



NYC TAXI TRIP DATA

EDA SUMMARY

Summary of Trends, Patterns, and Key Insights
Muhammad Huzaifa Mohsin
August 07, 2025

INTRODUCTION

The dataset contains detailed records of NYC taxi rides from March 2019, collected from yellow cab trips. Each row represents an individual taxi ride, capturing trip characteristics such as:

- Pickup & drop off timestamps
- Passenger count
- Trip distance (in miles)
- Fare amount, tips, tolls, and total fare
- Payment method (cash/credit card)
- Pickup and drop-off zones/boroughs

PROJECT OVERVIEW

- Timeframe: March 2019
- Total Rides: 6,433 | 14 Features
- Features: Pickup/Dropoff, Distance, Fare, Tip, Tolls, Zones, Boroughs, Payment Method
- Null Values: Only in payment & location zones
- No duplicate rows found
- Goal: To perform Exploratory Data Analysis (EDA) and uncover insights about taxi usage, fares, tips, and user behavior
- Tools Used: Python (Pandas, Matplotlib), Canva

KEY INSIGHTS

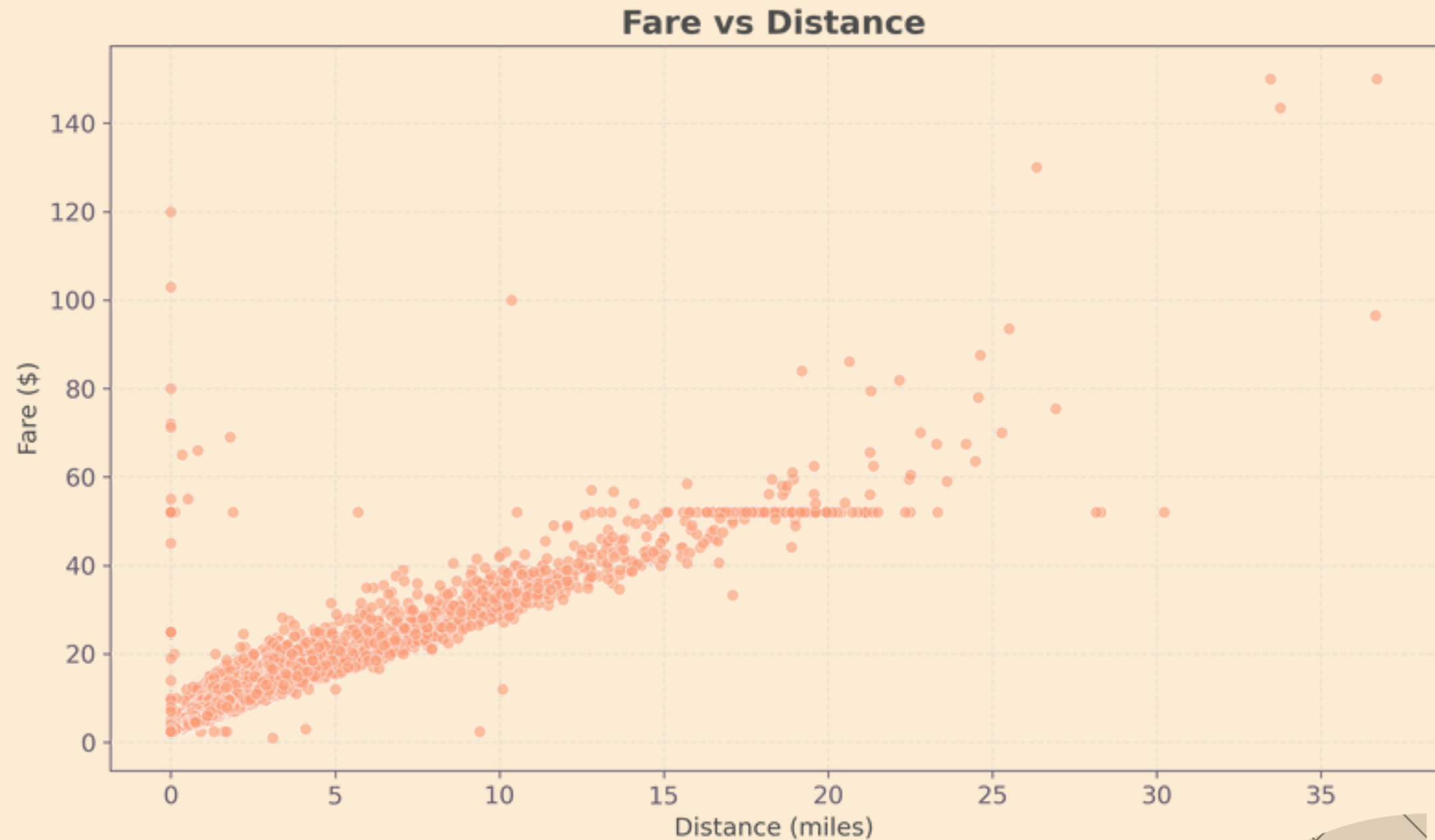
- Only credit card users tipped; cash users didn't.
- Fare increases with trip distance (strong positive correlation).
- Most rides occur between 15:00–20:00.
- Most common pickup zone: Midtown Centre.
- Friday is the busiest day.
- Average fare: \$13.09 | Average tip: \$1.98
- Longest trip distance: 36.7 miles

OUTLIER ANALYSIS

- Distance: 3 trips > 30 miles
 - Fare: 5 rides > \$100
 - Tip: Few tips > \$20
 - Total: Some totals > \$100
-
- No unrealistic values found. So, kept all data for analysis.

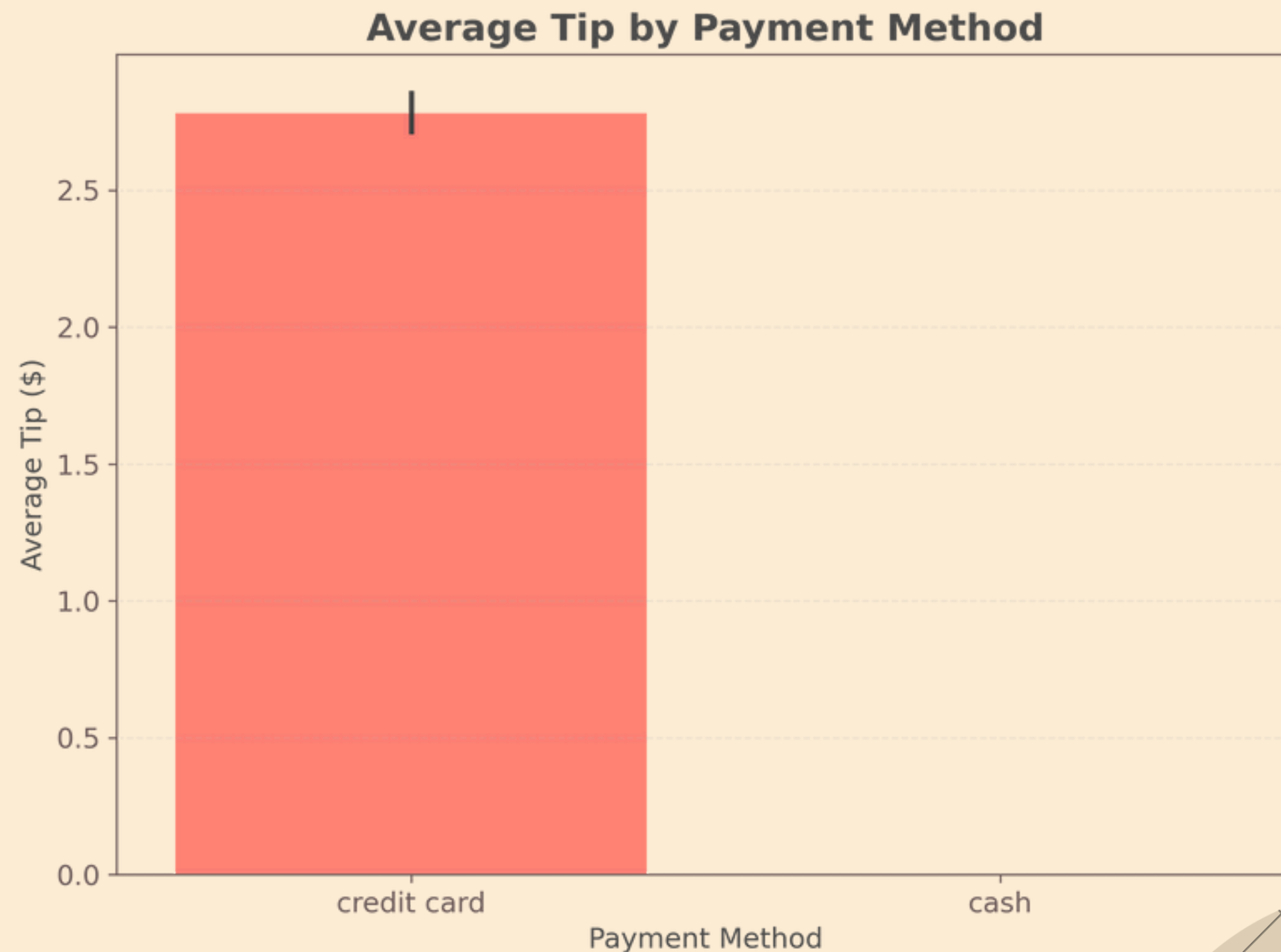
FARE VS DISTANCE

- There is a **strong positive relationship** between trip distance and fare. As the distance increases, the fare also increases.



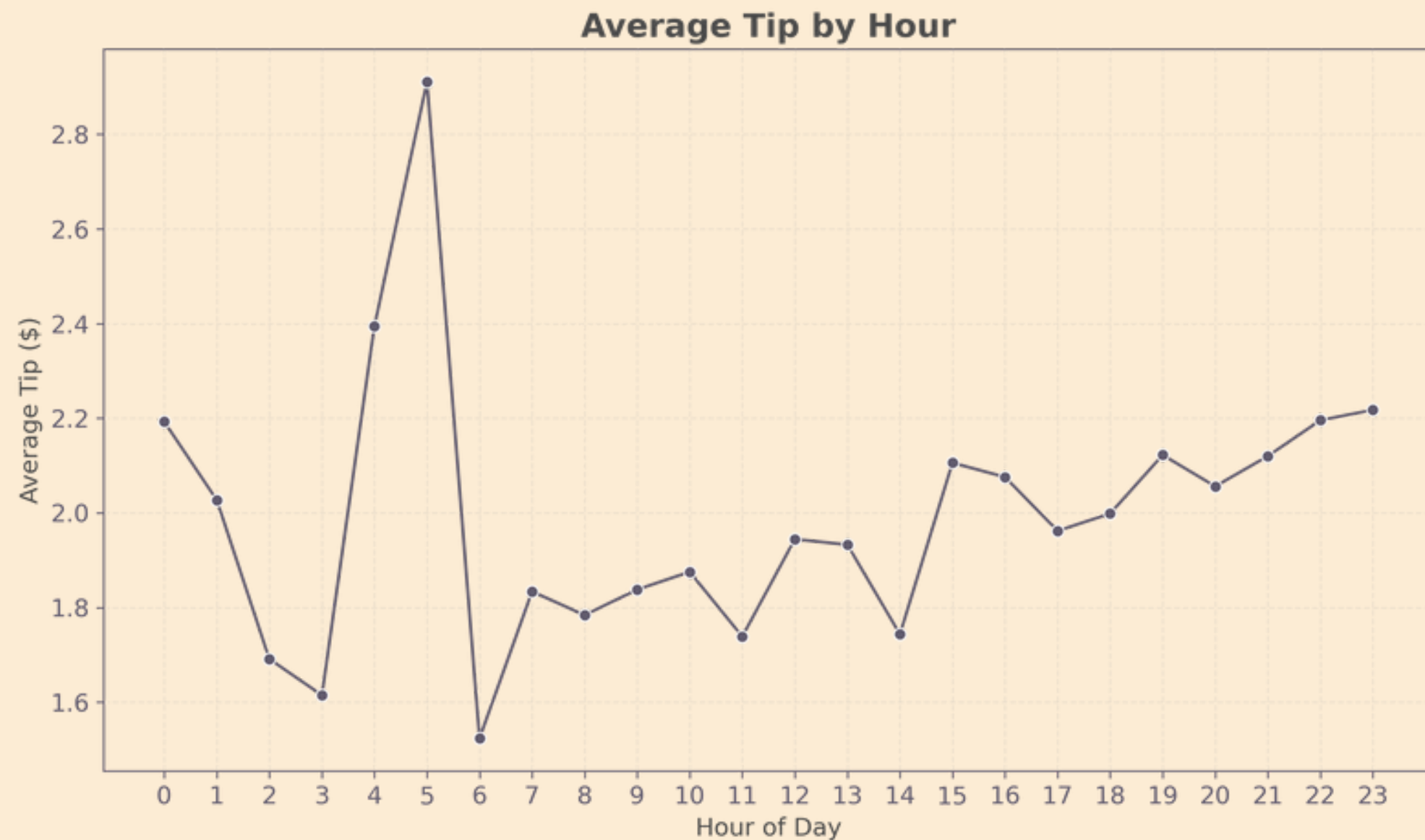
TIPS BY PAYMENT METHOD

- Only credit card users gave tips. There were no tips recorded for cash payments, suggesting that tipping is either not allowed, not recorded, or not practiced during cash transactions.



TIP BY HOUR OF DAY

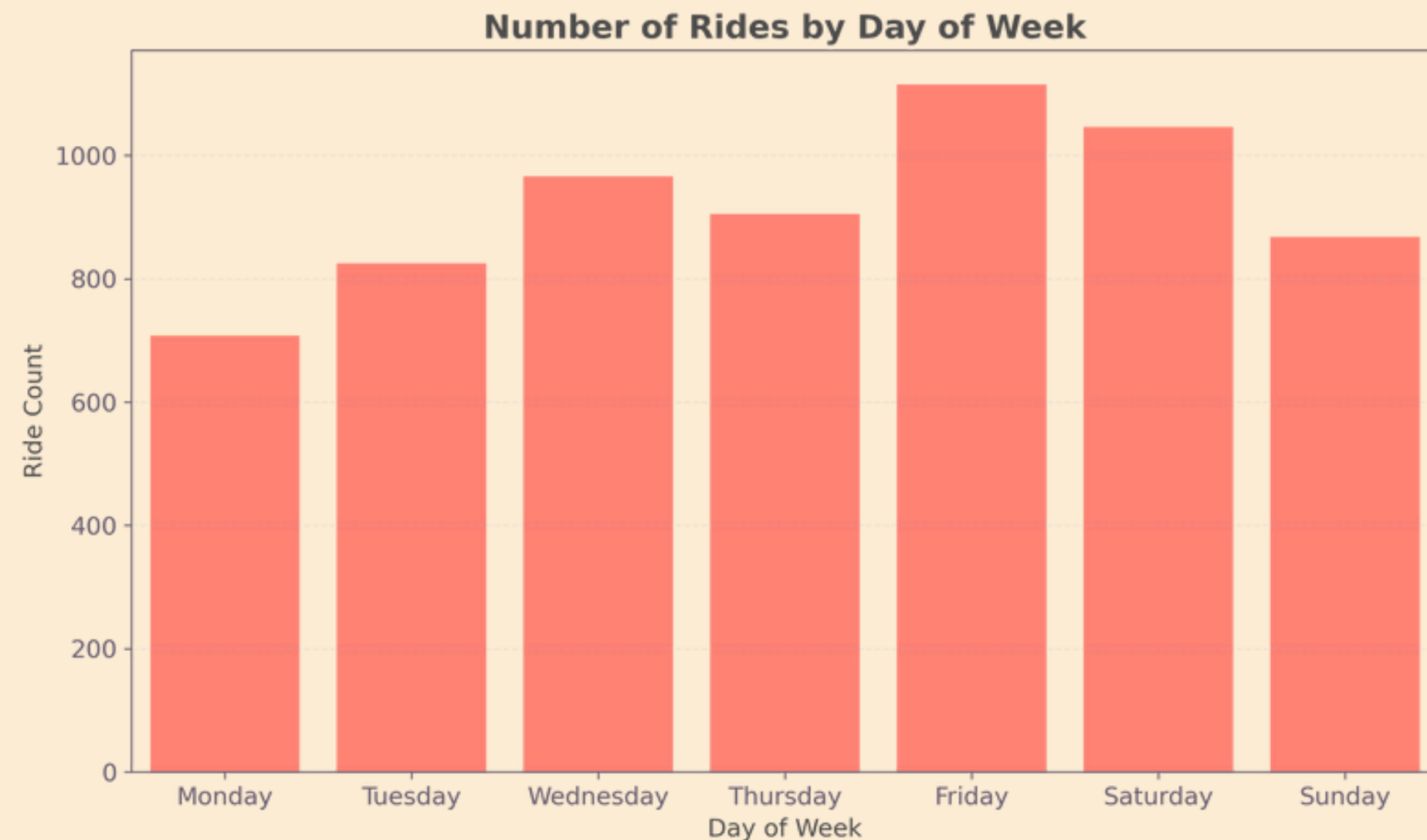
- Average tips are highest between 3 PM and 8 PM (15:00 to 20:00 hours), reaching up to \$2.20. This suggests that riders are more generous during the late afternoon and early evening, possibly due to workday fatigue, better earnings, or higher value trips.



RIDES BY DAY OF WEEK



- Friday had the **highest number** of rides, likely due to weekend plans or commuting rush.



CONCLUSION & RECOMMENDATIONS

- ✓ Data is clean and usable
- ✓ Cash payments are not associated with tipping
- ✓ Peak ride hours (15:00–20:00) and days (Friday) can be leveraged for marketing
- ✓ Fare and distance have a clear linear trend, good for fare prediction models

THANK YOU

