

**MCKV Institute of Engineering**

**Question Bank**

**Paper Code: PE-CS602C**

**Paper Name: Data Analysis and Visualization**

*The figures in the margin indicate full marks.*

*Candidates are required to give their answers in their own words as far as practicable.*

**Group – A**

**(Multiple Choice Type Questions)**

1. Choose the correct alternatives for the following questions:

i) The ………..plot is a combination of box and kernel density plots.

a)Violin , b)mesh c) Contour d)Quiver

ii) With respect to colouring scheme cool colours are

a) shades of blue & green b) shades of red & orange

c) there is no such colour d) any colour may be

iii) The value of arguments in subplot for creating multiple images in 2 row and 4 columns is ………..

a) 42…. b)24.. c)81…. d18….

iv) The box plot is also called

a) Quiver plot b) violin plot

c) whisker plot d) None of the above

v) The ………….value is used to draw cyan square on the line plot

a)cyansq b)sqcyan c)cs d)sc

vi) Example of dimensionality reduction strategy are i) PCA ii) Wavelet transform iii) regression iv) Feature selection v) histogram . Which of the options is true

1. i),ii) iii) b) i) ii) iv) c) iii) iv) v) d) i) iv) v)

vii) Python programming language, following is not a data visualization libraries

a) matplotlib b) seaborn c)plt d) ggplot

viii) Outlier data points are important in calculating mean of an ensemble:

a)True b) false c) no contributing d)can’t be said

ix) In Data pre-processing step missing categorical data can be filled up using

1. mean b) standard deviation c) proper manual entry d) can’t fill in

x) What type of join is used in blending?

a) Left join b) Full join c) Right join d)Inner join

xi) What are the file extensions in Tableau ?

a) Tableau Packaged Workbook (.twbx) b) Tableau Data Source(.tds) c) Tableau Workbook (.twb) d) All of the above

xii) Views in Tableau\_\_\_\_\_

a) Dashboard b) Story

c) Sheet d) All of above

xiii) Which of these is not a datatype in Tableau?

a) Integer b) Character c) Float d) String

xiv) Data Analysis is a process of

1. Inspecting data
2. Data Cleaning
3. Transforming of data
4. All of the mentioned above
5. Data Analytics uses \_\_\_ to get insights from data.
6. Statistical figures
7. Numerical aspects
8. Statistical methods
9. None of the mentioned above
10. What is true about Data Visualization?
11. Data Visualization is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts.
12. Data Visualization helps users in analyzing a large amount of data in a simpler way.
13. Data Visualization makes complex data more accessible, understandable, and usable.
14. All of the above

xvii)Which are pros of data visualization?

1. It can be accessed quickly by a wider audience.
2. It can misrepresent information
3. It can be distracting
4. None Of the above
5. Which method shows hierarchical data in a nested format?

a)Treemaps b)Scatter plots c)Population pyramids d)Area charts

1. Which of the following is false?
2. Data visualization include the ability to absorb information quickly
3. Data visualization is another form of visual art
4. Data visualization decrease the insights and take slower decisions
5. None Of the above
6. Which one of the following is not a valid data type in Tableau?
7. String b)Integer c)Boolean d)Complex

xxi)What type of join is used in blending?

1. Left join
2. Full join
3. Right join
4. Inner join

xxii) Amongst which of the following is must before using any technology to evaluate your data,

1. Study the dataset
2. Organize dataset
3. Remove impurities from data set
4. All of the mentioned above
   1. Data Identification → Data Acquisition & Filtering → Data Extraction → Data Validation & Cleansing, are the phases of?
5. Data Analytics Lifecycle
6. System Analysis and Design
7. Software Development and Life Cycle
8. None of the mentioned above
   1. Which is/are drawbacks of data visualization?
9. It conveys a lot of information in a small space.
10. It makes your report more visually appealing.
11. Visual data may be distorted or excessively used.
12. None Of the above
    1. Which method shows hierarchical data in a nested format?
13. Treemaps
14. Scatter plots
15. Population pyramids
16. Area charts
    1. Which of the following is false?
17. Data visualization include the ability to absorb information quickly
18. Data visualization is another form of visual art
19. Data visualization decrease the insights and take slower decisions
20. None Of the above
    1. Which of the following statements is true about Tableau?
21. Tableau is a relational database management system.
22. Tableau is a business intelligence and data visualization tool.
23. Tableau is a programming language used for statistical analysis.
24. Tableau is an operating system.
    1. Data can be visualized using?
       1. Graphs b) Charts c)Maps d)All of the above
    2. Views in Tableau\_\_\_\_\_

a)Dashboard b)Story c)Sheet d)All of above

* 1. Histogram and bar charts represents   a) Exactly Same information, b) Totally different information

c) Similar data trends d)No such graph exists

* 1. Which one of the following is not a valid data type in Tableau?

a) String b. Integer c) Boolean d) Complex

* 1. Which one of the following is not a type of join in Tableau?

a) Inner join b) Left join c) Right join d) Top join

* 1. What is a calculated field in Tableau?

a)A column in a data source that is created by performing a calculation on existing columns.

b) A chart that displays values as bars.

c) A type of data source that is optimized for live connections to data.

d) A type of join used to combine data from multiple tables.

* 1. Kkjhdkj
  2. jbkdj

**Group – B**

**(Short Answer Type Questions)**

**Module 1:**

1. Briefly state the importance of Data Visualization. [Module1/PE-CS602C.2, Understand, LOCQ][5]
2. Define data visualization. Mention different Data visualization techniques. [module1/CO1/Remember- LOCQ] [5]
3. Why data visualization is such a powerful tool ? [Module1/CO1/Understand-IOCQ][3]
4. Define Data Analytics. State any two factors/drivers that are involve in the growth of data analytics and hence mention how they are involve with this growth. [Module1/CO1/Remember-IOCQ] [1+2+2]
5. Define Data Analytics Lifecycle? [Module1/CO1/Remember-IOCQ] [2]
6. State different phases of data analytics lifecycle. [Module1/CO1/Remember-IOCQ] [3]
7. Mention how these phases are used to address big data analytics project. [Module1/CO1/Remember-IOCQ] [10]
8. Mention stages of visualizing data. Explain any three steps of visualizing data in data science process. [Module1/CO2/Remember-IOCQ] [3+(3x4)=15]
9. Explain 7 Gestalt principles of visual perception with the help of proper figure. [Module1/CO1/Remember-IOCQ] [10]
10. Explain Tufte’s Design Principles briefly. Explain the role of data ink ratio. Explain the situations where this principle contradicts the real situation. [Module1/CO1/Understand-IOCQ] [10]

**Module2: (11,13, 14,15)**

1. Briefly Explain about real time issues in data gathering and preparation [Module2/PE-CS602C.2, Understand, LOCQ][5]
2. Explain “Data Cleaning” steps used during data pre-processing [Module2/PE-CS602C.2, Remember, LOCQ] [5]
3. Explain the requirement of dimensionality reduction of a dataset. Mention different data reduction strategies used during data pre-processing? [module2/PE-CS602C.2, Understand, LOCQ] [5]

1. State PCA algorithm for dimensionality reduction. [Module2/CO2/Understand- IOCQ] [6]
2. Functions of matplotlib are similar to functions of which programming language/package. With respect to python write the functions in sequence available under matplotlib.pyplot for drawing a line with all axis heading, graph heading and visualizing the line with red squares. [Module2/CO2/Apply- IOCQ] [5]
3. Define COV(X,Y) and Calculate COV(X,Y) for the following dataset. Explain the meaning of COV(X,Y)=0 , COV(X,Y)>0 and COV(X,Y)<0 . [Module2/CO2/Apply, IOCQ] [2+4+3]

| X | 16 | 10 | 30 | 20 | 24 | 6 | 34 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | 10 | 7 | 17 | 12 | 14 | 5 | 19 |

1. Explain 5 principles of improving a vision of a plot/chart. Give example wherever required. [Module2/CO2/Understand-IOCQ] [5]
2. Explain the requirement of data cleaning during data preprocessing. [Module2/CO2/Understand-IOCQ][3]
3. What is meant by categorical variable ? Mention the string manipulation functions used for noise reduction of categorical variables with example. [Module2/CO2/-IOCQ][5]

**Module3:**

1. Diagrammatically represent Box& Whisker plot and mention the parameters that can be measured from it. [Module3/CO3/Understand-LOCQ] [5]
2. Define Analytics Architecture. Analytics -architecture can support a wide range of capabilities- mention and briefly discuss any four of them.
3. Differentiate between data exploration and explanation analysis with respect to visualization. [Module3/CO2/Understand/IOCQ] [5]
4. Discuss the contrast between the following: i) Stacked Bar chart and Grouped bar chart.

ii) Line chart and area chart. [Module3/CO3/Understand-LOCQ][2+2]

1. Mention any two ways by which you can visualize hierarchical data and compare between them. [Module3/CO3/Understand-LOCQ][6]
2. Discuss any three mistakes that we should avoid when designing data visualization. Define combination chat with proper example [Module3/CO3/Remember-LOCQ] [2+3]

**Module4: (25, 26)**

1. Is it possible to plot categorical variables using the functions of matplotlib? With respect to python write the functions in sequence available under matplotlib.pyplot for plotting 3 categorical variables in 3 different ways (line, bar, scatter) under one plot. Put heading and x-axis, y-axis properly. [Module4/PE-CS602C.3, Apply, IOCQ] [5]
2. "Visual representation of data should ethical " - Justify [Module4/CO3/Understand-IOCQ][5]
3. Depending on your dataset how will you select the chart type and colour scheme to be used? [Module4/CO3/Understand-IOCQ][5]
4. List five reasons to why data visualization is important. [Module4/CO3/Remember-IOCQ] [5]
5. Depending on your dataset how will you select the chart type and colour scheme to be used. [Module4/CO3/Understand/IOCQ] [5]

**Module 5: ( 30+31,38,39, 40,42,44, 46,33+47, 48)**

1. Mention few points in favour of using Tableau software. [Module5/CO4/Remember-LOCQ][2]
2. The Tableau Product Suite consists of several parts. Mention all of them. [Module5/CO4/Remember-LOCQ][3]
3. What are the file extensions used in Tableau? [Module5/CO4/Remember-LOCQ][2]
4. You are given a dataset and need to prepare a presentation based on the different types of plots. Is it possible to use different colors depending on certain condition on different part of line graph you are preparing? Mention the steps. Mention the steps to send your graphs from Tableau to power-point presentation. [Module5/CO4/Apply-IOCQ][4]
5. What data types does Tableau support? [Module5/CO4/Understand-LOCQ][2]
6. What are Measures and Dimensions? [Module5/CO4/Understand-IOCQ][2]
7. What are the different types of joins in Tableau? [Module5/CO4/Understand-IOCQ][2]
8. What is disaggregation and aggregation of data? [Module5/CO4/Understand-IOCQ][2]
9. What happens when you join two tables in Tableau and values do not match? [Module5/CO4/Understand-IOCQ][2]
10. Define Tableau. What are the different data types in Tableau? [Module5/CO4/Remember-IOCQ] [5]
11. Compare among Worksheet, Dashboard and Story with respect to tableau. [Module5/CO4/Understand/IOCQ] [5]
12. Mention few points in favour of using Tableau software. The Tableau Product Suite consists of several parts. Mention all of them. [Module5/CO4/Understand-LOCQ] [5]
13. Differentiate between Dual-axis and blended axis in Tableau? [Module5/CO4/Understand-IOCQ] [4]
14. How to create a calculated field in Tableau? [Module5/CO4/Understand-IOCQ][4]
15. Can we remove the "All options" from a Tableau auto-filter? If yes write down the steps. [Module5/CO4/Understand-IOCQ] [4]
16. What are the different types of joins in Tableau? [Module5/CO4/Understand-IOCQ][3]
17. Design a view in a map such that if a user selects any state the cities under that state have to show profit and sales using Tableau. [Module5/CO4/Understand-IOCQ] [15]
18. Mention the significance of using calculated fields in Tableau. Also explain the types of calculations that can be done through these fields. [Module5/CO4/Understand-IOCQ] [4+4]
19. Mention the significance of creating parameters in Tableau. [Module5/CO4/Understand-IOCQ] [3]
20. Explain LOD in details with respect to Tableau software. [Module5/CO4/Understand-IOCQ] [3]
21. What is the difference between a worksheet and a dashboard in Tableau? [Module5/CO4/Understand-IOCQ] [3]
22. What is a calculated field in Tableau? [Module5/CO4/Understand-IOCQ] [3]