Assignment

1. University of Arizona was sued for sex discrimination. Its admission data showed that men applying to graduate school at Arizona were more likely to be admitted than women. The graduate schools had just accepted 44% of male applicants but only 35% of female applicants. The difference was so great that it was unlikely to be due to chance.

   Experts analyzed the data set in detail and find out that discrimination has been carried in a way. Do you agree after analyzing the data set or you have a different conclusion? (Dataset: university.csv)

2. Analyze the Aircrash dataset and find out the following: (Dataset: Airplane_Crashes_Since_1908.csv)
   a. Yearly how many planes crashed? how many people were on board? how many survived? how many died?
   b. Highest number of crashes by operator and Type of aircrafts.
   c. 'Summary' field has the details about the crashes. Find the reasons of the crash and categorize them in different clusters i.e Fire, shot down, weather (for the 'Blanks' in the data category can be UNKNOWN) you are open to make clusters of your choice but they should not exceed 7.
   d. Find the number of crashed aircrafts and number of deaths against each category from above step.
   e. Find any interesting trends/behaviors that you encounter when you analyze the dataset.

Make the visualizations (graphs, pie etc) for the above assignments as you feel appropriate.

You are supposed to communicate your analysis and technique in simple terms for non-technical users as well as technical people as part of recruitment process. Please fit into "20 mins + 10 mins discussion" format that will be held via eConference software (Skype/Lync/WebEx etc.).

You are open to do analysis, clustering and visualization on any tool or set of tools of your choice and feel free to use any language if necessary like R, SQL, Aster-SQL etc.