last-mile delivery delays and increase efficiency overall. To further optimize fulfillment strategies based on category urgency, Amazon can introduce incentives such as Prime Fast Lane, offering priority delivery for Prime members on essential items, Bulk Order Discounts to encourage consolidated shipments, and a Driver Performance Bonus that rewards drivers for on-time grocery and perishable deliveries, ensuring efficiency and customer satisfaction.

Delivery Time by Age group - Line chart

Figure 6 - Delivery Time by Age Group

Here, the delivery time analysis dashboard is a friendly tool designed to explore how the age of the delivery agent may affect the time to take complete deliveries. It may provide an interactive way to analyze real-world data and help businesses and logistics teams make better decisions for workforce management.

As a core dashboard loads a data set and cleans by missing values for ensuring accurate analysis, it offers insight like basic statistics and correlations between agent age and delivery time and highlights the top 5 delivery time records. We can adjust the user and age range of agents using the simple slider and choose to view the correlation value and update our analysis with a click of a button.

This dashboard visually represents data for a scatter plot, which will show that the delivery time varies with age along the trend line to highlight patterns a table is displayed with filtered data for detailed exploration, and additionally it provides a correlation coefficient quantifies the relation between age and delivery efficiency, offering insights into whether older or younger agent will be faster or slower.

This tool is useful for logistic managers and delivery companies looking to optimize the workforce by identifying trends very efficiently. It may turn complex data into clear actionable insights, making it easier to improve performance.

Delivery time vs traffic condition - Boxplot

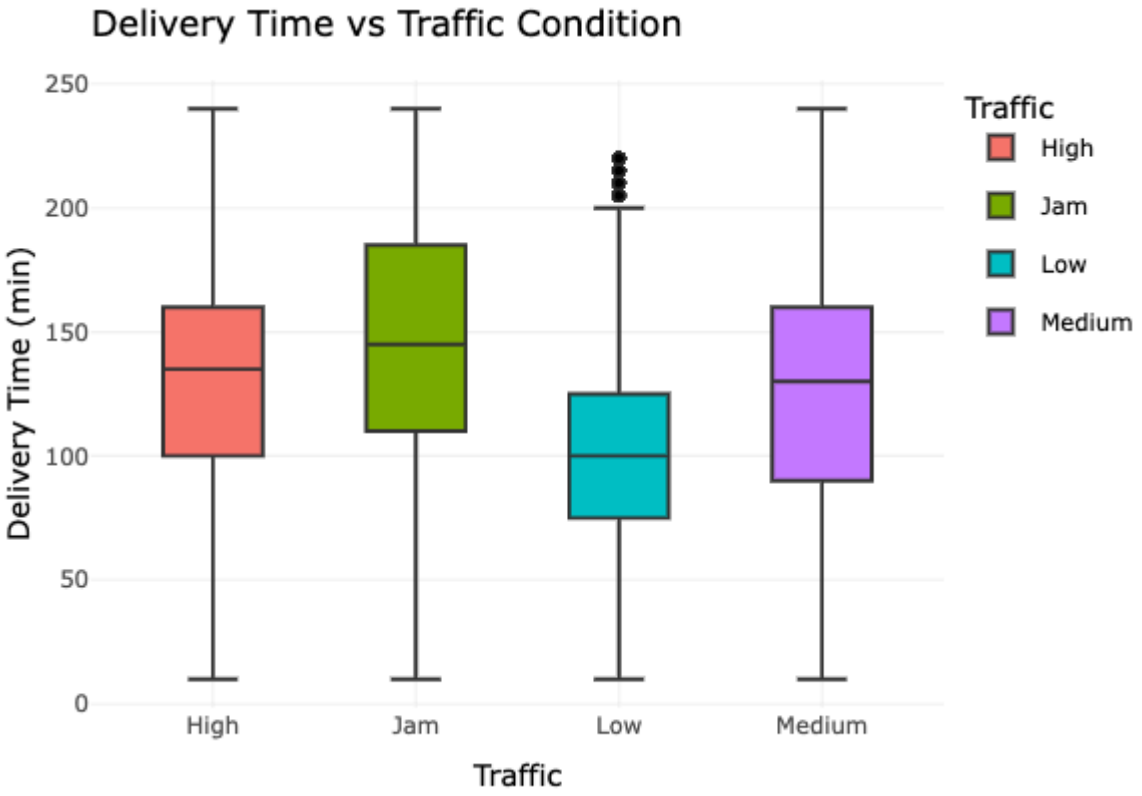

Figure 7 - Delivery Time vs Traffic Condition

The delivery time vs traffic condition compares the different delivery times across different traffic conditions, helping to provide a clear view of how congestion impacts the efficiency of delivery. The boxplot highlights that low traffic conditions lead to short delivery times, with a median of 100 minutes in comparison to when there are traffic jams where the deliveries extend to 150 minutes. Furthermore, scenarios with high traffic show multiple outliers, when some deliveries take longer than

usual, suggesting that traffic congestion has a huge impact on delivery delays. To better improve this, introducing real time traffic rerouting, dynamic pricing models and introducing AI powered dispatching strategies could be adopted to improve the efficiency. This would help encourage customers to schedule deliveries in advance and give them the option to request for certain time slots, which will lead to an enhancement in reliability as you provide them with the option of choosing an alternative time slot if there is peak congestion hours.

Delivery time distribution - Density Plot

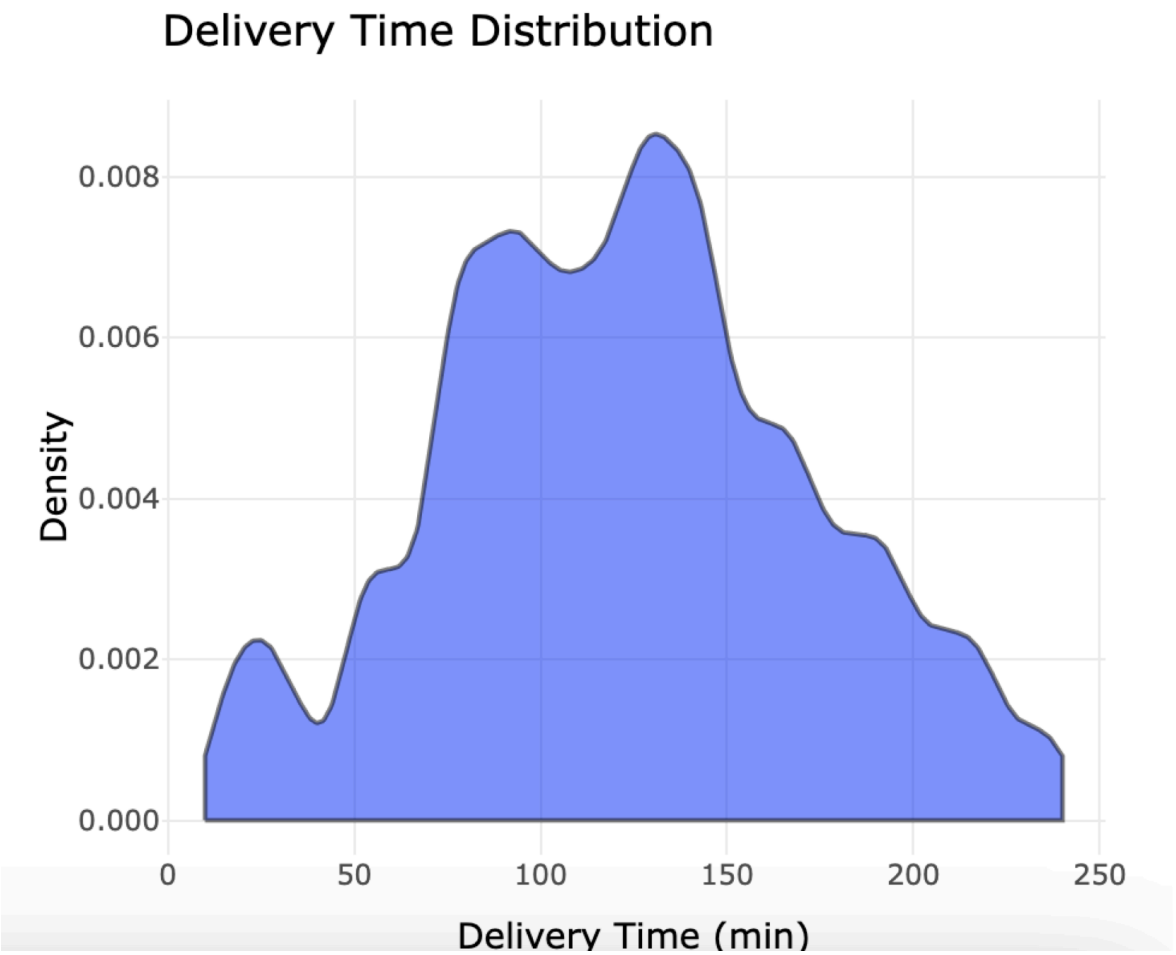

Figure 8 - Delivery Time Distribution (Density)

The delivery time distribution dashboard provides an insight into how all the delivery times are spread across the different orders. This density plot shows us that deliveries are completed within the 50 to the 200-minute range, and most deliveries are completed within the 100 and 150 minute range. It also shows that there is a small percentage of deliveries that take longer than 200 minutes to be delivered, this could be caused by delays due to factors such as traffic congestion, weather conditions or interruptions to the route of the delivery driver. In order to address these extensive delivery times, we suggest investigating outliers and optimize different delivery strategies for the specific locations

where deliveries take excessively long. Also implementing a priority delivery service and establishing different delivery drivers based on geographical routes to help reduce the delay time.

Weather Influence on delivery time - Pie chart

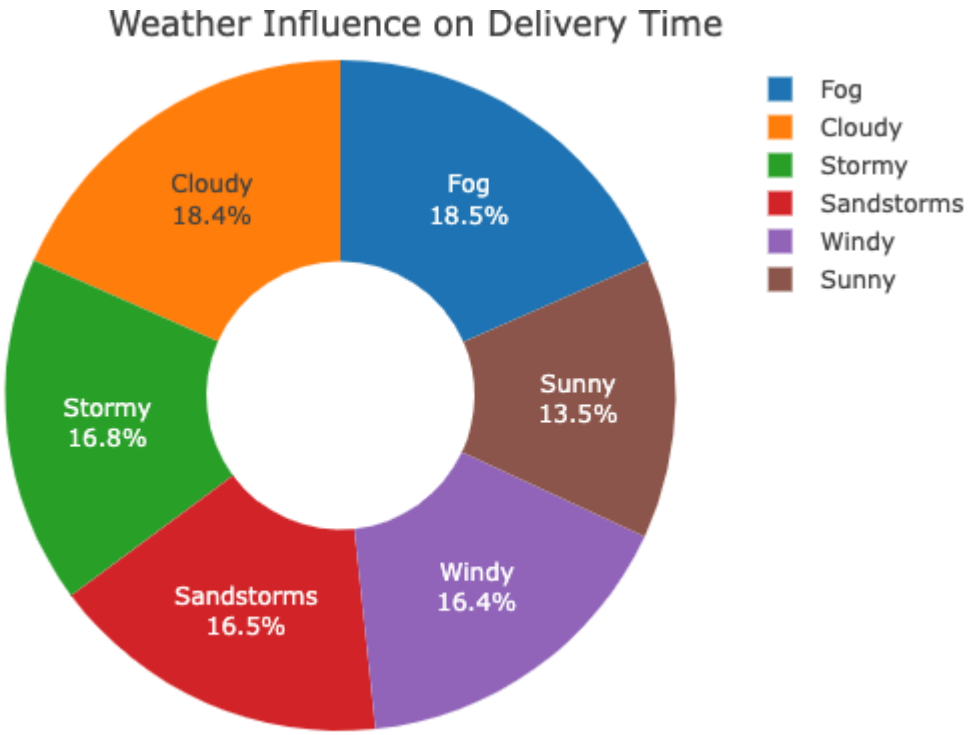

Figure 9 - Weather Influence on delivery time(%)

The Weather influence on delivery time dashboard shows us how the different weather conditions can impact the efficiency of the delivery. This specific pie chart demonstrates that most deliveries occur when there are foggy (18.5%) and cloudy (18.4%) conditions. An interesting stat, is that sunny days account for only 13.5% of total deliveries, this may be due to the fact that the dataset time range is from February to April, which are not the peak summer months. The findings indicate that hostile weather conditions frequently challenge delivery services, leading to possible delays and inefficiencies. In order to overcome these challenges, providing delivery drivers with increase hourly rate salary during extreme conditions and implementing a weather based dispatch planning could help the overall efficiency of each delivery.

High-Risk Orders(% late deliveries) by Weather, Traffic, Area and Vehicle - Heatmap

Another important aspect was to understand in which cases deliveries took longer than usual, and to understand why. However, for this, it was crucial to know the distance between the Amazon warehouse and the drop off location. While the coordinates were provided, it became apparent that they are highly inaccurate. First off all, while Amazon store (warehouse actually) coordinates were accurate, drop off coordinates were a total mess: some had negative latitude values, and even after fixing those, the coordinates did not seem realistically possible due to being in strange line-shaped patterns. This meant that trying to realistically identify which exact locations were experiencing the most delays was worthless. The only noticeable takeaway was that a big part of amazon warehouses were not too far from city centres.

Regardless, it was decided to make a heatmap dashboard with the known variables. For the dashboard, it was selected that deliveries with Delivery Time within the fourth quartile would be seen as “late”, or essentially taking too long. As it can be seen below (Figure X), most of late deliveries happen when there is “high” or “Jam” level of traffic, except for Semi-Urban areas, which seem to always have long deliveries and seemingly always have bad traffic (it should be considered that some data might be missing). Interestingly, it can be seen that vans are not doing worse that scooters or motorbikes, which is counter intuitive, and most likely is because routes are planned well. However, it should be mentioned that there is an increase in late deliveries during foggy or cloudy (potentially post rain) conditions. Additionally, it can be seen that deliveries are just slightly more frequent to be late within the Metropolitan area, compared to Urban and “Other”.

With this information known, it would be recommended to plan ahead of heavy traffic, which is very likely to be periodical, during rush hours. Additionally, the same applies for foggy and cloudy conditions. To decrease the number of late deliveries, there are two options: first option is to not increase the number of the overall driver fleet, and instead schedule them accordingly to conditions where there are more late deliveries. However, this could increase the number of late deliveries during times of more suitable conditions. Option two is to simply hire more drivers, and assign them specifically during rush hours in metropolitan areas, and if possible, once it is known that “counter efficient” weather is likely, however this option will come at an additional cost to the company.

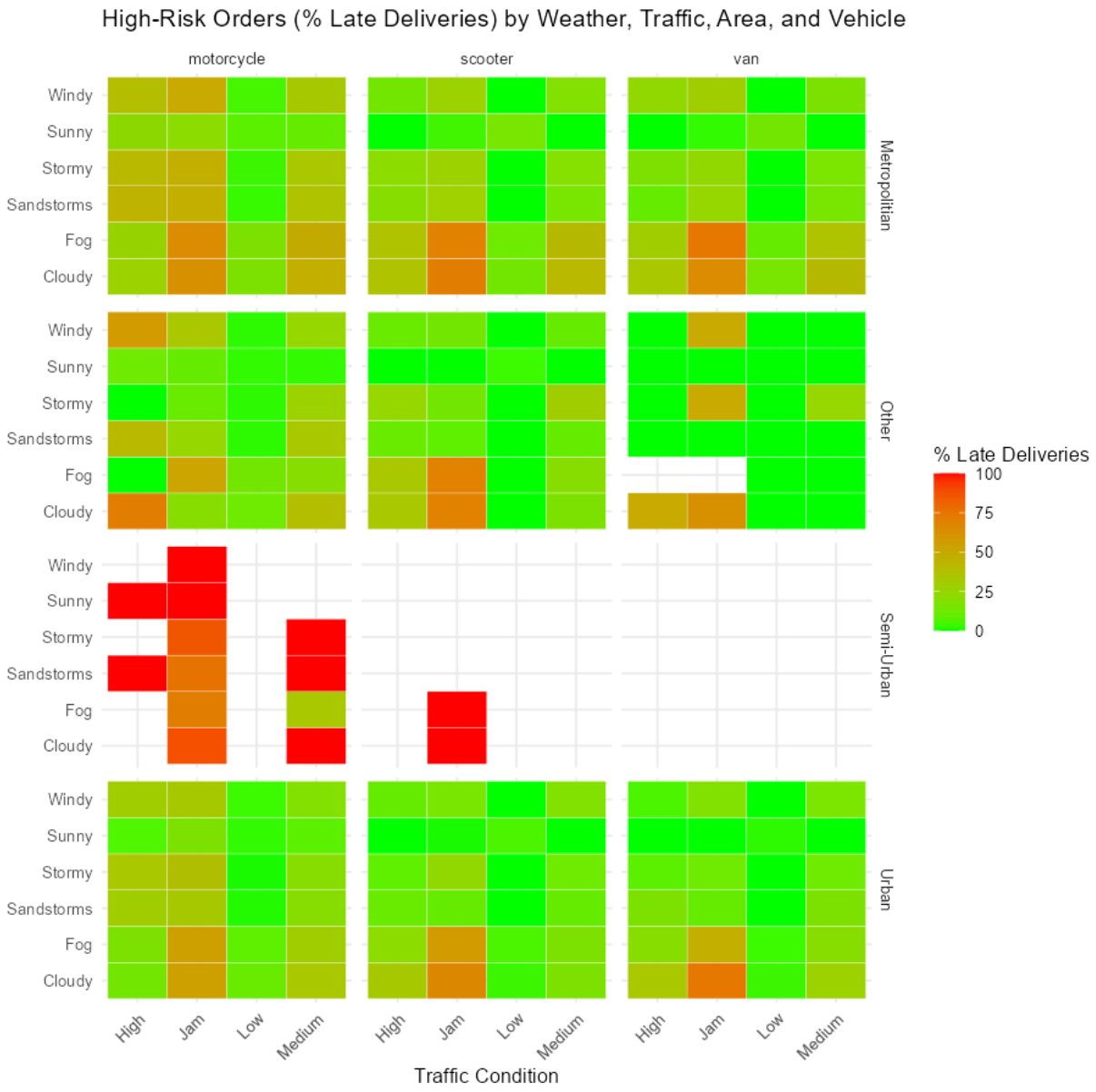

Figure 10 - Number of Late-Orders.

Conclusion

The Amazon Delivery Time Analysis Dashboard provides crucial insights of factors that affect the delivery efficiency such as the weather, traffic, vehicle type, and regional performance. The findings demonstrated in this assignment demonstrate that high traffic congestion and adverse weather conditions have a significant impact on the increment of delivery times, thereby, emphasizing the importance of real time route optimization and AI powered scheduling. In addition, allocating the correct vehicle per delivery has an influential impact on the speed of each delivery, having vans and scooters outperforming motorcycles in specifically urban areas, demonstrates that optimizing vehicle use can enhance efficiency of delivery. Metropolitan and semi urban regions experience the most delays, emphasizing on the need for localized fulfilment strategies such as priority routing per delivery journey. In order to address these challenges, introducing a strategic workforce planning, targeted incentives for drivers and providing customers with a scheduling option, should be implemented. Making use of these improvements, can help Amazon enhance their delivery speed,

improve their overall customer satisfaction and optimize their logistic network for better and more efficient operations.