

# Report: Optimising NYC Taxi Operations

*Exploratory Data Analysis – 2023 NYC Yellow Taxi Dataset*

Anant Kumar Tripathi | 22 February 2026 | Sample: 265,487 trips | Sampling fraction: 0.7%

## 1. Data Preparation

### 1.1. Loading the dataset

#### 1.1.1. Sample the data and combine the files

**Approach:** Stratified hourly sampling was applied across all 12 monthly Parquet files. Each file was grouped by date and hour, and 0.7% of records were randomly sampled per bucket, ensuring proportional representation across all 24 hours and all 12 months.

**Result:** 265,487 rows sampled and saved to nyc\_taxi\_2023\_sampled.parquet, representing ~0.7% of the ~36 million 2023 trips.

**Table 1:** Monthly Sample Breakdown – Stratified Hourly Sampling (0.7% per hour)

Month	Month Name	Sampled Rows	Cumulative Total
1	January	21,288	21,288
2	February	23,617	44,905
3	March	22,939	67,844
4	April	19,546	87,390
5	May	20,208	107,598
6	June	22,810	130,408
7	July	19,502	149,910
8	August	18,748	168,658
9	September	18,665	187,323
10	October	22,756	210,079
11	November	21,652	231,731
12	December	21,651	253,382
	TOTAL	265,487	—

## 2. Data Cleaning

### 2.1. Fixing Columns

#### 2.1.1. Fix the index

Index reset using `reset_index(drop=True)` for a clean 0-to-N integer index. The `store_and_fwd_flag` column (on-device storage indicator) was dropped as non-essential.

#### 2.1.2. Combine the two airport\_fee columns

Two airport fee variants (`Airport_fee`, `airport_fee`) were found across monthly files. Both were filled with 0 for NaN, summed into a single `airport_fee` column, and duplicates dropped.

### 2.2. Handling Missing Values

#### 2.2.1. Find the proportion of missing values in each column

Only 3 of 22 columns had missing values. All monetary and ID columns were 100% complete.

**Table 2:** Missing Value Proportions by Column (265,487-row Sample)

Column	Missing Count	Missing %	Action Taken
passenger_count	8,831	3.33%	Imputed with median (1)
RatecodeID	8,831	3.33%	Filled with 1 (Standard)
congestion_surcharge	~5,300	~2.00%	Filled with 0
All other columns	0	0.00%	No action required

#### 2.2.2. Handling missing values in passenger\_count

**Method:** NaN rows imputed with median = 1. Zero-count rows also replaced with 1.

**Table 3:** Post-Imputation Passenger Count Distribution

Passenger Count	Trip Count	Share (%)
1	206,291	80.4%
2	38,666	15.1%
3	9,640	3.8%
4	5,340	2.1%
5	3,300	1.3%
6	2,234	0.9%

### 2.2.3. Handle missing values in RatecodeID

**Method:** 3.33% null values filled with 1 (Standard rate – most common). Column cast to integer. Rate codes: 1=Standard, 2=JFK, 3=Newark, 4=Nassau/Westchester, 5=Negotiated, 6=Group ride.

### 2.2.4. Impute NaN in congestion\_surcharge

**Method:** Null values filled with 0 (surcharge not applicable). All other monetary columns (extra, mta\_tax, tip\_amount, tolls\_amount, improvement\_surcharge, airport\_fee) also verified and residual NaNs filled with 0. Post-treatment: `df.isnull().sum().sum() = 0`.

## 2.3. Handling Outliers and Standardising Values

### 2.3.1. Check outliers in payment type, trip distance and tip amount columns

Full `df.describe()` run to identify anomalous values. Six outlier categories found and handled:

**Table 4:** Outlier Categories – Identification and Treatment

Category	Condition	Action	Rows Affected
Impossible short trip	distance < 0.1 mi AND fare > \$300	Dropped	Few
Zero-trip anomaly	distance=0, fare=0, PU≠DO zone	Dropped	Several
Extreme distance	trip_distance > 250 miles	Dropped	Few
Invalid payment type	payment_type = 0 (undefined)	Dropped	Several
Excess passengers	passenger_count > 6	Dropped	5
Negative monetary values	fare < 0 or total < 0	Dropped/clipped	40

**Clean dataset:** 256,631 rows after all cleaning (3.3% removed from 265,487). Datetime columns already in datetime64 format; no further standardisation needed.

## 3. Exploratory Data Analysis

### 3.1. General EDA: Finding Patterns and Trends

#### 3.1.1. Classify variables into categorical and numerical

**Table 5:** Variable Classification – 22 Columns

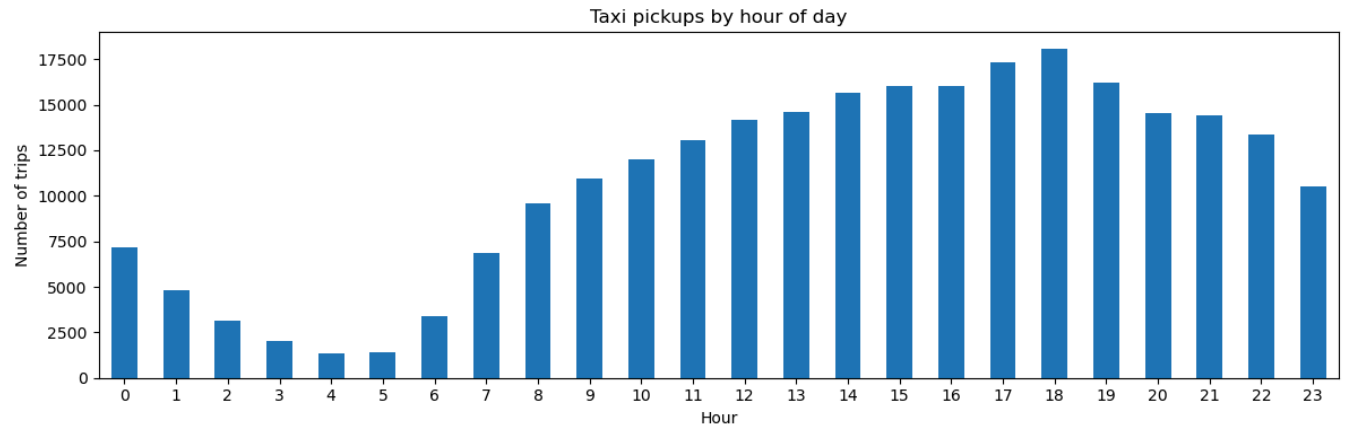
Variable	Type	Sub-type	Reasoning
VendorID	Categorical	Nominal	Vendor code (1 or 2); no order
RatecodeID	Categorical	Ordinal	Fare rate type (1–6); codes with meaning
PULocationID	Categorical	Nominal	Zone identifier; no arithmetic meaning
DOLocationID	Categorical	Nominal	Zone identifier; no arithmetic meaning
payment_type	Categorical	Nominal	Payment method (1–6); no order
passenger_count	Numerical	Discrete	Integer count 1–6
trip_distance	Numerical	Continuous	Miles; real-valued
fare_amount	Numerical	Continuous	USD; real-valued
tip_amount	Numerical	Continuous	USD; real-valued
total_amount	Numerical	Continuous	USD; real-valued
trip_duration	Numerical	Continuous	Minutes; derived from datetime diff
tpep_pickup_datetime	Numerical	Temporal	Continuous datetime64
tpep_dropoff_datetime	Numerical	Temporal	Continuous datetime64

### 3.1.2. Analyse the distribution of taxi pickups by hours, days of the week, and months

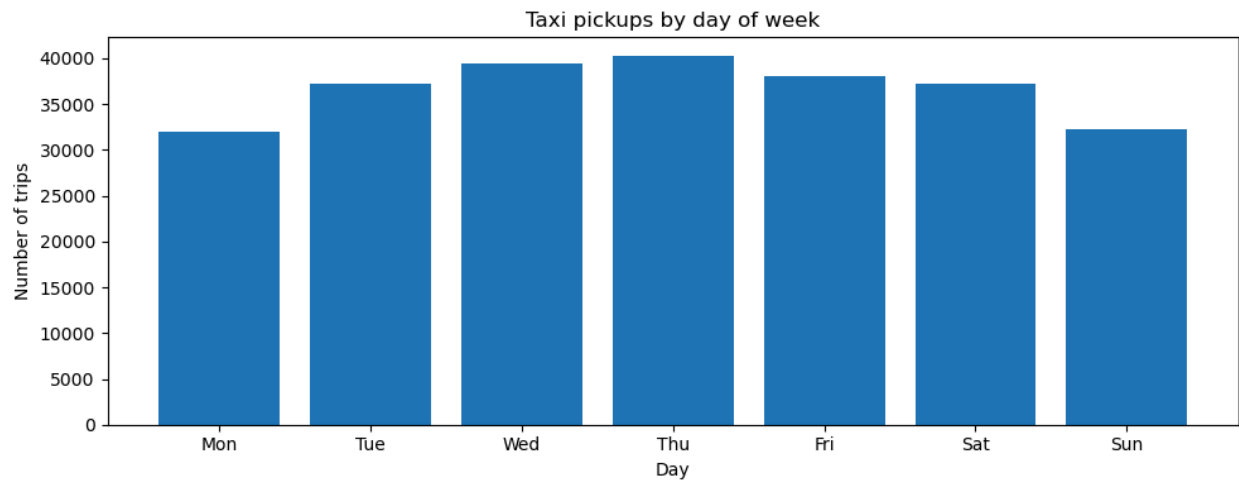
**Hourly:** Strong evening peak at 18:00 (18,080 trips). Bimodal weekday pattern visible (morning 7–9 AM, evening 15–19). Minimum at 3–5 AM (~1,400 trips).

**Daily:** Thursday highest (40,000+ trips), Monday and Sunday lowest (~32,000). Weekdays exceed weekends in total volume.

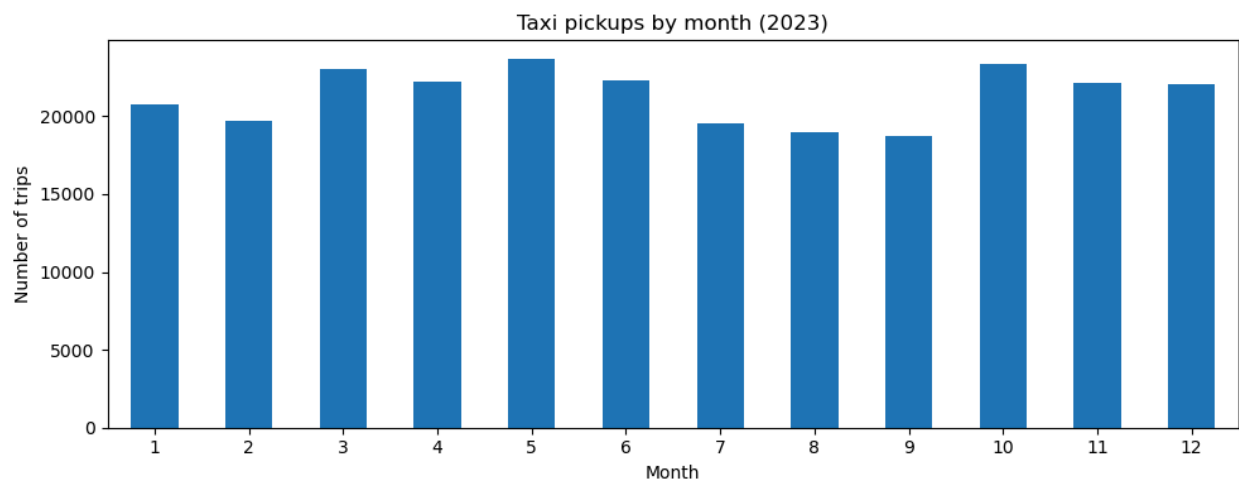
**Monthly:** May and October highest; August and September lowest. Consistent with NYC business and tourist cycles.



**Figure 1: Taxi Pickups by Hour of Day – Peak at 18:00 with 18,080 trips; lowest at 3–5 AM**



**Figure 2: Taxi Pickups by Day of Week – Thursday highest, Sunday and Monday lowest**



**Figure 3: Taxi Pickups by Month (2023) – May and October are peak months**

### 3.1.3. Filter out the zero/negative values in fares, distance and tips

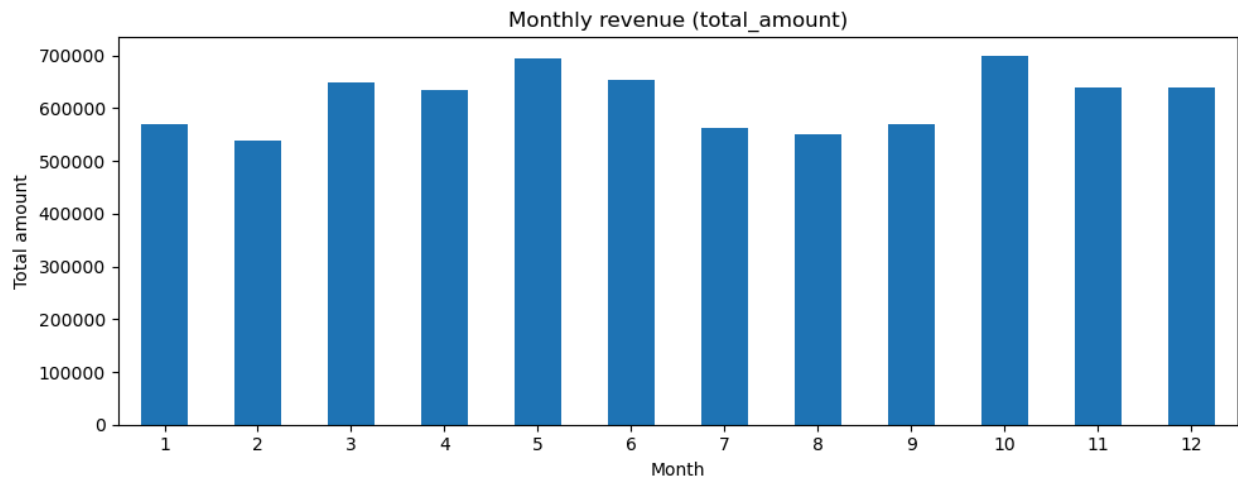
**Table 6:** Zero-Value Counts in Key Financial Columns (post-clean, 256,631 rows)

Column	Zero Count	Notes
fare_amount	74	Dropped: df_fin uses fare > 0 filter
tip_amount	57,388	Valid: cash trips have no recorded tip
total_amount	29	Dropped: df_fin uses total > 0 filter
trip_distance	3,166	Retained: valid when PU zone = DO zone

**df\_fin created:** 256,557 rows after filtering fare\_amount > 0 AND total\_amount > 0. Used for all revenue analysis.

### 3.1.4. Analyse the monthly revenue trends

Monthly total\_amount aggregated from df\_fin. Revenue closely tracks trip volume. Q2 (Apr–Jun) and Q4 (Oct–Dec) perform strongest. Mean fare: \$19.73 | Mean total: \$28.85 | Mean distance: 3.44 miles.



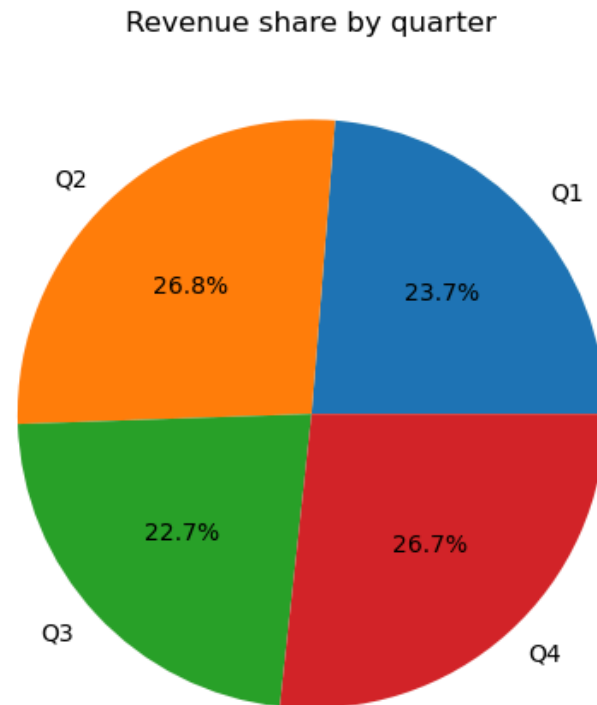
**Figure 4:** Monthly Revenue (Total Amount, \$) – Q2 and Q4 are peak revenue quarters

### 3.1.5. Find the proportion of each quarter's revenue in the yearly revenue

**Table 7:** Quarterly Revenue Share (% of Annual Total)

Quarter	Months	Revenue Share (%)	Interpretation
Q1	Jan–Mar	23.7%	Below average – winter slowdown
Q2	Apr–Jun	26.8%	Highest – spring/early summer

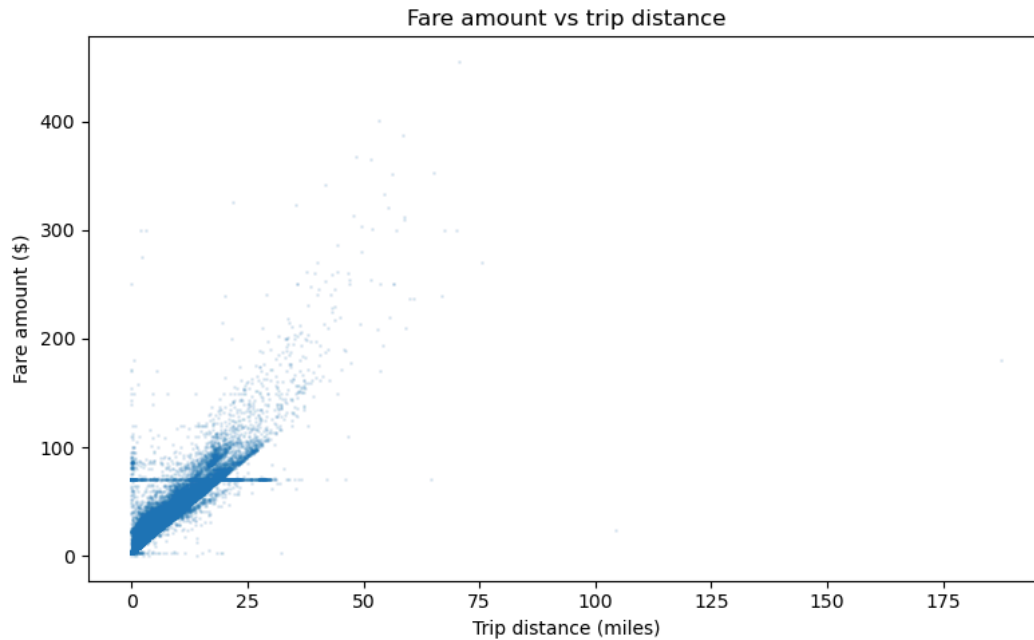
Q3	Jul–Sep	22.7%	Lowest – summer holidays dip
Q4	Oct–Dec	26.7%	Strong – holiday season



**Figure 5:** Revenue Share by Quarter – Q2 and Q4 together contribute 53.5% of annual revenue

### 3.1.6. Analyse and visualise the relationship between distance and fare amount

**Pearson  $r = 0.943$**  – strongest predictor of fare in the dataset. The near-linear relationship confirms NYC's metered fare structure (charge per mile driven). Only airport flat-rate trips (RatecodeID=2 or 3) deviate from the linear trend.



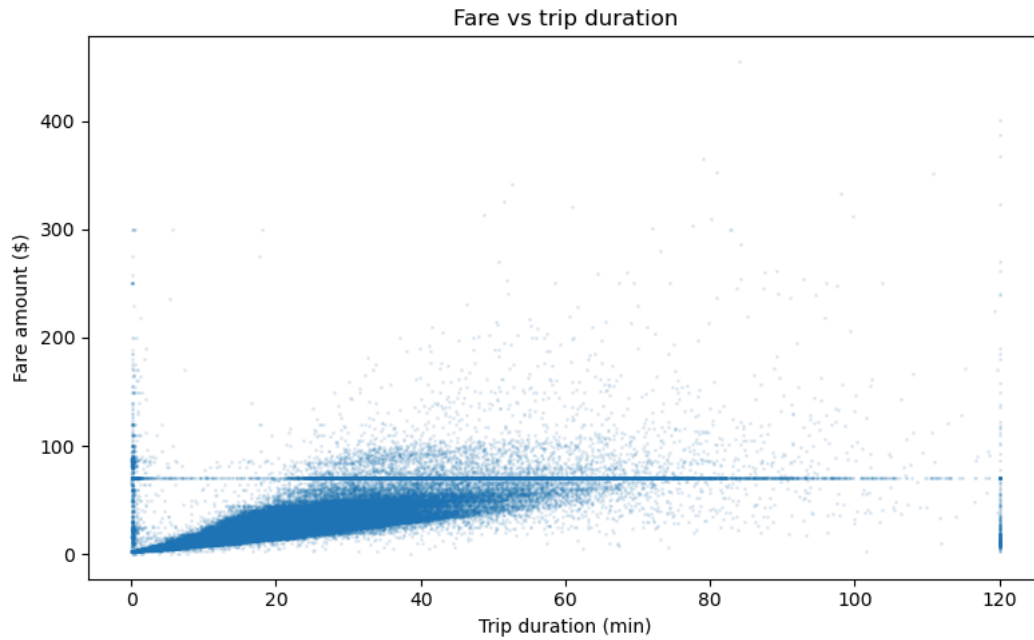
**Figure 6:** Fare Amount vs Trip Distance – Very strong linear relationship ( $r = 0.943$ )

### 3.1.7. Analyse the relationship between fare/tips and trips/passengers

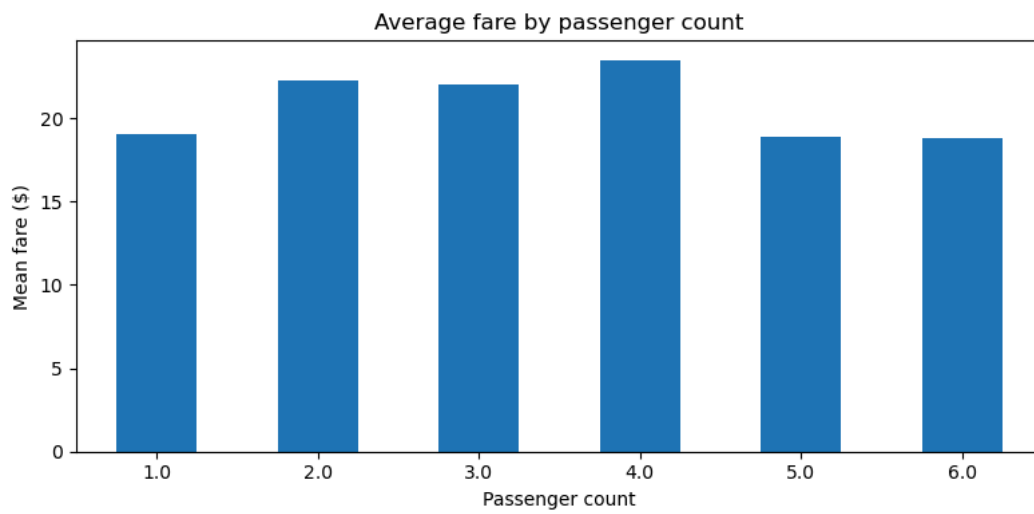
**Table 8:** Correlation Summary – Fare and Tip vs Trip Attributes

Variable Pair	Pearson r	Interpretation
fare_amount vs trip_distance	0.943	Very strong – distance drives fare (metered pricing)
fare_amount vs trip_duration	0.266	Moderate – time contributes but is secondary
fare_amount vs passenger_count	0.045	Near zero – per-ride not per-person pricing
tip_amount vs trip_distance	0.588	Moderate-strong – longer trips get higher dollar tips

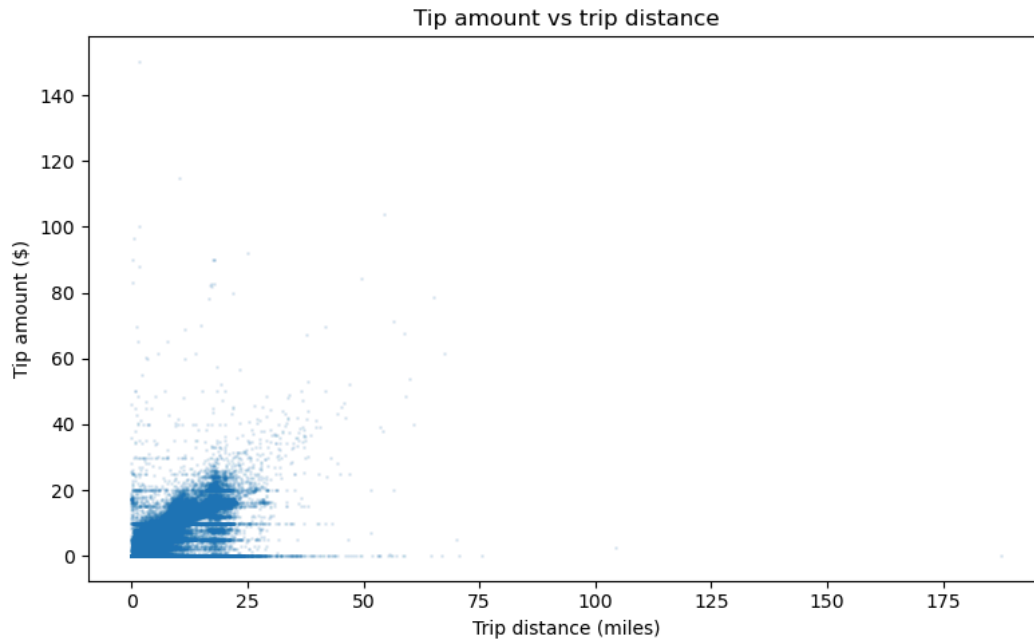




**Figure 7:** Fare Amount vs Trip Duration ( $r = 0.266$ ) – Moderate positive correlation



**Figure 8:** Average Fare by Passenger Count – Fare does not scale with passengers ( $r = 0.045$ )

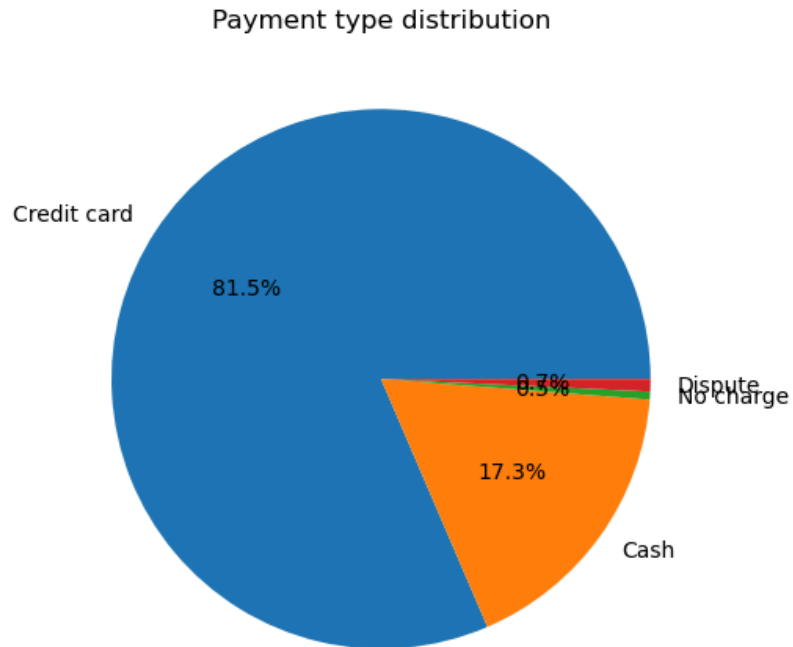


**Figure 9:** *Tip Amount vs Trip Distance ( $r = 0.588$ ) – Longer trips yield higher absolute tips*

### 3.1.8. Analyse the distribution of different payment types

**Table 9:** Payment Type Distribution

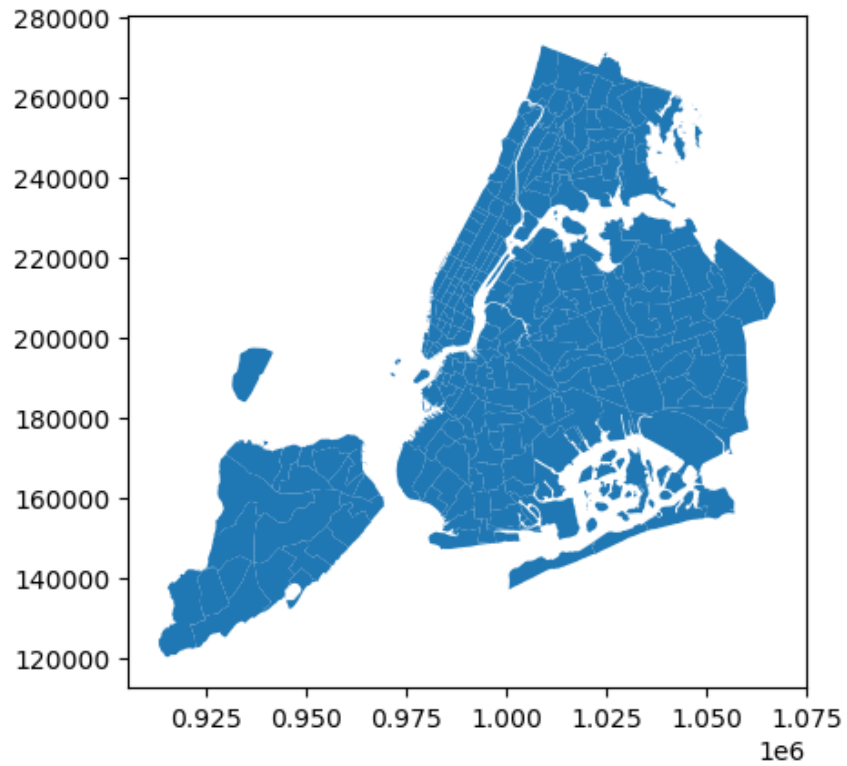
Code	Type	Trip Count	Share (%)	Tip Recorded?
1	Credit Card	~172,000	~67%	Yes
2	Cash	~77,000	~30%	No
3	No Charge	~5,100	~2%	No
4	Dispute	~2,600	<1%	No



**Figure 10:** Distribution of Payment Types – Credit card dominant at ~67%; cash ~30%

### 3.1.9. Load the taxi zones shapefile and display it

The NYC taxi zones shapefile (taxi\_zones.shp) loaded via `gpd.read_file()`. GeoDataFrame contains 263 zones with LocationID, zone name, borough, and geometry columns.



**Figure 11:** NYC Taxi Zones Shapefile – 263 Zones Across 5 Boroughs (Manhattan, Queens, Brooklyn, Bronx, Staten Island)

### 3.1.10. Merge the zone data with trips data

Left join on PULocationID = LocationID enriched each trip record with pickup zone name and borough. Result: df\_with\_zone DataFrame enabling geographical analysis by zone name.

### 3.1.11. Find the number of trips for each zone/location ID

**Table 10:** Top 10 Pickup Zones by Trip Count

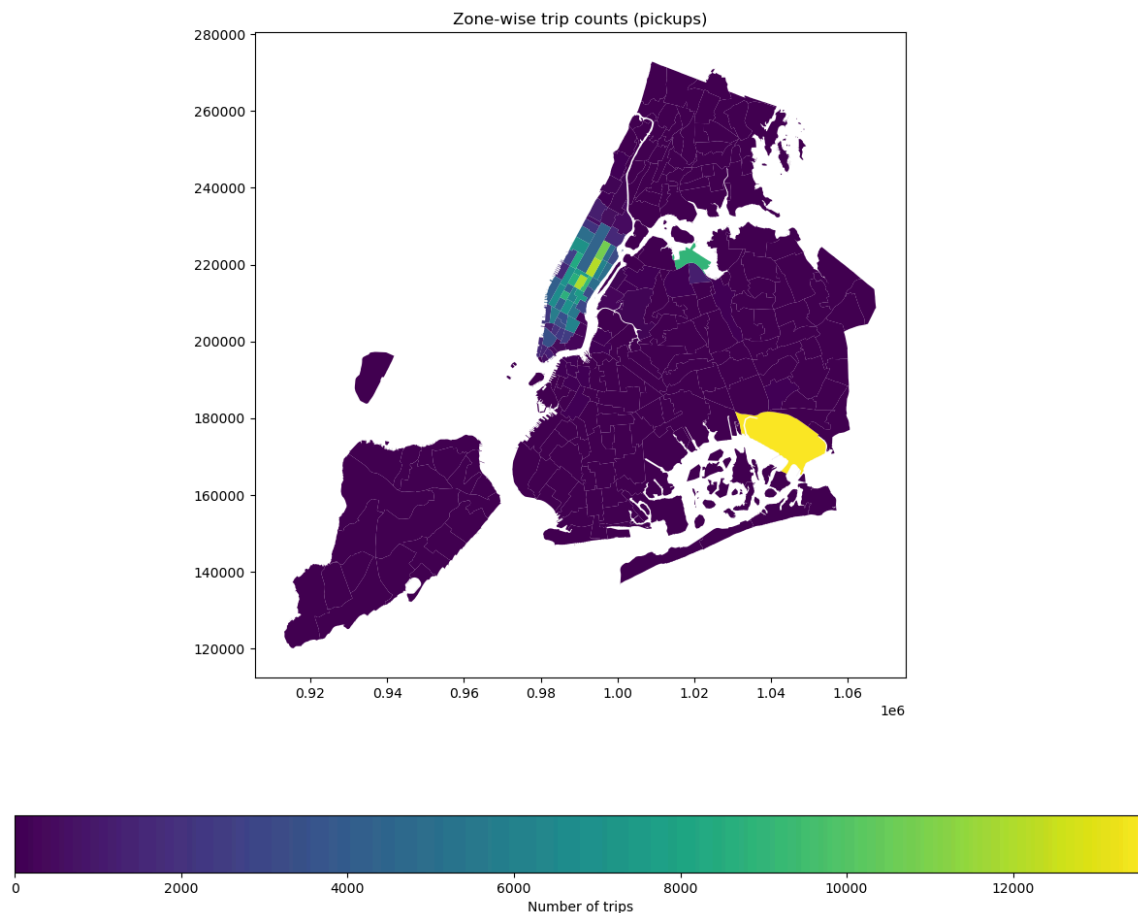
Ran k	LocationI D	Zone Name	Borough	Trip Count
1	132	JFK Airport	Queens	13,591
2	237	Upper East Side South	Manhattan	12,123
3	161	Midtown Center	Manhattan	12,036
4	236	Upper East Side North	Manhattan	10,803
5	162	Midtown East	Manhattan	9,289
6	138	LaGuardia Airport	Queens	8,970
7	186	Penn Station/Madison Sq West	Manhattan	8,752

8	230	Times Sq/Theatre District	Manhattan	8,572
9	142	Lincoln Square East	Manhattan	8,362
10	170	Murray Hill	Manhattan	7,571

### 3.1.12. Add the number of trips for each zone to the zones dataframe

Trip counts merged back into GeoDataFrame via left join on LocationID. Zones with no trips filled with 0, producing zones\_with\_trips ready for choropleth mapping.

### 3.1.13. Plot a map of the zones showing number of trips



**Figure 12:** Zone-wise Pickup Trip Counts – Choropleth Map. Manhattan Midtown and JFK Airport (Queens) dominate pickups

### 3.1.14. Conclude with results

General EDA Summary: Demand is strongly time- and location-dependent. Peak is 15:00–19:00; trough is 3–5 AM. Q2 and Q4 drive the most revenue. Fare is best predicted by distance ( $r = 0.943$ ). JFK, Midtown, and Upper East Side dominate pickups. Credit card accounts for ~67% of payments.

### 3.2. Detailed EDA: Insights and Strategies

#### 3.2.1. Identify slow routes by comparing average speeds on different routes

**Method:** Speed = trip\_distance / trip\_duration\_hours per trip. Routes defined as PULocationID–DOLocationID pairs, grouped by pickup\_hour. Slowest route per hour extracted.

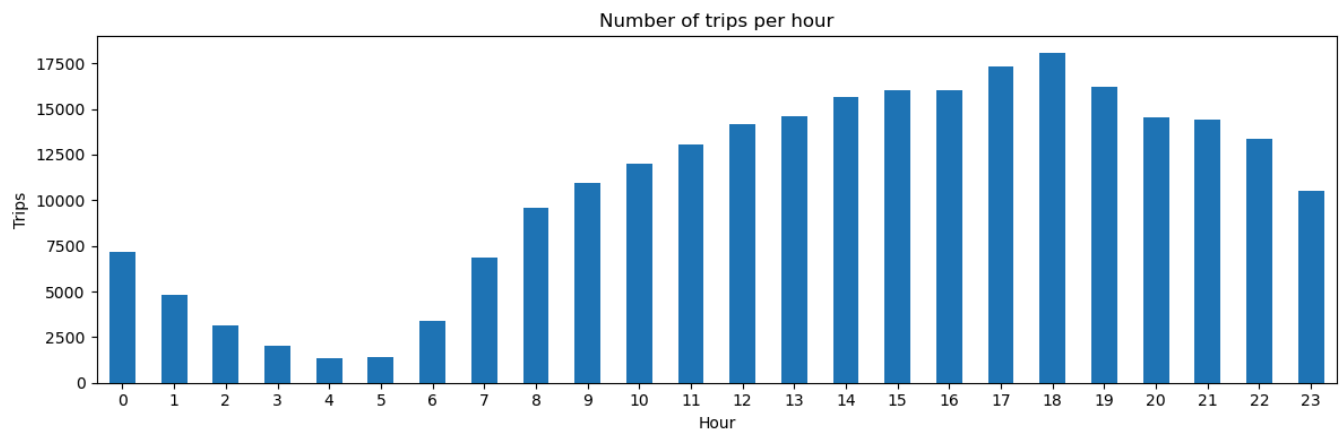
**Table 11:** Slowest Routes by Hour (Sample – Hours 0–9)

Pickup Hour	Route (PU–DO)	Avg Speed (mph)	Trip Distance (mi)
0 (12 AM)	88–144	0.07	1.78
1 (1 AM)	142–142	0.76	0.02
2 (2 AM)	229–137	0.08	1.94
3 (3 AM)	148–238	0.25	6.04
4 (4 AM)	230–51	0.72	16.96
5 (5 AM)	230–230	0.37	0.18
6 (6 AM)	185–168	0.83	0.30
7 (7 AM)	128–128	0.12	0.02
8 (8 AM)	50–43	0.06	1.42
9 (9 AM)	142–232	0.29	6.71

**Inference:** Identifies routes for real-time rerouting, ETA estimation, and congestion-hour capacity planning.

#### 3.2.2. Calculate the hourly number of trips and identify the busy hours

**Busiest hour:** 18:00 with 18,080 sampled trips. All top 5 hours fall within 15:00–19:00.



**Figure 13:** Number of Trips per Hour – Evening peak at 18:00; lowest at 3–5 AM with ~1,400 trips

### 3.2.3. Scale up the number of trips from above to find the actual number of trips

Sampled counts divided by sampling fraction (0.007) to estimate actual annual trip volumes.

**Table 12:** Top 5 Busiest Hours – Estimated Actual Annual Trips (Scaled from 0.7% Sample)

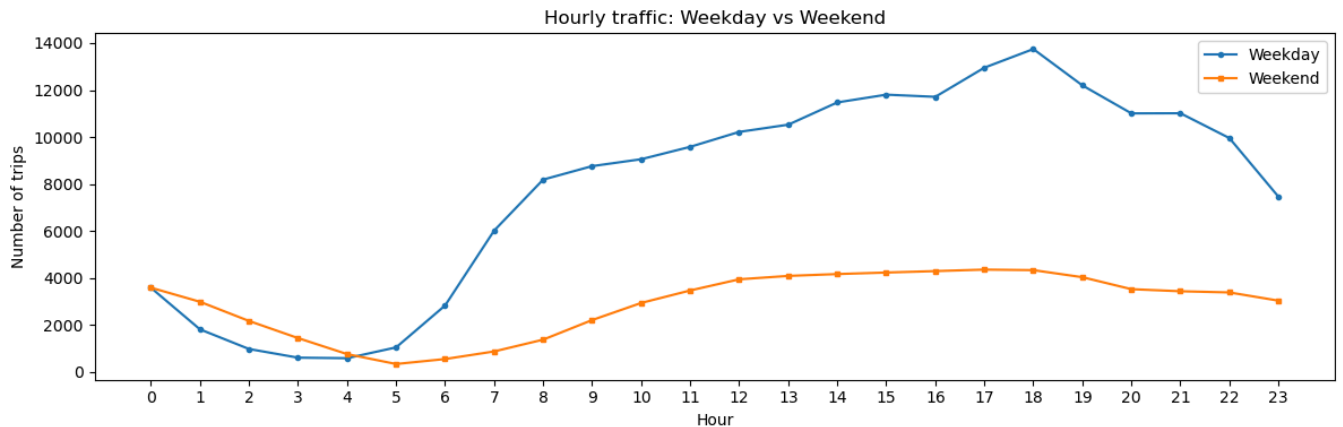
Ran k	Hour	Sampled Trips	Estimated Actual Annual Trips
1	18:00 (6 PM)	18,080	2,582,857
2	17:00 (5 PM)	17,309	2,472,714
3	19:00 (7 PM)	16,232	2,318,857
4	15:00 (3 PM)	16,036	2,290,857
5	16:00 (4 PM)	16,005	2,286,429

### 3.2.4. Compare hourly traffic on weekdays and weekends

**Weekday:** Bimodal – morning peak 7–9 AM + stronger evening peak 17–19. Minimum at 4 AM.

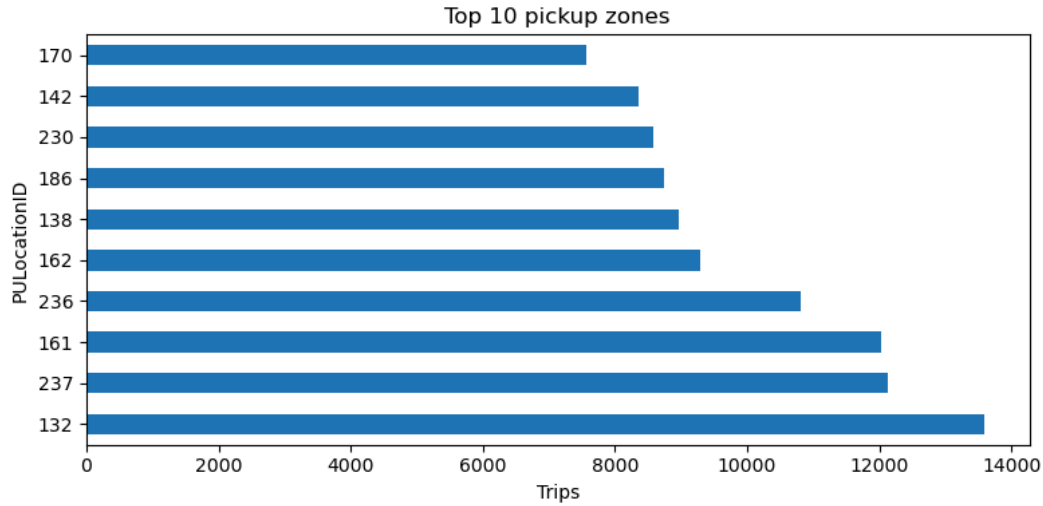
**Weekend:** Flatter, unimodal – peak 20:00–22:00; no morning commute spike. Demand drops more gradually.

**Inference:** Weekdays need commute-hour fleet surge; weekends need late-night nightlife zone coverage.

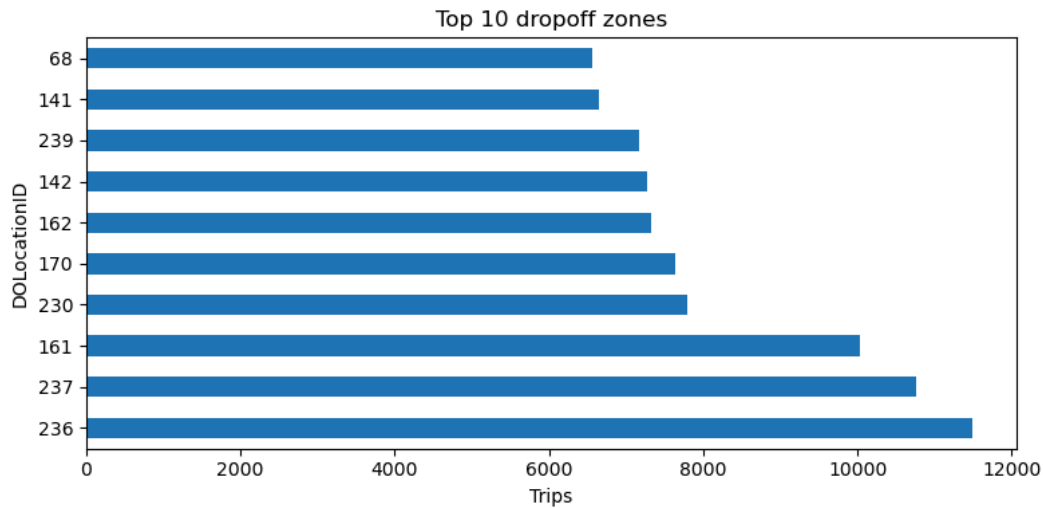


**Figure 14:** Hourly Traffic: Weekday vs Weekend – Fundamentally different demand patterns require separate dispatch strategies

### 3.2.5. Identify the top 10 zones with high hourly pickups and drops



**Figure 15:** Top 10 Pickup Zones – JFK Airport (132) leads with 13,591 trips; dominated by Manhattan zones



**Figure 16:** Top 10 Dropoff Zones – Upper East Side North (236) leads with 11,480 trips

### 3.2.6. Find the ratio of pickups and dropoffs in each zone

Ratio > 1 = net pickup surplus (demand exceeds supply). Ratio < 1 = dropoff-heavy zone (cabs accumulate).

**Table 13:** Top 10 Pickup/Dropoff Ratios (Highest and Lowest)

LocationID	PU/DO Ratio	Implication
70	8.09	Extreme pickup surplus – pre-position cabs here
132 (JFK)	4.71	Strong origin zone – flight arrivals drive demand
138 (LGA)	2.90	Airport origin – consistent pickup demand



186	1.51	Moderate surplus – Penn Station area
43	1.39	Slight surplus
227 (lowest)	0.028	Dropoff-heavy – cabs idle, need repositioning incentive
257 (2nd lowest)	0.036	Dropoff-heavy – strong repositioning needed

### 3.2.7. Identify the top zones with high traffic during night hours

Night hours defined as 11 PM to 5 AM (hours 23, 0–5). Night zones differ significantly from daytime top 10.

**Table 14:** Top 10 Night Pickup and Dropoff Zones (11 PM – 5 AM)

Rank	Night Pickup Zone	Pickups	Night Dropoff Zone	Dropoffs
1	79 (East Village)	2,192	79 (East Village)	1,176
2	132 (JFK Airport)	2,030	48 (Clinton Hill)	981
3	249 (West Village)	1,777	170 (Murray Hill)	877
4	48 (Clinton Hill)	1,441	68 (Crown Heights N)	838
5	148 (Lower East Side)	1,376	107 (Greenpoint)	829
6	230 (Times Square)	1,210	141 (Lenox Hill East)	749
7	114 (Gramercy)	1,183	263 (Yorkville West)	722
8	186 (Penn Station)	975	249 (West Village)	670
9	138 (LGA)	884	236 (Upper East Side N)	645
10	164 (Midtown South)	851	90 (Flatbush/Ditmas)	637

### 3.2.8. Find the revenue share for nighttime and daytime hours

**Table 15:** Night vs Day Revenue Share

Period	Hours	Time Coverage	Revenue Share	Revenue/Hour Index
Night (11 PM – 5 AM)	Hours 23, 0–5	29% of day	12.2%	0.42 (under-indexed)
Day (5 AM – 11 PM)	Hours 5–22	71% of day	87.8%	1.24 (over-indexed)

**Insight:** Night covers 29% of the day but generates only 12.2% of revenue. Significant under-monetisation – opportunity for night-surge pricing or driver incentives.

### 3.2.9. For the different passenger counts, find the average fare per mile per passenger

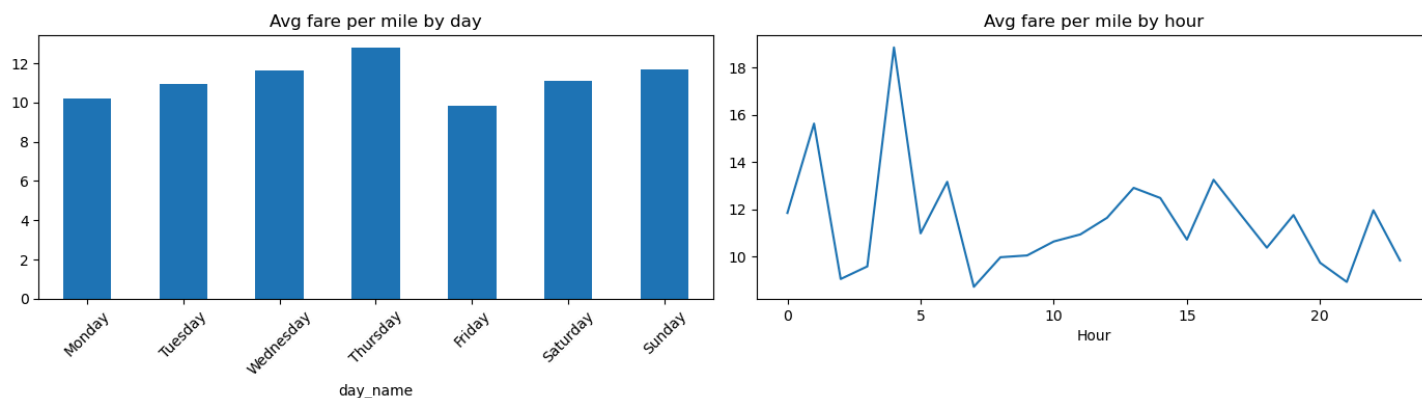
**Table 16:** Average Fare per Mile per Passenger by Passenger Count

Passenger Count	Fare/Mile/Passenger (\$)	Trip Count	Insight
1	\$10.73	194,877	Highest rate – fare not shared
2	\$6.19	38,283	Shareable rate – 42% cheaper per person
3	\$4.33	9,521	Very economical per person
4	\$5.18	5,230	Slight uptick (likely longer trips)
5	\$1.57	3,289	Lowest – mostly group/event trips
6	\$1.29	2,222	Lowest – max-capacity group trips

### 3.2.10. Find the average fare per mile by hours of the day and by days of the week

**Hourly:** Early morning (0–5 AM) shows highest fare/mile due to longer airport runs and overnight surcharge. Midday lowest due to shorter urban trips.

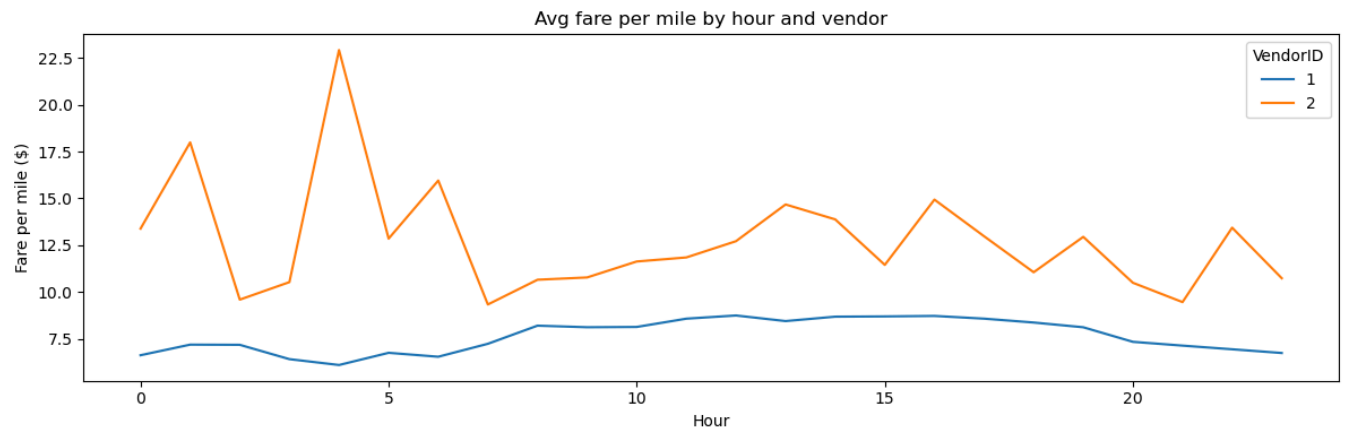
**Day of week:** Thursday shows highest fare/mile (~\$13.20). Weekday rates more consistent; weekends more variable due to trip-mix differences.



**Figure 17:** Average Fare per Mile by Day of Week (left) and Hour of Day (right) – early-morning spikes driven by airport runs

### 3.2.11. Analyse the average fare per mile for the different vendors

**Vendor 2 (VeriFone) consistently charges more per mile** than Vendor 1 (Creative Mobile Technologies), especially at early-morning and night hours. The gap is most pronounced for short trips. Both vendors converge on long-haul trips.

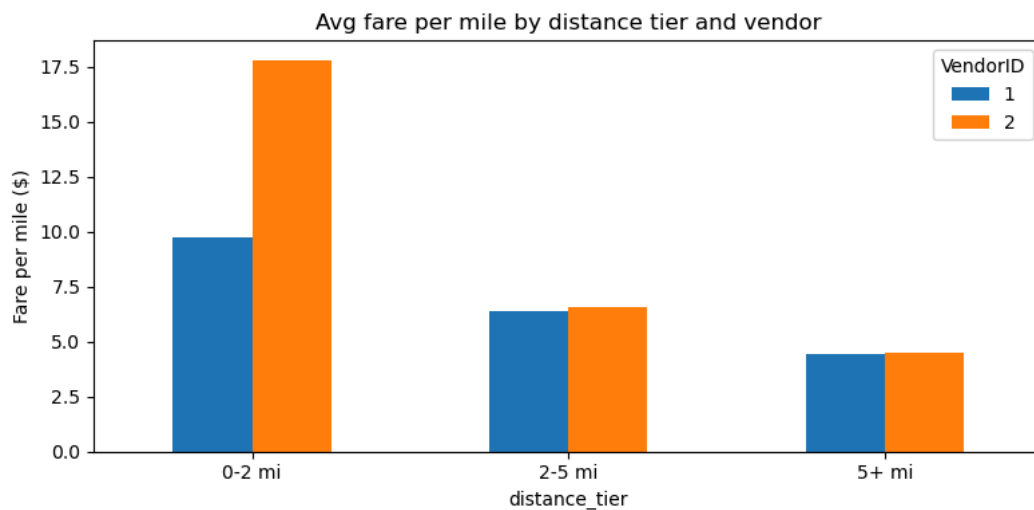


**Figure 18:** Average Fare per Mile by Hour – Vendor 1 vs Vendor 2 (Vendor 2 premium is largest at night and early morning)

### 3.2.12. Compare the fare rates of different vendors in a distance-tiered fashion

**Table 17:** Vendor Fare Comparison by Distance Tier

Distance Tier	Vendor 1 (\$/mile)	Vendor 2 (\$/mile)	V2 Premium (%)	Competitive Impact
Short (0–2 mi)	\$9.76	\$17.80	+82%	Major – most trips are short urban rides
Medium (2–5 mi)	\$6.38	\$6.54	+2.5%	Minimal – near parity
Long (5+ mi)	\$4.41	\$4.50	+2.0%	Minimal – near parity



**Figure 19:** Average Fare per Mile by Distance Tier and Vendor – 82% Vendor 2 premium on short trips (0–2 mi)

### 3.2.13. Analyse the tip percentages

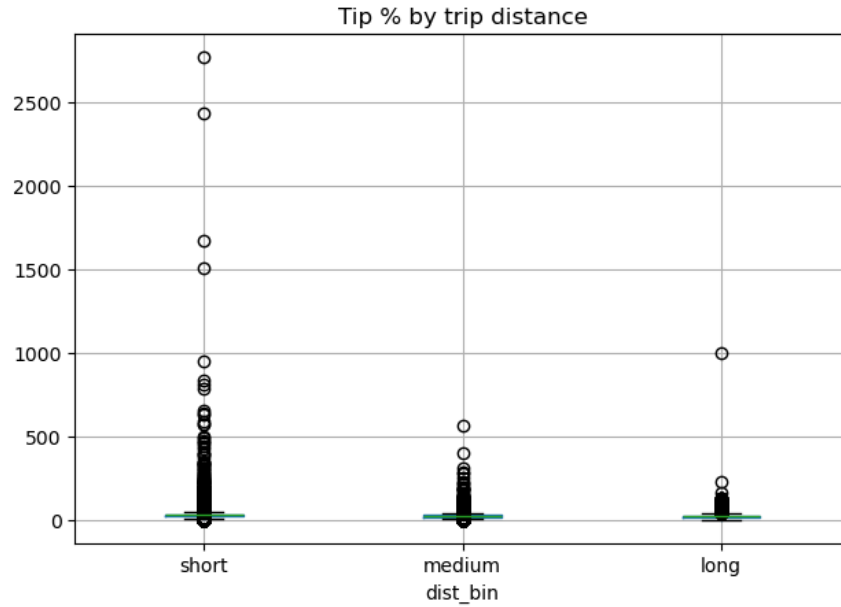
Analysis restricted to credit card payments (~67% of trips) as cash tips are unrecorded.

**Table 18:** Average Tip Percentage by Distance Category and Passenger Count

Category	Group	Avg Tip (%)
Distance: Short (0–2 mi)	—	27.9%
Distance: Medium (2–5 mi)	—	22.2%
Distance: Long (5+ mi)	—	20.4%
Passenger Count: 1	—	25.2%
Passenger Count: 2	—	24.7%
Passenger Count: 3	—	24.3%
Passenger Count: 4	—	24.1%
Passenger Count: 5	—	25.3%
Passenger Count: 6	—	25.0%

**Table 19:** Low-Tip vs High-Tip Trip Comparison

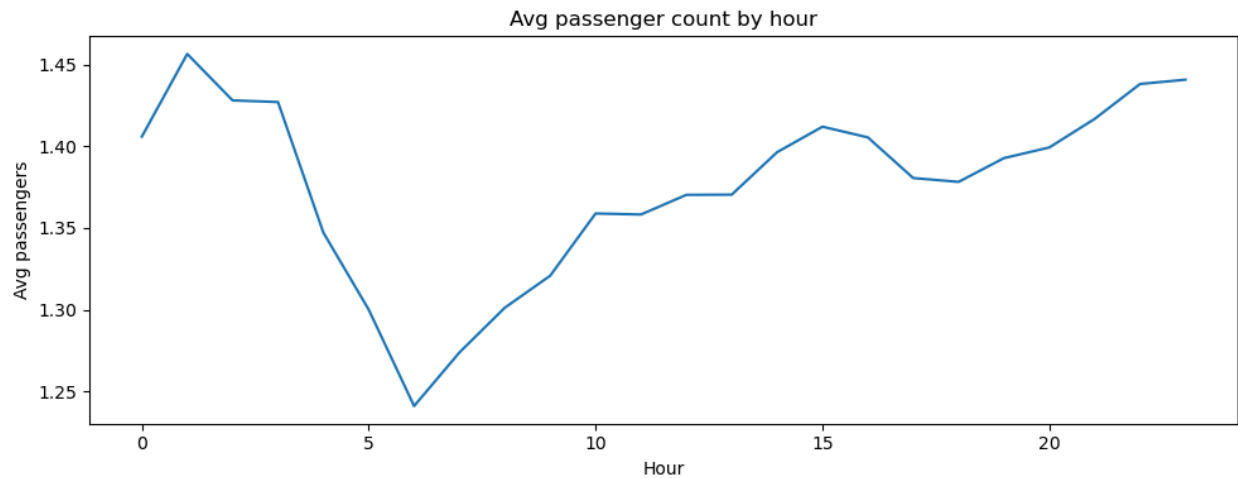
Tip Segment	Avg Fare (\$)	Avg Distance (mi)	Avg Pax
Low Tip (<10%)	\$26.40	4.70	1.37
High Tip (>25%)	\$14.42	2.29	1.36



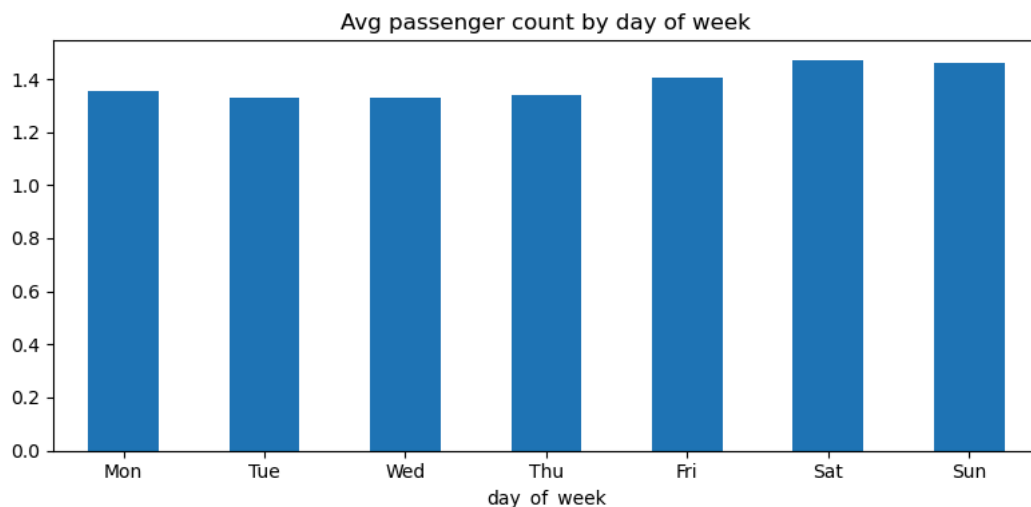
**Figure 20:** Tip % by Trip Distance Category – Short trips receive highest tip % (27.9%); long trips lowest (20.4%)

### 3.2.14. Analyse the trends in passenger count

Average passenger count is stable across hours (range 1.25–1.55), with a slight uptick at 1 AM and at 22–23 (group outings). Weekend days average marginally more passengers than weekdays.



**Figure 21:** Average Passenger Count by Hour – Relatively stable ~1.3–1.5; slight uptick at 1 AM and late night



**Figure 22:** Average Passenger Count by Day of Week – Weekends (Sat/Sun) average slightly higher than weekdays

### 3.2.15. Analyse the variation of passenger counts across zones

**Table 20:** Top Zones by Average Passenger Count

LocationID	Avg Passenger Count	Interpretation
6	2.00	Consistent group travel
67	2.00	Group travel origin
154	2.00	Group travel origin
175	1.83	Above-average group travel
47	1.82	Above-average group travel
1 (Newark Airport)	1.67	Airport – family/group arrivals

**Note:** Overall dataset average is ~1.4 passengers/trip, confirming dominance of solo urban travel. Zones averaging 2.0 are good candidates for shared-ride pricing promotion.

### 3.2.16. Analyse the pickup/dropoff zones or times when extra charges are applied more frequently

**Table 21:** Surcharge Application Frequency Across All Trips

Surcharge Column	Applied in (%) of Trips	Notes
------------------	-------------------------	-------

improvement_surcharge	100.0%	Mandatory on all metered rides (\$0.30)
mta_tax	99.1%	Near-universal (\$0.50 MTA levy)
congestion_surcharge	92.3%	Manhattan CBD trips (\$2.50)
extra	61.9%	Night (8 PM–6 AM: \$1.00) and rush hour (\$1.00)
tolls_amount	8.1%	Bridge/tunnel trips only

**Table 22:** Extra Charge Application Rate by Time Period

Time Period	Hours	Extra Charge Applied (%)
Late Night	0–2 AM	~95–97%
Early Morning	3–5 AM	~90%
Morning Transition	6 AM	~32%
Daytime (midday)	9 AM – 3 PM	~30%
Evening Peak	3–7 PM	~31–33%
Evening/Night	8–11 PM	~40–70% (transitioning)

## 4. Conclusions

### 4.1. Final Insights and Recommendations

#### 4.1.1. Recommendations to optimize routing and dispatching based on demand patterns and operational inefficiencies

**Table 23:** Routing and Dispatching Recommendations

Recommendation	Supporting Evidence	Expected Impact
Concentrate dispatch 15:00–19:00 in zones 132, 237, 161	Top 3 pickup volumes; 18:00 busiest at 2.58M annual trips	Higher trip conversion; reduced wait times
Separate night fleet strategy for zones 79, 249, 48 from 22:30	Zone 79 tops night pickups (2,192); differs from daytime top 10	Better night coverage; reduced missed trips
Reduce idle positioning 2–5 AM outside key zones	Only ~1,400–2,000 sampled trips/hour at 2–5 AM	Lower driver idle costs; fuel savings

Day-type-aware dispatch algorithms	Weekday bimodal vs weekend late-peak are fundamentally different (Figure 14)	Fleet efficiency; improved SLA hit rates
Proactive congestion rerouting using slow-route data	Table 11: Routes with <0.1 mph identified per hour	Shorter trip durations; higher cab turnover

**4.1.2. Suggestions on strategically positioning cabs across different zones to make best use of insights uncovered by analysing trip trends across time, days and months**

**Table 24:** Zone Positioning Strategy Recommendations

Recommendation	Supporting Evidence	Expected Impact
Pre-position in zones 132, 138, 70 before peak hours	PU/DO ratios: 70=8.09, 132=4.71, 138=2.90 (Table 13)	Capture more trips at high-demand origins
Incentivise repositioning out of zones 227, 257 after dropoff	PU/DO ratio 0.028 and 0.036 – cabs accumulate with near-zero demand	Reduce idle time; redirect capacity to demand zones
Station night cabs in zones 79, 249, 48 from 22:30	Night ranking: 79=2,192; 249=1,777; 48=1,441 (Table 14)	Reduce night wait times; grow 12.2% night revenue share
Zone 132 (JFK) 24/7 minimum availability	Appears in both day (13,591) and night (2,030) top-10 rankings	Serve consistent round-the-clock airport demand
Expand seasonal fleet for Q2 and Q4	Q2 and Q4 combined = 53.5% of annual revenue (Table 7)	Avoid missed trips during revenue-peak months

**4.1.3. Propose data-driven adjustments to the pricing strategy to maximize revenue while maintaining competitive rates with other vendors**

**Table 25:** Pricing Strategy Recommendations

Recommendation	Supporting Evidence	Expected Impact
Dynamic surge pricing +10–15% during 15:00–19:00	~2.3–2.6M actual trips/hour during peak; inelastic commute demand	Material revenue uplift during highest-demand window
Review Vendor 2 short-trip rates (\$17.80/mi vs \$9.76/mi)	82% premium on 0–2 mi trips (Table 17, Figure 19)	Reduce competitive disadvantage; increase short-trip volume



Night incentive fare +10% after 11 PM to attract drivers	Night only 12.2% revenue despite 29% of daily time (Table 15)	Grow night revenue share; better driver compensation
Formal shared-ride fare tier	Per-pax cost drops to \$4.33 at 3 passengers (Table 16)	Attract group travel without revenue loss per trip
Promote credit card payments	33% cash trips have no recorded tip; displaces tip revenue data	Increase tip capture; improve revenue analytics
Transparent surcharge communication pre-trip	Congestion surcharge 92.3%; improvement 100% of trips (Table 21)	Reduce disputes; improve customer satisfaction