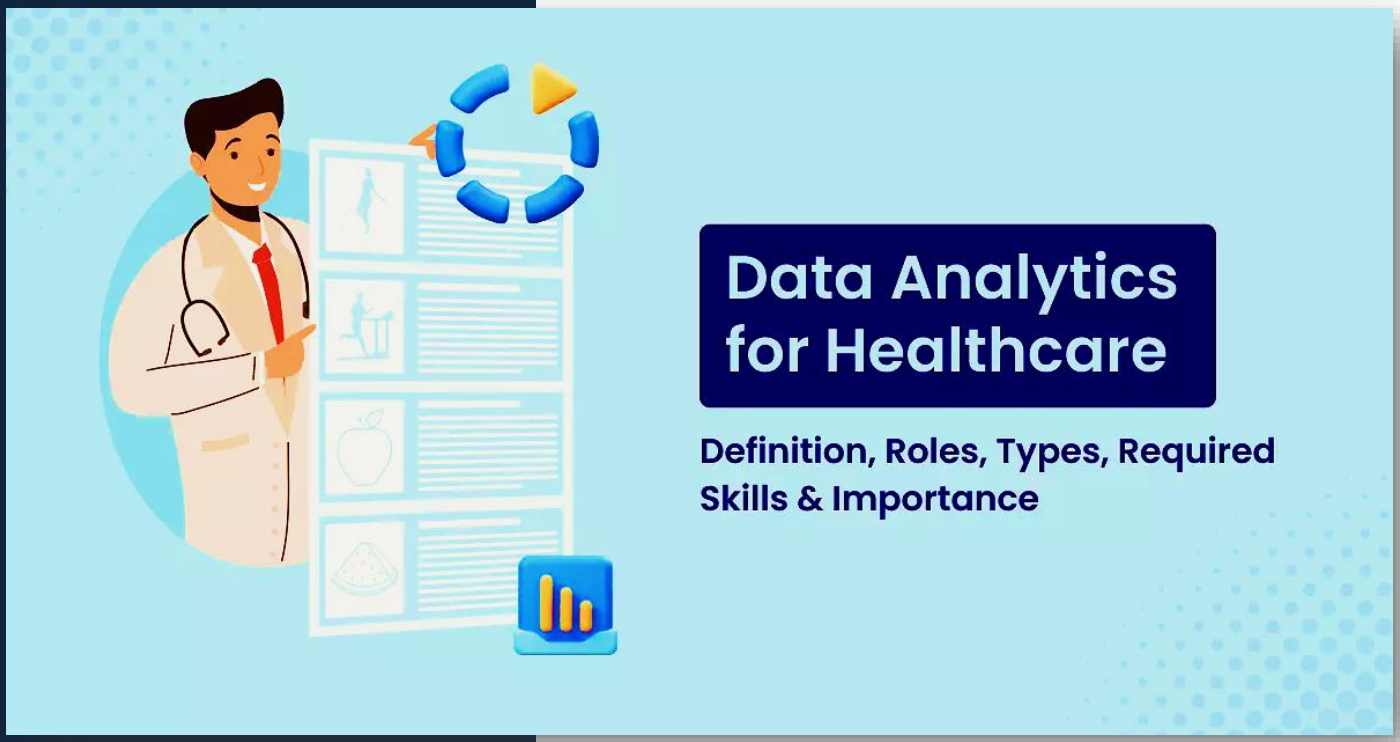
رواد مصر الرقمية

وزارة الاتصالات وتكنولوجيا المعلومات

Healthcare Data Analysis Project



م	الاسم	QR
1	محمد محمود فاروق أبوزيد	
2	صفا احمد صلاح خميس	
3	هاجر محمد حامد الفقي	
4	احمد رجب عبدالله السيد	
5	حسام الدين رفعت محمد سيداحمد	
6	مصطفى كارم محمود على	

Index

Healthcare Data Analysis Project:

Executive Summary

The report provides an in-depth and comprehensive analysis of a healthcare dataset containing patient records from multiple hospitals across the United States. The project employs leading data analytics techniques to identify patterns in demographic, clinical and financial data to support better decision making in healthcare.

Project Overview

Purpose of the Project

The overall objective of this study is to perform a comprehensive analytical evaluation of patient record data to derive actionable insights into healthcare delivery models, patient medical outcomes and related financial results. In this paper, an attempt is made to perform an analysis on the correlations between the demographic (age, sex), clinical (chronic disease, BMI) and financial (billing amounts, length of stay) variables.

Dataset Description

The dataset contains patient records from various states and hospitals in the US, with a sample of 28 records (which can be scaled to thousands of records for the full analysis). The data is structured in a table where each row corresponds to one patient record, which includes unique identifiers, demographic information, health indicators, lifestyle factors, clinical data, and financial information. Patient anonymity is preserved by the de-identification of data and by concentrating on aggregate trends in order to meet privacy requirements like HIPAA.

Key Columns in Detail

- **PatientID**
- **Demographics**:
    - State: The location of the hospital (e.g., Massachusetts, Washington, Indiana).
    - Hospital: The name of the institution (e.g., Sons and Miller, Kim Inc).
    - Sex: Male or Female.
    - Age:
    - AgeCategory: Classifications such as “Age 30 to 34” or “Age 80 or older.”
    - HeightInMeters:

- WeightInKilograms:
- RaceEthnicityCategory: Categories such as “White only, Non-Hispanic” or “Hispanic.”

- **Health Indicators**:
  - GeneralHealth: Ratings such as “Excellent,” “Good,” or “Poor.”
  - BMI: Automatically calculated from height and weight (e.g., 30.73, indicating obesity if >30).
  - Blood Type: E.g., “B-,” “A+,” “AB+.”
  - Binary indicators (0/1 or No/Yes)
    - for chronic conditions:
      - HadHeartAttack .
      - HadAngina HadStroke .
      - HadAsthma .
      - HadSkinCancer .
      - HadCOPD (chronic obstructive pulmonary disease).
      - HadDepressiveDisorder (depression).
      - HadKidneyDisease .
      - HadArthritis .
      - HadDiabetes .
    - Sensory/Functional Difficulties:
      - DeafOrHardOfHearing (deafness or hearing difficulty).
      - BlindOrVisionDifficulty .
      - DifficultyConcentrating DifficultyWalking .
      - DifficultyDressingBathing DifficultyErrands (difficulty performing daily errands).
- **Lifestyle Factors**:
  - SmokerStatus: E.g., “Never smoked,” “Former smoker,” “Current smoker.”
  - ECigaretteUsage: E.g., “Never used e-cigarettes,” “Not at all (right now).”
  - AlcoholDrinkers: Binary (1 for Yes, 0 for No).
  - HIVTesting: Binary.
  - FluVaxLast12: Binary (recent flu vaccination).
  - PneumoVaxEver: Binary (pneumonia vaccination ever received).
  - TetanusLast10Tdap: Details such as “Yes, received Tdap” or “No.”
  - HighRiskLastYear: Binary.
  - CovidPos: Binary (COVID-19 positivity).
  - ChestScan: Binary.

- **Admission and Clinical Information**:
  - Date of Admission: MM/DD/YYYY format .
  - Admission Type: “Urgent,” “Emergency,” or “Elective.”
  - Discharge Date: MM/DD/YYYY format.
- **Financial and Outcome Data**:
  - Billing Amount: Monetary value in USD .
  - Length of Stay: Calculated as the difference between admission and discharge dates (in days).

Tools and Technologies

The project incorporates several tools to handle data from the raw form to the end visualizations, focusing on seamless integration and performance.

- **Microsoft Excel:** Used to import initial data, cleaning (removal of repetitions, error correction, deal with lost values), and accounts.
- **Power BI:** Used for data modeling, complex accounts with DAX, interactive information
- **Tableau:** It is used for more advanced perceptions