Assignment 1 - Data Quality Report

Marks are going to your reasoning / understanding and complete handling strategies for poor data not that you can calculate the mean etc.

Deliverable: Upload a single word document to Moodle.

Using the sample data file below create a Data quality report. Your report should contain the following:

(1) Tabular reports for each **continuous feature** containing the following information:

* min,
* max,
* median,
* mode
* 1st quartile
* 3rd quartile
* standard deviation
* total number of instances
* % missing value for feature
* cardinality of each feature

(2) Tabular reports for each c**ategorical feature** containing the following information:

* total number of instances
* % missing value for feature
* cardinality of each feature
* mode
* mode frequency
* mode percentage
* 2nd mode
* 2nd mode frequency
* 2nd mode percentage

(3) Histograms of each feature

(4) Discussion on histogram distribution of each feature

(5) Identify data quality issues from above and discuss potential handling strategies (pg 16 of notes)

(6) Using a scatter plot diagram look at the relationships between each feature in relation to each other.

Use the notes Lecture 5 - Data Exploration to assist you in answering the above.