AI-Driven Data Analytics Curriculum – Version 8
Module Number: *1* **Module Name:** *Excel*
Module End Assignment Title: *AI-Powered Healthcare Data Insights and Dashboarding*

Problem Statement:

The healthcare industry generates vast amounts of data daily, providing valuable insights for healthcare providers and policymakers to improve patient care, allocate resources effectively, and manage healthcare costs. This project aims to analyze a comprehensive healthcare dataset comprising medical examinations, hospitalization details, and customer profiles to extract insights into patient health profiles, medical histories, and healthcare costs. By exploring relationships between various health metrics, identifying trends, and visualizing key patterns, we aim to deliver actionable insights to healthcare stakeholders for informed decision-making through rigorous data cleaning, transformation, exploration, and analysis.

Dataset Download:

https://drive.google.com/uc?export=download&id=1zelh7bZrE7F290QtTABHgHYn4B7JDbZO

Project Steps and Objectives:

Data Cleaning: *(5 marks)*

1) Check for the number of missing values marked with '?' in each column of the "Medical Examinations" table and "Hospitalization Details" table.

2) Fill in the missing values of 'month' with Sep and 'year' with its average, rounded to the nearest integer.

3) Determine the most frequently occurring values in the 'smoker', 'hospital tier', and 'city tier' columns, and fill in the missing values accordingly.

4) If any 'State ID' values are missing, consider filling them with 'Unknown' or using another appropriate strategy.

Data Transformation: *(5 marks)*

1) Split the 'names' column in the "Customer Names" table into 3 meaningful columns: 'Title', 'First Name', and 'Last Name'.

   2) Convert the "NumberOfMajorSurgeries" column in the "Medical Examinations" table to numerical data by replacing non-numeric characters with meaningful numerical values.

3) Check for inconsistencies in the 'Heart Issues' and 'Smoker' columns and propose corrective actions if necessary.

4) Create a new column named "Weight Status" that categorizes BMI into different categories as below:

BMI	Weight Status
Below 18.5	Underweight
18.5 – 24.9	Normal Weight
25.0 – 29.9	Overweight
30.0 and Above	Obesity

5) Merge the 'year', 'month', and 'date' columns in the "Hospitalization Details" table into one column named 'Date of Birth' and format it in the 'DD-MMM-YYYY' custom format.

7) Calculate the 'Age' of each customer based on their 'Date of Birth' and the date of collection of the dataset, which is 8th June 2023.

8) Format the 'charges' column as currency ($).

AI-Driven Data Exploration, Analysis & Visualization: *(15 marks)*

Activity 1: Data *Preparation – (2 marks) (should be done manually)*

➢ Create a new sheet named "Healthcare", and combine all three tables into one, using Customer ID as the common column, utilizing XLOOKUP/Data Model manager.

➢ Retain the following necessary columns: Customer *ID, First Name, BMI, HBA1C, Heart Issues, Any Transplants, Cancer History, Number of Major Surgeries, Smoker, Weight Status, Diabetes Status, Date of Birth, Charges, Hospital Tier, City Tier, State ID, and Age*.

Activity 2: Manual Dashboard Build (5 marks) *(should be done manually)*

● Create a pivot table manually (Insert → PivotTable) using the combined healthcare dataset.
● **Add charts:**
  ○ *Pie/Donut Chart:* Cancer history distribution among smokers vs non-smokers.
  ○ *Column Chart:* Charges by Weight Status.
● ***Draw the insight manually*** *and marked it down in the sheet*

Activity 2: AI-Powered Summary Statistics & Visualization(5 marks) *(AI-driven task)*

Automatically generate and visualize overall healthcare statistics.

1. Highlight the combined dataset (new sheet named "Healthcare").
2. Use **Quick Analysis**→ to calculate dynamic counts and averages
   ● Total customers count,
   ● Total hospitalisation count,
   ● Total healthcare costs,
   ● Average costs and medical metrics

3. Use **Quick Analysis → Recommended Charts** (Column/Bar/Doughnut) to **visualise distributions** of all pivots created in the previous step using AI tools.

- **Expected Output:**
  - Pivot chart/quick analysis chart visualising totals.

Activity 4: Dashboard Creation (3 marks)

- Build an interactive dashboard that consolidates all key insights using the above visualisations. Ensure visual clarity and ease of interpretation for all chart types.
- *Insert slicers* **for weight status and diabetes status,** and wherever possible.
- Arrange charts and slicers into a simple dashboard layout
- Add a heading (e.g., *Healthcare Dashboard – AI Analysis*).
- Ensure consistency in chart formatting for clear interpretation.