# **Hospital Management Data Transformation Challenge**

## **Background**

You have been provided with a Hospital Management Excel dataset containing information about patients, doctors, treatments, billing, and appointments. However, as with many real-world datasets, this information contains several data quality issues that need to be addressed before meaningful analysis can be performed.

## **Task**

Using Power Query Editor in Power BI, perform the necessary data transformations to clean and prepare the dataset for analysis. You will need to address the following common data issues:

1. Handle Missing Values

* Identify and address null or blank values in patient names and doctor names
* Develop an appropriate strategy for missing billing amounts (consider using averages, medians, or flagging records)
* Document your approach for handling each type of missing value

2. Remove Duplicate Records

* Identify and eliminate duplicate patient records
* Create a log showing how many duplicates were removed
* Ensure unique identifiers are maintained throughout the dataset

3. Fix Inconsistent Formatting

* Standardize date formats across appointment and admission dates
* Remove extra spaces from text fields
* Correct inconsistent capitalizations in names and departments
* Normalize phone number formats

4. Address Outliers

* Identify statistical outliers in billing amounts and patient age
* Determine whether outliers represent errors or legitimate values
* Implement an appropriate strategy for handling each type of outlier

5. Correct Data Types

* Ensure all columns have appropriate data types (e.g., dates as date type, numbers as numeric)
* Convert any text stored as numbers or dates stored as text
* Implement data type validation checks

## **Dataset**

<https://docs.google.com/spreadsheets/d/1l0J2eHi3VIki7XrTWpApo6Dv2ydRm2c0/edit?usp=drive_link&ouid=103661808625668826597&rtpof=true&sd=true>