**Intro to Data Science**

**Robertson**

**Final Project**

**Due: Wednesday, Dec 14th**

**Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**(print first and last name legibly)**

In pairs (or a group of three with instructor permission), produce a report on a data set of your choosing.

1. Present the scope of your project as a five-minute presentation slide show. This is similar to what you would do to get funding/the go ahead to proceed with your project. What are you going to do and what results are you hoping to find
2. Import and clean your data so it can be used by linear regression, association rule mining, support vector machines, and k means. You may need to produce different cleaned data sets.
3. Run at least 15 experiments. You must run at least 2 experiments of each type. Some of these experiments can produce bad results but you will be expected to have at least one successful experiment as you will need to write about the implications of it in your paper.
4. Put together a five-to-ten-page paper detailing your results. Briefly analyze the results of each experiment in this paper. A template with appropriate headings will be provided.

|  |  |
| --- | --- |
| **RUBERIC** |  |
| **Presentation** | **15 points** |
| **Successfully import your data** | **10 points** |
| **Transform your data** | **20 points** |
| **Comment your code** | **10 points** |
| **Run 15 experiments with comments** | **20 points** |
| **Paper** | **25 points** |
|  |  |
|  |  |
|  |  |
| **PENALTIES** |  |
| **Forget to submit code** | **-40 points** |
| **Forget to submit report** | **-40 points** |