Exploratory Data Analysis – D207

Gooden, Nina S. [ID #: 009823504]

Dr. David Gagner

Exploratory Data Analysis – D207

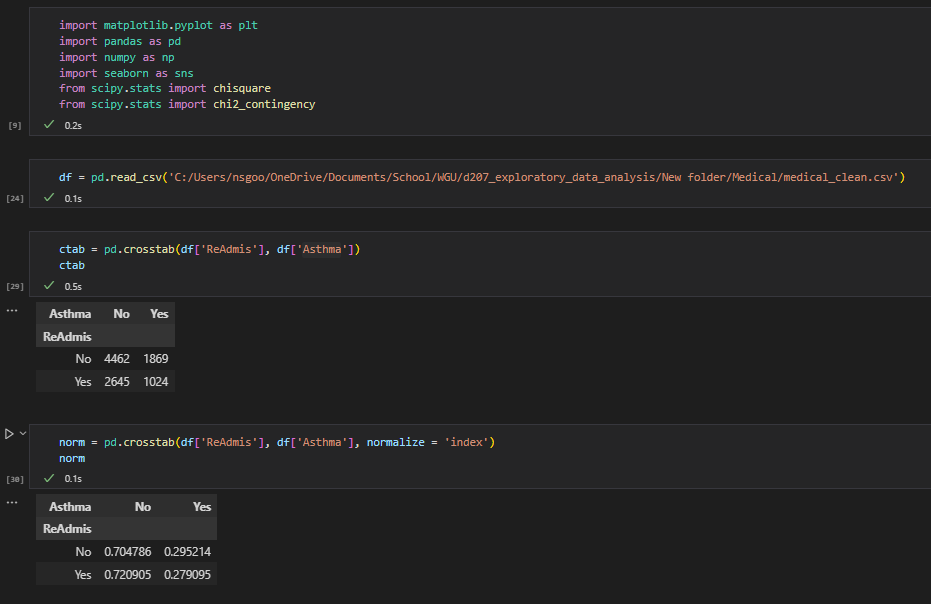
This analysis will explore the medical data of a potentially real-world organization. In creating this analysis, I will present an evaluation using one of the allowed measurement techniques, to identify key performance indicators to answer a specific question.

## **A. Describe an organizational situation/issue that the data will address:**

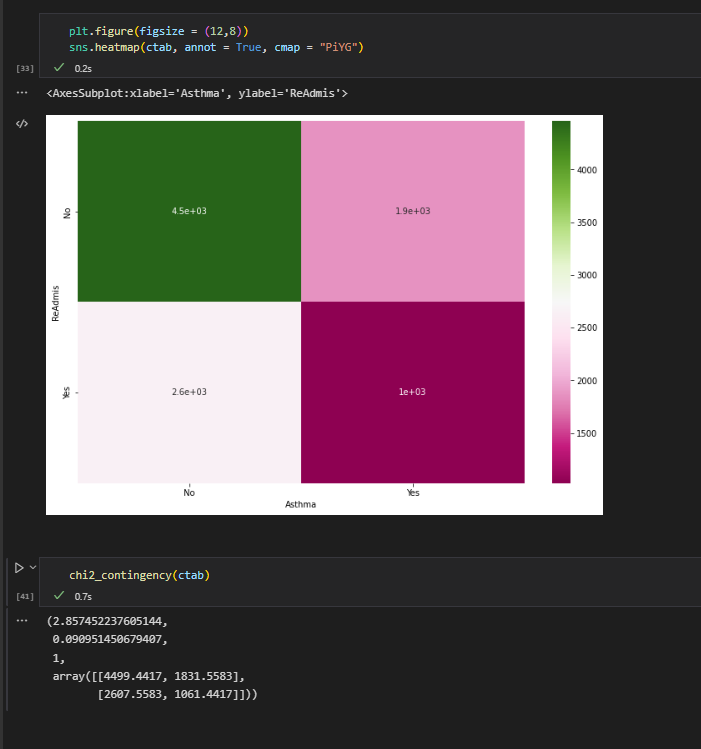
1. Research Question
   1. Does having asthma have an association with readmission likelihood? What other variables may potentially contribute to readmission?
2. Organizational Benefits
   1. By identifying whether or not patients with asthma have a higher chance of being readmitted, stakeholders may be able to offset that likelihood with more pointed care.
   2. Provide centers of adjustments to lower readmission.
   3. Determine if changes need to be made to processes in order to avoid future legal action due to readmission.
3. Relevant Data from Set
   1. The most relevant variable to our analysis is “Readmis,” which serves as our statis variable for analysis. “Asthma” is another categorical variable that I will be exploring in order to satisfy my initial research question. During the evaluation process, I identified “Initial\_days,” “BackPain,” “Initial\_admin,” “Overweight,” “Age,” and “Doc\_visits” as also having association/dependence with the readmission variable.

## **B. Data Analysis**

1. Code for Technique



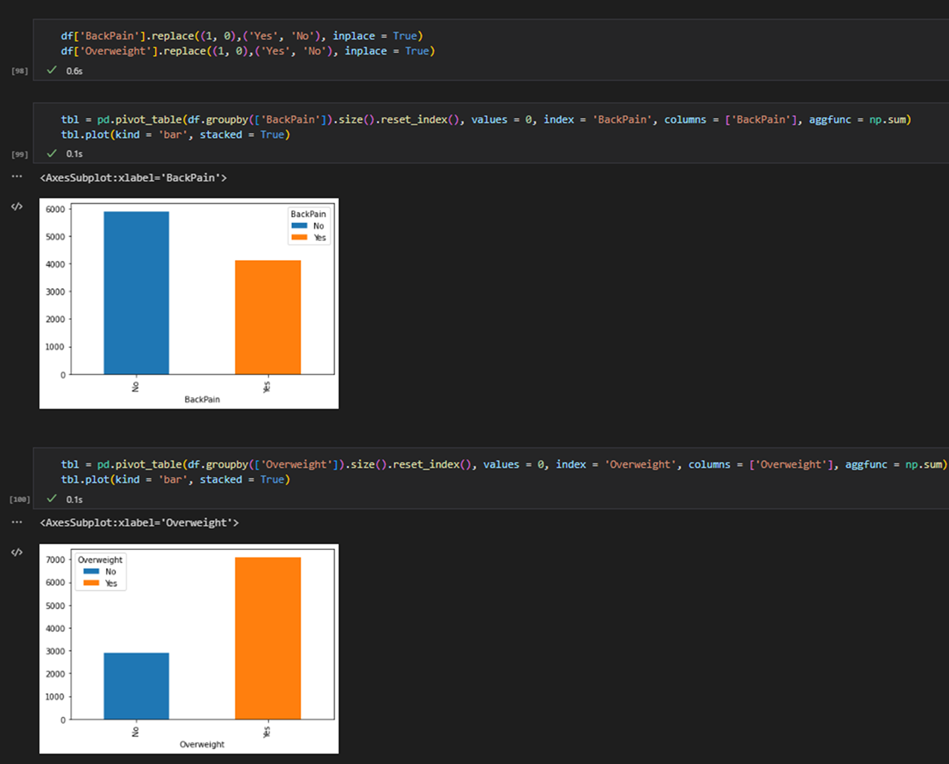
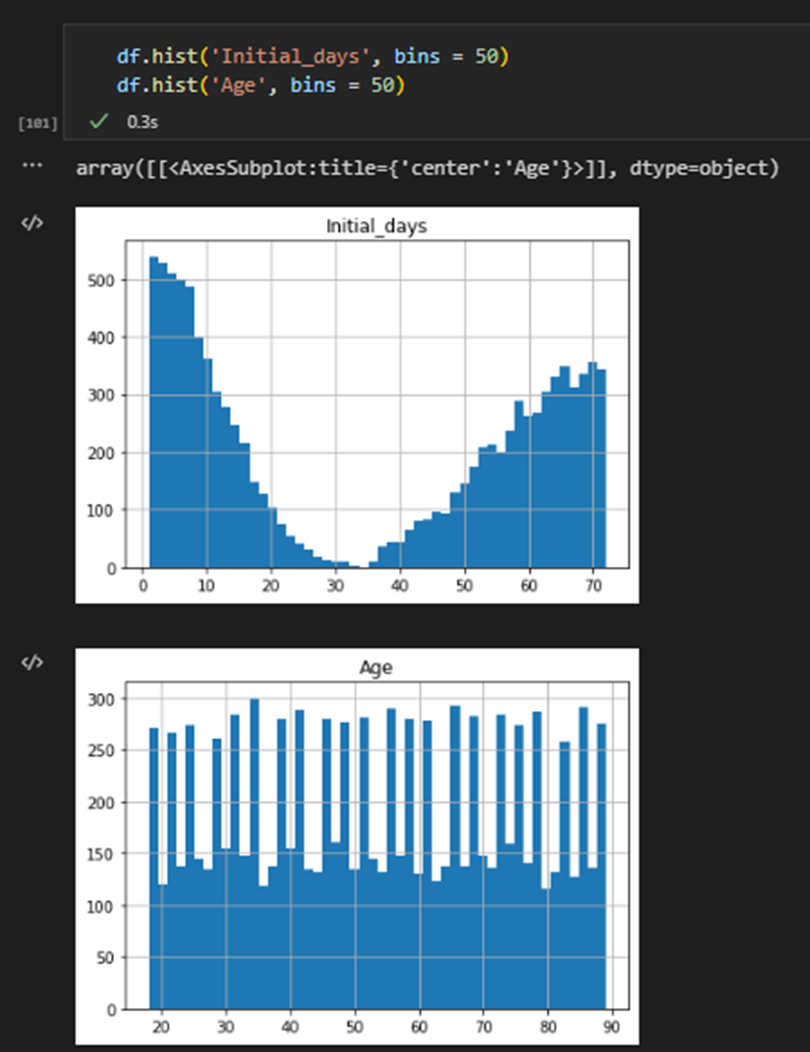
1. Output and Results



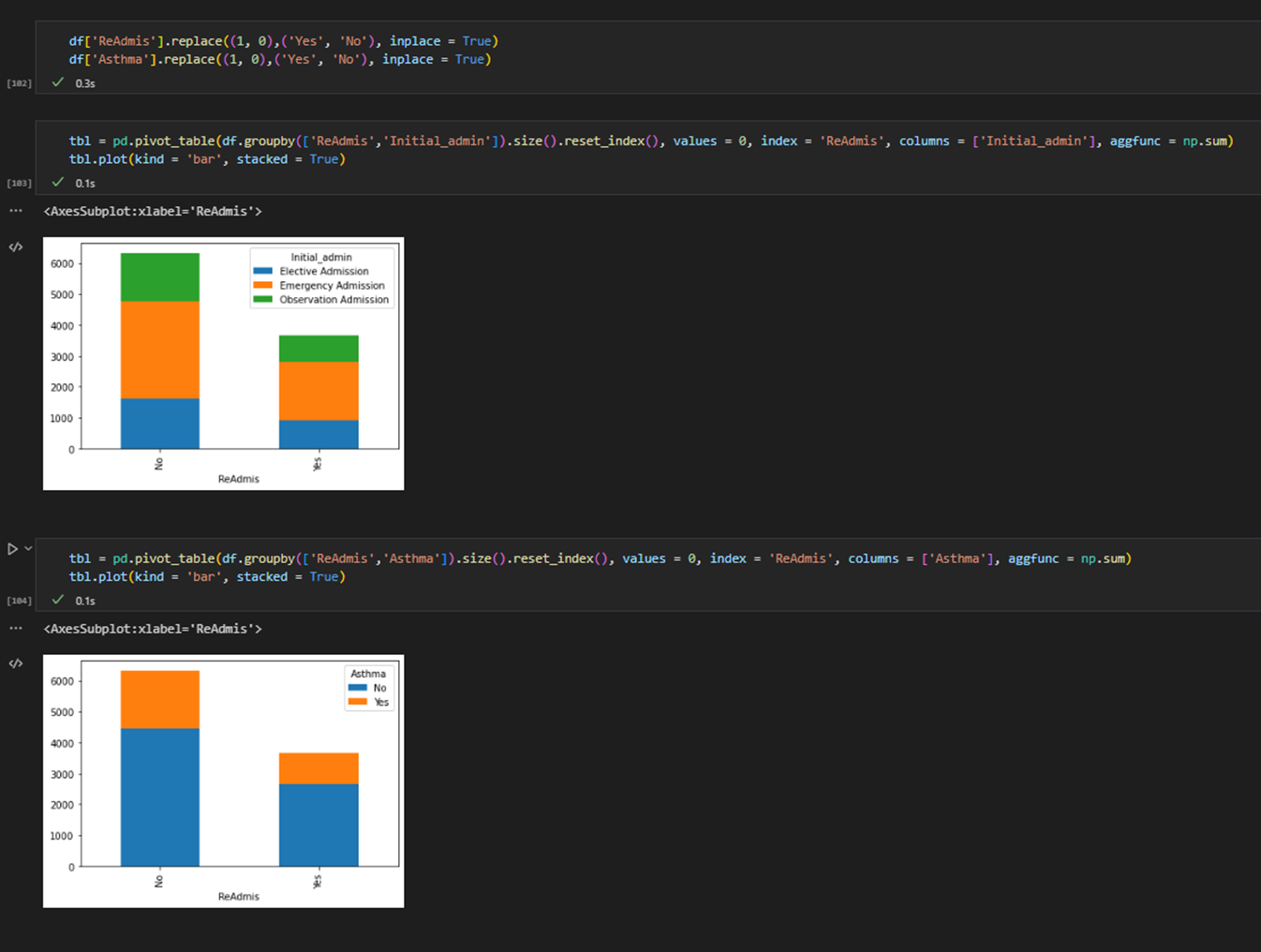
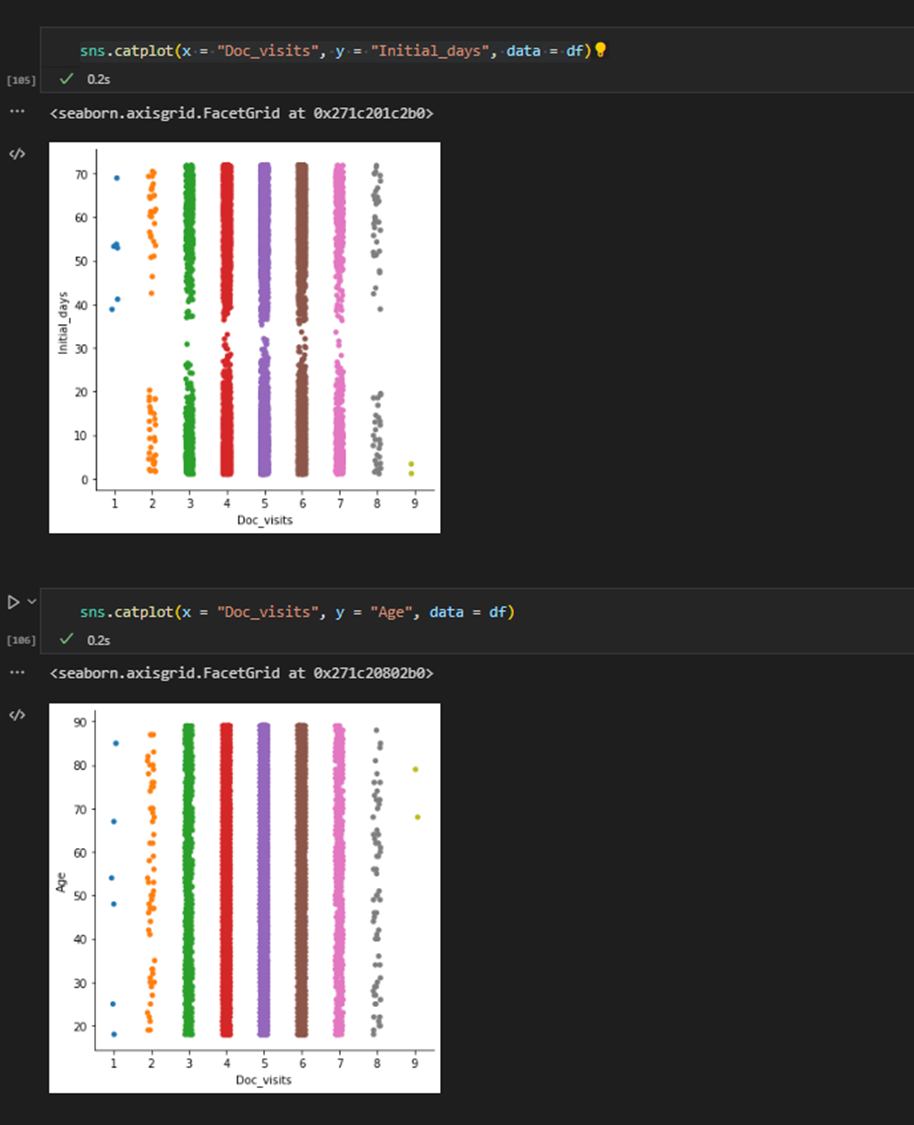
The p-value here is much lower than the statistic, which means the likelihood that the variables have a correlation is high.

1. Justification
   1. Because we are looking at potential associations between readmission and a number of potential correlations, I first looked at the type of variable “Readmis” was. As a radio, categorical dependent variable, it was absolutely vital to my assessment. I was specifically interested in the potential correlation between readmission and asthma, so the chi-square was the most appropriate option.

## **C. Univariate Statistics**

1. Categorical Variables
   1. BackPain, Overweight
   2. BackPain showed the potential for readmission correlation. Being overweight is a hot-button issue in the medical community. I wanted to check this column to see if there was merit for that concern.
   3. 
2. Continuous Variables
   1. Initial\_days and Age
   2. Initial\_days, and Age showed the potential for readmission correlation.
   3. 

## **D. Bivariate Statistics**

1. Categorical Variables
   1. Initial\_admin, Asthma
   2. Initial\_admin and Asthma showed the potential for readmission correlation. In order to get an accurate view of these variables, I needed to look at them broken down further.
   3. 
2. Continuous Variables
   1. Initial\_days, Age
   2. Initial\_days and Age showed the potential for readmission correlation. I used Doc\_visits as a comp because the spread for the secondary variable was easy to categorize visually.
   3. 
   4. Chart, histogram

      Description automatically generated

## **E. Implication Summary**

1. **Results:** The chi-square test identified the “Asthma” variable as having a possible association with “Readmis.”
   1. Patients with asthma were found to have significant probabilities of being readmitted.
   2. In addition, duration of Initial\_days and age were also found to have the potential for correlation.
2. **Limitations:** With and all statistical evaluations, it is vital that we remember that correlation and causation are not one and the same.
   1. Data evaluated is from a single industry source, with set resources and may not apply to all medical institutions.
   2. Varying degrees of correlation were noted and should be further examined for in-depth analysis.
3. **Recommendations:** The primary research question is focused on whether or not asthmatic patients are readmitted more often. A potential counter to this identified concern would be to provide patients with asthma with additional resources to offset and not any specific challenges.
   1. Additionally, spend additional resources on patients over a certain threshold of age in order to decrease the likelihood for readmittance—reporting and observing this trend may be vital to avoiding lawsuits.
   2. Lastly, reduce number of days patients are held after being initially admitted to the hospital.

## **F. Panopto**

## **G. Third-party Code References**

Seaborn.histplot#. seaborn.histplot - seaborn 0.12.0 documentation. (n.d.). Retrieved September 30, 2022, from https://seaborn.pydata.org/generated/seaborn.histplot.html

https://stackoverflow.com/questions/72009138/pandas-dataframe-label-columns-encoding

pandas.DataFrame.sort\_values — pandas 1.5.0 documentation. (n.d.). Retrieved September 30, 2022, from https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sort\_values.html

GeeksforGeeks. (2022, September 28). Python program to convert a list to string. Retrieved September 30, 2022, from https://www.geeksforgeeks.org/python-program-to-convert-a-list-to-string/?ref=leftbar-rightbar

scipy.stats.chi2\_contingency — SciPy v0.14.0 Reference Guide. het.as.utexas.edu/HET/Software/Scipy/generated/scipy.stats.chi2\_contingency.html. Accessed 09 Oct. 2022.

“Chi-squared Test With Scipy: What’s the Difference Between Chi2\_Contingency and Chisquare?” Cross Validated, 5 Aug. 2014, stats.stackexchange.com/questions/110718/chi-squared-test-with-scipy-whats-the-difference-between-chi2-contingency-and.

scipy.stats.chi2\_contingency — SciPy v1.9.2 Manual. docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2\_contingency.html. Accessed 11 Oct. 2022.

## **H. In-text Citations**

None.