Course Project – Summary Document

Shaun Pritchard

Rasmussen College

QMB4400

Tina Harclerode

April 11, 2021

**Course Project – Summary Document**

Looking at the data file we have 14 category features that give information on the patent, the patient's age, the type of element or disease they have uses schedule data, the neighborhood they are located, the patients age, gender, whether or not they received an SMS text, and whether or not they showed up for the appointment or not. From observing the data, I found that there are 87531 no shows in total out of 110528 features totaling 22997 no shows or 79.19% in the "No show" feature column. 677 Null values in the "No-show" feature column would have to be cleaned. Also, other dataset feature columns need to be cleaned to make sure that the analysis is correct Justified.

The data provides some useful feature columns that could determine useful correlations. understanding the correlation between the age of the patients not showing up could give insight it is whether or not it was younger patients more prone to be no-shows or older patients? The data potentially show us insights on the type of disease the patient is dealing with which might have some reasoning as to why they no show. Maybe physical limitations or dependencies? The data gives us locations that can be used to identify correlations of no-shows specific to neighborhoods.

The data shows potential correlations between those patients who have received SMS text or not. Missing SMS text could be a contributing factor two missing the appointments. The data shows the potential to determine and if more males or females are prone to being no-shows. We could also use the data to do Inferential analysis on multiple feature variables that could determine probability up scheduling, specific times of appointments that are not being met concerning the patient gender, location, and medical issue. In-turn the data can be used for forecasting or identifying specific outliers? Actions could be met to facilitate internal better be accustomed to patient care and less missed appointments. Overall, the data should be sufficient enough to tell us what determinant factors or underlying trends for the high quantity of no-shows.

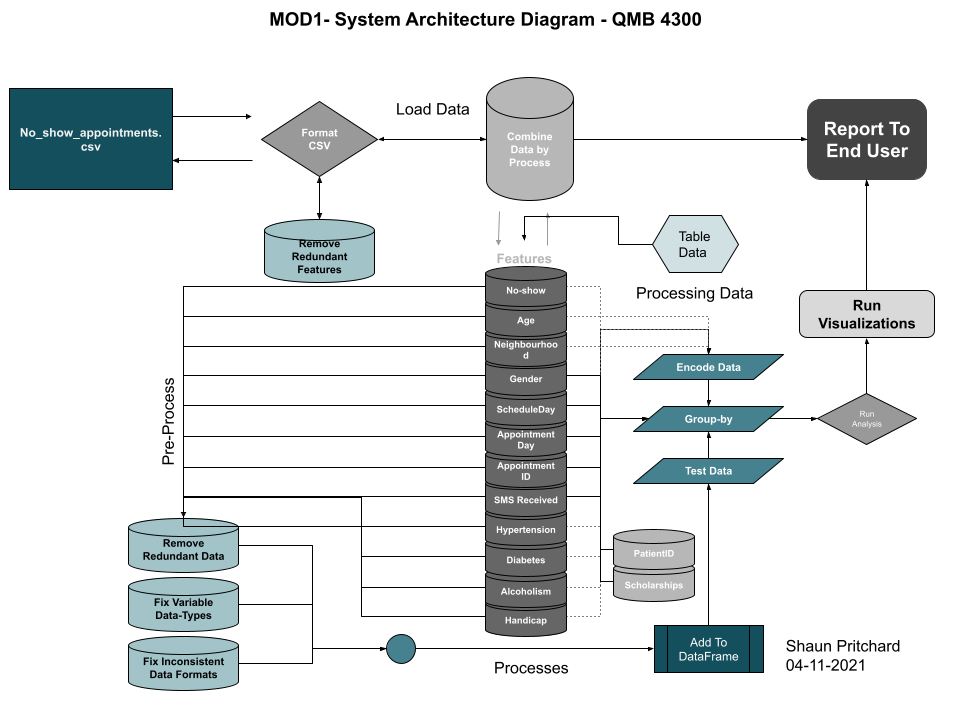
**Missing data:**

Yes, I would say there are several features is missing from the data set. For starters, there is no name associated with the patient ID. By having the patients name, we could determine if there are any patterns or Trends with specific patients who are missing appointments and being inconsistent with their patient care. Having data that describes features for the types of appointments that are being missed would be something of Great Value to this project. This would help the under the correlations between missing appointments based on the type of appointments being received by the patients. We do not have the contextual name of the appointments, but we do have identifying appointment numbers that could be used to determine which appointment types are more frequent no-shows. Also, that more detailed information surrounding the patient medical issue would provide a more feasible depiction of the issue. Also, having a phone number or contact information would be important to this dataset. It could potentially be a factor caused due to inaccurate contact information creating conflicts with scheduling or notification of prior scheduled appointments.

**Justification for additional data:**

By having the patients name data, we could do a detailed descriptive analysis to find out which patients are missing more than others. This would be able to pinpoint specific patients who might be having issues with the times being scheduled. This could lead to the provider implementing solutions to schedule appointments contingent on the specifications of the patient. Having more detail on the actual patient medical issue could be a determining factor in scheduling specific patients with specific issues at the wrong time. Inaccurate contact information can lead to patients missing appointments. Therefor making sure that the contact information is readily available and correct to show correlations between patients who might need to be updated. More detailed location information could help to see if there are any correlations in patients who m9is from specific areas.

**System architecture diagram:**

****