Mini Project on Used Cars Data

Sub-Topics	Max Marks
1.1 Data manipulation using Python (30 marks) 1.2 Statistical Analysis using Python (30 marks) 1.3 Exploratory Data Analysis using Python (40 marks)	**100**

Domain:
Automobile Industry

About:
There is an automobile company XYZ which aspires to enter the Used-car market by setting up their company locally to give competition to their counterparts.

Challenges:
They want to understand the factors affecting the pricing of cars in the market, since those may be very different from the new car market. Essentially, the company wants to know:

- Which variables are significant in predicting the price of a used car?
- How well those variables describe the price of a car

Based on various market surveys, the consulting firm has gathered a large dataset of different types of used cars across the market.

What is Expected?
Being a data analyst, you must come up a first step document that lists output of your exploratory analysis, any issues or problems you may see with data that need follow up, and some basic descriptive analysis that you think highlights important outcomes/findings from the data. Based on your findings, the next level of analysis will be charted out.

.Data Dictionary:

Column Name	Description
Sales_ID	Sales ID
name	Name of the used car
year	Year of the car purchase
selling_price (target variable)	Current sellling price for used car

km_driven	Total km driven
Region	Region where it is used
State or Province	State or Province where it is used
City	City where it is used
fuel	Fuel type
seller_type	Who is selling the car
transmission	Transmission type of the car
owner	Owner type
mileage	Mileage of the car
engine	engine power
max_power	max power
seats	Number of seats
sold	used car sold or not

Here are some indicative types of analysis you can perform. Please note that this is not an exhaustive list, you may add more

Task 1.1(Data Manipulation using Python)

Here are some indicative types of analysis you can perform. Please note that this is not an exhaustive list, you may add more

- Come up with appropriate results for the following:

    a. Which brands are selling most?
    b. Are there specific locations selling more?
    c. Which factors are more important in deciding cars' selling price? Ex. kms driven or type of owner or fuel type?

Task 1.2 (Statistical Analysis using Python)

- Descriptive statistics for both numerical and categorical and draw few insights from them.
- Perform relevant hypothesis testing (t, chi-Square, Anova tests)

TASK 1.3 (Exploratory Data Analysis)

Data Preparation/Analysis tasks including (but not limited to) the following.

- Univariate, Bi- Variate Analysis and Multi- Variate Analysis
- Missing values identification and treatment
- Outlier analysis and treatment
- Data scaling using min-max and/or Z-score normalisation
- Data transformation
- Feature Engineering

Deliverables/Submission guidelines

1. You have to prepare a power point presentation with screenshots of outputs (10 -15 slides)
2. Mention Problem Statement and Your approach to the problems
3. You need to submit all the code files .
4. All comments/inferences/insights/reasons for doing a particular tasks etc should be written as a 'markdown text', but **NOT** using a comment lines with # or ''.
5. Submit the code file as HTML file format (Also, submit Jupyter Notebook).
6. Upload all the deliverables in the LMS