

Project Report :Logistics & Supply Chain Optimization

1. Executive Summary

This report outlines a comprehensive data analysis project focused on a logistics network. The project demonstrated a full data science lifecycle, from data preparation and advanced modeling using **SQL** to dynamic visualization and strategic storytelling with **Power BI** and **Tableau**. The primary goal was to uncover key factors impacting delivery performance, costs, and risks to provide actionable insights for process optimization and a more resilient supply chain.

2. Methodology and Skills Demonstrated

The project was executed in distinct phases, showcasing a wide range of skills.

Data Preparation & Advanced Analysis (SQL)

- **Data Cleaning and Manipulation:** Used **SQL** to load and clean a raw logistics dataset, addressing data type and formatting issues.
- **Data Modeling:** Designed a **star schema** by creating Fact and Dimension tables to structure the data for efficient analysis.
- **Feature Engineering:** Engineered new columns like `Is_Weekend` and `Time_of_Day_Category` to enrich the dataset for predictive insights.
- **Window Functions:** Applied advanced functions like `LAG()` and a rolling `AVG()` to perform time-series analysis and track performance trends.

Data Visualization & Reporting (Power BI & Tableau)

- **Power BI:** Built a comprehensive, multi-page dashboard. Visuals included a map to identify high-congestion areas, a clustered bar chart to analyze costs by delay category, and a treemap to assess the impact of risk classification on delivery deviation.
- **Tableau:** Created a visually engaging dashboard with a focus on storytelling. Visuals included a packed bubble chart to analyze the distribution of shipping costs by risk level and a line chart to track average delivery deviation over time.
- **Dashboard Refinement:** Implemented interactive slicers, KPI cards, and clear titles to make the dashboards professional and easy for a business audience to understand.

3. Key Insights and Recommendations

The analysis revealed several key findings and led to the following actionable recommendations:

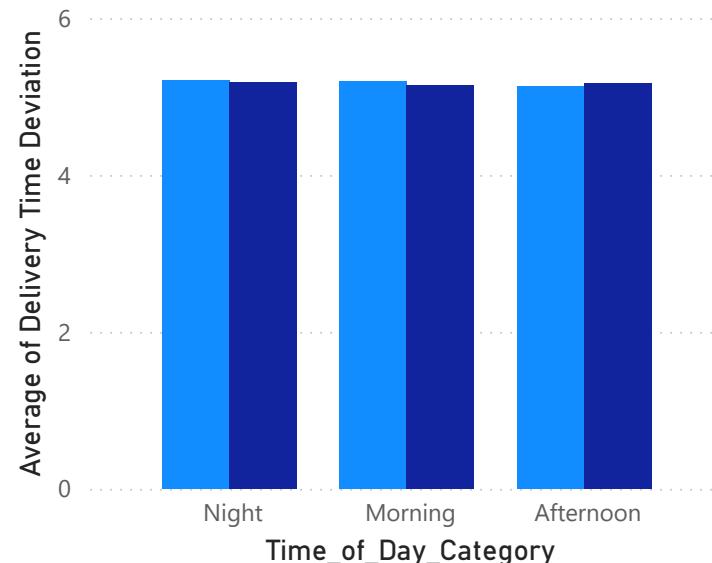
- **Cost and Efficiency:** A strong correlation was found between major delivery delays and higher average shipping costs and fuel consumption.
- **Risk Mitigation:** The analysis demonstrated that high-risk routes are directly linked to higher delivery deviations. The dashboards visually identified geographic hotspots of congestion that are a primary cause of these delays.
- **Strategic Planning:** The predictive analysis showed that while **Delay Probability** exists, it is not a strong predictor of **ETA Variation**. This insight suggests that a more robust predictive model or a focus on real-time monitoring of external factors is needed.
- **Process Optimization:** The findings indicate that issues with order fulfillment are not solely tied to inventory levels, suggesting a need to investigate other internal bottlenecks (e.g., loading/unloading processes).

Overall Disruption Likelihood Score

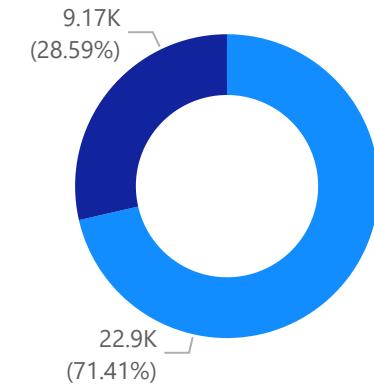


Average of Delivery Time Deviation by Time_of_Day_Category and Is_Weekend

Is_Weekend ● 0 ● 1



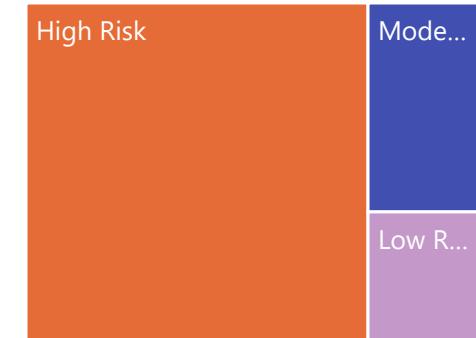
Count of Is_Weekend by Is_Weekend



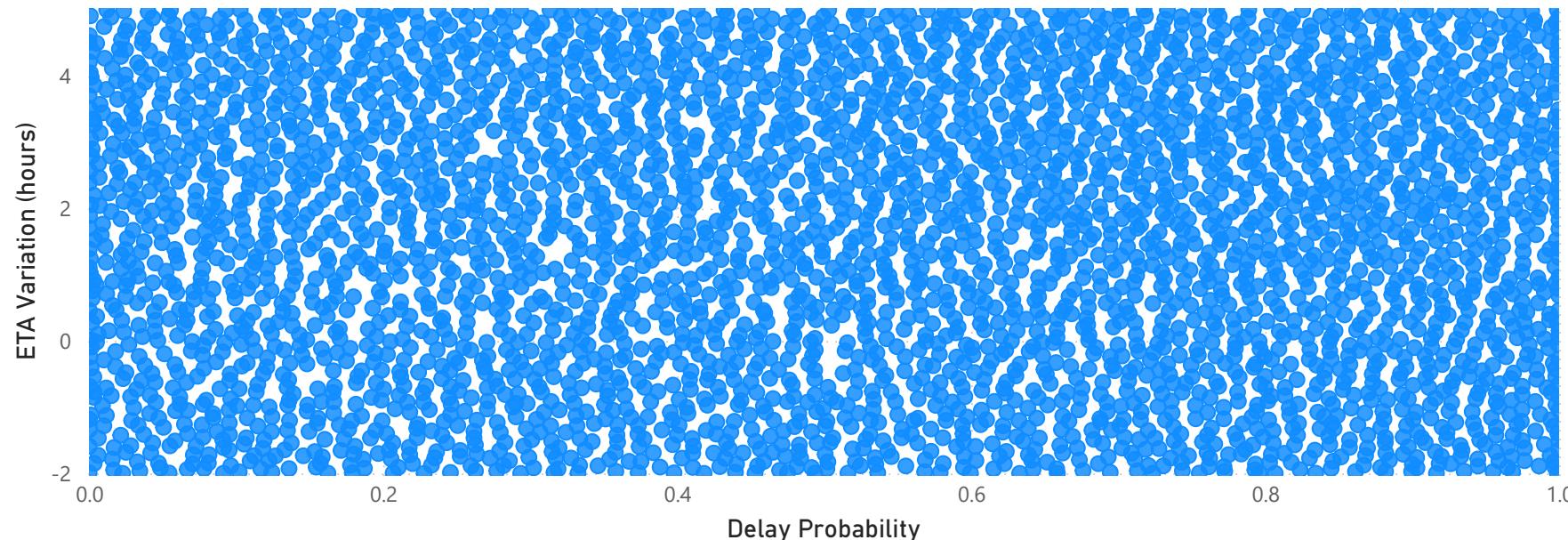
Time_of_Day_Category

- Afternoon
- Morning
- Night

Impact of Risk Classification on Delivery Deviation



Delay Probability vs. ETA Variation



Average of Delivery Time Deviation by Delivery Time Deviation

10.00

Average of Disruption Likelihood Score by Disruption Likelihood Score

1.00

Traffic Congestion Level, Vehicle GPS Latitude and Vehicle GPS Longitude

Traffic Congestion Level ● 3.22067... ● 5.80059... ● 6.30550... ● 1.50841... ● 5.12051... ● 6.69948... ● 9.57906... ● 0.00013... ● 0.00014... ● 0.00014... ● 0.00020... ● 0.00032... ● 0.00054... ● 0.00063... ▶

