Presentation

Introduction:

Good evening,

We'll be presenting an analysis on a healthcare dataset. I'm going to briefly give an overview on our data set and our objective.

This dataset contains just over 55,000 patient records and 15 columns, capturing various aspects of patient information and hospital data, including:
Demographic Information, Medical Information, and Billing/Insurance Information

And looking through the dataset, we realized that there are some limitations. A couple of them being that we don't know the level of insurance that the patients had, nor the itemized expenses in the billing amount.

But with those limitations in mind, we asked these questions as we analyzed this data set

1. What is the impact of a patient's information on their billing amount
2. What is the relationship between different elements of the dataset
3. Based on our findings, who could best utilize the analysis we provided

---

SOPS (Matt): Technical Aspects of the Data set

Data cleaning: The original dataset retrieved was a healthcare data set of hospital/clinic operations with the focus on treating Diabetes, Hypertension, Cancer, and Asthma. The first step of our analysis was cleansing the data. We first tested the code to ensure our selected year of 2024 revealed no errors and to ensure the data only reflected 2024 results. After cleansing the file, we began removing unnecessary columns that were not required for our analysis goal. Lastly, we performed a check for duplicates then began analyzing the remaining data.

(Rosy slide):

We used this chart to see which hospital can be seen as the most expensive from a patient's perspective. We can assess that hospital costs all vary and that Lamb Ferguson is the most costly amongst other hospitals. Averaging a higher charge to insurance providers along with its patients.

(Chris Slide):

On this slide it indicates hospitals with the highest numbers of stays within the 30 day average. In our dataset we've concluded that the average number of stays is approximately a month. With that, we can conclude for people who are in need of urgent or emergency stays, which hospitals will be able to accommodate them.

(Andre slide):

Over here we can see the cost differences amongst the three different kinds of admission types. Urgent being the highest among the three. The difference we've indicated between Urgent and emergency is time and sensitivity. Urgent being something like the ICU unit and emergency being the emergency room for paramedics. Main reason elective admissions cost almost as high as two imperative admission types is being within insurance providers elective treatments that are not considered medically necessary, which means patients will more than likely pay out of pocket. (Cosmetic surgeries are a good example)

(Andre slide): On this slide it indicates the mean of ages and which insurance providers cover them. We've figured out the average age in the data set is people in between the 50's and 60's. With CIgna being the highest of insurance coverage. Which is plausible because Cigna is within the same parent group as Medicare. Difference being CIgna is typically a plan whenever someone chooses a medical plan with Medicare advantage, supplement, or higher premiums.

(Sara slides):

1. This pie chart is showing The distribution of the provider insurance that are used across hospitals. From there we can see that they are close to each other.

- Cigna is the most used by patients across hospitals.
- Aetna is the lowest used by patients across hospitals.

2. This bar chart is showing: Average billing Costs that are covering each insurance.

The billing amounts for the five insurance providers are quite close to each other medicate is providing the most coverages for medical expenses.

(Fun facts on the insurance companies for **Sara**):
Medicare: Medicare is a federal health insurance program in the United States for people age 65 or older and younger people with disabilities.

- Because it is government funded it can explain why not too many of the elderly in our data set in the ages of 65 and older use medicare. The co pay must be higher and because of that they opted into using more advantageous plans like cigna the others.

- Bluecross blue shield and Aetna tend to be more of use because of that with the young as well.

-It can also explain why for (Sara second slide) medicare has a higher billing amount, since it is known to cover the elderly. Which is plausible to say they need extra treatment.

#10 slide (Alma) On this slide we have the blood type vs Avg Billing Cost. In this visualization, we are looking to see if a rarer blood type may influence a patient's billing cost by a noticeable margin. So, what we found was that, despite AB - and B - being the rarer blood types, our data suggests that it does not necessarily correlate to a higher billing cost.

(Jackson Slide):
In this slide the scatter plot showcases how despite the age there is no correlation to the Billing amount. Like in the previous slide when it comes to the billing amount it varies more on the patient's medical condition along with the insurance provider that the patient has.These outliers may reflect unique treatments, limited care, or specific insurance factors.Additional analysis is needed to determine the impact of factors like treatment type, insurance coverage, or medical conditions on these costs.

Last Slide analysis:

In our final analysis we've noticed that the data set is limited. However, it did not stop us from searching for our end goal. This data set can be used for patients, insurance providers, and hospitals. In this data set we were able to find which insurance provider's may be the best fit for lower billing amounts for patients alongside which hospitals can be a better fit for those in need of immediate accommodation and care. For hospitals and insurance providers it can give insight on which admission type along with which medical condition was more profitable for them. We calculated the majority of the correlations by finding the mean, and by doing so it can help them select their target demographic. Which condition and based on admission type they will profit off by.

Final Report:

Summary Report

1. **What is the impact of a patient's information on their billing amount?**
    - There is no strong correlation between patient-specific details (e.g., age, gender) and billing costs, with the exception of patients aged 60+ who have arthritis. Billing appears standardized regardless of these factors.
2. **What are some relationships between different elements of the dataset?**
    - The dataset did not reveal substantial relationships between elements, such as insurance providers, blood types, or admission types. Information is inconclusive and lacks enough depth to draw definitive conclusions.
3. **Based on our findings, who could best utilize the analysis we provided?**
    - This dataset can be useful for patients, insurance providers, and hospitals. Patients may gain insights into cost-effective insurance options and hospital choices. Insurance providers and hospitals can identify trends in admission types or conditions that drive profitability, though more detailed data is needed for comprehensive recommendations.

Summary of Key Findings

Average Billing Costs by Hospital

Lamb Ferguson Hospital has the highest average billing costs from a patient's perspective.

Insurance Providers Distribution

Cigna is the most common insurance provider (20.3%), while Aetna is the least common (19.7%).

Length of Stay (LOS)

Most stays last 15-16 days, with arthritis requiring up to two weeks on average. Elective admissions account for 52% of stays and require more observation and recovery time.

Impact of Admission Type on Costs

Urgent admissions have the highest average billing costs compared to elective and emergency types.

Blood Type and Billing

A-blood type patients incur the highest costs, but no clear correlation exists between blood types and billing. Rare blood types may have slightly higher costs.

Insurance Provider Billing Costs

Medicare patients face the highest average costs. There is little variation in billing among other providers.

Age and Billing Costs

Billing for most patients falls between $2,000 and $4,000. Outliers are linked to specific treatment types, insurance coverage, or unique conditions.

Conclusion

While the dataset offers some insights, its limitations (e.g., lack of specific alignments, diagnostic details, and facility locations) restrict its utility for deeper analysis. To enhance its value, additional data on treatment types, geographic factors, and patient demographics would be necessary. This analysis serves as a starting point for decision-making by patients, hospitals, and insurance providers.