Big Data and Analytics: Final Project

Part One: Data Analytics

Case Study: Used Car Agency

1. Introduction:

A manager of used care agency would like to create a system that can be used to facilitate his work in the agency and all his employees and agents. Currently he is using spreadsheet for collecting and saving all the data about the cars that he had in his agency.

The idea is to create an online system to help him to search for a car, by various different values and inputs. Your role is to ***do a Business Intelligence, and Data Analytics*** for his agency using R programming language and provide the agency CEO about your prospective of the business.

2. Data description

The data that already stored in the excel file is dataset contains information about used cars that already sold or not by the agency. The dataset are as follows:

1. Name, (Model)
2. Year
3. Selling price
4. Km driven
5. Fuel type

6. Seller type
7. Transmission type
8. Owner (First or second)
9. Mileage (Fuel consume per km)
10. Engine capacity
11. Maximum horse power
12. Engine torque
13. Number of seats
14. Name of dealer

The agency has also three dealers working full time:

1- Mr. Henry Spelman

2- Ms. Anny Grouper

3- Mr. David Soliman

3. Project Requirement

Your job is to do data analysis on the dataset, using R programming language, and excel in the following steps:

1- ***Statistical Analytics using Excel:*** After cleaning the data set in the ways that you see it is possible and benefit for your analysis, Uses excel data shit to do Statistical Analysis of the Car Agency data, *you are free to provide your view of that*. The analysis must be done only using excel, and the results can be shown on graphs, tables or both, Please explain in details your results, ***(Results or code without your discussion will not be accepted)***. 10 points.

2- ***Perspective analysis using R:*** using R programming language uses the same data set and cleaned using R in the ways you see it will benefit your analysis. Do perspective analysis and show your results in Tabular form and graphs forms using ggplot2 package. Please explain in details, your results means, compare your results with point .1 ***(Results or code without your discussion will not be accepted)***. 15 points.

3- ***Predicts Analytics using R:*** using R programming language uses the same data set and cleaned using R in the ways you see it will benefit your analysis. Do some predictive about the Car Businesses for the Car Agency and how the sails will be in the future, using ***Linear Prediction or Logistic Regression***, By using one feature prediction and then multi features prediction and show your results in Tabular form and graphs using ggplot2 package. Please explain in details, what are the results, compare your results with point .2 ***(Results or code without your discussion will not be accepted)***. 15 points.

4- ***Prepare a power presentation (***10 points for the final presentation and Q and A)

Notes:

1. Write your analysis in the reports using graphs of your choices (Histogram, bar chart, scatter chart, bobble chart …) it's up to you to decide the number of charts, and how you create them.

2. Cleaning and Normalization: It is better to clean and normalize your data before used in prediction and perspectives analysis.

3. Report Information: Please write in details a report that contained the following information (Home work without a repost will not accepted). (30 points penalty)

4. Provide your R code, and excel file in separated files.
5. Answer the following questions using R programming: (5 points penalty)
    - Find out which car model are more sailed.
    - Which dealers' sale more cars, and his total sales?
    - What is the average price for each car model?
    - What is the newest car and the oldest car?
    - Which is the best car related to price, and low mileage?
    - What is the total revenue from Individual sailing?
    - What is the famous care type for each agent?
6. Predict the any car price based on the odometer, (2 points penalty)
7. Predict the any car price based on the odometer and year of manufacture. (2 points penalty)

Notes:

1- Please save your Homework compressed zip file according:

Group Name – Final Project.zip

2- if you have several documents, or saved as pdf file if it is a single file.

3- Upload the power point presentation, including group names, and groups members.

4- Please submit your homework as zip file before 12 July 12PM using the team upload link in your account (One upload for team group members will be sufficient).

5- You presentation data will be:

- DBDS: 13.07.2022 at 13:00 PM
- SE : 14.07.2022 at 12:00 PM

6- Not showing to presentation (Fail the course).

Good Luck

Prof. Dr. Rand Kouatly