DAF Graduation Project

Uber: Optimizing Revenue per Trip

September 2025

Table of Contents

1. Introduction

1.1. Business Context

Uber, a global leader in ride-hailing, serves millions of daily trips, from short city commutes to longer airport transfers. While trip volume is high, the strategic focus has shifted to improving financial efficiency by maximizing the revenue generated from each individual trip.

1.2. Problem Statement

The objective of this analysis is to dissect Uber's trip-level data to identify the key drivers of **`Revenue per Trip`**. By analyzing trip attributes, payment behaviors, and operational patterns, this report will provide actionable, data-driven recommendations to help Uber strategically increase earnings per ride.

---

2. Analytical Approach

2.1. Key Metrics and Hypotheses

The primary metric is **`Revenue per Trip`**, supported by **`Tip Rate`** and **`Revenue per Mile`**. Seven hypotheses were established to guide the analysis, covering the impact of payment methods, trip distance, time patterns, passenger count, and location.

2.2. Dataset Overview

The analysis was conducted on a large-scale dataset containing over 10 million trip records. Key fields included pickup/dropoff datetimes and locations, trip distance, fare and tip amounts, total amount, passenger count, payment method, and rate code.

2.3. Methodology and Workflow

This analysis was conducted with a **"memory-first" mindset** due to the dataset's multi-gigabyte size and the memory constraints of the Google Colab environment. A memory-efficient workflow was implemented:

- **Sequential Merging:** Datasets were merged sequentially, and redundant DataFrames were immediately deleted from RAM (`del df`, `gc.collect()`) to prevent system crashes.
- **Optimized Lookups:** The `.map()` method was used as a more memory-efficient alternative to `pd.merge()` for adding location names.
- **Strategic Sampling:** For visualizations (histograms, boxplots), a large, random sample

of 1 million rows was used to ensure fast processing while maintaining visual accuracy. For precise statistical calculations, the full, clean dataset was always used.

---

3. Data Cleaning and Preparation

3.1. Handling Missing Values and Duplicates

The initial dataset was systematically cleaned to ensure data integrity. Over **15,000 duplicate trip records** were removed, along with rows containing critically missing data points (e.g., pickup datetimes).

3.2. Managing Outliers and Anomalies

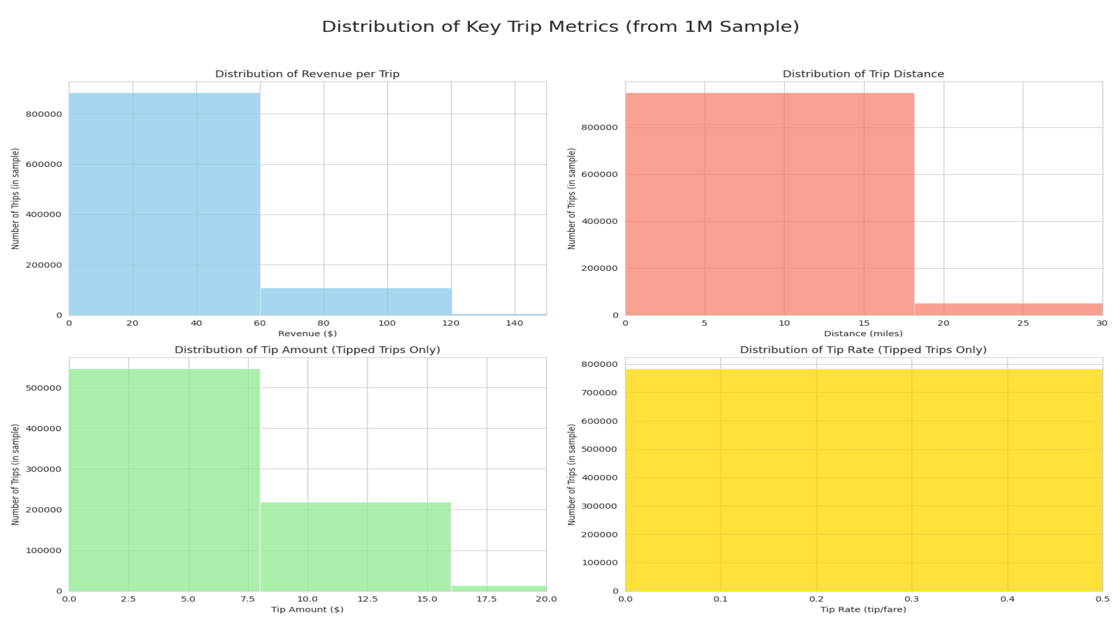
Illogical entries were filtered out, including over **15,000 trips** with negative fare amounts, zero passengers, or invalid dropoff times. Extreme outliers were identified and filtered out of visualizations to prevent distortion.

---

4. Exploratory Data Analysis (EDA)

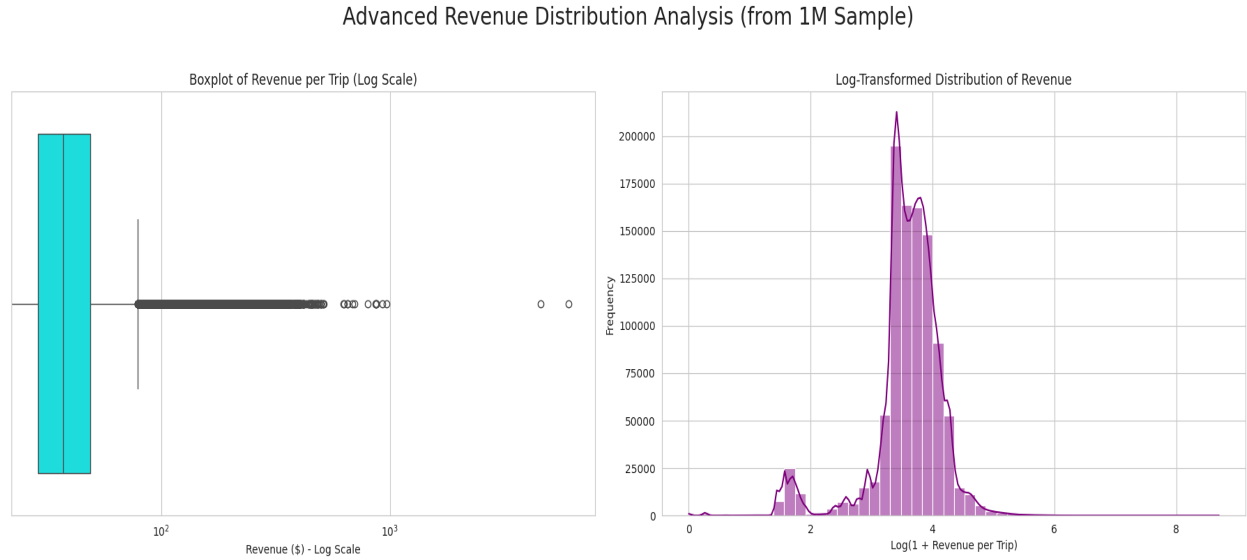
4.1. Distribution of Key Metrics

Visual analysis of key metrics revealed that revenue, distance, and tip amounts are all heavily **right-skewed**, indicating a business model driven by a high volume of common, lower-value trips, with a long tail of less frequent but more lucrative high-value trips.



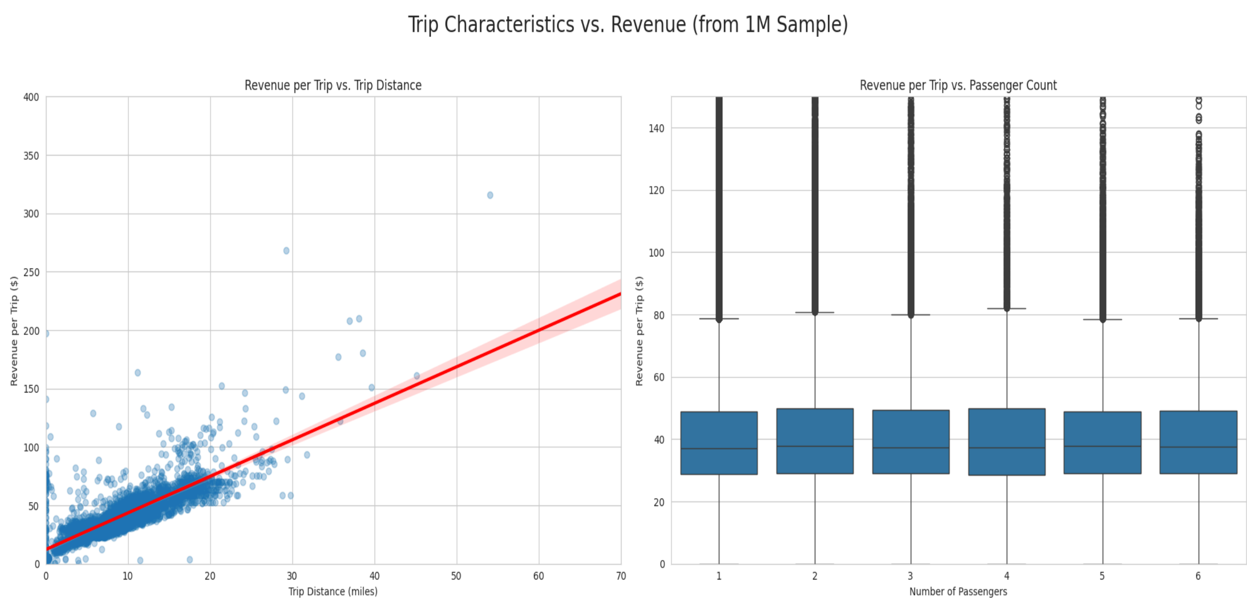
4.2. Revenue Distribution Analysis

A deeper quantile analysis confirmed the skewed nature of revenue. While a median trip earns **$38.20**, the top 1% of trips generate over **$114.96—**at least **3 times more** than a typical trip.



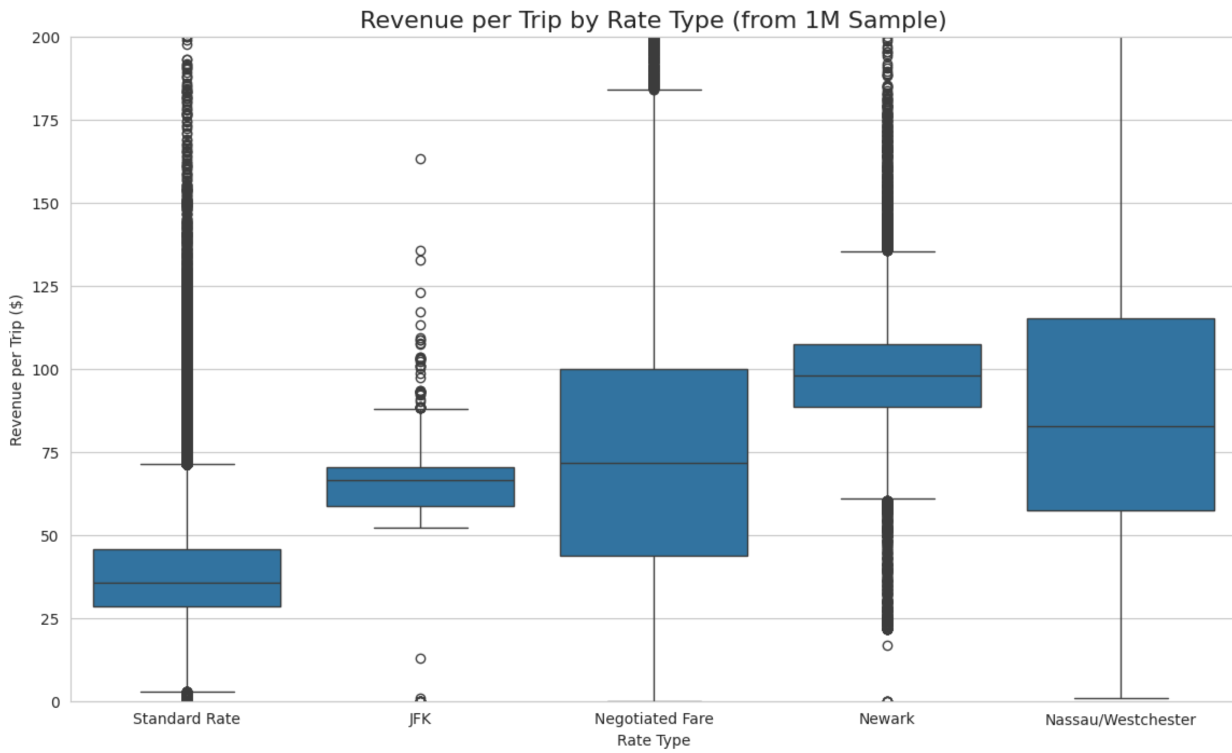
4.3. Impact of Trip Characteristics

Analysis of trip attributes revealed a **strong, positive correlation** between trip distance and revenue. Conversely, **passenger count had no meaningful impact**. The most significant factor proved to be the **rate code**, with special fares for airports (`Newark`) generating dramatically higher revenue.



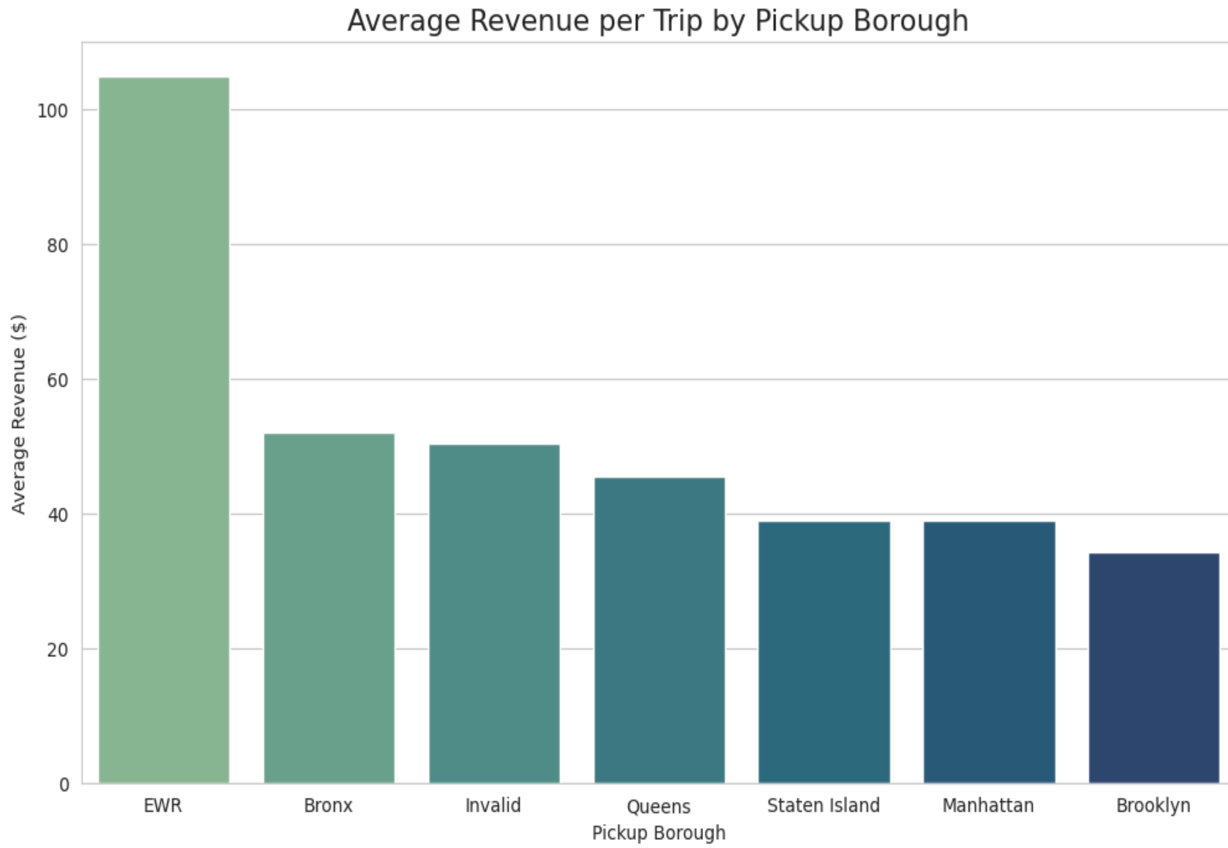
4.4. Payment Type & Tipping Behavior Analysis

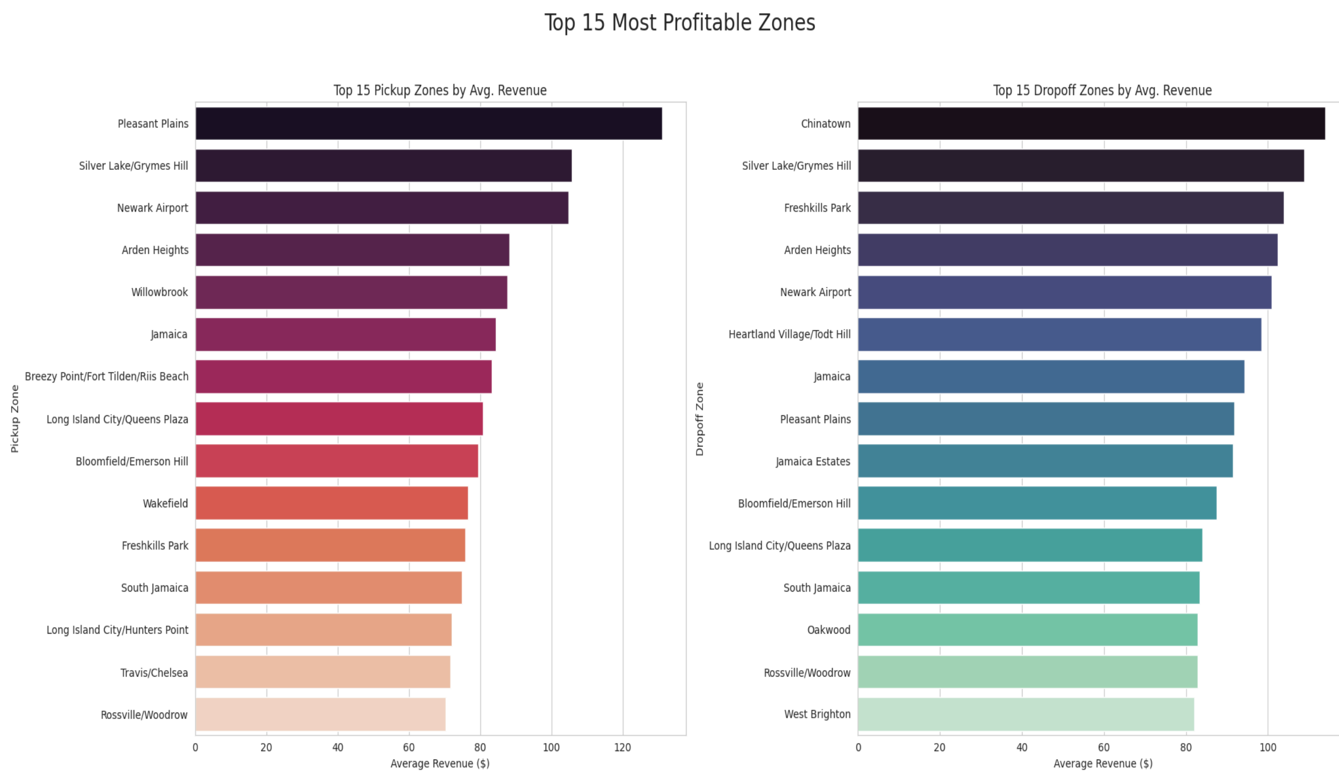
On average, trips paid with a **Credit Card ($42.88)** earn over 40% more than trips paid with **Cash ($30.38)**. This difference is driven almost entirely by tipping: **97% of credit card trips** included a recorded tip, compared to virtually zero for cash trips.



4.5. Location-Based Insights

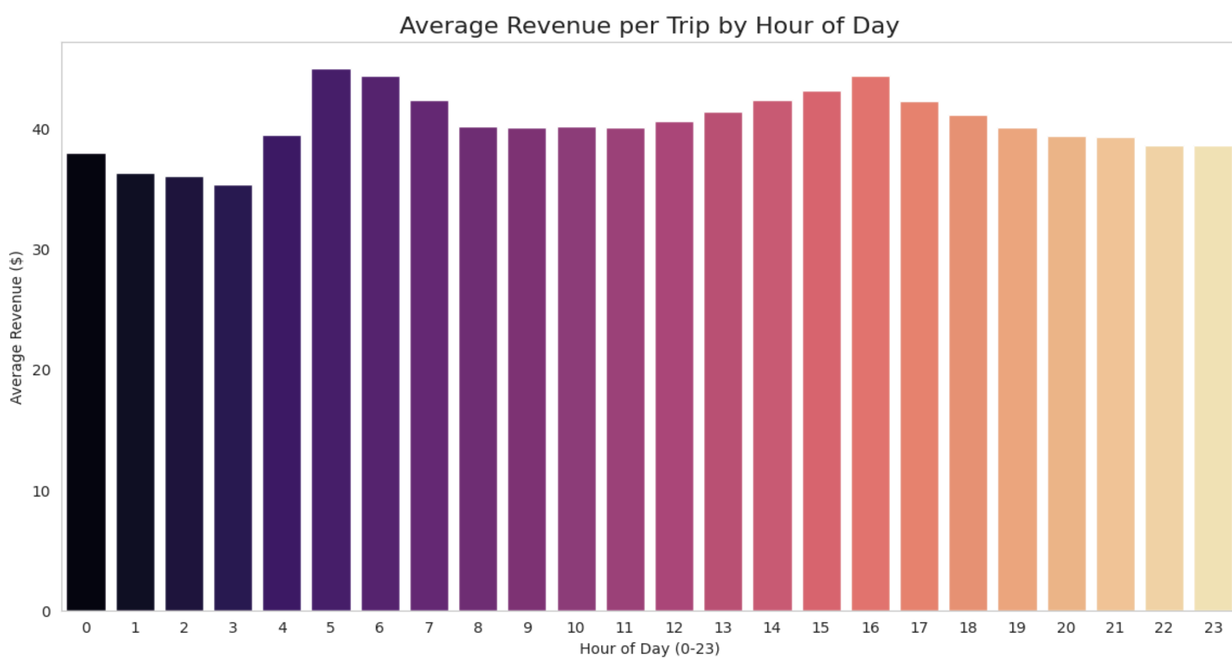
Geographic analysis showed that **EWR (Newark Airport)** was by far the most lucrative area. The data reveals that the highest average revenue comes from **long-distance, inter-borough trips**, often originating from the city's geographical extremes (e.g., Staten Island).

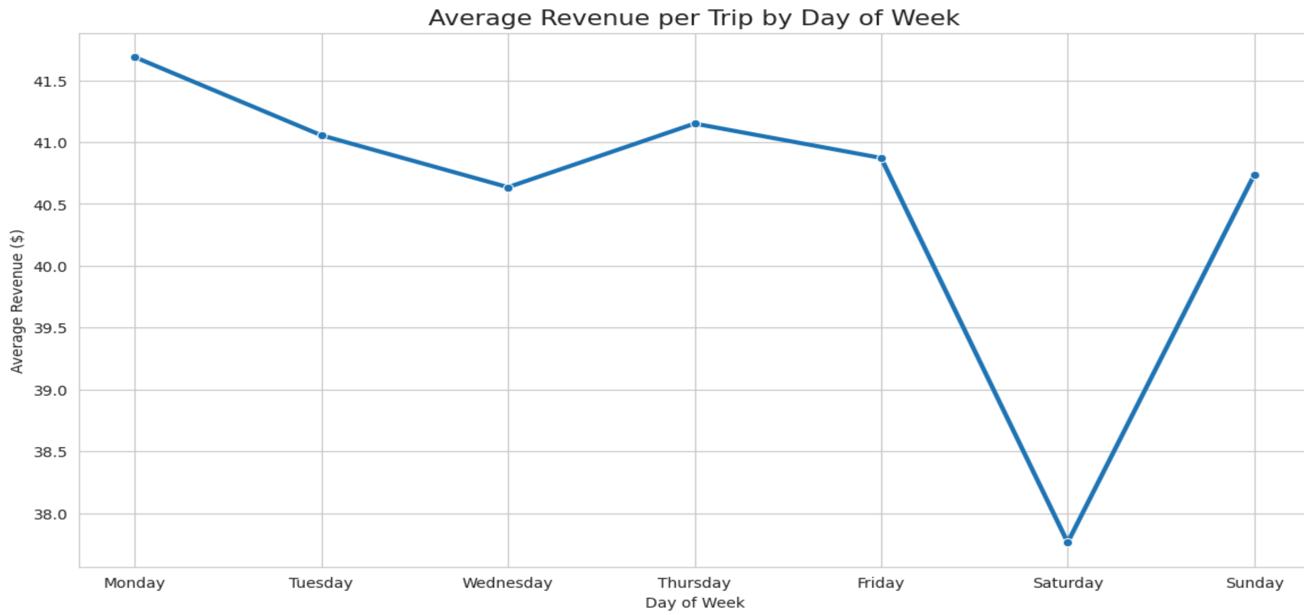




4.6. Temporal Patterns

Contrary to common assumptions, **weekdays were more profitable on average** than weekends, with Saturday showing the lowest average revenue. The most lucrative periods were the **weekday commute hours**, peaking at **6 AM** and **4 PM (16:00)**.





5. Hypothesis Scorecard: Final Results

#	Hypothesis	Verdict	Rationale
H1	**Payment** (Card > Cash)	✅ **Supported**	Card trips earn >40% more on average due to a 97% recorded tip rate.
H2	**Distance** (Longer = More Revenue)	✅ **Supported**	A strong, direct positive correlation exists between distance and revenue.
H3	**Weekend** (Weekend > Weekday)	❌ **Not Supported**	Saturday is the *least* profitable day on average.
H4	**Passenger Count**	❌ **Not Supported**	There is no significant difference in revenue based on the number of passengers.
H5	**Rate Type** (Special > Standard)	✅ **Supported**	Airport (`Newark`) rates generate up to 3x the median revenue.
H6	**Borough** (Manhattan > Others)	❌ **Not Supported**	Manhattan has lower average revenue due to a high volume of short trips.
H7	**Late-Night** (Night > Day)	❌ **Not Supported**	Late-night hours are the *least* profitable; peaks occur during commutes.

6. Summary and Recommendations

6.1. Key Findings

- **Primary Revenue Drivers:** The most powerful predictors of a high-revenue trip are a long **trip distance** and a special **rate type**, particularly airport fares.
- **The Power of Digital Payments: Credit card** is the most valuable payment method, adding substantial revenue almost exclusively through in-app tipping.
- **The Long-Haul Outer Boroughs:** The most profitable trips are long-haul journeys that often originate from **airports** or the city's geographical extremes (e.g., Staten Island).
- **Debunking Common Myths:** The highest average revenue comes from **weekday commute hours**, not weekends or late nights.

6.2. Business Recommendations

1. **Launch an "Airport Specialist" Program:** Create incentives (e.g., weekly bonuses) for drivers who complete a high number of trips to/from **Newark, JFK, and LaGuardia** to focus supply on these high-value zones.
2. **Optimize the Dispatch Algorithm for Long Hauls:** Prioritize matching drivers with requests from top-earning, distant pickup zones (e.g., in Staten Island).
3. **Enhance In-App Tipping Features:** A/B test dynamic tipping suggestions based on trip distance or rate type (e.g., "A $20 tip is common for airport trips like this").
4. **Implement Strategic "Commuter" Pricing:** Introduce small, targeted price adjustments during the proven most profitable hours **(6-7 AM and 4-5 PM on weekdays)**.
5. **Empower Drivers with Data:** Use the driver app to share anonymized insights, like a weekly "Opportunity Heatmap" showing the most profitable zones.

Here's a shorter, more concise version:

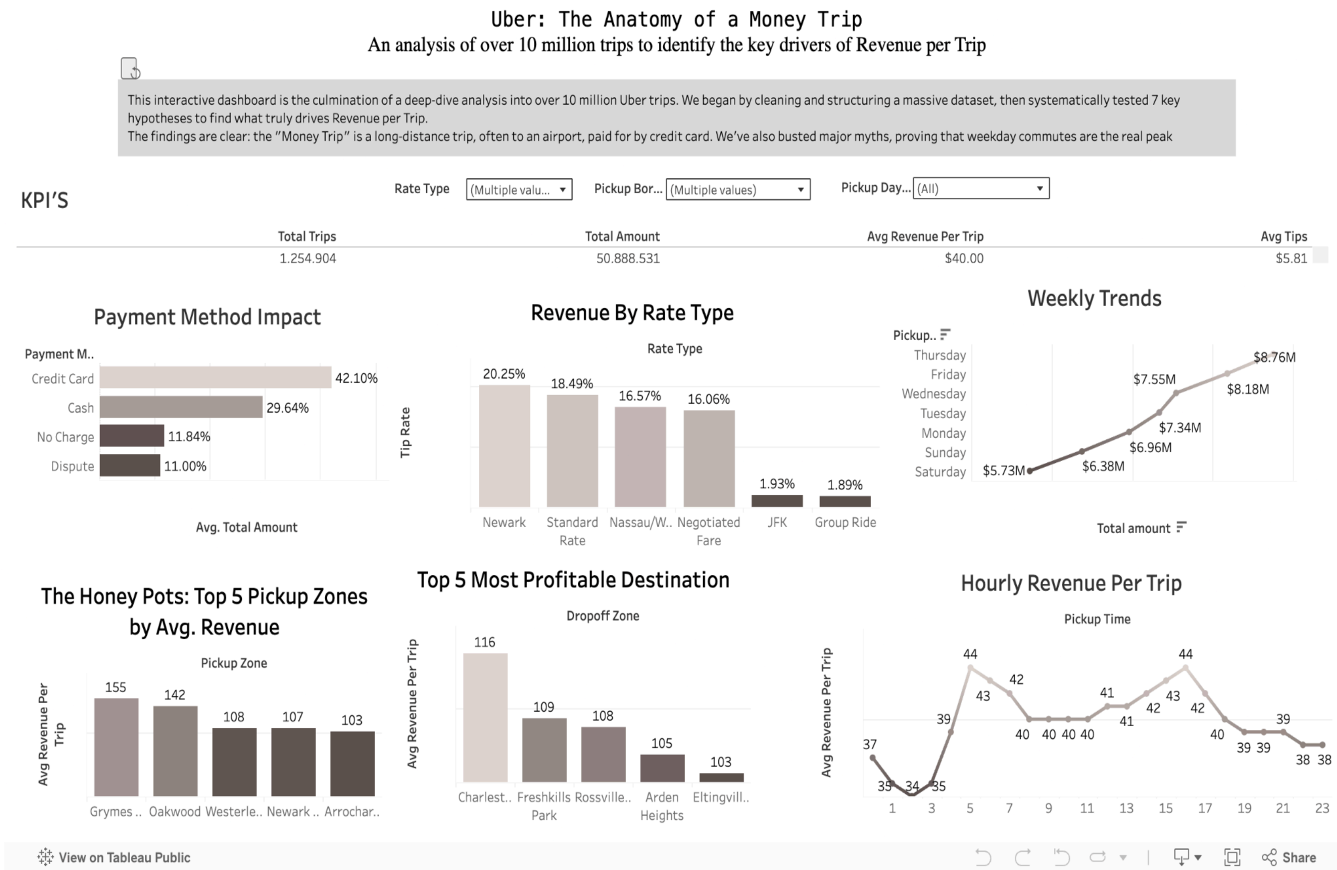
---

7. Interactive Dashboard: Tableau Visualization

To make the findings from this analysis accessible and actionable for business stakeholders, an **interactive Tableau dashboard** was developed. This dashboard consolidates all key insights into a single, user-friendly interface that enables dynamic exploration of revenue patterns.

7.1. Dashboard Overview

The dashboard, titled **"Uber: The Anatomy of a Money Trip"**, provides real-time filtering capabilities by **Rate Type**, **Pickup Borough**, and **Day of Week**, allowing users to explore revenue drivers across different operational scenarios.



7.2. Key Components

The dashboard includes the following visualizations:

- **KPI Summary Panel:** Displays critical metrics including Total Trips (1.25M), Total Revenue ($50.9M), Average Revenue per Trip ($40.00), and Average Tips ($5.81)

- **Payment Method Impact:** Shows that Credit Card payments generate 40% more revenue than Cash, driven primarily by higher tip capture rates

- **Revenue by Rate Type:** Confirms Newark Airport trips as the highest revenue generator, accounting for over 20% of total revenue

- **Weekly Trends:** Reveals that Thursday is the most lucrative day ($8.76M), while Saturday generates the lowest revenue ($5.73M), contradicting common weekend assumptions

- **Top Pickup & Dropoff Zones:** Identifies Staten Island locations and Newark Airport as the most profitable "honey pots," with average revenues exceeding $100 per trip

- **Hourly Revenue Patterns:** Demonstrates clear peaks during weekday commute hours (6-7 AM and 4-5 PM), with average revenues of $43-44 per trip

7.3. Strategic Value

The dashboard's interactive filters enable stakeholders to answer specific business questions in real-time and validate strategic decisions. This transforms the analysis from a static report into a dynamic decision-support tool for operations teams, driver experience managers, and pricing strategists.

---

8. Next Steps

1. **Operational Review:** The `anomalies_for_review.csv` file should be sent to the Data Engineering team to investigate the root cause of data errors.
2. **Controlled A/B Testing:** Implement recommendations #3 (Tipping) and #4 (Pricing) in a limited manner to measure their direct impact on revenue.
3. **Predictive Modeling:** Use the key features from this analysis to build a machine learning model to predict the potential revenue of a trip.

9. Appendix

8.1. Analysis of Anomalies and Edge Cases

An investigation of edge cases idIn the report, identified over 13,000 trips with zero distance but a significant fare. These were found to be overwhelmingly categorized as a **`Negotiated Fare`**, suggesting they are pre-arranged flat-rate trips rather than data errors. Extreme outliers were also isolated and exported for further review.