**HEALTHCARE DATA ANALYSIS**

**STEPS FOLLOWED**

**Step 1** : I created a hcare database in SQL Server and imported the data

**Step 2** : **Understanding the data**

1. The table had 15 columns and 55,505 rows
2. The dataset consisted of a single table with columns containing details about patients, including their age, gender, blood group, medical condition, date of admission, treating doctor, hospital of admission, insurance provider, billing amount, room number, type of admission, discharge date, medications prescribed, and test results.

**Step 3 :** **Checked for abnormalities in the data**

The dataset contained 1,078 rows, including duplicates, with accurate data types across all columns. However, the following issues were identified and addressed:

1. Billing\_Amount Column : Contained one NULL value and excessive decimal places.
2. Name Column : Some names were improperly formatted, with inconsistent upper and lower case letters, and included titles.
3. Hospital Column: Entries had formatting issues, such as leading and trailing 'and', as well as commas within names.
4. Data was filtered to ensure records were properly formatted and free from inconsistencies

**Step 4 : Cleaning the data**

1. Handling Duplicates :

* Added a new column in the table healthcare1 to identify duplicate records.
* Created a separate table containing only duplicate rows, ordered using a clustered index.
* Deleted alternate rows from the duplicate table and removed all duplicates from the healthcare1 table.
* Combined the cleaned tables using UNION ALL, effectively eliminating duplicated rows.

2. Addressing NULL Values :

* Replaced NULL values in the Billing\_Amount column with the average billing amount.

3. Name Formatting :

* Removed titles from names.
* Converted all names to lowercase and capitalized the first letter of each name.

4. Cleaning Hospital Names **:**

* Removed unwanted commas from hospital names.
* Eliminated unnecessary occurrences of "and" at the beginning or end of hospital names.

5. Standardizing Billing Amount :

* Rounded values in the BIling\_Amount column to 2 decimal places.

**Step 5 : Assigning patient\_id for each patient**

**Step 7 : Creating dimensional tables**

Designed and developed dimensional tables, including DimBlood\_Type, DimAdmissionType, DimAgeCategory, DimInsuranceProvider, DimMedicalCondition, DimMedication, and DimTestResults, with primary key IDs assigned to each table for efficient data management and integration.

**Step 8 : Primary Key**

Created a primary key, Patient ID, for the table to ensure unique identification of each record and maintain data integrity.

**Step 9 : Foreign keys**

* Establishing foreign key relationships between the FactHealthcareData table and the dimensional tables by linking their respective primary keys to ensure referential integrity and enable efficient data retrieval and analysis
* This step created a star schema (one to many relationship)

**Step 10 : Visualisation**

* Used Power BI to create insightful and interactive data visualizations
* Removed unwanted columns from fact table such as Blood Type, Admission Type, Age Category, Insurance Provider, Medical Condition,Medication and Test Results(as these are not supposed to be in Fact Table)
* Developed measures in Power BI to calculate the total number of patients and the average billing amount
* Created a calculated column in Power BI using the DATEDIFF function to determine the length of stay and subsequently developed a measure to calculate the average length of stay.
* Utilized various visualizations, including bar charts, column charts, pie charts, donut charts, treemaps, cards, and slicers, to perform analyses effectively and present data insights appropriately.