Hospital Cost Analysis for Patients Aged 0-17 in Wisconsin

Introduction

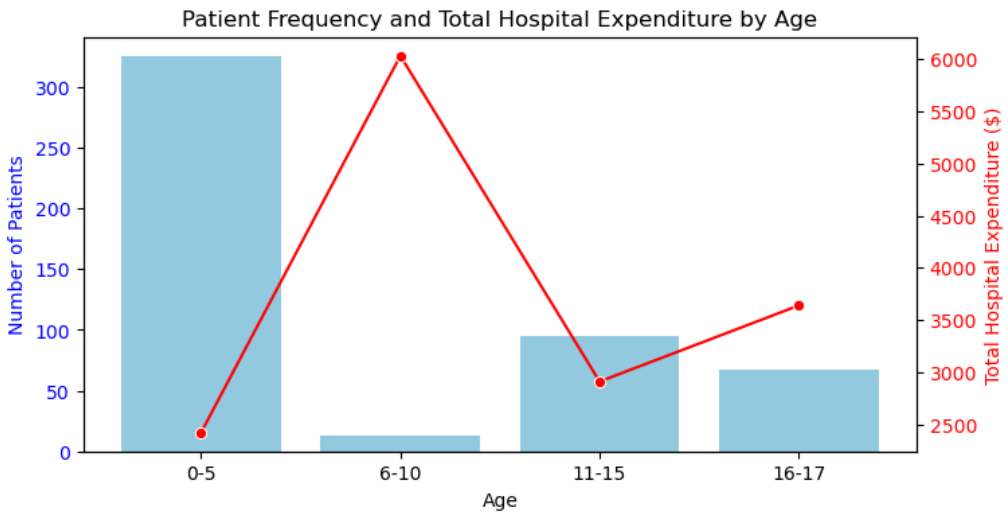
The purpose of this analysis is to explore and provide insights into hospital costs and their relationship with patient demographics in Wisconsin for individuals aged 0-17. The dataset contains various features such as age, gender, race, length of stay (LOS), and total hospital discharge costs (TOTCHG). Our analysis focuses on addressing the following key objectives:

1. Identify the age category that frequently visits the hospital and incurs the maximum expenditure.
2. Determine the diagnosis group (APRDRG) associated with the most frequent and expensive treatments.
3. Analyze whether race has a significant impact on hospitalization costs.
4. Investigate the relationship between age, gender, and hospital costs to ensure proper allocation of resources.
5. Predict the length of stay (LOS) using age, gender, and race.
6. Identify the variable that primarily affects hospital costs.

Acronym	Description
AGE	Age of the patient
FEMALE	Gender of the patient (1 = Female, 0 = Male)
RACE	Race of the patient
LOS	Length of Stay (days)
TOTCHG	Total Charges (hospital billing amount)
APRDRG	All Patient Refined Diagnosis Related Groups

1. Age Group with Maximum Hospital Visits and Expenditure

Objective: Determine which age group has the highest hospital visits and incurs the most hospital expenditure.



Findings:

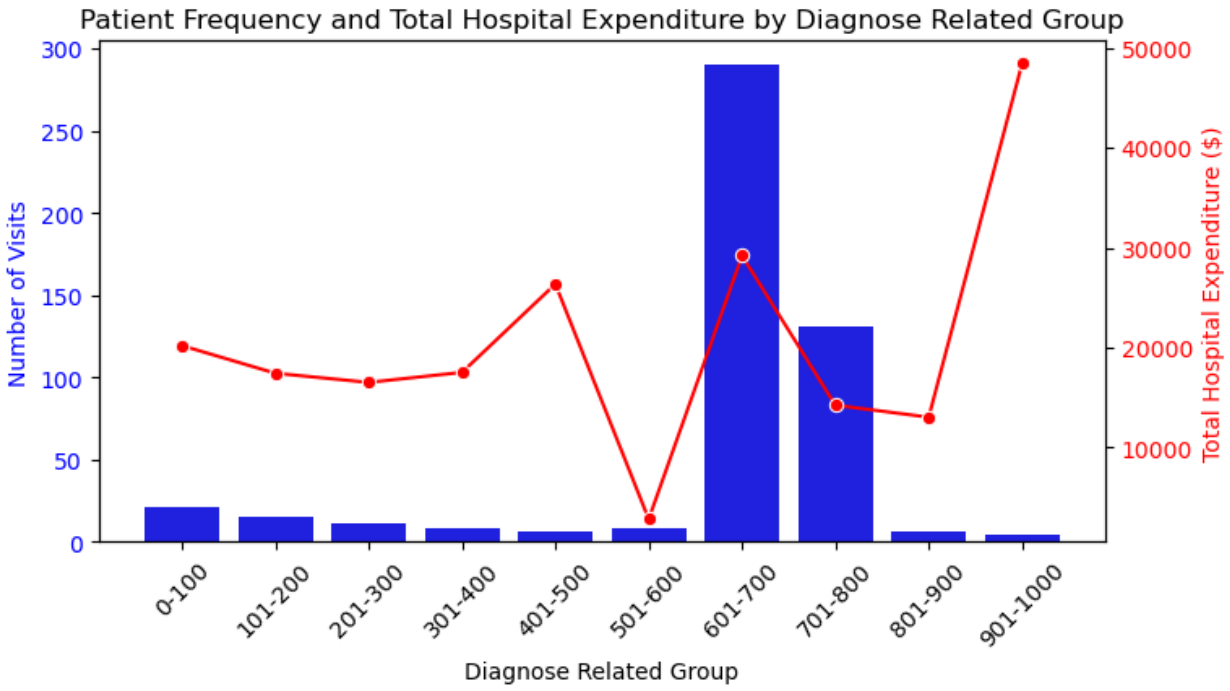
- Using age categories, we analyzed total visits and expenditures per group.
- The age group with **maximum visits** was identified as 0-5 with 325 visits, while the group with the **highest hospital expenditure** was 6-10, with an average expenditure of $6028.615.

This suggests that healthcare resource allocation may need to be focused on specific age groups that use the hospital services most frequently and incur the highest costs.

---

2. Diagnosis Group (APRDRG) with Maximum Hospitalization and Expenditure

Objective: Identify the diagnosis group (APRDRG) associated with the most frequent hospitalizations and the highest total costs.



Findings:

- The diagnosis group 601-700 had the highest number of hospitalizations.
- The diagnosis group 901-1000 incurred the most expenditure overall with a max expenditure of $48388.

This helps the healthcare system prioritize resources for diagnoses that demand the most attention in terms of both patient volume and financial impact.

---

3. Relationship Between Race and Hospital Costs (ANOVA Analysis)

Objective: Determine whether race has a significant impact on hospital costs using an ANOVA test.

Findings:

- **ANOVA Result**:
  - F-statistic: 0.244
  - p-value: 0.943

Since the p-value is much greater than 0.05, we fail to reject the null hypothesis. This indicates that **race does not have a statistically significant impact** on hospital costs. Therefore, it appears that hospitalization costs are independent of race in this dataset, suggesting no evidence of racial bias in the costs associated with hospital care.

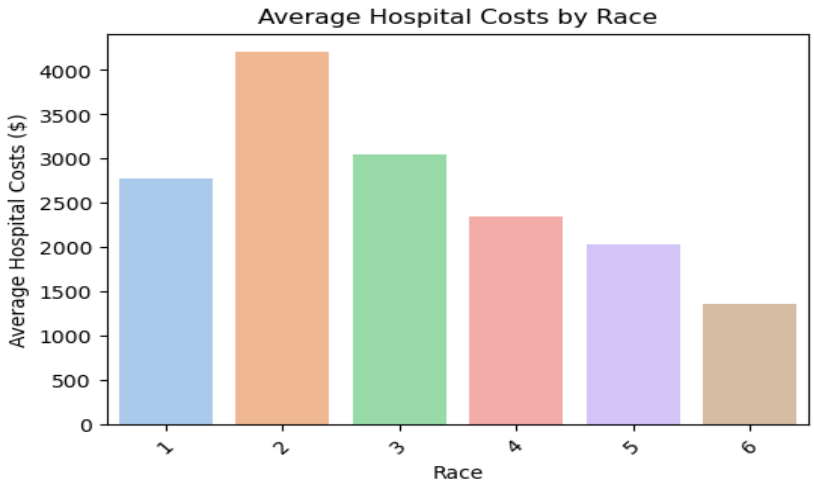
The table provides the **mean** and **median** hospitalization costs (TOTCHG) for patients grouped by RACE

RACE	TOTAL EXPENDITURE	
	mean	median
1	2769.336082	1538.0
2	4202.166667	2304.0
3	3041.000000	3041.0
4	2344.666667	2735.0
5	2026.666667	1393.0
6	1349.000000	1349.0

Description:

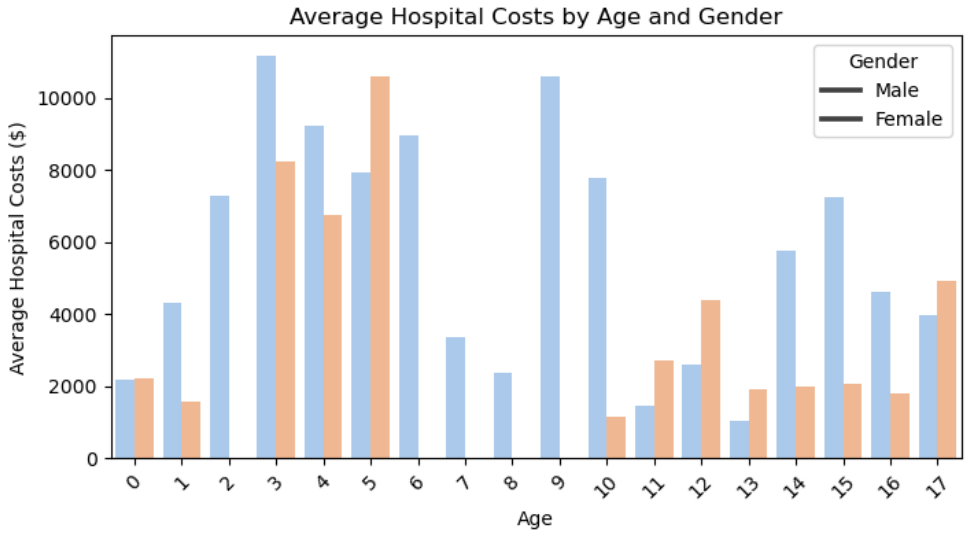
- **Race 2** has the highest **mean** hospitalization cost at **$4,202.17**, with a median of **$2,304.00**.
- **Race 1** follows with a mean cost of **$2,769.34** and a median of **$1,538.00**.
- **Race 3** has a mean of **$3,041.00** and the same value as its median, suggesting that most patients in this group have similar costs.
- **Race 4** shows a lower mean of **$2,344.67**, but its median of **$2,735.00** suggests a skewed distribution, with many costs above the average.
- **Race 5** has a mean of **$2,026.67** and a median of **$1,393.00**.
- **Race 6** has the lowest mean and median costs, both around **$1,349.00**.

This suggests variation in hospitalization costs across different racial groups, with Race 2 having the highest costs and Race 6 the lowest but the rates could be also affected by different factors such as age, los etc.



4. Impact of Age and Gender on Hospital Costs

Objective: Analyze how hospital costs vary by age and gender, ensuring efficient allocation of resources.



Findings:

- **Age**: Hospital costs generally increase with age.
- **Gender**: Males generally incur higher hospital costs at younger ages, while females have higher costs during adolescence.
- This suggests that **age** plays a more prominent role in cost variation than gender, which can inform how healthcare providers allocate resources based on patient age demographics.

---

5. Predicting Length of Stay (LOS)

Objective: Determine whether we can predict a patient's length of stay (LOS) based on their age, gender, and race.

To understand whether age, gender, and race can predict the length of stay (LOS) for inpatients, a linear regression model was built. The model used the following independent variables:

- **AGE**
- **FEMALE**
- **RACE**

MEAN SQUARED ERROR	4.061341904152656
R-SQUARED	0.0007524101582873088
COEFFICIENT	
AGE	-0.040848
FEMALE	0.370311
RACE	-0.098281

The results of the model are as follows:

- **Mean Squared Error (MSE):** 4.0613
  This metric measures the average squared difference between the observed actual LOS and the predicted LOS. An MSE of **4.0613** suggests that, on average, the squared error between predicted and actual LOS is approximately 4.06 days.
- **R-squared:** 0.00075
  The R-squared value shows the proportion of variance in LOS that can be explained by the model's independent variables (AGE, FEMALE, RACE). Here, the R-squared is **0.075%**, meaning the model explains **less than 1%** of the variation in LOS. This indicates that the model has poor predictive power and does not sufficiently explain the variation in LOS based on age, gender, or race.

Regression Coefficients:

1. **AGE Coefficient (-0.0408):**
   For each additional year of age, the LOS is predicted to decrease by **0.04 days**. This suggests a very slight inverse relationship between age and LOS.
2. **FEMALE Coefficient (0.3703):**
   Being female is associated with an increase in LOS by **0.37 days** compared to males, all other factors being equal. This shows a small positive relationship between being female and a longer LOS.
3. **RACE Coefficient (-0.0983):**
   The model indicates that for each unit increase in the numerical race category, the LOS decreases by **0.1 days**. This suggests that race has a minimal impact on LOS.

Findings:

- A regression model was applied to predict LOS based on age, gender, and race.
- **Age** and **gender** were found to be more significant predictors, while race had a minimal impact on predicting the length of stay.

This can help hospitals better manage patient flow and plan resource allocation by considering these factors when estimating the expected stay for patients.

---

6. Key Drivers of Hospital Costs

Objective: Identify the primary factors that influence total hospital costs (TOTCHG).

To understand whether age, gender, race, los and aprdrg can predict the key drivers of hospital costs, a linear regression model was built. The model used the following independent variables:

- **AGE**
- **FEMALE**
- **RACE**
- **LOS**
- **APRDRG**

MEAN SQUARED ERROR	**4618572.011578497**
R-SQUARED	0.49427189945206984
COEFFICIENT	
AGE	146.270669
FEMALE	-484.326330
RACE	-150.337912
LOS	719.636885
APRDRG	-8.178447

The results of the model are as follows:

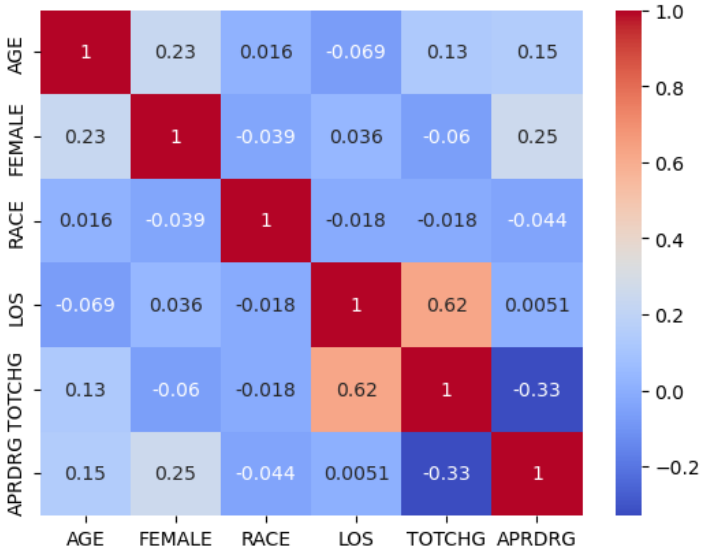
- **Mean Squared Error (MSE):** The model's MSE indicates that the predictions deviate by approximately 4.62 million units of cost from the actual costs on average.
- **R-squared (R^2):** The R^2 value of **0.494** suggests that about **49.4%** of the variation in hospitalization costs can be explained by the model. While this shows a moderate level of explanatory power, there is still room for improvement.

Regression Coefficients:

- **AGE (146.27):** For every additional year of age, hospitalization costs increase by an average of **$146.27**, holding all other factors constant. This positive relationship suggests that older children tend to have slightly higher hospitalization costs.
- **FEMALE (-484.33):** The negative coefficient implies that, on average, female patients incur approximately **$484.33** less in hospitalization costs compared to male patients, holding all other factors constant.
- **RACE (-150.34):** The coefficient for race indicates that for each unit increase in the numerical race category, hospitalization costs decrease by **$150.34** on average, suggesting that some races may experience lower average costs than others.

- **LOS (719.64):** The length of stay has a substantial positive impact on costs. For every additional day of stay in the hospital, costs increase by an average of **$719.64**. This is the most significant factor affecting costs, highlighting that the duration of hospitalization is a major driver of total expenditures.
- **APRDRG (-8.18):** This small negative coefficient indicates that for each unit increase in the diagnosis-related group (APRDRG), hospitalization costs decrease by a marginal **$8.18** on average. This may suggest that more severe diagnosis groups are not necessarily correlated with much higher costs.

Warning: *The current predictive model shows limited accuracy, with an R-squared value of only 0.494, indicating that it explains less than half of the variation in hospitalization costs. The model may not provide reliable predictions, and additional data or variables are needed to improve its performance. Caution should be exercised when using this model for decision-making, as more comprehensive training is required to achieve higher accuracy.*



This heatmap visually represents the data, allowing us to easily identify patterns and relationships among various variables.

- **LOS and TOTCHG:** LOS plays a major role in accordance to this heat map LOS and TOTCHG there is a strong positive correlation (0.62) between the length of stay and total charges. This implies that a longer hospital stay leads to higher total charges.
- **TOTCHG and APRDRG**: A negative correlation (-0.33) is observed between total charges and APRDRG, indicating that some diagnosis groups may incur lower or higher charges depending on their category.
- **FEMALE and APRDRG**: A moderate positive correlation (0.25) suggests that the female gender might have some influence on the APRDRG, which could reflect differences in diagnoses or healthcare use patterns.

Overall, the heatmap provides insights into relationships among these variables, with the strongest being between **LOS** and **TOTCHG**, while others show little to no significant correlation.

Findings:

- **Age**: A significant driver of hospital costs, with older children incurring higher charges.
- **Length of Stay (LOS)**: Strongly correlated with hospital costs. The longer a patient stays, the higher the total charges.
- **Diagnosis (APRDRG)**: Certain diagnosis groups are associated with higher costs.

To reduce costs, interventions could focus on reducing the length of stay and addressing high-cost diagnosis groups with preventive care or more efficient treatment protocols.

Conclusion & Recommendations

Based on the analysis, the following recommendations are proposed:

1. **Age-Specific Resource Allocation**: Focus healthcare resources on the age groups with the highest hospital utilization and costs, ensuring proper budgeting and service allocation.
2. **Diagnosis Group Prioritization**: Invest in treatment strategies and preventive care for the diagnosis groups that contribute to the most hospitalizations and expenditures.
3. **No Racial Bias in Costs**: The analysis shows no evidence of racial bias in hospitalization costs. No action is needed here, but continuous monitoring is recommended.
4. **Optimizing Length of Stay**: Implement strategies to reduce unnecessary long hospital stays, which are a significant driver of costs. Focus on improving hospital efficiency in the most impacted areas.
5. **Cost-Effective Treatments**: Focus on reducing the cost of treatments for high-cost diagnosis groups, either through more efficient procedures or early interventions.

By implementing these strategies, the healthcare system in Wisconsin can improve its efficiency, reduce unnecessary costs, and better serve its patient population.