



**JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA**  
**PGDM / PGDM (M) / PGDM (SM)**  
**IV TRIMESTER (Batch 2021-23)**  
**END TERM EXAMINATION, NOVEMBER 2022**  
**SET - 2**

Course Name	Programming for Business Analytics (PBA)	Course Code	<b>20823</b>
Max. Time	<b>2 hours</b>	Max. Marks	<b>40 MM</b>

**INSTRUCTIONS:**

- a. Attempt all the questions on a single Jupyter Notebook
- b. The data for the case is available on Moodle.
- c. Write down your Roll no., course name and course code on top of Jupyter Notebook
- d. Save your Jupyter notebook with .ipynb extension and as pdf file
- e. Upload both the files on Moodle.
- f. Label the files as PBA\_roll no (for example: PBA\_23)
- g. This is an open book exam. Students may refer to the codes

**Read the case below and answer the questions given by analyzing the data using Python.**

**Case: Estimating Price of Used Cars**

A large XYZ car dealership offers purchasers of new XYZ cars the option to buy their used car as part of a trade-in. In particular, a new promotion promises to pay high prices for used XYZ cars for purchasers of a new car. The dealers then sell the used cars for small profit. To ensure a reasonable profit, the dealer needs to be able to predict the price that the dealership will get for the used cars. For that reason data were collected on all previous sales of used cars at the dealership. The description of each variable is given in the file "PBA\_Set2\_2022.csv". It has 1436 records containing details on 36 attributes, including *Price*, *Age*, *Kilometers*, *HP* and *other specifications*. The goal is to predict the price of a used Car based on its specifications.

Mr. Kanwaljeet Singh has hired you as the analytics consultant. He asked you to "identify" and "quantify" the factors for estimating the price of used cars in a multivariate fashion. Let's help Mr. Singh in carrying out the analysis.

**The data description is given on next page:**

Variable	Description
Id	Record_ID
Model	Model Description
Price	Offer Price in Rupees
Age_09_22	Age in months as in September 2022
Mfg_Month	Manufacturing month (1-12)
KM	Accumulated Kilometers on odometer
Fuel_Type	Fuel Type (Petrol, Diesel, CNG)
HP	Horse Power
Met_Color	Metallic Color? (Yes=1, No=0)
Color	Color (Blue, Red, Grey, Silver, Black, etc.)
Automatic	Automatic ( Yes=1, No=0)
CC	Cylinder Volume in cubic centimeters
Doors	Number of doors
Cylinders	Number of cylinders
Gears	Number of gear positions
Weight	Weight in Kilograms
Mfr_Guarantee	Within Manufacturer's Guarantee period (Yes=1, No=0)
Guarantee_Period	Guarantee period in months
ABS	Anti-Lock Brake System (Yes=1, No=0)
Airbag_1	Driver_Airbag (Yes=1, No=0)
Airbag_2	Passenger Airbag (Yes=1, No=0)
Airco	Airconditioning (Yes=1, No=0)
CD_Player	CD Player (Yes=1, No=0)
Central_Lock	Central Lock (Yes=1, No=0)
Powered_Windows	Powered Windows (Yes=1, No=0)
Power_Steering	Power Steering (Yes=1, No=0)
Tow_Bar	Tow Bar (Yes=1, No=0)

**Analyze the data and write the answers of the following questions:**

1. Identify the dependent variable in the case. **(2 marks)**
2. Mr. Singh wants to understand, what are the key factors that are significant in estimating the price of used cars? Which data science technique will you use? Why? **(4 marks)**
3. Analyze the data and perform Data Preprocessing. **(5 marks)**
4. Analyze the data and perform Data Exploration. **(5 marks)**
5. Build multiple regression models. Evaluate the results and answer the following questions: **(7 marks)**
  - a. List all important features for predicting the car's price? **(3 marks)**
  - b. What appears to be the most important car specifications for predicting the car's price? **(2 marks)**
  - c. Give the estimated regression equation from the output. **(2 marks)**

- d. Comment upon the model fit. **(3 marks)**
  - e. Evaluate the predictive accuracy of the model by examining its performance on the validation (test) data set. **(2 marks)**
6. Prepare a report by summarizing the usefulness of the study. **(5 marks)**

**Note:** Interpretation of all the outputs should be written by putting comments on the Jupyter notebook.

\*\*\*\*\*