

### Lab3 – Hagay Ringel

The following study concerns the field of taxis in New York City. The reason why I chose this topic is because first of all it is an interesting topic: Who among us does not want to know what the price for a taxi consists of? How much distance does the tip amount affect? Does the number of passengers matter on the price and if so, by how much? All of these questions interest me personally but I think it is relevant to every New York resident who uses taxis no matter how often - as we have all noticed the rise in taxi prices (yellow taxi, Uber, Lift, etc.). In addition, I think the information from this work will also be useful for the taxi drivers as maybe it will reflect to them the data on other taxi drivers and they will see data that may improve their services and even their profitability.

At this research I checked the correlation between the total amounts that was paid to the taxi driver to factors such as drop off/pick up locations, distance, fare, tip and etc. As expected, there is a high level of correlation since the R Squared is 0.99.

In order to check the accuracy of this model I calculated the Mean Absolute Error (MAE), which is 0.6. It means that there is a chance of 60% the model and the prediction is wrong.

The result for the Mean Squared Error (MSE) is 0.72 which is not that good as the data points are far from the fitted line which means that the accuracy is low.

Then, I calculated the Root Mean Squared Error (RMSE) which is necessary to find out what is the standard deviation of our prediction. The result I got is 0.85 which is pretty high (I expected for 0.2-0.5) and means that there may be a big distance between the predictions to the real result in case of mistake in the model.

To sum it up, despite of the last numbers that show a high standard deviation, moderate level of accuracy and a great distance between the points of the data, I believe that the high level of correlation (99%) pretty much ensures that the model is correct and accurate.