|  |  |  |
| --- | --- | --- |
| **Category** | **Points** | **Description** |
|  |  |  |
| Business Understanding | 10 | Describe the purpose of the data set you selected (i.e., why was this data collected in the first place?). 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? How would you measure the effectiveness of a good prediction algorithm? Be specific. |
| Data Meaning Type | 10 | Describe the meaning and type of data (scale, values, etc.) for each attribute in the data file. |
| Data Quality | 15 | Verify data quality: Explain any missing values, duplicate data, and outliers. Are those mistakes? How do you deal with these problems? Give justifications for your methods. |
| Simple Statistics | 10 | Visualize appropriate statistics (e.g., range, mode, mean, median, variance, counts) for a subset of attributes. 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 | 15 | 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 | 15 | 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 | 10 | 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 | 5 | Are there other features that could be added to the data or created from existing features? Which ones? |
| Exceptional Work | 10 | You have free reign to provide additional analyses. One idea: implement dimensionality reduction, then visualize and interpret the results. |
|  |  |  |
| Total | 100 |  |