# LAST-MILE DELIVERY DELAY ANALYSIS USING PYTHON EDA FOR A U.S. REGIONAL LOGISTICS FIRM

## 1. Background

A regional logistics firm based in the southeastern United States, operating across five states, faced recurring complaints regarding delayed deliveries in their last-mile operations. Despite having tracking data and dispatch logs, their internal team lacked the tools and bandwidth to identify the root causes behind these delays.

We were hired to conduct an in-depth exploratory data analysis using Python to surface key delay patterns, route-level anomalies, and operational inefficiencies—supporting decision-making in scheduling, routing, and vendor management.

## 2. Objective

- To identify consistent patterns and outliers contributing to delivery delays
- To compare performance across delivery hubs, drivers, zip codes, and time-of-day
- To visualize the delay structure using interactive charts and heatmaps
- To provide actionable insights to help reduce delivery time variance and improve SLA adherence

## 3. Data Used

The firm provided **4 months of operational data** containing **57,200 delivery records** with fields including:

- Delivery_ID
- Driver_ID
- Hub_Location
- Scheduled_Delivery_Time
- Actual_Delivery_Time
- Delay_Minutes (calculated)
- Route_Zip

- Weather_Condition (Clear, Rain, Fog, Snow)
- Delivery_Window (Morning / Afternoon / Evening)
- Package_Type (Standard, Fragile, Oversized)

Additional metadata about driver shift lengths and delivery zones was also integrated for enriched analysis.

# 4. Methodology

**4.1 Data Preprocessing**

- Converted time columns to datetime format and calculated Delay_Minutes
- Filtered out invalid and duplicate entries (e.g., negative delivery time)
- Flagged deliveries as On_Time or Delayed using a ±10-minute threshold
- Created derived variables:
    - Hour_of_Delivery
    - Is_Weekend
    - Weather_Flag (binary delay risk due to rain/fog/snow)

**4.2 Exploratory Analysis**

- Univariate analysis: Delay frequency, mean/median delay per category
- Bivariate comparisons:
    - Delay_Minutes by Weather_Condition and Delivery_Window
    - Hub-wise delay rates and driver-level on-time performance
- Multivariate analysis using heatmaps to analyze:
    - Hub_Location × Delivery_Window × Delay_Minutes
    - Zip code clusters using geospatial mapping

**4.3 Tools Used**

- Python libraries: pandas, numpy, matplotlib, seaborn, plotly, geopandas
- Report generated using Jupyter Notebook, exported to APA-style PDF format
- Geo-mapping done via interactive plotly.express choropleth map

## 5. Key Results

- **Average Delay**: 14.2 minutes
- **Most delayed deliveries** occurred during the **Afternoon Window (1–4 PM)** across all hubs
- **Driver fatigue correlation**: drivers in the final hour of their shift were 2.3× more likely to be delayed
- **Weather-related delays** accounted for 29% of all late deliveries; rain was the dominant factor
- **Top 10 ZIP codes** accounted for 38% of all reported delays—largely urban congestion zones
- Hub C had the lowest on-time rate (61%) and the highest standard deviation in delay time

## 6. Report Output

- **Deliverables**:
  - Python Jupyter Notebook with complete preprocessing, visualization, and trend analysis code
  - Interactive HTML dashboard with plotly visualizations and ZIP-level performance breakdown
  - PDF Report (18 pages) including:
    - Hub and driver performance comparison tables
    - Delay heatmaps by time-of-day and weather
    - Recommended time slot reallocation based on actual delay trends
  - Excel summary table for operational leadership with:
    - Top delayed zones
    - High-risk weather periods
    - Suggested hub-level targets for Q2

## 7. Business Impact

- Dispatch team revised driver scheduling to avoid back-to-back fatigue blocks

- Introduced new delivery window definitions based on empirical delay data
- Geo-targeted route optimization started for ZIP clusters with above-average congestion
- SLA compliance improved by 8% in the following quarter
- Delay analytics used to renegotiate vendor penalties and adjust incentives in high-risk weather zones

## 8. Future Scope

- Integrate real-time traffic data to enrich future EDA refreshes
- Develop a machine learning model to predict delay risk before dispatch
- Deploy a Tableau or Power BI version of the delay dashboard for executive teams
- Extend the framework to analyze **missed delivery attempts** and customer satisfaction data