Find the nearest cluster with the maximum number of pickups in the following 10-minute bin interval.

Taxi is the most suitable or popular means of commuting worldwide. Precisely and empirically estimating the demand for taxis is critical for taxis or any ridesharing companies like Uber. 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 predict 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 fit a simple linear model which will give an appropriate estimation of the number of pickups in the nearby clusters of the client.

Parameters that we consider for getting the trend w.r.t to time.

Time-bin:

Why directly not consider the exact time?

Time in a bin bucket will give a finite approximation of the number of pickups. We can't expect pickups in every subsequent time stamp w.r.t seconds or minutes. So estimating the taxi demand for a bin bucket (by default 10) will give a good approximation. Moreover, it will also take the driver to reach the nearby cluster by 10 minutes at least. We need the current time bin to predict the pickups in the next bin. Each time bin will have a specific pattern associated with it.

Time-bin plays a crucial role in analyzing trends.

Pickups

From the graph, it is clear that the peaks will reach only in the mornings and evenings.

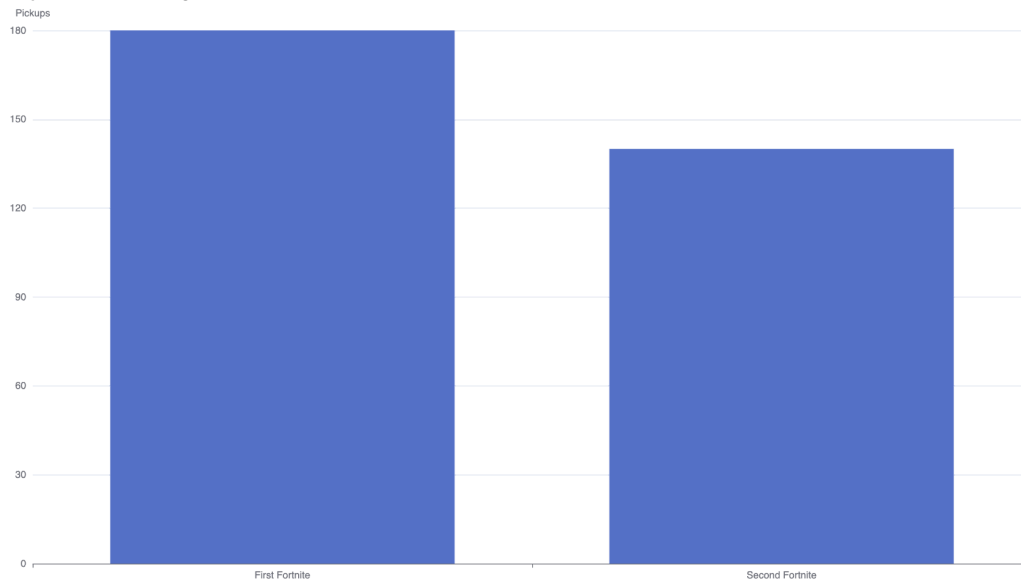**Cluster:**

The current cluster of the driver.

Reason:

We need the driver's cluster to find the near k clusters as the trend for each cluster will vary. Some would have linear or non-linear patterns.

Cluster threshold:

The threshold is the number of neighboring clusters to consider by default, it is 5.

Fortnite:

Fortnite vs pickups in the month of august 2021

We can see the first Fortnite has many pickups compared to the second because the payday falls in the first Fortnite.
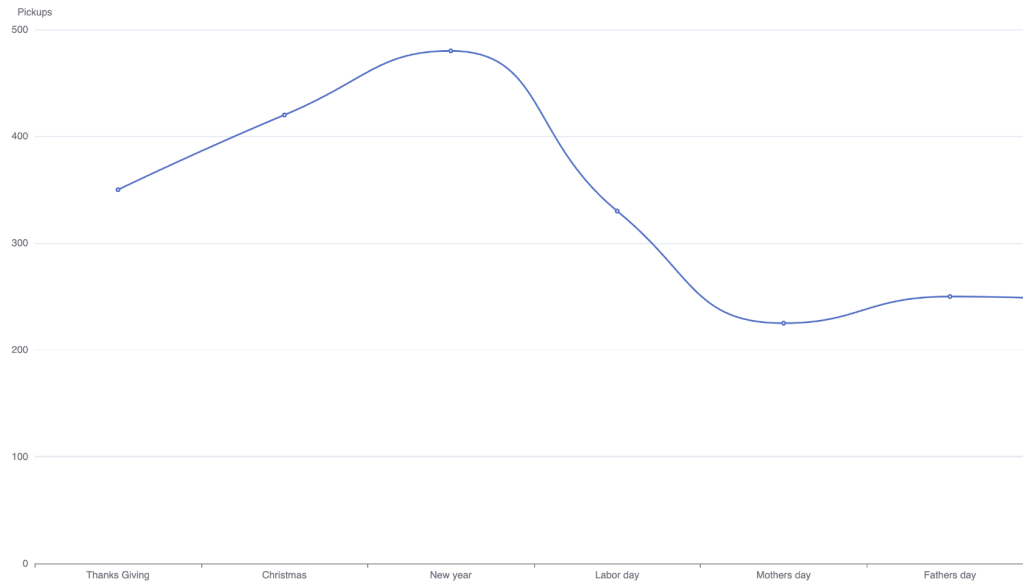
Day categorization:

It contains discrete values ('weekday,' 'weekend,' 'holiday')

The pattern in each cluster will vary in terms of the above values.

From the graph above, we can see that the average number of pickups on the weekends is very high compared to the weekdays.

Pickups vs popular holidays

Pickups

500 —

400 —

300 —

200 —

100 —

0 —

Thanks Giving   Christmas   New year   Labor day   Mothers day   Fathers day

Popular holidays usually have a lot of pickups; the rest of the other holidays have fewer pickups as people prefer to celebrate it at home on unpopular holidays.

Day name:

It is considered to analyze the past trend based on the name of the weekday or weekend. For example, if we find the previous window trend for Saturday (weekend), we can't just go to the past weekend data, which includes Sunday.

Weather data:

Precipitation will influence the pickups. They are highly correlated with one another.

Weather condition vs pick up frequency in each day
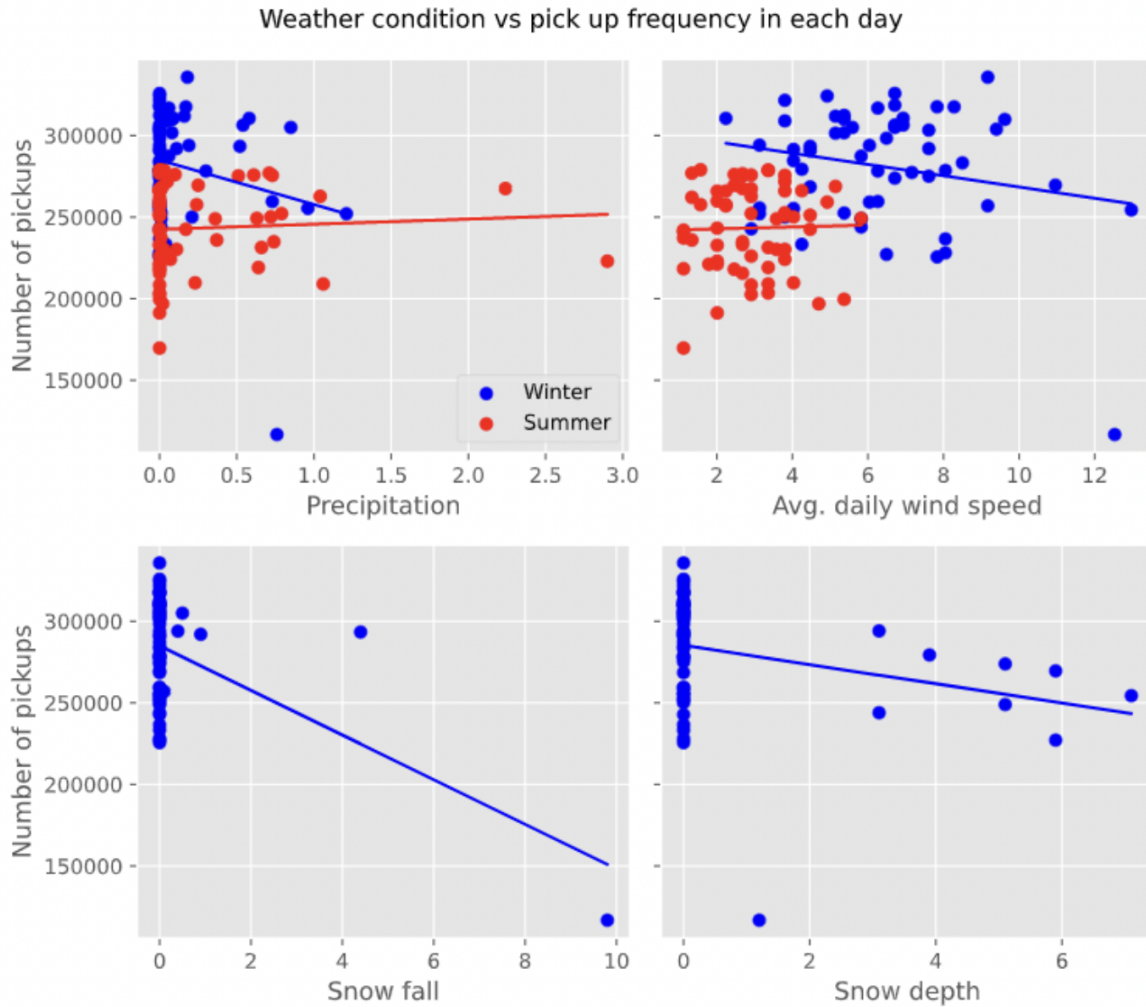
You can see from the above graphs how the weather impacts the pickups.

Note:

Some features will get ignored based on the data.

For example, there is no point in considering the day name to analyze the previous trends

if it is a holiday. For instance, if Christmas occurred on Monday no point in analyzing

Christmas in the last window, which occurred on Monday.

Let's take a sample point and see what kind of pattern we are getting (The data displayed in the graph is not obtained from the real data)

Given:

Taxi driver X at the kth cluster on a working day Friday with normal weather at 37th-time bin (an assumption that the intervals of the day are segmented by 10-minute bins for example, from 12:00 AM to 1:00 AM, there are 1 to 6 bins, so totally in a day there are (6 * 24) bins)

We are considering the maximum neighborhood clusters to be considered by default to be 5.

We get the previous data points w.r.t to the parameters obtained

1st neighboring cluster of the kth cluster

Pickup counts at 38th timebin on Friday for nearest cluster

There is a linear pattern observed over time, so with the help of linear regression, we could fit a line to give the prediction value to the taxi driver.

Pickup counts at 38th timebin on Friday for nearest cluster

From the above graph, the taxi driver could find that the number of pickups at the requested bin for the 1st nearest cluster could be 600 based on past trends.

Like this, we will obtain the patterns for the remaining 4 clusters, and the taxi driver will decide the cluster he will prefer next.