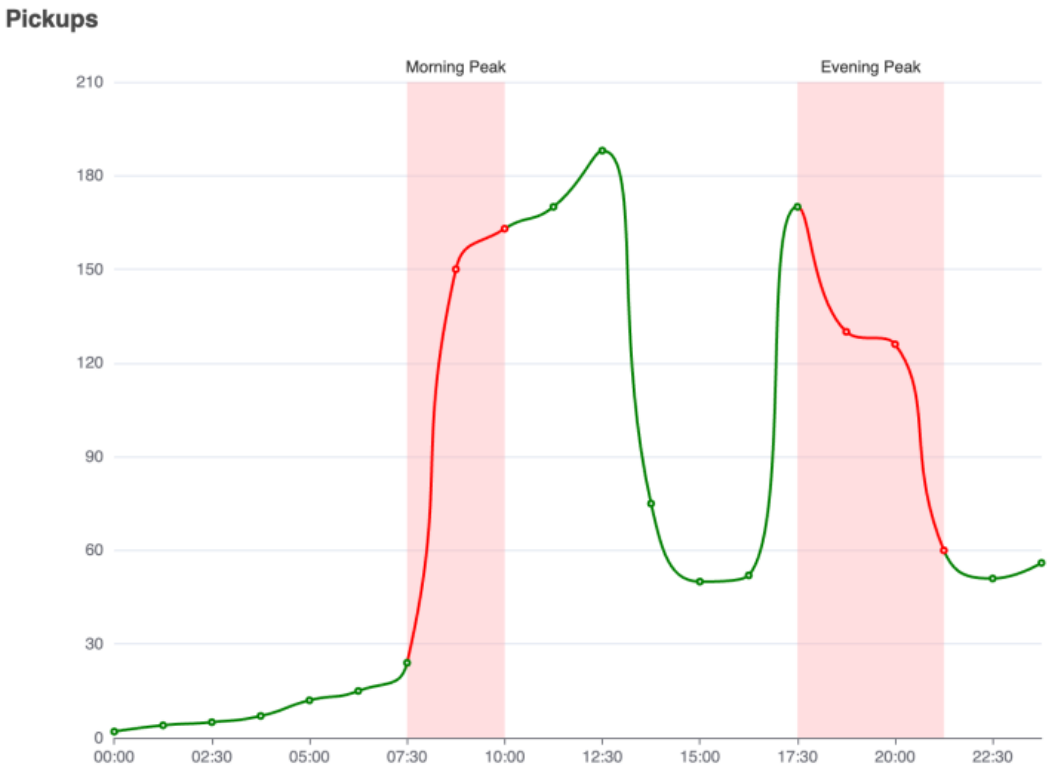
Complex Trend Queries

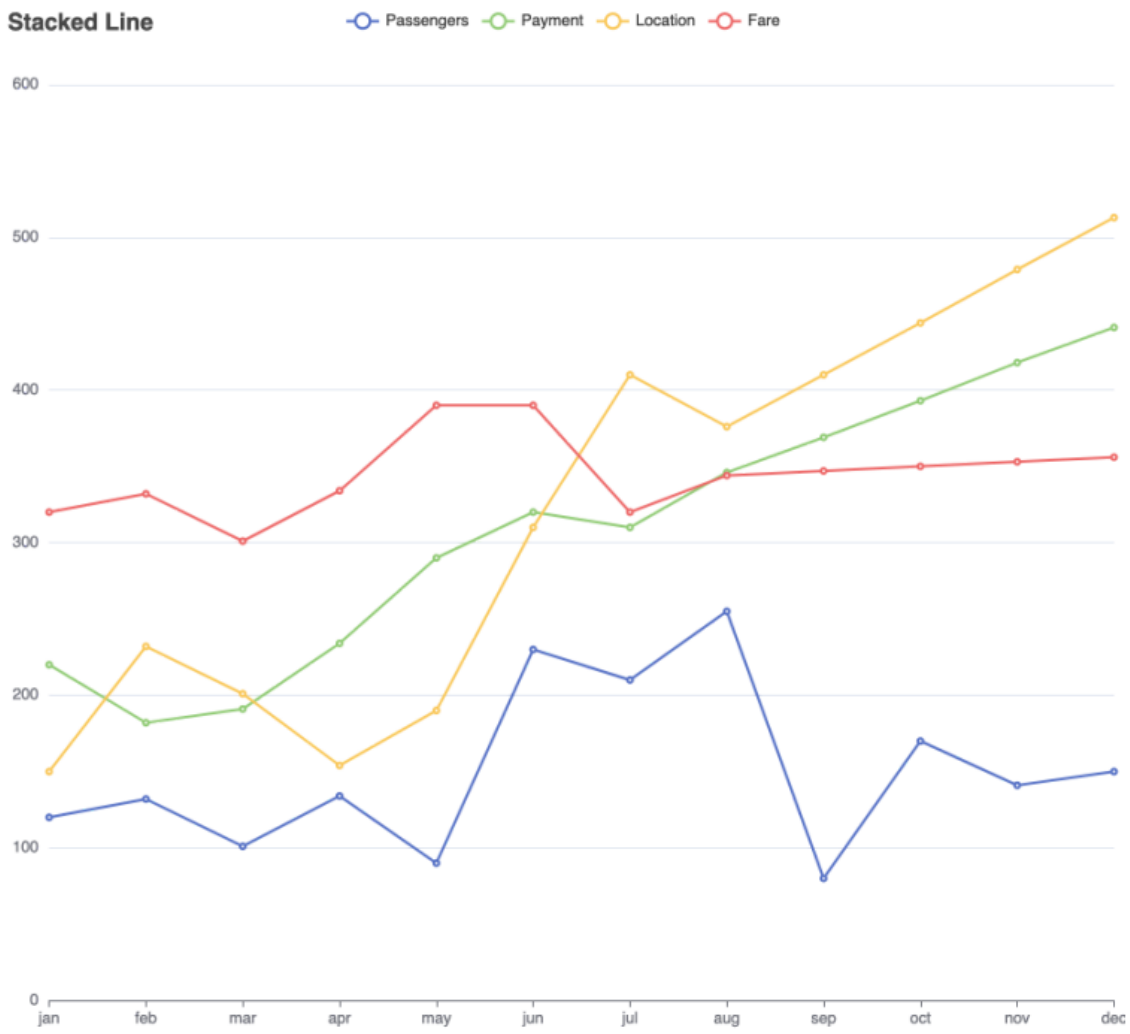
Query #1: Find the trend of passengers over time for the given time bin at the current and k - nearest clusters to maximize the pickups.

Taxi is the most suitable or popular means of commuting worldwide. Precisely and empirically estimating the demand for taxi passengers is critical for any ridesharing companies like Uber, OLA, and Lyft. Taxis efficaciously assigns their fleet to some pre-defined stands and minimizes passengers' waiting time, thus increasing their overall satisfaction and customer retention. Nowadays, trip information is available in the database, which we can use to analyze the patterns and trends in passenger demands in specific areas, maximizing the taxi driver's profit. The query will derive the past trends based on the parameters retrieved from the client. We will also fit a simple linear model, giving an appropriate estimation of the number of pickups in the nearby clusters of the client.



Query #2: Find the effect of Uber and Lyft on Yellow Cabs in NYC over time.

Uber and Lyft's growth in NYC saw a decline in the value of Yellow Cab medallions, which ended up ruining investments once seen as solid as a bar of gold. With passing time, a lot of cab drivers lost their medallions as they lost riders to Uber and Lyft. With the data available to us, we can easily find a trend to visualize the decline of yellow cabs with the growth of mobile based taxi services like Uber and Lyft. This query will focus on comparing the decline over years for a combination of parameters which will be based on user input and will include location, ride fares, tips and number of rides. This will give users a comprehensive overview and hopefully a better insight into this particular trend.



Query #3. Find the overall trend of cab trips during major holidays across a calendar year.

Across a calendar year, we can easily observe from the data available to us that the average population in NYC changes from holiday to holiday and how that affects an average cab driver. At holidays like Thanksgiving and Christmas where all the major tourist locations, shops and mall closed, people might not prefer to go out at all. This could help drivers and riders recognize the hot spots across the city during the holidays, if they ever decide to go out or drive a cab to earn some extra bucks.

Query #4: Track the number of passengers over time before and after COVID to see the effects that COVID may have had on people taking a taxi.

Another potential database query that can be used for our application would be tracking the number of passengers over time before and after the COVID-19 pandemic to observe the effects that the virus may have had on people taking a taxi.

Time is by far the most important variable to consider for our query and data analysis because it helps us determine the trends that we can draw conclusions from. We will have three main time periods based on the data we have: before the pandemic, during the pandemic, and after the pandemic. The years 2009-2019 will comprise the “before the pandemic” timeframe, years 2020 and 2021 will represent “during the pandemic”, and 2022 will be considered “after the pandemic”.

Variables to consider: Number of passengers: One of the primary variables we will consider in this query is the number of passengers that took rides on the NYC taxicabs. With the data given to us, we can find trends in how the number of passengers changed due to potential factors such as the COVID-19 pandemic by fitting that information into a linear graph. Number of COVID-19 cases in NYC : Another potential variable we can consider is the number of COVID-19 cases in NYC for the time that the pandemic took place. By representing the number of cases in a linear graph, we can make connections between how the number of cases could affect peoples’ reactions to the virus’ severity and their unwillingness to take a taxicab causing a decrease in the number of passengers.
By representing both the datasets on the number of COVID cases in NYC and the number of passengers against these timeframes, we can make astute observations on any cause effect relationships that may have occurred.

Query #5: Different payment methods used over time by passengers

One complex database query that our application can utilize is tracking the amount of different payment methods over time used by passengers for taxis to determine the effects the introduction of cash-less payments has had on the modern world. It is known that as time progresses, so does human innovation and inventions to better suit the needs of people and improve convenience.

The primary parameters that we plan on following are the different kinds of payment methods used by passengers in the transactions for the taxi as well as how many were used during our observed timeframe. The payment types recorded in the dataset include credit cards, cash, no charge, dispute, unknown, and voided trip.

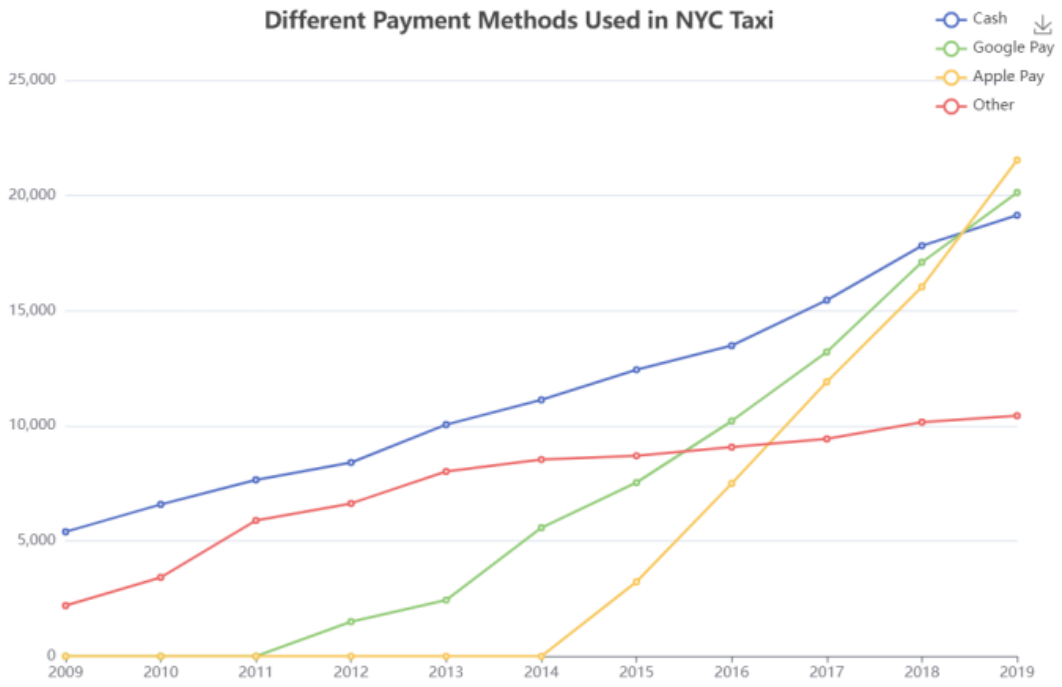


Figure 1: Sample Diagram of Potential Graph of NYC Taxi Payment Methods

In the case of our dataset, we have a considerable amount of taxi data from the years 2009 to 2022 so we'd be able to utilize the total numbers of the different payment methods used and plot them by year. As indicated by the sample plot, Figure 1, above, each payment method would be plotted linearly on a line graph showing the change of their counts over time.

Query #6: Find the trend of taxi driver income compared to that of inflation

https://www.bloomberg.com/news/articles/2022-05-23/nyc-taxi-drivers-call-for-first-fare-increase-in-a-decade?leadSource=uverify%20wall

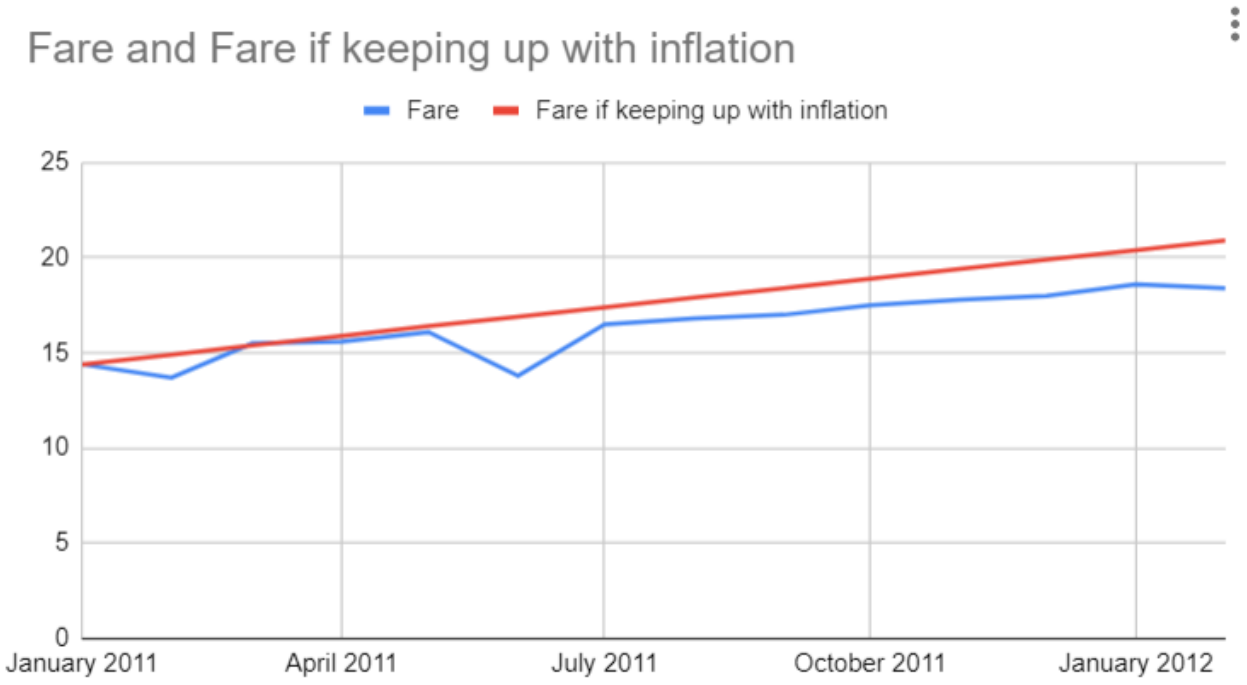
The fares of taxi drivers have become a hot topic with many unions fighting for increased fares. This is especially so with the rise of inflation. We must look at the change in income and prices over time, benchmarked to inflation. Taxi wages depend on a number of things, from the average fare they get per trip, and how many trips they

can take per day. It also depends on how much they get tipped. Tipping culture for taxis isn't always a percentage of the fare, sometimes it can take the form of an informal "keep the change" style tip. Do surcharges happen more or less often than before? What about extra fees? All things considered, when you add up all the sources of income, are taxi wages keeping up with inflation?

There are several things we need to do to create our final change over time comparision:

- Change of average fare amount over time (main source of income)
- Change of average number of trips per day over time (main source of income)
- Change of average tip over time (Do tips increase with inflation, has tipping culture itself changed. Tipping culture for taxi's isn't always a percentage based value, often times it takes the form of "keep the change")
- Change of average use of Extra Fees over time. E.G a driver may get the extra fee on average 1.2 times a day. How has this average extra fees per day changed over time.
- Change of improvement surcharge over time. E.G a driver may get the improvement Surcharge on average 1.2 times a day. How has this average extra fees per day changed over time.

Finally, combine these totals to show the change in average daily wage and compare it the average daily wage needed to keep up with inflation.



Average Fare over time