Forecasting Demand for Taxi and Rideshare Services in NYC

Lucas Fern
Student ID: 1080613

August 15, 2021

1 Introduction

As a result of technological innovation and the profitability of minimising employee bargaining power, the gig economy has exploded in market share over the last decade. In 2018, 24% of US citizens aged 18+ earned some income from gig work, with a majority of this work performed in the ridesharing industry with companies such as Uber and Lyft. [1]

This report will assume the perspective of a taxi or rideshare company with the objective of forecasting demand for pick-ups and drop-offs in various locations around New York City (NYC). A Deep Neural Network with architecture as shown in Figure 1 is used as such a model was expected to be able to identify complex relationships between the input features.

In this report the term 'For Hire Vehicles' (FHVs) will be used to refer collectively to all ridesharing services (including High Volume For Hire Vehicles). Taxi and ridesharing services in the city will also be collectively referred to as Mobility Services.

2 Dataset

This report primarily uses data published by the NYC Taxi & Limousine Commission (NYCTLC) [2] which records statistics about each taxi and FHV trip taken within the city. The relevant subset of features extracted from this data for the analysis is outlined in Section 2.2. The NYCTLC also provides a file describing the shape of each of the locations around the city, which is used for visualisation purposes.

In addition to this taxi and ridesharing data, hourly weather reports taken at JFK Airport (in the South-East of NYC) will be used as these are expected to be highly correlated with demand for taxi services. Weather data is made publicly available courtesy of the US National Center for Environmental Information's Integrated Surface dataset. [3] From this dataset multiple features including the temperature, wind speed and direction, and atmospheric pressure will be used to provide the model with a detailed representation of the hourly weather conditions.

2.1 Range Selection

This report assumes the perspective of a ridesharing company with the objective of generating accurate forecasts of demand for their service. As a result it is essential that the features selected as predictors for the model are available for FHV trips. This is a significant constraint, as until early 2017, Uber - who provide a significant proportion of FHV trips - were legally challenging the requirement to release trip data to the Taxi and Limousine Commission. [4] Because of this the data before 2017 for FHV

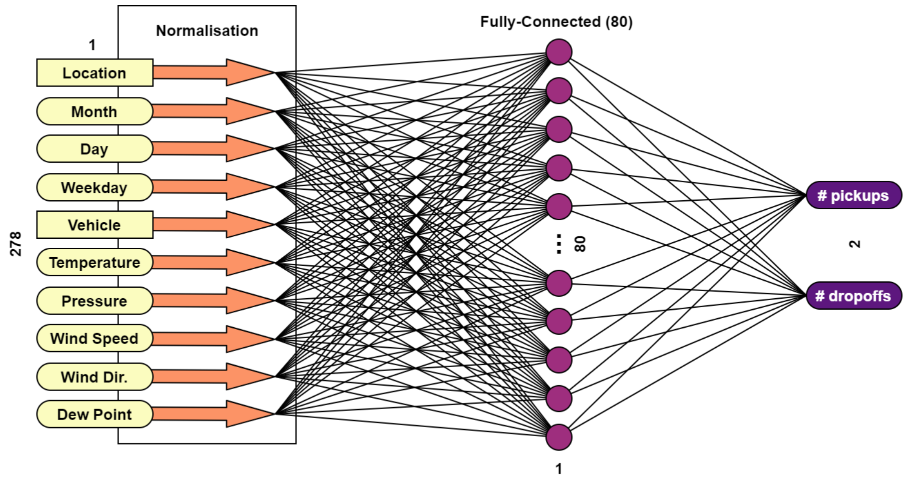


Figure 1: The Architecture of the Neural Network Model, Including Inputs and Outputs.

trips is limited and unreliable and it was therefore decided to use data from January 2017 onwards.

In addition to this, New York City went into lockdown due to the COVID-19 pandemic in March of 2020. It is clear from Figure 2 that the demand for mobility services decreased in the leadup to the lockdown as the public sentiment shift in favour of staying home. To avoid any impact of fluctuating demand due to the pandemic, data is therefore only used up until the end of December 2019. This 3 year range results in a dataset of over a billion (1,057,052,914) unique taxi and rideshare trips.

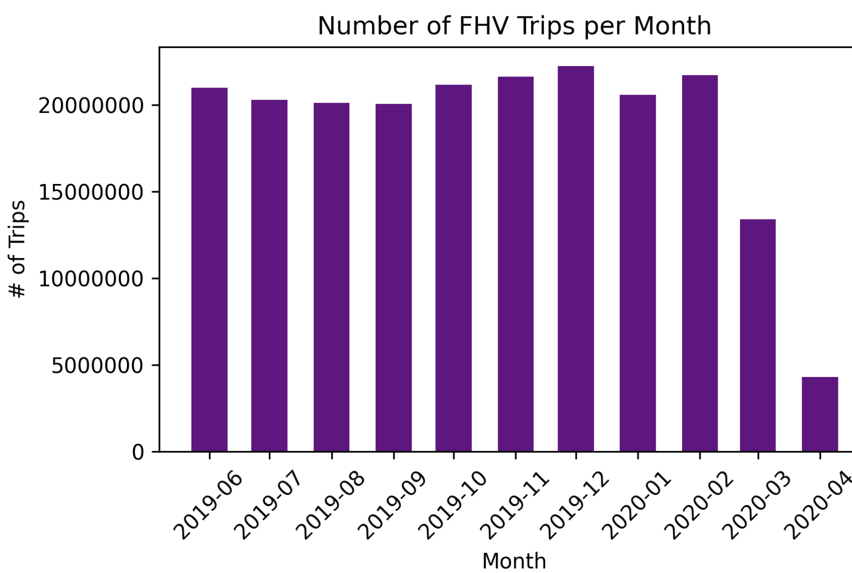


Figure 2: The number of FHV trips taken between June 2019 and April 2020.

2.2 Pre-processing

Though the format of both datasets used is fairly well defined there exist outliers in the data which do not follow the expected structure. This section will outline where these inconsistencies are found in the data and what steps were taken to eliminate them.

In addition to this, the format of taxi and rideshare data published by the NYTLC are not consistent with each other. The strategy used to minimise the effect of this is discussed in Section 2.2.1.

2.2.1 Feature Selection

When training a neural network, it is desirable to have observations for all features of all instances. Having such a complete dataset allows the model to learn relationships between the features and output with a greater degree of confidence. It is less important to remove insignificant features as these will be appropriately weighted in the final model.

In order to ensure that all instances in the training dataset had observations for a high proportion of the features, it was necessary to discard the features which were not available in the FHV data. This included features such as the itemised payment breakdown, number of passengers, and trip distance, which *were* available for the traditional taxi trips.

Many features were discarded from the weather data due to either being too specific to the location at which the weather observations were made (JFK Airport), or clearly having no legitimate causal relationship with the demand for taxi services (*i.e. Wave Height*).

Ultimately, the following features were retained from the taxi data:

- Date and Time,
- Day of Week,
- Pick-up Location ID,
- Drop-off Location ID,
- Vehicle Type.

And from the weather data:

- Date and Time,
- Wind Direction,
- Wind Speed,
- Temperature,
- Dew Point,
- Atmospheric Pressure.

Note that the day of week was extracted from the date as this is expected to be highly correlated with the demand for mobility services. The year is also discarded in the final dataset so the model will not be constrained to making predictions for the years present in the training data.

2.2.2 Outlier Detection

Various techniques were used to detect and remove outliers from both datasets. This section will summarise the types of outliers which were detected and how the offending instances were managed.

Trips with dates outside the defined range were included in some of the mobility service datasets. There were less than 100 of these in over a billion instances and all were eliminated.

Weather reports outside the specified date range were detected, but no action was taken as these will be removed when performing an inner join with the taxi data.

Trips which spanned more than 5 hours were considered to be outliers since Google Maps estimates it would take only 3 hours to drive the entire perimeter of NYC, therefore it is reasonable to expect any round trip within the city to be completed within this time. In total 1,374,880 instances were removed for this reason.

Trips with a negative duration exist when the drop-off time precedes the pick-up time. 19,379 of these were identified and removed.

Trips with a pick-up/drop-off location ID out of the range 1-263 represented outliers as they either started or ended at an undefined location. 68,706,340 ($\approx$ 6.499% of) trips were removed for this reason, accounting for the majority of outliers detected. It is possible that some or many of these trips started or ended outside of the city, but since this behaviour was not defined in the usage guide these were removed.

2.2.3 Imputation

Imputation was not required after the feature selection and outlier detection steps. There were very few hours during the years for which weather reports were unavailable, and considering the size of the dataset it was considered reasonable to discard the records which could not be matched to an hourly weather report.

No taxi records were missing pick-up or drop-off times after filtering outliers, and it was unnecessary to impute missing locations due to the aggregation steps outlined in the following section.

2.3 Aggregation

This report aims to model the demand for each type of taxi and rideshare pick-up and drop-off in each of the locations of NYC in a given hour of a year. Because of this, it is unnecessary to retain information about each individual trip, and instead the model will be trained using aggregated statistics.

For the model to be able to predict the number of pick-ups and drop-offs by a given vehicle type, in a given location, at a given hour, it is sufficient to train the model using instances which store each of these features, and are labelled with the total number of pick-ups and drop-offs in said location during the hour. For clarification about the inputs and outputs of the model refer to Figure 1.

The aggregation step was performed by grouping on all of the features simultaneously, and counting the number of pick-ups and drop-offs in each group. When performing aggregation the granularity of the time attribute was decreased from the nearest minute to the nearest hour.

3 Preliminary Analysis

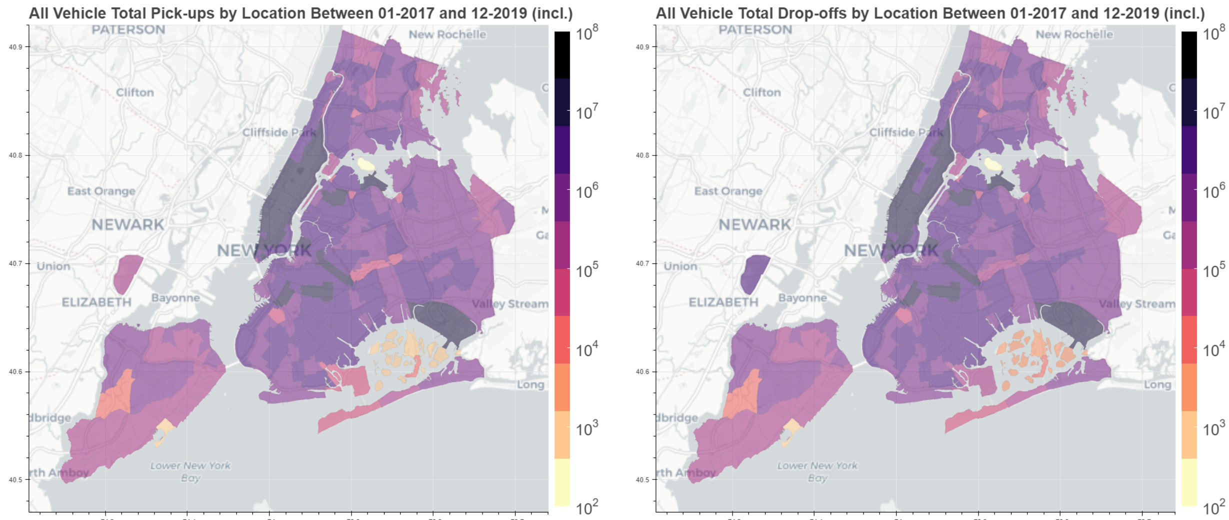
This section will investigate the relationships which exist between individual features and the hourly amount of trips taken. For some visualisations samples of the dataset were used. In these cases it was verified that the samples were representative of the entire dataset. No additional refinement or feature selection was required on the dataset as a result of this analysis.

3.1 Location ID

This feature is most clearly correlated with the amount of trips taken per hour. Figure 3 shows beyond any doubt that the amount of trips taken to and from a suburb depend heavily on it's location. This relationship is so extreme that suburbs such as those in central Manhattan and at the airports have more than 10,000 times the pick-ups of the least travelled locations. Locations with a large amount of pick-ups also consistently have a proportionally large amount of drop-offs, which is unsurprising.

3.2 Weekday

The day of week feature was generated from the date of each trip due to an expectation of high correlation with demand for mobility services. Here we will look at the significance of the weekday attribute for predicting mobility service demand at JFK airport, though the attribute is similarly significant when predicting demand in other locations.



(a) Pick-ups (b) Drop-offs

Figure 3: Total Trips by all Vehicle Types by Location Between 01-2017 and 12-2019 (incl.)

Figure 4 shows the relationship between the day of the week and the amount of trips taken to and from JFK Airport, one of the locations in the dataset. It is intuitively and visually evident that the demand for travel to the airport is highly dependent on the day of the week, and the ANOVA shown in Table 1 demonstrates this relationship empirically with a p-value for the weekday attribute of 1.29×10^{-58}.

Table 1: ANOVA for Weekday as a Predictor for Mobility Service pick-ups at JFK Airport.

	SS	DF	F	$\Pr(> F)$
Weekday	2.40e+09	6	54.2	1.29e–58
Residual	8.07e+09	1088		

3.3 Weather

Weather data was joined with the mobility records since certain weather patterns were expected to cause fluctuations in demand for taxi and rideshare services. In this section the effect of weather records will be analysed independently of other features, since features like the month and time of day are expected to be highly correlated with the weather. Note however that this correlation will not cause issues in the final neural network model since independence of features is not assumed.

Table 2: ANOVA for Weather Predictors in the Presence of Confounding Factor Location.

	SS	DF	F	$\Pr(> F)$
Location ID	2.09e+12	256	11038.45	0
Wind Speed	7.37e+08	1	994.35	7.44e-218
Temperature	6.15e+08	1	829.47	3.94e-182
Dew Point	4.61e+08	1	621.14	6.00e-137
Pressure	1.37e+08	1	185.15	3.74e-42
Residual	2.03e+11	274233		

A linear model was determined to be appropriate for predicting the number of pick-ups per location based on the weather conditions. The confounding factor of location was introduced into the model and Table 2 shows that each of the weather attributes are statistically significant in it's presence. It was unexpected that Dew Point and Pressure would have such a strong relationship, but these attributes have a causal relationship with rainfall, which has a clear link to demand.

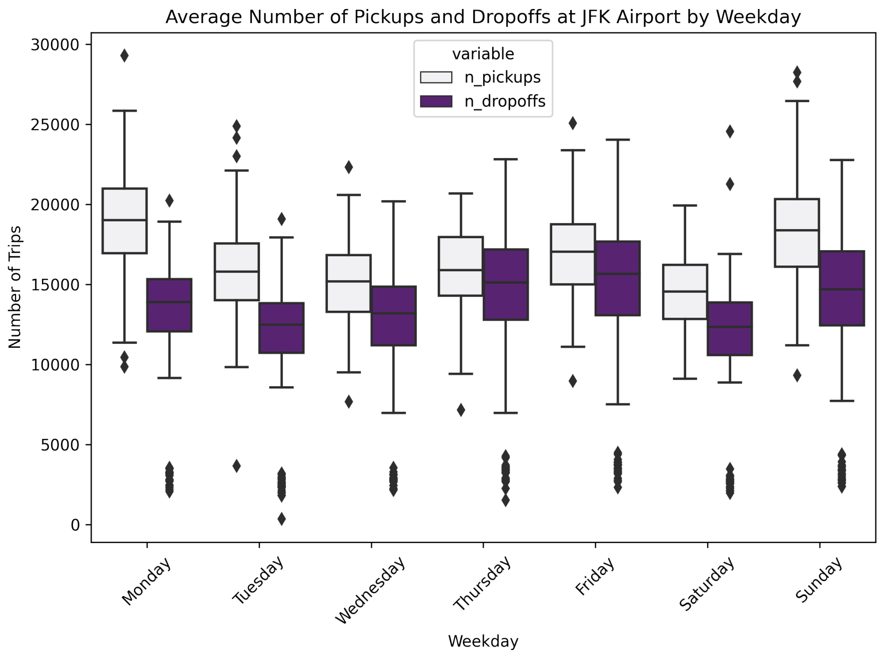


Figure 4: Average Number of pick-ups and Drop-offs at JFK Airport by Weekday.

4 Neural Network Model

The neural network takes a 1×178 tensor as input, and produces a continuous regression output for the two targets, the number of pick-ups and drop-offs in the given location during the provided hour of the year. The exact inputs to the model are shown in Figure 1, where the vehicle type and location ID are one-hot encoded due to being categorical attributes.

Considering the tabular format of the data a shallow network architecture performed better in testing. Ultimately, a single, fully connected hidden layer with 80 neurons was used in the final model. Mean Absolute Error (MAE) was used as a loss metric since it is highly interpretable, and minimising the sum of the prediction error of the two outputs is desirable.

The final model was trained on 8.3 Million instances in batches of 32, and validation and test sets were used to refine and test the model each containing 1.8 Million instances. Training on a this sizeable dataset took 10 epochs to converge and ultimately resulted in a MAE of 19.70 on the test set. This indicates that predictions for the amount of pick-ups and drop-offs from a given location were on average accurate to within 20 trips.

4.1 Prediction and Error Analysis

The Mean Absolute Error is not a terribly useful metric for assessing the quality of the model in isolation. This section will contextualise this measurement of error and perform sample predictions to evaluate the usefulness of the model from a business perspective.

Recall from Section 1 that this report assumes the perspective of a mobility service provider wanting to predict the demand for their service over time. The efficacy of the model at making such predictions will now be demonstrated with a randomly selected time from the test set.

The model was used to predict the demand for pick-ups and drop-offs in a single location at all times throughout a randomly selected day. The result is shown in Figure 5 for the hours which had ground truth labels. From this Figure it appears that the model overestimates the demand for all services slightly in early hours of the morning and underestimates the amount of drop-offs in morning peak hour. In the evening the model predicts the difference between pick-up and drop-off counts with relatively high accuracy. Overall it is entirely plausible that such a model could be used effectively when pathing drivers around NYC to meet fluctuating hourly demand.

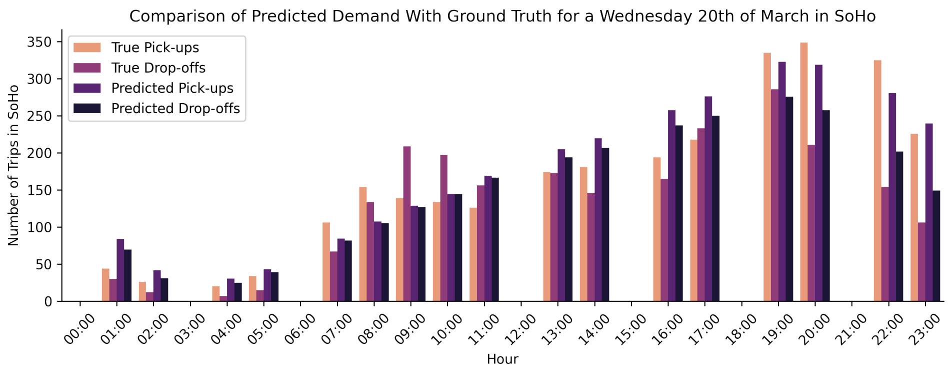
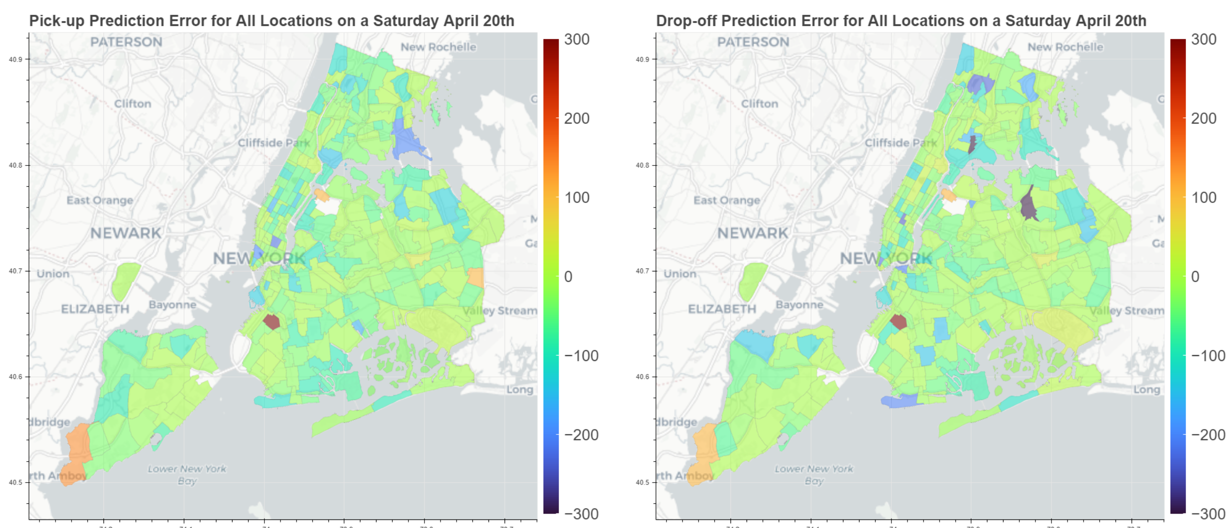


Figure 5: Comparison of Predicted and True Trip Count in SoHo by Hour on Wednesday March 20th.

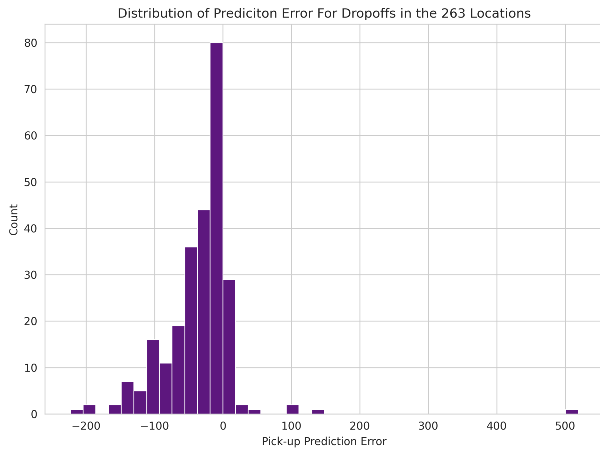
Considering a broader perspective, Figure 6 shows the error in prediction accuracy in each location at a different randomly selected time. Figure 7 shows how these errors are distributed, which makes it clear that while the majority of predictions are centred around the ground truth, there appears to be a systematic bias towards underestimating the demand for both pick-ups and drop-offs. In this example there are individual locations which demonstrate extreme over and underestimates. Notably the location which is significantly overestimated in both cases is a nature reserve in Greenwood Heights which does not have publicly accessible roads. This indicates that the model is not sensitive enough to this specific location attribute and therefore more training data would improve prediction accuracy.



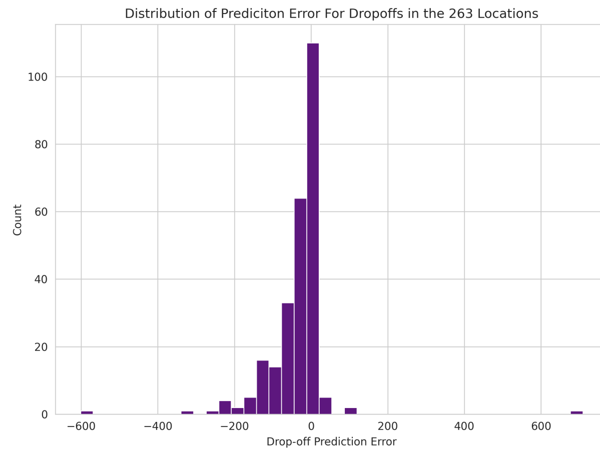
(a) Pick-ups

(b) Drop-offs

Figure 6: Difference Between Predicted and True Trip Count by Location on a Saturday April 20th.



(a) Pick-ups



(b) Drop-offs

Figure 7: Distribution of Prediction Error in Figure 6.

5 Discussion

Considering this fairly simple model is able to predict demand for locations in the city with an average error of less than 20 trips per hour it is clear that a neural network can be effective at forecasting demand for mobility services. This model is surprisingly accurate considering that the distribution of hourly rides per location is centred at 53.4 with a relatively high standard deviation of 109.5.

This report would suggest that businesses in the industry should look at producing such a model of their own. This model could then be used by the business to determine where passengers were likely to request rides, and this information could be used to determine incentives for drivers to encourage them to travel to locations expecting surges in demand. This has the potential to result in reduced idle time and consequentially increased profitability.

The investigation in this report is limited by the amount of features provided in data by For Hire Vehicle companies, therefore it is suggested that such an exercise would be especially useful for their business model considering the immense amount of data which is collected but not published about each trip. Training on such data would enable much more significant and accurate predictions, and considering external predictors beyond the weather has the potential to make such a model extremely effective. It is recommended that future versions of such a model would consider features such as ongoing events and road conditions within the city.

It is highly recommended that future investigations go to additional effort to clean outliers from the data. It is suspected that a small amount of extreme observations such as those which caused overestimates in Figures 6 and 7 are also responsible for the systematic bias towards underestimation, as the underestimation compensates for these extreme overestimates on average. It is surprising that the bias is so evident considering the normalisation of each attribute before training, but this appears to be the most likely cause to investigate.

In addition to investigating modifications to the neural network model presented here it is advisable that any business wishing to implement such a system investigate more interpretable models such as the Generalised Linear Model, as these provide significantly greater potential to evaluate the influence of individual features and their interaction. Said models may be able to produce comparably accurate results considering the lack of complex interaction between the relatively small amount of features in the dataset.

References

[1] "The gig economy," *Edison Research*, p. 2, December 2018.

[2] New York City Taxi and Limousine Commission, "TLC trip record data." https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page, 2017-2019. [Online; accessed 1-August-2021].

[3] National Centers for Environmental Information, "Integrated surface dataset." https://www.ncei.noaa.gov/access/search/data-search/global-hourly, 2017-2019. [Online; accessed 31-July-2021].

[4] A. Marshall, "NYC forces uber and lyft to provide more data on passenger pickups and drop-offs." https://www.wired.com/2017/02/ubers-coughing-data-nyc-fix-commute/, Feb 2017.