Proposal

Team Beta

Dataset to be Used: https://www.kaggle.com/code/malaychand/healthcare-dataset/input
Project Title: Patient Cost Optimization
Project Description/Outline: Using patient information to provide recommendations to potential patients broken down by cost

Breakdown of Tasks:
-Data Clean
    -Delete Potential Duplicates

Display demographic information
-Table Information:
    -Current year (2024)
    -Name
    -Age
    -Gender
    -Blood Type
    -Medical Condition
    -Hospital
    -Billing Amount
    -Date of Admission
    -Discharge Date
    -Test Results
    -Admission Type

Cost Considerations:
*Billing Cost divided by Duration of Stay
*Averages of Billing Cost by Patient Count at medical facility

Visualizations:
Hospital vs Billing Cost
Medical Condition vs Billing Cost
Admission Type vs Cost

Conclusion: Recommending potential patients a hospital to go based on cost

Research Questions to Answer:
1) Find out Hospital Recommendations based on Cost based on patient information
2) Cost Ranges per Admission Type
3)

Source:

Healthcare dataset from Kaggle

Focuses on patient info (age, name, sex), Medical info (blood type, medical condition, admission type, room number, doctor, hospital) and dates (admission date, departure date)

Changed admission and released to date time format because they were originally objects

Filtered the admission date to only include 2024 as the year

Datasheet went from having over 55,000 different rows of data to under 5,000

Removed two columns: medication was not related to their medical condition and room number because they were not pertinent to what we were trying to discover. Medication was inconclusive whether it was related to their medical condition.

We then determined each patient's length of stay with a formula

Visualizations:

Average Billing Costs by Hospital:

- Used to see which hospital could be seen as the most expensive
- Lamb Ferguson is represented as the highest average billing cost per patient

(If used) Average Billing Costs by Hospital (Descending)

- Used to see which hospital could be seen as the least expensive
- For negative values:
    - Patients may be uninsured (hospitals lose money)
    - Hospitals have financial assistance programs to help low-income patients

Total Days per Hospital:

- Hospitals with the highest TOTAL number of days
- According to the dataset, PLC Davis had the most occupancy

This could suggest

    - Less room for new patients

- Busier Doctors, nurses, and other staff
- Atmosphere may inhibit the patients' recovery

Total Billing Cost per Hospital

· Used to see which hospital made the most in the year 2024

· Smith PLC made the most in the year with close to $175,000.00

Admission type vs. Average Billing Cost

· Elective, Emergency, and Urgent

· Very close but Urgent has the highest average billing cost

Blood Type vs. Average Billing Cost

· A+ blood type had the highest average billing cost

· O+ most common blood type

· AB negative, B negative some of the rarest blood types in the US

· Lack of correlation, it could be said rare blood types could result in a higher billing cost

Distribution of insurance providers

· Very close but Cigna is the most common insurance provider at 20.3% of patients having it

· Least popular is Aetna at 19.7%

Average Billing Costs by Insurance Provider

· Medicare is shown to have the highest average billing cost per patient

· Little variance in cost between insurance providers

Age vs Billing Cost

· Most results fall between a billing cost of $2,000 and $4,000 across all ages with some outliers

Average Length of Stay by Medical Condition

· Little variance of average length of stay between admission type (all were between 15-16 days)

- Elective admission had the highest average stay

Average Length of Stay by Medical Condition

- Believe it or not, arthritis had the highest average length of stay at over 16 days
- Cancer had the lowest average length of stay
- Across all medical conditions, the average length of stay was between 14 - 17 days
    - It could be inferred that patients with the conditions reflected on the dataset should prepare to spend a similar amount of time in the hospital