

New York City TLC Preliminary Data Executive Summary

Commission Prepared by Automatidata

Project Overview

The NYC Taxi & Limousine Commission has consulted with Automatidata to build a regression model to predict taxi cab fares. The Automatidata team performed preliminary data analysis to find the best variables for model training and to ensure the data was clean.

Details

Key Insights

There are inaccurate values for trip distance and the total amount fields.

51810714	33.96
40523668	33.92
49894023	32.72
76319330	31.95
94052446	30.83
...	
63574825	0.00
47368116	0.00
91619825	0.00
78660848	0.00
58079289	0.00

Name: trip_distance, Length: 22699, dtype: float64

- Found unusual values
- Determined total_amount and trip_distance to be most impactful fields for model training
- Showed statistics of variables and relationships between them

29059760	-120.30
14668209	-5.80
24690146	-5.80
58395501	-5.30
109276092	-5.30
91187947	-4.80
75926915	-4.80
55302347	-4.30
28459983	-4.30
97329905	-4.30
43859760	-3.80
105454287	-3.80
833948	-3.80
57337183	-3.30
26005024	0.00
108016954	0.00
49670364	0.00
19022898	0.00
58713019	0.30
37675840	0.31

Name: total_amount, dtype: float64

Next Steps

- Begin exploratory data analysis.
- Begin data cleaning.
- Explore the data further with statistics.
- Build a regression model.