

Uber Trip Analysis Report Summary (Jan–Feb 2015)




Objective:

The primary goal of this project is to analyze Uber's operational data from January and February 2015 in New York City, identify key trends in rider demand, evaluate dispatch base performance, and provide insights to improve operational efficiency such as trip distribution, base utilization, and vehicle deployment.

Dataset Overview

- Source: NYC TLC FOIL Data
- File Used: Uber-Jan-Feb-FOIL.csv
- Period Covered: January and February 2015
- Columns:
 - dispatching_base_number: The Uber base/company that dispatched trips
 - date: Date of recorded activity
 - active_vehicles: Number of vehicles operating that day
 - trips: Total Uber trips completed that day

Exploratory Data Analysis (EDA) Summary

-  Daily Trip Volume
 - Trips increased gradually across both months, peaking in early February.
 - A clear daily seasonality is visible, with certain days showing consistent higher demand.
-  Weekly Patterns
 - Most trips occurred on Thursdays and Fridays, while Sundays had the least.
 - Week 6 (early February) recorded the highest overall trip volume.
-  Base Performance

- Top 3 most active bases by total trips:
 1. B02512 (Unter)
 2. B02764 (Danach-NY)
 3. B02682 (Schmecken)
- B02512 consistently outperformed other bases across all dates.

4. 🚗 Vehicle Efficiency

- Measured as: $\text{trips} / \text{active_vehicles}$
- B02598 and B02764 had the highest average trips per vehicle, indicating better utilization.

5. 📦 Distribution of Trips

- Daily trip counts per base had significant variation.
- Boxplots show that B02512 had a high number of days with very large trip volumes.

6. 🔥 Correlation Insights

- There is a strong positive correlation between the number of active vehicles and total trips.
 - This implies operational scalability: more cars = more completed trips.

7. 📊 Stacked Area Chart

- Visually demonstrates how different bases contributed to the total daily trips.
- B02512 consistently led in trip contribution over time.

📁 Tools Used:

- Python: Data cleaning, transformation, and visualization
- Libraries: pandas, matplotlib, seaborn
- *(Optional extensions can include Power BI, Excel, or Plotly for dashboards)*

📌 Key Insights

- High-performing bases like B02512 and B02764 played a critical role in daily trip volume.
- Weekend dips suggest reduced demand or fewer working vehicles.
- Vehicle-trips efficiency can help optimize deployment strategy by base.
- Seasonal behavior is present despite only 2 months of data.

Suggested Improvements or Extensions

- Include location data (locationID, Lat, Lon) for spatial/geographic mapping.
- Extend dataset to cover longer duration (full year) for richer time-series analysis.
- Add weather or holiday data to explain anomalies in trip behavior.