# Team Members

|  |  |
| --- | --- |
| 22010323 | أمنية محمد عبد الهادي أحمد زايد |
| 22011609 | نور الدين سامر عبد الرزاق حسن |
| 22011606 | محمد عماد مبروك محمد |
| 22011547 | سارة شريف حسن محمد |
| 22011610 | مازن أحمد محمد أحمد محمد |
| 22011589 | محمد مصطفى محمد عبد المجيد |
| 22011611 | مهند حسام إبراهيم خليل |
| 22011612 | حسام الدين رمضان عبد اللطيف محروس |
| 22010030 | أحمد محمد أحمد |
| 22010028 | أحمد فرج |
| 22010343 | زياد محمد إبراهيم شلبي |

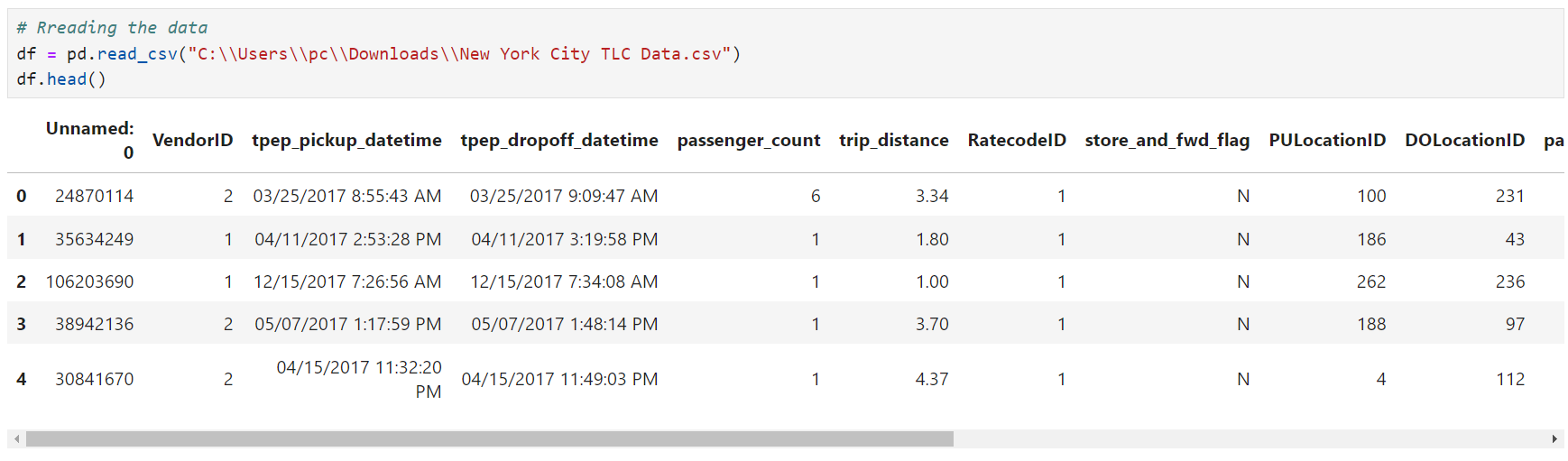
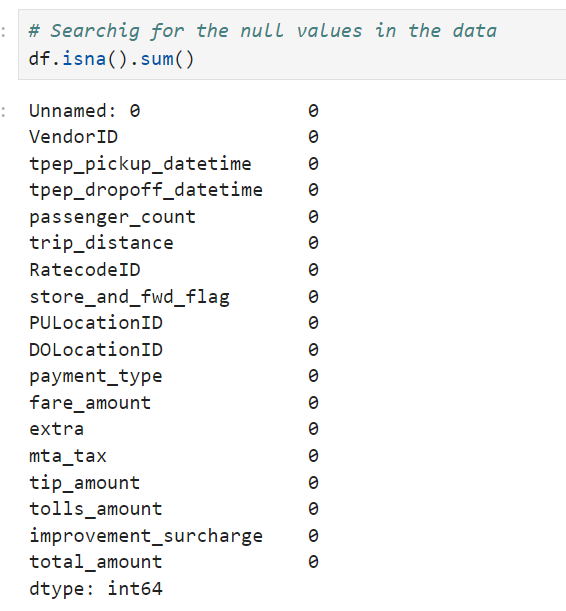
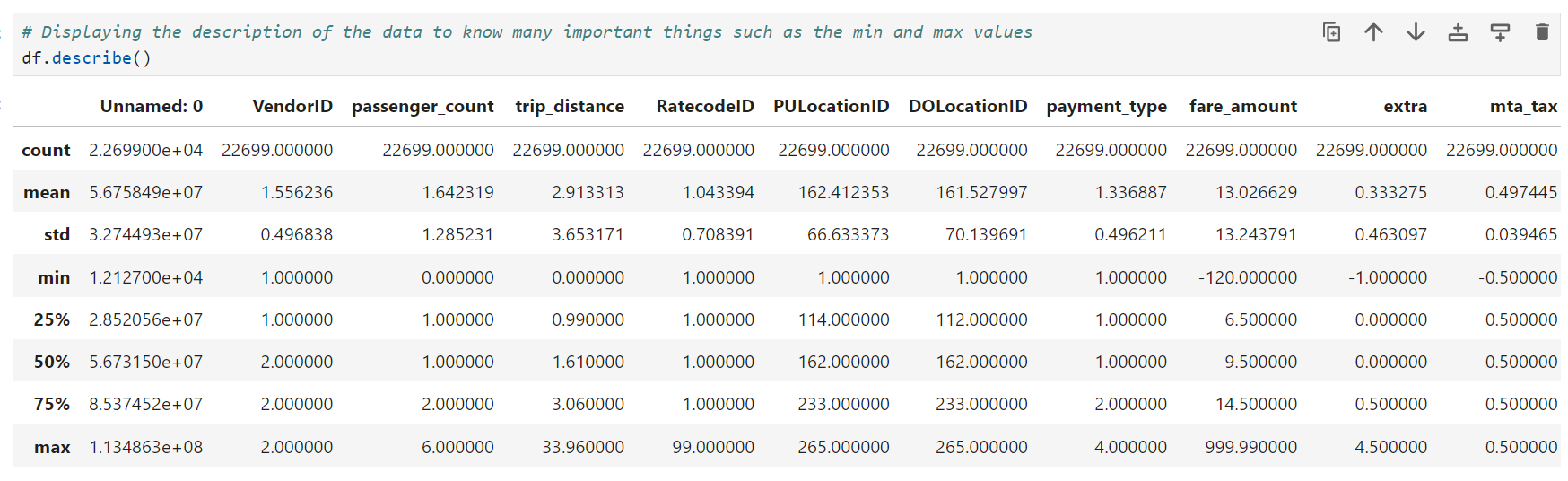
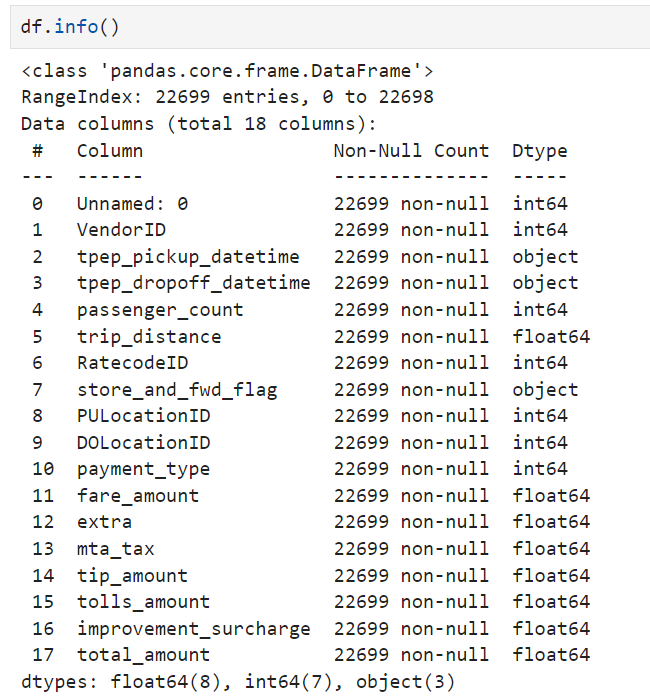
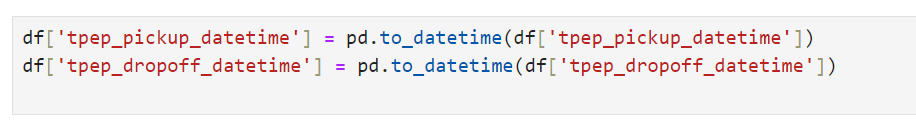
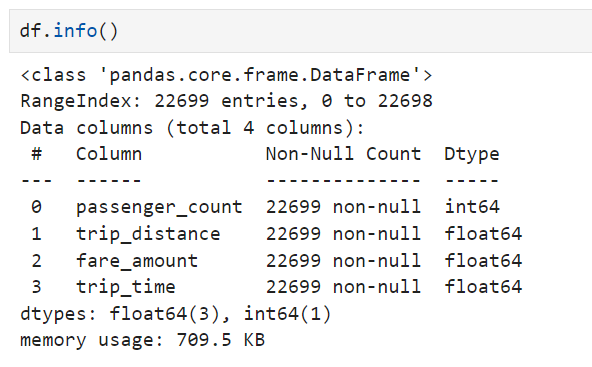
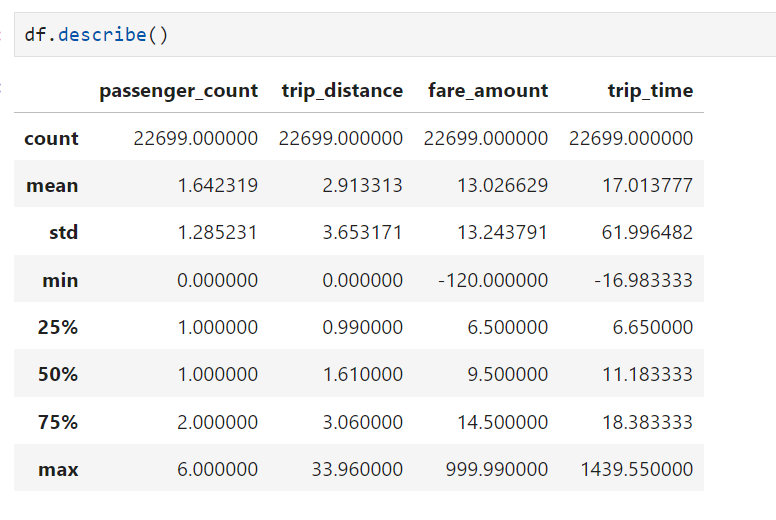
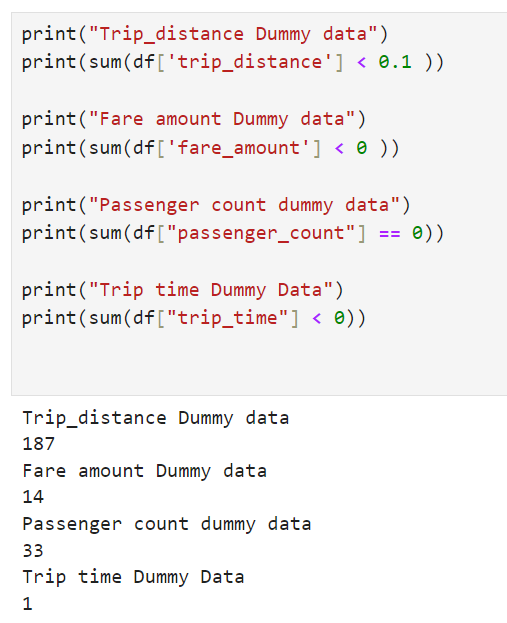
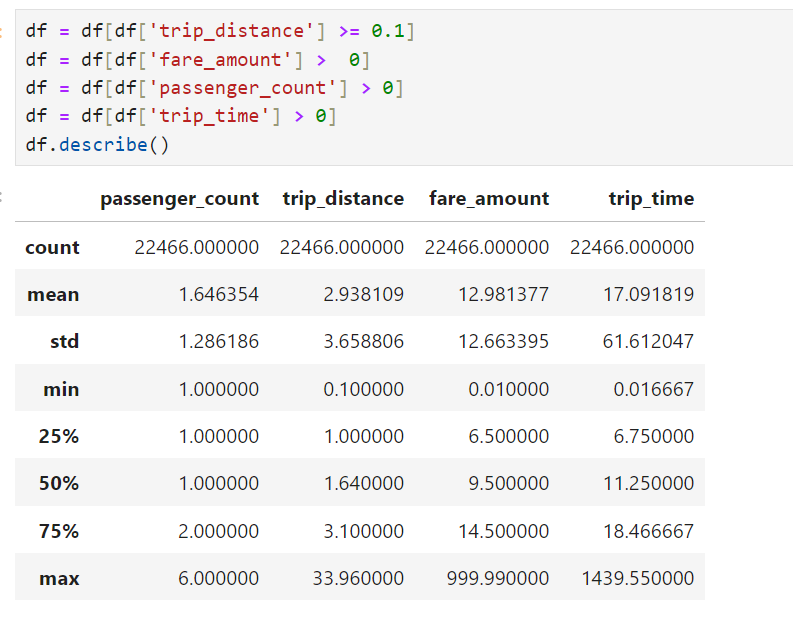
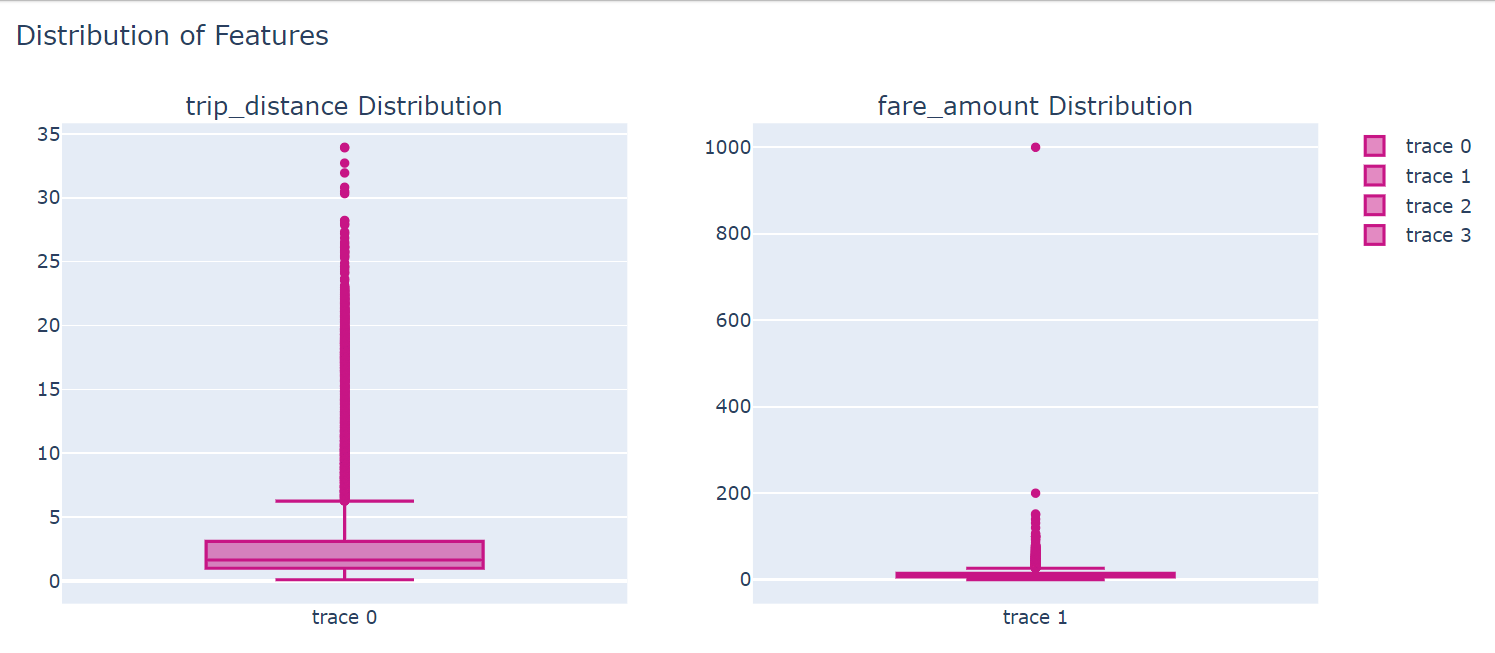
Overview

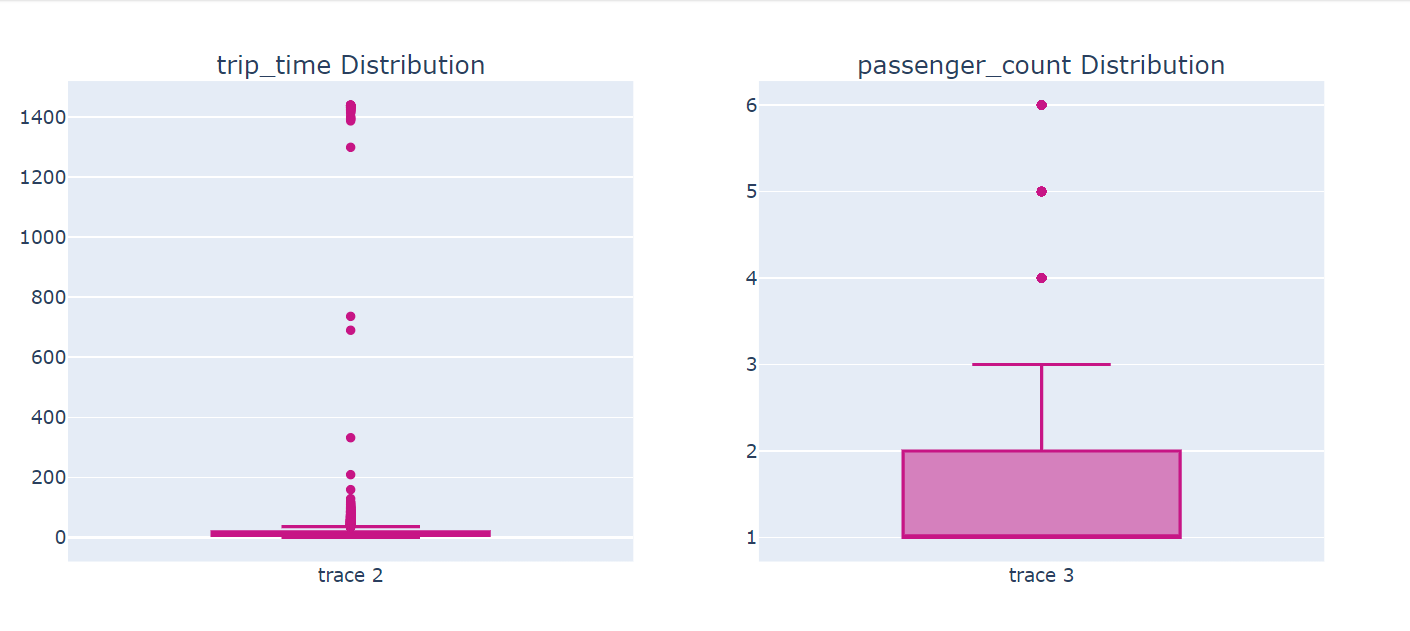
This dataset contains information about the taxi and limousine rides in NYC.

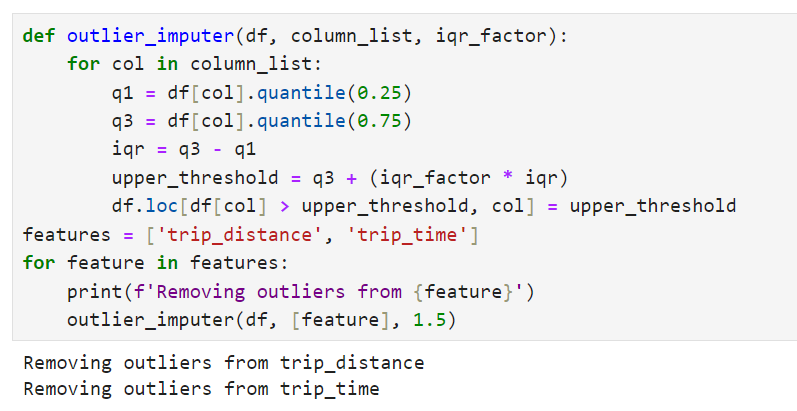
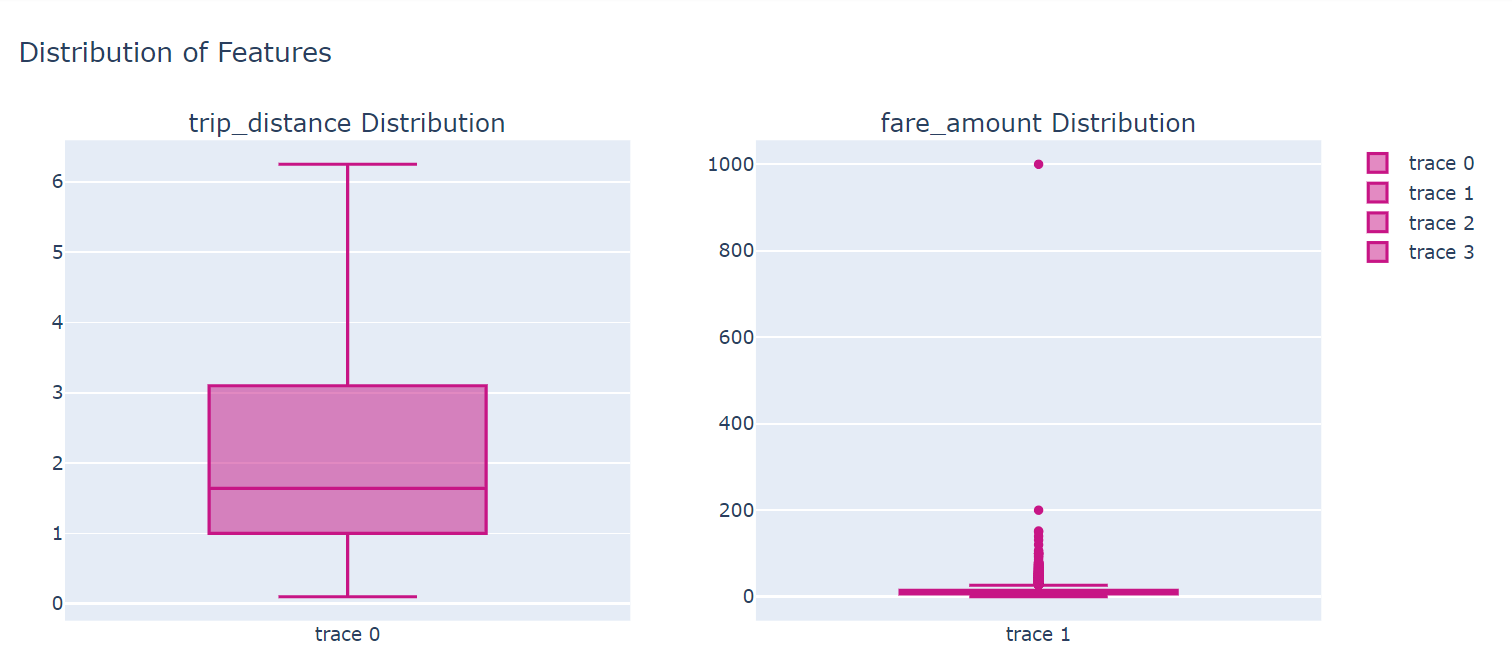
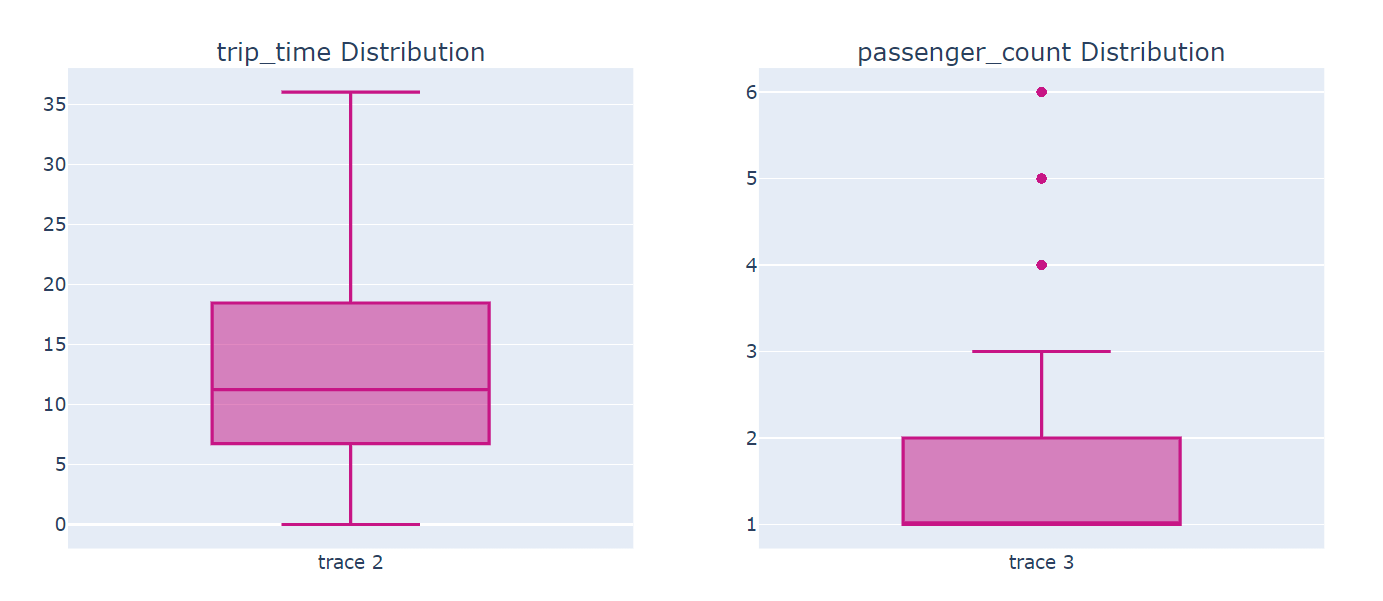
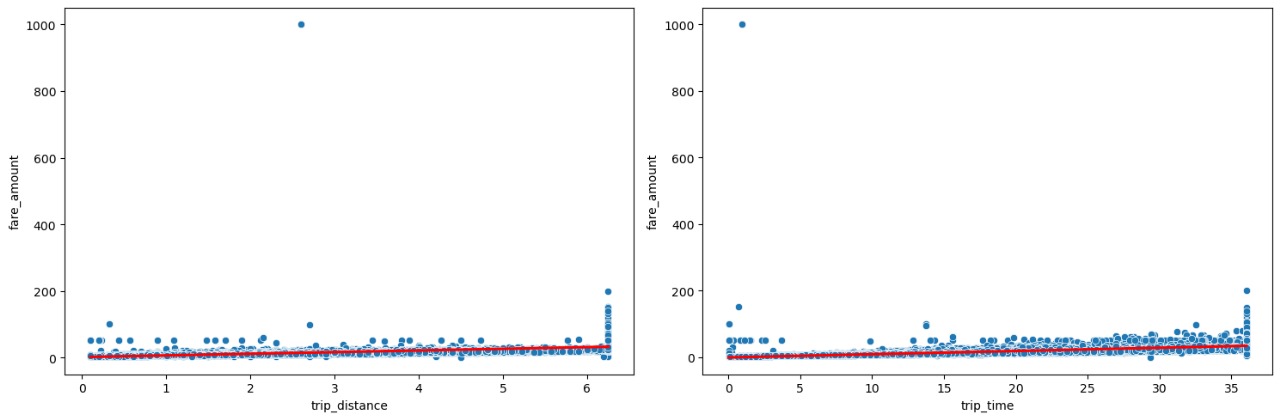
Objective:

Our goal is to predict the fair price of the trip based on the attribute(s) that contribute the most in the fair price of the ride.

And we will explain the process through a series of steps.

* Importing and displaying the data  
  
* Searching for null values  
  
* Displaying summary statistics  
  
* Displaying data frame info  
  
* Converting date columns into datetime types to use them later  
  
* Adding a new feature by calculating the trip time  
  
* Reducing the data frame so that it only contains necessary columns  
  
* Displaying info about our features  
  
* Displaying summary statistics of the features  
  
* Counting dummy data  
  
* Dropping dummy data and displaying the new summary statistics  
  
* Visualizing the distribution of our features to detect outliers  
  



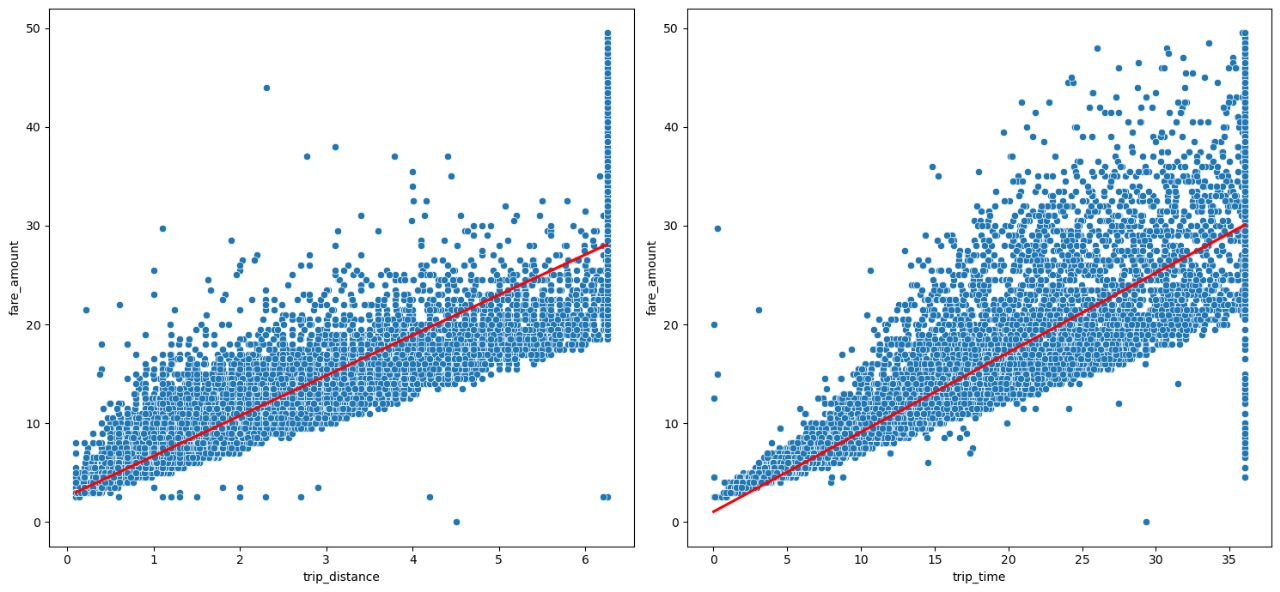
* Getting rid of the outliers by setting them equal to the maximum value of each column  
  
* Displaying the distribution after removing the outliers  
    
  
* Plotting fair price (dependent variable) against trip distance and trip time (independent variables)  
  
* We can sort of see the linear relationship here, but we’ll do the test first.
* Testing and comparing the actual data against the predicted values  
    
  A graph showing a line of red and blue dots

  Description automatically generated

Mean Squared Error: 23.171155902054664

R-squared: 0.7998669234406937

* Plotting the histogram of fair price  
  
* Because the histogram is skewed and doesn’t look right, we will keep only the frequent and normal values that are under $50.  
  A white rectangle with red text

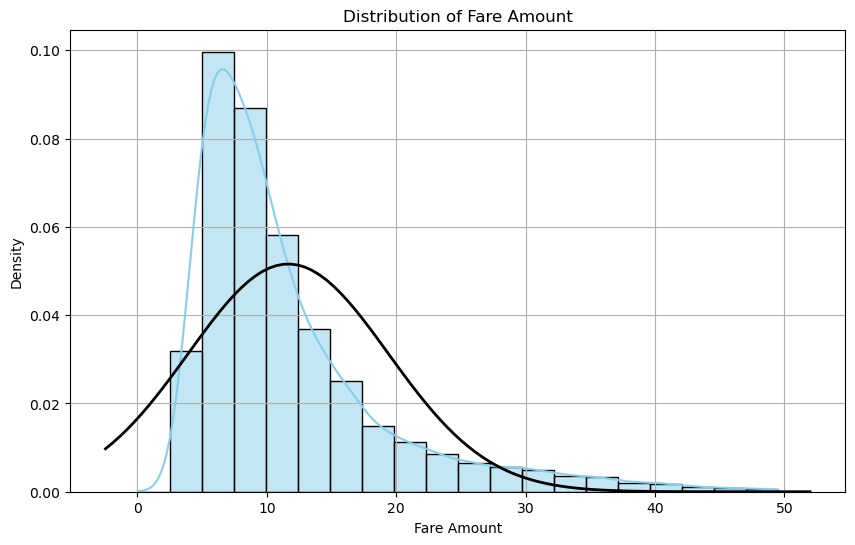
  Description automatically generated
* Plotting the trip distance and trip time against fair amount  
  
* Now we can see the linear relationship more clearly.
* Calculating the mean squared error and the coefficient of determination R2

A computer screen shot of a program

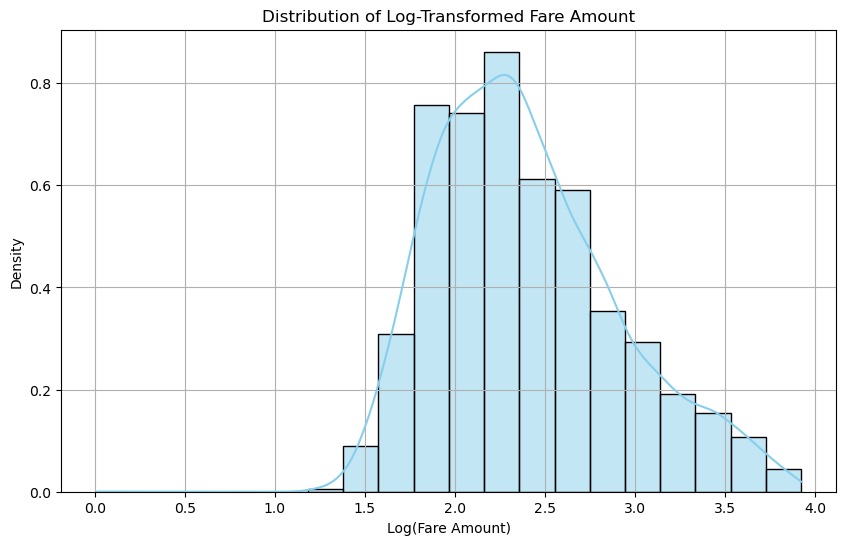
Description automatically generated

* Plotting the distribution of fair price  
  A computer screen shot of a code

  Description automatically generated



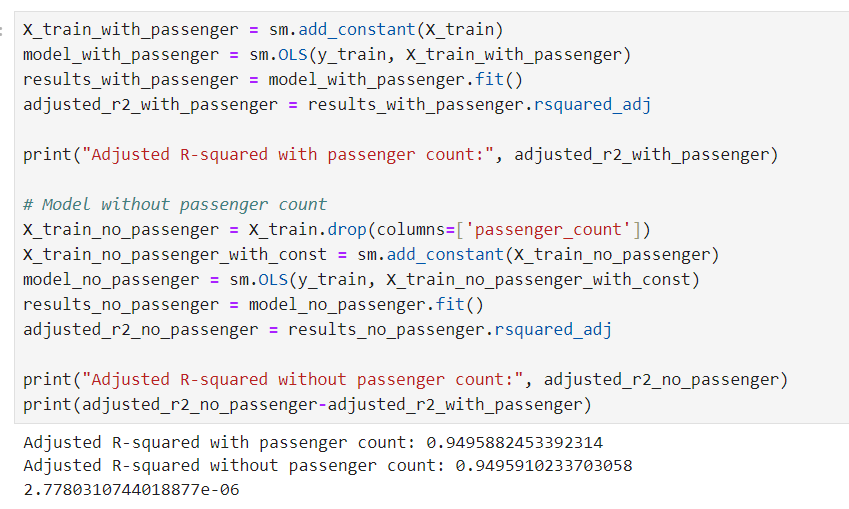
* Now the histogram looks much better but still it is a little skewed so we’ll use log to normalize it.  
  A computer code with colorful text

  Description automatically generated
* Plot the distribution of normalized fair price  
  
* Testing and comparing the actual data against the predicted values  
    
  A graph with red and blue dots

  Description automatically generated

Mean Squared Error: 4.762193168885501

R-squared: 0.9506388500987996

* Calculating the coefficient of determination R2 with and without the passenger count.  
  
* Adjusted R2 without the passenger count is greater than adjusted R2 with the passenger count.
* Plotting the new model without the passenger count.  
  A close-up of a computer code

  Description automatically generated

A graph showing a red and blue line

Description automatically generated

* Calculating the model coefficients to get the model equation  
  A screenshot of a computer code

  Description automatically generated

Fare amount = 1.64 + 0.14\*Trip Distance + 0.03\*Trip Time

* A function to predict the fair price based on the trip distance and the trip time  
  A screen shot of a computer code

  Description automatically generated