Prof Meeting Minutes #3

Date/Time	19th Jan 2017, 10am to 11am
Venue	School of Information System, Meeting Room 4.3
Attendees	Aishwarya Agarwal, Nasrullah Bin Khairullah and Zoey Loh
Agenda	1. Share our exploratory analysis with Prof Kam 2. Share our variable selection process and findings

S/N	Notes/Task	Actor	Follow up Action
1.	Our exploratory data analysis showed us missing values for variables like race, visit number and revised nationality. Our approach to input these missing values was through sorting, but Prof showed us an easier way to impute the missing values, a change we'll make.	Nas	Carry out the change recommended by Prof
2.	In the case of the names of the attending doctors and psychologists, we saw many variations of the same name. So we decided to recode the names of the doctors to avoid any overlap.		
3.	There were also a few anomalies in terms of the genders of the patients. Based on the other details recorded, we imputed the gender of the patients. Prof also suggested us to ask our sponsor for clarifications about these anomalies		
4.	For the age variable, when we ran a distribution check, we observed skewed data. Though our data is primarily for children, we observed some information for people above the age of 25 as well. Prof suggested us to contact our sponsor and ensure the focus of the study to be on patients who are 20 years and below. Subsequently, we can remove the outliers in the data.	Nas	Send an email to clarify (Completed)
5.	The next step was variable selection. Based on Prof's suggestion, our dependent variable consists of three outcomes, no-show, cancellation and attendance. Prof told us to focus on analyzing the relationship between the dependant variable and other independent variables. Prof told us that it is essential to ensure that our independent variables are not biased, and has a random distribution	All	Relook at how we analyze the relationship
6.	For the purpose of modeling, Prof told us about two different approaches we could take, a patient approach or an episode approach. For the patient approach, we would have to aggregate out the	All	Prepare the data accordingly to the 2 approaches.

patient's based on their number of visits. Each patient would be considered as one data point. For the episode way, each visit is treated individually. The problem here would be that a patient could be repeated several times, depending on the number of visits.

Prof recommended us to come up with a strategy to choose one of the two ways. This would depend on the perspective we choose, the patient's or the hospitals and also the kind of solution we want to present