

JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA
PGDM (G/SM/M); FOURTH TRIMESTER (Batch 2019-21) SET 1
END TERM EXAMINATIONS, OCTOBER 2020

Course Name	Data Visualisation	Course Code	BA 401
Max. Time	2 hours	Max. Marks	40 Marks

INSTRUCTIONS: Attempt all questions.

- The ggplot2 package offers a powerful graphics language for creating elegant and complex plots. Its popularity in the R community has exploded in recent years. “ggplot2” allows you to create graphs that represent both univariate and multivariate numerical and categorical data in a straightforward manner. Grouping can be represented by color, symbol, size, and transparency “mt cars” data in r has its details as follows.

Variable	Description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (lb/1000)
qsec	1/4 mile time
vs	V/S
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

By taking mtcars data, draw the graphs between following variables and explain the significance of graph.

- Boxplot of cyl, am and mpg
- Scatterplot of cyl, am and mpg
- Horizontal barchart of hp, am and gear
- Bubble chart of mpg, wt and cyl

(4 X 4 = 16 marks)

2. Tableau is a data analytics and visualization tool used widely in the industry today. Many businesses even consider it indispensable for data-science-related work. In real life, data can be complex, tedious, and sometimes it can be a difficult task to uncover patterns. Consider a situation where you have some sales data belonging to your company and you want to discover a pattern in terms of the spending capacity of the consumers. If you could uncover distinct groups or associations in the data, you could target the different groups accordingly to get maximum sales. The basic idea behind this intuition is called clustering and Tableau has an inherent feature which can automatically cluster similar data points based on certain attributes. By using Tableau, analyse the data in file “clusters” by drawing individual and group of clusters. (12 marks)

3. Python programming has become one of the most sought after programming languages in the world, with its extensive amount of features and the sheer amount of productivity it provides. The *pandas* package is the most important tool for data visualization at the disposal of Data Scientists and Analysts working in Python today. Working with Pandas in Python, enables you to tap into the power of the various other features and libraries such as NumPy, SciPy, Matplotlib, etc. By using Pandas, visualize the following in excel sheet “Superstore” and share your analysis:

- a. Profit on chairs and tables in respective category in each region
- b. Quantity of each category sold in each region
- c. Sales and Profits of different categories by shipping mode

Mention complete commands along with graphs. Please ensure that each graph is different. (12 marks)

JAIPURIA INSTITUTE OF MANAGEMENT, NOIDA
PGDM (G/SM/M); FOURTH TRIMESTER (Batch 2019-21) SET 2
END TERM EXAMINATIONS, OCTOBER 2020

Course Name	Data Visualisation	Course Code	BA 401
Max. Time	2 hours	Max. Marks	40 Marks

INSTRUCTIONS: Attempt all questions.

1. SPSS is used by market researchers, health researchers, survey companies, government entities, education researchers, marketing organizations, data miners, and many more for the processing and analyzing of survey data. Most top research agencies use SPSS to analyze and visualize data so that they can get the most out of their research projects.
By using SPSS, visualize and hence analyse data given in file loandata among the following variables
 - a. "marital";"education";"default": barchart
 - b. "marital";"age";"default": scatterplot
 - c. "age";"education";"default": boxplot (4 X 3 = 12 marks)
2. Bakery business is one of the largest food processing industry. Bakery products are popular in all parts of the country. The business is very lucrative as one can start with it by just renting a space. Choosing the correct product and right marketing strategy are the key factors of this business. Choose your products, as per your financial capabilities and demand aspects. The very first step to start a bakery is, drafting a good plan. By using Pandas, visualize the following in excel sheet "Bakery" and share your analysis:
 - a. A comparative analysis of cakes and coffee sold with (out) promotions
 - b. A comparative analysis of cakes and coffee sold on week days and weekends
 - c. A comparative analysis of all 5 products on Saturday and Sunday
 - d. A comparative analysis of all 5 products on Sunday with (out) promotions (4 X 4 = 16 marks)

3. The datafile "Monthly Sales" contains the data of different products sold in different months at different locations and different prices. By using ggplot2, visualize and explain the following:
- a. Comparative sales of Tape 1 & Tape 10 in different quarters
 - b. Comparative sales of all products in different locations- downtown or uptown
 - c. Sale of products (in Q1 and Q2) whose price is exactly 10. (4 X 3 = 12 marks)