|  |  |
| --- | --- |
| **Business Understanding** | Describe the purpose of the data set you selected (i.e., why was this data collected in the first place?).   * Track crime * Classification of crimes * Response times   Describe how you would define and measure the outcomes from the dataset. That is, why is this data important and how do you know if you have mined useful knowledge from the dataset?   * Predict response time and arrests made * Victim profile analysis – minors, females etc. * Police performance – qtr, month etc. * Seasonality / crime cycles / * Response time vs location/community * Population density  How would you measure the effectiveness of a good prediction algorithm? Be specific. * Cross validation * Dividing the data set into test and train |
| **Data Meaning Type** | Describe the meaning and type of data   * Scale * Values * Types |
| **Data Quality** | Verify data quality:   * Explain any missing values * Duplicate data * Outliers * Typos * Standardization * Normalization  Are those mistakes? * Why do you feel they are? * Reasons?  How do you deal with these problems? * Deletion * Imputation  Give justifications for your methods. – WHY? |
| **Simple Statistics** | Visualize appropriate statistics   * range * mode * mean * median * variance * counts * quartiles for a subset of attributes. * Box plots * Violin * Scatter * Correlation matrices * Bar charts  Describe anything meaningful you found from this or if you found something potentially interesting. Note: You can also use data from other sources for comparison.   Explain why the statistics run are meaningful. |
| **Visualize Attributes** | Visualize the most interesting attributes (at least 5 attributes, your opinion on what is interesting).   Important: Interpret the implications for each visualization.   Explain for each attribute why the chosen visualization is appropriate. |
| **Explore Joint Attributes** | Visualize relationships between attributes: Look at the attributes via scatter plots, correlation, cross-tabulation, group-wise averages, etc. as appropriate.   Explain any interesting relationships. |
| **Explore Attributes and Class** | Identify and explain interesting relationships between features and the class you are trying to predict (i.e., relationships with variables and the target classification). |
| **New Features** | Are there other features that could be added to the data or created from existing features?   Which ones? |
| **Exceptional Work** | You have free reign to provide additional analyses.   One idea: implement dimensionality reduction, then visualize and interpret the results. |