# Automatida

Taxi fares estimation project

**Problem to solve:**

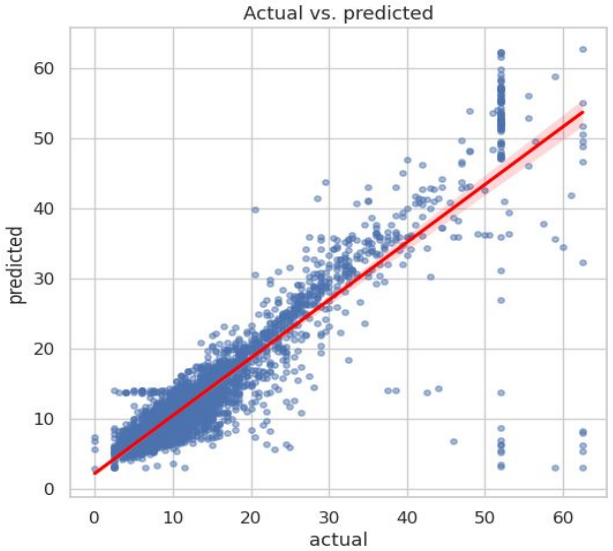
The New York City Taxi & Limousine Commission contracted Automatidata to predict taxi cab fares.

## Response:

I chose to create a multiple linear regression (MLR) model based on the type and distribution of data provided. The MLR model showed a successful model that estimates taxi cab fares prior to the ride.

## Impact:

The model performance is high on both training and test sets, suggesting that the model is not over-biased and that the model is not overfit. Model provides framework for predicting the estimated fare amount for taxi rides.




The scatter plot shows a linear regression model plot illustrating predicted and actual fare amount for taxi cab rides.

**Model metrics:**

- Net model tuning resulted in:
  - ✓ R^2 0.83, meaning that 83% of the variance is described by the model.
  - ✓ MAE 2.12
  - ✓ MSE: 17.52
  - ✓ RMSE 4.18

## Insights/next steps:

- The model provides a generally strong and reliable fare prediction that can be used in downstream modeling efforts.
- The feature with the greatest impact on fare amount was mean distance between pickup and drop-off point. The model reveals a mean increase of $2 for each additional mile, however this result can’t be 100% trus due to high correlation between some features.
- Request additional data for further exploration