Adetutu B.

Exploratory Analysis

A1: Organization Question

Is there a significant relationship between a patient's likelihood to be readmitted and their initial admission type?

A2: Stakeholder Benefits

Addressing this research question can benefit stakeholders by helping the hospitals save money, improve patient outcomes, and implement better strategies to manage hospital resources over time. Understanding patterns in readmission data in relation to patient admission types can enable providers to identify patients with a higher risk of readmission earlier in their initial stay. In addition to allowing proactive identification, insights from this data could enable healthcare staff to better allocate resources and services to patients with a higher likelihood of readmission and give the hospitals a chance to improve or develop new patient discharge and follow-up protocols. This is important because stakeholders benefit from reduced readmission rates. Using insights from this data could not only help the hospitals implement ways to lower costs related to patient care, but reducing readmission will also lower penalty fees imposed on hospitals with excessive preventable readmission rates.

A3: Relevant Data

Of the 52 total columns in the medical data set, I will mainly only need 2 variables to research my question: **ReAdmis** (a qualitative binary variable stating whether or not the patient was readmitted within 1 month of release) and **Initial_admin** (a qualitative variable identifying a patient's initial admission type: emergency, elective, observation).

B1: Dataset Code

I have attached my code in a separate file titled: **analysis_code.ipynb**

The code for this section can be found under the header '**B1 & B2: Dataset Code & Chi-Squared Test Output Results**'.

B2: Chi-Squared Test Output Results

In B1 I performed a chi-square test to analyze my research data and calculated the chi-square statistic, p-value, degrees of freedom, and expected frequencies of the contingency table. The chi-square statistic, which shows the measured difference between my observed and expected frequencies, was calculated to be approximately (**$X^2 \approx 3.88997$**). My p-value, which helps

determine the significance of my chi-square statistic, was approximately (**p ≈ 0.69**). In order to define the critical chi-square distribution values, I calculated the degrees of freedom which were (**dof = 6**). Lastly my **expected** values, which were used to calculate the chi-square statistic, provide a benchmark to compare the observed frequencies for discrepancies. For these values, an array was printed/output from my code: "**array ([[1585.2824, 3203.486, 1542.2316, 6331.], [918.7176, 1856.514, 893.7684, 3669.], [2504., 5060., 2436., 10000.]])**". Regarding the significance of my results in relation to my hypothesis, I will talk more about this later in section E1.

B3: Analysis Technique Justification

I chose to complete a chi-square test because I wanted to analyze the association of two categorical variables. I wanted to figure out if there was a statistically significant relationship between patients who were readmitted and the type of admission they initially came to the hospital under during their prior stay. For me, the chi-square test was the best option in this scenario because of the type of question and type of data (categorical and discrete) I had.

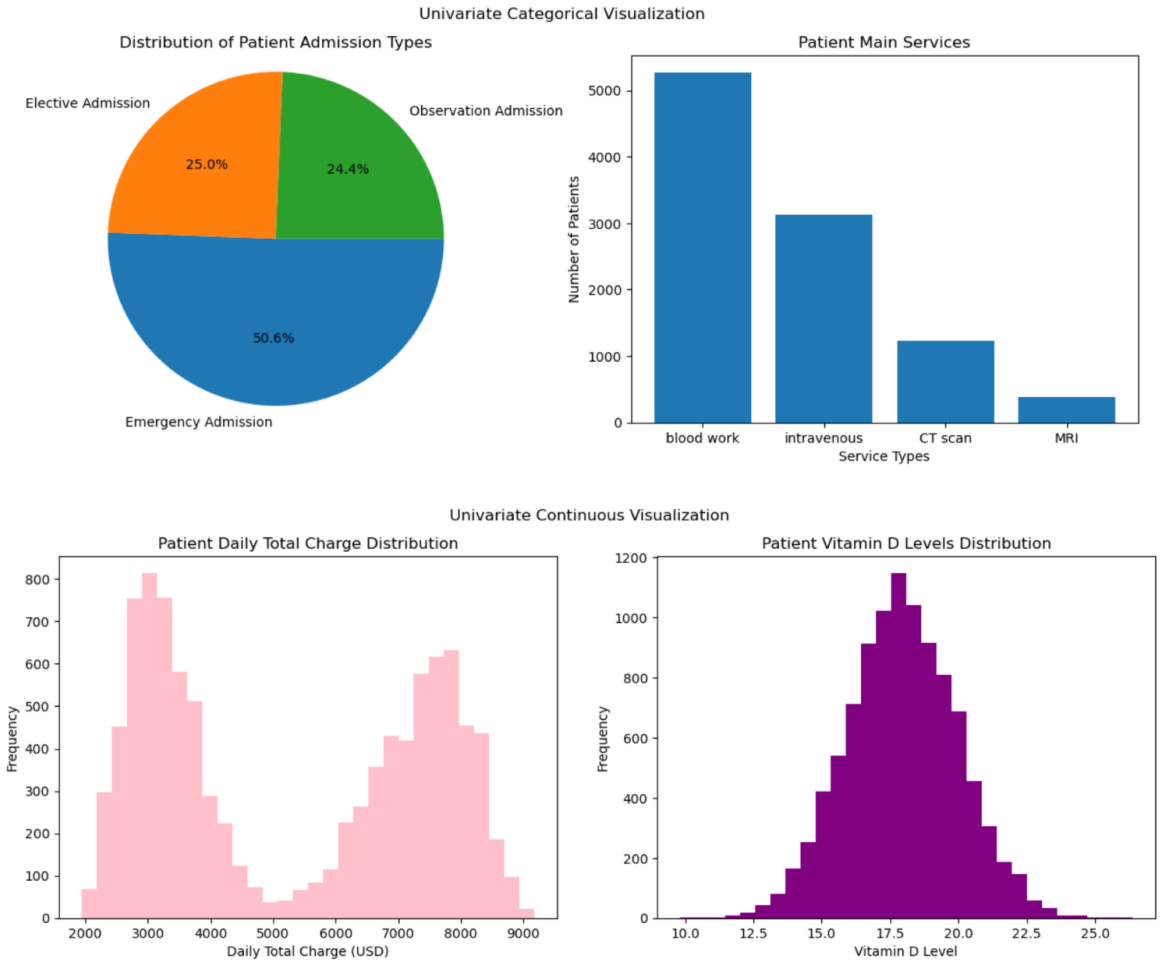
C: Univariate Statistics

This section's code can be found in my separately attached file: **analysis_code.ipynb**
Under section '**C: Univariate Statistics**' is the code for the univariate distribution of categorical variables: **'initial_admit_type'** and **'services'**. Code for univariate distributions of continuous variables: **'daily_charge'** and **'vitamin_d_level'** can also be found here.

C1: Univariate Visualization

Visualization for the univariate statistics can be found in the same file: **analysis_code.ipynb** in the '**C1: Univariate Visualization**' section.

Direct images of the visualizations for your convenience:



D: Bivariate Statistics

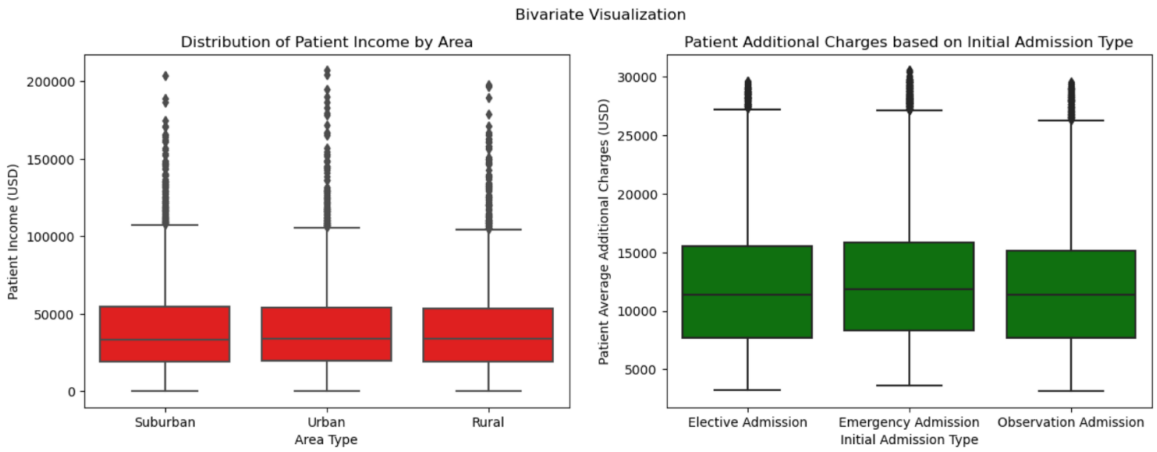
The code in this section is also in my separately attached file: **analysis_code.ipynb**

In section '**D: Bivariate Statistics**' you can find code for the distribution of categorical variables: **'initial_admit_type'** and **'area_type'** and continuous variables: **'additional_charge'** and **'income'**. These variables will be used to show the relationships between "area_type vs income" and "initial_admit_type vs additional_charge".

D1: Bivariate Visualization

The bivariate statistics visuals can be found in the same file: **analysis_code.ipynb** under the header '**D1: Bivariate Visualization**'.

A direct image of the visualization for your convenience:



E1: Hypothesis Test Results

H_0 *: There is not a statistically significant relationship between a patient's likelihood of readmission and their initial admit type.*

H_A *: There is a statistically significant relationship between a patient's likelihood of readmission and their initial admit type.*

My p-value (**p ≈ 0.69**) represents my probability of getting the observed result under the assumption of a true null hypothesis. To determine whether this p-value was statistically significant I compared it against an **alpha value of 0.05**, which provides a criterion for assessing statistical significance with the acceptance that there is a 5% chance of incorrectly concluding there is significance (LaMorte, 2019). In comparison to the alpha value of 0.05, my p-value of 0.69 was the greater value (NOT below 0.05). So, we should fail to reject the null hypothesis (accept the null hypothesis). This means that **there is not enough evidence to claim there is a statistically significant relationship between a patients' likelihood of being readmitted and their initial admission type**.

E2: Analysis Limitations

There are a few limitations that came to mind during the process of analysis. The most obvious being the sample size of only 10,000 responses. Considering that this data is being used to analyze/investigate not just one hospital, but a chain of medical hospitals across the United States, I felt that the sample size was relatively small in comparison to the scale of the company. Analyzing a bigger sample size may enable me to improve the reliability and validity of my results and reduce possible errors/biases in the long run. Another limitation I faced was the lack of information about a patient's reason for hospitalization. My goal was to analyze if there is an association between the likelihood of readmission and the initial admission type. In

addition to a patient's admission type, knowing why someone was initially admitted and subsequently readmitted could greatly change how my data insights are interpreted and utilized in the business. Understanding a patient's specific reason(s) for admission could reveal possible underlying factors that impact initial admission and readmission rates. It could also allow me to provide stakeholders with more specific insights on the cause or severity of hospital readmissions through further analysis and questions like: Was the patient's readmission type the same as their initial admission type? What symptoms/complications constitute readmission? Was the patient readmitted due to complications from their initial treatment, or is this a new illness or complication that resulted after discharging from their initial stay? While I know the intent of not providing the original reason for hospitalization in the data was to compare readmission to other factors, I still believe this additional context would allow analysts to provide more nuanced recommendations for targeted patient interventions, care protocol adjustments, and effective preventative measures in order to address stakeholder concerns regarding patient readmission rates.

E3: Next Course of Action

My hypothesis test results showed no evidence of a statistically significant relationship between my two variables. However, I would still recommend that data continue to be collected/reported to increase the data sample size for further analysis in the future. I would also suggest beginning to record data on a patient's readmission type to be analyzed in addition to the initial admission type and whether or not they were readmitted. This honors the intent of keeping a patient's reason for admission out of the data pool while still providing more context to analysts and stakeholders what the severity of a patient's readmission was if they did return to the hospital within a month's time.

F: Code Web Sources

Pie charts#. Pie charts - Matplotlib 3.9.2 documentation. (n.d.). https://matplotlib.org/stable/gallery/pie_and_polar_charts/pie_features.html

G: References

LaMorte, W. W. (2019, May 16). *Module 7 - Comparing Continuous Outcomes*. P-Values. https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module7-T-tests/Module7-ttests3.html