Find and show the hourly trends in taxi pickups.
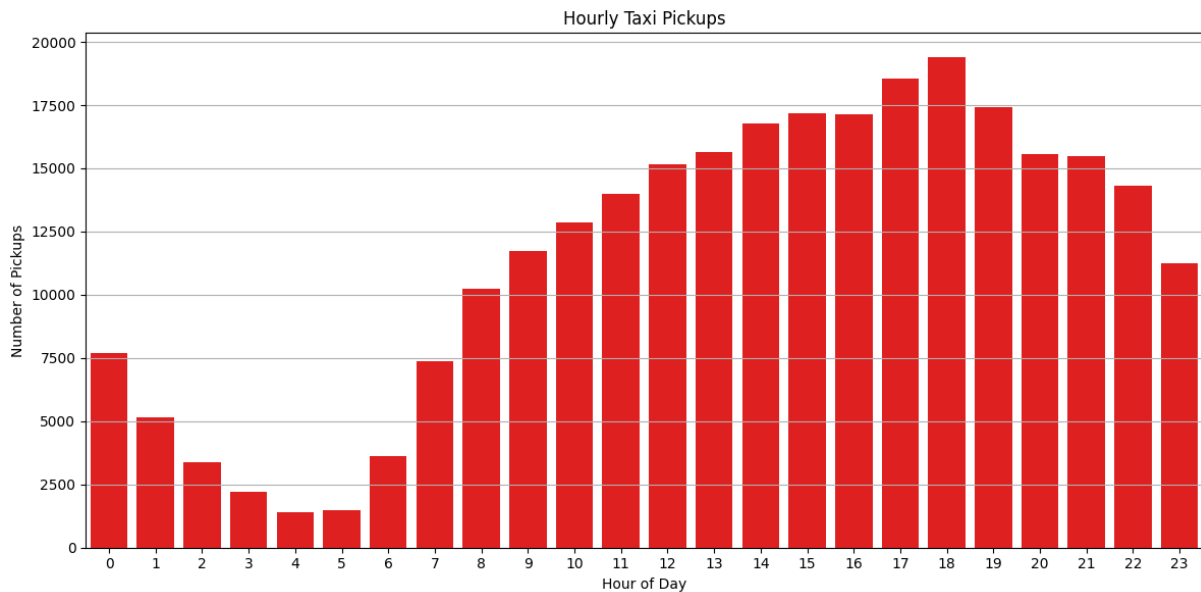

Afternoon to Early Evening (2 PM – 7 PM)
This is the peak demand period, with the highest pickup counts. Hour 18 (6 PM) has the highest activity, followed by 5 PM and 7 PM. This aligns with evening commute, shopping, and dining activity.

Late Morning to Early Afternoon (10 AM – 2 PM) Moderate pickup volume, potentially linked to business and personal travel.

Early Morning (12 AM – 5 AM) Significantly lower activity, likely due to fewer passengers on the road. These hours may represent airport runs, late nightlife, or shift changes.

Morning Commute (7 AM – 9 AM) There's a noticeable pickup in activity, showing the morning rush. However, it's still lower than the evening peak, suggesting more evening ride-hailing in NYC.

Find and show the daily trends in taxi pickups (days of the week)

Taxi Pickups by Day of the Week
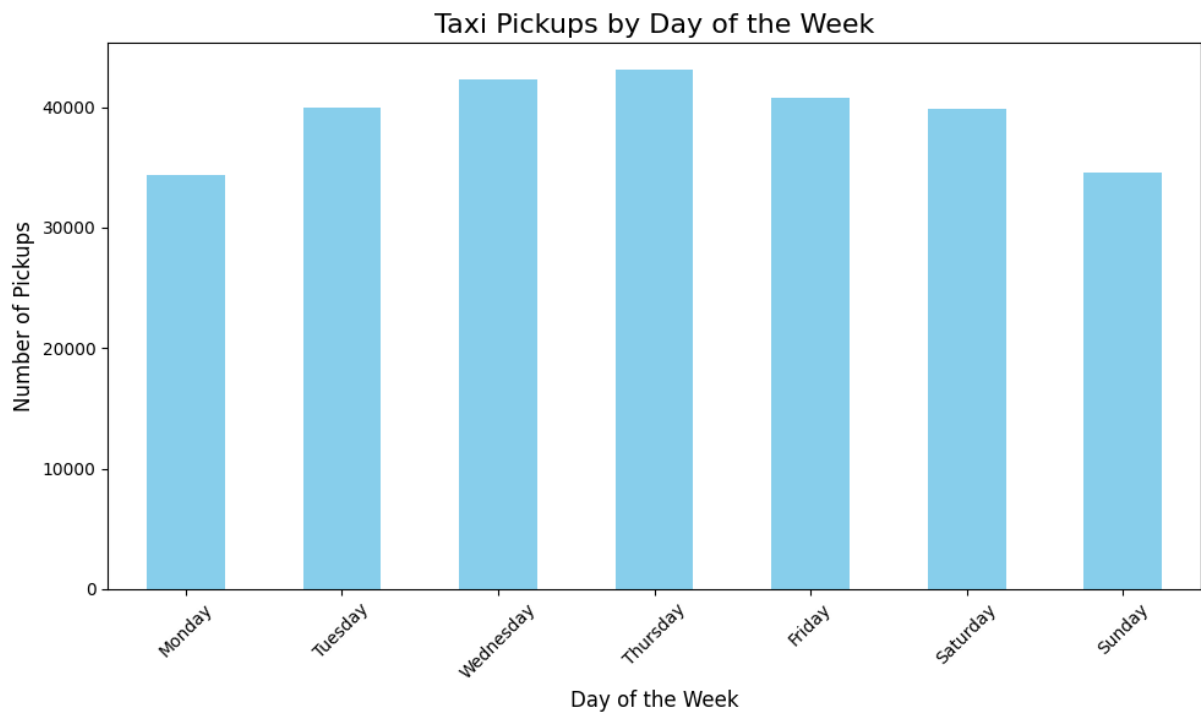
arrow_upward

arrow_downward

link

comment

edit

delete

more_vert

Thursday is the busiest day. A lot of people take taxis on this day. Maybe they go out or have late work events.

Wednesday is also very busy. It's in the middle of the week, so many people are moving around the city.

Tuesday, Friday, and Saturday are also busy, but a little less than Wednesday and Thursday.

Monday is the slowest day. Maybe people stay home more or take fewer trips after the weekend.

Sunday is also a bit quiet, probably because it's a rest day and not many people are going to work or events.

Show the monthly trends in pickups

Taxi Pickups by Month
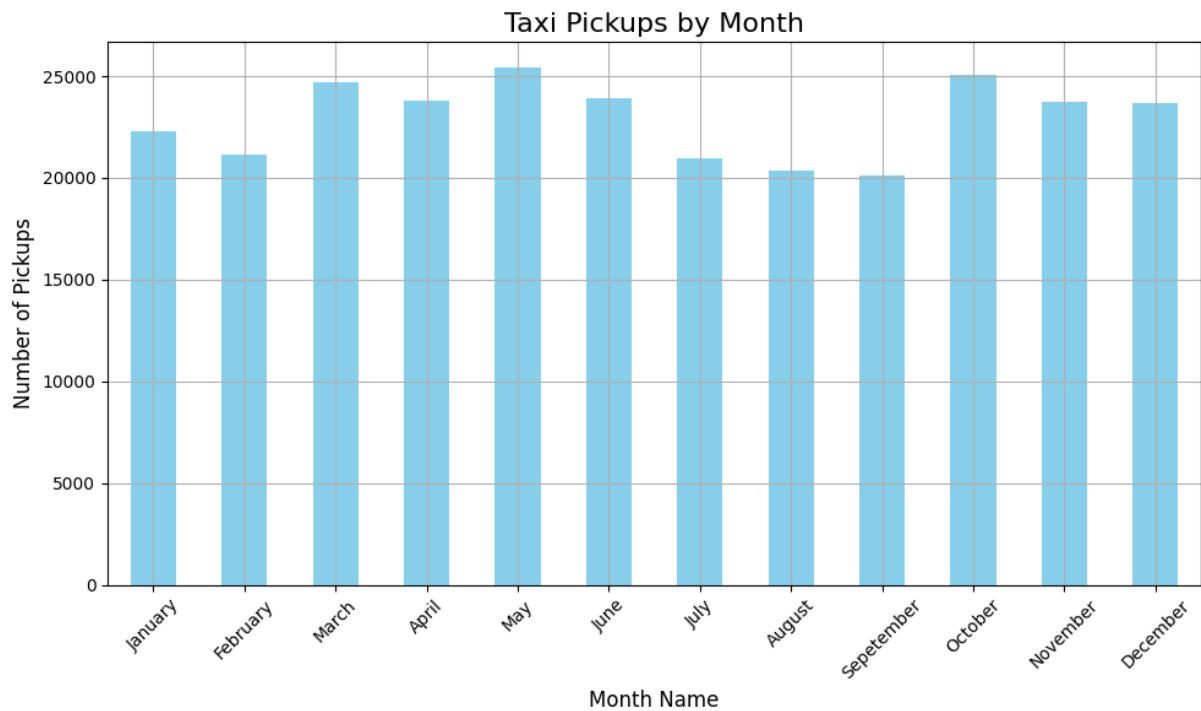
May and October had the most taxi rides. These months seem very busy, maybe because of nice weather or people traveling.

March and April also had a lot of pickups. This could be because the weather is getting better after winter.

Summer months (July, August, September) had fewer rides. Maybe some people are on vacation or using other ways to travel.

January and February had fewer rides too — could be due to cold weather or people staying home more.

December had a good number of rides — maybe due to holidays and people going shopping or visiting family.

Group data by month and analyse monthly revenue

Monthly Revenue Trend (total_amount)
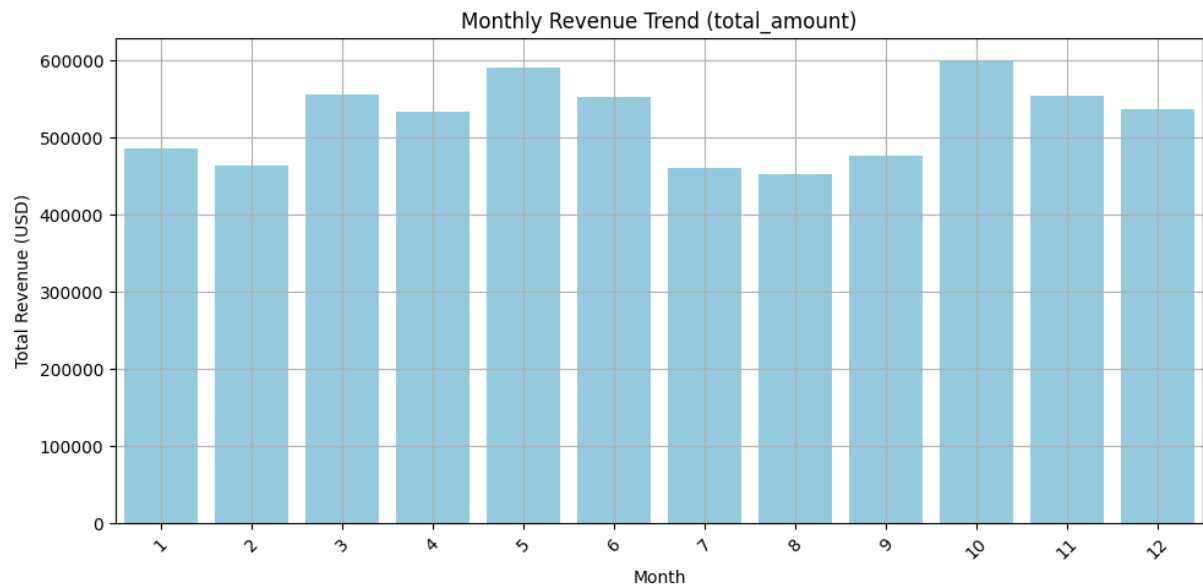
October had the highest revenue — about 598,646. That means a lot of people took taxis or took longer/more expensive rides.

May was also very high, followed by March, November, and June.

August and July had the lowest revenue, even though summer is busy with tourists. This might mean rides were shorter or fewer people used taxis.

February and January had less revenue, possibly because of cold weather and fewer people traveling.

Show the proportion of each quarter of the year in the revenue

Proportion of Revenue by Quarter
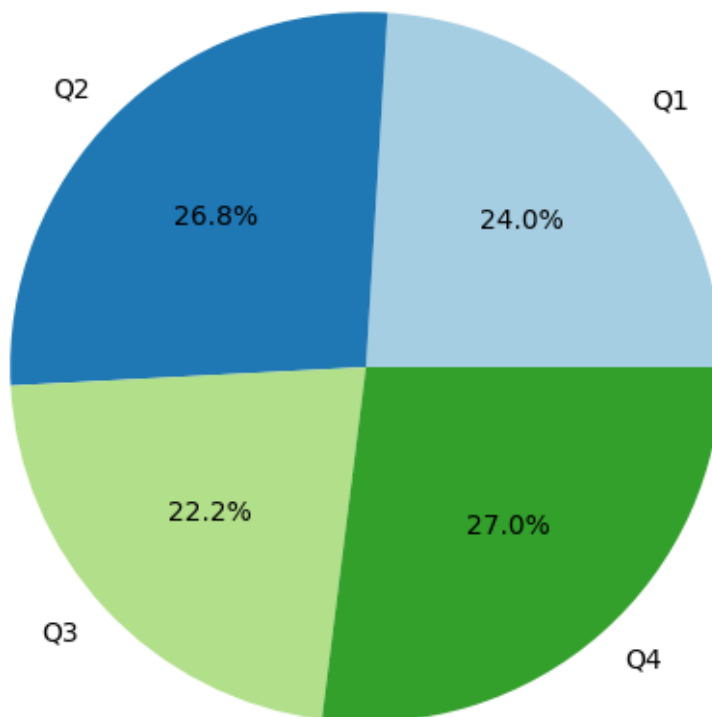
Q4 (October to December) had the highest revenue, slightly more than Q2. This could be due to fall events, holidays, and people traveling more.

Q2 (April to June) also had strong revenue, possibly from good weather and spring tourism.

Q1 (January to March) had less revenue, likely due to winter weather and slow travel.

Q3 (July to September) had the lowest revenue overall — maybe fewer rides, vacations, or more competition from other transport options.

Show how trip fare is affected by distance
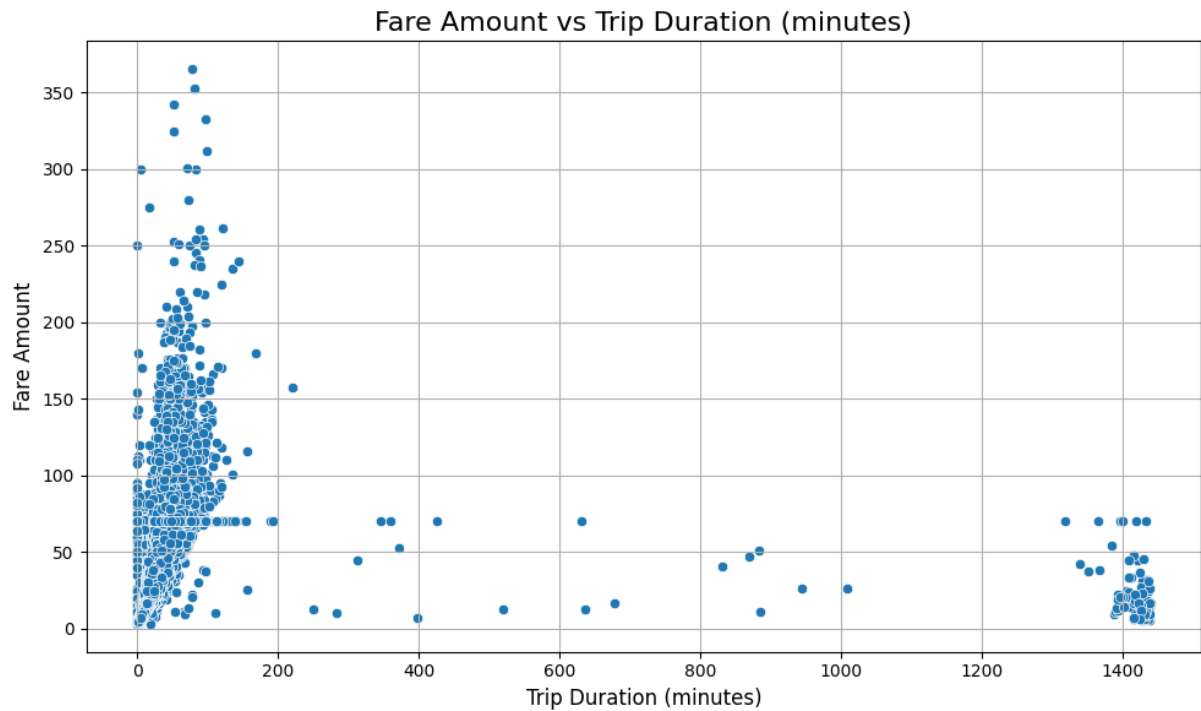
Correlation between Trip Distance and Fare Amount

Trip Fare vs Trip Distance
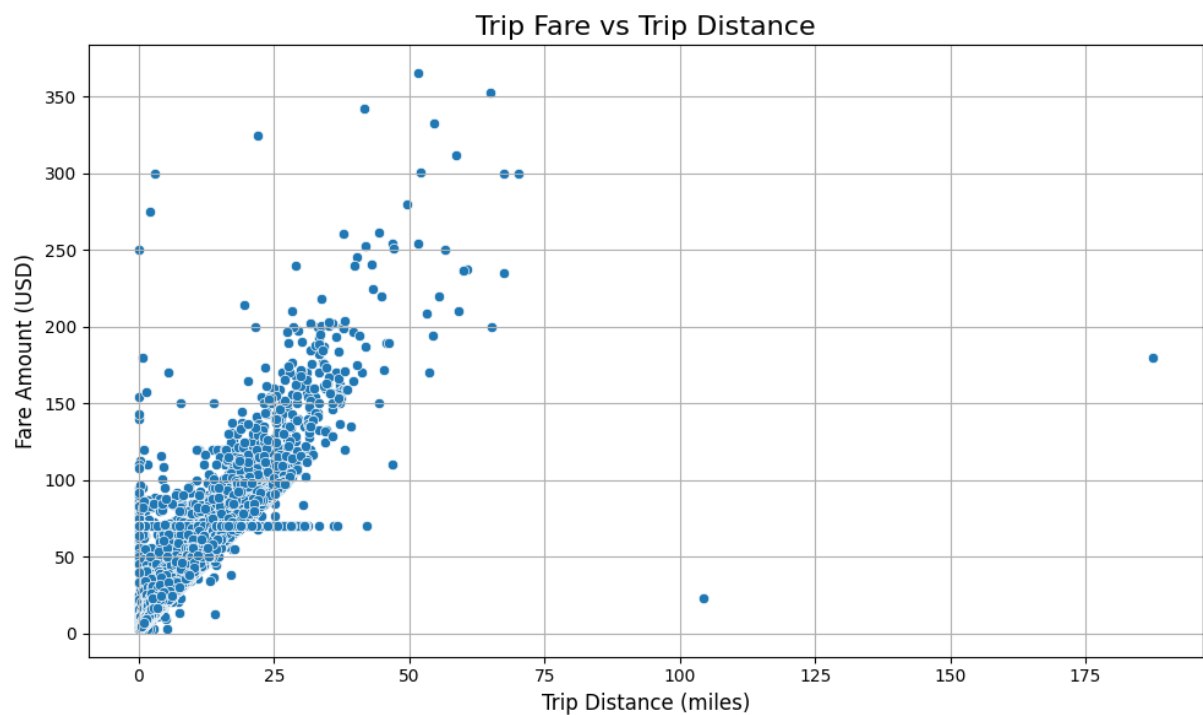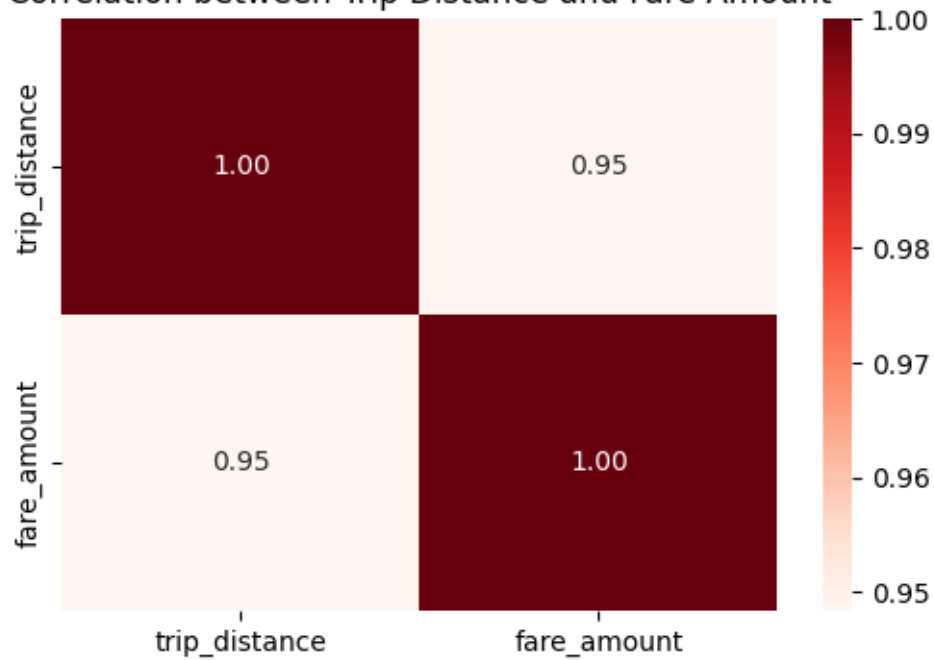
A correlation of 0.95 means that as trip distance increases, the fare amount also increases — almost directly.

Longer rides almost always cost more.

Show relationship between fare and trip duration

Fare Amount vs Trip Duration (minutes)

The plot shows that most trips last less than 150 minutes, and fares mostly range from 10 to 150

Some trips have very long durations (over 1000 minutes) but low fares — these are likely errors or special cases (e.g., parked or waiting time not billed).

There is a general upward pattern, especially for trips under 200 minutes — meaning that as duration increases, fare tends to increase too. A correlation of 0.3320 between fare_amount and trip_duration means there is a positive but weak relationship — fares do go up as duration increases, but not as strongly as they do with distance.

Show relationship between fare and number of passengers

Fare Amount vs Passenger Count
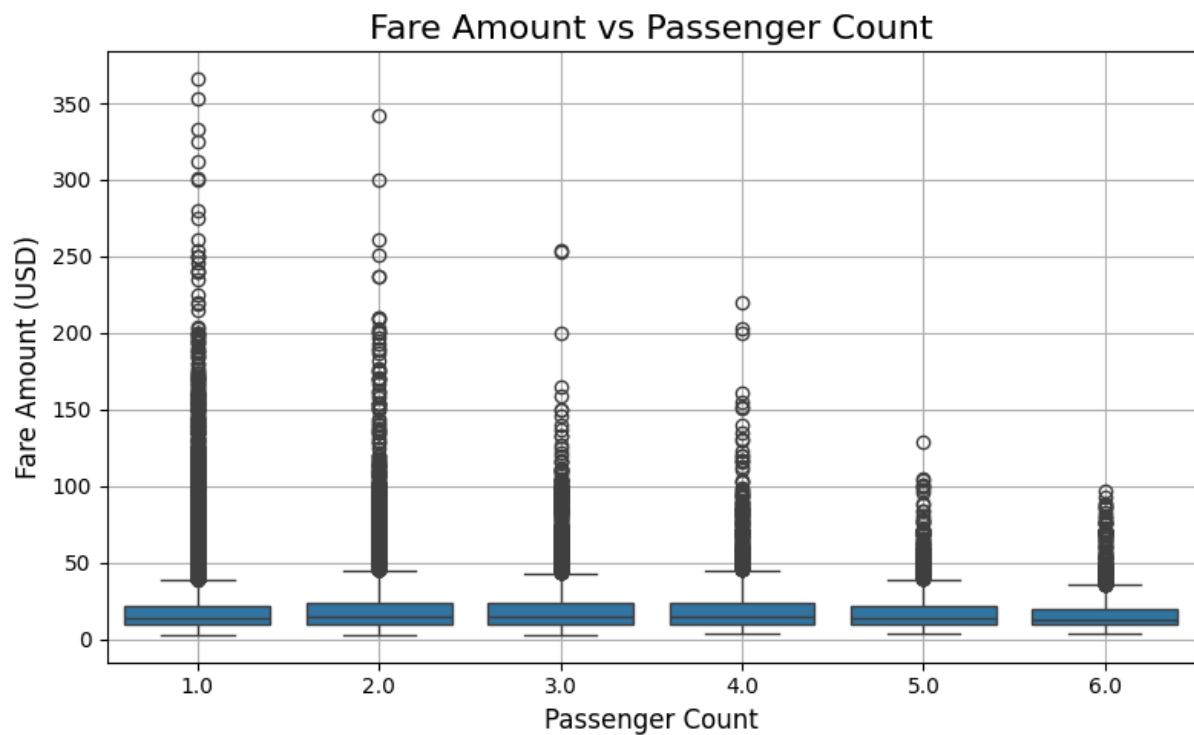
This is a weak positive correlation.

So, adding more passengers might increase the fare slightly (e.g., group airport trips), but not consistently. Fare is not strongly affected by how many passengers are in the taxi. This suggests passenger count is not a major factor in fare pricing for most trips.

Show relationship between tip and trip distance


Tip Amount vs Trip Distance
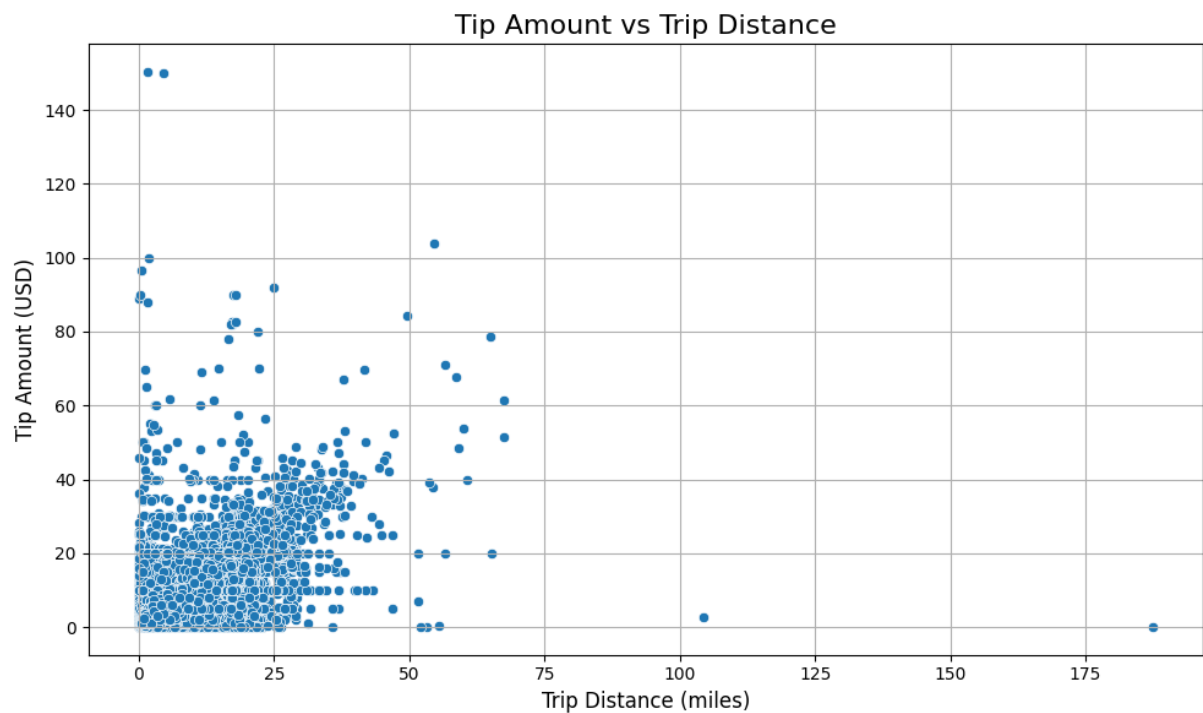
This shows that most trips are under 25 miles and tips are mostly under 40. We can see a positive trend as the trip distance increases, the tip amount also tends to increase. Some outliers show very high tips (over 100), but these are rare. This is a strong positive correlation.

It means that longer trips often receive higher tips — maybe because:

They cost more overall

Analyse the distribution of different payment types (payment_type).
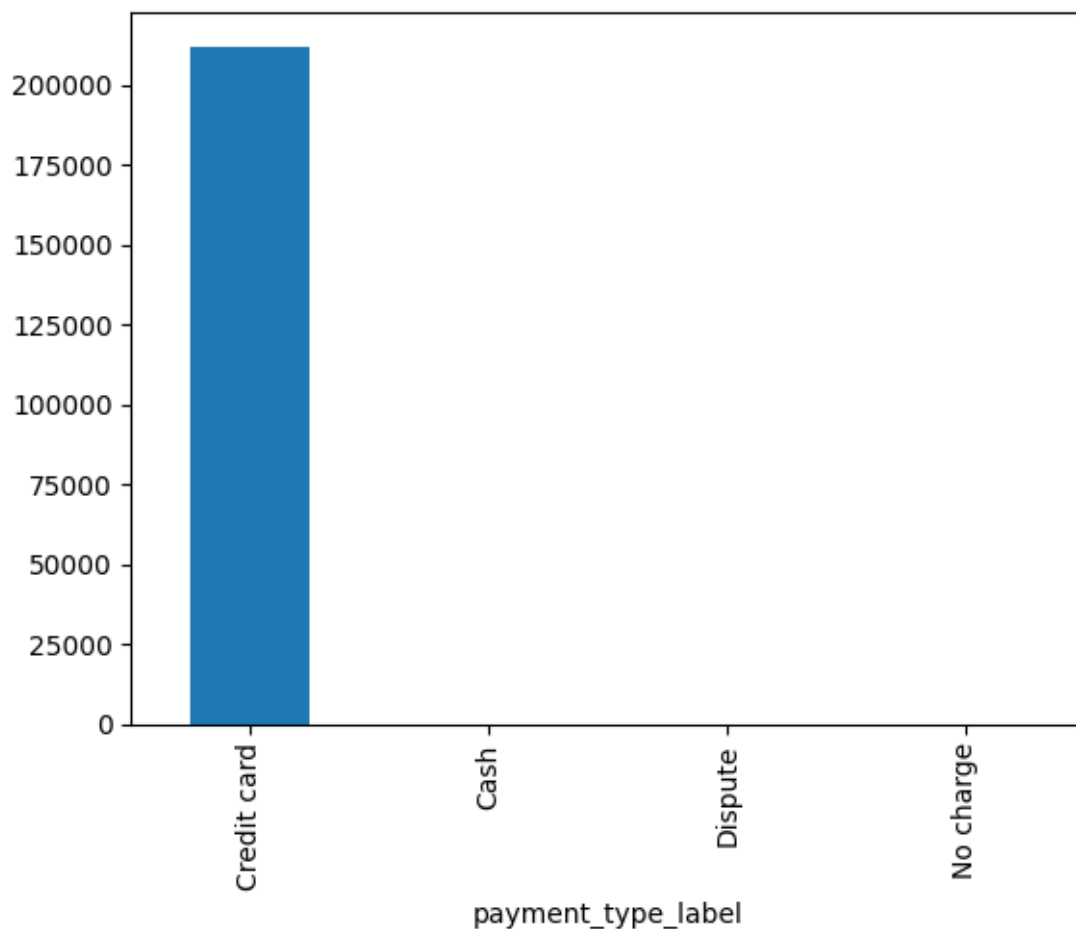
Credit cards are clearly the most preferred payment method in NYC taxis.
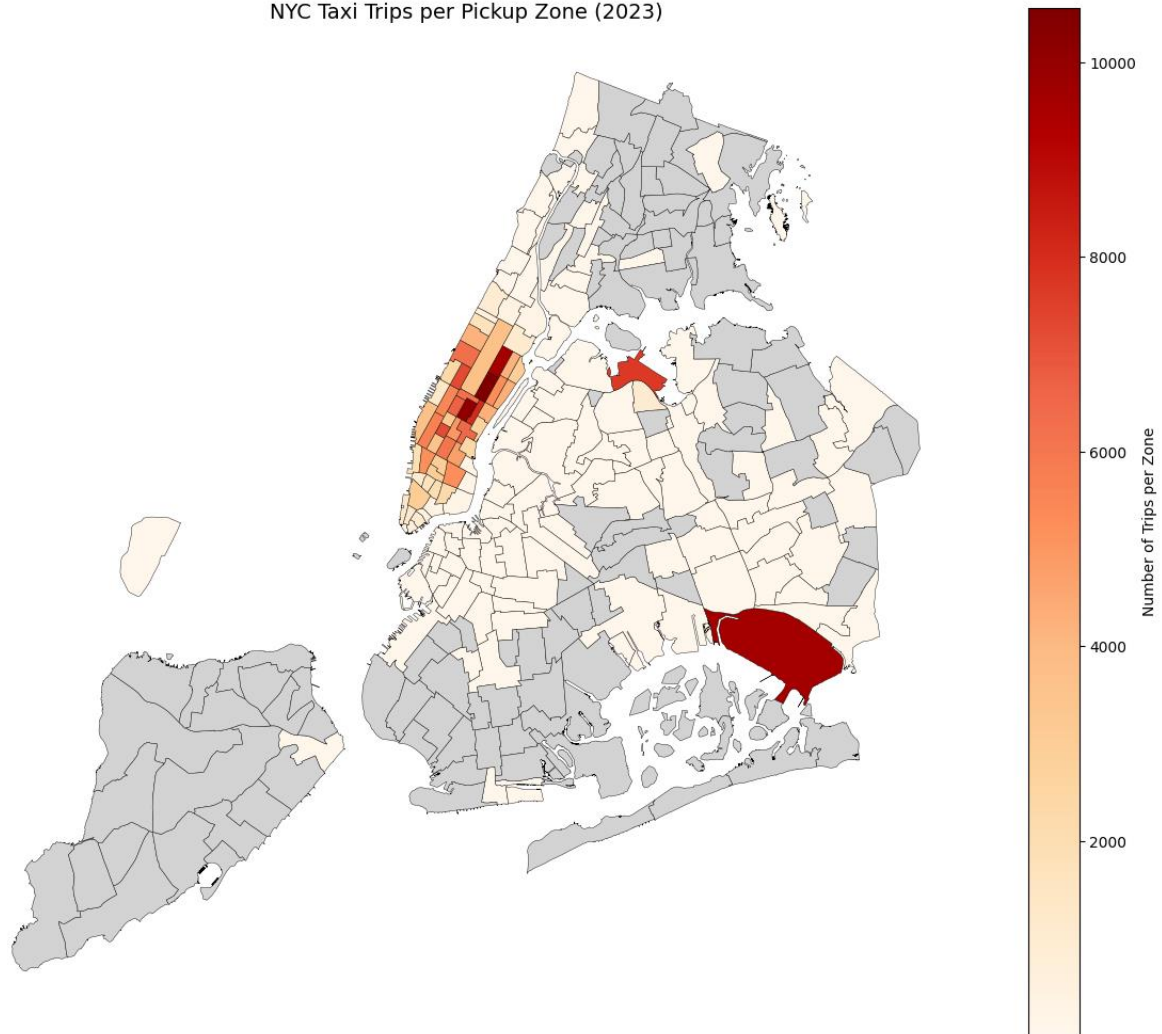
This might be due to:

More people using apps or card terminals.

Business travelers needing receipts.

Convenience and safety over carrying cash.

Define figure and axis

NYC Taxi Trips per Pickup Zone (2023)

Upper East Side (both North and South) is very busy — May be lots of people live or work there, and many seem to take taxis from that area.

Midtown areas (Center and East) are also very active —May be this makes sense because they are full of offices, hotels, and shops.

Airports (JFK and LaGuardia) are big pickup points — May be likely from people arriving in the city and needing taxis to their destinations.

Penn Station and Times Square are important spots too — May be full of tourists, commuters, and city visitors.

Murray Hill and Lincoln Square East also show strong activity — May be these are popular residential and nightlife areas.

You can consider the following points:

- Busiest hours, days and months

- Trends in revenue collected

- Trends in quarterly revenue

- How fare depends on trip distance, trip duration and passenger counts

- How tip amount depends on trip distance

- Busiest zones

Busiest hours, days and months: Busiest Hours: Most pickups happen between 2 PM to 7 PM, with a clear peak at 6 PM. Early mornings (0–5 AM) have the lowest activity.

Busiest Days: Thursday and Wednesday are the busiest days, followed closely by Friday. Sunday and Monday are the least busy.

Busiest Months: May and October had the highest number of trips, while August and September were among the lowest. **Trends in revenue collected**

Highest in October, followed by May and March. July and August collected the least revenue.

*Trends in Quarterly Revenue *

Q4 (Oct–Dec) and Q2 (Apr–Jun) had the highest revenue (~27% each of the total).

Q3 (Jul–Sep) had the lowest revenue (~22%), possibly due to summer vacations.

How fare depends on trip distance, trip duration and passenger counts

Trip distance has a strong effect on fare (correlation ≈ 0.95) — longer trips = higher fare.

How tip amount depends on trip distance Trip duration has a weak effect on fare (correlation ≈ 0.33) — longer time doesn't always mean higher fare.

Passenger count has little to no effect on fare.Tip amount increases with trip distance (correlation ≈ 0.79).

Passengers tend to tip more on longer rides — possibly airport or out-of-city trips.

Busiest Pickup Zones

Upper East Side South and Midtown Center top the list.

JFK and LaGuardia Airports are also major pickup points.

Other high-demand zones: Times Square, Penn Station, Murray Hill, and Lincoln Square.

Many outer zones had very few or zero pickups, especially in the Bronx and Staten Island. Focus taxi availability in central Manhattan and airport zones, especially during weekday afternoons and evenings.

Detailed EDA: Insights and Strategies

Visualise the number of trips per hour and find the busiest hour
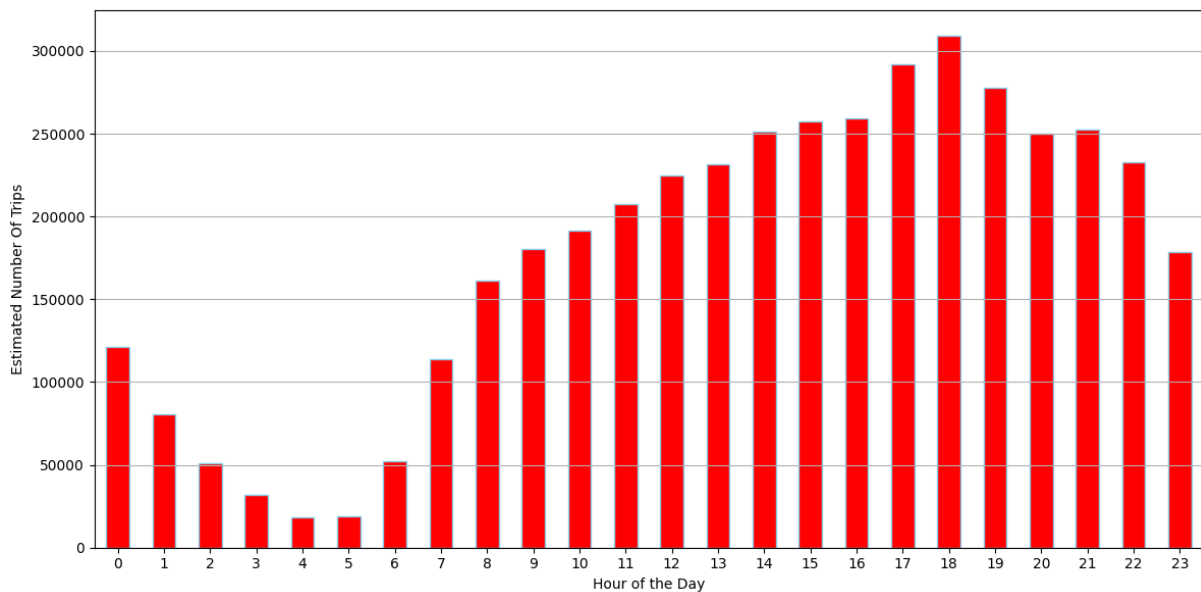

The busiest hour is 6 PM — this is likely due to people leaving work, going out, or catching evening events.

Rush hour traffic starts building around 3 PM and continues strong until 8 PM.

The quietest time is early morning, especially between 3 AM to 6 AM.

There is also a morning rise starting at 7 AM, with steady growth until midday.

Identify top 10 zones with high hourly pickups. Do the same for hourly dropoffs. Show pickup and dropoff trends in these zones.

Top 10 Zones by Hourly Pickups

Top 10 Zones by Hourly Dropoffs
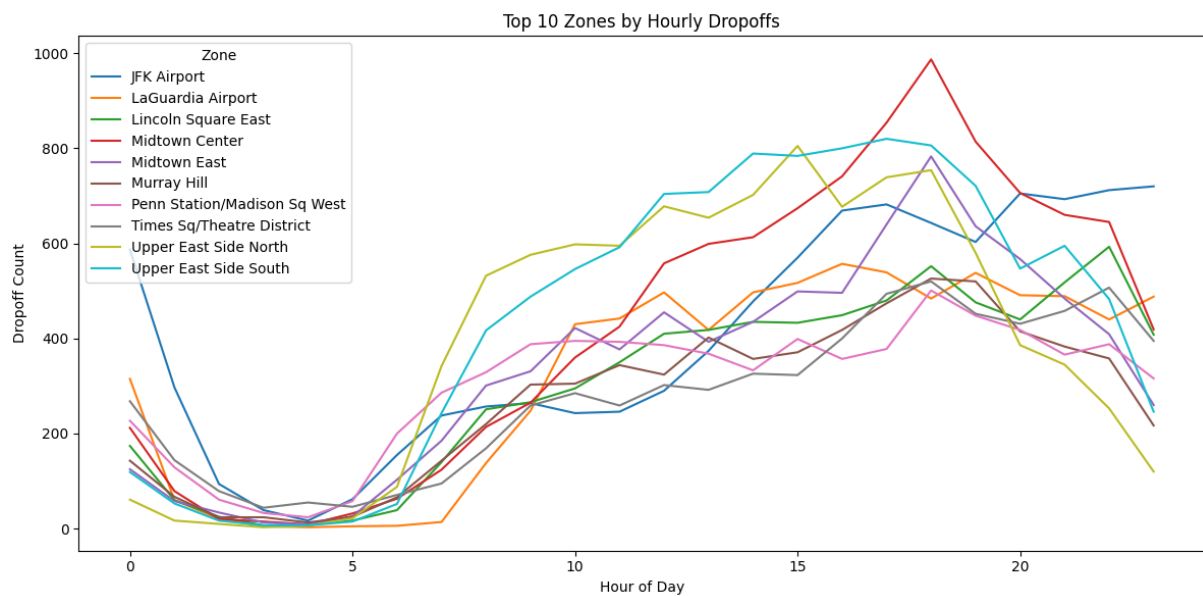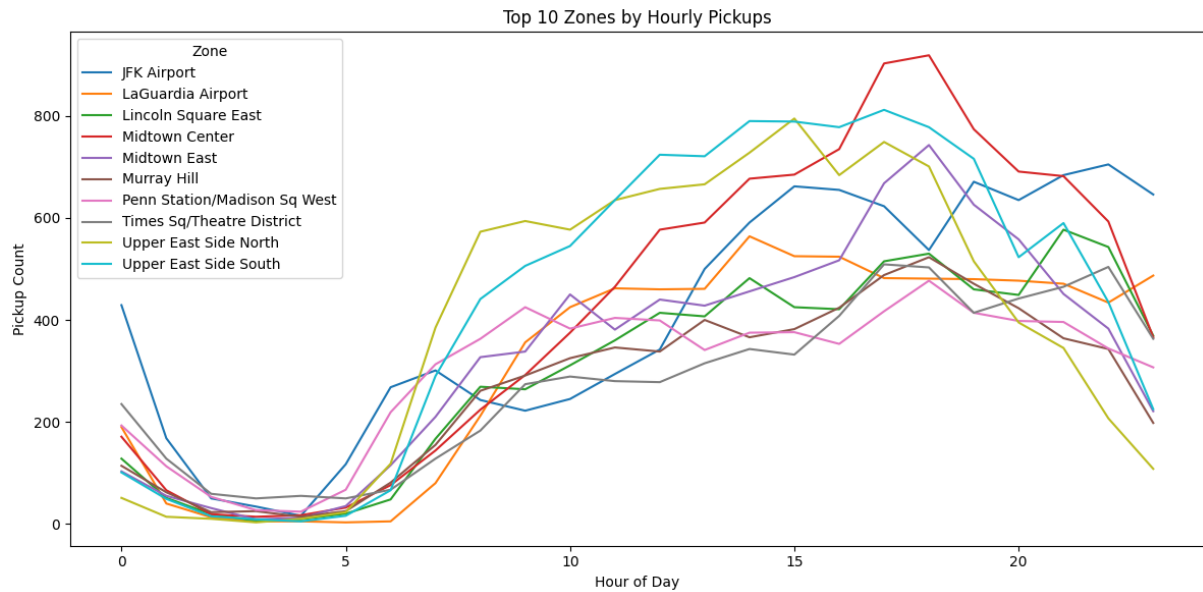
Pickup Trend: Most zones show gradual increase from morning, peaking between 15:00 to 18:00.

Midtown Center and Upper East Side South peak strongly in late afternoon.

Airports show more stable or early pickups — JFK has a high spike around midnight and early morning.

Dropoff Trend: Similar shape to pickups but with peaks more spread out.

Midtown Center hits highest dropoffs around 17:00–18:00, likely reflecting workers or evening crowds.

Airports (JFK and LaGuardia) have early morning dropoff peaks — showing arrivals for flights.

Residential areas like Upper East Side show strong evening dropoffs, suggesting people going home.

Analyse the fare per mile per passenger for different passenger counts
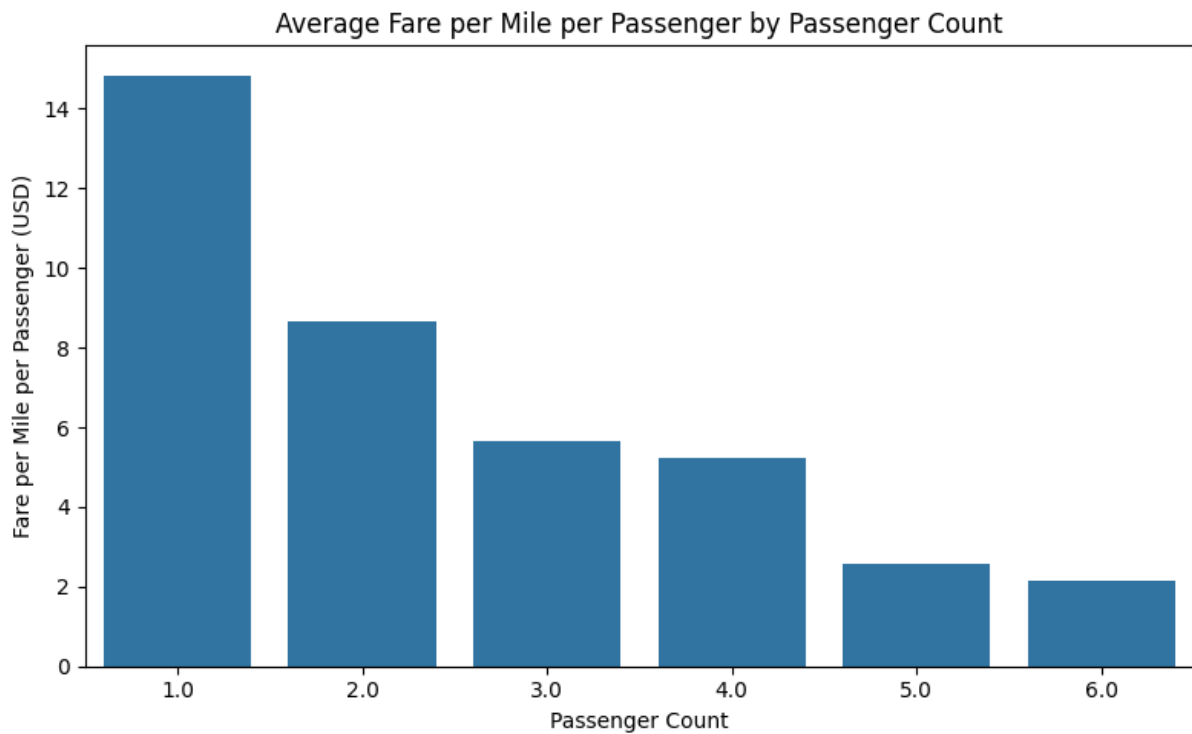
Solo riders pay the most per mile per person — nearly 7x more than a group of 6.

As passenger count increases, the cost per person drops sharply.

The biggest savings happen when going from 1 to 2 passengers, and again from 2 to 3+.

For groups of 5 or 6, the fare is split efficiently, costing just 2–2.5 per mile per person.

Compare the average fare per mile for different days and for different times of the day

Average Fare per Mile by Hour and Day of Week
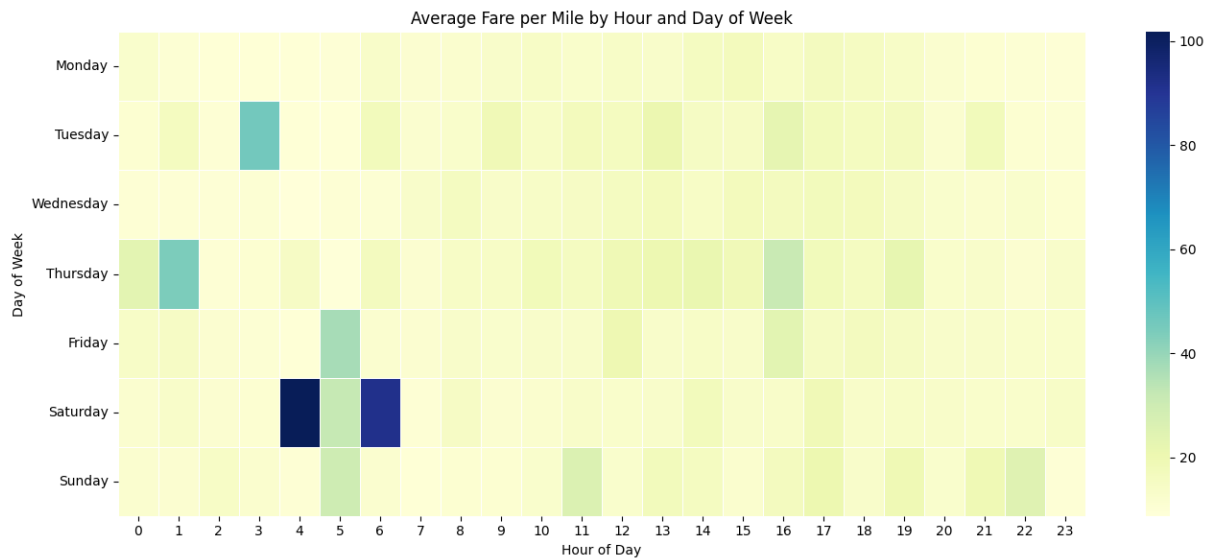
Fare per mile is higher late at night and early morning, especially on weekends.

These high values are often due to short trips where base fare dominates, or night surcharges.

During the day, pricing is more predictable — possibly due to regular commuting and longer business-related trips.

Compare fare per mile for different vendors

Average Fare per Mile by Vendor and Hour
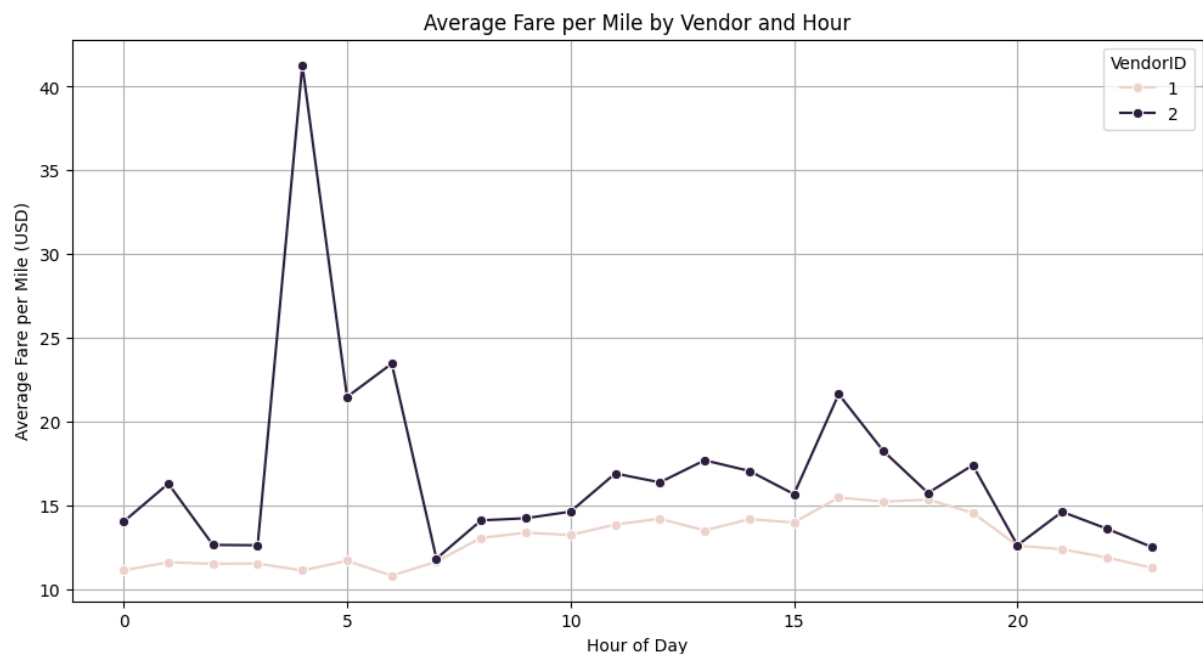
Vendor 1 has a steady, consistent pricing model, slightly increasing during high-demand hours. Vendor 2 charges significantly higher fares per mile, especially in early morning

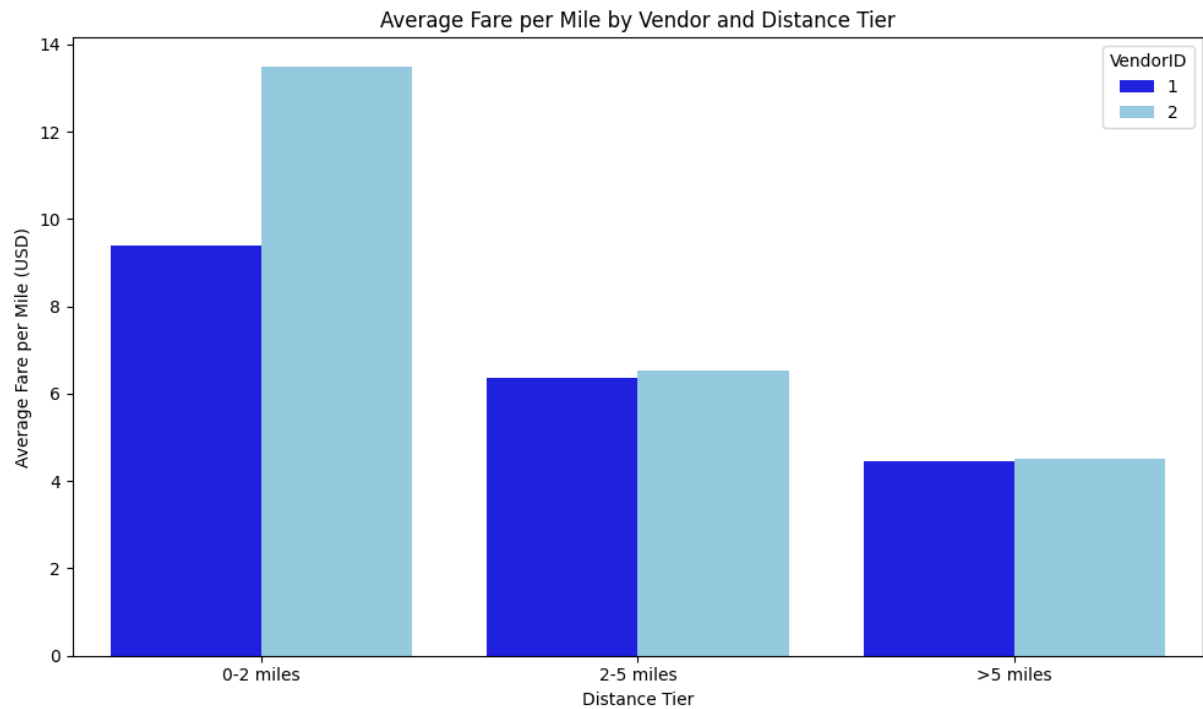and afternoon. These may be caused by: Short-distance trips, inflating fare-per-mile Surcharge patterns , Small sample sizes during odd hours

Defining distance tiers

Average Fare per Mile by Vendor and Distance Tier

Short rides are where the biggest pricing difference exists.

Vendor 2 seems to charge more aggressively for short distances.
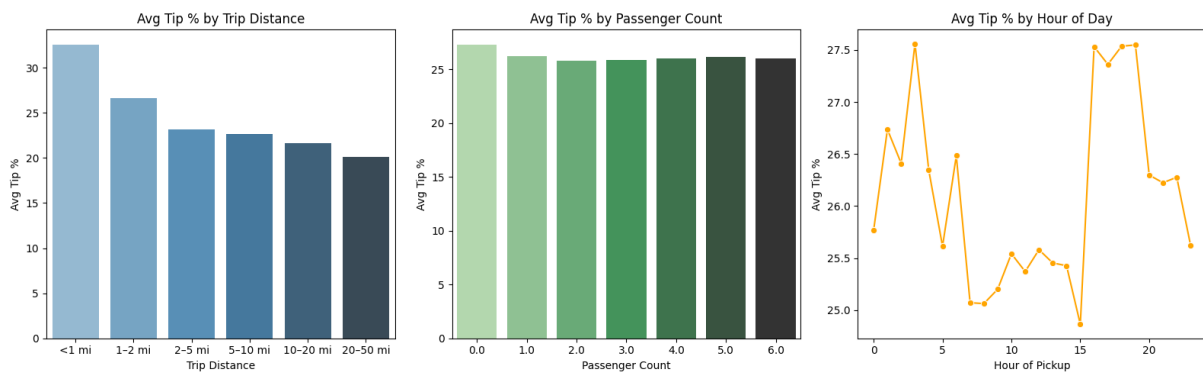
Longer trips normalize fare rates, suggesting base fares have less impact when distance is high.

This analysis can help passengers and regulators:

Passengers may prefer Vendor 1 for short trips.

Fleet managers or policymakers may monitor Vendor 2 is short-trip pricing behavior.

Analyze tip percentages based on distances, passenger counts and pickup times

Short trips (<1 mile) receive the highest tip percentage (around 32%). Tip percentages remain very stable (around 26 to 27%) regardless of how many passengers Tip % gradually declines as trip distance increases:

1 to 2 miles are around 27%

5 to 10 miles are around 23%

20 to 50 miles are around 20% are on board. Slight dip for 0 passengers, which might represent errors or no-tips cases (e.g., test rides or airport vouchers). Tips hover around 25 to 27.5%, with a few small peaks:

Around 2 AM and 18 to 20 PM show slightly higher tipping.

Lowest tipping happens around 15:00 (3 PM) and early morning (8 AM).

Compare trips with tip percentage < 10% to trips with tip percentage > 25%
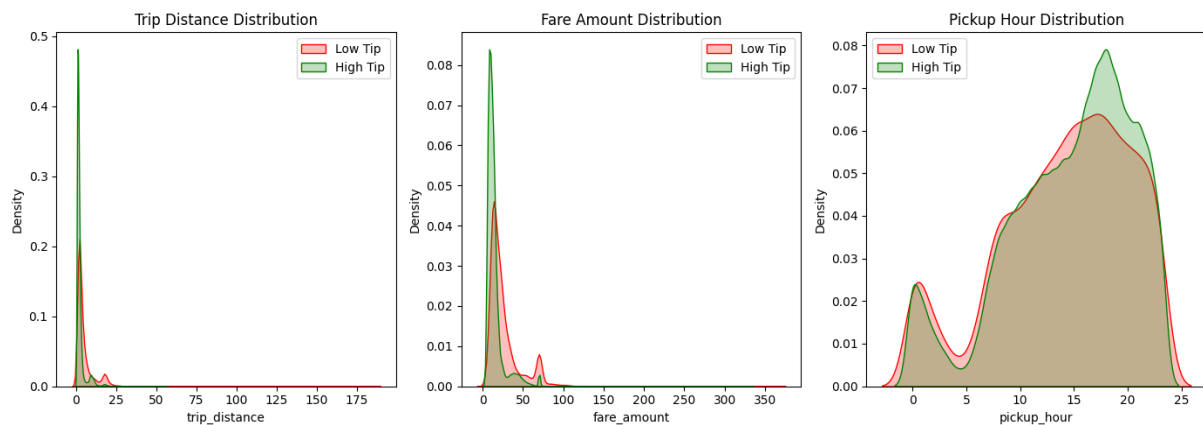

Shorter, cheaper trips tend to earn better tips — possibly because they feel faster, smoother, and more personal.

Longer trips get worse tip percentages, even though they generate more revenue.

Tipping behavior is more emotional than logical — not tied to fare or passenger count, but possibly to ride experience or convenience

See how passenger count varies across hours and days

Average Passenger Count by Hour and Day of Week
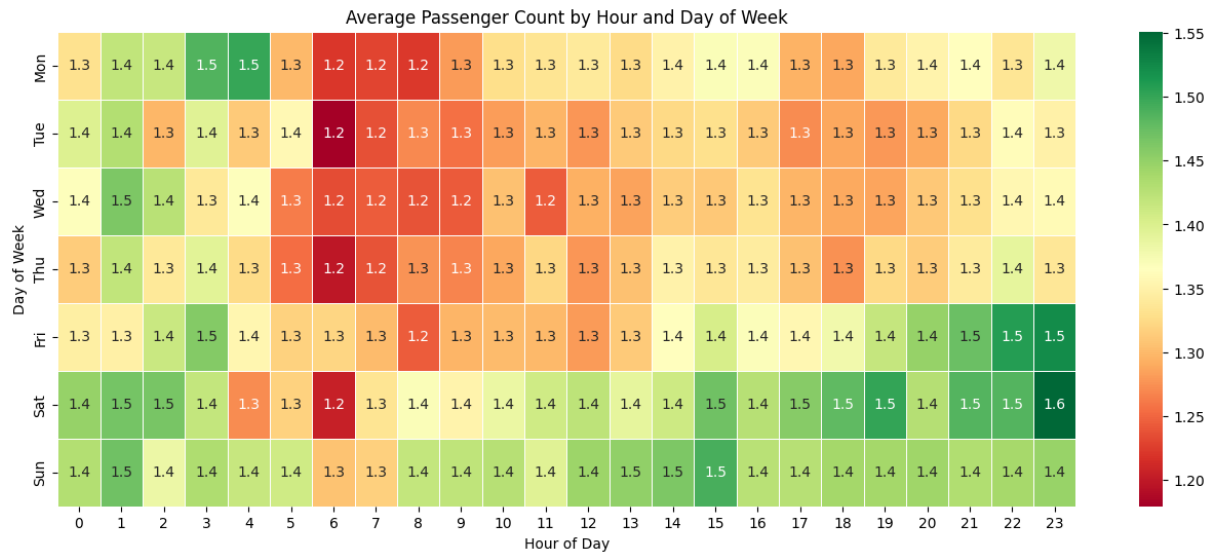
Saturday late night trips tend to have more passengers, perhaps because people ride together after events.

Weekdays are dominated by solo riders, especially in the early morning hours.

How often is each surcharge applied?

Rate of Congestion Surcharge Application by Hour
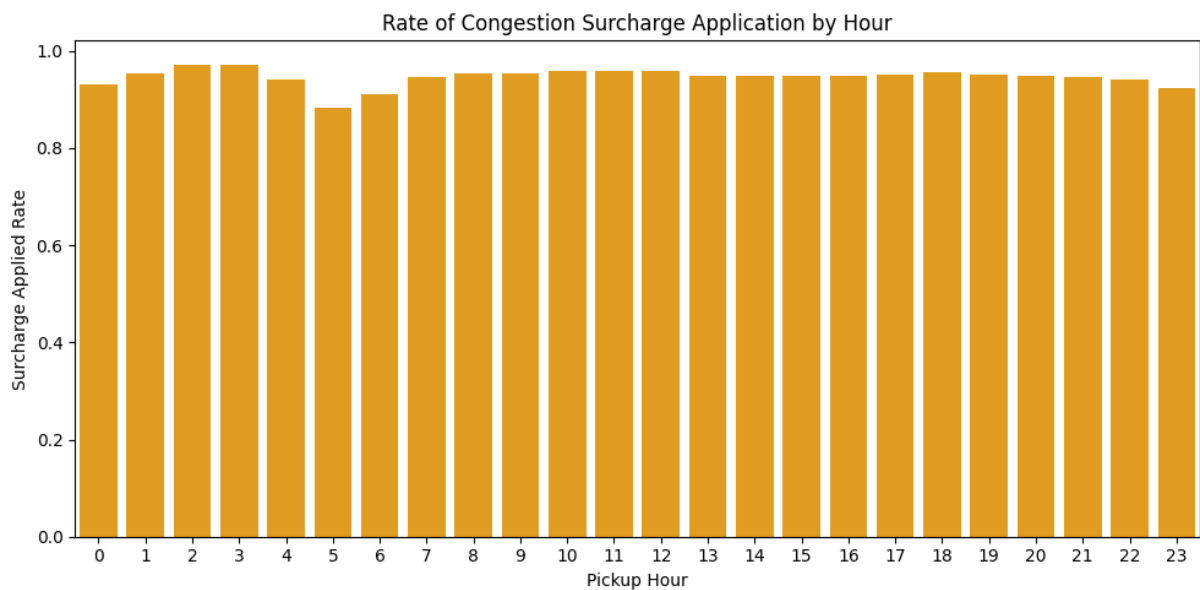
The surcharge is applied in over 90% of trips across nearly all hours.

Peak rate is seen between 2 AM and 3 AM, possibly due to fixed surcharge zones or fewer trips concentrated in Manhattan.

5 AM shows the lowest application rate (around 88%).

A small dip also occurs at 23 PM and 6 AM, just under 90–91%.

Final Insights and Recommendations

Conclude your analyses here. Include all the outcomes you found based on the analysis.

Based on the insights, frame a concluding story explaining suitable parameters such as location, time of the day, day of the week etc. to be kept in mind while devising a strategy to meet customer demand and optimise supply.

Busy Times and Days The busiest time of the day is in the late afternoon and early evening (3 PM – 7 PM).Thursday and Wednesday are the busiest days, while Sunday and Monday are the least busy.Most of the revenue (88%) comes during the daytime (6 AM – 10 PM). Only 12% comes at night

** Popular Pickup Zones** The most active pickup places are:Upper East Side South Midtown Center,JFK and LaGuardia Airports. People often get picked up from airports in the morning and dropped off at night.

Fare Patterns Fares go up as trip distance increases — longer trips cost more.But fare per mile is highest for short trips because of the minimum base fare.Vendor 2 usually charges more per mile than Vendor 1, especially for short trips.

Tipping Behavior Riders give the highest tip % for short trips (<1 mile) — sometimes over 30%. Longer trips get lower tip % (about 20%), even though the total fare is higher. Tip % doesnot change much based on the number of passengers. Passengers also tip a bit more during evening and late-night hours.

Passenger Count Patterns Most rides have 1–2 passengers. On weekends (especially Saturday nights), there are slightly more passengers per ride — up to 1.6 on average. Weekday mornings are mostly solo riders, likely commuters.

Surcharges The congestion surcharge is applied in over 90% of trips, at all times of the day.

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

Send More Taxis During the Busiest Hours (3 to 7 PM) Most trips happen between 3 PM and 7 PM, especially at 6 PM (hour 18) with 15,456 trips. Dispatch more cars during these hours to reduce wait time and missed bookings. Also, these hours have more solo rides and higher tipping — good for driver earnings.

Focus on Mid-Week Days (Wednesday & Thursday) Thursday and Wednesday are the busiest days, each with over 33,000 trips. Add more drivers on these days compared to Monday and Sunday, which have the lowest demand.

Airport Routing During Afternoon and Evening Airport pickups (from JFK & LaGuardia) are highest between 2 PM and 10 PM. Make sure enough taxis are available near airports in this window to meet demand.

Watch Vendor Pricing Vendor 2 charges more per mile than Vendor 1. Try to balance dispatching fairly so customers aren't overcharged — especially for short trips where fare per mile is highest.

Expect Solo Riders in the Evening Solo trips are highest at 6 PM, followed by 5 to 8 PM. Plan more vehicles for individuals at these times — especially in downtown and commuter-heavy zones.

Tipping Is Best Between 3 PM and 7 PM Tip percentage is highest at 3 PM to 7 PM, peaking at 3 PM. Encourage drivers to work these hours — better chance of earning good tips.

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

Weekdays (especially Wednesday and Thursday)

Afternoons to evenings (3 PM to 8 PM)

Spring and Fall months (April, May, October)

Reason: These areas are always high in demand, especially during commute and event hours.

Great zones for solo rides, consistent tipping, and frequent pickups.

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

Lower Fare per Mile for Very Short Trips-Introduce a "micro-ride" flat fare or reduced per-mile pricing for trips under 2 miles.This makes pricing look fairer and encourages more short-trip demand.

Add Dynamic Pricing Based on Time of Day & Day of Week-Apply slight peak-hour surcharges (e.g., +5%) to these windows.This boosts revenue without major fare hikes, and aligns with how demand naturally rises.

Encourage Group Rides with Passenger-Based Discounts-Offer group fare bundles ,Promote carpooling in residential and nightlife areas to boost efficiency and total fare.

Reward Off-Peak and Long-Distance Trips with Discounts-Offer 5–10% off for trips longer than 8–10 miles during quiet hours (midnight to 6 AM).This encourages use and keeps your fleet working.

Match Vendor 2 Smartly on Competitive Routes-Match Vendor 2's pricing only in high-demand zones (e.g., Midtown, JFK), but keep lower base rates in low-demand zones.Helps win business in hot spots without overcharging elsewhere.

Use Tip Trends to Influence Driver Behavior, Not Pricing-Keep base fares steady in tipping hours, but offer driver bonuses or shift rewards during those hour.