**Analyzing the Clinical and Financial Data of Patients**

Step 1: Collected data from 4 excel files (Demographics, Bill ID, Bill Amount, Clinical Data). Based on Patient ID, I added the Demographics info of the patient in Clinical Data. With Bill reference, the patient ID’s total sum of amount paid per date is known. The Total amount is added to the clinical data.

Step 2: The 3400 Data is generated from the dataset.

Step 3: The Patient ID with latest date of admission is taken out for calculating the medical history, symptoms, medications taken by the patients. The 3000 Data is identified after removing the previous date of admission.

**Python coding in Jupyter**

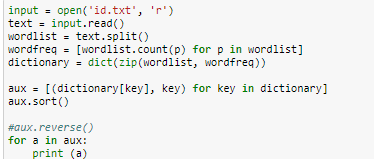
Step 4: In the Jupyter Notebook, the function NLTK WordNetLemmatizer is used to analyze the occurrences of medical history, preop medication and symptoms of the patients.



The symptom 4, medical history 2 and Preop medication 3 is highly prevalent among the patients.

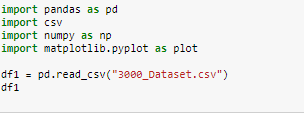
Step 5: The df function will display the contents of the table which has no header.

  
Step 6: Collected the Patient ID in separate text file to see the occurrences of individual IDs using the dictionary function.



The patient ID ‘4e46fddfa404b306809c350aecbf0f6a’ is frequently admitted for four times.

Step 7: The dataset with the header is analyzed in python.



Step 8: The occurrences of the race which has unique id is obtained. The Chinese race is found to be the high frequency.



Step 9: The occurrences of the gender which has unique id is obtained. The Male is found to be highly frequent



Step 10: The occurrences of the resident status which has unique id is obtained. Among all, Singaporeans are coming often.



Step 11: The occurrences of id with different date of admission is printed.



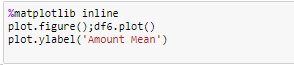
Step 12: The mean value for all the columns is obtained with the resident status as unique. The total amount paid by the foreigners is quite high compared to other status.



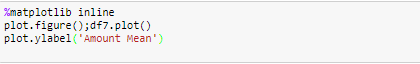
Step 13: The mean value for all the columns is obtained with the race as unique. The total amount is contributed by the Malay people.



Step 14: Plotted the graph among the race and their total amount mean.



Step 15: Plotted the graph among the resident status and their total amount mean.



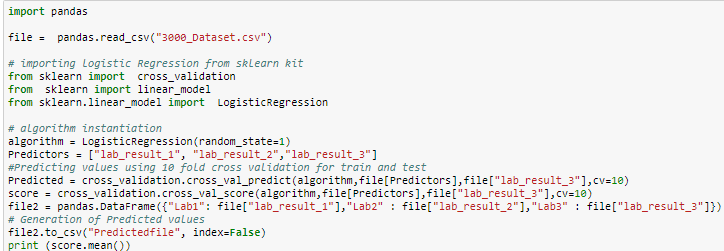
Step 16: The mean value for the lab result 1, lab result 2 and lab result 3 is obtained with the gender as unique.



Step 17: Plotted the graph among the gender and their lab result 1,2 and 3.



Step 18: The values of lab result 1,2 and 3 are predicted using cross validation of 10 folds and linear model logistic regression.

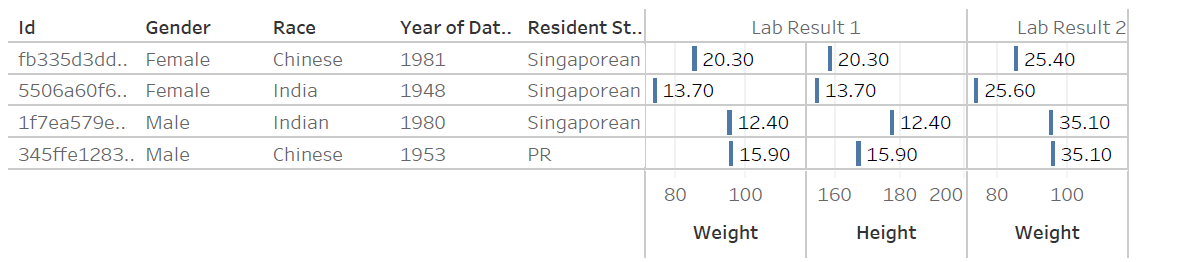


Step 19: With the random column as label in new excel file, the grid search algorithm is used to calculate the accuracy, precision and recall values of the table.

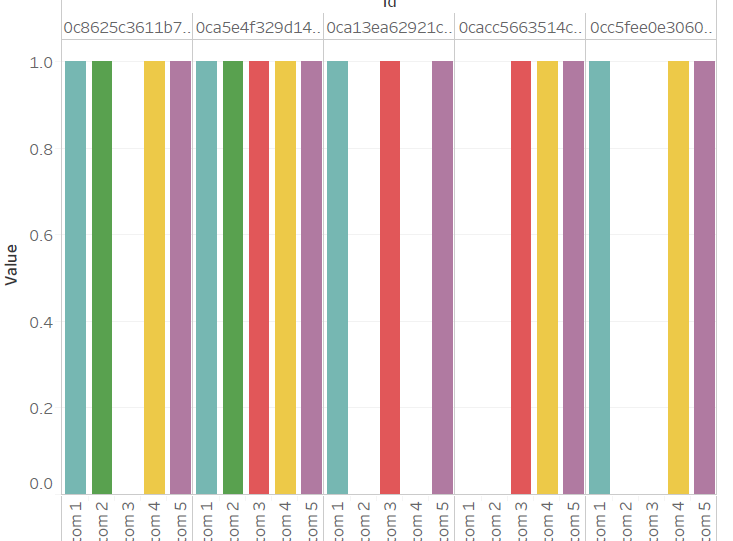


**Visualization in Tableau**

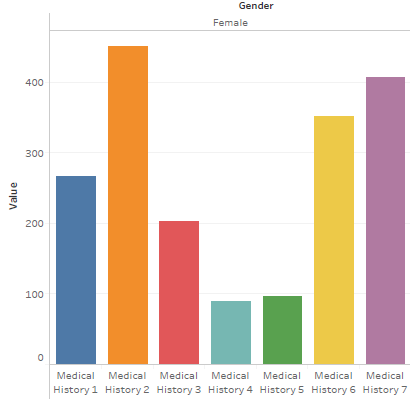
* The highest value of Lab result 1, 2 and 3 along with the ID, Race, Date of Birth, Resident status, height and weight is obtained.



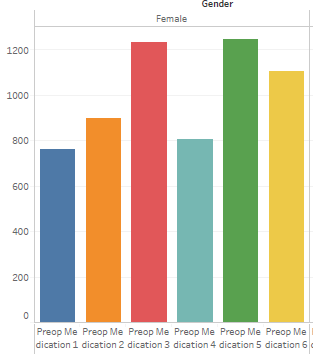
* The Patients with any or all the Symptoms 1,2,3,4 and 5 is easily visualized through the bar chart form



* The Medical history 1,2,3,4,5,6 and 7 which is prevalent mostly in male or female is identified



* The Preop medication 1,2,3,4,5 and 6 which is prevalent mostly in male or female is identified



* The count of Symptoms with year of admission and gender is shown in tabular form. It is found that symptom 4 is common with both Male and Female.

