

QUESTION BANK FOR LAB EXTERNALS

1. A) Define Data Analytics? Explain variable, measurement and data. Why data is important?
B) Write a Python code to perform following array operations using numpy library.
Arrays: A=[1,2,3,4,5], B=[6,7,8,9,10]
 - Perform basic arithmetic operations on two arrays
 - Find min, max, square root, exponential of given array elements
 - Find mean, exp, std deviation of given array

2. A) What are the different functions used for data cleaning using pandas in python?
B) Implement a program to do Data Cleaning and Data Manipulation Operations on the dataset '**datasetExample.csv**'. Perform following operations:
 - Identify the datatype of each column
 - Check and remove all duplicate records
 - Check and remove all duplicate columns

3. A) How to work with missing data in python?
B) Write a Python code to perform Data pre-processing task on given dataset '**pre-process_datasample.csv**' (eg. Checking and Handling Missing Data, handling inappropriate data)
 - Checking and Handling Missing Data on all Columns
 - Identify Null values in a column
 - Handle missing values
 - Check and Handle Categorical Columns

4. A) What is Hypothesis testing? Explain Null and Alternate Hypothesis.
B) Implement the python code for hypothesis testing of following data:
A sample of 25 individuals is taken, and their cholesterol levels are measured.
Cholesterol Levels (mg/dL): 205, 198, 210, 190, 215, 205, 200, 192, 198, 205, 198, 202, 208, 200, 205, 198, 205, 210, 192, 205, 198, 205, 210, 192, 205.
Populations Mean = 200
Population Standard Deviation (σ): 5 mg/dL(use critical value approach)

5. A) Explain Simple Linear Regression.
B) Write a Python code to implement Simple Linear Regression on given data:

x = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

y = [1, 3, 2, 5, 7, 8, 8, 9, 10, 12]

6. A) Illustrate variance and standard deviation with an example. Also give uses of standard deviation

B) Python code to provide summary statistics (Mean, median, mode, standard deviation) on a given dataset

num= [1, 3, 4, 5, 7, 12, 34, 56, 74, 324, 355, 546]

7. A) Explain the steps to perform Hypothesis Testing? Also give decision rule for p-value approach.

B) Implement the python code for hypothesis testing of following data:

You are determining if the average age of 10 people is 30 or otherwise.

ages = [45, 89, 23, 46, 12, 69, 45, 24, 34, 67]

8. A) What is Exploratory Data Analysis?

B) Write a Python Program to Implement Exploratory Data Analysis on **CARS** dataset.

➤ Modifying data

➤ Removing unwanted data

➤ Retrieving data

➤ Getting Statistical Information

➤ Draw Plot/Graph

9) A) What is hypothesis testing. Explain type I and type II errors.

B) T-test (sigma of the population is unknown)

Experian Marketing Services reported that the typical American spends a mean of 144 minutes (2.4 hours) per day accessing the Internet via a mobile device. (Source: The 2014 Digital Marketer, available at ex.pn/lkXJfX.) In order to test the validity of this statement, you select a sample of 30 friends and family. The results for the time spent per day accessing the Internet via mobile device (in minutes) are stored in InternetMobileTime

a) Is there evidence that the population mean time spent per day accessing the Internet via mobile device is different from 144 minutes? Use the p-value approach and a level of significance of 0.05.

b) What assumption about the population distribution is needed in order to conduct the t test in (a)