

Hamma Lakhdar University

Master 2 – Artificial Intelligence and Data Science (IA & Data Science)

Big Data TP3 – Healthcare Data Analysis with Power BI

Prepared by: **Hassan Zuebidi**

Academic Year: 2025

Date: November 2025

1. Introduction

This practical session (TP3) focuses on analyzing healthcare-related data using Power BI. The goal is to extract meaningful patterns and visualize relationships between medical indicators such as age, blood pressure, cholesterol level, and diabetes status. Unlike real hospital records, the dataset used here was **generated using Artificial Intelligence (AI)** to simulate realistic but privacy-safe patient information. It allows students to explore data analysis and visualization techniques without any ethical or confidentiality concerns.

2. Objectives

- 1 Understand how to import, clean, and prepare healthcare data in Power BI.
- 2 Build meaningful KPIs and dashboards for medical insights.
- 3 Interpret the correlation between age, BMI, blood pressure, and diabetes.
- 4 Identify seasonal variations in cholesterol levels throughout the year.
- 5 Demonstrate storytelling and data interpretation using Power BI visuals.

3. Dataset Description

The dataset, named **healthcare_dataset.csv**, contains 1,200 synthetic patient records generated through AI-based data synthesis. Each record includes demographic information and health indicators. The variables include: **PatientID**, **Age**, **Gender**, **BMI**, **BloodPressure**, **CholesterolLevel**, **HasDiabetes**, and **VisitDate**. The time range covers medical visits between January 2023 and November 2024. All fields were complete with no missing values, ensuring clean analysis readiness.

4. Tools and Technologies

The main tool used in this TP is **Microsoft Power BI Desktop**. It allows data modeling, visualization, and interactive dashboard creation. Additional technologies and expressions used include: DAX formulas for creating calculated columns and measures. Power Query Editor for

cleaning and transforming the dataset. Dynamic slicers to filter by Age, Gender, Year, Month, and BMI. Cards, line charts, and pie charts to illustrate relationships.

5. Methodology

The analysis followed a structured workflow: Imported the dataset (CSV file) into Power BI. Converted the VisitDate into Date format and extracted **Month Name** and **Year**. Created DAX measures such as **Average Age** and **Diabetes Rate**. Designed KPIs and charts to summarize key health metrics. Added slicers for interactivity and user-driven exploration.

6. Results and Discussion

The final dashboard displayed several key findings: **Average Age**: 52.7 years – the dataset represents a balanced middle-aged population. **Diabetes Rate**: 25.9% – roughly one in four patients had diabetes. **Gender Distribution**: 50.7% Female and 49.3% Male, showing a balanced sample. **Blood Pressure**: Increased gradually with age, consistent with medical expectations. **Cholesterol Levels**: Showed clear seasonal patterns. The **highest cholesterol level** was recorded in **January (13.8K)**, followed by October (11.2K). This spike corresponds with **New Year celebrations**, when people typically consume rich, fatty foods such as roasted turkey, pork, and butter-based desserts. The lowest levels appeared in **May (8.4K)**, suggesting post-holiday dietary moderation.

7. Conclusion

This TP demonstrated how Power BI can be effectively used to analyze and visualize healthcare data. Although the dataset was AI-generated, the patterns observed are realistic and educational. It showed the correlation between age and blood pressure, the prevalence of diabetes, and the influence of seasonal habits on cholesterol levels. Power BI proved to be a powerful tool for transforming raw data into meaningful insights through interactive dashboards.

8. Recommendations

- 1 Include geographical data for mapping healthcare trends.
- 2 Add correlation visuals between BMI, blood pressure, and diabetes.
- 3 Develop predictive models using Power BI or Python integration.
- 4 Expand the dataset to include more years for trend analysis.

9. Executive Summary

The TP3 Healthcare Analysis project demonstrated how AI-generated medical data can be transformed into actionable insights using Power BI. Through KPI cards, line and pie charts, the report highlighted health indicators across age and gender, and revealed seasonal cholesterol variation—particularly the spike in January due to high-fat festive diets. The work shows that, with proper data visualization and analytical reasoning, even synthetic data can be a valuable educational tool in the Big Data and Health Analytics field.