Assignment 1

Audit File: CARS 1993

Sukanya Mukherjee

128041217

BAN100ZBB

Statistics for Analytics

Samaneh Gholami

Assignment 1 (Audit) File: CARS 1993

Question: You are a new consultant at work, your manager gave you a dataset a client has sent over. She wants you to perform a data audit to ensure the data is received and understood properly before analysis.

Given: attached Excel dataset.

Outcome: a data audit report consisting of the following:

1. Metadata description
2. Descriptive statistics
3. Your observation, commentary and/or questions you would like to find out about the data from the client

Submission:

Prepare the report using PowerPoint, please export as PDF. Submit the PDF under the Assignment 1 drop box.
Grading policy are work-place standards -- you want to work professionally, right?

A - you did a very thorough job and asked good questions about the data. I'm impressed with the newly hired consultant.

B - you did a better than average job, made some good observations, asked some thoughtful questions.

C - it's an average job, meaning anybody could've done it, while technically correct, there's nothing special about the work, you don't stand out from other analysts.

D - your work has some technical flaws in it, it can be fixed but I've lost faith in your ability to do work unsupervised, you are going to need more supervision than normal, not a good first impression

F - awful work, there's nothing worth saving, I give the dataset to someone else to audit. We made a mistake hiring you, most likely you won't make it pass probation

F - if you are late. Late at work is unacceptable. If the team can't count on you to deliver on schedule, it's unacceptable. Work will not wait for you.

Weight of assignment in course: this is a warm-up assignment, it's only 10% of the course.

Good Luck.

Answer: Code

Statistics.sas

CODE LOG RESULTS OUTPUT DATA

```
1  libname carslib "~";
2  /*autocorrects the column name and uses SAS standards*/
3  options validvarname=v7;
4  /*importing the data from excel file*/
5  proc import
6          datafile="/home/u59406283/File CARS_1993.xlsx"
7          out=carslib.cars1993
8          dbms=xlsx replace;
9  run;
10 /*Metadata description-contents*/
11 proc contents data=carslib.cars1993;
12 run;
13 /*Descriptive statistics-means,freq,univariate*/
14 proc means data=carslib.cars1993;
15 run;
16 proc freq data=carslib.cars1993;
17 run;
18 proc univariate data=carslib.cars1993;
19 run;
```

Results:

https://odamid-usw2.oda.sas.com/SASStudio/sasexec/submissions/7a426356-5307-472b-9293-691e001a7fd1/results

OR the file : Statistics-results.html

1. ## Metadata description

Metadata is data dictionary information, i.e., data about data.
With the CONTENTS procedure in SAS we can list information about our SAS datasets, including the following:

- Dataset name
- Date of creation
- Number of observations and number of variables
- Whether the dataset is sorted, compressed, indexed
- Information for each variable
    a) Position (sequential number) in the dataset
    b) Variable name
    c) Type

d) Length
e) Format and Informat
f) Label

2. Descriptive statistics

PROC MEANS:
We can get descriptive statistics by using proc means
PROC FREQ:
We can use proc freq to produce frequency tables.
PROC UNIVARIATE:
We can use proc univariate to get more detailed summary statistics

3. Your observation, commentary and/or questions you would like to find out about the data from the client

Data analysis:

1) **Observation/Commentary:** Firstly, **basic units of all columns are missing**. Unit of measurement is quite needed for correct estimation or research of the components. For example: Currency value for prices and volume measurement for fuel etc.
2) **Question:** Chevrolet Corvette and Mazda RX-7 have **missing values** for Luggage/Rear Room columns in the data provided. Upon research, I found that both have some space for luggage/rear room. Is there any reason behind this or it was a miss?
3) **Observation/Commentary:** Similarly, all cars in Van category, have **no values** in Luggage column. As we do not know if, it denotes number of luggage a car can contain or how much room in length, width, height, the car contains to store the luggage, we aren't able to further ask questions regarding this.
4) **Question:** Which **continent/country/region** is this data being referred to?
5) **Question:** What **age-group** is this data relevant to? Is there a **specific trend** in age-group which buys any **specific type of cars**?
6) **Question/Observation:** What is the value of **number of cars sold** in 1993 and how many in which **months**? This can further let us know about the affect of **seasonal sales of each car type/name**.
7) **Observation/Commentary:** Frequency table shows us that Chevrolet and Ford have the **highest percentage of cars sold** – 8.60 while BMW, Infiniti, Plymouth, Saturn, Saab, and Suzuki have **least percentage of cars sold** – 1.08.
8) **Observation/Commentary:** Frequency table shows us that **midsize and small** category cars are preferred over **Large and Van** category cars while buying.
9) **Observation/Commentary:** Frequency table shows us that, **About 23%** of the cars are having an **average of 18-19 km per litre in city** while **26-28 km per litre on highway**.
10) **Observation/Commentary:** About 53% of the cars have 4 cylinders.
11) **Observation/Commentary:** About 44% of the cars have 5 passenger seats.