**Healthcare Cost and Utilization Analysis in Wisconsin Hospitals**

**1. Executive Summary**

This report analyzes hospital patient data for Wisconsin patients aged 0-17 to understand healthcare costs and utilization patterns. Key findings include that 9-year-olds have the highest mean expenditure, APRDRG 640 is the most frequent diagnosis, and length of stay (LOS) and diagnosis-related group (APRDRG) are the most important factors affecting hospital costs. Race was not found to be significantly related to costs, and age, gender, and race are poor predictors of LOS. The analysis provides insights for resource allocation, cost management, and quality improvement initiatives.

**2. Data Overview**

* **Observations:** 500 hospital records
* **Variables:** AGE, FEMALE, LOS, RACE, TOTCHG, APRDRG
* **Data cleaning:**
* Converted FEMALE, RACE, and APRDRG to factors
* Removed one row with missing RACE value

**3. Descriptive Analysis**

**3.1 Patient Statistics**

* The most frequent age category is 0 years old.
* The age category with the highest mean expenditure is 9 years old (Mean Cost = 10574).

*Table 1: Sorted Age and Expenditure Summary*

| **AGE** | **Mean\_Cost** |
| --- | --- |
| 9 | 10574.00 |
| 3 | 10183.00 |
| 5 | 9254.00 |
| 6 | 8964.00 |
| 4 | 7996.00 |
| 2 | 7298.00 |
| 10 | 6117.00 |
| 17 | 4599.00 |
| 15 | 3853.00 |
| 1 | 3774.00 |
| 12 | 3661.00 |
| 7 | 3362.00 |
| 14 | 2586.00 |
| 16 | 2384.00 |
| 8 | 2370.00 |
| 0 | 2212.00 |
| 11 | 1781.00 |
| 13 | 1730.00 |

**3.2 Diagnosis and Treatments**

* The most frequent diagnosis-related group is APRDRG 640 (Count = 266, Mean Cost = 1642).
* APRDRG 911 has the highest mean expenditure (35382.72).

*Table 2: APRDRG Summary with Count and Mean Cost*

| **APRDRG** | **Count** | **Mean\_Cost** |
| --- | --- | --- |
| 640 | 266 | 1642.00 |
| 754 | 37 | 1599.00 |
| 753 | 36 | 2210.00 |
| 758 | 20 | 1748.00 |
| 751 | 14 | 1548.00 |
| 755 | 13 | 859.00 |

**3.3 Cost Analysis by Race**

* ANOVA showed no significant relationship between race and hospitalization costs (F(5, 493) = 0.2437, p = 0.9429).

**3.4 Cost Analysis by Age and Gender**

*Table 3: Mean Cost by Age and Gender*

| **AGE** | **FEMALE** | **Mean\_Cost** |
| --- | --- | --- |
| 0 | 0 | 2198.00 |
| 0 | 1 | 2230.00 |
| 1 | 0 | 4328.00 |
| 1 | 1 | 1561.00 |
| 2 | 0 | 7298.00 |
| 3 | 0 | 11164.00 |
| 3 | 1 | 8223.00 |
| 4 | 0 | 9230.00 |
| 4 | 1 | 6762.00 |
| 5 | 0 | 7923.00 |
| 5 | 1 | 9254.00 |
| 6 | 0 | 8964.00 |
| 7 | 0 | 3362.00 |
| 8 | 0 | 2370.00 |
| 9 | 0 | 10574.00 |
| 10 | 0 | 6117.00 |
| 11 | 0 | 1781.00 |
| 12 | 0 | 3661.00 |
| 13 | 0 | 1730.00 |
| 14 | 0 | 2586.00 |
| 15 | 0 | 3853.00 |
| 16 | 0 | 2384.00 |
| 17 | 0 | 4599.00 |

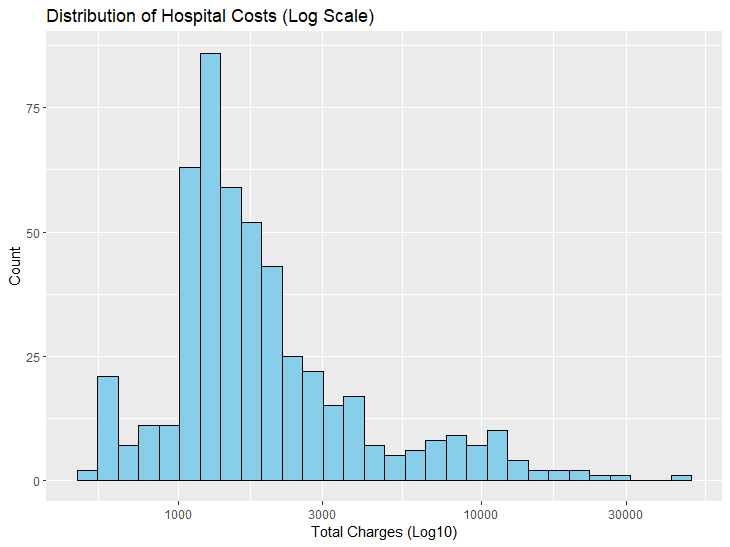
**3.5 LOS Prediction**

* Linear regression model (LOS ~ AGE + FEMALE + RACE) had very low explanatory power (R-squared = 0.0087).
* None of the predictors were statistically significant at α = 0.05.

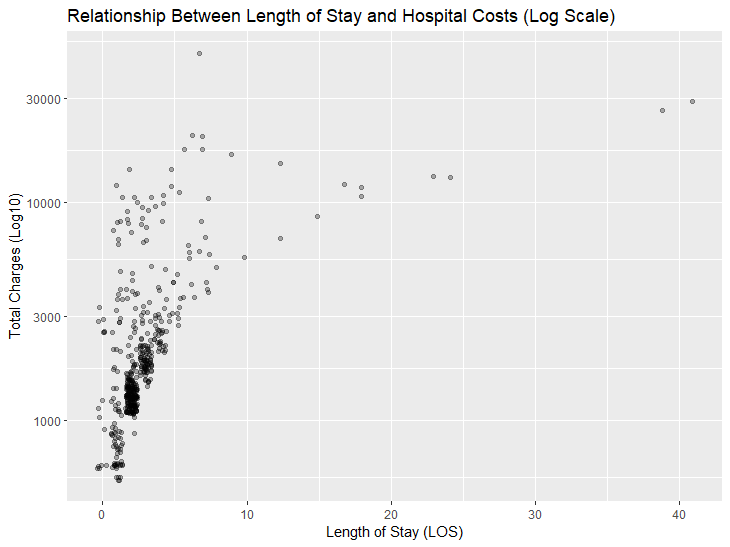
**3.6 Variable Importance**

* Linear regression model (TOTCHG ~ AGE + FEMALE + LOS + RACE + APRDRG) had high explanatory power (R-squared = 0.965).
* LOS and APRDRG were the most important predictors of hospital costs.

**4. Visualizations**

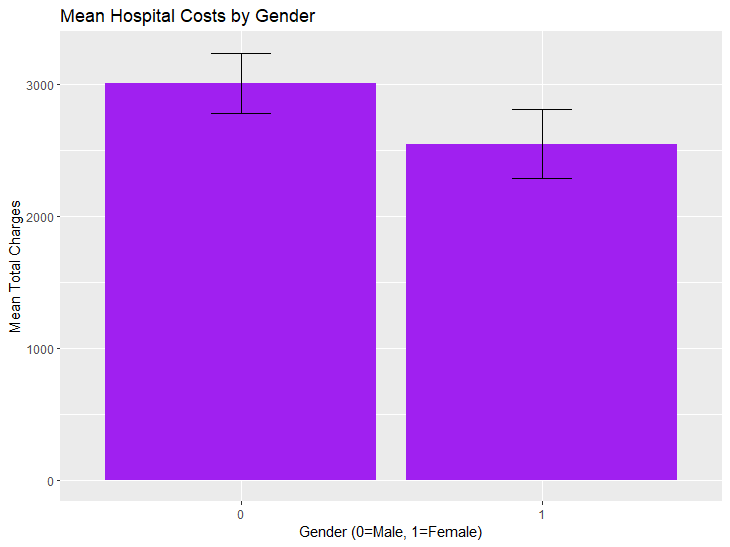
**4.1 Distribution of Hospital Cost****

*Caption:* "The distribution of hospital costs is highly right-skewed, with most patients incurring relatively low charges and a few patients with very high costs. The log scale helps visualize the distribution more clearly."

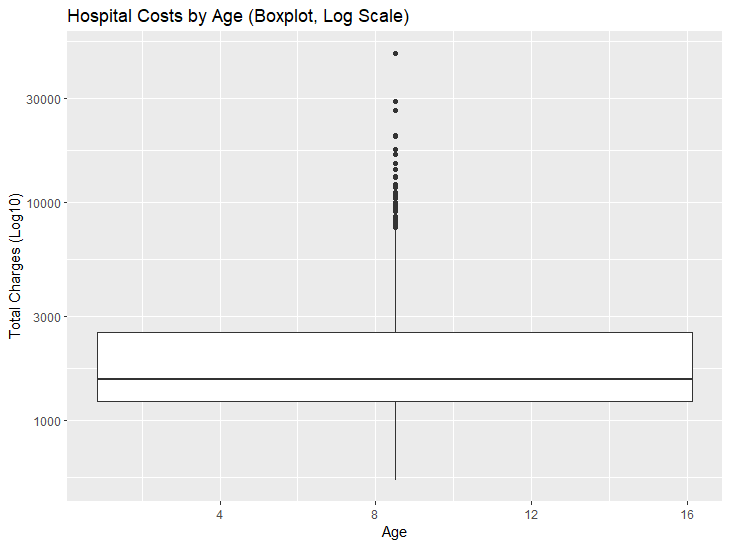
**4.2 Relationship Between Length of Stay and Hospital Costs**

*Caption:* "There is a positive relationship between length of stay and total hospital charges, with longer stays generally associated with higher costs. The log scale helps to visualize the relationship, but the high spread indicates that other factors also influence costs."

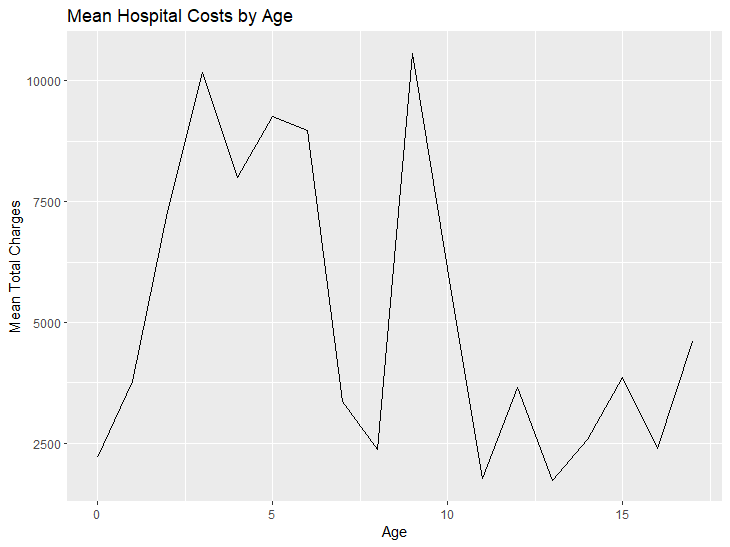
**4.3 Mean Hospital Costs by Gender (with Error Bars)**

**

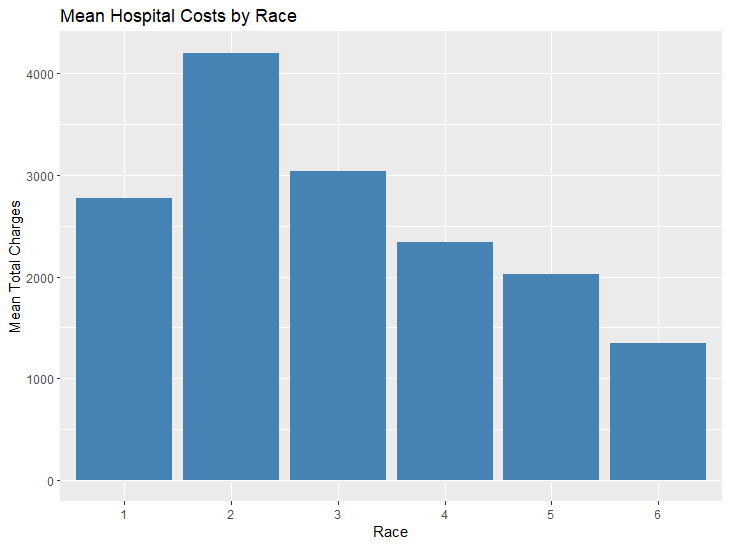
*Caption:* "This barplot shows the mean hospital costs for each gender. The overlapping error bars suggest that the difference in mean costs between genders may not be statistically significant."

**4.4 Hospital Costs by Age (Boxplot, Log Scale)**

*Caption:* "This boxplot shows the distribution of hospital costs for each age group. The log scale helps to visualize the central tendency and variability across age groups."

**4.5 Mean Hospital Costs by Age**

*Caption:* "This line plot shows the distribution of hospital costs for each age group. The log scale helps to visualize the central tendency and variability across age groups."

**4.6 Mean Hospital Costs by Race**

*Caption:* "This bar plot shows the mean hospital costs for each race category. ANOVA indicated that these differences are not statistically significant."

**5. Recommendations**

* Focus on managing length of stay and diagnosis-related groups to control hospital costs.
* Further investigate APRDRG 911 to understand why it has such high costs.
* Allocate resources based on age and gender, as these factors influence cost severity.
* Given the lack of relationship between race and costs, ensure equitable treatment across all racial groups.

**6. Appendix**

* R code
* Regression summary tables
* Additional plots/tables

**References**

* Agency for Healthcare Research and Quality (AHRQ) (dataset source)
* R version 4.5.0 and packages: tidyverse, readxl, car, lubridate

**End of Report**