Tipping habits for New York City Taxis

John Melleuish
Student ID: 1171367
Github repository

August 20, 2022

1 Introduction

Tipping in the United States is often seen as a national passion especially in the restaurant industry. The suggestion of not tipping your server would spark disagreements and judgemental looks amongst the party your with. But is it the same with taxi drivers? Do people see tipping as compulsory? and if so what external factors influence someone level of generosity?

This report will investigate from a taxi drivers perspective what kind of environment can they put themselves in, in order to maximise their chances of receiving a gratuity.

Two datasets were used in this analysis; NYC yellow and green taxis dataset (184397591 rows x 19 columns) as well as weather data from LaGaurdia Airport (28361 rows $\times$ 102 columns). Both Yellow and Green NYC taxis were chosen for this investigation due to the relevant datasets containing almost identical variables captured including tip amount. Ride Share data (Uber, Lyft, private limousines etc.) was available, however they did not make the amount tipped per ride available and thus couldn't be use.

The timeline of analysis was taken from January 2018 through to December 2019. This period was chosen as it is the most recent data not effected by the COVID-19 Pandemic. 2020 on wards despite being newer data, has a whole range of factors that effect the reliability of the conclusions that will be drawn from it. As we move out of the pandemic, taxi drivers will find data without a pandemic asterisk far more valuable as people decisions around tipping would've no doubt changed.

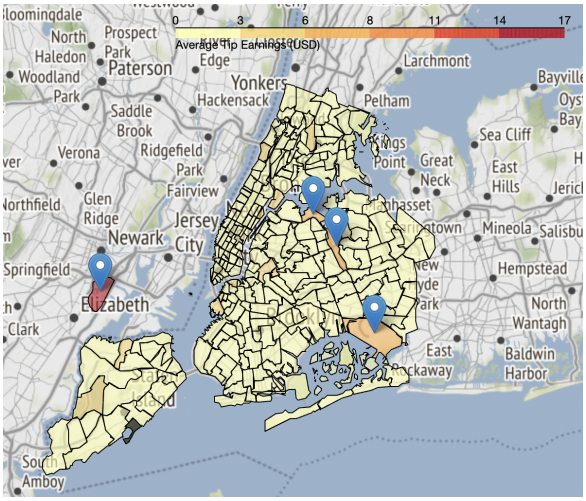


Figure 1: Distribution of Tips across NYC

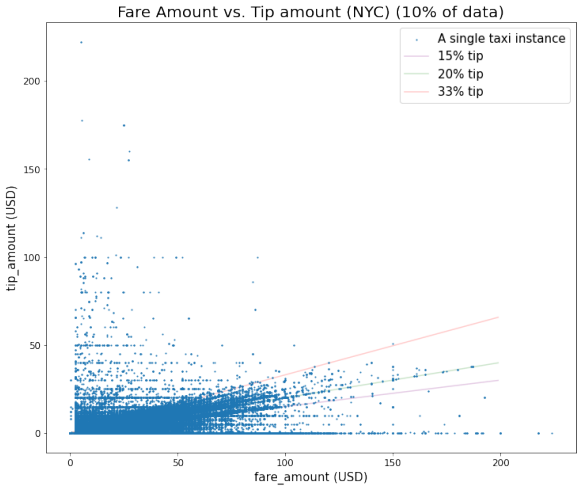


Figure 2: Fare Amount vs. Tips

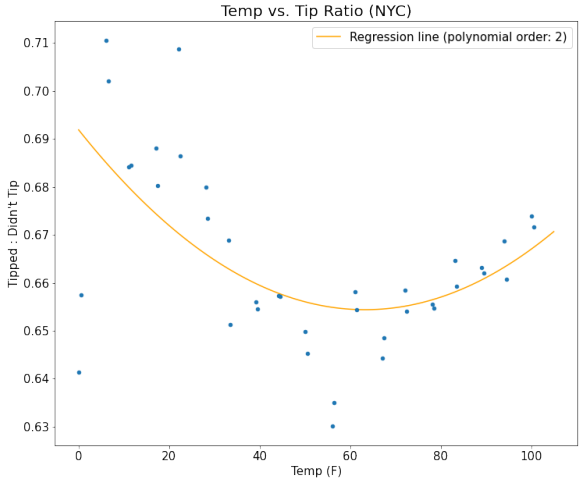


Figure 3: Temp vs. Tips Ratio

2 Preprocessing

2.1 Taxi Dataset

2.2 Weather Dataset

3 Analysis

3.1 Geospatial Visualisation

3.2 Fare Amount

3.3 Temperature

3.4 Time of Day

3.5 Trip Distance (optional)

4 Modelling

4.1 Neural Network

[h]

4.2 Linear Regression

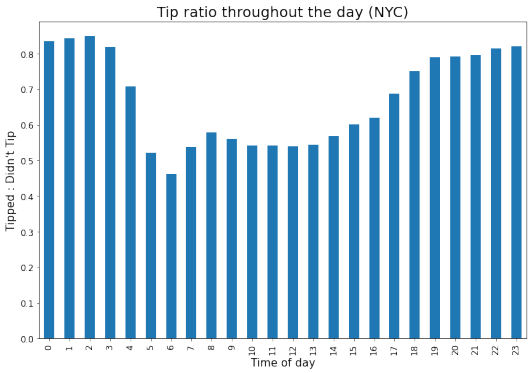
[h]

4.3 Ridge Regression

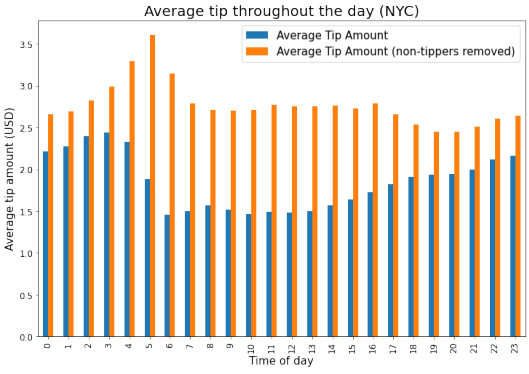
[h]

5 Recommendations

6 Conclusion



[Tip Ratio]



[Average Tip Amount]

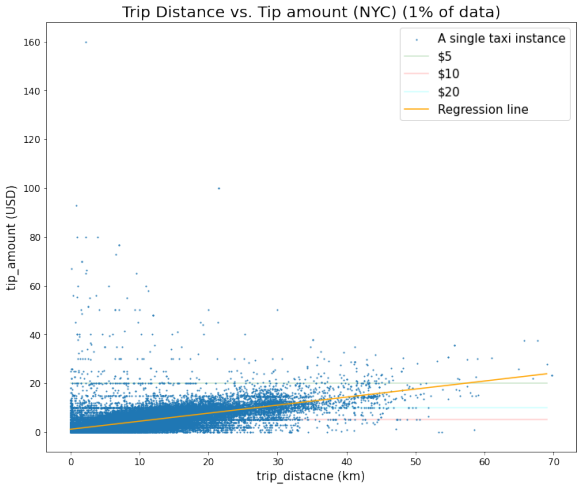


Figure 4: Trip Distance vs. Tips

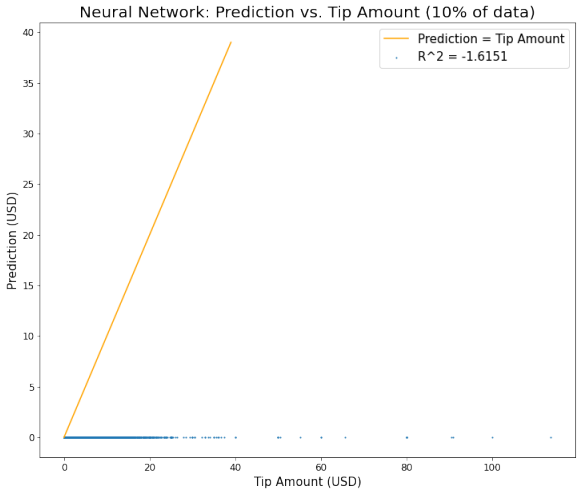


Figure 5: Neural Network

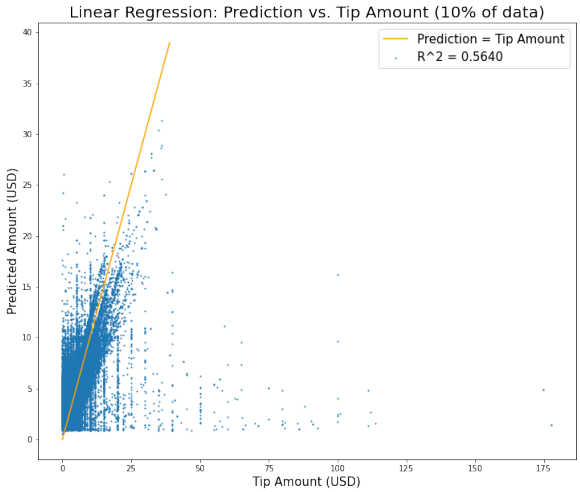


Figure 6: Linear Regression

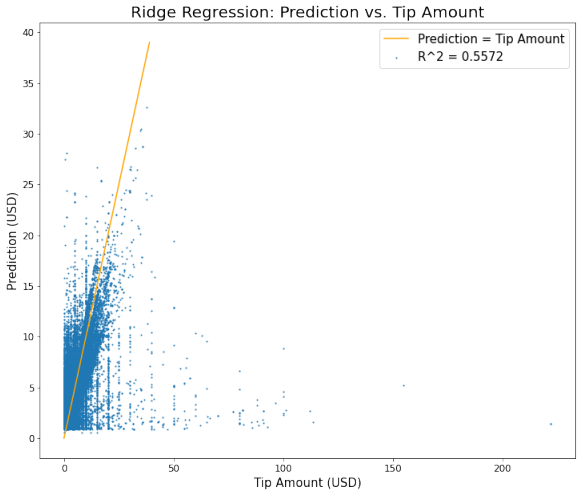


Figure 7: Ridge Regression