DAPT 622 (Statistics II) Assignment #2

Due: By 11:59pm on Sunday, February 10

As a team, you are to perform the tasks listed in Parts I and II below. Organize your analyses and conclusions in a typed document (Word, Excel, PowerPoint, LaTex formats are all fine). Be sure to include all relevant software output in your submission.

This assignment is to be submitted electronically **by 11:59pm on Sunday, February 10, 2019** via Blackboard**.** All team member names should be included somewhere in this document.

Part I. Consider the candy bar data again. A study was conducted to investigate the nutrition content of a variety of candy brands and products. The data set includes the following variables (most of which should be self-explanatory):

Brand – Candy brand
Name – Specific name of candy product (e.g., Snickers, Kit Kat)
Oz/pkg – Weight in ounces per package
Calories
Total fat g
Saturated fat g
Cholesterol g
Sodium mg
Carbohydrate g
Dietary fiber g
Sugars g
Protein g
Vitamin A %RDI
Vitamin C %RDI
Calcium %RDI
Iron %RDI

Perform a Factor Analysis on this data by carrying out the following steps:

1) How many factors should be retained (based on eigenvalues and a Scree plot)? You may use any "Factoring method" or "Rotation method" you wish. The defaults, of course, are fine.

2) Provide the rotated factor loadings matrix. You should notice that (as compared with Principal Component Analysis) fewer variables load onto each factor. Interpret (as best as possible) your factors. Can you give an overall name to each factor?

3) Plot factor scores for your first two factors. Color the points by Brand. Are there any natural groupings of observations?

Part II. Consider the demographics data (Demographics.jmp or Demographics.txt). Each of the 440 rows contains demographic information on particular localities in the United States. The following is a brief description of the variables.

Variable Name	Variable Description / Label
County_name	County, parish or city name
State	
Region_num	1=Northeast, 2=North central,3=South, 4=West
Location	East or West of Mississippi River
Square_miles	Land Area (Square Miles)
Population	Population in County/City/ Parish
Pop_Size_Group	Small, Medium, Large
Pct_Age18_to_34	% = (# aged 18-34) / (population)
Pct_65_or_over	% = (# aged 65 & up) / (population)
Num_physicians	Active Nonfederal Physicians
Num_hospital_beds	Hospital Beds, Cribs, Bassinets
Num_serious_crimes	Serious Crimes
Pct_High_Sch_grads	% = (# 12 yrs school) / (# 25 & over)
Pct_Bachelors	% = (# bachelor's) / (# 25 & over)
Pct_below_poverty	% = (# < poverty) / (population)
Pct_unemployed	% = (# unemployed) / (population)
Per_cap_1990income	Per Capital Income
Total_personal_income	Total Personal Income

1) Perform a **linear discriminant analysis** using the highlighted variables for discriminating the variable Pop_Size_Group.

2) Show a plot of the two discriminant functions with points colored by Pop_Size_Group. Do both discriminant functions appear necessary to describe group separation?

3) Give the scoring coefficients for the two discriminant functions. Which variables appear most important for describing group separation?

4) Provide a classification matrix and the percent of observations misclassified.