MODULE FOR SUMMER INTERNSHIP PROGRAMME 2025
(BY **MANTRA ASSOCIATES & E&ICT ACADEMY IIT GUWAHATI**)
on
DATA ANALYTICS AND VISUALIZATION TECHNIQUES

OBJECTIVE:

1. To introduce the core concepts of data analytics and visualization, emphasizing practical applications.

2. To develop skills in data cleaning, preprocessing, and feature engineering for analytics tasks.
3. To teach the fundamentals of data visualization and storytelling using modern tools.
4. To explore statistical analysis and its role in drawing actionable insights from data.
5. To work with Python libraries (e.g., pandas, matplotlib, seaborn, Plotly) for data manipulation and visualization.
6. To implement end-to-end analytics workflows, including exploratory analysis, visualization, and reporting.

OUTCOME:

1. **Data Preparation**: Demonstrate the ability to clean, preprocess, and transform raw datasets for analysis.
2. **Statistical Proficiency**: Apply statistical concepts to interpret datasets and draw meaningful conclusions.
3. **Visualization Expertise**: Create engaging and interactive visualizations to effectively communicate insights.
4. **Tool Mastery**: Gain proficiency in Python libraries and visualization tools like Tableau and Power BI.
5. **Insight Generation**: Design analytics workflows to identify trends, patterns, and outliers in diverse datasets.
6. **Real-World Application**: Complete hands-on projects simulating real-world data analytics challenges.

DURATION: ONE MONTH **(60 HOURS)**

PREREQUISITES:

1. **Basic Programming Knowledge**: Familiarity with Python basics (e.g., loops, functions, and lists).
2. **Mathematical Foundations**: Understanding of basic statistics, probability, and algebra.
3. **Spreadsheet Skills**: Basic familiarity with tools like Excel for working with tabular data.

INTERNSHIP STRUCTURE BREAKDOWN

DAY NO. & DATE	TOPICS TO BE COVERED	TIME DURATION
DAY 1 (TUESDAY) 01-07-2025	Introduction to Data Analytics, Types of Data, Importance of Visualization.	2.5 HRS
DAY 2 (WEDNESDAY) 02-07-2025	Python basics for data manipulation, Working with Jupyter Notebooks, Installing libraries (pandas, matplotlib).	2.5 HRS
DAY 3 (THURSDAY) 03-07-2025	Introduction to pandas, Loading and exploring datasets, Handling missing data.	2.5 HRS
DAY 4 (FRIDAY) 04-07-2025	Data cleaning techniques, Handling categorical and numerical data, Outlier detection.	3.5 HRS (MCQ TEST 1)
DAY 5 (MONDAY) 07-07-2025	Basics of statistics: Mean, Median, Mode, Standard deviation, Variance.	2.5 HRS
DAY 6 (TUESDAY) 08-07-2025	Data grouping and aggregation, Pivot tables in pandas, Merging and joining datasets.	2.5 HRS
DAY 7 (WEDNESDAY) 09-07-2025	Introduction to data visualization, matplotlib basics, Line and bar charts.	2.5 HRS
DAY 8 (THURSDAY) 10-07-2025	Advanced visualization with seaborn: Heatmaps, Pair plots, Box plots.	2.5 HRS
DAY 9 (FRIDAY) 11-07-2025	Storytelling with data: Choosing the right chart, Color palettes, Annotation techniques.	3.5 HRS (MCQ TEST 2)
DAY 10 (MONDAY) 14-07-2025	Perform EDA and visualize trends on some well known dataset	2.5 HRS
DAY 11 (TUESDAY) 15-07-2025	Introduction to Plotly, Creating interactive line and scatter plots.	2.5 HRS
DAY 12 (WEDNESDAY) 16-07-2025	Dashboard building with Plotly Dash: Layouts, Callbacks, Interactivity.	2.5 HRS
DAY 13 (THURSDAY) 17-07-2025	Tableau basics: Loading datasets, Building simple dashboards.	2.5 HRS
DAY 14 (FRIDAY)	Advanced Tableau: Filters, Calculated fields, Actions.	3.5 HRS (MCQ TEST 3)

18-07-2025		
DAY 15 (MONDAY) 21-07-2025	Create an interactive visualization dashboard for some dataset.	2.5 HRS
DAY 16 (TUESDAY) 22-07-2025	Advanced topics: Time series analysis basics, Moving averages, Trend lines.	2.5 HRS
DAY 17 (WEDNESDAY) 23-07-2025	Case study: Visualizing world happiness trends using the **World Happiness Report Dataset**.	2.5 HRS
DAY 18 (THURSDAY) 24-07-2025	Analyzing air pollution trends using the **Air Quality Index Dataset**.	2.5 HRS
DAY 19 **(FRIDAY)** 25-07-2025	Data storytelling techniques: Combining narrative and visuals with insights from the **HR Analytics Dataset**.	3.5 HRS **(MCQ TEST4)**
DAY 20 (MONDAY) 28-07-2025	Hands-on project: Explore and visualize NYC Taxi data trends.	2.5 HRS
DAY 21 (TUESDAY) 29-07-2025	Predictive analytics introduction: Build a simple regression model using the **Ames Housing Dataset**.	2.5 HRS
DAY 22 (WEDNESDAY) 30-07-2025	Case study wrap-up: Perform a complete EDA and visualization workflow using the **Superstore Dataset**.	2.5 HRS
DAY 23 (THURSDAY) 31-07-2025	DOUBT CLEARING SESSION	**1 HR & VALEDICTORY SESSION**

PROJECTS TO BE ASSIGNED TO THE INTERNS (MIN. 10):

1. Analyze customer demographics using the **Titanic Dataset.**
2. Perform sales trend analysis with the **Superstore Dataset**.
3. Visualize global population growth using the **World Bank Dataset**.
4. Create a dashboard for COVID-19 data visualization using **Johns Hopkins Dataset**.
5. Identify factors influencing house prices using the **Ames Housing Dataset**.
6. Perform EDA and visualization on the **Iris Dataset**.
7. Build a correlation heatmap for features in the **Heart Disease Dataset**.
8. Analyze loan approvals using the **Lending Club Dataset**.
9. Visualize crime data trends using the **US Crime Rates Dataset**.
10. Analyze employee retention using the **HR Analytics Dataset**.
11. Create a visual report on traffic patterns using the **NYC Taxi Dataset**.
12. Explore air pollution data using the **Air Quality Index Dataset**.
13. Analyze and visualize world happiness rankings with the **World Happiness Report**.
14. Visualize stock market trends using **Yahoo Finance Stock Data**.
15. Build an interactive data dashboard using **Power BI or Tableau.**