**Project Overview**

This project involved building a complete data pipeline to analyse New York City taxi trip data. The goal was to extract meaningful monthly insights on revenue, trip behavior, and passenger volume using a structured, multi-layered approach.

**Pipeline Structure**

**1. Bronze Layer**

* Raw Parquet files loaded and chunked for memory efficiency
* Data cleaned and stored in PostgreSQL

**2. Silver Layer**

* Millions of rows processed in 83 chunks
* Key columns selected and transformed
* Stored in a clean, query-ready format

**3. Gold Layer**

* Monthly metrics calculated:
  + total\_revenue
  + avg\_trip\_pct
  + total\_trips
  + total\_passengers
* Exported to CSV for visualization

**Dashboard Visuals**

**1. Monthly Revenue Trend (Line Chart)**

* Shows how total revenue fluctuates month by month
* Clear upward and downward patterns visible
* Y-axis formatted in millions ($)

**2. Average Tip Percentage by Month (Column Chart)**

* Highlights tipping behavior across months
* Some months show unusually high tip percentages
* Y-axis formatted as percentage (%)

**3. Trips vs Passengers per Month (Stacked Bar Chart)**

* Compares number of trips to passenger volume
* Useful for identifying ride-sharing trends
* Y-axis formatted in thousands

**Key Insights**

* Revenue peaks in certain months, possibly due to seasonal demand
* Trip percentages vary widely — may reflect rider satisfaction or fare types
* Passenger counts often exceed trip counts, suggesting shared rides

**Final Output**

* Dashboard saved as: nyc\_taxi\_dashboard.pbix
* Exported to PDF for reporting: nyc\_taxi\_dashboard.pdf