|  |  |
| --- | --- |
| **HEALTH CARE ANALYSIS**    Using MySQL to practice, develop and showcase data analysis skills in the context of the healthcare industry |  |

**📝 PROJECT SUMMARY:**

This project contains data from 28.697 patient. Through data analysis, I wish to identify trends in medical conditions by age and gender, billing patterns and insurance provider.

**🎯DATASET:**

I downloaded the healthcare dataset from Kaggle.

Here is the link: <https://www.kaggle.com/datasets/prasad22/healthcare-dataset>

**Columns informations:**

* **Name:** This column represents the name of the patient
* **Age:** The age of the patient at the time of admission, expressed in years.
* **Gender:** Indicates the gender of the patient, either "Male" or "Female."
* **Blood Type:** The patient's blood type, which can be one of the common blood types (e.g., "A+", "O-", etc.).
* **Medical Condition:** This column specifies the primary medical condition or diagnosis associated with the patient.
* **Date of Admission:** The date on which the patient was admitted to the healthcare facility.
* **Doctor:** The name of the doctor responsible for the patient's care during their admission.
* **Hospital:** Identifies the healthcare facility or hospital where the patient was admitted.
* **Insurance Provider:** This column indicates the patient's insurance provider
* **Billing Amount:** The amount of money billed for the patient's healthcare services during their admission
* **Room Number:** The room number where the patient was accommodated during their admission.
* **Admission Type:** Specifies the type of admission, which can be "Emergency," "Elective," or "Urgent"
* **Discharge Date:** The date on which the patient was discharged from the healthcare facility, based on the admission date and a random number of days within a realistic range.
* **Medication:** Identifies a medication prescribed or administered to the patient during their admission
* **Test Results:** Describes the results of a medical test conducted during the patient's admission. Possible values include "Normal," "Abnormal," or "Inconclusive," indicating the outcome of the test.

**🔹 STEPS AND ANALYSIS:**

1. Data Loading

* Create a new schema, load the dataset
* Create a staging table
* Explore the dataset

2. Data Cleaning 🛠

* Finding the duplicates
* Handling missing values
* Delete less relevant columns for efficiency

3. Exploratory Data Analysis **📊**

**📌** Most Common & Least Common Condition

* Most common condition: Arthritis (4867 cases)
* Least common condition: Obesity (4709 cases)
* The differences between conditions are small, meaning all these conditions need similar attention in healthcare planning.

**📌** Find the common Medical Condition per year

* Based on my findings, there were 28.697 patients admitted and diagnosed with a medical condition.
* Year 2020 having the highest amount of admission. Cancer, obesity and hypertensions recorded the highest rates that year.

**📌** Which age group is most affected by each condition?

* Middle aged group occupies the first position, for all medical conditions.
* Middle aged individuals (40-64) might have the highest count because they are at the peak risk for lifestyle-related diseases like obesity, hypertension and diabetes.
* Second place is held by Adults, and the last place by Seniors, both affected by Arthritis. Adults (18-39) are generally healthier, but still susceptible to common illnesses.

**📌** Medical Condition per gender

* There are no major differences between genders.

**📌** What are the top conditions for males and females across different age groups?

* For both, female and male, arthritis, cancer and hypertension might be dominant in adults groups.
* Hypertension and arthritis might be dominant in middle-aged groups.
* Seniors likely have chronic conditions like arthritis, asthma

**📌** Average Billing for Admission Type

* Elective admissions have the highest average cost (~$25,800), slightly above Emergency and Urgent admissions.

**📌** What is the medical bill for each medical condition?

* Based on my analysis, the highest billing amount was for arthritis.

**📌** Show top 3 most expensive conditions per admission type

* Elective: cancer, diabetes, hypertension
* Emergency: asthma, arthritis, obesity
* Urgent: diabetes, arthritis, hypertension

📌 What is the average billing amount per provider?

* Highest Billing: Aetna ($25.740)
* Lowest Billing: Blue Cross ($25,537)

🔍 **KEY FINDINGS**:

1. Arthritis is the most common condition. Knowing this information helps hospital prioritize resources and treatment focus.
2. Year 2020 having the highest amount of admission.
3. Middle-aged people might have the most cases for all medical conditions
4. Elective admissions have the highest average costs (~$25,800).
5. Insurance Provider Aetna has the highest average billing ($25.740).