# **Case Study: Understanding Rider Behavior by Payment Type in NYC Taxi Trips**

## Introduction

This case study aims to analyze publicly available taxi trip data from New York City to uncover patterns in rider behavior based on payment types. The objective is to draw actionable insights that can help improve the city’s taxi services and guide strategic decision-making. The study replicates the Cyclistic-style data analysis process by focusing on how different types of users (in this case, cash vs. card payers) behave across several trip metrics.

## Business Question

The main question guiding this analysis is:

**“How do card-paying and cash-paying taxi users in New York City use taxi services differently?”**

From this central question, the following sub-questions emerge: 1. What percentage of rides are paid via card vs. cash? 2. Do card users take longer or more expensive trips? 3. Are there differences in ride patterns by day of the week between card and cash users?

## Dataset Overview

* **Source**: bigquery-public-data.new\_york\_taxi\_trips.tlc\_green\_trips\_2022
* **Period**: Year 2022
* **Data Access Method**: Google BigQuery (Sandbox)
* **Payment Types Included**:
  + 1: Credit Card
  + 2: Cash
  + (Other payment types such as “No charge,” “Dispute,” and “Unknown” were excluded.)

## Step-by-Step Documentation

### 1. Data Extraction and Cleaning

* Connected to the BigQuery public dataset tlc\_green\_trips\_2022.
* Filtered the dataset to include only valid rides:
  + trip\_distance > 0
  + fare\_amount > 0
  + trip duration between 1 and 240 minutes
  + payment\_type in (1, 2)
* Calculated additional fields:
  + trip\_duration\_min using TIMESTAMP\_DIFF
  + day\_of\_week using EXTRACT(DAYOFWEEK)

### 2. Data Validation

* + Ensured there were no NULLs in key fields after filtering.
  + Checked logical consistency between start and end times.
  + Validated payment\_type field to only include 1 (Card) and 2 (Cash).

### 3. Summary Statistics

* + **Total Trips**: 625,801
  + **Average Trip Duration**: 29.84 minutes
  + **Average Fare**: $28.97
  + **Card Share**: 63%

### 

### 4. Data Visualization

Using Google Sheets:

#### Chart 1: Average Trip Duration by Payment Type

* Type: Column chart
* Insight: Card users take slightly longer trips on average than cash users.

#### Chart 2: Rides by Day of Week

* Type: Grouped bar chart
* Insight: Card usage is more consistent across weekdays, while cash usage spikes on weekends.

#### Chart 3: Average Fare by Payment Type

* Type: Donut chart
* Insight: Card users pay slightly more per trip on average.

#### KPI Section

Displayed key figures using clean, minimal “stat cards” with icons:

* Total Trips: 625,801
* Avg Trip Time: 29.84 mins
* Card Share: 63%
* Avg Fare: $28.97

### 5. Limitations of the Data

* + The data only includes green taxi trips from 2022, excluding yellow taxis and ride-hailing services.
  + Some payment types were excluded due to ambiguity or missing values.
  + Data is anonymized; no demographic or geographic personalization is possible.
  + Average calculations may be slightly skewed by outliers, even after filtering.

## Final Recommendations

1. **Promote Digital Loyalty Programs**: With 63% of users paying by card, introducing rewards for digital payments could increase retention.
2. **Targeted Incentives for Cash Users**: Offer discounts or perks to encourage cash users to switch to card, improving transaction speed and tracking.
3. **Implement Time-Based Fare Strategies**: With an average ride duration of ~30 minutes, shorter rides could be promoted via fixed fares, while longer rides remain at standard pricing.

## Conclusion

This case study successfully applied the Cyclistic data analysis framework to NYC taxi data. By focusing on payment type behavior, the analysis revealed meaningful differences that can inform operational and marketing strategies. With further improvements in data scope and granularity, this approach can be expanded to enhance city-wide transportation planning and policy-making.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Attatchment:*

>> [sql query cleaned data\sql scripts for data cleaning.txt](sql%20query%20cleaned%20data/sql%20scripts%20for%20data%20cleaning.txt)