

Analysis of decline in Views of Hotstar show

Introduction

Hotstar is a streaming platform that release new shows on a regular basis, similar to traditional television. On the basis of data, the platform has noticed a decline in views of a particular show. The objective of this study is to investigate the factors that may be contributing to this decline and provide recommendations for improvement.

Data Description

The dataset used for this study is a CSV file named “mediacompay.csv” that contain information of 80 days of a particular show. The dataset contain the information from March 2017 to May 2017. The data include the following columns:

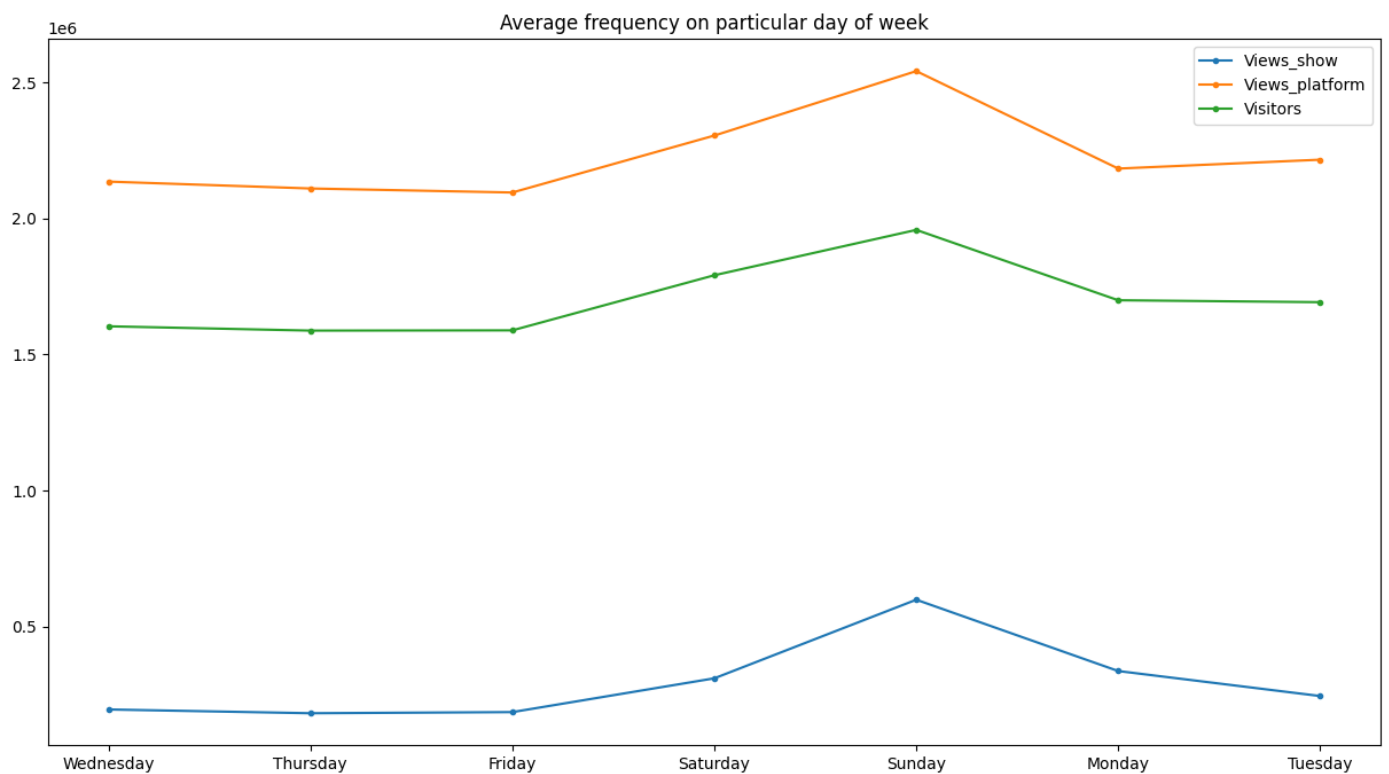
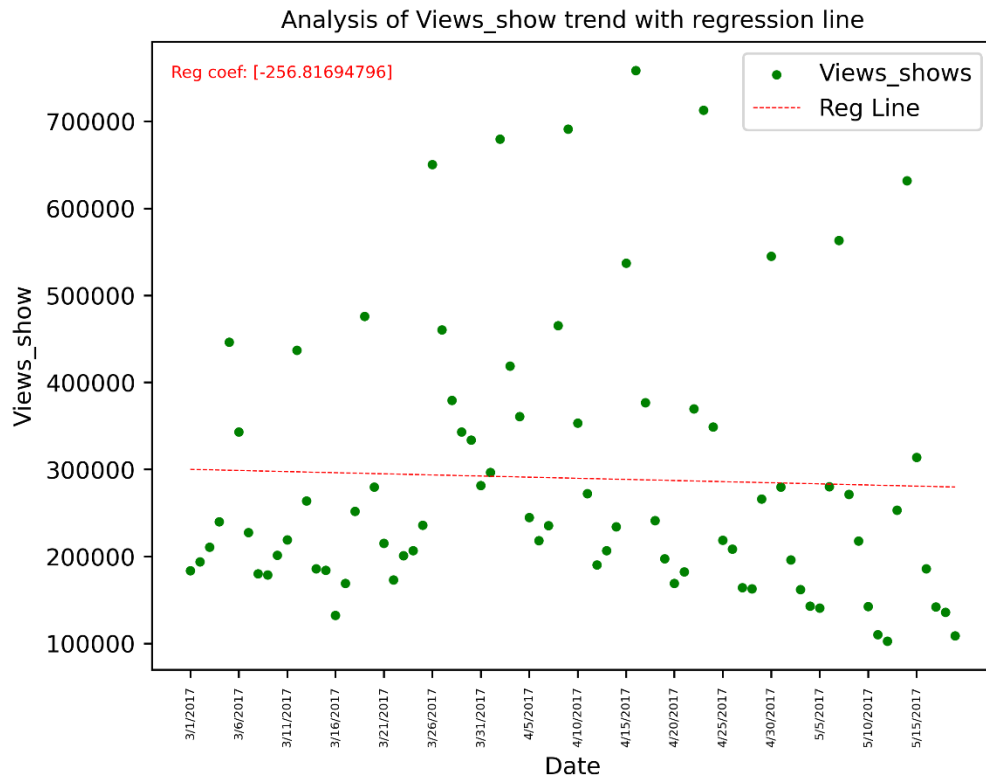
- Date
- Views_show: The number of views for the show on that date
- Visitors: The number of unique visitors for the show on that date
- Ad_impression: The amount spends on advertising for the show on that date
- Cricket_match_india: A binary column indicating whether or not there was a cricket match in India on that date (0 = no match, 1 = match)
- Character_A: A binary column indicating whether or not there was a change in the main character of the show on that date (0 = no change, 1 = change)
- Unnamed: 7: Null column

Methodology

- To investigate the factors contributing to the decline in Views, we used a combination of descriptive statistics, data visualization and regression analysis.
- Data visualization were used to understand the distribution of Views and to identify patterns in the data.
- Data visualization techniques such as line charts and scatter plots were used to better understand the data and to identify patterns.
- Time series analysis was used to examine the trend of views over time.
- The impact of the day of the week on the Views was also analysed.
- The impact of the Cricket match on the Views was also analysed.

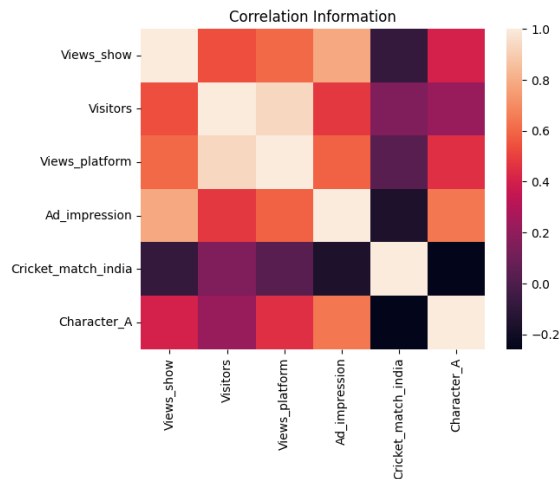
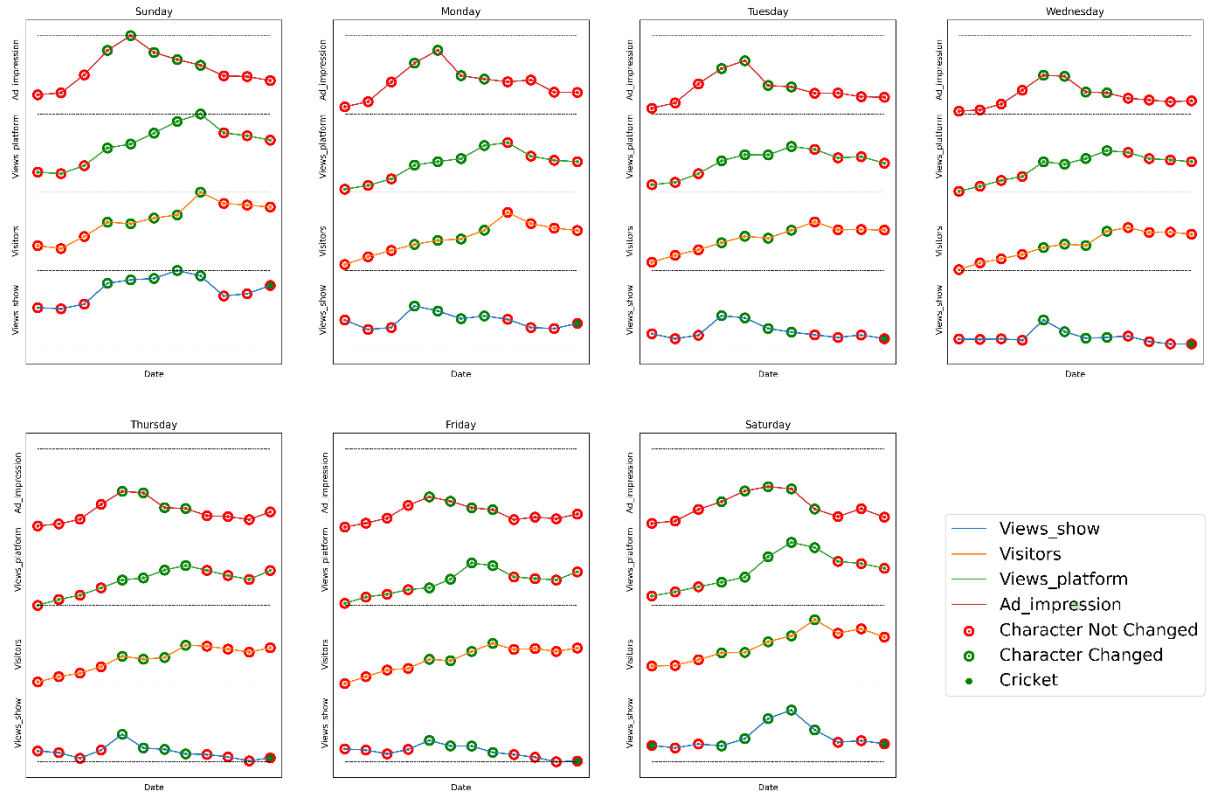
Results

- Views of Hotstar shows tend to decline over time and shows on weekends tend to have higher Views than those on weekdays.



- A change in the main character of the show has a significant positive impact on Views and Cricket match has negligible impact on views.

Complete Data Visualization in Single plot on every day of week



	Views_show	Visitors	Views_platform	Ad_impression	Cricket_match_india	Character_A
Views_show	1.000000	0.535290	0.604279	0.785673	-0.083154	0.402533
Visitors	0.535290	1.000000	0.935832	0.478598	0.147210	0.223093
Views_platform	0.604279	0.935832	1.000000	0.587003	0.036575	0.452375
Ad_impression	0.785673	0.478598	0.587003	1.000000	-0.163936	0.644354
Cricket_match_india	-0.083154	0.147210	0.036575	-0.163936	1.000000	-0.258199
Character_A	0.402533	0.223093	0.452375	0.644354	-0.258199	1.000000

- Regression model was also developed with the following hypothesis which can predict the views.
 - Independent Variable
 - Visitors
 - Views_platform
 - Ad_impression
 - Cricket_match_india
 - Character_A
 - Day
 - Dependent Variable
 - Views_show
 - Days was encoded into nominal order to apply the regression analysis

Output of Regression Model

On Train Data

```

=====
                        OLS Regression Results
=====
Dep. Variable:          Views_show      R-squared:                0.930
Model:                  OLS             Adj. R-squared:           0.912
Method:                 Least Squares   F-statistic:             53.10
Date:                  Wed, 16 Nov 2022 Prob (F-statistic):       1.03e-21
Time:                  22:37:28         Log-Likelihood:          81.800
No. Observations:      56              AIC:                    -139.6
Df Residuals:          44              BIC:                    -115.3
Df Model:              11
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-1.3958	0.374	-3.732	0.001	-2.150	-0.642
x1	-0.2016	0.197	-1.025	0.311	-0.598	0.195
x2	0.2410	0.214	1.126	0.266	-0.190	0.672
x3	0.3877	0.085	4.562	0.000	0.216	0.559
x4	-0.0127	0.028	-0.454	0.652	-0.069	0.044
x5	0.0352	0.034	1.025	0.311	-0.034	0.104
x6	0.1247	0.036	3.458	0.001	0.052	0.197
x7	0.1297	0.034	3.867	0.000	0.062	0.197
x8	0.4293	0.048	9.008	0.000	0.333	0.525
x9	-0.0424	0.031	-1.362	0.180	-0.105	0.020
x10	0.0085	0.038	0.224	0.824	-0.068	0.085
x11	-0.0288	0.030	-0.958	0.344	-0.090	0.032

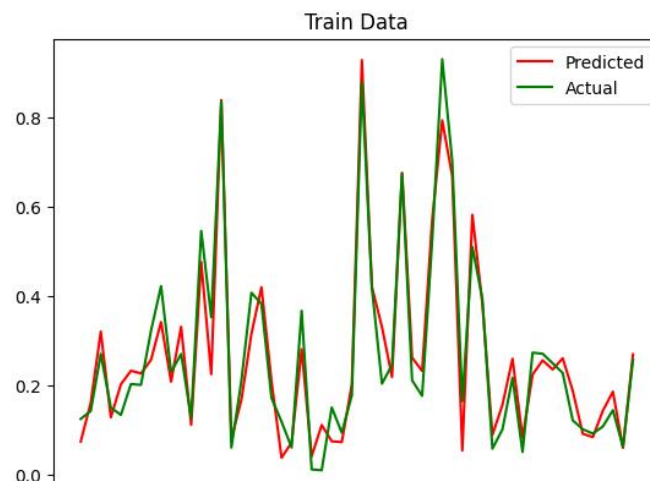
```

=====
Omnibus:                2.084    Durbin-Watson:           1.912
Prob(Omnibus):          0.353    Jarque-Bera (JB):         1.876
Skew:                   0.440    Prob(JB):                 0.391
Kurtosis:               2.833    Cond. No.                 232.
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.



On Test Data

OLS Regression Results

```

=====
Dep. Variable:          Views_show    R-squared:                0.988
Model:                  OLS           Adj. R-squared:           0.976
Method:                 Least Squares  F-statistic:             87.27
Date:                  Wed, 16 Nov 2022  Prob (F-statistic):      1.07e-09
Time:                  22:37:29       Log-Likelihood:          49.664
No. Observations:      24            AIC:                    -75.33
Df Residuals:          12            BIC:                    -61.19
Df Model:              11
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-2.2386	0.448	-4.992	0.000	-3.216	-1.262
x1	-1.4804	0.206	-7.198	0.000	-1.928	-1.032
x2	1.5473	0.195	7.924	0.000	1.122	1.973
x3	0.2338	0.122	1.913	0.080	-0.032	0.500
x4	0.2529	0.061	4.119	0.001	0.119	0.387
x5	-0.1391	0.054	-2.575	0.024	-0.257	-0.021
x6	0.3196	0.044	7.281	0.000	0.224	0.415
x7	0.2924	0.047	6.184	0.000	0.189	0.395
x8	0.5476	0.052	10.450	0.000	0.433	0.662
x9	0.0090	0.042	0.216	0.832	-0.081	0.099
x10	0.1104	0.037	2.957	0.012	0.029	0.192
x11	0.0425	0.040	1.055	0.312	-0.045	0.130

```

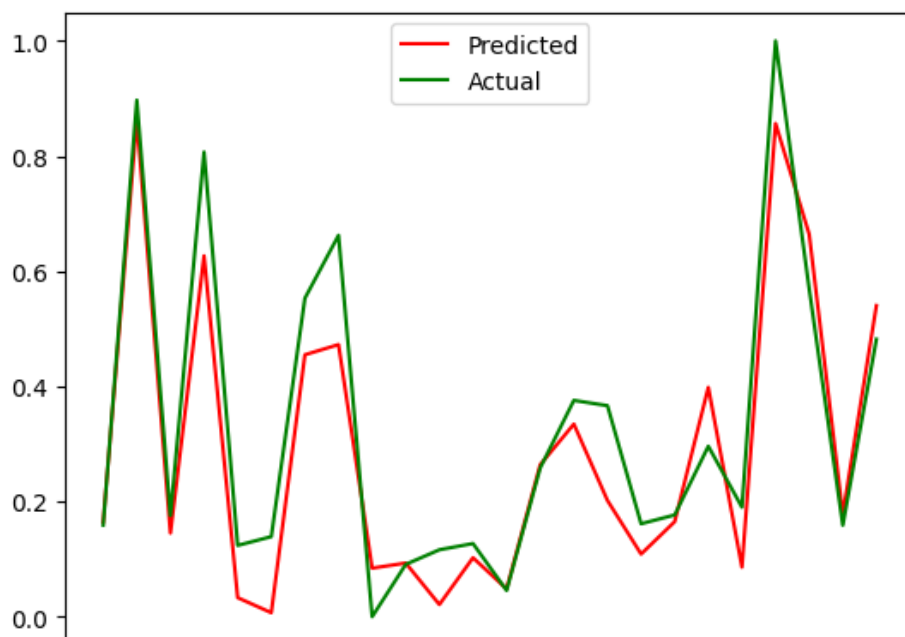
=====
Omnibus:                2.684    Durbin-Watson:              1.648
Prob(Omnibus):          0.261    Jarque-Bera (JB):          1.211
Skew:                   0.228    Prob(JB):                  0.546
Kurtosis:               4.001    Cond. No.:                 257.
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Test Data



On Complete Data

OLS Regression Results

```

=====
Dep. Variable:          Views_show      R-squared:                0.927
Model:                  OLS             Adj. R-squared:           0.915
Method:                 Least Squares   F-statistic:             78.04
Date:                  Wed, 16 Nov 2022 Prob (F-statistic):       4.09e-34
Time:                  22:37:30         Log-Likelihood:          106.94
No. Observations:      80              AIC:                    -189.9
Df Residuals:          68              BIC:                    -161.3
Df Model:              11
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-1.5755	0.281	-5.599	0.000	-2.137	-1.014
x1	-0.5916	0.148	-4.004	0.000	-0.886	-0.297
x2	0.6595	0.156	4.221	0.000	0.348	0.971
x3	0.3030	0.075	4.063	0.000	0.154	0.452
x4	0.0046	0.028	0.165	0.869	-0.051	0.060
x5	0.0081	0.030	0.266	0.791	-0.053	0.069
x6	0.1781	0.033	5.346	0.000	0.112	0.245
x7	0.1711	0.030	5.650	0.000	0.111	0.231
x8	0.4928	0.040	12.183	0.000	0.412	0.574
x9	-0.0207	0.028	-0.734	0.465	-0.077	0.036
x10	0.0439	0.031	1.439	0.155	-0.017	0.105
x11	-0.0122	0.029	-0.426	0.671	-0.069	0.045

```

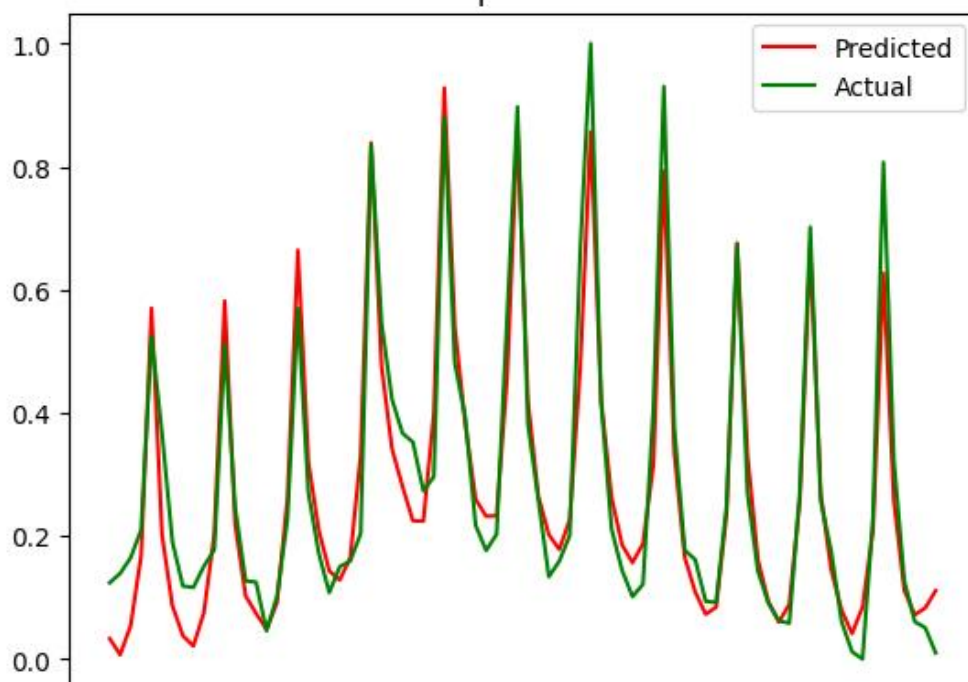
=====
Omnibus:                2.785      Durbin-Watson:            1.090
Prob(Omnibus):          0.248      Jarque-Bera (JB):         2.760
Skew:                   0.428      Prob(JB):                 0.252
Kurtosis:               2.694      Cond. No.:                187.
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

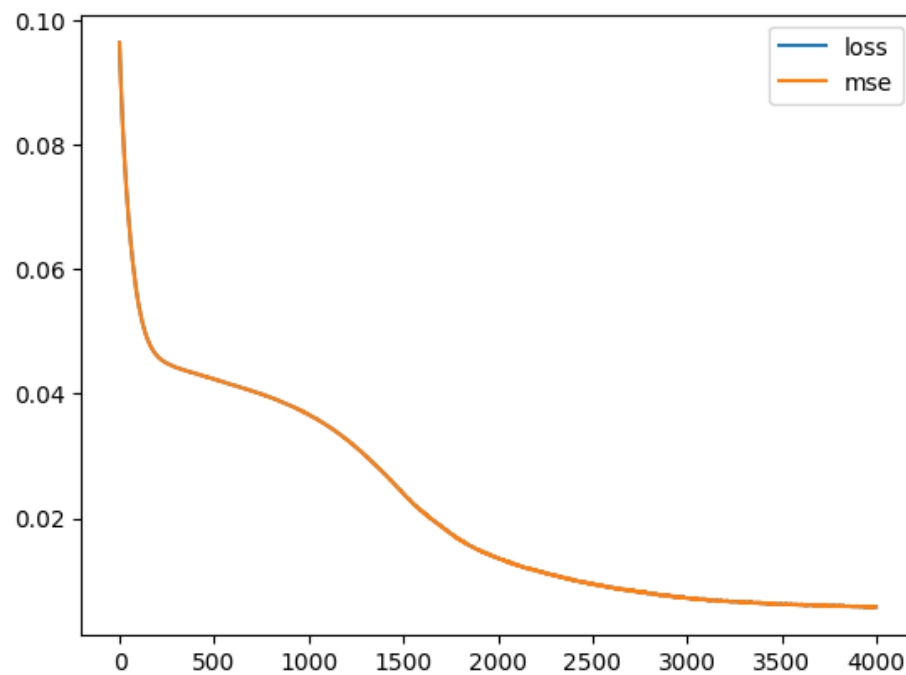
Complete Data



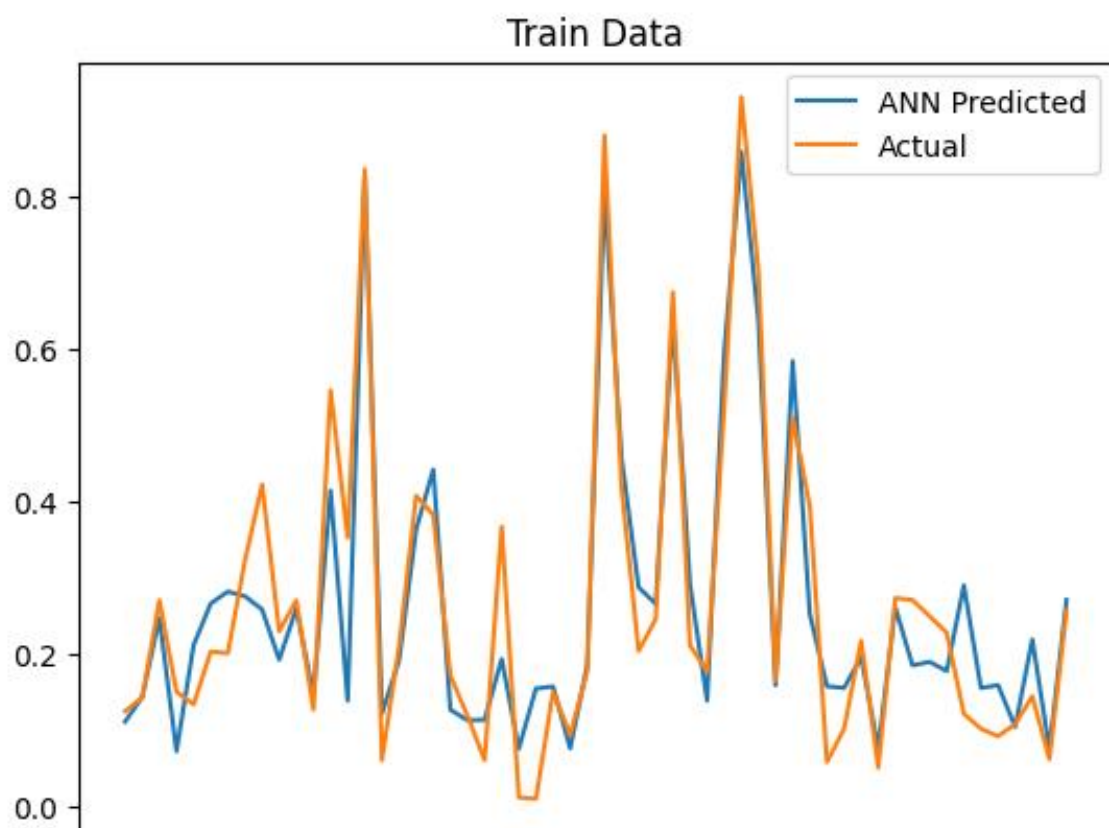
- Model was also developed using ANN and compared with regression model

Output of ANN model

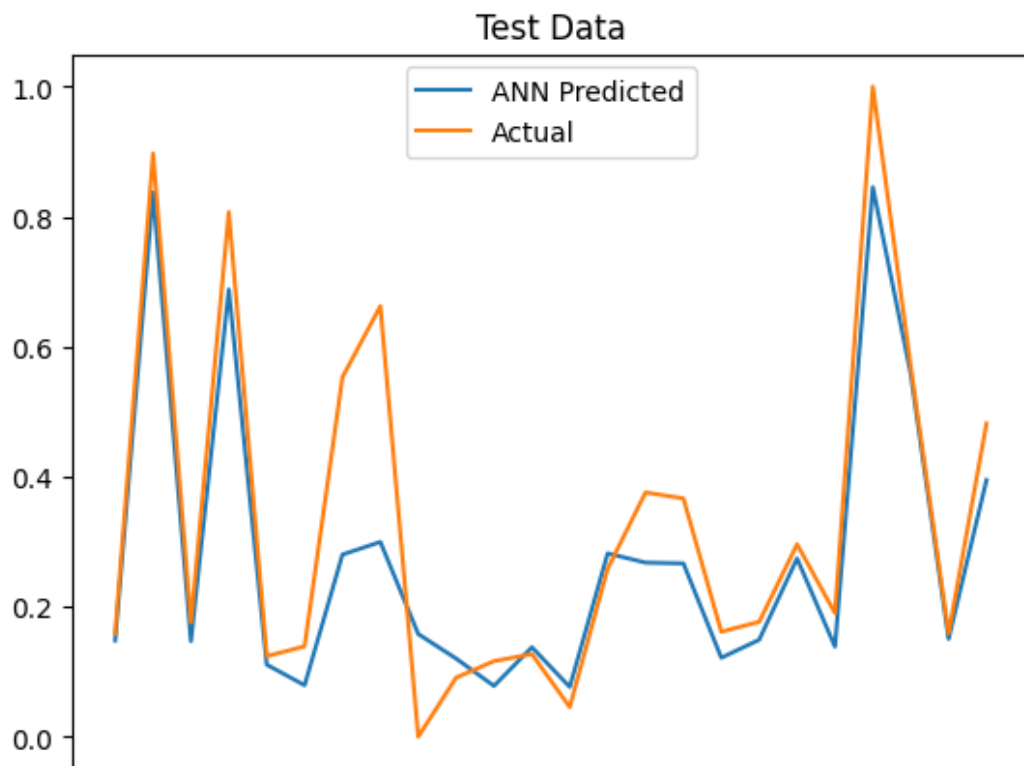
Loss



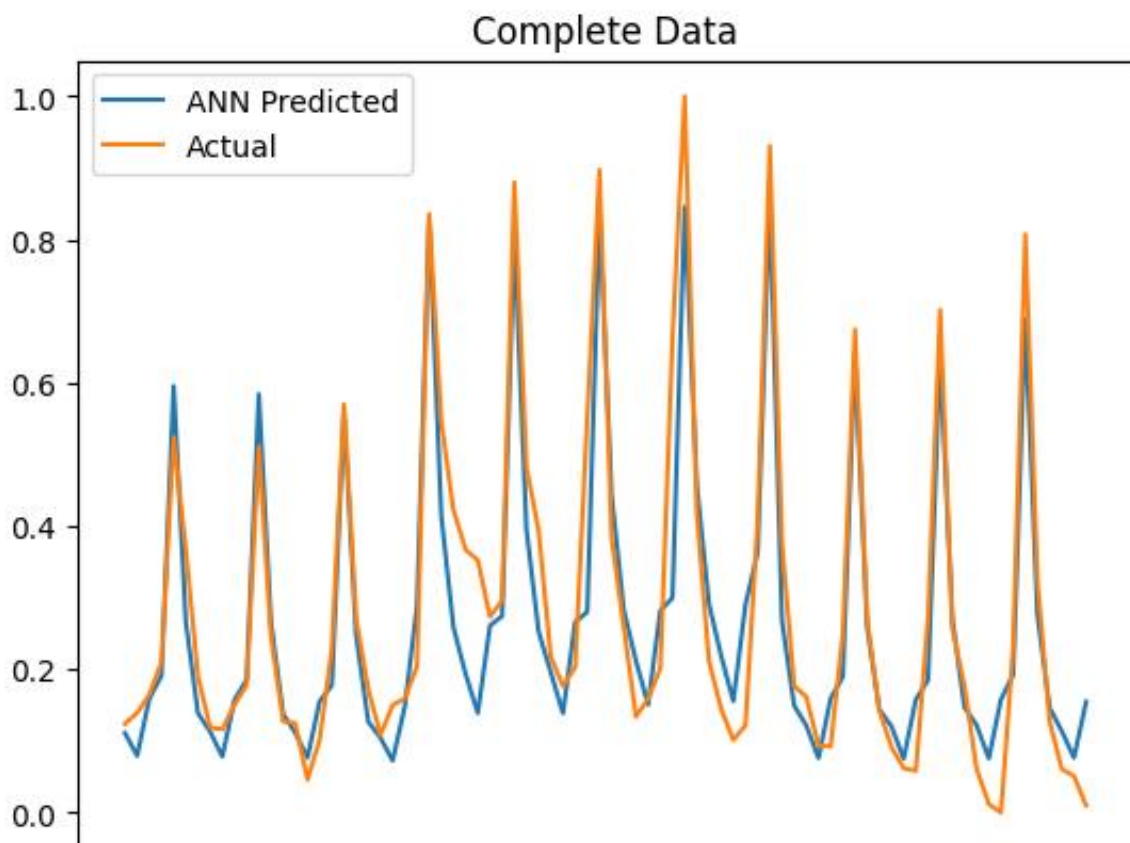
On Train Data



On Test Data

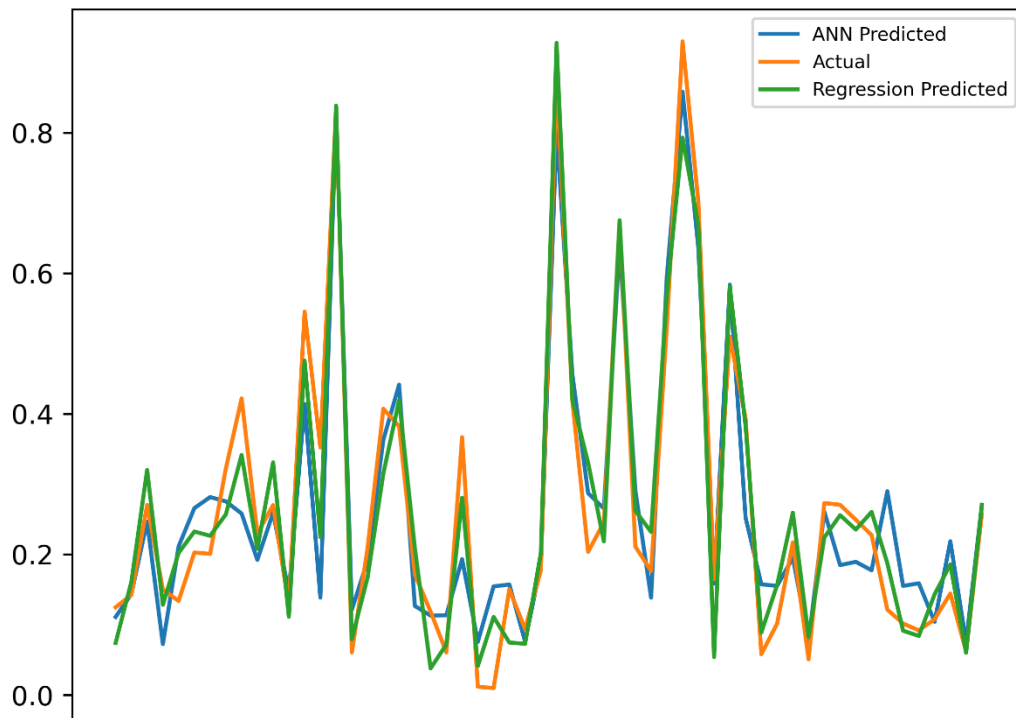


On Complete Data

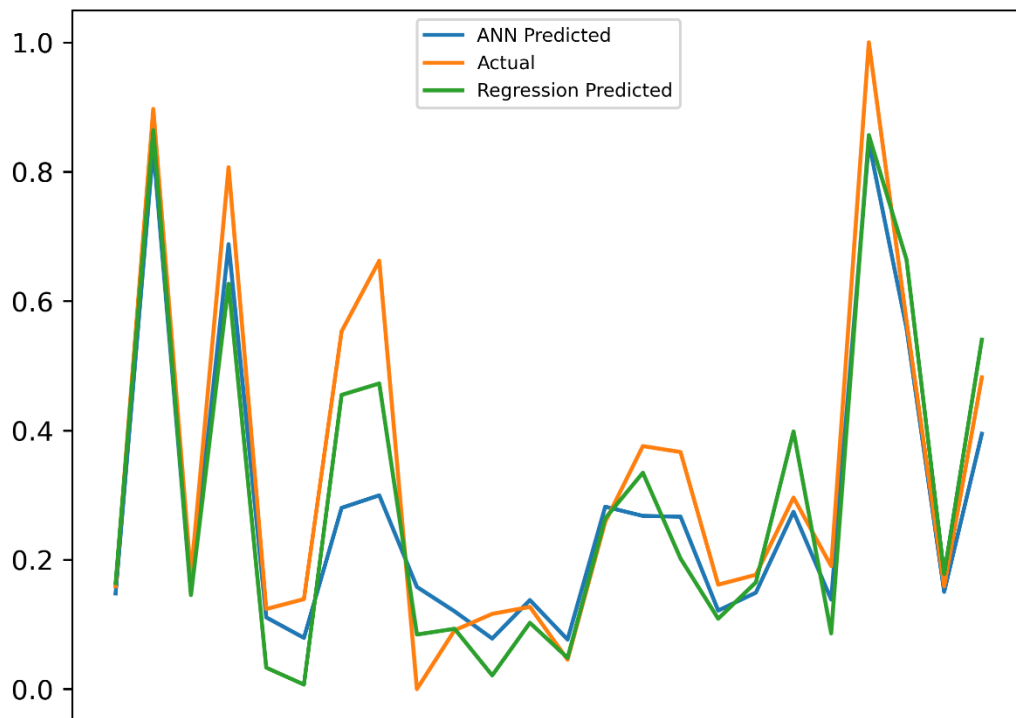


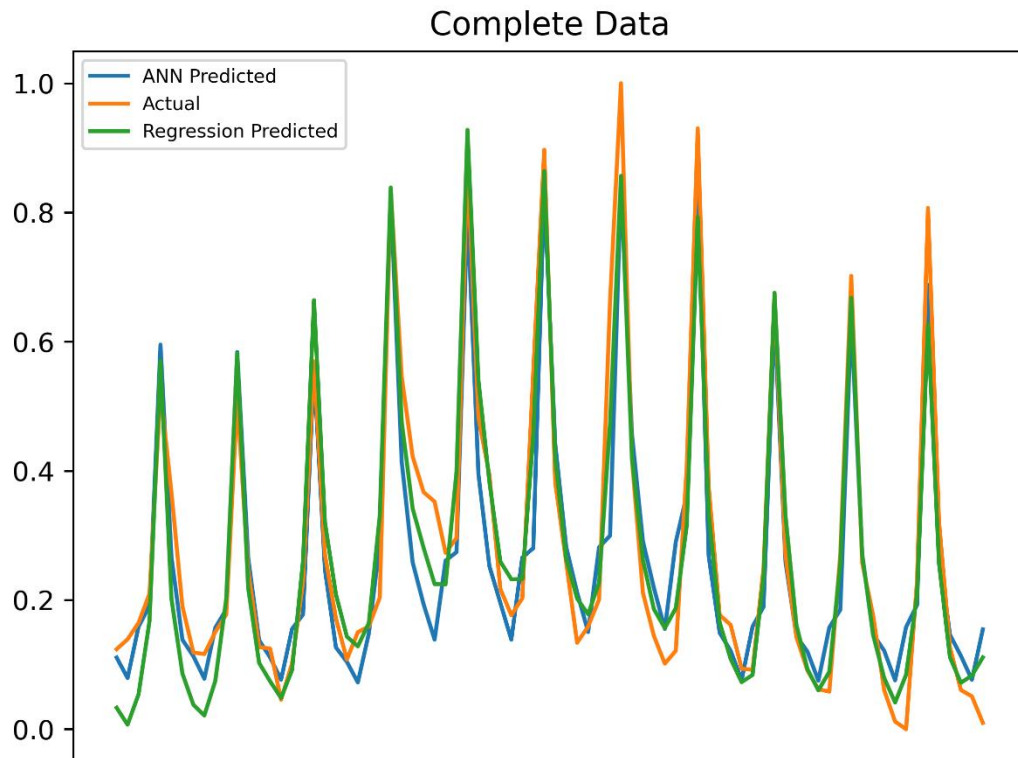
ANN Model vs Regression Model

Train Data



Test Data





Conclusion

The results of this study suggest that Hotstar should focus on promoting their shows more effectively by increasing their marketing efforts and by targeting the audience more effectively. Additionally, Hotstar should consider to change the main character of the show.

Team Details

GROUP: I

Name	Branch	Semester/Year
PINTU RAJ	Information Technology	5 th Sem /3 rd year
Ayush Kumar	Information Technology	5 th Sem /3 rd year
Sudhir Kumar	Information Technology	5 th Sem /3 rd year