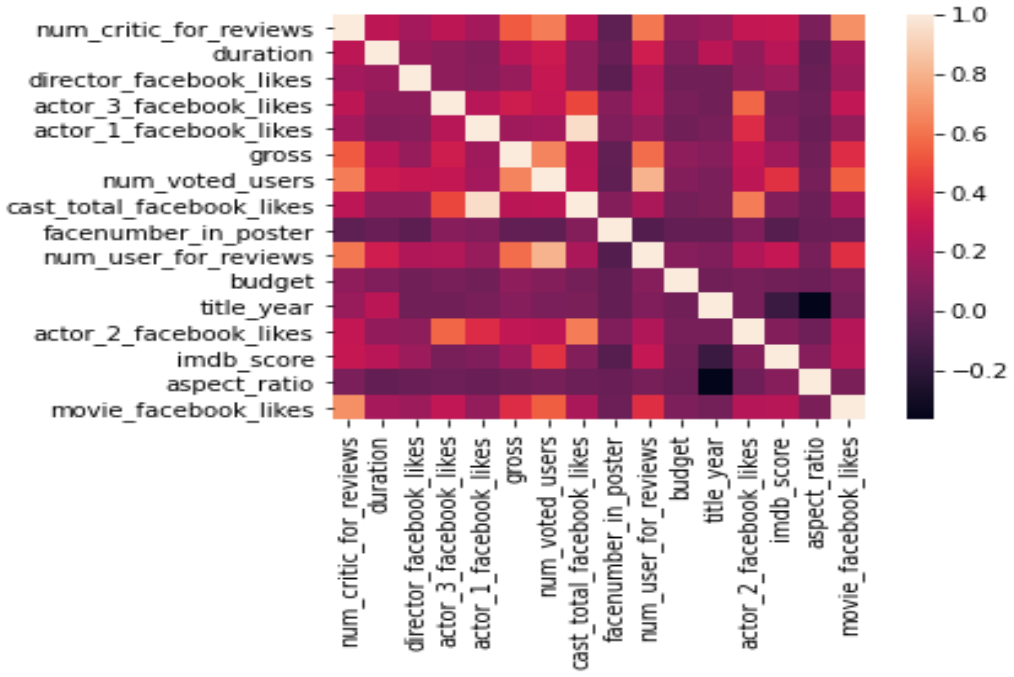
SI&MT Assignment-2 Documents For output Result

Dataset Specifications: -

- Dataset Name – Movie Metadata (https://www.kaggle.com/karrrimba/movie-metadatacsv)
- This dataset consists of 5043 rows and 28 columns

Pairwise Correlation: -

Output of pairwise correlation of all columns in the data frame

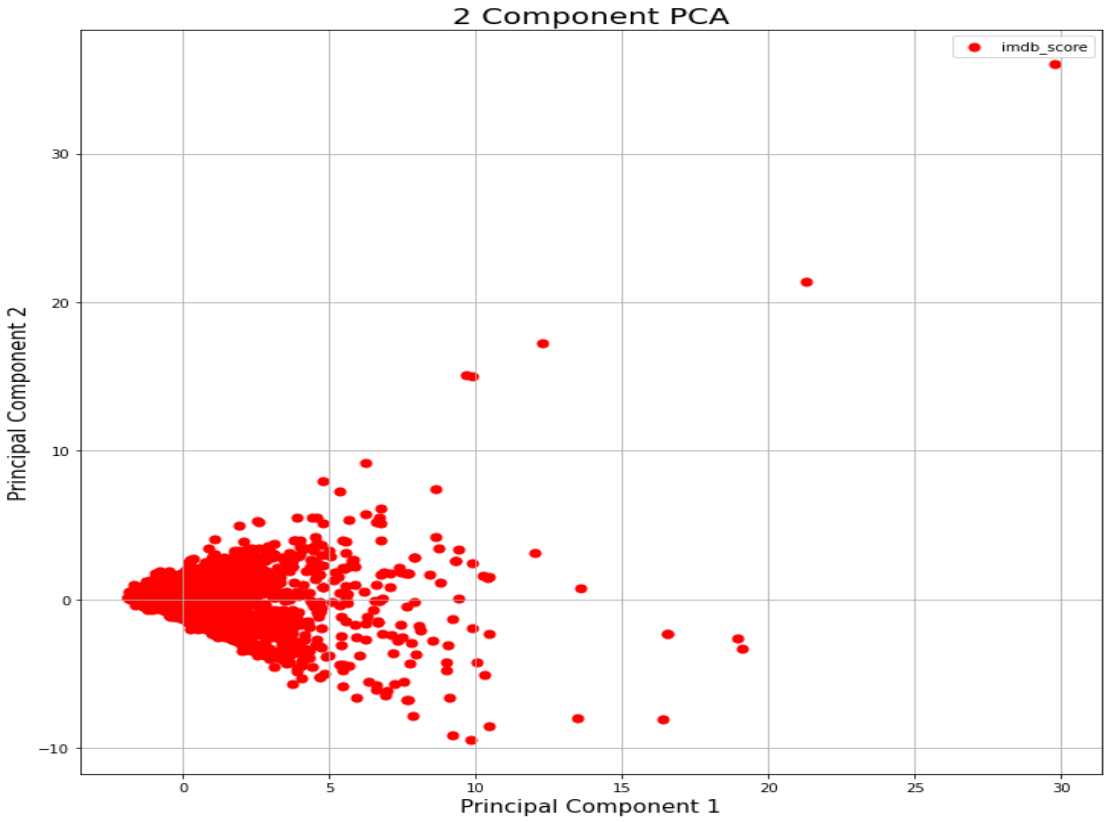


For PCA: -

- I am taking here 2-component PCA and target is “imdb_score”
- Output-

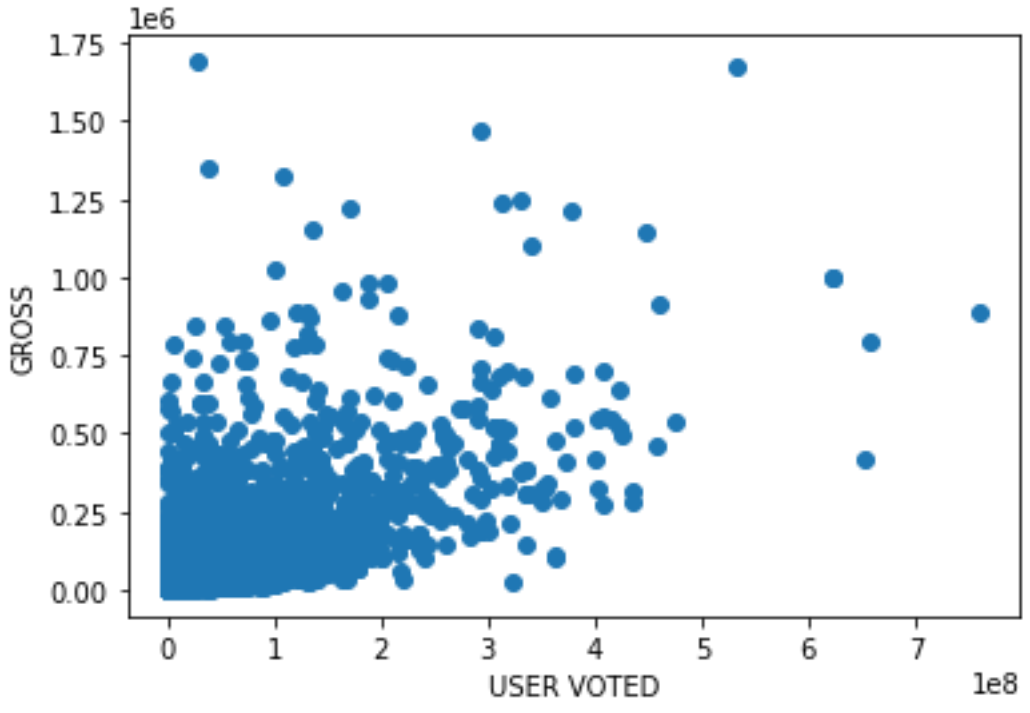
	principal component 1	principal component 2	imdb_score
0	9.206763	-9.148960	7.9
1	5.395706	-1.161434	7.1
2	3.383085	-3.049702	6.8
3	18.955423	-2.648398	8.5
4	-1.906675	0.154353	7.1
5	0.952058	-1.781310	6.6
6	6.636790	-1.543452	6.2
7	1.595273	-2.294432	7.8
8	13.586557	0.760371	7.5
9	7.062331	0.826880	7.5

SI&MT Assignment-2 Documents For output Result



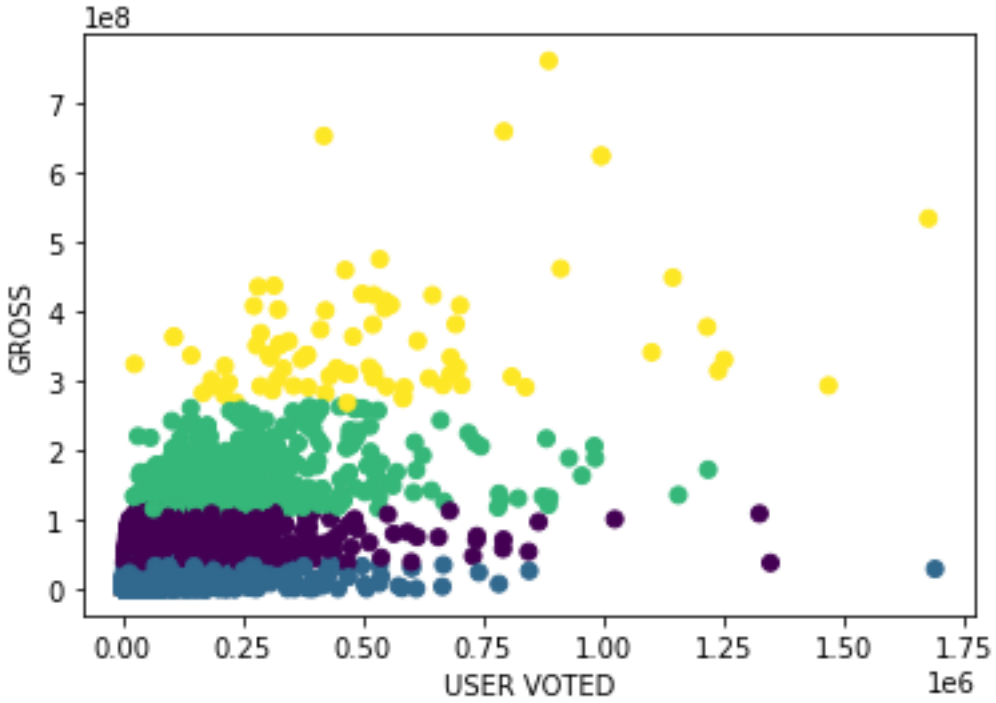
CLUSTERING: -

- I am using K-Means algorithm for clustering and take parameter k=4 .
- Before Clustering the data points-



SI&MT Assignment-2 Documents For output Result

- After Clustering the data points-



Linear Regression: -

- I use KNN algorithm for linear regression and we taking k=5
- Test and training accuracy-
    - Performance after using LinearRegression()

```
Training score:  0.1964408701403445
Testing score:  0.24023135006071816
```

    - Performance after using KNeighborsRegressor()

```
Training score:  -0.05122383424657917
Testing score:  0.3362987934034837
```