Course Code	Course Title	L	T	P	C
PMDS604L	**Exploratory Data Analysis**	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. The course introduces the methods for data preparation and data understanding.
2. It covers essential exploratory techniques for understanding multivariate data summarizing it through statistical and graphical methods.
3. Supports to summarize use of predictive analytics, data science and data visualization.

Course Outcomes

At the end of this course, students will be able to:

1. Understand various data formats, sources and storage mechanisms.
2. Prepare the missing data and manage data wrangling and manipulation
3. Explain and present the findings in the Data Sets, after the Analysis is complete
4. Demonstrate the data visualization and make interpretations
5. Construct the data story using various software tools.

Module:1	**Introduction to Exploratory Data Analysis**	**4 hours**

Data Analysis - Exploratory Data Analysis and Data Science Process - Responsibilities of a Data Analyst - Data Analytics vs. Data Analysis - Types of Data - Understanding Different Types of File Formats - Languages for Data Professionals - Overview of Data Repositories - Data Marts, Data Lakes, ETL, and Data Pipelines - Foundations of Big Data - Identifying Data for Analysis

Module 2	**Data Wrangling**	**4 hours**

Data Sources – Data Loading, Storage and File Formats - Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, interacting with Web APIs, Interacting with Databases – Data Wrangling - Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting - Tools for Data Wrangling - Data Cleaning and Preparation - Handling Missing Data, Data Transformation, String Manipulation

Module:3	**Data Analysis**	**4 hours**

Statistical summary measures, data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, contingency tables, n-D Statistical data analysis.

Module:4	**Outlier Analysis**	**4 hours**

Outliers and Outlier Analysis - Outlier Detection Methods - Proximity-Based Approaches – distance metrics, Mahalanobis distance, Outlier Detection in High-Dimensional Data.

Module:5	**Data Visualization**	**4 hours**

Intro to data visualization - Visualization Tools - Getting started with Tableau Desktop – Connecting to the dataset - Creating charts – Creating common visualizations (bar charts, line charts etc.) - Filtering and sorting data - Adding Titles, Labels, and descriptions - Publish your work to Tableau Cloud - Interactivity with text and visual tooltips - Interactivity with actions (filter, highlight, URL) – Assembling dashboards from multiple charts

Module:6	**Exploratory Visualization Techniques**	**4 hours**

Introduction to data Visualization libraries – Customizing plots for effective communication-Interactive visualization tools - Geographic visualization - Text and sentiment Analysis.

Module:7	Insights of Data Visualization	4 hours

Introduction to Power BI - Understanding Desktop - Understanding Power BI Report Designer - Report Canvas, Report Pages: Creation, Renames - Report Visuals, Fields and UI Options - Experimenting Visual Interactions, Advantages - Reports with Multiple Pages and Advantages - Pages with Multiple Visualizations - PUBLISH Options and Report Verification in Cloud - Adding Report Titles-Report Format Options.

Module:8	Contemporary Issues	2 hours
	Total Lecture hours	**30 hours**

Text Book(s)

1	McKinney, W., Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython, 2017, 2nd Edition, O"Reilly Media.
2	Suresh Kumar Mukhiya and Usman Ahmed, Hands-On Exploratory Data Analysis with Python, 2020, Packt Publishing.

Reference Book(s)

1	O"Neil, C., and Schutt, R., Doing Data Science: Straight Talk from the Frontline by, 2013, O"Reilly Media.
2	Alberto Ferrari and Marco Russo, Introducing Microsoft Power BI, 2016, Microsoft Press, Washington.
3	Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, The Big Book of Dashboards, 2017, John Wiley & Sons.
4	Ryan Sleeper, Practical Tableau, 2018, O" Reilly Media.
5	Roger F Silva, Business Intelligence Clinic: Create and Learn, 2018, Create and Learn

Mode of Evaluation: CAT, Assignment, Quiz and FAT

Recommended by Board of Studies	15-02-2024		
Approved by Academic Council	No. 73	Date	14-03-2024