

CS3352 FOUNDATIONS OF DATA SCIENCE

QUESTION BANK

UNIT -1

PART A

1. What is Data preparation?
2. Define Data Warehousing
3. Define Data Mining
4. Define Regression
5. Define SVM
6. Define Data Science
7. Define Data Exploration

PART B & PART C

1. Explain briefly about Data Preparation process in detail.-3 steps of data preparation like data cleaning, data integration, data transformation in detail, challenges and benefits
2. Explain briefly about Facets of data-types of data with an examples and necessary diagrams
3. Explain briefly about data mining and Data warehousing (definition, advantages, disadvantages, types, architecture)
4. Explain briefly about Data mining-Data mining definition, types of data mining algorithms like a priority algorithm, kmeans clustering, classification, decision trees, support vector machine, Regression, Association rules classifier
5. Explain briefly about Steps involved on data mining same as data science process steps, benefits
4. Explain briefly about data science process
5. Explain briefly about Data exploration

UNIT – 2

PART A

1. Define Mean, Median Mode
2. Define Quartile ranks
3. Define Sample Mean Population
4. Define zscore
5. What is Frequency distribution problem?
6. What is Cumulative, Relative frequency?
7. What is Range
8. What is variance?
9. Define Standard Deviation.

PART B

1.Explain about theTypes of Descriptive statistics in detail-4 types

measures of frequency-count,percentage,frequency-steps

. Measures of central tendency/averages-mean,Median,mode

Measures of variability/dispersion-range,variance and standard deviation,kurtosis and skewness

Measures of position-percentile, quartile ranks

2.Explain about zscore,properties, finding proportions

3. Explain about sample mean,population mean

4.Explain briefly about frequency distribution with its types.

5. Explain about Cumulative, Relative

6. Explain about Rules to find frequency distribution

7. Explain about Frequency distribution problem.

UNIT – 3

PART A

1.Define Correlation.

2.Define Standard Error

3.Define Least Square regression

4.What is regression towards mean?

5.What are the types of Correlation.

PART B & PART C

1. Explain about types of correlation and scatter plots.

2. Explain about correlation,regression and standard error problem.

3. Explain about correlation coeffiecient

3. Explain about least square regression definition,least square regression line,example

4. Explain about regression towards mean

UNIT – 4

PART A

1. Define Broadcasting on numpy arrays.
2. What is fancy indexing?
3. What is Slicing?
4. What is indexing?
5. What is merge, join and concat?
6. Define Masks.
7. What is Pandas Series

PART B

1. Explain about Pandas missing data -Nan, none
Ways to handle missing data-detect null values, non null values, fill values and drop values
Example program
2. Explain about computation on numpy
Ufuncs- definition, trigonometric, Statistical fns, arithmetic fns
Broadcasting-defn, 3 rules
Fancy indexing
Comparisons
Masks
Example for each.
3. Explain about Fancy indexing in numpy
4. Explain about hierarchical indexing in pandas
5. Explain about merge, join and concat datasets
6. How to create pandas objects
By pandas series
By pandas datagram
By pandas index
7. Explain briefly about aggregation and grouping in pandas.

8.explain about numpy arrays

slicing,indexing,concatenation,splitting of arrays,creating arrays from list,scratch,array attributes

UNIT – 5

PART A

1.What is Data Visualisation?

2.Define Matplotlib

3.Define Subplots

4.Define Customisation.

5.What is 3d Geographical data?

PART B

1.Explain about Data visualization -introduction to matplotlib

2.Explain about Advantages of data visualization

3.Explain about Types of plots line and scatter plots

4.Explain about Sub Plots

5.Explain about Customisation

6.Explain about Data visualization with seaborn

7.Explain about Types of plots

8.Explain about types of projections in geographical base map

9.Explain about 3D geographical data