School of Computer Science Engineering and Technology

Course- BTech
Course Code- CSET-214
Year- 2024-25
Date-

Type- Specialization Core
Course Name- Data Analysis using Python
Semester- Odd
Batch-

Lab # No. (8) Regular Expression, Histogram, Bar chart, Pie Chart

CO Mapping

Lab No.	Name	CO1	CO2	CO3
8	**Regular Expression, Histogram, Bar chart, Pie Chart**	➹	➹	➹

Introduction:

Descriptive statistics refers to a branch of statistics that involves summarizing, organizing, and presenting A RegEx, or Regular Expression, is a sequence of characters that forms a search pattern. It can be used to check if a string contains the specified search pattern. Data visualization is a field in data analysis that deals with visual representation of data. It graphically plots data and is an effective way to communicate inferences from data. Python offers several plotting libraries, namely Matplotlib, Seaborn and many other such data visualization packages with different features for creating informative, customized, and appealing plots to present data in the most simple and effective way.

1. WAP in Python using regex to accomplish the following task.

   a. Create pattern for regex to find phone numbers like 666-232-3403 or (666) 232 3403 in text "Call us at 666-232-3403 or (666) 232 3403 and find your details".
   b. Data: " The unicorn is an element of a myth, it is magical and unique. An igloo, an island, and an umbrella are all interesting examples. " Construct a regex that captures words that start with a vowel (a, e, i, o, u), can have any number of characters from a-z after the vowel (including zero), and end with a consonant (any letters that are not a, e, i, o, u).
   c. Data = "The items are priced at $3.45, $23.32, and $400. Additional fees of $5.99 apply." Find the price item for example: $3.45 or $400.
   d. Data = {'company_code': ['Abcd','EFGF', 'zefsalf', 'sdfslew', 'zekfsdf'],
      'date_of_sale': ['12/05/2002', '16/02/1999', '25/09/1998', '12/02/2022', '15/09/1997'],
      'sale_amount': [12348.5, 233331.2, 22.5, 2566552.0, 23.0] }
      Create two new columns in the dataframe indicating the length of the string present in 'company code' and 'sale_amount'.
   e. In the above data set create a column that contain year on the date_of_sale (using regex)

2. WAP in Python using suitable data visualization technique to accomplish the following task.
   Data : Housing csv
   a. Plot the distribution of house prices in the dataset in 30 bins with blue color?
   b. Demonstrate relation between house prices and area (square feet) of the house using scatter plot?

c. Display the correlation between all numerical features such as price, area, bedrooms, etc..
d. Count houses with and without guest rooms and whether they are located on a main road or not. And visualize using heat map.
e. Plot furnishing status distribution using pie plot

3. WAP in Python using iris dataset (**from sklearn.datasets import load_iris**)

a. Plot the distribution of sepal lengths across all iris species?
b. Compare sepal area and petal area for each species using Scatter Plot.
c. Plot the pairwise relationships between all features for different iris species.
d. Demonstrate how are the different features correlated with each other.
e. Plot the average values for each feature in the iris dataset for each species using Bar Plot.