

NYC Taxi Trip Duration Analysis

Sagnik Bhattacharya

December 28, 2025

Abstract

This report presents a production-oriented exploratory data analysis of the NYC Taxi Trip Duration dataset. The objective is to uncover temporal demand patterns, trip behavior, and distance characteristics that can support operational and business decision-making. The analysis integrates data cleaning, feature engineering, SQL-based querying, and an interactive Excel dashboard for stakeholder-friendly insights.

1 Dataset Attribution

This project uses the **NYC Taxi Trip Duration** dataset made publicly available by **Yasser H.** on Kaggle.

- Source: Kaggle
- Dataset: NYC Taxi Trip Duration
- Author: Yasser H.
- URL: <https://www.kaggle.com/datasets/yasserh/nyc-taxi-trip-duration>

The dataset is used solely for educational and analytical purposes.

2 Introduction

Urban mobility data plays a critical role in understanding transportation demand and commuter behavior. This report analyzes NYC taxi trip data to identify peak usage periods, weekday versus weekend trends, and trip distance distributions. The insights derived aim to replicate a real-world analytics workflow suitable for business intelligence and data analyst roles.

3 Dataset Description

The dataset contains over **1.45 million** taxi trips with the following key attributes:

- Pickup datetime
- Pickup hour and weekday
- Passenger count
- Trip distance (kilometers)
- Trip duration (seconds)
- Weekend indicator

4 Methodology

The analysis followed a structured workflow:

1. Data ingestion and validation
2. Data cleaning and anomaly handling
3. Feature engineering for temporal analysis
4. Exploratory analysis using visual analytics
5. KPI design and dashboard development

5 Data Cleaning and Feature Engineering

- Removed trips with zero or invalid duration values
- Validated passenger count and trip distance ranges
- Engineered time-based features such as pickup hour, weekday, and weekend flag
- Created trip distance buckets for categorical analysis

6 Exploratory Analysis

6.1 Taxi Demand by Hour

Taxi demand shows a clear temporal pattern, with peak activity observed during evening commute hours (17:00–19:00), indicating strong work-related travel behavior.

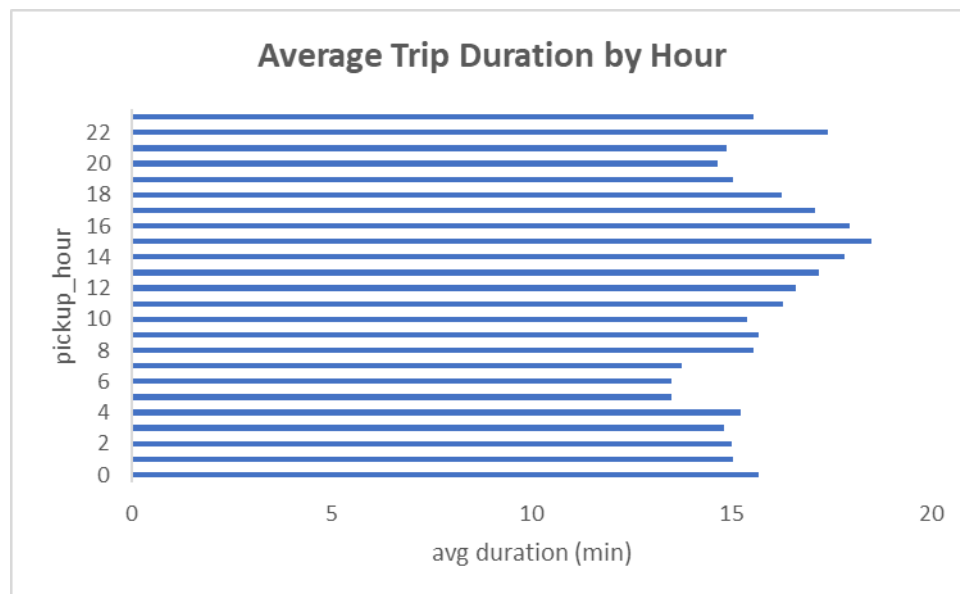


Figure 1: Trips by Hour

6.2 Weekend vs Weekday Trips

Weekday trips dominate overall taxi usage, suggesting that taxis primarily support routine commuting rather than leisure travel.

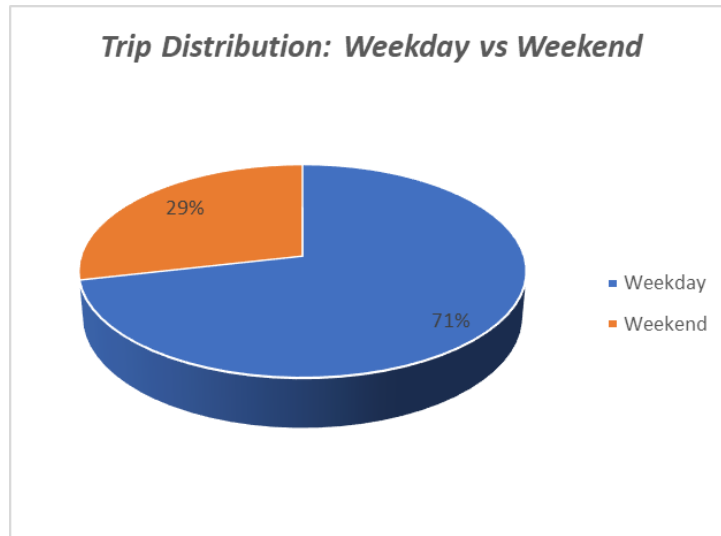


Figure 2: Weekend vs Weekday Trips

6.3 Trip Distance Distribution

The majority of trips fall under short-distance categories (≤ 2 km), highlighting the role of taxis in short urban mobility rather than long-distance travel.

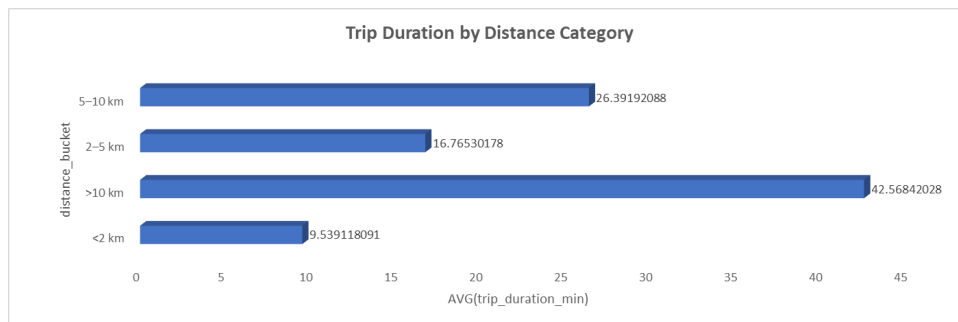


Figure 3: Trip Distance Distribution

7 Key Performance Indicators

The following KPIs were derived to support business interpretation:

- Total number of trips
- Peak demand hour
- Percentage of weekend trips
- Distribution of trips by distance category

8 Dashboard

An interactive Excel dashboard was developed to allow non-technical stakeholders to dynamically explore trip demand by time, distance, and weekday category.

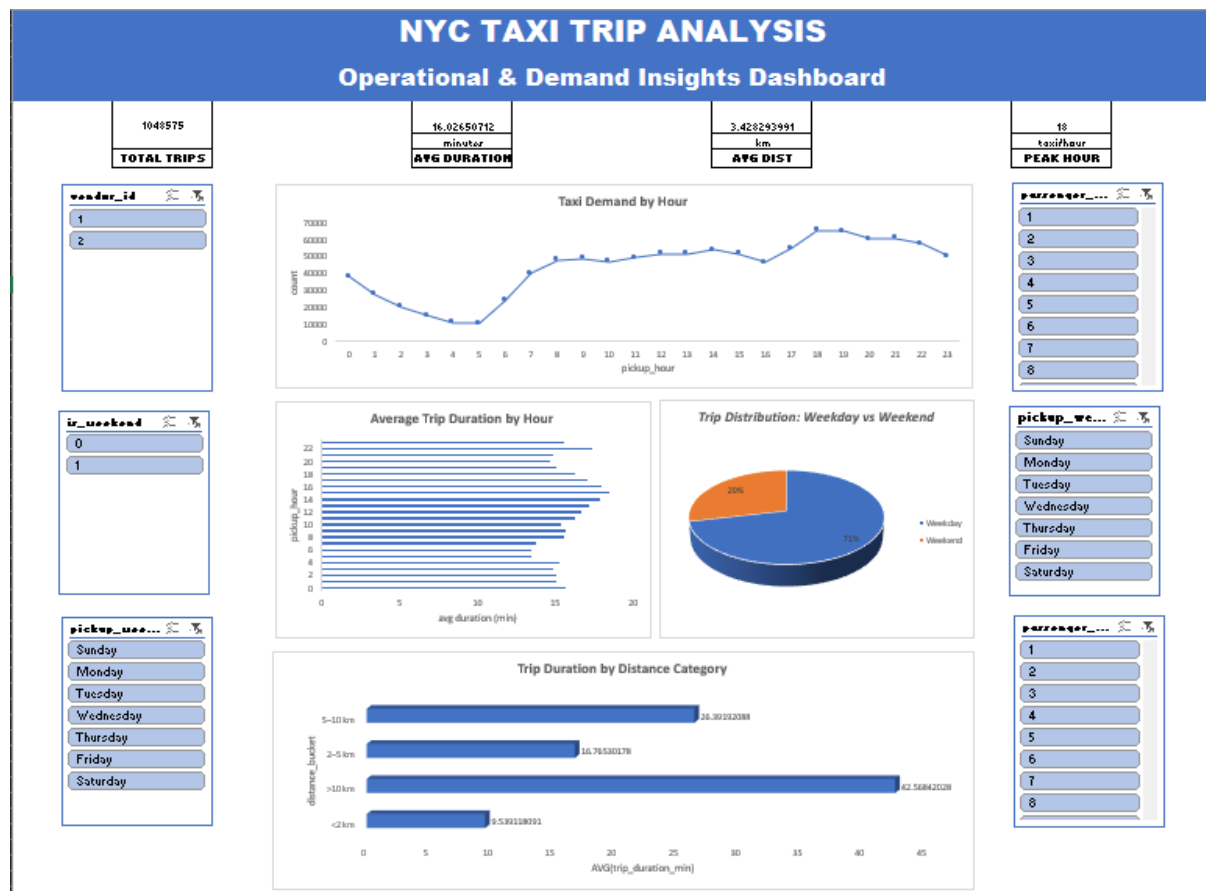


Figure 4: Excel Dashboard Overview

9 SQL Analysis

Interview-grade SQL queries were written to:

- Identify peak demand hours
- Compare weekday and weekend trip volumes
- Analyze trip distance distributions
- Aggregate trip metrics for KPI computation

10 Conclusion

This project demonstrates an end-to-end data analytics workflow using a large real-world dataset. The findings reveal strong weekday commuter demand and a predominance of short-distance taxi trips. The combination of data cleaning, SQL analysis, and dashboarding makes this project suitable for production environments and data analyst portfolios.