Will studying the area of NYC during 2020 and the pandemic help to determine how a TLC yellow cab driver can maximize their daily earnings? Since the start of the pandemic, taxi drivers in NYC have not had the same foot traffic as before. Tourists stopped traveling and a majority of local people have been forced to commute less. Between the months of March and June, NYC was hit at its hardest in the pandemic and so were all daily workers. With enough unique data about a taxi's trip I can generalize a model to predict the highest amount of money a taxi driver can make.

The problem I am trying to solve is the one about predicting a maximized hourly earning for yellow cab drivers. My client would be the yellow cab drivers because this model should help reveal which features affect their ability to make the most amount of money on any given day. The data I will be using consists of TLC data from January 2020 to June 2020. This data also contains features like, surcharge, location, pick up date and time, drop off date and time, and additional columns that I have added from studying the data. The data contains types in datetime , continuous, and categorical.

Through the use of Exploratory Data Analysis I came to find that there were certain features from the original data that had a better correlation to the total amount of a trip. These features include: trip_distance, fare_amount, tolls_amount, tip_amount. In addition, I included the pick up areas and the drop off areas as the locations also impacted how much money a trip would make.

The Regression model used was a Random Forest Regressor with a max depth of 9 and n estimators of 5. This helped produce prediction prices for over 1000 entries, which were then grouped together, by hour, to create the chart below.

Average amounts sold by the hour
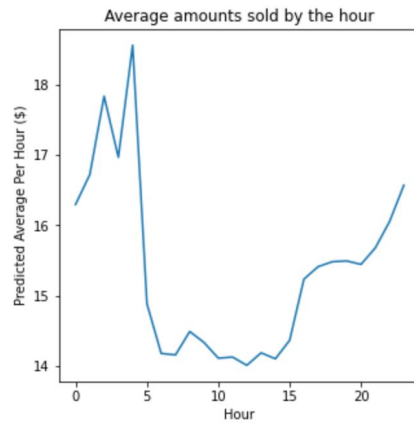
This chart demonstrates which hours makes the most money and which hours make the least.

- Hours between midnight and 5am are the ones where the **most** are made

- Hours after 5am and 3pm are where the **least** ones are made

- After 3pm we can see the amount made begin to increase again.

This can help a taxi driver know that the hours, where most taxi drivers are asleep and when it is rush hour, are the ones where most of the money is being made. It can help to show them that working within those hours means they can increase their hourly pay. The area with the most frequent pick ups also makes a difference in making the most money.
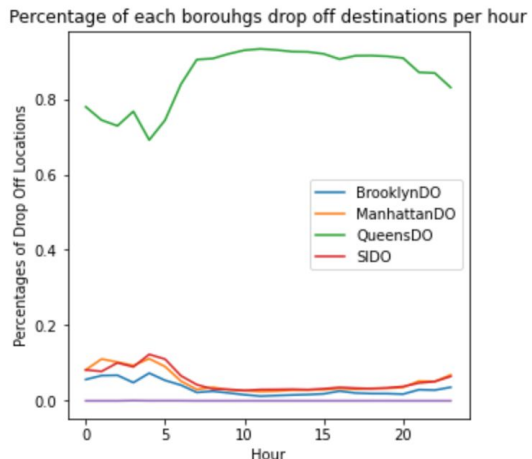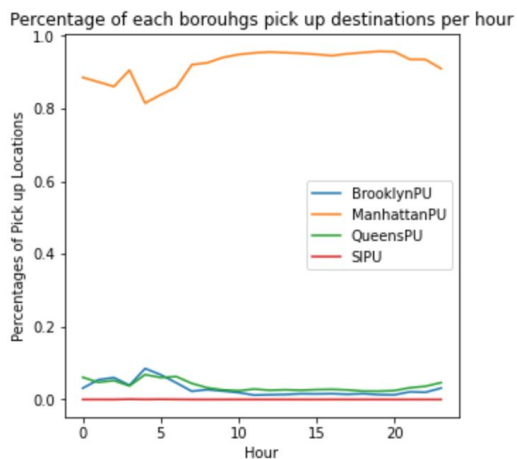
These graphs here show the percentage of each borough's pickups and dropoffs. It states that:

- The area with the **most** pickups are in Manhattan from around 6am to midnight

- The areas with the **least** pickups are Queens, Staten Island and brooklyn
- The area with the **most** drop offs are in Queens at around the same hours as the **most** pickups.
- The areas with the **least** drop offs are Staten Island, Brooklyn, and Manhattan.

This information can be helpful in finding which boroughs provide the most money so that taxi drivers can be in the most profitable destination at all hours.

Recommendations:

1. A yellow cab taxi driver should avoid driving between the hours of midnight and 5am because even though they make the most out of each trip, they are still only getting a very small amount of customers within those hours. They can expect the lowest number of pickups in those hours even when driving in Manhattan, the area with the most number of pickups

2. A yellow cab taxi driver should focus their hours between 6am and before midnight to be able to maximize their hourly earnings. It would mean that they would receive a lot more trips in that hour that will ultimately add up to more than working during the hours that make the most amount in a single trip. In addition, choosing drop off locations also matters so dropping off people in Queens would contribute to maximizing their hourly rate.