**Capstone Project Summary**

**Name: Anthony Pistocchi**

**Project Title: Readmission: A Healthcare Analysis**

**Data Collection Source:** *Describe how the data was collected. Include website links.*

*any updates from the previous**analysis.*

*The data was collected from Canvas, Capstone 2.*

**Project Summary:** *This should be at least 2-3 paragraphs (1-2 pages)***.** *Describe project goals and objectives. Include information regarding the target audience.*

*Examples:*

* *How will they benefit from your findings?*
* *Discuss any updates from the previous analysis including methodology for organizing and cleaning of the data.*
* *Was there any additional research or data added from preliminary analysis?*
* *Describe any statistical methods used.*
* *Any outliers or inconsistencies?*
* *Describe findings and conclusions.* 
  + *Include recommendations.*
  + *Discuss next steps if any.*

The audience will benefit from the findings of the analysis by being educated on why the readmission rates of patients is high at a hospital. For the methodology, the data was organized with Excel and Python, mostly with Python code to see the stats and info. Using Excel, I checked the data for any sort of missing or duplicated data. I made sure to filter some data columns just to get a look. I also made sure to have a sort of working dictionary for some of the columns as well. I then moved to Python to get a full grasp of the data and ran another full data quality check. Most of the tools used were in Python, as any sort of data checks done were in Python. The data did not need much cleaning, but some cleaning was done on Excel. At the moment, I have not converted data types but I will be doing that if needed on Python and SQL. No missing values were replaced, since there would have been no impact on the analysis. There are certain columns that are being excluded, such as most of the medication’s columns, the only medication column that I chose to keep was the ‘Insulin’ column, the rest of the data is staying. When exploring the data I did not see any outliers or inconsistencies.

The size of the data was 101766 rows and 50 columns. The missing values were numbered at 181,168, but the missing values were from columns that were not used in the final data set. The missing values were not replaced or deleted from the original dataset, all of that data was kept as it was provided. While getting the data checked and cleaned on Python, the amount of duplicated values was found to be zero. When the data was explored, there were no outliers or inconsistencies found in the dataset. The data that was provided had no dates given since this is patient data. The findings of the analysis suggests that diabetes and age have a big effect on readmission rates with patients. Taking a look at the demographics of the patients shows that the age of the patients who were over 51 years old had a much denser population compared to younger ages. Also, the same readmission rates were effected by diabetes, with patients who were not only prescribed a diabetic medication, but with insulin injections as well. Just with emergency patients, a good chunk of patients who had diabetes and who were over 51 years old had the biggest population. In conclusion, with the data that was analyzed, the readmission rates were found to be influenced by many different factors such as diabetes management, demographics such as age, and other complications.