R. N. G. PATEL INSTITUTE OF TECHNOLOGY (RNGPIT)

ISROLI-AFWA BARDOLI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

YEAR- III SEMESTER – V

3150713- PYTHON FOR DATA SCIENCE

TERM- 211

LABORATORY MANUAL

Vision

To create a holistic teaching -learning ecosystem with a vision to prepare globally competent computer professionals for the development of society and the nation.

Mission

M1: To ensure effective and efficient teaching learning process for sound fundamentals.

M2: To impart technical skill through effective hands on practice on advanced technologies and Value education.

M3: To prepare industry ready professionals by industry partnerships and entrepreneurs activities.

M4: To inculcate human values, research capabilities, leadership abilities and ethics in young minds.

Program Educational Objectives

PEO1: The graduates will be able to interact with their peers in industries and society as an engineering professionals, entrepreneurs and leaders to set up technical ambience in the society

PEO2 The graduates will be able to reveal professionalism, ethical attitude, strong communication skills and maintain good team work spirit in their profession.

PEO3 The graduate will be able to utilize their skills with strong foundation to prepare them for higher learning and research activity.

PEO4: The graduate will be able to build up an ability to analyze the requirements, comprehend professional specifications, design and provide engineering

PROGRAM OUTCOMES

PO1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8 Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life- long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

COURSE OUTCOMES

1. Apply various Python data structures to effectively manage various types of data.
2. Explore various steps of data science pipeline with role of Python.
3. Design applications applying various operations for data cleansing and transformation.
4. Use various data visualization tools for effective interpretations and insights of data.
5. Perform data Wrangling with Scikit-learn applying exploratory data analysis.

LIST OF EXPERIMENTS

1. Develop Programs To Understand The Control Structures, Branching Programs, Strings
2. Develop programs to learn different types of structures (list, dictionary, tuples) in python
3. Setting up Python for Data Science
4. Uploading, Streaming, and Sampling Data Using Pandas
5. Connecting Pandas to a PostgreSQL Database with SQLAlchemy
6. Handling Missing Data
7. Working with Data Shaping
8. Working with Data Visualization
9. Working with Data Wrangling
10. Mini Project

INDEX

PEN:

NAME:

SUBJECT: Python for Data Science (3150713)

SR. NO.	DATE	TITLE	PG NO.	GRADE	SIGNATURE
1		Develop Programs To Understand The Control Structures, Branching Programs, Strings			
2		Develop programs to learn different types of structures (list, dictionary, tuples) in python			
3		Setting up Python for Data Science			
4		Uploading, Streaming, and Sampling Data Using Pandas			
5		Connecting Pandas to a PostgreSQL Database with SQLAlchemy			
6		Handling Missing Data			
7		Working with Data Shaping			
8		Working with Data Visualization			
9		Working with Data Wrangling			
10		Mini Project			

1.	**Develop programs to understand the control structures of python**
	1.1 Write a Python Program to find those numbers which are divisible by 7 and multiple of 5, between 1500 and 2700.
	1.2 Write a Python program to construct the following pattern, using nested for loop. *
	1.3 Write a Python program that accepts a word from user and reverse it (without using the reverse function)
	1.4 Write a Python program to check whether an alphabet is a vowel or consonant.
	1.5 Write a Python program to find reverse of given number.
	1.6 Write a Python program to check whether the given no is Armstrong or not using.
	1.7 To write a Python program to find first n prime numbers.
	1.8 Write a Python program to print Fibonacci series upto n terms.
	1.9 Give the output of following Python code:
	a) myStr = 'GTU is the best University' print myStr [15 : : 1] print myStr [-10 : -1 : 2]
	c) I=[(x, y) for x in [1,2,3] for y in [3,1,4] if x != y] print I
	d) str1 = 'This is Pyhton' print "Slice of String : ", str1[1 : 4 : 1] print "Slice of String : ", str1[0 : -1 : 2]
2.	**Develop programs to learn different types of structures (list, dictionary, tuples) in python**
	2.1 To write a Python Program to find the maximum from a list of numbers.

b)

	2.2 Write a Python program which will return the sum of the numbers in the array, returning 0 for an empty array. Except the number 13 is very unlucky, so it does not count and number that come immediately after 13 also do not count. Example : [1, 2, 3, 4] = 10 [1, 2, 3, 4, 13] = 10 [13, 1, 2, 3, 13] = 5
	2.2 Write a Python program which takes a list and returns a list with the elements "shifted left by one position" so [1, 2, 3] yields [2, 3, 1]. Example: [1, 2, 3] → [2, 3, 1] [11, 12, 13] → [12, 13, 11]
	2.3 Write a program to convert a list of characters into a string
	2.4 Write a Python program 1) To generate a list except for the first 5 elements, where the values are square of numbers between 1 and 30(both included) 2) To generate a list of first and last 5 elements where the values are square of numbers btween 1 and 30.
	2.5 Write a python program to print numbers given in the list after removing even numbers from it.
	2.6 Write a program to count the numbers of characters in the string and store them in a dictionary data structure.
	2.7 Write a program to use split and join methods in the string and trace a birthday with a dictionary data structure.
	2.8 Write a python program to sort a dictionary by value
3	**Setting up Python for Data Science**
	3.1 Installing Anaconda on Windows
	3.2 Working with Jupyter Notebook • Creating a new notebook • Opening existing notebooks • Using Google Drive for existing notebooks • Using GitHub for existing notebooks • Using local storage for existing notebooks • Saving notebooks Using Drive to save notebooks • Using GitHub to save notebooks
	3.3 Performing Common Tasks • Creating code cells • Creating text cells • Creating special cells Executing the Code • Viewing Your Notebook
	3.3 Defining the code repository • Defining a new folder • Creating a new notebook

	• Adding notebook content • Exporting a notebook • Removing a notebook • Importing a notebook
	3.4 Understanding the following datasets with its code. • load_boston(): Regression analysis with the Boston house-prices dataset • load_iris(): Classification with the Iris dataset • load_diabetes(): Regression with the diabetes dataset • load_digits([n_class]): Classification with the digits dataset • fetch_20newsgroups(subset='train'): Data from 20 newsgroups • fetch_olivetti_faces(): Olivetti faces dataset from AT&T
4.	**Uploading, Streaming, and Sampling Data Using Pandas**
	4.1 Uploading small amounts of data in Colors.txt into memory.
	4.2 Streaming large amounts of data in Colors.txt into memory.
	4.3. Retrieve every odd number record from the file Colors.txt (Data Sampling)
	4.4 Select random samples from the file Colors.txt
	4.5 Read the csv file named **titanic.csv** and print the values.
	4.6 Read the Excel file named **values.xls** file and parse the values and print it.
	4.7 Write a python script to read the image stored in local storage as well as on the specific URL. Also perform the following operations on image. a) Displaying the image information b) Cropping the image c) Resizing the image d) Flatening the image
	4.8 Read the data from the given XMLData.xml file
5.	**Connecting Pandas to a PostgreSQL Database with SQLAlchemy**
	5.1 Create a SQLAlchemy Connection with Postgres Database and retrieve data from the table in Postgresql
	5.2 Create a table in PostgreSQL and retrieve the data using **read_sql_query().**
	5.3 Create a SQL table from data in a CSV. The CSV containing **NYC job data**

	5.4 Create DataFrame from SQL Table and read the data using **read_sql().**
6	**Handling Missing Data**
	6.1 Find the Missing Data in the given file cars.csv and print it.
	6.2 Impute the missing data with all the methods (mean, median and most_frequent)
	6.3 Delete all the missing data entry in cars.csv
7.	**Working with Data Shaping**
	7.1 Apply bags of words model on the following statements. • Review 1: This movie is very scary and long • Review 2: This movie is not scary and is slow • Review 3: This movie is spooky and good
	7.2 Apply TF-IDF Model on the following statements. • Review 1: This movie is very scary and long • Review 2: This movie is not scary and is slow • Review 3: This movie is spooky and good
8.	**Working with Data Visualization**
	8.1 Plot a line graph: You have to pass only one list of two points, which will be taken as y axis co-ordinates. For x axis it takes the default values in the range of 0 to 1, 2 being the length of the list [5, 15] and plot the graph.
	8.2 Plot line graph with multiple lines with label and legend for the following values: Rice [5, 15] Oil [3, 6] Wheat [8.0010, 14.2] Coffee [1.95412, 6.98547, 5.41411, 5.99, 7.9999] Also Mark the Line graph with Marker
	8.3 Plot scatter with marker size of 20 and 10 for x1 and x2 for the following data: x1 = [214, 5, 91, 81, 122, 16, 218, 22] x2 = [12, 125, 149, 198, 22, 26, 28, 32]
	8.4 Plot pie graph for the following values and also turn on the axis of graph. India 31% Canada 19% Japan 15% Australia 14% Russia 21%

	8.5 Read the data from Topic_Survey_Assignment.csv file and Plot the bar graph. Also Put the percentage values on top of each bar
	8.6 Read the data from tips.csv file and Plot the box graph. plot day on X-Axis and ploy total_bill on Y-Axis
	8.7 Plot the Geographic Data from california_cities.csv file using basemap.
	8.8 Here is a Dataset of various Indian cities and the distances between them in edge_list.txt. Draw the Graph for the data.
9.	**Working with Data Wrangling**
	9.1 Load the Boston dataset from sklearn library concerns the housing prices in housing city of Boston. The dataset provided has 506 instances with 13 features. Split the data into training and testing sets. Train the model with 80% of the samples and test with the remaining 20%. • Predict the house prices for testing dataset.
10.	**Mini Project related to Data Science Applications.**