Page No.:
Date: / /

Q.1) Attempt any eight of following.

1) What is Data science?
→ Data science is deep study of massive amount of data, which involves extracting meaning full insights from raw, structured & unstructured data.

2) Define Data source?
→ A Data source is ~~empty~~ simply the source of the data. A Data source in data science in the initial location where data that is being used come from

3) What is missing values?
→ Some values in the data may not be filled up for various reason is called as missing value.

4) List the virualization libraries in python
→ i) matplotlib v) Leather ix) Geoplotlib
ii) seaborn vi) plotly x) Gleam.
iii) Ggplot vii) Bokeh
iv) missingno viii) pygal

5) List application of Data science.
→ i) image Recognition speech recognition
ii) Gaming world.
iii) Internet search
iv) Transport
v) Heltheare
vi) Recommandation system.
vii) Risk Detection.

Date:

6) What is Data transformation?
→ Data transformation is the process of converting raw data into a format that would be more suitable for data analysis.

7) Define Data cleaning
→ Data cleaning is the process of correcting of removing incorrect incomplete or duplicate data with a dataset.

8) Define standard Deviation?
→ Standart deviation is a measure of how dispersed the data is in relation to mean.

Q.A Attempt any four of the following
1) List the tools for data scientist
→ The tools are following is
i) Pytorch ii) Numpy iii) Pandas iv) SQL v) Excel vi) Tableu vii) Tensor flow iv) Apache spark

2) What is Data cube?
→ i) A Data cube is a multi-dimensinal array of values. A data cube generally used to easily interpret Data.
ii) Data cube is especially useful when representing data together with dimensions as certain measures of business requirments.

3) What are the purpose of Data visualization?
→ i) Data visulization is the presentation of data in graphical format.
ii) Data visulization is a generic term is used which describes any attempt to help understanding of data by providing visual representation.

iii) It is the very important part of the data any analysis. We can it to explore our data.

4) Define statistical data analysis
→ i) statistical data analysis is the collection & inter interpretation of data in order to uncover pattern & trends.
ii) It is component of data analytics.
iii) It is the science of collecting exploring & presenting large amount of data to discover underlying pattern & trends.

Q.3) Attempt any 2 of the following
1) What are the measures of central tendancy? Explain any 2 of them in brief.
→ i) One of the simplest & yet important measure of statistical analysis is to find one such value that describe the charactersic of the entire huge set of data.
ii) A measure of central tendancy is a summary statistic that represent the point or typical value of dataset.
iii) There are mainly 3 ways to measure central tendancy, mean, median, mode as describe below
i) mean :-
i) It is the popular & widely used measure of repressering the entire data by one value.
ii) Mean is the ratio of the sum of an observations in the data to the total number of observ
2) Median :-
i) The median of a distribution with a discrete random variable depends on whether the number of terms in distribution is even or odd.

Page No.:
Date: / /

Q.4) Attempt any 2 of the following
1) Explain different data format in brief
→ i) In data science the data appears in different size of & shapes it can be numerical data, text, audio, video or a few other types of data.
ii) some data formats in data science are explained:
a) Integers :- Integers data types may be different sizes & may not be allowed to contain negative value.

b) Floats :- A floating point has decimal point even if that decimal point value is 0.

c) Text Data :- Text Data type is know as string in python, or object in pandas, string can contain number & or characters.

d) Dense Numberical Arrays :- Most image files or sound files consist mostly of dense arrays of numbers, packed adjacent to each other in ~~not~~ many

e) Compressed or Archieved Data :- many data files when stored in a particular format take up a lot more spaced compared to the file in question logically needs

f) CSV format :- CSV stand for comma separated value CSV files are the commanly used data format for data science.

3) What is venn Diagram? How to create it? Explain with example.

→ i) A venn diagram is a also called primary diagram set diagrams or logic diagram is a diagram that shows all possible logical relationships beth a finite collection of different set.

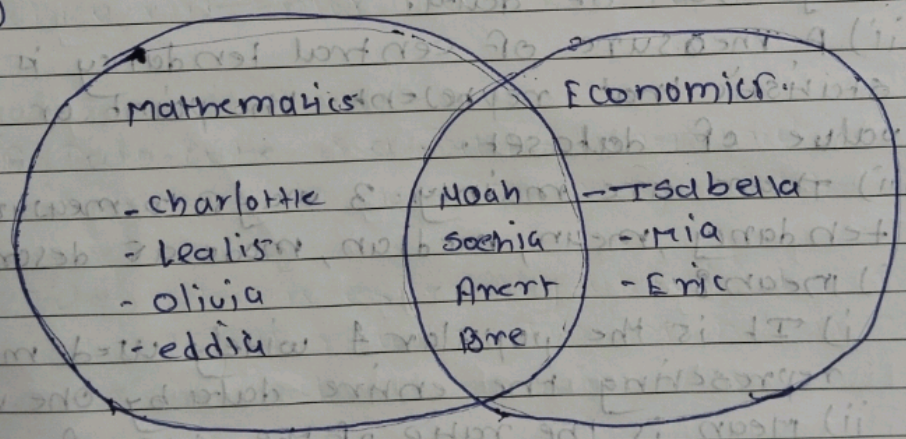
ii) Each set is represented by circle. The circle size sometimes represents the important of the group but not always.

iii) A venn diagram that visually display all the possible logical relationship beth collection of set.

iv) A venn diagram is we a) visually org apes b) compar contrast.
c) Find correlation & predict probability.

v) The simplest way to create a venn diagram is by drawing two overlapping circles.

vi)



2) Write detail notes on Basic data visualization tools? (9.5)

→ i) Data visualization tools such as Histogram, Bar charts, scatter plots, line chart area plots, pie charts, Donut chart if so ?

ii) Histogram :- A hirstogram is a graphical display of data using base of different height

iii) Bar charts :- Bar chart is are used for comparing the quantities of different categories or groups.

iv) Line plot :- The line plot is a two-dimensional plotting of values connect to the order.

v) Scatter plot :- A scatter plot is a 2-D chart showing the comparison of two variables scattered across 2 axes.

vii) Pie chart :- A pie chart show the proportion or percentage of a data elements in a circular form.

vi) Area plot :- Area charts are used to plot data trends over a while to show how a value changing.

viii) Donut chart :- A donut chart is an extension of a pie chart.

Page No.:
Date: / /

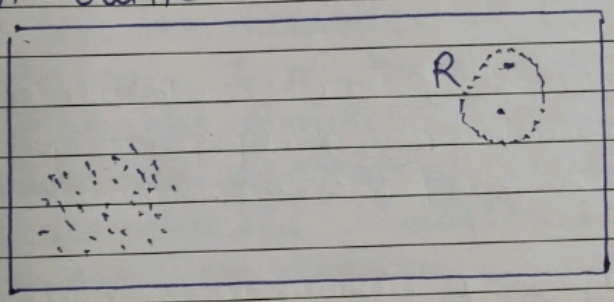
Q.5) Attempt any one of the following.

i) What is Outlier & state typer of outliers?

→ i) Outliers as a very important aspect of data analysis.

ii) An outlier may indicate an experimental error or it may be due to variability in the measurment

i) Global Outlier :- If an individual data point can be considered anamalour with respect to the rest of the data, then the donut is as a point outlier.



ii) Contextual outlier :- If on individual data instance is anamalour in a specific context or condition then it in termed as contextual outlier.

iii) collective outlier :- If a collective of data point is anamalour with respect to the entire data set it is termed as a collective outlier.