Executive Summary

Milestone 2 of the Automatidata Claims Classification Project

Overview

New York City TLC seeks to use travel data to perform a linear regression model to help estimate and predict fares for future trips. We are conducting an initial exploration of the data before structuring and organizing it..

Objective

Perform an initial scan of the data submitted by New York TLC. Observe the integrity of the data, its trends and analyze if there are any trends on the variables or unusual data.

Results

In the initial exploration of the data, no missing data were found. However, strange behaviors were found in the data such as fares with negative or free fares. Data such as long trips with decreased fares were also observed.

The variables that could be much more useful to plot are the fare and trip distances variables. It is recommended to obtain more information about the trips to understand the data extraordinary events presented by the data, such as promotions, etc.

Next Steps

The next steps would be to explore the correlations between the distance and fare variables in order to have an approximate amount. One could also classify these variables by various schedules in order to readjust with extra fares for overnight trips, among others.

Executive Summary

Milestone 2 of the Tik Tok Claims Classification Project

Overview

Tik Tok users have the ability to submit reports that identify videos and comments that contain user complaints. These reports identify content that should be reviewed by moderators. For this reason, Tik Tok is working on a predictive model to determine whether a video contained a claim or an opinion, reducing response time and improving customer service.

Objective

Conduct an initial exploration of the data obtained from the Tik Tok videos to ensure data integrity. In addition to exploring the relationships of the variables, also explore whether there is any difference between videos classified as claim and opinion.

Results

In the initial exploration of the data, missing data was found in some of the variables, in addition to the need for more variables for the elaboration of the model.

On the other hand, the classifications of claim and opinion retain, in a first approach, the same trends, with the videos classified as claim having the highest number of likes, shares and views. These trends were also analyzed separating by author status (under review, banned, reviewed), observing the same trends regardless of status. However, it was found that banned authors have higher engagement than the rest of the status. Regardless of whether it was an opinion or a claim.

Next Steps

In case of having more data and variables add them to the data set and if possible complete the missing data.
On the other hand, it is necessary to specify in the data set the criteria for banning the authors, since there is no initial information on the criteria to verify the authors.

Executive Summary

Milestone 2 of the Waze Claims Classification Project

Overview

Waze's current project is to create a model to predict users who within a one-month period abandon the app. The first step is to analyze the data obtained by Waze from the activity of its users to understand the factors that influence users' decision to leave the app.

Objective

Perform an exploratory analysis of the data provided by Waze, ensuring the integrity of the data and if sufficient data is available for the analysis, also establish if there is any initial structure or trend in the data.

Results

The data set provided has the necessary variables for the first approach to analyze if there are patterns in the abandonment of users on a monthly basis.

Additionally, a first trend was found in the data. It was observed that a number of users who abandoned the application have in common a higher number of hours driven in their trips and a lower number of trips in less time than the average user of the application.

The hypothesis is that these “experienced” users drop out because the app may not fully satisfy the particular needs of this group of users.

Next Steps

To confirm this first observation, more data are needed to ensure that it is a real trend and then to perform a first approach with a correlation analysis to establish the hypothesis and corroborate trends.