Team Beta

Healthcare dataset Written Report:

Throughout the project, we came up with three questions that motivated our search and analysis. Firstly, what is the impact of a patient's information on their billing amount? While there was no direct correlation between patient-specific details (with the exception of patients aged 60+ who have arthritis), we found that billing appeared to be standardized across the board. There were small marginal differences in billing costs when it came to insurance providers and medical conditions. The only substantial evidence found was related to elderly patients with arthritis in particular.

The second question of interest that drove us was: What are the 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. Its limitations made it difficult for us to draw strong conclusions or correlations among the different elements, which we deduced as a limitation of the dataset.

Lastly, our third question was: 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.

Some key findings we discovered after cleaning the dataset were interesting, despite the challenges and limitations within the data. We dissected seven major components to guide our research. These components helped direct our analysis of the dataset.

One of the components we decided to follow up on was the average billing cost by hospital. The graph compares the average billing costs across hospitals, with Lamb Ferguson being the most expensive. It’s important to note that these hospitals may represent premium options for the patients, contributing to the high costs. The billing gap between the highest and lowest priced hospitals is approximately $8,000, showing the variability. This data can guide patient decision-making, hospital benchmarking, and insurance planning. However, the reasons behind these cost differences—such as case complexity, treatment types, or regional factors—warrant further investigation.

The second component of our research was the distribution of insurance providers. The pie chart illustrates the distribution of market share among five major insurance providers: Cigna, Medicare, UnitedHealthcare, Blue Cross, and Aetna. Each provider has nearly equal market shares, with only a minimal difference of 0.6% between the largest and smallest providers. This highlights a highly competitive and evenly distributed market. Cigna’s narrow lead could reflect slight advantages in service quality, pricing, or network coverage, while Aetna’s slightly lower share may indicate areas for potential improvement. This balanced distribution suggests a diverse range of customer preferences, despite the split among providers. The chart underscores a dynamic marketplace where even small changes in offerings or customer

satisfaction could significantly shift the balance. Overall, the data suggests a diversified and competitive insurance landscape.

The length of stay for patients averaged 15-16 days, with those having arthritis requiring up to two weeks on average. Elective admissions account for 52% of stays and generally require more observation and recovery time. One of the key factors we found interesting was the impact of admission costs, which was another variable we decided to explore. A trend we discovered was that urgent admissions had the highest average billing costs compared to elective and emergency types.

Patients with A-blood type incurred the highest costs, although there was no clear correlation between blood types and billing. Rare blood types might result in slightly higher costs, but the relationship remains unclear. When it comes to insurance providers, Medicare patients faced the highest average billing costs, while billing amounts showed little variation among other providers.

The last key component in our analysis was the relationship between age and billing costs. The scatter plot shows the relationship between patient age and average billing costs. Billing for most patients falls between $2,000 and $4,000, and there is no clear trend indicating a strong correlation between age and billing. This suggests consistent costs regardless of age, though there are some outliers with patients under 20 and over 80. These outliers may reflect treatment types, limited care, or specific insurance factors. However, additional analysis is needed to determine the impact of factors like treatment type, insurance coverage, or medical conditions on these costs. This data highlights the overall consistency of billing.

In 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. From this dataset, we were able to identify which insurance providers may be the best fit for lower billing amounts for patients, as well as which hospitals may be more suitable for those in need of immediate accommodation and care. For hospitals and insurance providers, the data provides insights into which admission types and medical conditions are more profitable. By calculating the majority of the correlations and finding the mean, this information can help them select their target demographics—specifically which conditions and admission types they could profit from most.