Exploratory Data Analysis – D207

Gooden, Nina S. Gooden [ID #: 009823504]

Dr. David Gagner

Exploratory Data Analysis – D207

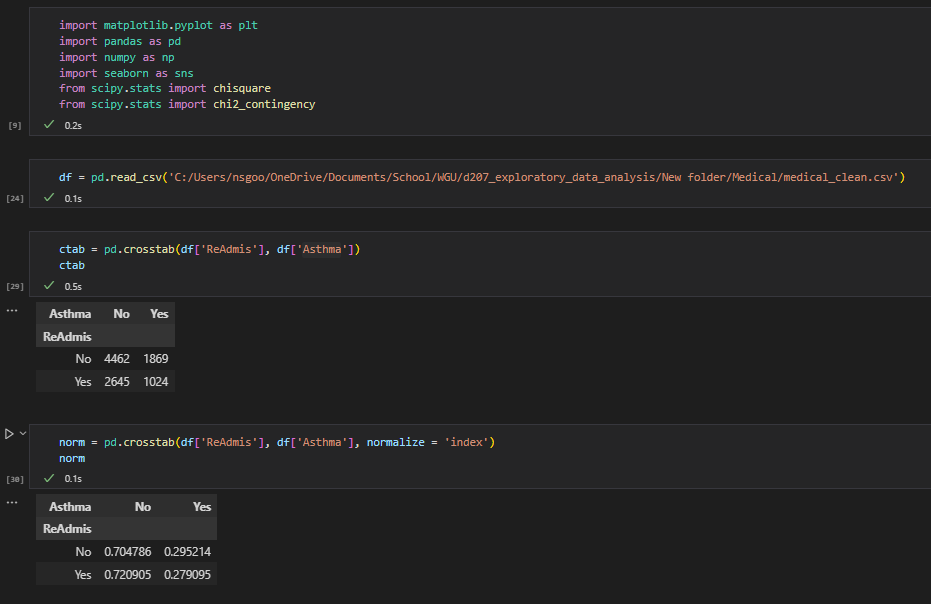
This analysis will explore the medical data of a potentially real-world organization. In creating this analysis, I will present an evaluation using one of the allowed measurement techniques, to identify key performance indicators to answer a specific question.

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

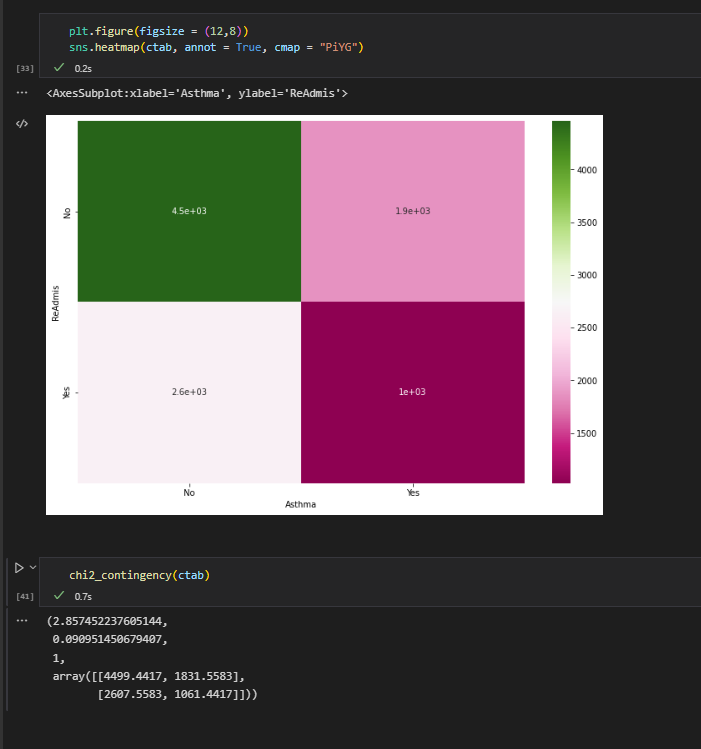
1. Research Question
   1. Does having asthma have an association with readmission likelihood?
2. Organizational Benefits
   1. In 2018, the cost of readmissions cost hospitals more than $41B (LaPoint, 2018). Identifying whether patients with asthma consistently contribute to readmissions will allow stakeholders to take action to mitigate cost.
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.

## **B. Data Analysis**

1. Code for Technique
   1. When comparing two categorical variables, the chi-square test makes the most sense. To run this test, I first used the *pd.crosstab* command from *pandas* to convert the data into a contingency table with the number of frequencies for each answer of Yes and No.
   2. Next, I normalized the data into a second contingency table.

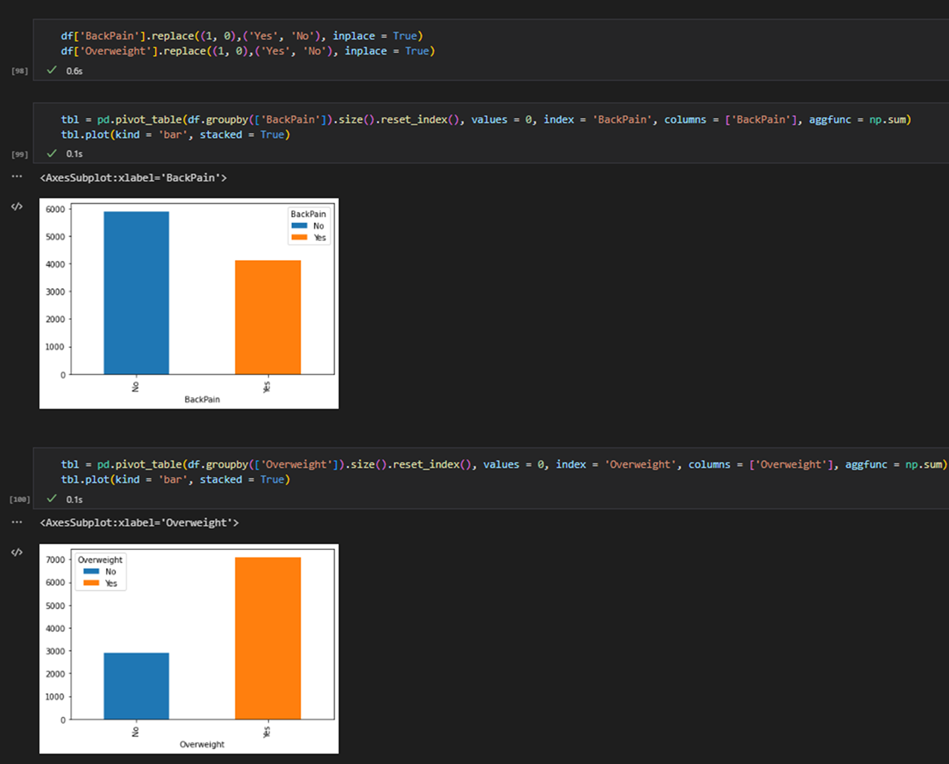
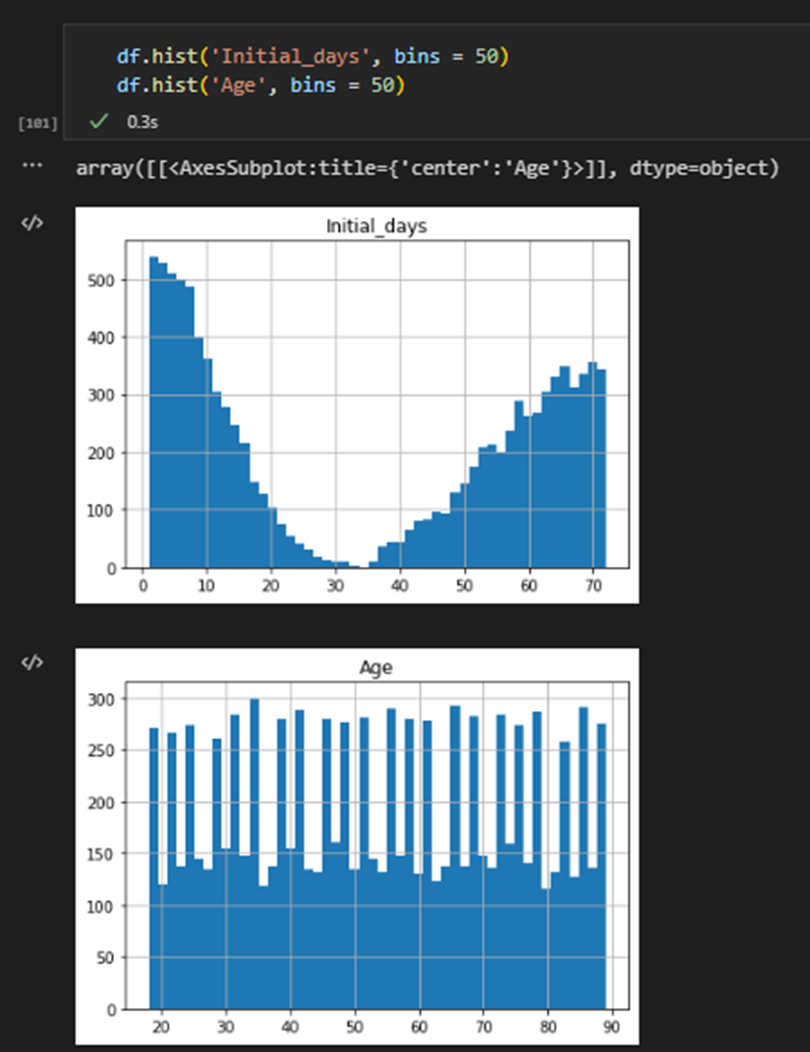


1. Output and Results
   1. In order to better understand what the contingency table was telling me, I visualized the output with a heatmap.
   2. I used the chi-square test of independence, *chi2\_contigency* to call the test statistic and p-value. The p-value is .09 and much lower than the statistic, which means I can reject the null hypothesis—“Readmis” and “Asthma” are correlated.

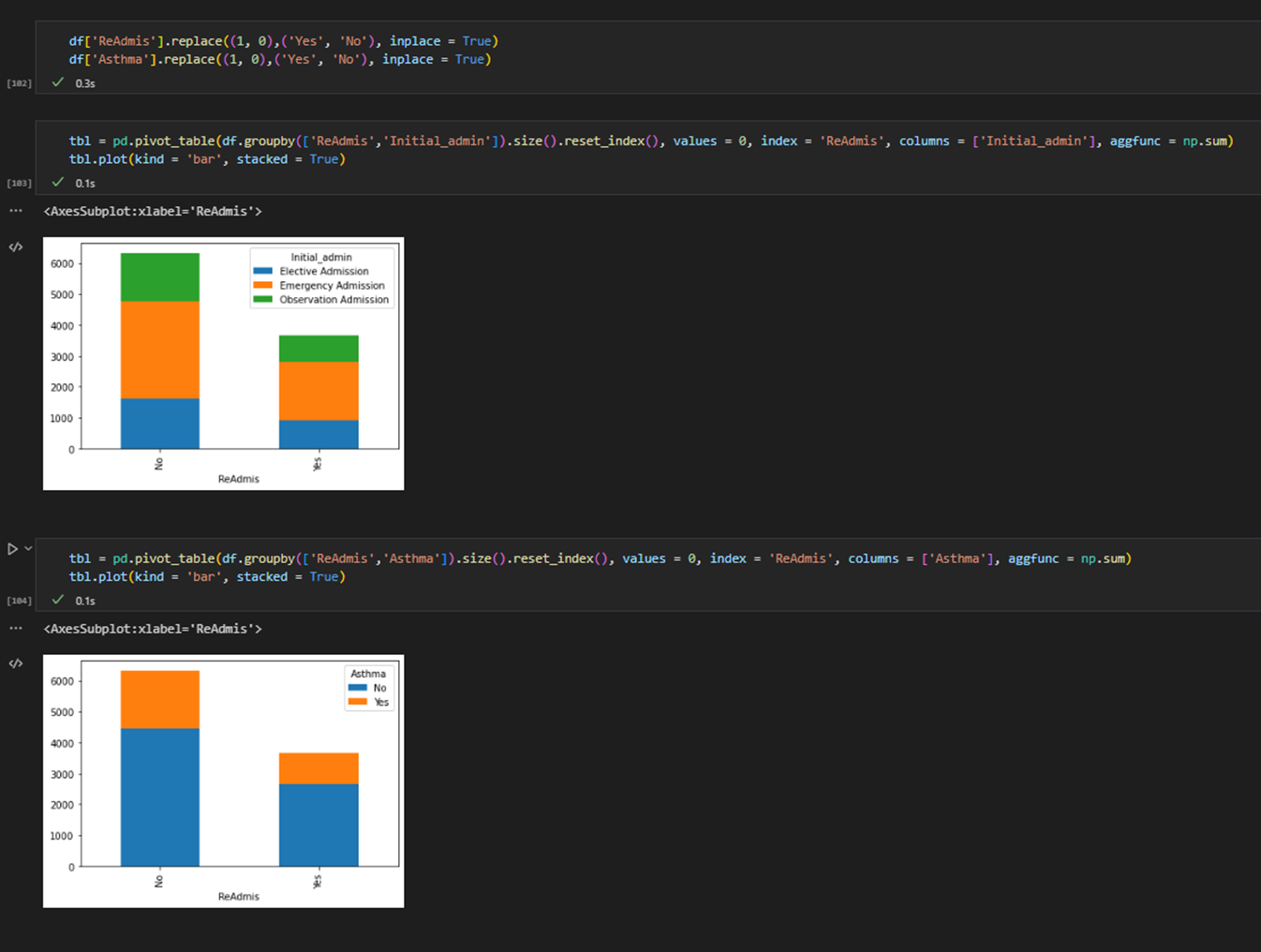
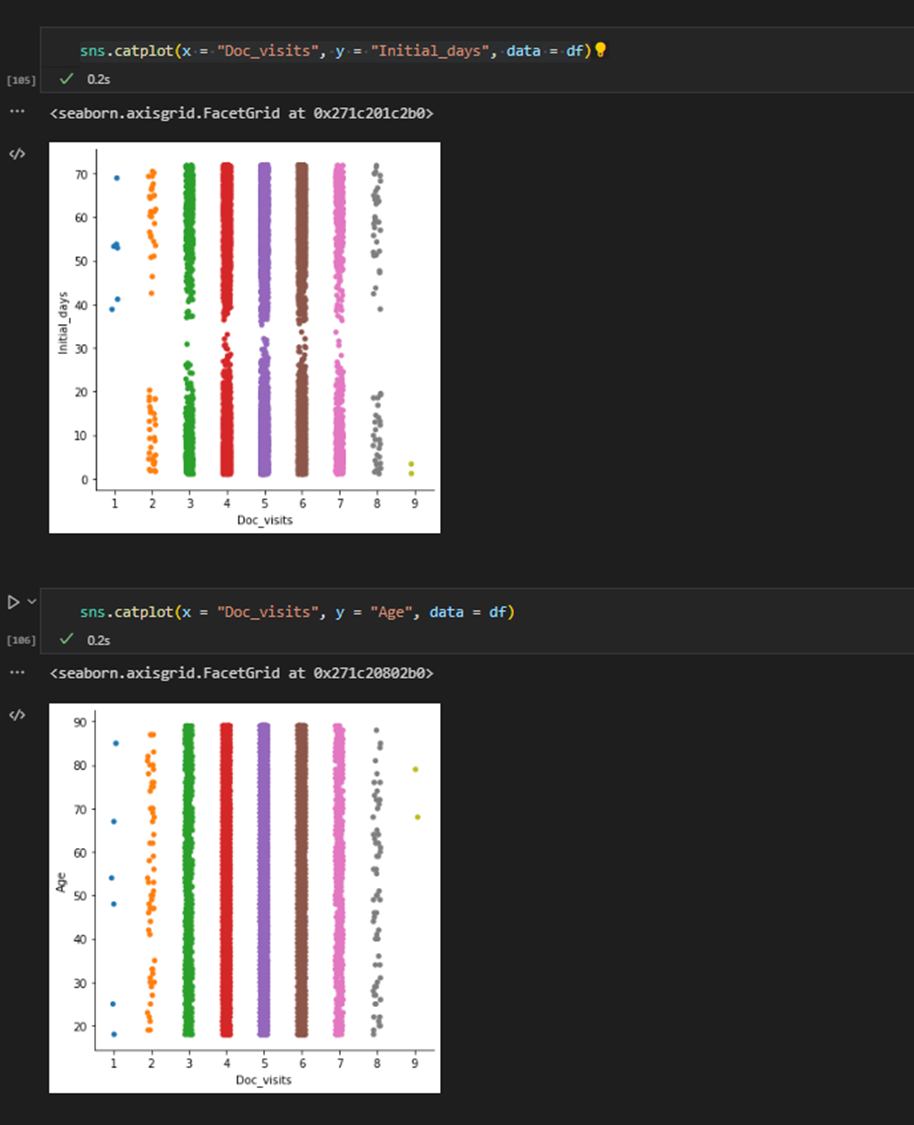


1. Justification
   1. “Readmis” and “Asthma” are both categorical variables and I wanted to test whether or not they were related to one another. As such, the chi-square test was the most appropriate option.

## **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:** Conducting the chi-square test brought the following conclusions:
   1. Per the chi-square test, a relationship exists between Asthma and Readmis, as the p-value of .09 is less than the significance level of 2.85.
2. **Limitations:** There are a number of limitations presented by the data as it is and the test I used:
   1. We would not be able to use the chi-square test on non-categorical variables, should the scope of our test expand.
3. **Recommendations:** Because Asthma was found to contribute to Readmis, this medical center should take mitigation action for patients with this illness.

## **F. Panopto**

https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=551b7951-b5b0-43e1-8f34-af2c0175e1eb

## **G. Reference web sources**

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. Acknowledge sources**

LaPointe, J. (2019) 3 strategies to reduce hospital readmission rates, costs, Rev Cycle Intelligence. Agency for Healthcare Research and Quality (AHRQ). Available at: https://revcycleintelligence.com/news/3-strategies-to-reduce-hospital-readmission-rates-costs (Accessed: October 12, 2022).