Report for OYC Project

Group 10

10/24/2021

❖ Story

We are an OYC company that serves customers to buy the cars used they want with the required features. We work to meet all customer requirements and provide them according to his specific budget and desired features.

❖ Assumption

- o When we choose the best price of the car with the good version, we help our client with budget business.
- o When we choose the car has low walk, we help our client to buy a good car with good price.

❖ Target project

Our goal is to reach a correct prediction through data pulled from an electronic car purchase website, which is to provide a car at an affordable price for each customer with the features ex: the walking of the car, the version.

❖ Data description

- o **Car_name:** name of car
- o **Car_price:** price of car
- o **Car_location:** location of car place such (Riyadh, Dammam, Jeddah)
- o **Car_brand:** brand of car such (Toyota, Nissan ,…)
- o **Car_walk:** the walk of the car by km
- o **Car_version:** the version of the car such as(2016 to 2021)

We convert the data type text to an integer to apply regression and prediction by 'astype' function and dummies to convert the categorical column by 0,1.

❖ Tools
- o Beautiful Soup and requests and csv and Itertools for web scraping
- o Sklearn.linear and sklearn.preprocessing and PolynomialFeatures and Sklearn.pipeline for represent linear Regration
- o Pandas with python language for clean and explore data.
- o Matplotlib library for data visualization

❖ Steps execute the project:

We are scraping the data from the ksa.motory.com website for Selling used cars by beautifulsoup web scraping tools:

o We did scrape the car name and price in all pages on the website by for and while loop.

o After scraping the name and price and saving in the list we did scraping other features such brand, walk, version and ……, by for loop to collect all products links and use for loop to follow each link to collect features in the table and saving in lists.

o After scraping all features, we did add all features on the dictionary then add them into the data frame to read and view data easily.

o After the data frame step, we did clean and explored data by pandas' tool (EDA).

o After the EDA step, we did apply the linear regression steps to view what features affect on cars.

o After train we did the test then found best model.

o After the linear regression step, we did measure fewer errors and validations.

o After the validation train model.