**Data Mining Tools and Techniques**

**Homework – Analyzing Data, Posing Questions**

**Summary**:

For this assignment, I will be working with different types of data sets. My goal is to understand and characterize one data set and pose data mining question against three/four datasets.

There are total of 4 different datasets:

Dataset 1: Car Dealership Sales

Dataset 2: USA accident Data

Dataset 3: Volcano Eruption Data

Dataset 4: Airbnb

For dataset 1:

1. I provided descriptive statistic for each attribute as they apply. This may include minimum, maximum, average, etc.
2. For each attribute, explain whether it is descriptive, discrete, continuous, or discontinuous?
3. Is this a supervised or unsupervised data set? If supervised, what is the class variable(s)?
4. Is this data set time-series, temporal, spatial data, sequence, some combination, or none of these?
5. What measurement scale is the dependent variable (if it exists, it is usually the right-most column)?

For dataset 2 to 4:

I formulate 3 specific questions against datasets. In that, there are one questions based on classification, regression and clustering type.