

Automatidata: data analysis

Cleaning and sorting data by using python

The main goal of this project is to prepare, clean and understand provided data. First step, after importing the data set, was to analyze data structure, find variables including outliers and means. The next step was to analyze the results. There is a clear connection between several variables, like between travel distance and amount paid.

Key Insights

There are some strange outliers, for example trip distance with negative values or extremely large trip distance without expected payment.

Details

trip_distance	fare_amount
2.60	999.99
0.00	450.00
33.92	200.01
0.00	175.00
0.00	200.00
32.72	107.00
25.50	140.00
7.30	152.00
0.00	120.00
33.96	150.00

Next Steps

Conducting complete exemplar analysis of provided data is recommended. Another thing is identify reasoning behind strange outliers.