Milestone 1 Report

* Preprocessing techniques:

1. Removing any rows that contains NULL values
2. Select only related columns based on correlation
3. Apply label Encoder to categorical features
4. Featuring scaling

* Regression techniques:

1. Multi Linear Regression (MSE=1.85)
   * #feature=3
2. Polynomial Regression (MSE=1.6)
   * Polynomial degree=input
   * using Label encoder

* Dataset:

1. size after applying pre-processing=412102 record
2. Training set =70 %record
3. Testing set=30% record

* Feature Selection:

|  |  |  |
| --- | --- | --- |
| Feature | Status |  |
| Hotel\_Address | Discarded |  |
| Additional\_Number\_of\_Scoring | Discarded |  |
| Review\_Date | Discarded |  |
| Average\_Score | Used |  |
| Hotel\_Name | Discarded |  |
| Reviewer\_Nationality | Discarded |  |
| Negative\_Review | Discarded |  |
| Review\_Total\_Negative\_Word\_Counts | Used |  |
| Total\_Number\_of\_Reviews | Discarded |  |
| Positive\_Review | Discarded |  |
| Review\_Total\_Positive\_Word\_Counts | Used |  |
| Total\_Number\_of\_Reviews\_Reviewer\_Has\_Given | Discarded |  |
| Tags | Discarded |  |
| days\_since\_review | Discarded |  |
| lat | Discarded |  |
| lng | Discarded |  |

