MGMT 571 Homework 1

1. The dataset contains an input variable called "Car Brand" which includes all the car brands in the market. What measurement level is the most suitable to describe this variable?
   (a) Binary
   (b) Nominal
   (c) Ordinal
   (d) Interval

2. Consider the following table, what is the Chi-square (χ^2) statistic and the degree of freedom for testing the independence of two variables Student and Voting?

	Voting		
Student	A	B	C
Yes	60	100	30
No	60	90	20

3. Customer retention is important in view of the customer lifetime value, meaning that customers who stay with a company for a very long time give more value to the company than customers who often switch service providers. Churn, also called attrition, is a term used to indicate a customer leaving the service of one company in favor of another company. Churn analysis is prevalent in competitive industries such as mobile communication services, internet services, and retail banking services because customers may easily switch among companies looking for new features, promotion and other factors. There have been many case studies available online. The churn data set was taken from the UCI Repository of Machine Learning Databases at the University of California, Irvine. The data set contains information on 20 variables for 3333 customers, along with an indication of whether or not that customer churned. The 20 potential predictor variables are as follows:

   - State: for the 50 states and the District of Columbia
   - Account length: how long account has been active
   - Area code: phone area code
   - Phone number: essentially a surrogate for customer ID
   - International Plan: yes or no
   - VoiceMail Plan: yes or no
   - Number of voice mail messages
   - Total day minutes: minutes customer used service during the day
   - Total day calls
   - Total day charge: perhaps based on foregoing two variables
   - Total evening minutes: minutes customer used service during the evening

- Total evening calls
- Total evening charge: perhaps based on foregoing two variables
- Total night minutes: minutes customer used service during the night
- Total night calls
- Total night charge: perhaps based on foregoing two variables
- Total international minutes: minutes customer used to make international calls
- Total international calls
- Total international charge: perhaps based on foregoing two variables
- Number of calls to customer service

Use data sets **Churn.sas7bdat** or **Churn.csv** from Brightspace and answer the following questions. **For each question, please provide your short answers and the evidence from SAS EM outputs if needed**. Remember we only use the above 20 potential predictor variables and one target variable ("Chrun"). The other variables should be set to "Rejected" when you read in the data.

(a) Create a project in SAS EM using your last name + Churn, e.g., **Sun Churn**.
(b) Create a library and import the **Churn.sas7bdat** data to the library.
(c) Also, try to import the csv data **Churn.csv** directly using File Import node.
(d) Determine the variable role and measurement level for each variable.
(e) Plot a 2D scatter between **Total international charge** and **Total international minutes**. Discuss your findings.
(f) Use a graph to determine visually whether there are any outliers in **Total day calls**. Discuss how you will deal with the outliers.
(g) Does a 2D scatter plot between **Number of calls to customer service** and **Total day charge** (group by the target variable) reveal any information?
(h) Utilize the Chi-square table with a significance level of 0.05 to determine which, out of the 20 predictor variables, are useful in predicting customer churn.

Variable	Variable Role	Measurement Level
State		
Account length		
Area code		
Phone number		
International Plan		
VoiceMail Plan		
Number of voice mail messages		
Total day minutes		
Total day calls		
Total day charge		
Total evening minutes		
Total evening calls		
Total evening charge		
Total night minutes		
Total night calls		
Total night charge		
Total international minutes		
Total international calls		
Total international charge		
Number of calls to customer service		
Churn		