**The project is concerned with:**

* Data Cleaning
* Feature Engineering (transformation)
* Exploratory Analysis

**HW1 Excerpt**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Question # | Question |  |  |  |
| 5 | What problem are you trying to solve? |  |  |  |
| 6 | Describe the attributes of the dataset.  If collecting data, estimated time to finish data collection |  |  |  |
| 7 | Give example of records.  Why is this a noisy data set (highlight noisy attributes) |  |  |  |
| 8 | What kind of feature engineering opportunity do you see?  Examples:   * Extract portion of a column * Transformation of values |  |  |  |

**Final Rubric Excerpt**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sections** | **Questions** |  |  |  |
| Dataset and Motivation | * How/why was the dataset collected? * Description of the metadata of your dataset |  |  |  |
| Task Definition/ Research Question | * What real world problem are you trying to solve? * What are the input and output of your analysis |  |  |  |
| Literature Review | * What other work has been done in this area, and how is our work novel compared to others? * Give context to our work |  |  |  |
| Quality of Cleaning | * Data cleaning and type conversion activity. Share anything unusual faced. * What did we do about missing values and why was that choice justified? * New feature/attribute creation and data summary statistics and interpretation. Note: anything in our report we need to tell why we put it in there-how does it help answer our question. |  |  |  |
| Visualization | * Data visualization activity – (box, bar, violin, pair, etc. to see relationships and distribution). For plot: what does it tell you about the data   + Answer your question?   + Inform the model you should use? * Describe anything you find in the data after each visualization. * What data visualization helped you understand about data distribution * What we did about possible outlier as per the data distribution visualization. Did we confirm w/client re: actually an outlier or put a disclosure in our notebook if we decide to remove it? |  |  |  |
| Choice of Model and Why | *I’m not sure where this fits but he said it’s part of the deal. I think it might be part of the dataset and motivation which he mentioned moving to the end.* |  |  |  |