GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering
Subject Code: 3150713
Semester – V
Subject Name: Python for Data Science

Type of course: Open elective

Prerequisite: Programming concepts, Statistical and numerical methods

Rationale: The data collected by organization needs insights to take the decisions, for predictions as well as for finding hidden patterns inside the data. Python is an appropriate language supporting all the features and libraries to perform data science activates. This subject covers the overview of the python with emphasis on various python data structures and various libraries like Pandas, NumPy, Matplotlib for performing various data science function including data preparation, cleaning, exploratory analysis and visualization

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks				Total Marks
L	T	P	C	Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
2	0	2	3	70	30	30	20	150

Contents:

Sr. No.	Topics	Teaching Hrs.	Module Weightage
1	**Overview of Python and Data Structures:** Basics of Python including data types, variables, expressions, objects and functions. Python data structures including String, Array, List, Tuple, Set, Dictionary and operations them.	06	20
2	**Data Science and Python:** **Discovering the match between data science and python:** Defining the Sexiest Job of the 21st Century, Considering the emergence of data science, Outlining the core competencies of a data scientist, Linking data science, big data, and AI , Understanding the role of programming, Creating the Data Science Pipeline, Preparing the data, Performing exploratory data analysis, Learning from data, Visualizing, Obtaining insights and data products, Understanding Python's Role in Data Science, Considering the shifting profile of data scientists, Working with a multipurpose, simple, and efficient language, Learning to Use Python Fast ,Loading data, Training a model, Viewing a result. **Introducing Python's Capabilities and Wonders:** Why Python?, Grasping Python's Core Philosophy, Contributing to data science, Discovering present and future development goals, Working with Python, Getting a taste of the language, Understanding the need for indentation, Working at the command line or in the IDE, Performing Rapid Prototyping and Experimentation, Considering Speed of Execution, Visualizing Power, Using the Python Ecosystem for Data Science, Accessing scientific tools using SciPy, Performing fundamental scientific computing using NumPy, Performing data analysis using pandas, Implementing machine learning using Scikit-learn, Going for deep learning with Keras and TensorFlow, Plotting the data using	04	15

	matplotlib, Creating graphs with NetworkX, Parsing HTML documents using Beautiful Soup.		
3	**Getting Your Hands Dirty With Data:** **Understanding the tools:** Using the Jupyter Console, Interacting with screen text, Changing the window appearance, Getting Python help, Getting IPython help, Using magic functions, Discovering objects, Using Jupyter Notebook, Working with styles, Restarting the kernel, Restoring a checkpoint, Performing Multimedia and Graphic Integration, Embedding plots and other images, Loading examples from online sites, Obtaining online graphics and multimedia. **Working with Real Data:** Uploading, Streaming, and Sampling Data, Uploading small amounts of data into memory, Streaming large amounts of data into memory, Generating variations on image data, Sampling data in different ways, Accessing Data in Structured Flat-File Form ,Reading from a text file Reading CSV delimited format, Reading Excel and other Microsoft Office files, Sending Data in Unstructured File Form, Managing Data from Relational Databases, Interacting with Data from NoSQL Databases, Accessing Data from the Web. **Conditioning Your Data:** Juggling between NumPy and pandas, Knowing when to use NumPy, Knowing when to use pandas, Validating Your Data, Figuring out what's in your data, Removing duplicates, Creating a data map and data plan, Manipulating Categorical Variables, Creating categorical variables, Renaming levels, Combining levels, Dealing with Dates in Your Data, Formatting date and time values, Using the right time transformation, Dealing with Missing Data, Finding the missing data, Encoding missingness, Imputing missing data, Slicing and Dicing: Filtering and Selecting Data, Slicing rows, Slicing columns, Dicing, Concatenating and Transforming, Adding new cases and variables, Removing data, Sorting and shuffling, Aggregating Data at Any Level. **Shaping Data:** Working with HTML Pages, Parsing XML and HTML, Using XPath for data extraction, Working with Raw Text, Dealing with Unicode, Stemming and removing stop words, Introducing regular expressions, Using the Bag of Words Model and Beyond, Understanding the bag of words model, Working with n-grams, Implementing TF-IDF transformations, Working with Graph Data, Understanding the adjacency matrix, Using NetworkX basics.	10	30
4	**Data Visulization:** **Visualizing Information:** Starting with a Graph, Defining the plot, Drawing multiple lines and plots, Saving your work to disk, Setting the Axis, Ticks, Grids, Getting the axes, Formatting the axes, Adding grids, Defining the Line Appearance, Working with line style, Using colors, Adding markers, Using Labels, Annotations, and Legends, Adding labels, Annotating the chart, Creating a legend.	04	15

	Visualizing the Data: Choosing the Right Graph, Showing parts of a whole with pie charts, Creating comparisons with bar charts, Showing distributions using histograms, Depicting groups using boxplots, Seeing data patterns using scatterplots, Creating Advanced Scatterplots, Depicting groups, Showing correlations, Plotting Time Series, Representing time on axes, Plotting trends over time, Plotting Geographical Data, Using an environment in Notebook, Getting the Basemap toolkit, Dealing with deprecated library issues, Using Basemap to plot geographic data, Visualizing Graphs, Developing undirected graphs, Developing directed graphs.		
5	**Data Wrangling:**	06	20
	Wrangling Data: Playing with Scikit-learn, Understanding classes in Scikit-learn, Defining applications for data science, Performing the Hashing Trick, Using hash functions, Demonstrating the hashing trick, Working with deterministic selection, Considering Timing and Performance, Benchmarkin, with,timeit, Working with the memory profiler, Running in Parallel on Multiple Cores, Performing multicore parallelism, Demonstrating multiprocessing. **Exploring Data Analysis:** The EDA Approach, Defining Descriptive Statistics for Numeric Data, Measuring central tendency,Measuring variance and range ,Working with percentiles, Defining measures of normality, Counting for Categorical Data, Understanding frequencies, Creating contingency tables, Creating Applied Visualization for EDA ,Inspecting boxplots, Performing t-tests after boxplots, Observing parallel coordinates, Graphing distributions, Plotting scatterplots ,Understanding Correlation, Using covariance and correlation, Using nonparametric correlation, Considering the chi-square test for tables ,Modifying Data Distributions, Using different statistical distributions, Creating a Z-score standardization, Transforming other notable distributions.		

Reference Books:

1. Python for data science for dummies 2nd Edition, John Paul Mueller, Luca Massaron, Wiley
2. Programming through Python, M. T. Savaliya, R. K. Maurya, G. M. Magar, STAREDU Solutions
3. Pandas for everyone :Python Data Analysis, Daniel Y. Chen, Pearson
4. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, Arno D.B. Meysman, et al., Minning
5. Applied Data Science with Python and Jupyter: Use powerful industry-standard tools to unlock new, actionable insights from your data, , Packt
6. Data Analytics, Anil Maheshwari , McGrawHill
7. Data Science From Scratch: First Principles with Python, Joel Grus , SPD
8. Star Data Science Specialist, STAR CERTIFICATION

GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering
Subject Code: 3150713

Course Outcomes:

Sr. No.	CO Statement	Marks % weightage
CO-1	Apply various Python data structures to effectively manage various types of data.	20
CO-2	Explore various steps of data science pipeline with role of Python.	15
CO-3	Design applications applying various operations for data cleansing and transformation.	30
CO-4	Use various data visualization tools for effective interpretations and insights of data.	15
CO-5	Perform data Wrangling with Scikit-learn applying exploratory data analysis.	20

List of Practical:

Practical should be performed by students based on
- Use of Python Data Structures
- Using NumPy and Panda for Data Analysis
- Matplotlib for Visulization

Web Resources:
- www.anaconda.com
- www.python.org
- www.w3schools.com
- https://www.learnpython.org/
-