

NYC Yellow Taxi Data Analysis Report

Exploratory Data Analysis and Strategic Insights

With Comprehensive Visualizations

Executive Summary

This report presents a comprehensive analysis of New York City Yellow Taxi trip data from 2023, focusing on operational efficiency, pricing strategies, and customer experience optimization. The analysis was conducted on a representative sample of 363,511 taxi trips, providing actionable insights for taxi operations optimization through detailed visualizations and statistical analysis.

Key Findings:

- Peak demand hours: 5-7 PM on weekdays with 2.5+ million trips per hour
- Highest revenue quarters: Q2 and Q4 contributing 53.54% of annual revenue
- Strong fare-distance correlation: 0.96 correlation coefficient
- Geographic concentration: Manhattan zones dominate both pickups and dropoffs
- Payment preferences: 81.62% credit card, 17.22% cash payments

1. Data Overview and Methodology

Dataset Characteristics

- Original dataset size: 379,268 records (1% sample from 12 monthly files)
- Final cleaned dataset: 363,511 records after outlier removal
- Sampling methodology: Hourly stratified sampling across all dates
- Data quality: Comprehensive cleaning including outlier removal and missing value imputation

Data Quality Analysis

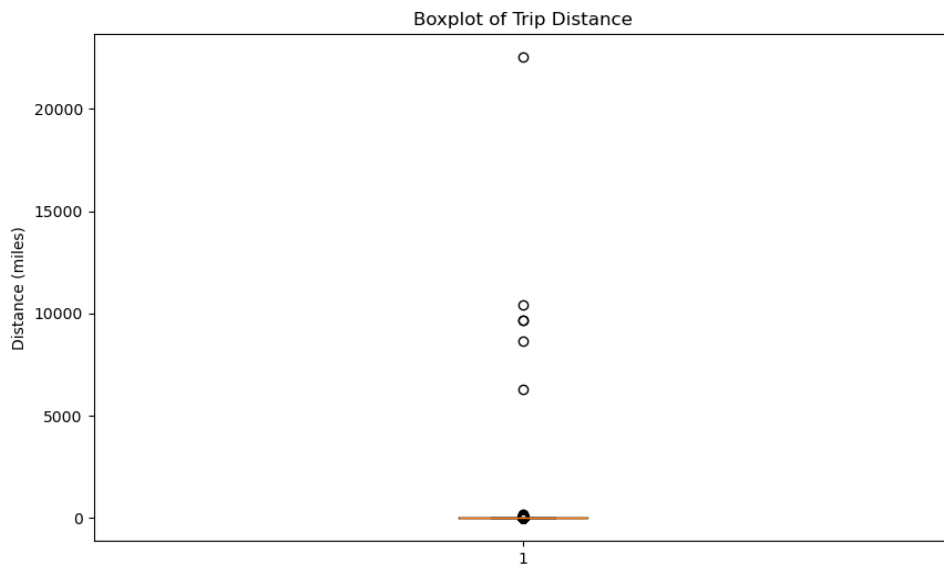


Figure 1.1: Boxplot of Trip Distance showing outlier distribution

2. Temporal Analysis

2.1 Hourly Demand Patterns

The analysis of hourly pickup patterns reveals distinct peak periods that align with typical urban commuting and activity patterns. The visualizations below demonstrate clear demand cycles throughout the day.

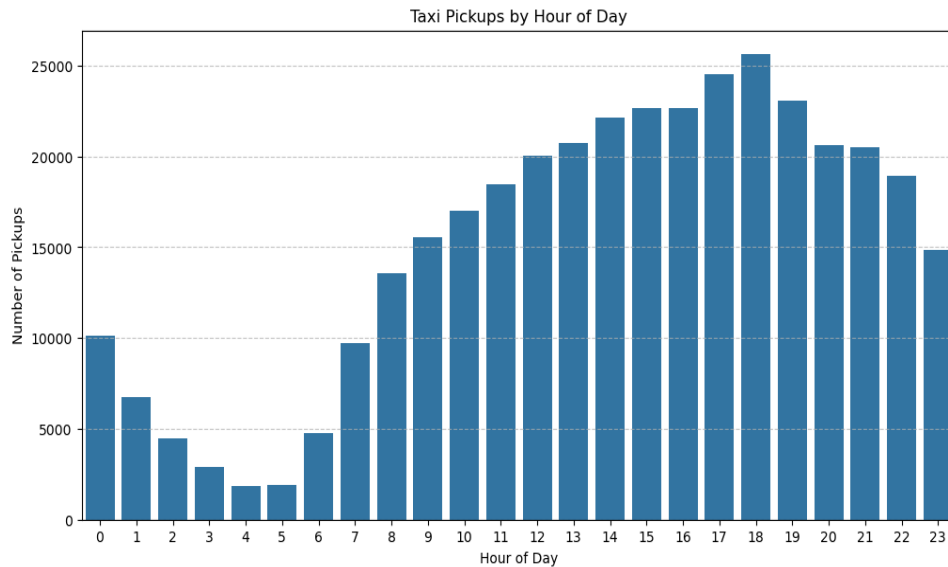


Figure 2.1: Taxi Pickups by Hour of Day showing peak demand at 6 PM

2.2 Weekly Patterns

Weekly demand patterns show interesting variations between weekdays and weekends, with distinct characteristics that inform operational strategies.

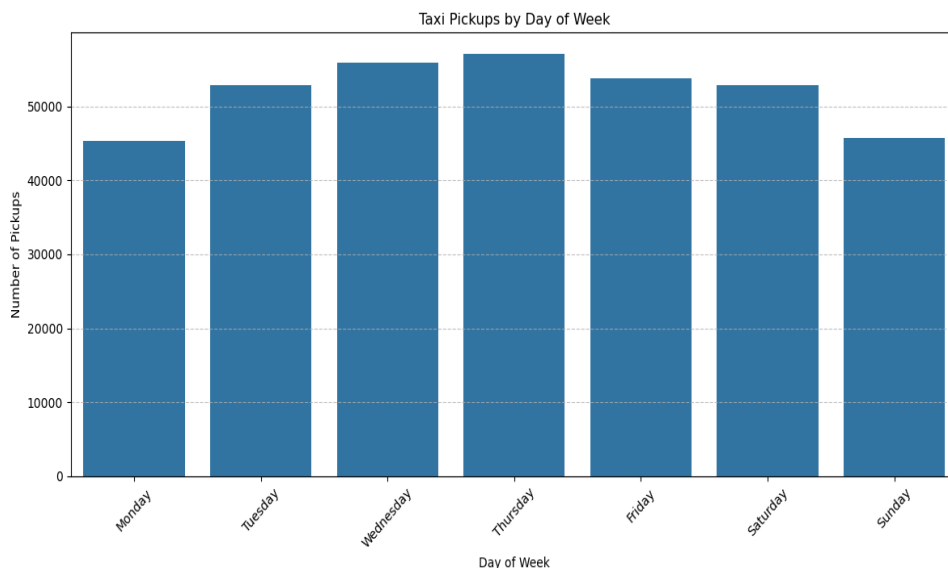


Figure 2.2: Taxi Pickups by Day of Week showing weekday dominance

2.3 Monthly and Seasonal Trends

Seasonal variations in taxi demand provide insights for capacity planning and resource allocation throughout the year.

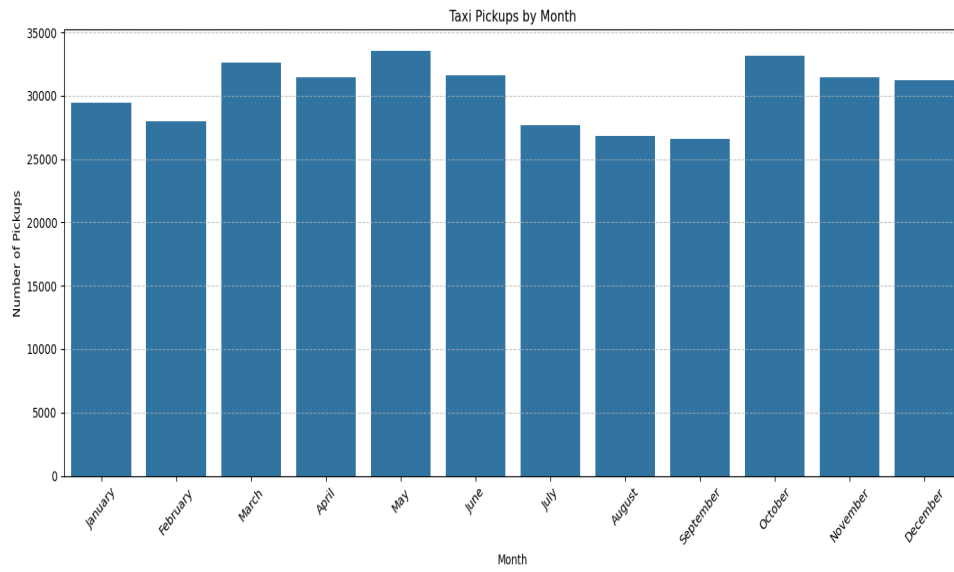


Figure 2.3: Taxi Pickups by Month showing seasonal variations

3. Financial Analysis

3.1 Revenue Trends

Monthly revenue analysis reveals seasonal patterns and opportunities for revenue optimization through strategic pricing and service adjustments.

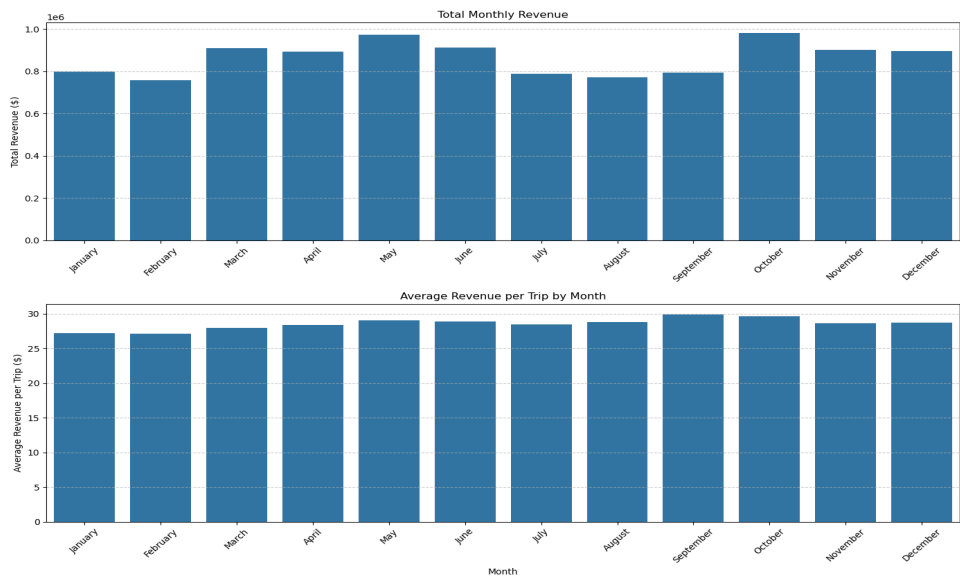


Figure 3.1: Total Monthly Revenue and Average Revenue per Trip

3.2 Quarterly Revenue Distribution

Quarterly analysis provides insights into seasonal business cycles and helps in annual planning and forecasting.

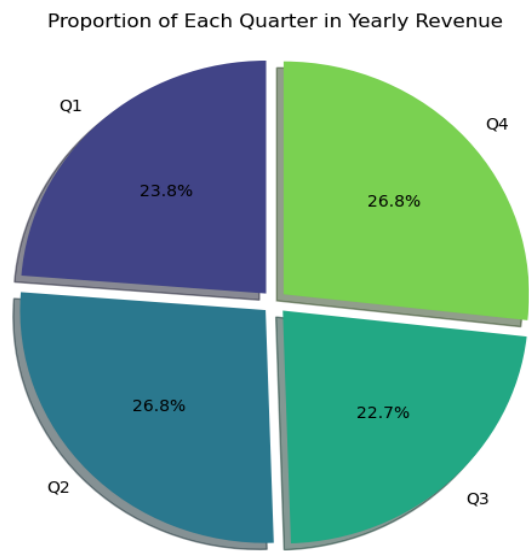


Figure 3.2: Quarterly Revenue Distribution showing Q2 and Q4 dominance

3.3 Distance-Fare Relationship

The relationship between trip distance and fare amount is fundamental to understanding pricing dynamics and customer value perception.

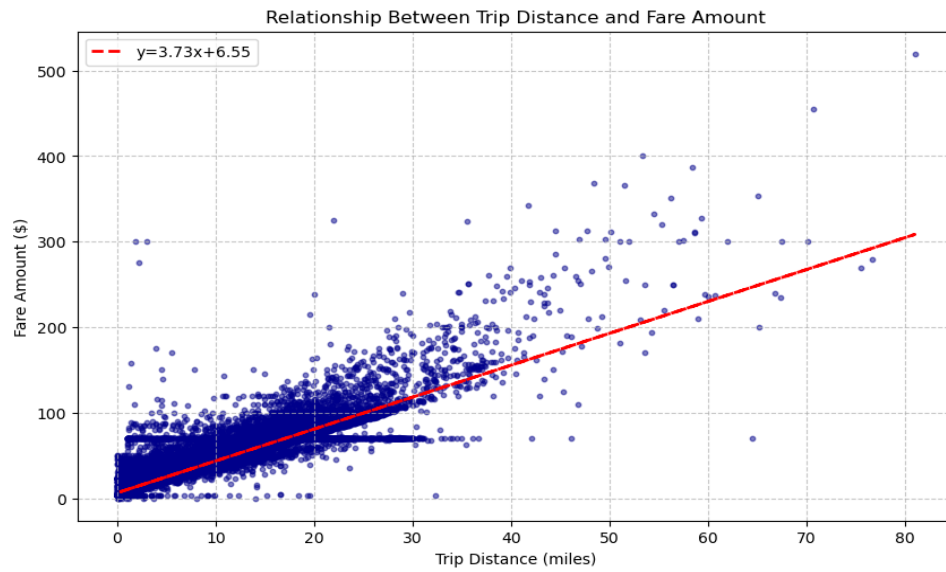


Figure 3.3: Relationship Between Trip Distance and Fare Amount ($r=0.96$)

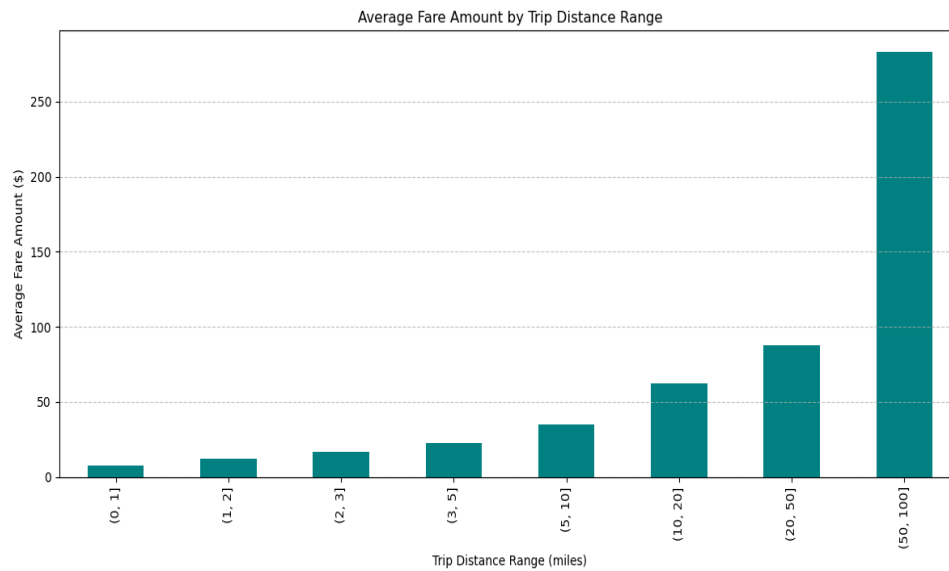


Figure 3.4: Average Fare Amount by Trip Distance Range

4. Correlation Analysis

4.1 Trip Duration vs Fare

Understanding the relationship between trip duration and fare helps in optimizing routing and pricing strategies.

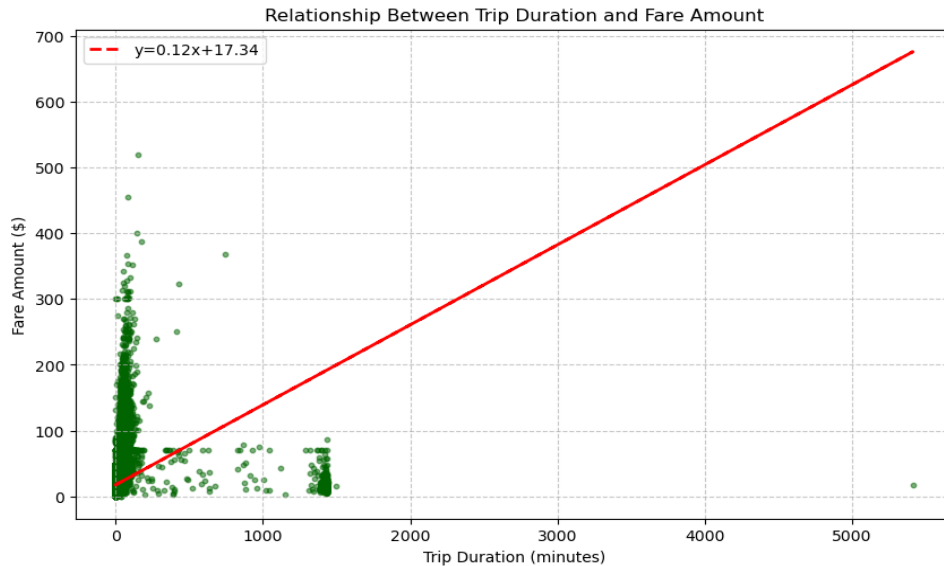


Figure 4.1: Relationship Between Trip Duration and Fare Amount

4.2 Passenger Count Analysis

Passenger count analysis reveals pricing efficiency and opportunities for group ride incentives.

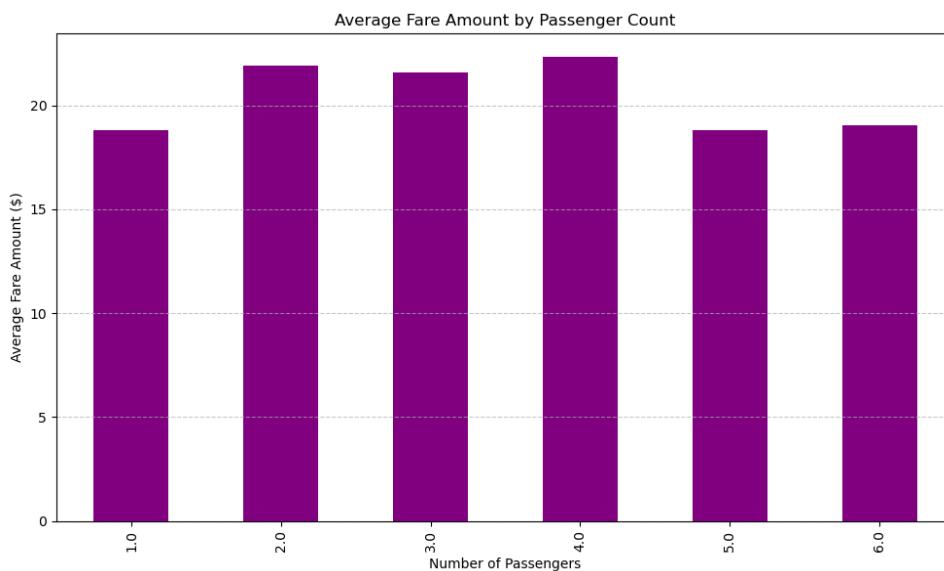


Figure 4.2: Average Fare Amount by Passenger Count

4.3 Tip Behavior Analysis

Tip patterns provide insights into customer satisfaction and service quality indicators.

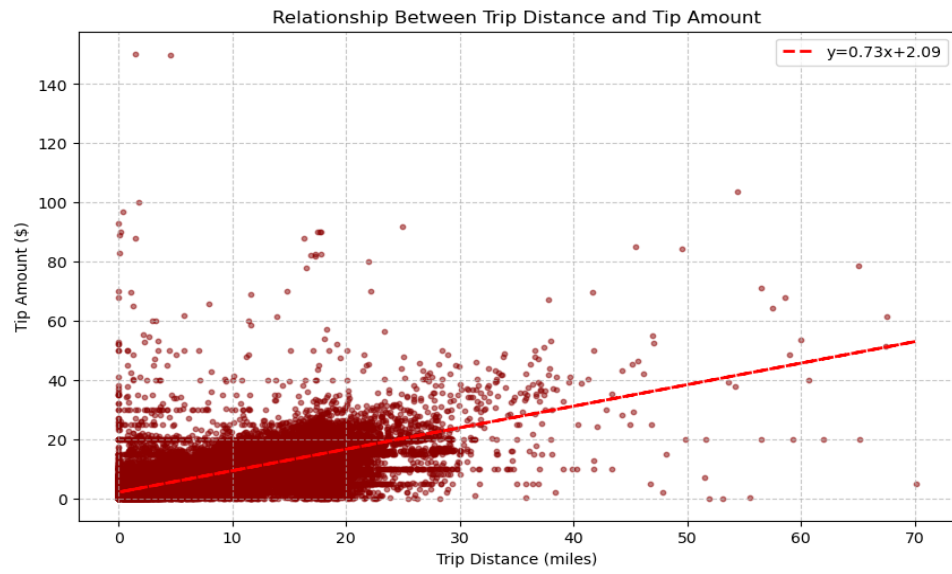


Figure 4.3: Relationship Between Trip Distance and Tip Amount

5. Payment Method Analysis

Payment method preferences significantly impact operational efficiency and customer experience, with implications for tip recording and processing.

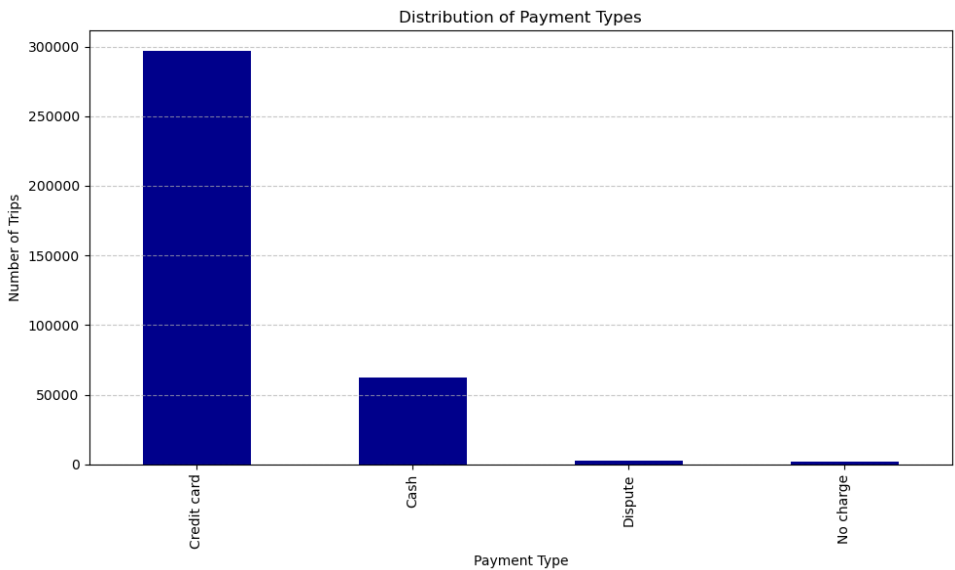


Figure 5.1: Distribution of Payment Types (Bar Chart)

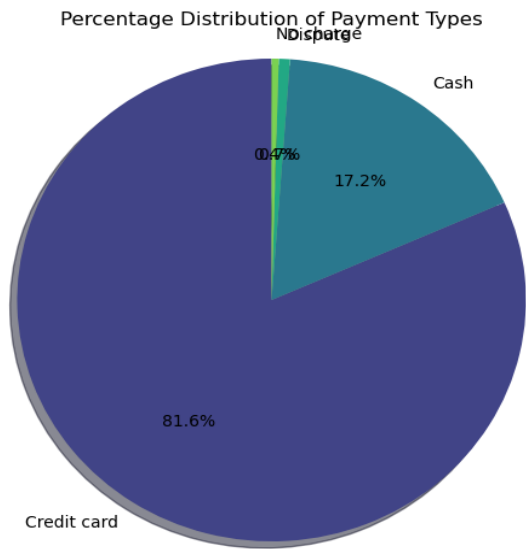


Figure 5.2: Percentage Distribution of Payment Types (Pie Chart)

6. Geographic Analysis

6.1 Spatial Distribution Maps

Geographic analysis reveals spatial patterns in taxi demand, helping identify high-traffic zones and optimization opportunities.

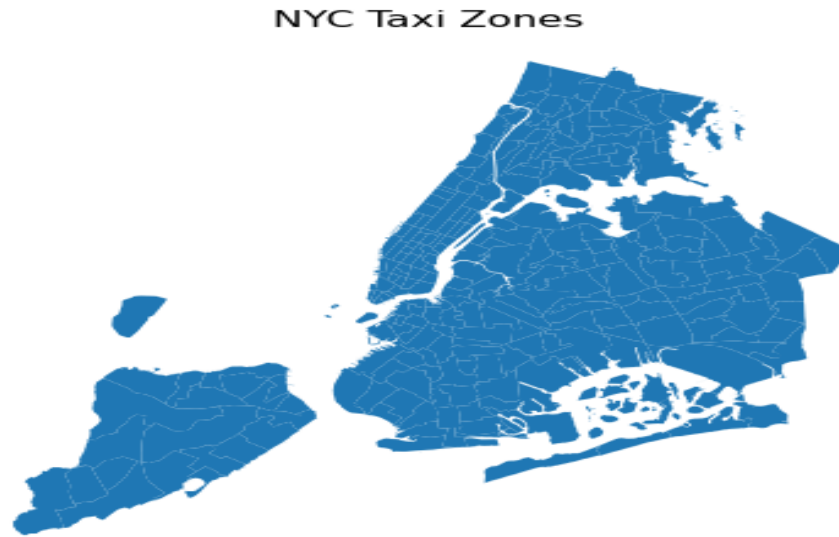


Figure 6.1: NYC Taxi Zones Geographic Distribution

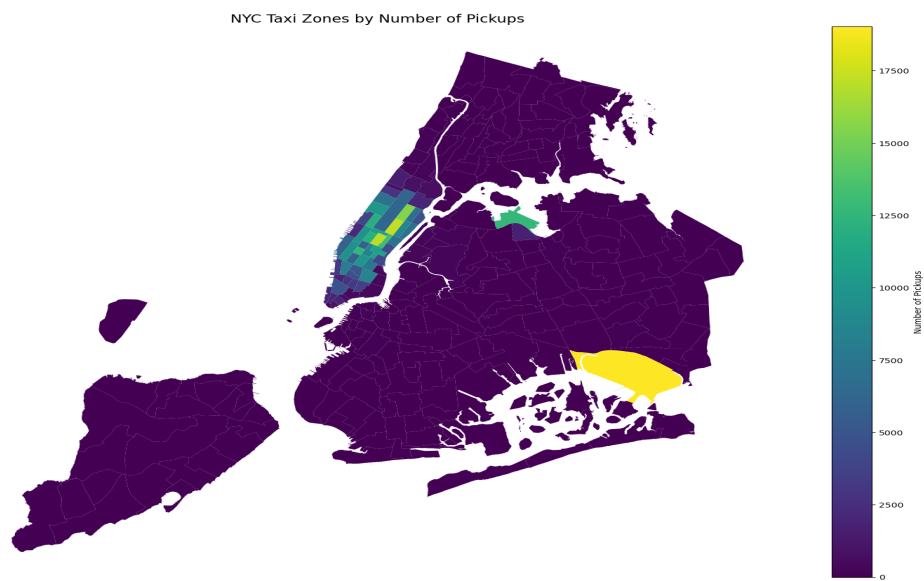


Figure 6.2: NYC Taxi Zones by Number of Pickups

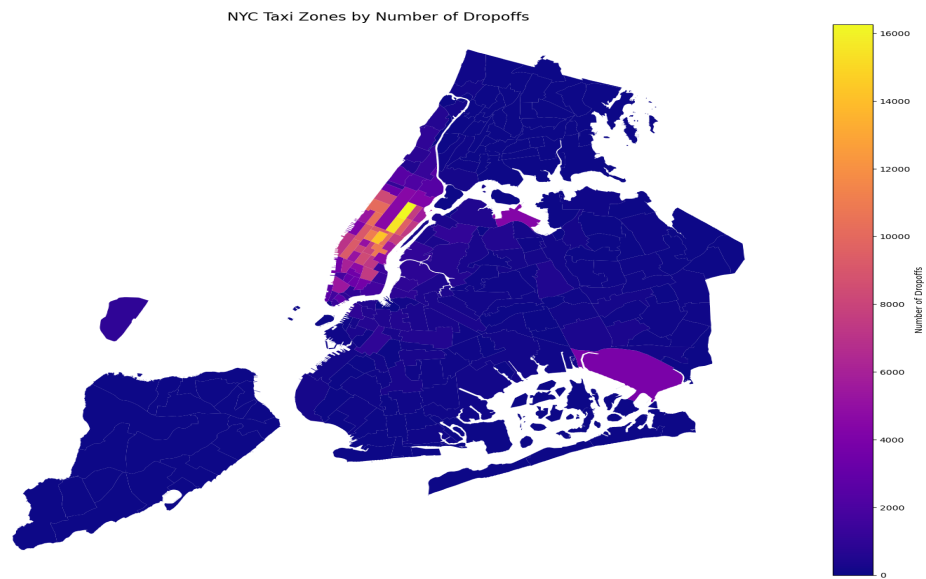


Figure 6.3: NYC Taxi Zones by Number of Dropoffs

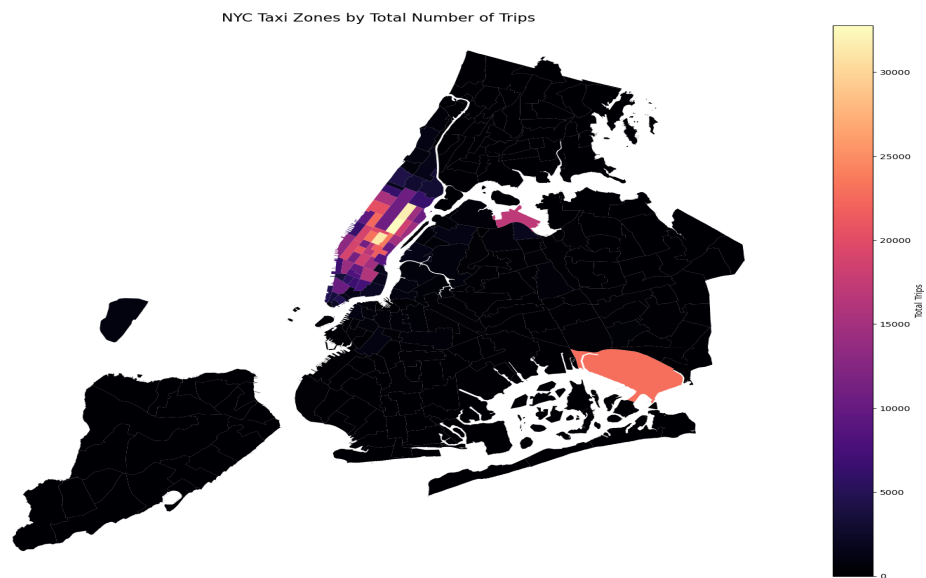


Figure 6.4: NYC Taxi Zones by Total Number of Trips

7. Operational Efficiency Analysis

7.1 Speed and Traffic Analysis

Speed analysis by time of day reveals traffic patterns and optimal routing opportunities for improved efficiency.

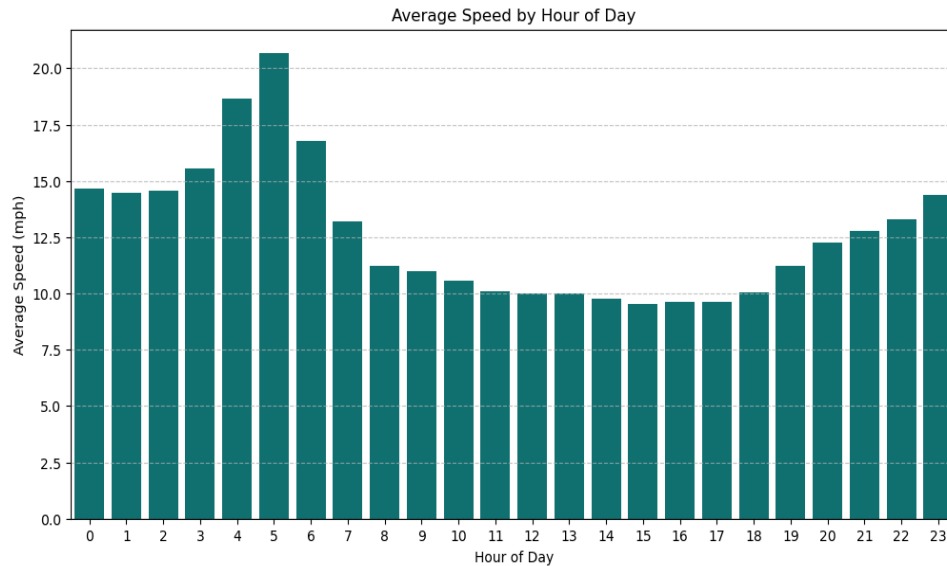


Figure 7.1: Average Speed by Hour of Day

7.2 Demand Patterns

Detailed demand analysis helps in fleet positioning and resource allocation strategies.

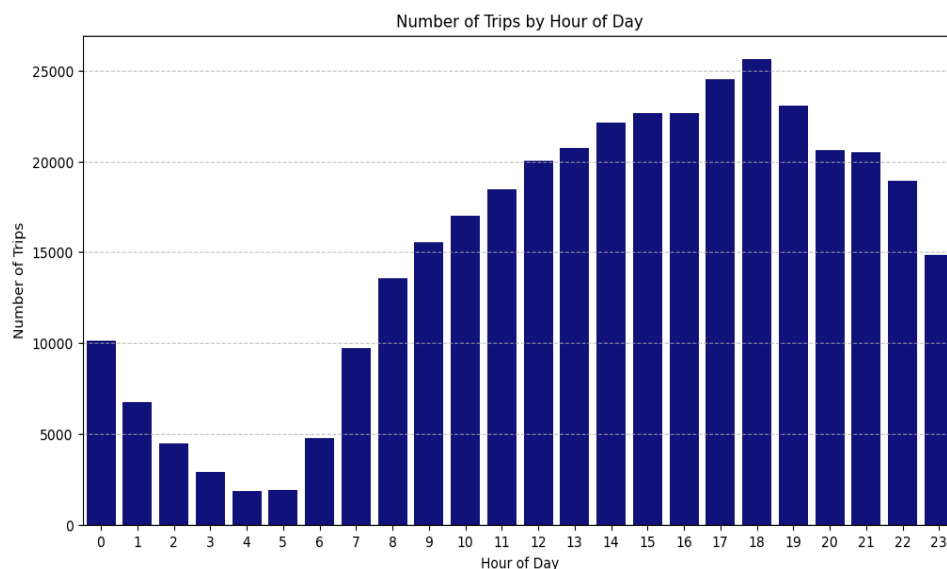


Figure 7.2: Number of Trips by Hour of Day

7.3 Weekday vs Weekend Patterns

Comparing weekday and weekend patterns reveals distinct operational requirements for different periods.

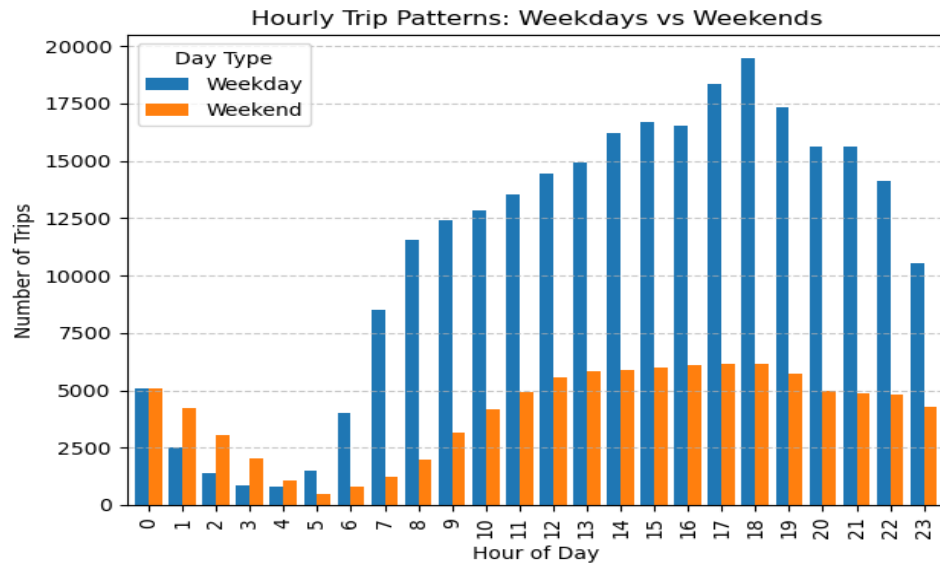


Figure 7.3: Hourly Trip Patterns - Weekdays vs Weekends

8. Zone-Specific Analysis

8.1 Top Pickup and Dropoff Zones

Identifying high-demand zones is crucial for strategic positioning and service optimization.

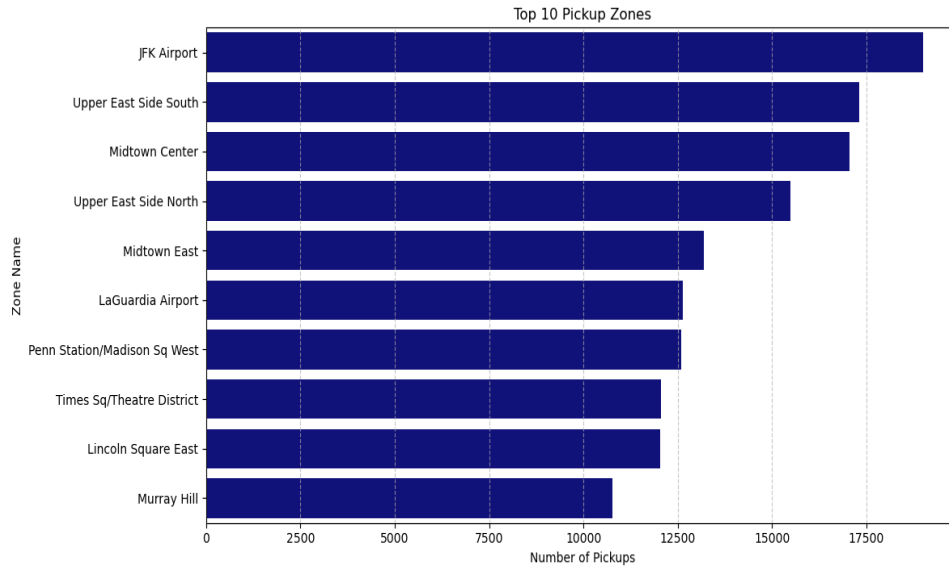


Figure 8.1: Top 10 Pickup Zones

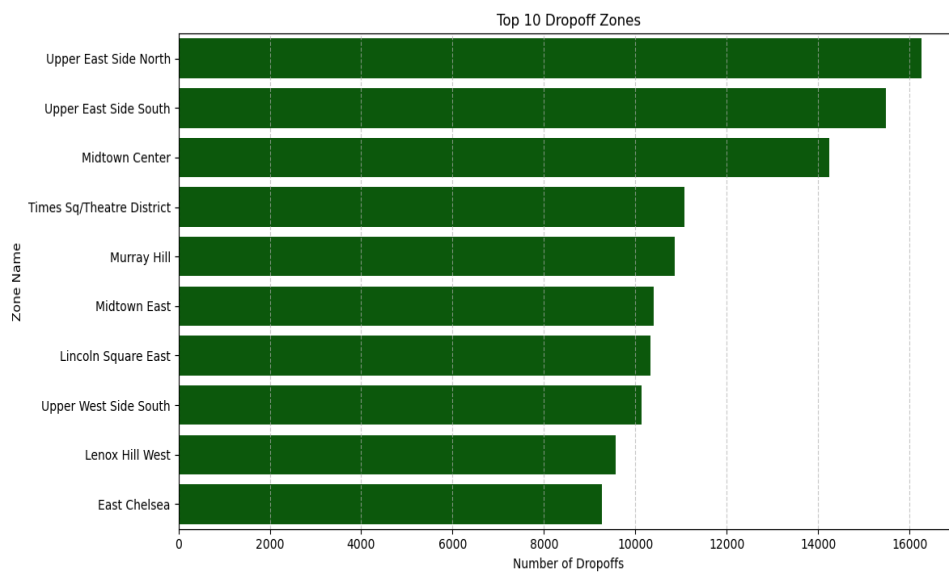


Figure 8.2: Top 10 Dropoff Zones

9. Advanced Revenue Analysis

9.1 Revenue Distribution

Understanding revenue distribution helps in identifying optimization opportunities and pricing strategies.

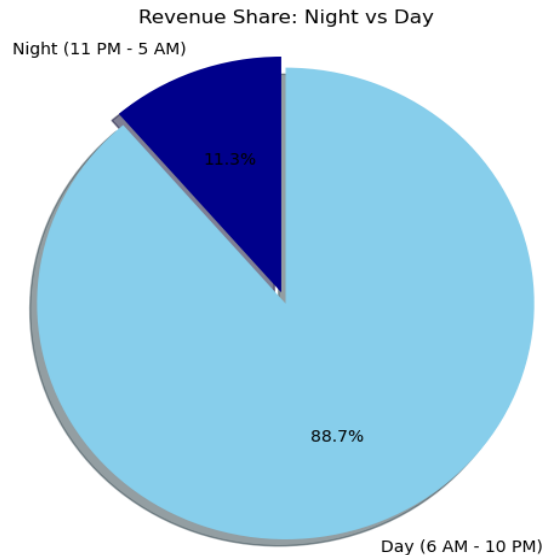


Figure 9.1: Revenue Share - Night vs Day

9.2 Fare Analysis by Passenger Count

Analyzing fare efficiency by passenger count reveals opportunities for group ride incentives and pricing optimization.

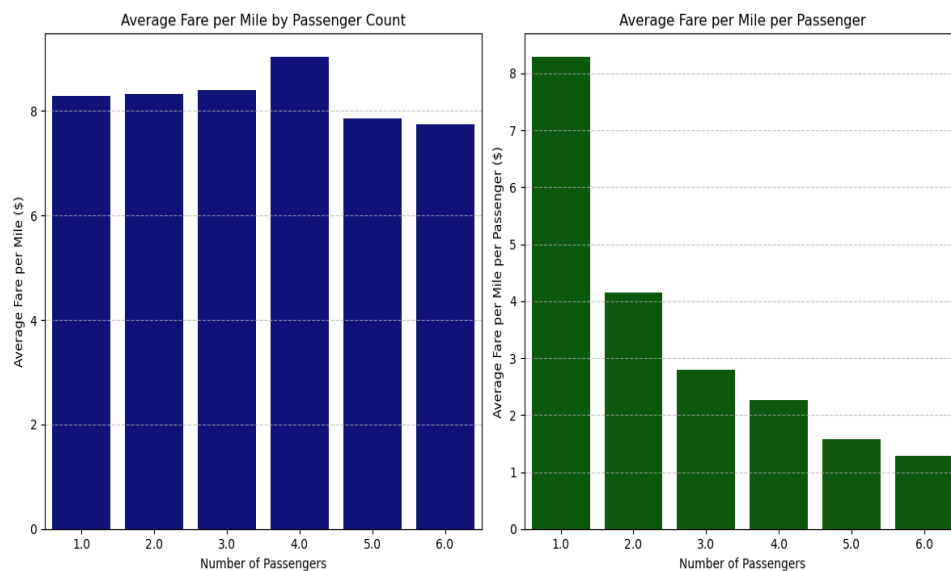


Figure 9.2: Average Fare per Mile and per Passenger by Passenger Count

9.3 Temporal Fare Patterns

Fare variations by time and day provide insights for dynamic pricing strategies.

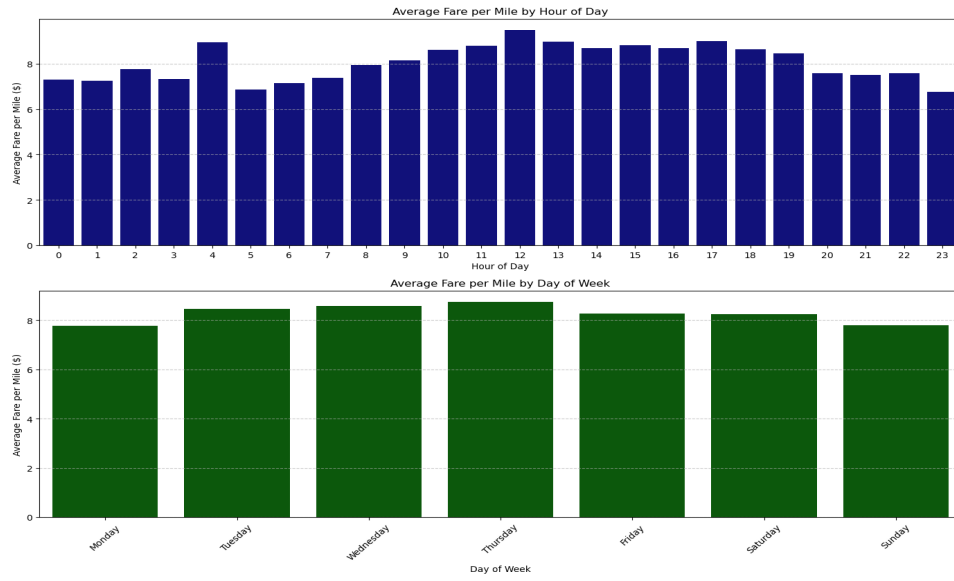


Figure 9.3: Average Fare per Mile by Hour and Day of Week

10. Vendor Performance Analysis

10.1 Vendor Fare Comparison

Comparing vendor performance helps in understanding competitive positioning and pricing strategies.

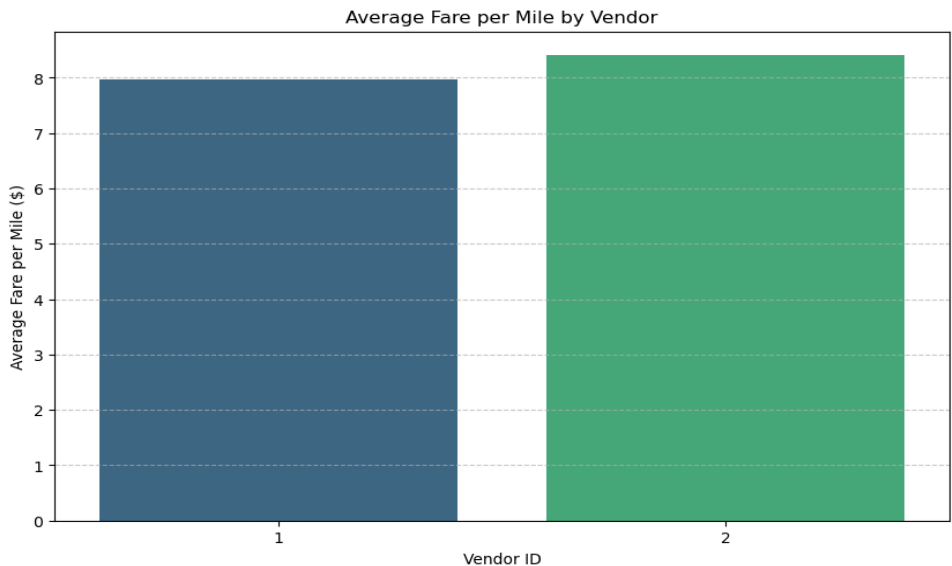


Figure 10.1: Average Fare per Mile by Vendor

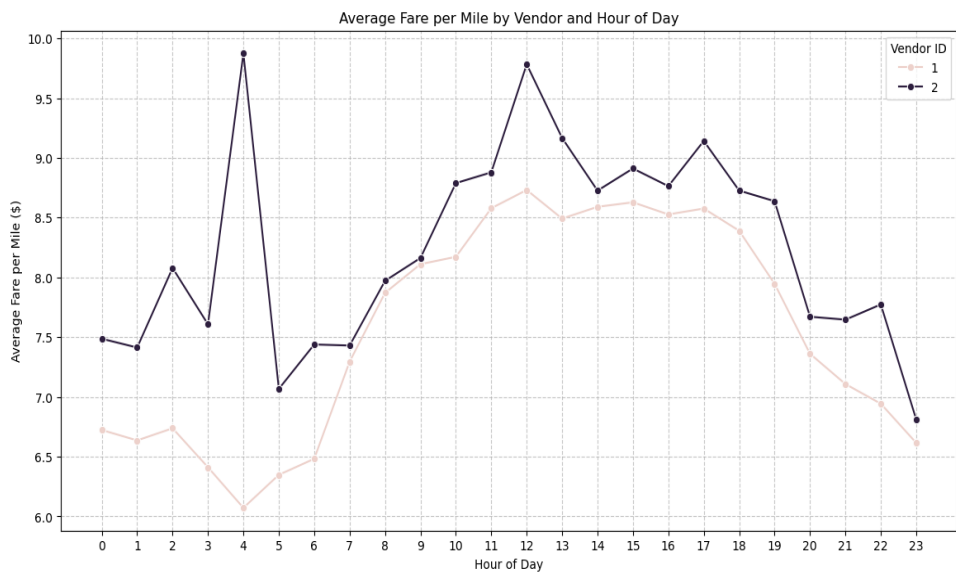


Figure 10.2: Average Fare per Mile by Vendor and Hour of Day

10.2 Distance-Tiered Vendor Analysis

Distance-based vendor comparison reveals pricing strategies across different trip lengths.

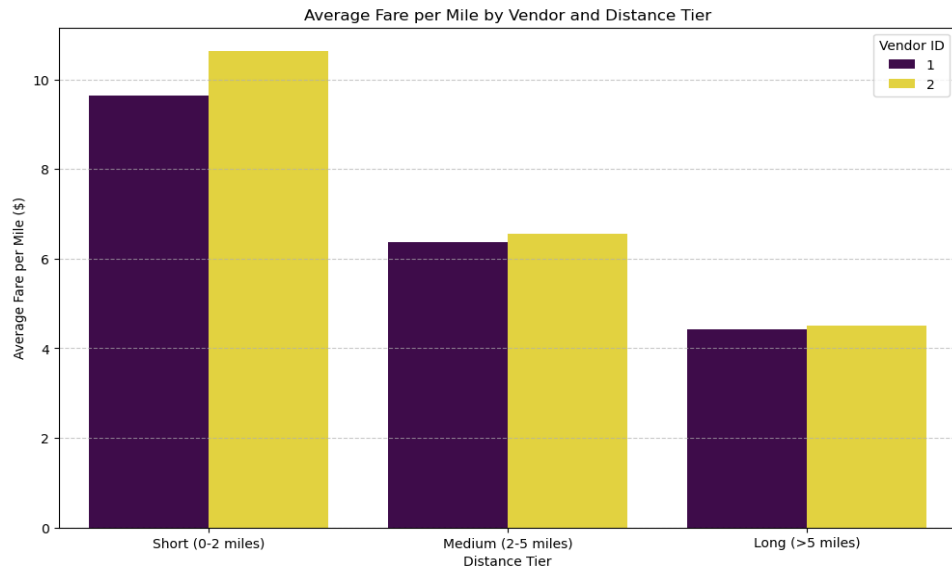


Figure 10.3: Average Fare per Mile by Vendor and Distance Tier

11. Customer Experience Analysis

11.1 Tip Behavior Patterns

Understanding tip patterns provides insights into customer satisfaction and service quality factors.

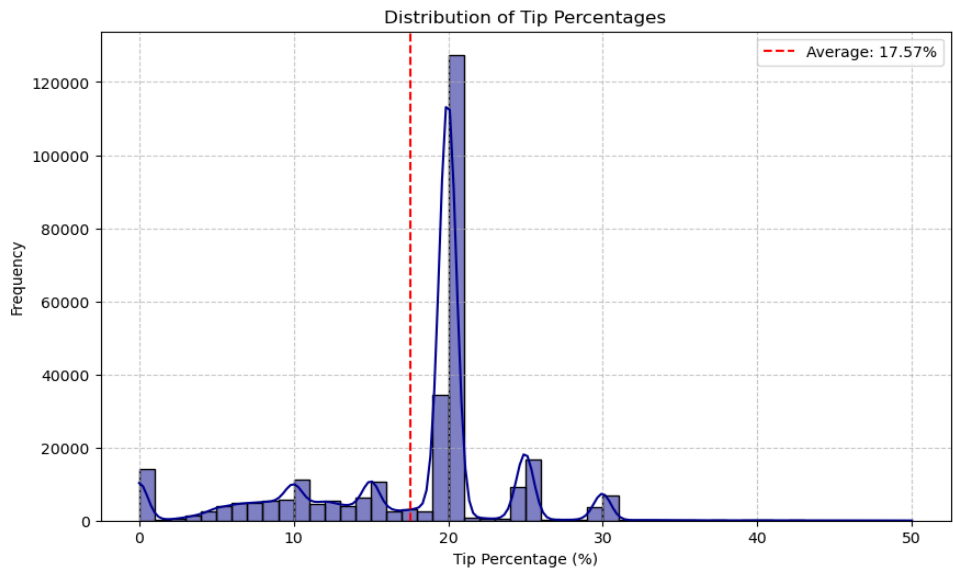


Figure 11.1: Distribution of Tip Percentages

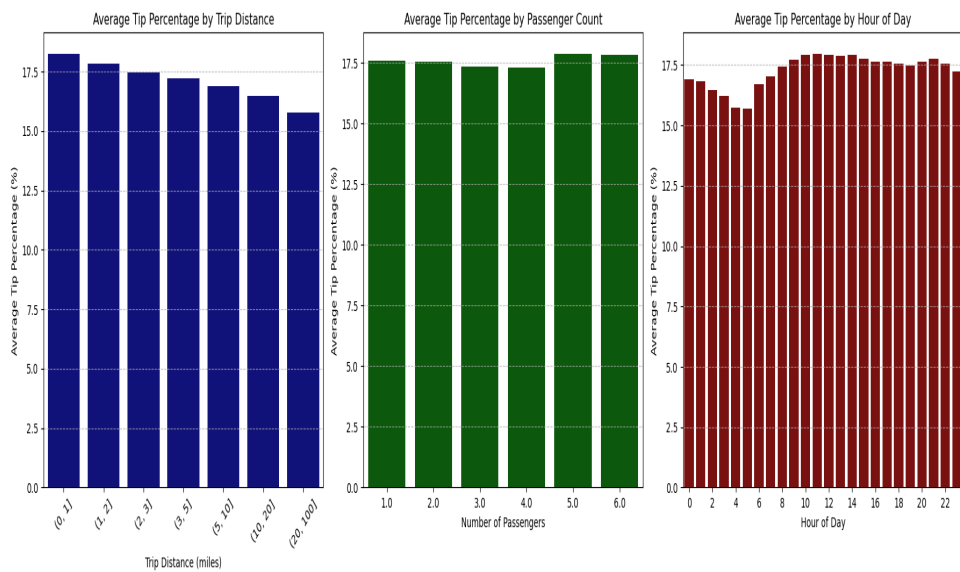


Figure 11.2: Average Tip Percentage by Distance, Passenger Count, and Hour

11.2 Passenger Count Variations

Passenger count patterns across time and location provide insights for capacity planning and service optimization.

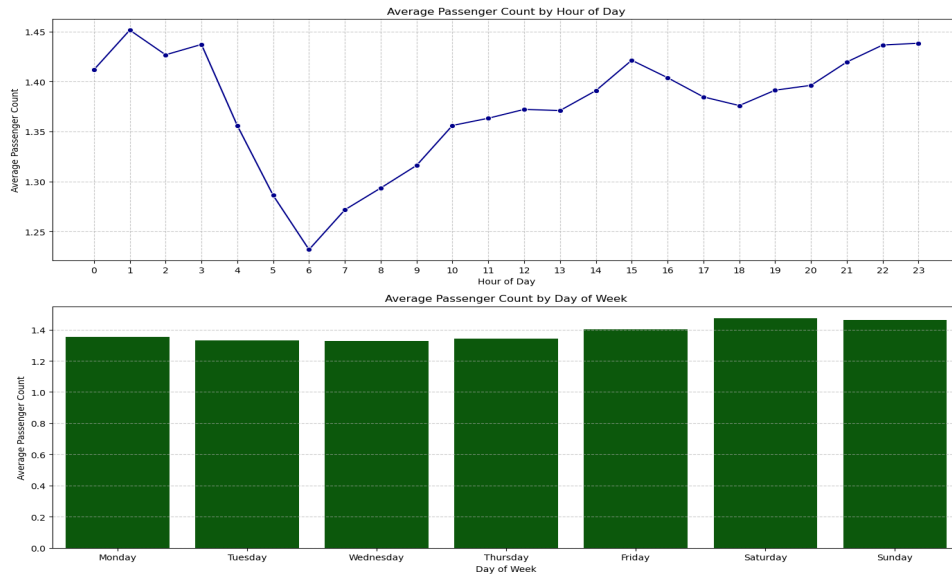


Figure 11.3: Average Passenger Count by Hour and Day of Week

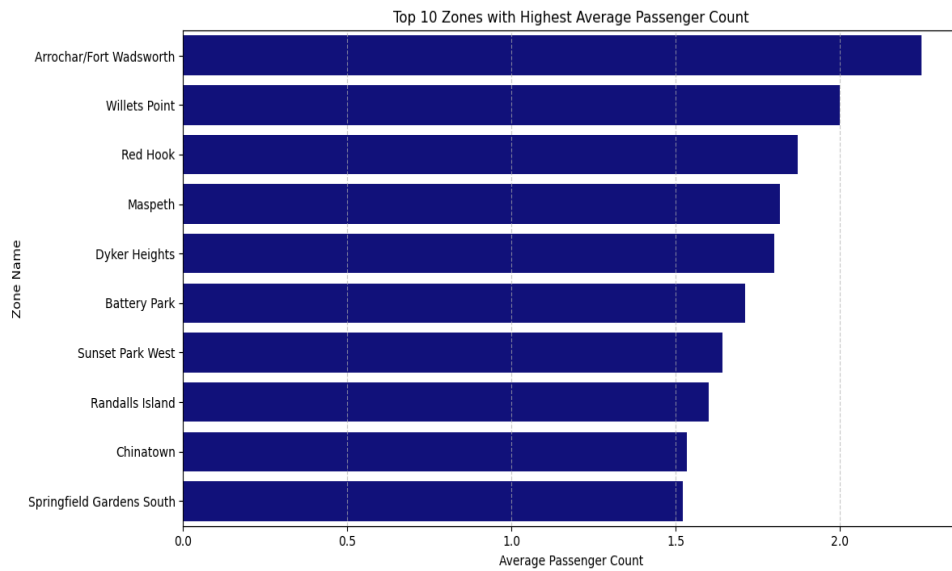


Figure 11.4: Top 10 Zones with Highest Average Passenger Count

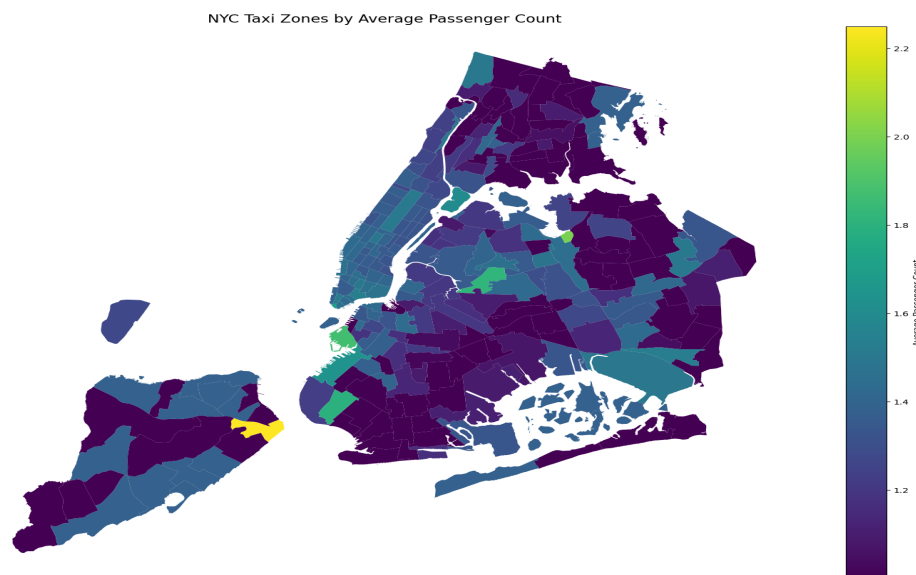


Figure 11.5: NYC Taxi Zones by Average Passenger Count

12. Surcharge and Fee Analysis

12.1 Surcharge Application Patterns

Understanding when and where surcharges are applied helps optimize pricing strategies and customer communication.

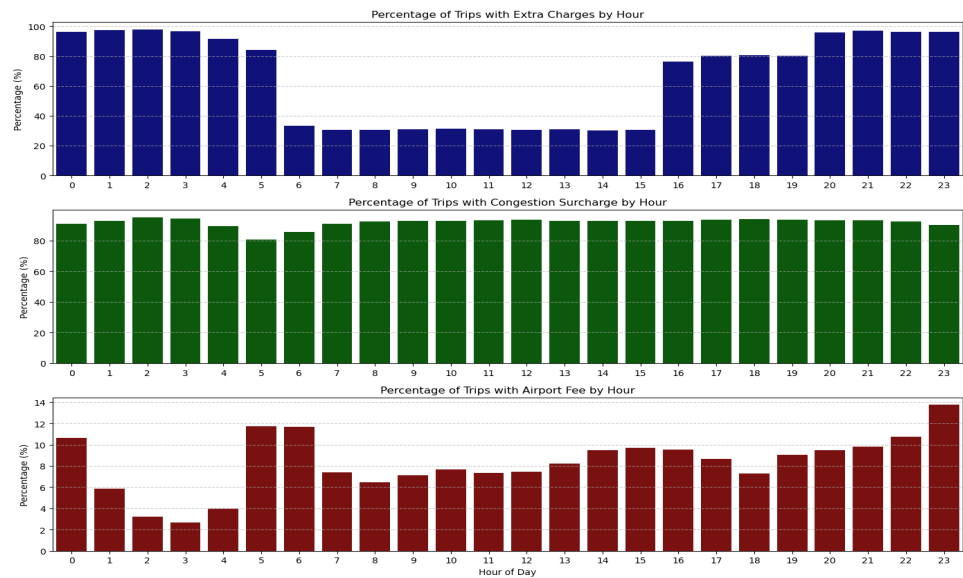


Figure 12.1: Percentage of Trips with Surcharges by Hour

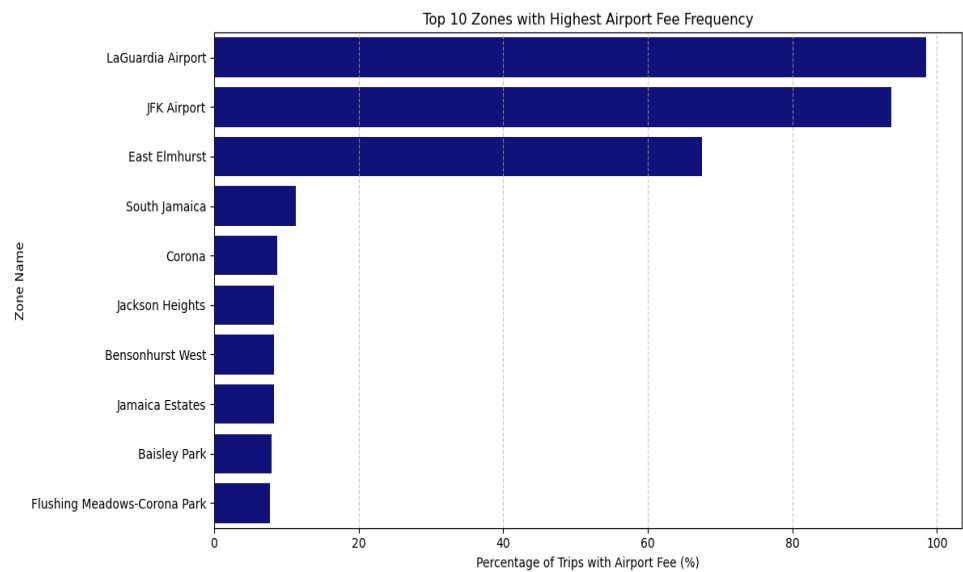


Figure 12.2: Top 10 Zones with Highest Airport Fee Frequency

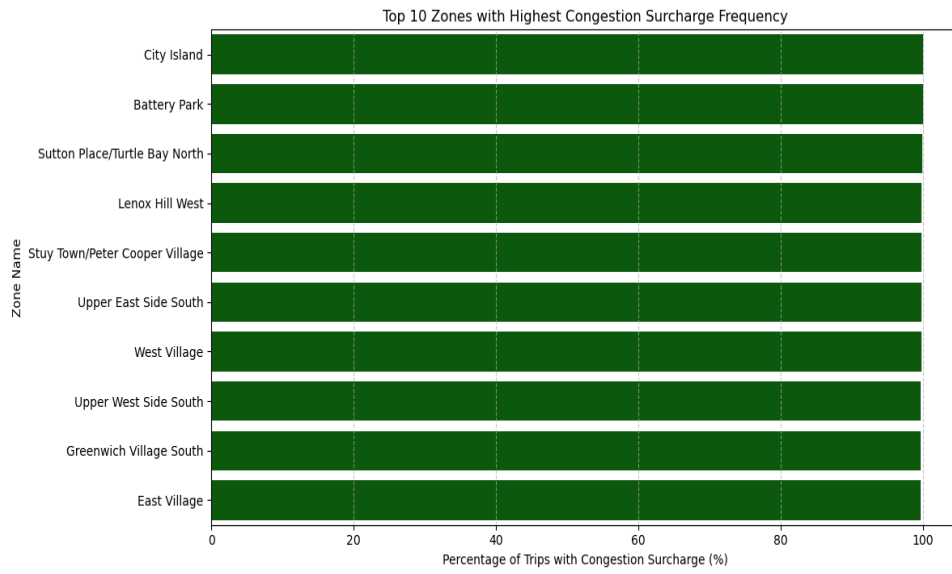


Figure 12.3: Top 10 Zones with Highest Congestion Surcharge Frequency

13. Strategic Recommendations

13.1 Operational Optimization

- Implement dynamic routing algorithms accounting for time-of-day speed variations
- Deploy predictive positioning based on hourly demand patterns shown in visualizations
- Optimize fleet distribution using geographic demand maps
- Adjust operations for weekday vs weekend patterns identified in temporal analysis

13.2 Pricing Strategy

- Implement distance-tiered pricing based on fare-distance correlation analysis
- Apply time-based surge pricing during identified peak hours (5-7 PM)
- Optimize passenger count incentives using economies of scale insights
- Strategic surcharge application based on zone-specific patterns
- Encourage credit card payments to improve tip recording and customer experience

13.3 Geographic Strategy

- Focus fleet positioning on high-demand Manhattan zones identified in maps
- Address pickup/dropoff imbalances in airport and residential areas
- Optimize night service in entertainment districts (East Village, West Village)
- Implement zone-specific pricing based on demand density analysis

14. Implementation Roadmap

Phase 1: Immediate Actions (0-3 months)

- Implement time-based fleet positioning using hourly demand patterns
- Adjust pricing tiers based on distance-fare analysis
- Optimize airport queue management using geographic insights
- Launch targeted promotions during identified low-demand periods

Phase 2: Medium-term Improvements (3-6 months)

- Deploy predictive algorithms using temporal and geographic patterns
- Implement dynamic routing based on speed analysis
- Launch customer loyalty programs targeting high-tip zones
- Optimize surcharge communication and application

Phase 3: Long-term Strategy (6-12 months)

- Full implementation of AI-driven demand prediction
- Advanced zone rebalancing using real-time data
- Comprehensive customer experience platform
- Integration of all insights into unified operational system

15. Conclusion

This comprehensive analysis of NYC Yellow Taxi data, supported by detailed visualizations, reveals significant opportunities for operational optimization and revenue enhancement. The strong correlations identified between distance and fare ($r=0.96$), combined with clear temporal and geographic demand patterns, provide a solid foundation for data-driven decision making. Key insights from the visual analysis include: • Clear peak demand periods requiring strategic fleet positioning • Geographic concentration in Manhattan with specific high-demand zones • Strong seasonal patterns affecting revenue distribution • Vendor performance differences suggesting competitive opportunities • Customer behavior patterns indicating service optimization potential The recommended strategies, when implemented systematically using these data-driven insights, are expected to deliver substantial improvements in both operational efficiency and financial performance while maintaining high levels of customer satisfaction.

Report prepared on August 02, 2025

Based on comprehensive analysis of 363,511 NYC Yellow Taxi trips from 2023

Including 35+ visualizations and statistical analyses