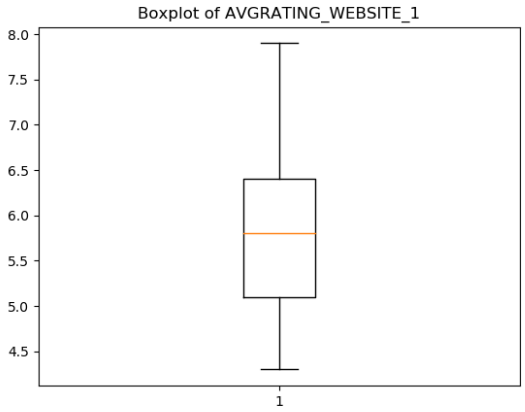
1. Data Description

Maximum & Minimum Score for 4 features:

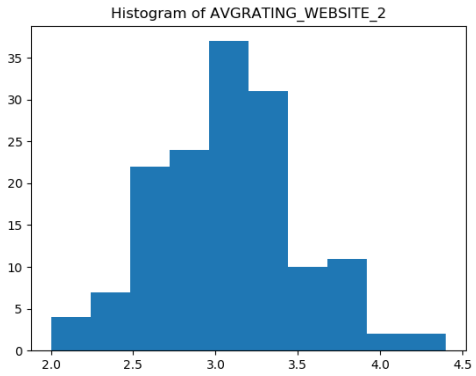
```
2.4836985805578653
3.080455435688643
1.7798692259486224
1.707378808019094
```

2. Data Visualization

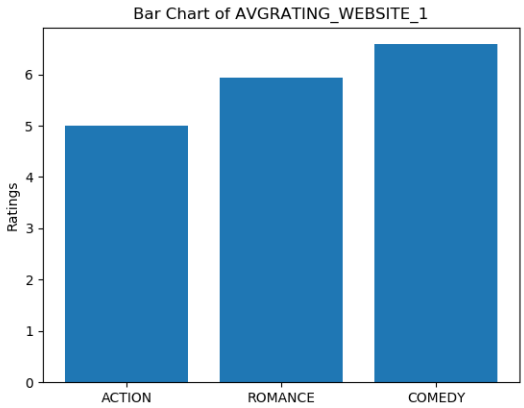
1)



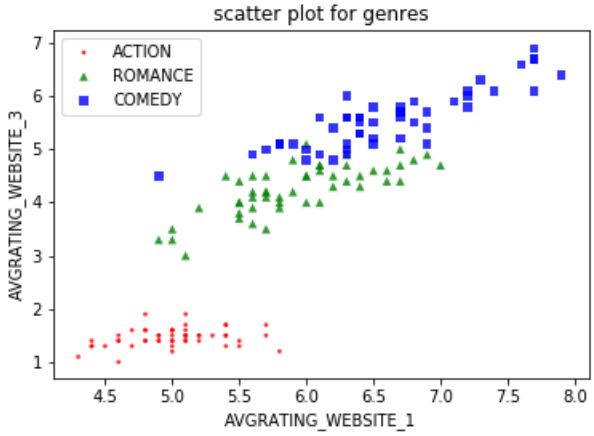
2)



3)



4)



5)

```
KL Divergence score is 0.4658250319477962
```

3. Data Cleaning and Integration

1)

```
The correlation coef. of avgrating_site 1 2  are : -0.11756978413300208
The correlation coef. of avgrating_site 1 3  are : 0.8717537758865831
The correlation coef. of avgrating_site 1 4  are : 0.8179421748583496
The correlation coef. of avgrating_site 2 3  are : -0.4284401043305394
The correlation coef. of avgrating_site 2 4  are : -0.36543079410332724
The correlation coef. of avgrating_site 3 4  are : 0.9627460246236463
```

2)

```
The correlation coef. of Z scores 1 2  are : -0.11756978413300198
The correlation coef. of Z scores 1 3  are : 0.8717537758865832
The correlation coef. of Z scores 1 4  are : 0.8179421748583491
The correlation coef. of Z scores 2 3  are : -0.42844010433054
The correlation coef. of Z scores 2 4  are : -0.36543079410332724
The correlation coef. of Z scores 3 4  are : 0.962746024623647
```

3)
The above results are the same. This is because the correlation coefficient measures the trends of data closest to the linear regression line. The Z scores measures the units of standard deviations away from the regression line. Two sets of data are same units(Zscores) away from the linear regression line so the correlation coefficient are the same.