Predictive Modelling
Business report

Context:

An over-the-top (OTT) media service is a media service offered directly to viewers via the internet. The term is most synonymous with subscription-based video-on-demand services that offer access to film and television content, including existing series acquired from other producers, as well as original content produced specifically for the service. They are typically accessed via websites on personal computers, apps on smartphones and tablets, or televisions with integrated Smart TV platforms.

Presently, OTT services are at a relatively nascent stage and are widely accepted as a trending technology across the globe. With the increasing change in customers' social behavior, which is shifting from traditional subscriptions to broadcasting services and OTT on-demand video and music subscriptions every year, OTT streaming is expected to grow at a very fast pace. The global OTT market size was valued at $121.61 billion in 2019 and is projected to reach $1,039.03 billion by 2027, growing at a CAGR of 29.4% from 2020 to 2027. The shift from television to OTT services for entertainment is driven by benefits such as on-demand services, ease of access, and access to better networks and digital connectivity.

With the outbreak of COVID19, OTT services are striving to meet the growing entertainment appetite of viewers, with some platforms already experiencing a 46% increase in consumption and subscriber count as viewers seek fresh content. With innovations and advanced transformations, which will enable the customers to access everything they want in a single space, OTT platforms across the world are expected to increasingly attract subscribers on a concurrent basis.

Objective:

ShowTime is an OTT service provider and offers a wide variety of content (movies, web shows, etc.) for its users. They want to determine the driver variables for first-day content viewership so that they can take necessary measures to improve the viewership of the content on their platform. Some of the reasons for the decline in viewership of content would be the decline in the number of people coming to the platform, decreased marketing spend, content timing clashes, weekends and holidays, etc. They have hired you as a Data Scientist, shared the data of the current content in their platform, and asked you to analyze the data and come up with a linear regression model to determine the driving factors for first-day viewership.

Data Dictionary:

- visitors: Average number of visitors, in millions, to the platform in the past week
- ad_impressions: Number of ad impressions, in millions, across all ad campaigns for the content (running and completed)
- major_sports_event: Any major sports event on the day
- genre: Genre of the content
- dayofweek: Day of the release of the content
- season: Season of the release of the content
- views_trailer: Number of views, in millions, of the content trailer
- views_content: Number of first-day views, in millions, of the content

Important questions:

1. What does the distribution of content views look like?
2. What does the distribution of genres look like?
3. The day of the week on which content is released generally plays a key role in the viewership. How does the viewership vary with the day of release?
4. How does the viewership vary with the season of release?
5. What is the correlation between trailer views and content views?

1. Exploratory Data Analysis

- Problem definition, questions to be answered - Data background and contents - Univariate analysis - Bivariate analysis - Answers to the key questions provided - Insights based on EDA

2. Data preprocessing

- Duplicate value check - Missing value treatment - Outlier treatment - Feature engineering - Data preparation for modeling

3. Model building - Linear Regression

- Build the model and comment on the model statistics - Display model coefficients with column names

4. Testing the assumptions of linear regression model

- Perform tests for the assumptions of the linear regression - Comment on the findings from the tests

5. Model performance evaluation

Evaluate the model on different performance metrics

6. Actionable Insights & Recommendations

- Comments on significance of predictors - Key takeaways for the business

1. Exploratory Data Analysis:

– The Name of the dataset is “ottdata.csv”.

– The Dataset contain a total of 1000 rows and 8 columns.

– There are no Duplicates present in the Data.

– There are no missing values present in the dataset.

– There no irregularities present in the dataset.

– There are 4 Categorical Columns: genre, dayofweek, season.

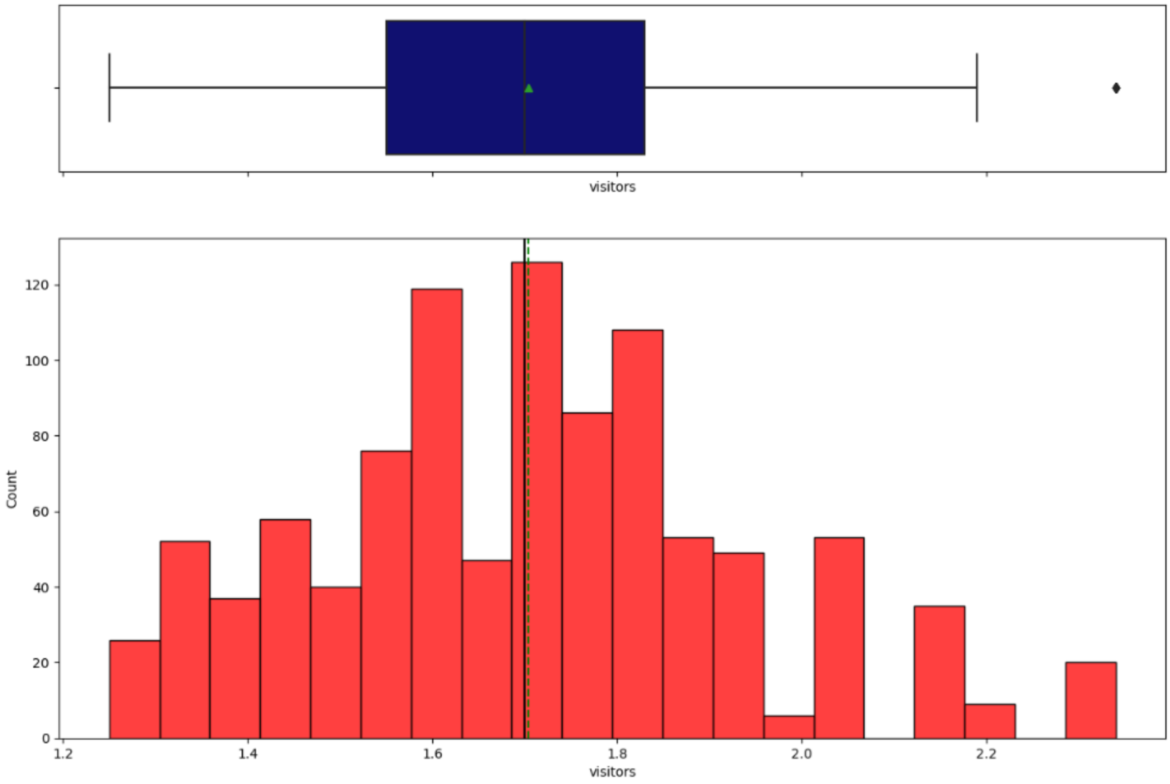
– And there are 5 Numerical column: visitors, ad_impressions, major_sports_event, views_trailer, views_content

– The Datatypes present are :

- Int64.
- Object.
- Float64.

Univariate Analysis:

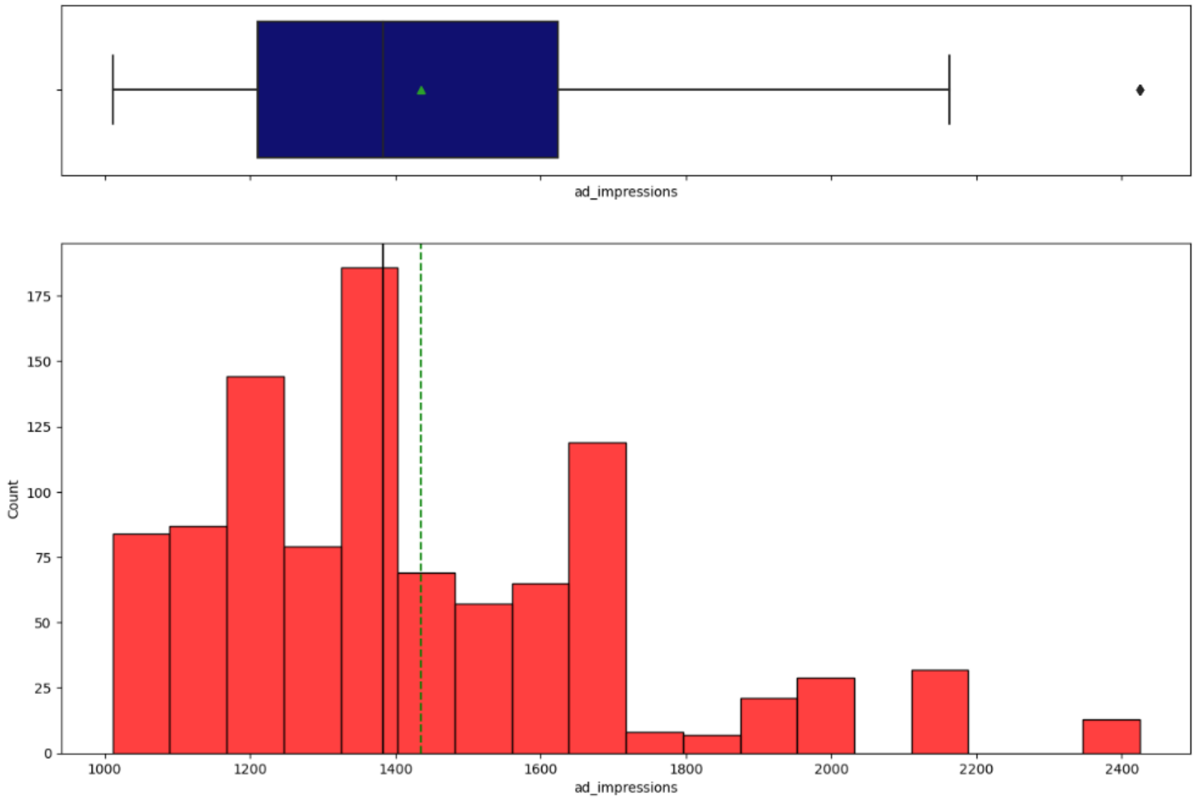
1. Visitors:



- The mean value of visitors that is somewhere around 1.7 million visitors.

- Most of the data lies between 1.5 to 1.9 million visitors.

2. **Ad_impressions:**

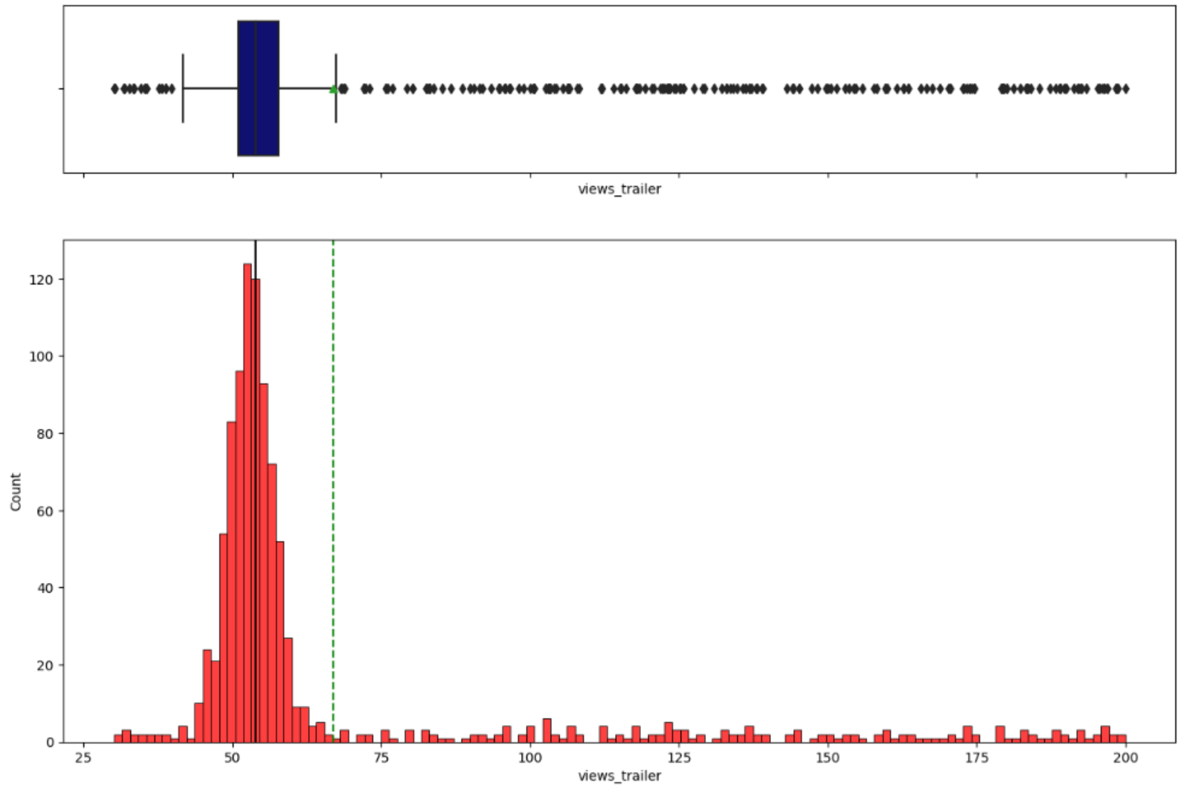


- The mean value of Ad impressions is somewhere around 1400 to 1500 million `ad_impressions`.

- Most of the data lies betwee 1200 to 1700 million `ad_impressions`.

- There is a slight right skew present in the plot.

3. Views_trailer:

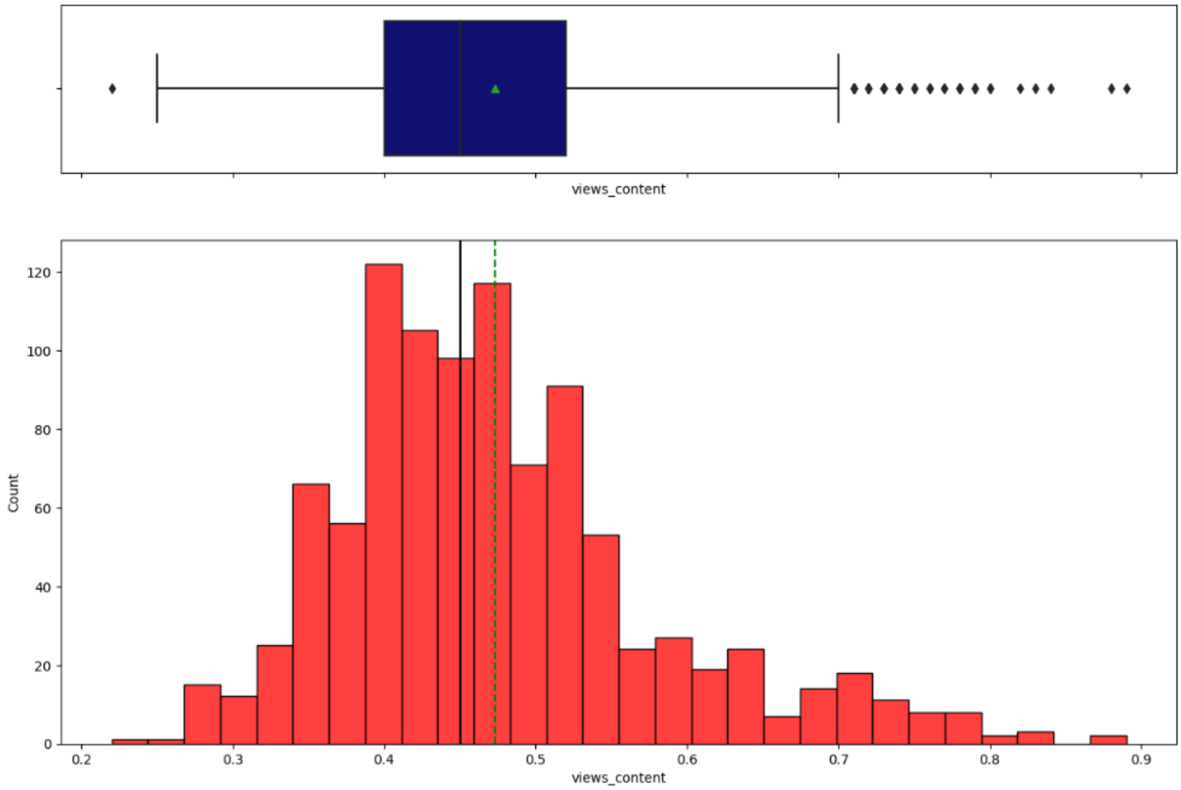


- The mean value of is found somewhere betwee 60 to 75 million views.

- Most of the values lies around 50 million views.

- There are lot of outliers found in the dataset but its okay to havr 200 million views on trailer.

4. Views_content:



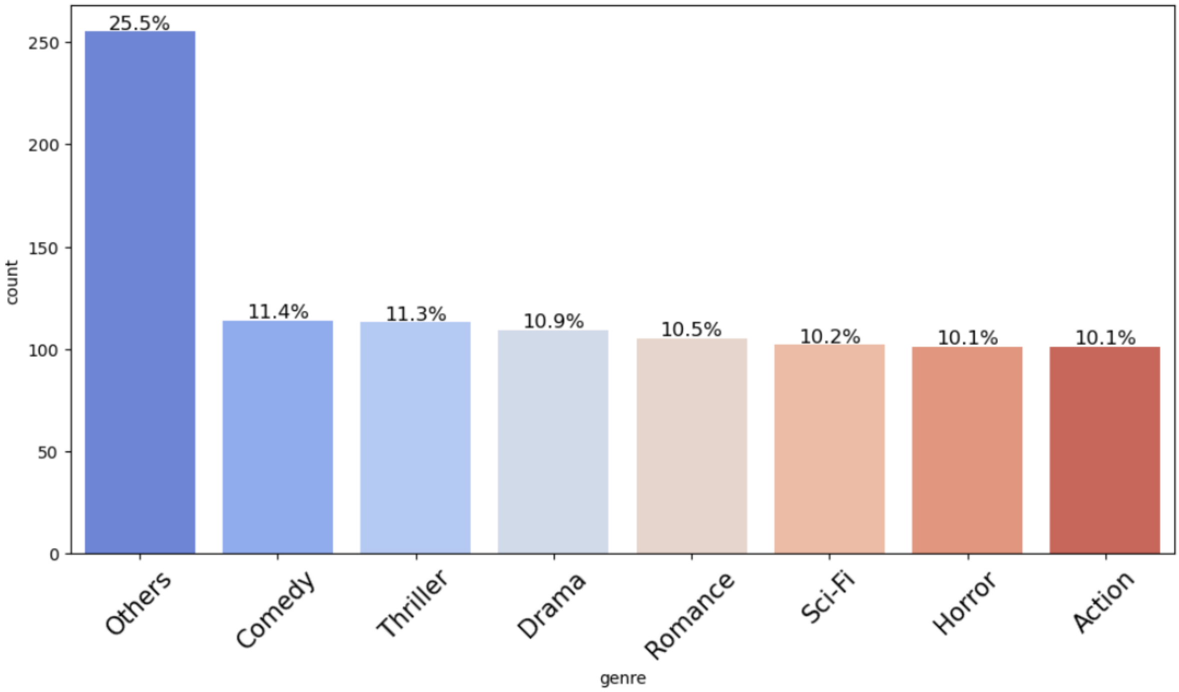
- The mean value lies between 0.4 to 0.5 million view of the content.

- Most of the values lies between 0.4 to 0.6 million views.

- There are some of outliers present in the dataset might need to take a look on it.

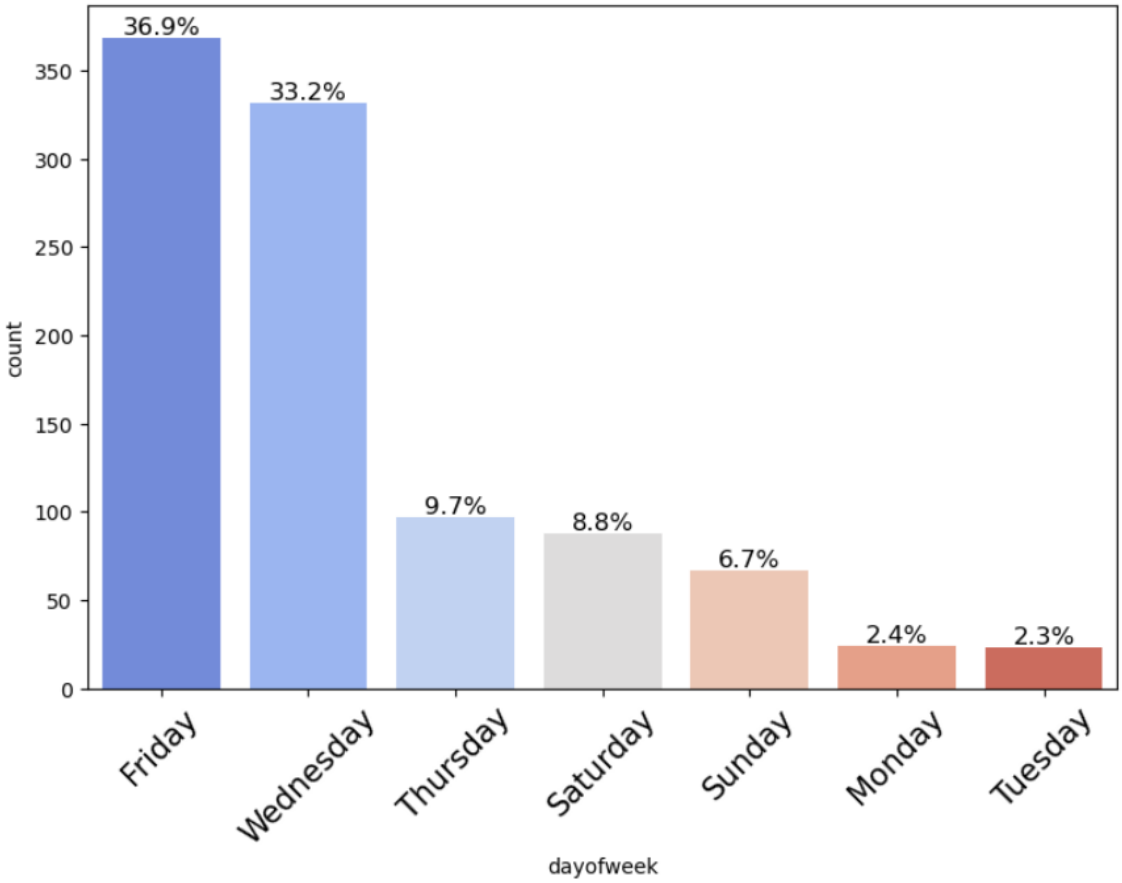
Categorical columns:

1. Genre:



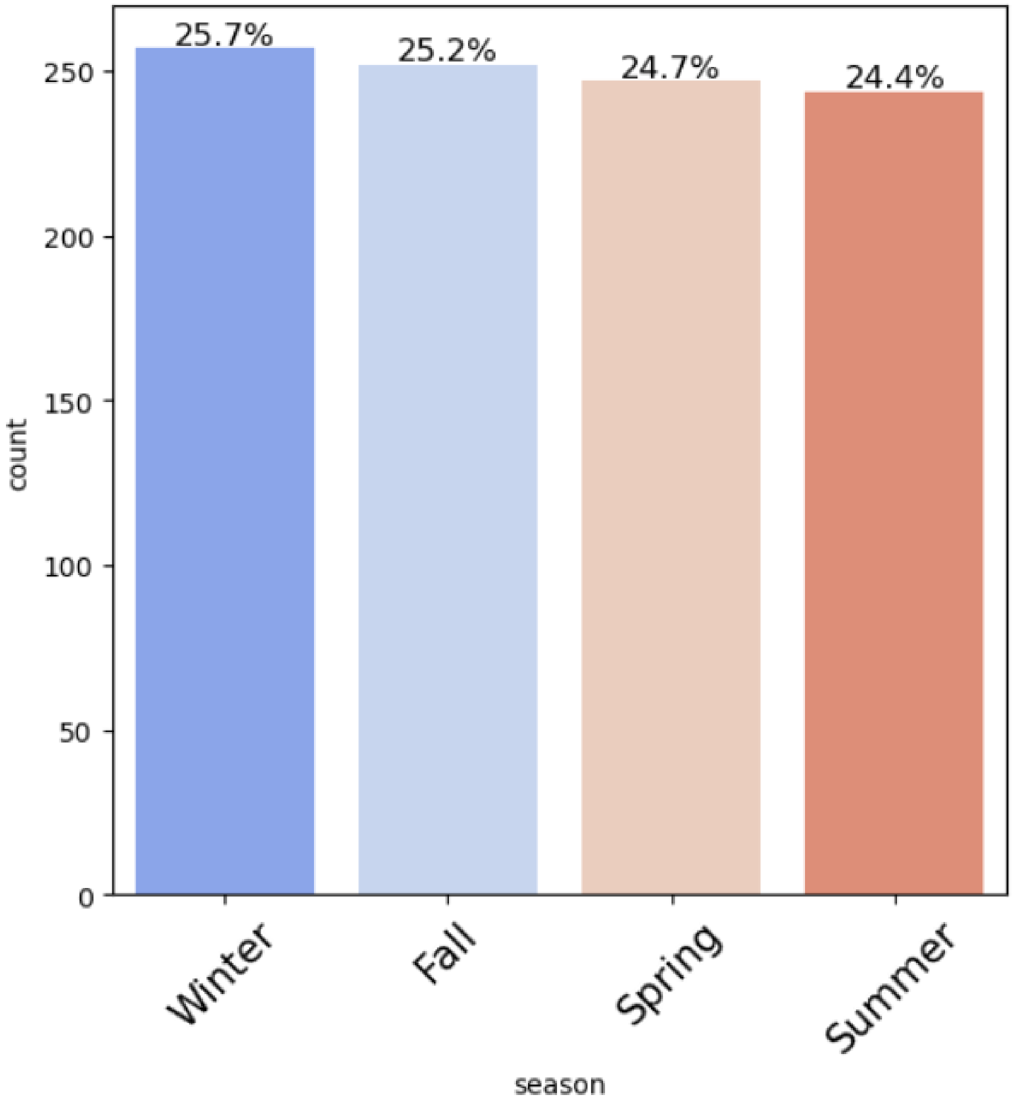
- The other genre are higher in number compaed to other columns
- All other columns are almost equal in number.
- Almost 25% of the data is in other genre category.

2. Dayofweek:



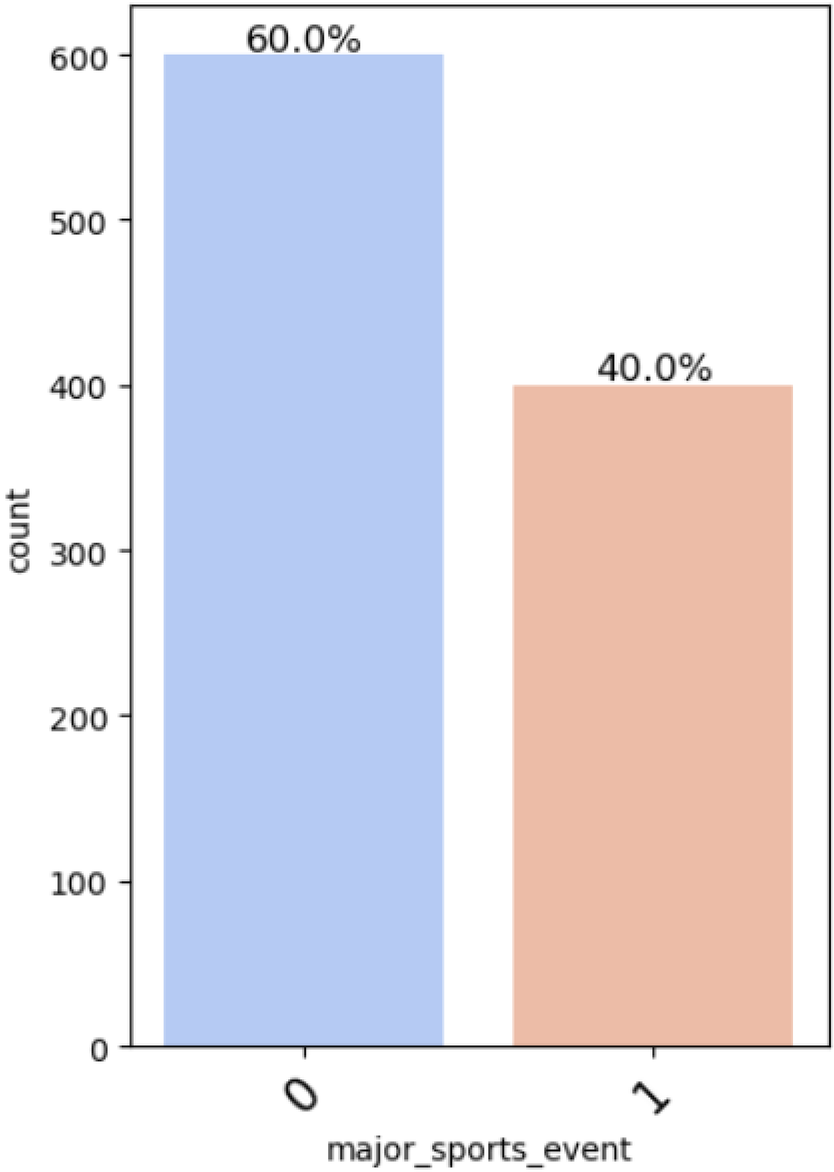
- Movie released on Friday are more in number compared to other day releases.
- Almost 36.9% of themoviews are released in Friday.
- Monday and Tuesday have very low percentage of movie releases of 2.4% and 2.3%.
- Other Three days also have very loew amount of movies released.

3. Season:



- There are uniform almost uniform distribution seen in the season column.
- Where Winter have 25.7%, Fall have 25.2%, Spring have 24.7% and Summer have 24.4%.

4. Major_sports_event:

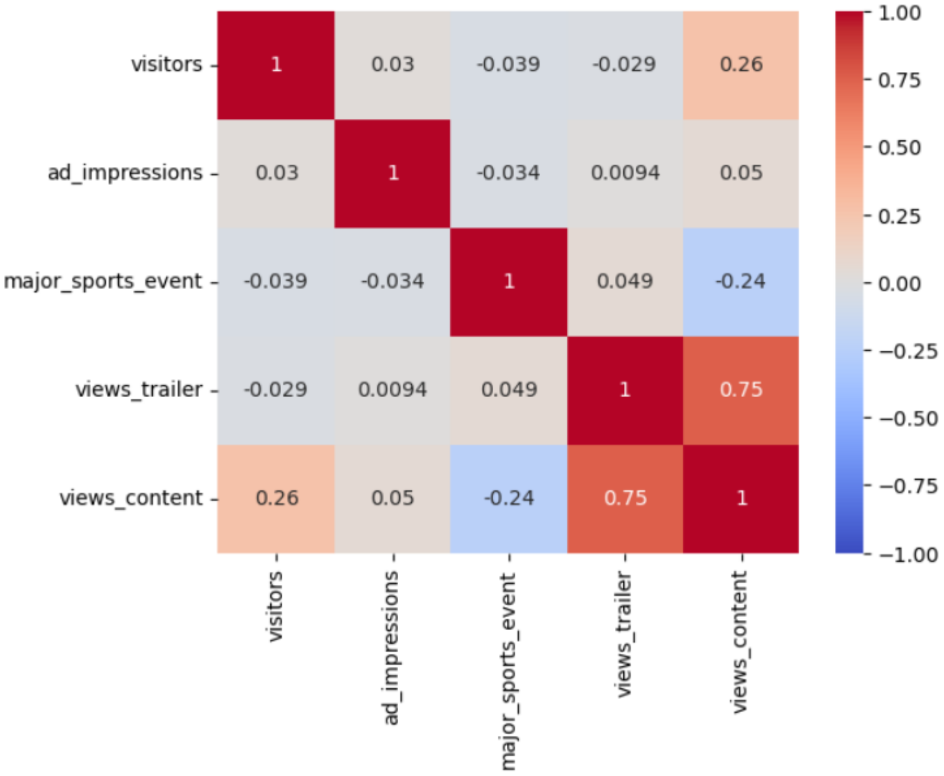


- In the aboce Graph 0 represent the there are no Sports event at the time of release and 1 represent there was a major sports event at the tym of release.

- We can clearly see that 60% of the movies released when there where no major sports event.

- Only 40% of movies are released when there where major sports event.

Bivariate Analysis:



- we can clearly see that the views_trailer` and views_content are highly correlated with eachother.

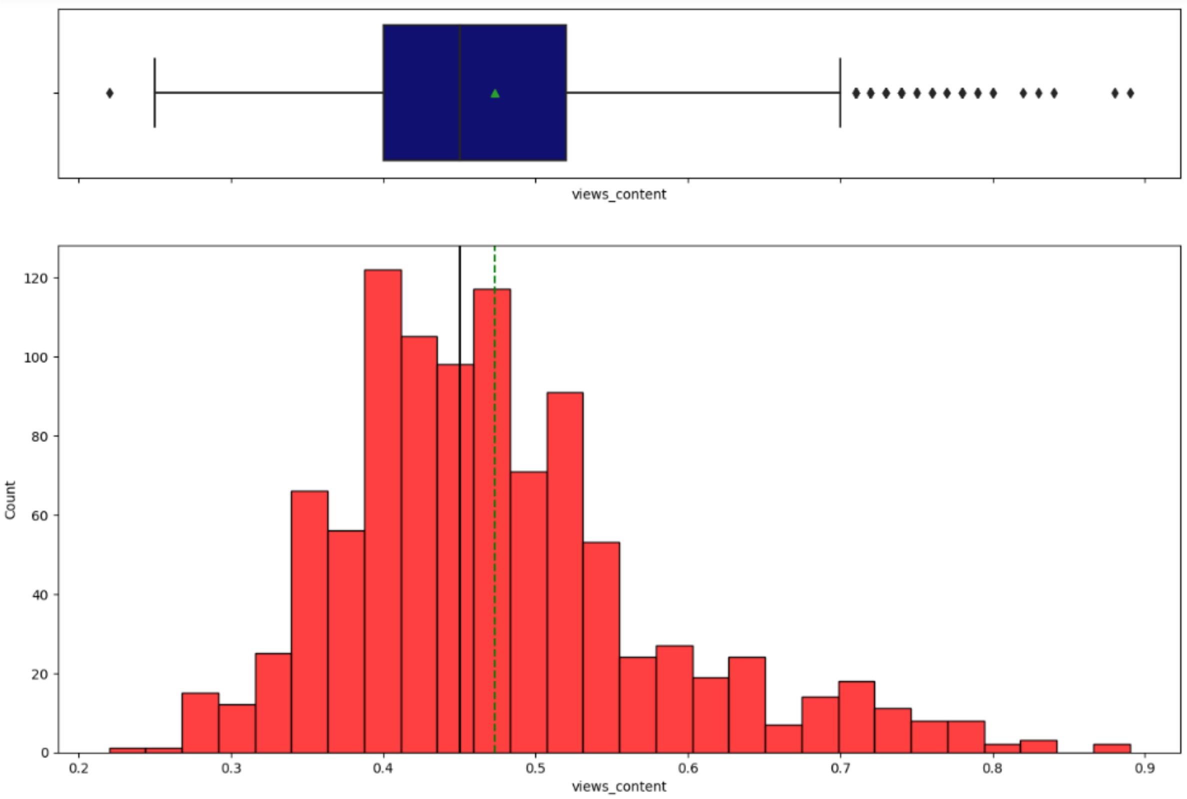
- Also the content views and visitors have some correlation.

- All other values don't have much of a correlation.

- views_content and major_sports_event have some negative correlation between eachother.

Important Questions:

1. What does the distribution of content views look like?



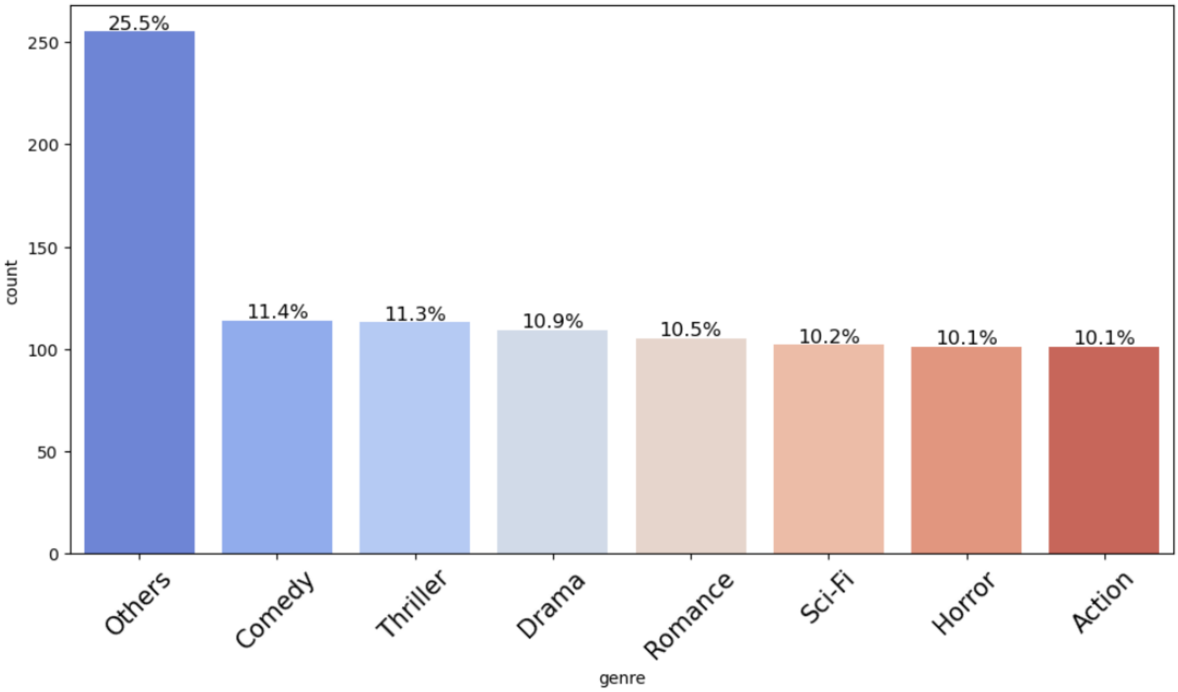
- The mean value lies between 0.4 to 0.5 million view of the content.

- Most of the values lies between 0.4 to 0.6 million views.

- There are some of outliers present in the dataset might need to take a look on it.

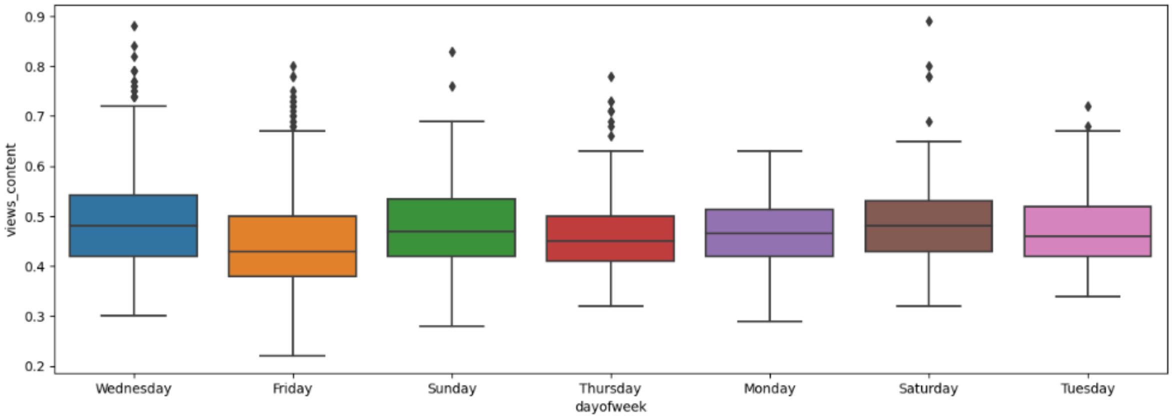
- Seems the distribution is normal distribution.

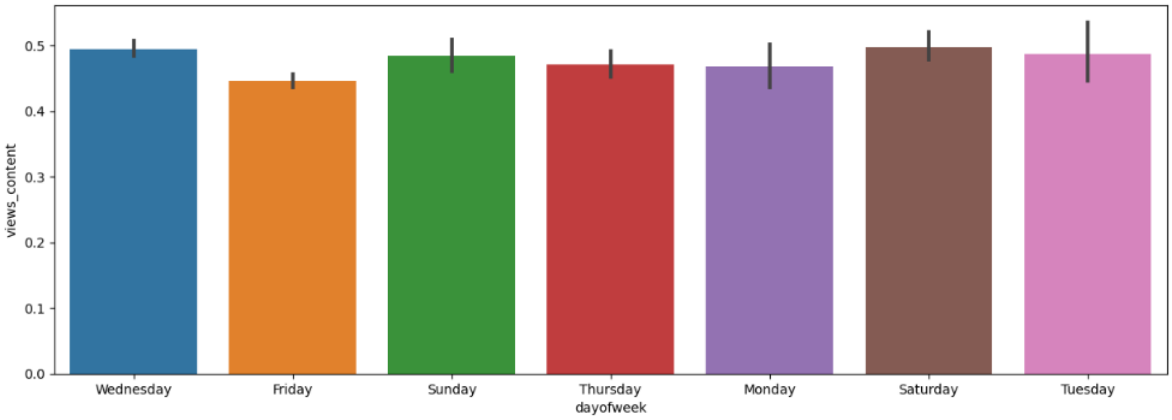
2. What does the distribution of genres look like?



- The other genre are higher in number compaed to other columns
- All other columns are almost equal in number.
- Almost 25% of the data is in other genre category.

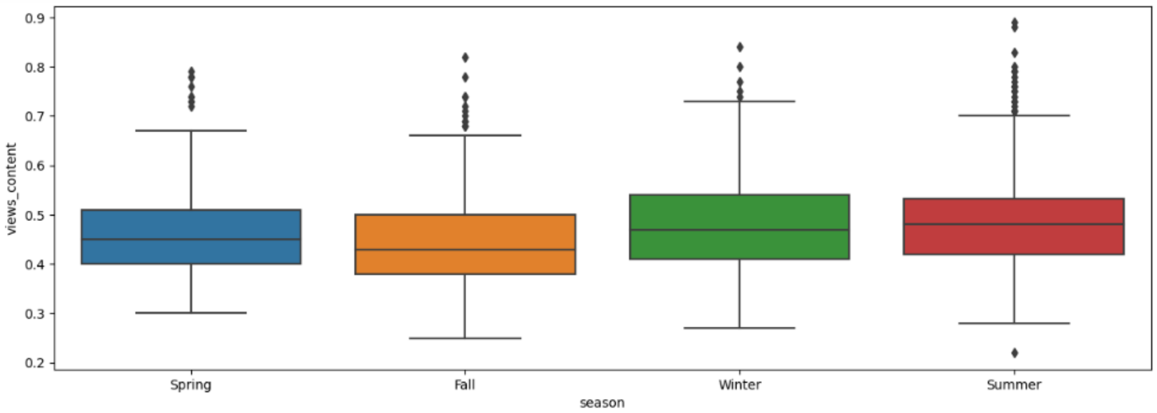
3. The day of the week on which content is released generally plays a key role in the viewership. How does the viewership vary with the day of release?

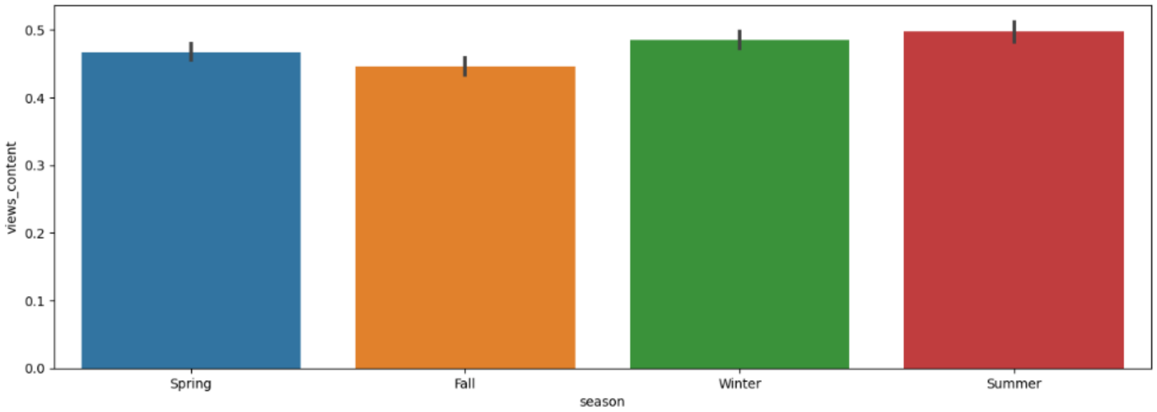




- The day of the week column have a considerable amount of impact on the viewership as the number of people viewing the content is more in Wednesday and less in Friday.
- It is interesting to see this kind of relationship where on Friday there 0.4 million to 0.6 million views.
- Whereas in all the other days the median value is almost close to 0.5 million views.
- We can clearly see that there are manu outliers in Friday and Wednesday which are ok in this case as the views can be 0.9 million which is quite normal.

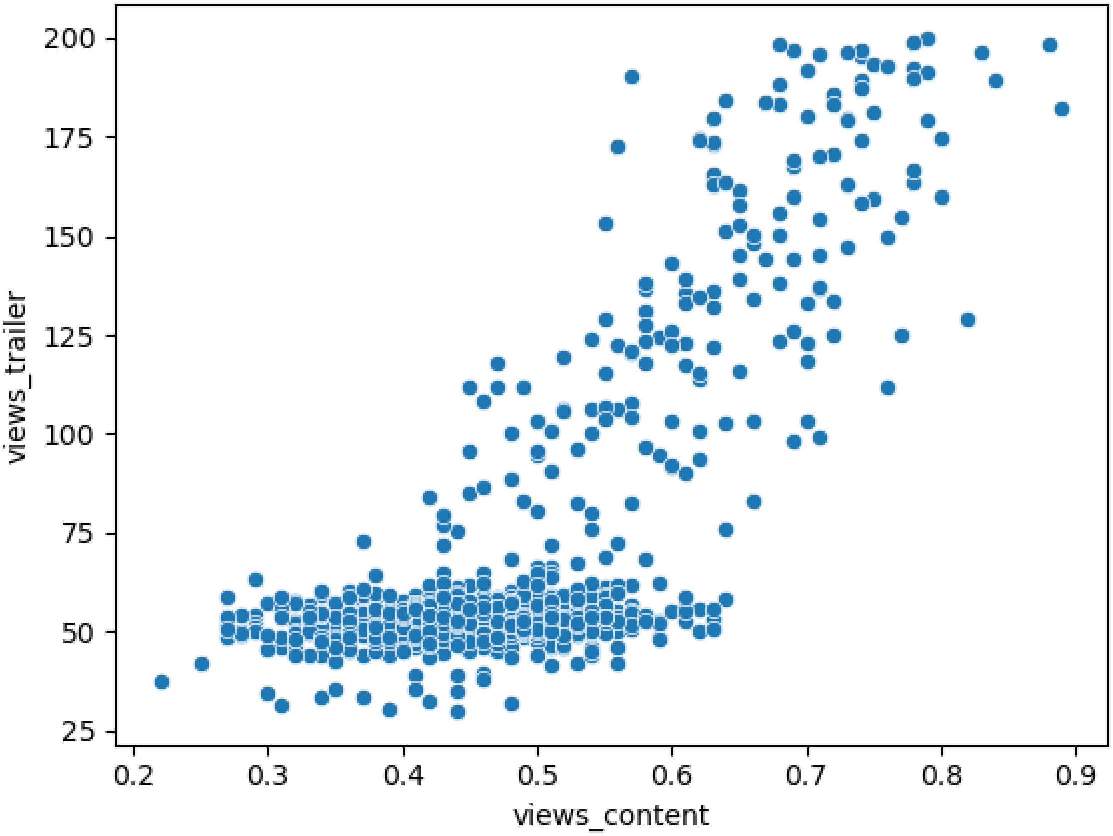
4. How does the viewership vary with the season of release?





- From the above plot we can conclude that the movies which are released on winter and summer have higher median value near 0.5 million.
- We can see there are outliers in all the seasons which is ok as there can be that high value of view.
- Summer have highest value ok nearly 0.9 million views_content which is the highest.
- Fall has the lowest views content which is nearly between 0.2 to 0.3 millions views.

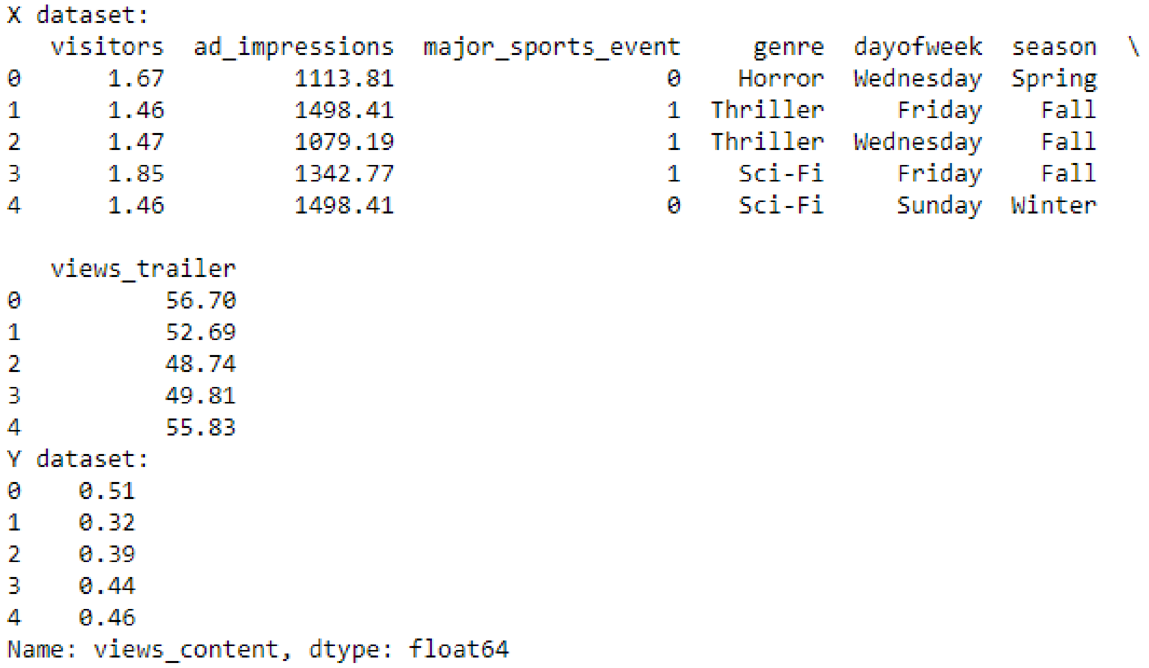
5. What is the correlation between trailer views and content views?



- From the above plot we can clearly see that the correlation between the trailer views and content views are positive correlation.
- we can conclude that they are highly correlated.

2. Data preprocessing

- We are creating new datasets with variable that needs to be predicted (Y dataset) and another dataset without the variable that need to be predicted (X dataset).



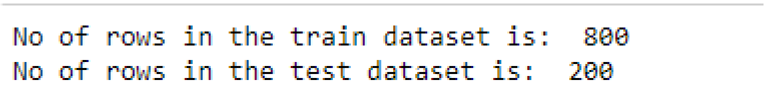
```
X dataset:
   visitors  ad_impressions  major_sports_event     genre  dayofweek  season  \
0      1.67         1113.81                   0    Horror  Wednesday  Spring
1      1.46         1498.41                   1  Thriller     Friday    Fall
2      1.47         1079.19                   1  Thriller  Wednesday    Fall
3      1.85         1342.77                   1    Sci-Fi     Friday    Fall
4      1.46         1498.41                   0    Sci-Fi     Sunday  Winter

   views_trailer
0          56.70
1          52.69
2          48.74
3          49.81
4          55.83
Y dataset:
0    0.51
1    0.32
2    0.39
3    0.44
4    0.46
Name: views_content, dtype: float64
```

- Next step is we are assigning the constant value in a variable and creating dummy variables for categorical variable present in the dataset.

- After assigning the constant and dummy variable we are splitting the dataset for training and testing data.

- 80% of the data is used for training the model and 20% ok data is used for Testing the model.

Splitting dataset to Training and Testing dataset:

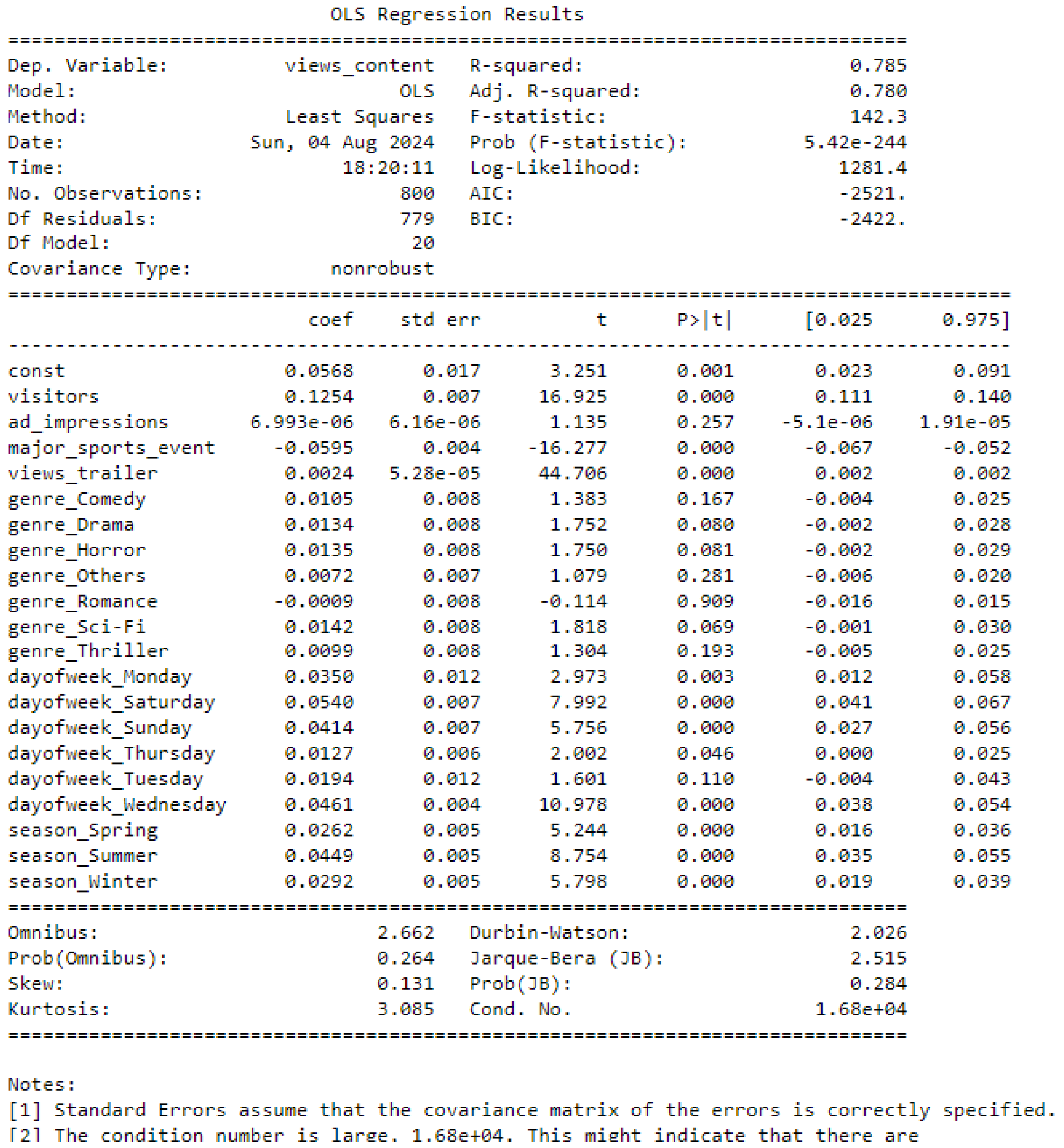


```
No of rows in the train dataset is:  800
No of rows in the test dataset is:  200
```

- 80% of 100 is 800 so 800 data is used for training.

- 20% of the dataset which is 200 data is used for testing the efficiency of the model.

3. Model building - Linear Regression



```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:           views_content   R-squared:                       0.785
Model:                             OLS   Adj. R-squared:                  0.780
Method:                  Least Squares   F-statistic:                     142.3
Date:                 Sun, 04 Aug 2024   Prob (F-statistic):          5.42e-244
Time:                         18:20:11   Log-Likelihood:                 1281.4
No. Observations:                  800   AIC:                            -2521.
Df Residuals:                      779   BIC:                            -2422.
Df Model:                           20                                         
Covariance Type:             nonrobust                                         
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                   0.0568      0.017      3.251      0.001       0.023       0.091
visitors                0.1254      0.007     16.925      0.000       0.111       0.140
ad_impressions       6.993e-06   6.16e-06      1.135      0.257    -5.1e-06    1.91e-05
major_sports_event     -0.0595      0.004    -16.277      0.000      -0.067      -0.052
views_trailer           0.0024   5.28e-05     44.706      0.000       0.002       0.002
genre_Comedy            0.0105      0.008      1.383      0.167      -0.004       0.025
genre_Drama             0.0134      0.008      1.752      0.080      -0.002       0.028
genre_Horror            0.0135      0.008      1.750      0.081      -0.002       0.029
genre_Others            0.0072      0.007      1.079      0.281      -0.006       0.020
genre_Romance          -0.0009      0.008     -0.114      0.909      -0.016       0.015
genre_Sci-Fi            0.0142      0.008      1.818      0.069      -0.001       0.030
genre_Thriller          0.0099      0.008      1.304      0.193      -0.005       0.025
dayofweek_Monday        0.0350      0.012      2.973      0.003       0.012       0.058
dayofweek_Saturday      0.0540      0.007      7.992      0.000       0.041       0.067
dayofweek_Sunday        0.0414      0.007      5.756      0.000       0.027       0.056
dayofweek_Thursday      0.0127      0.006      2.002      0.046       0.000       0.025
dayofweek_Tuesday       0.0194      0.012      1.601      0.110      -0.004       0.043
dayofweek_Wednesday     0.0461      0.004     10.978      0.000       0.038       0.054
season_Spring           0.0262      0.005      5.244      0.000       0.016       0.036
season_Summer           0.0449      0.005      8.754      0.000       0.035       0.055
season_Winter           0.0292      0.005      5.798      0.000       0.019       0.039
==============================================================================
Omnibus:                        2.662   Durbin-Watson:                   2.026
Prob(Omnibus):                  0.264   Jarque-Bera (JB):                2.515
Skew:                           0.131   Prob(JB):                        0.284
Kurtosis:                       3.085   Cond. No.                     1.68e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.68e+04. This might indicate that there are
```

1.Adjusted. R-squared:

- It reflects the fit of the model.

- Adjusted R-squared values generally range from 0 to 1, where a higher value generally indicates a better fit, assuming certain conditions are met.

- In our case, the value for adj. R-squared is 0.780, which is good.

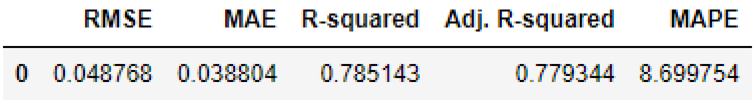
2.const coefficient:

- It is the Y-intercept. It means that if all the predictor variable coefficients are zero, then the expected output (i.e., Y) would be equal to the const coefficient.

- In our case, the value for const coefficient is 0.0568.

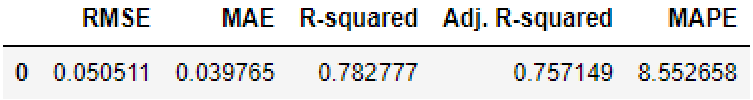
Modal Performance Check:

Performance of Training Dataset:



	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.048768	0.038804	0.785143	0.779344	8.699754

Performance of Testing dataset:



	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.050511	0.039765	0.782777	0.757149	8.552658

Observations:

- The training R^2 is 0.78, so the model is not underfitting.

- The train and test RMSE and MAE are comparable, so the model is not overfitting either

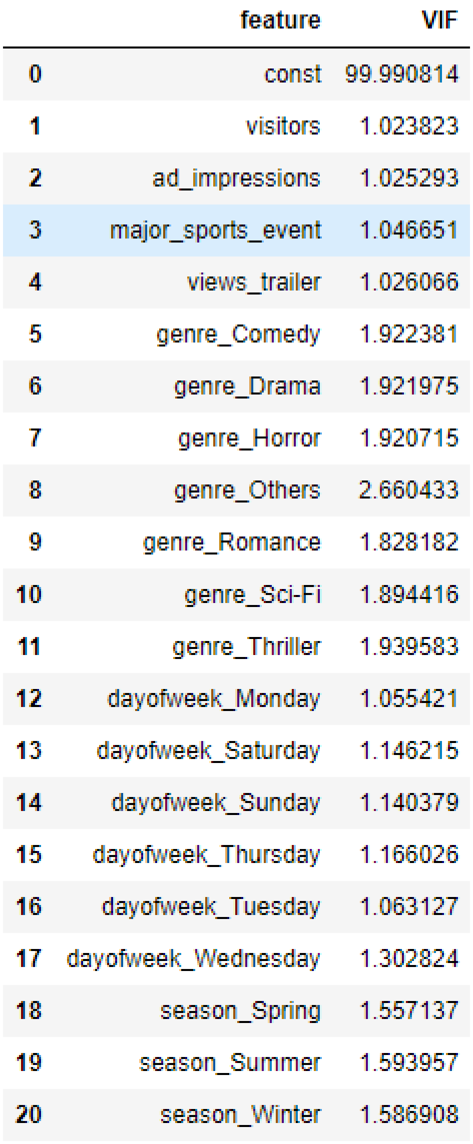
- MAE suggests that the model can predict content views within a mean error of 0.39 on the test data.

- MAPE of 8.5 on the test data means that we are able to predict within 8.6% of the content views.

4. Testing the assumptions of linear regression model

Test for Multicollinearity:

Checking the VIF value to check the Multicollinearity of the dataset.

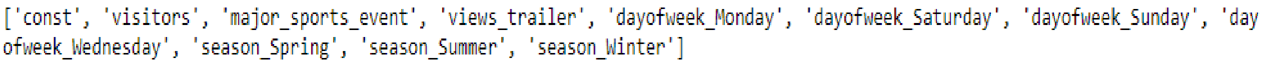


	feature	VIF
0	const	99.990814
1	visitors	1.023823
2	ad_impressions	1.025293
3	major_sports_event	1.046651
4	views_trailer	1.026066
5	genre_Comedy	1.922381
6	genre_Drama	1.921975
7	genre_Horror	1.920715
8	genre_Others	2.660433
9	genre_Romance	1.828182
10	genre_Sci-Fi	1.894416
11	genre_Thriller	1.939583
12	dayofweek_Monday	1.055421
13	dayofweek_Saturday	1.146215
14	dayofweek_Sunday	1.140379
15	dayofweek_Thursday	1.166026
16	dayofweek_Tuesday	1.063127
17	dayofweek_Wednesday	1.302824
18	season_Spring	1.557137
19	season_Summer	1.593957
20	season_Winter	1.586908

- No multicollinearity is present among the variables as all VIF values are below 5.

Dealing with high p-value variables

- If p-values are greater than 0.05, we will drop these variables.

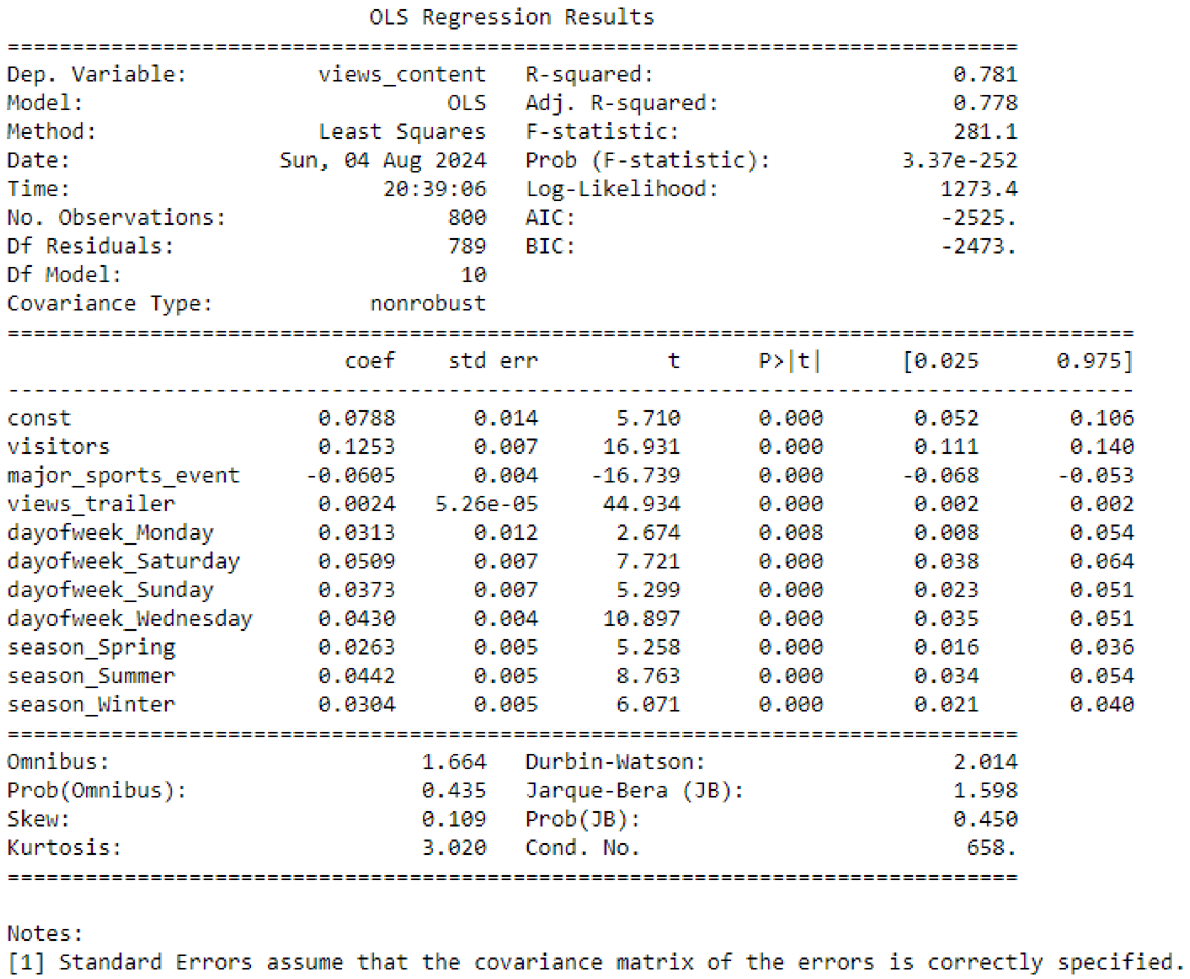


```
['const', 'visitors', 'major_sports_event', 'views_trailer', 'dayofweek_Monday', 'dayofweek_Saturday', 'dayofweek_Sunday', 'day
ofweek_Wednesday', 'season_Spring', 'season_Summer', 'season_Winter']
```

- The above columns have the P value less than > 5.

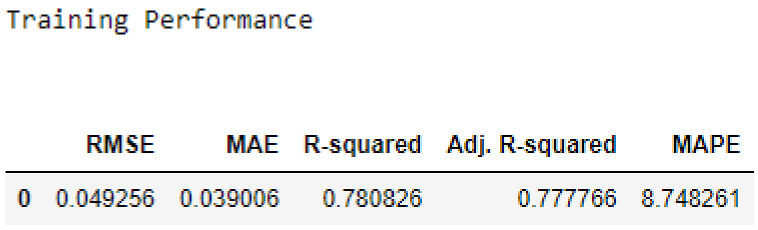
Building the Model:

The model consist of p value < 0.05 is created and the values as below,



```
                            OLS Regression Results                            
==============================================================================
Dep. Variable:          views_content   R-squared:                       0.781
Model:                            OLS   Adj. R-squared:                  0.778
Method:                 Least Squares   F-statistic:                     281.1
Date:                Sun, 04 Aug 2024   Prob (F-statistic):          3.37e-252
Time:                        20:39:06   Log-Likelihood:                 1273.4
No. Observations:                 800   AIC:                            -2525.
Df Residuals:                     789   BIC:                            -2473.
Df Model:                          10                                         
Covariance Type:            nonrobust                                         
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                   0.0788      0.014      5.710      0.000       0.052       0.106
visitors                0.1253      0.007     16.931      0.000       0.111       0.140
major_sports_event     -0.0605      0.004    -16.739      0.000      -0.068      -0.053
views_trailer           0.0024   5.26e-05     44.934      0.000       0.002       0.002
dayofweek_Monday        0.0313      0.012      2.674      0.008       0.008       0.054
dayofweek_Saturday      0.0509      0.007      7.721      0.000       0.038       0.064
dayofweek_Sunday        0.0373      0.007      5.299      0.000       0.023       0.051
dayofweek_Wednesday     0.0430      0.004     10.897      0.000       0.035       0.051
season_Spring           0.0263      0.005      5.258      0.000       0.016       0.036
season_Summer           0.0442      0.005      8.763      0.000       0.034       0.054
season_Winter           0.0304      0.005      6.071      0.000       0.021       0.040
==============================================================================
Omnibus:                        1.664   Durbin-Watson:                   2.014
Prob(Omnibus):                  0.435   Jarque-Bera (JB):                1.598
Skew:                           0.109   Prob(JB):                        0.450
Kurtosis:                       3.020   Cond. No.                         658.
==============================================================================

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

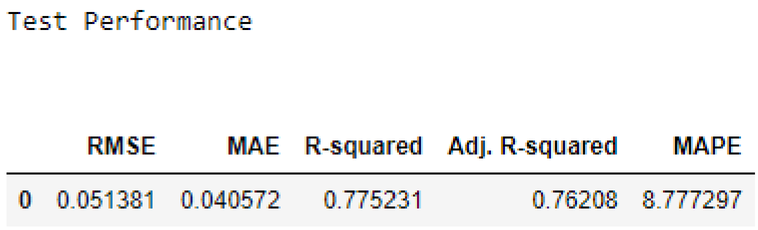
Performance of Training Dataset:



Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.049256	0.039006	0.780826	0.777766	8.748261

Performance of Testing Dataset:



```
Test Performance
```

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.051381	0.040572	0.775231	0.76208	8.777297

-The model is able to explain ~78% of the variation in the data.

- The train and test RMSE and MAE are low and comparable. So, our model is not suffering from overfitting.

- The MAPE on the test set suggests we can predict within 8.8% of the anime ratings.

- Hence, we can conclude the model *olsmodel_final* is good for prediction as well as inference purposes.

Test for lineraity and normality

Why the test?

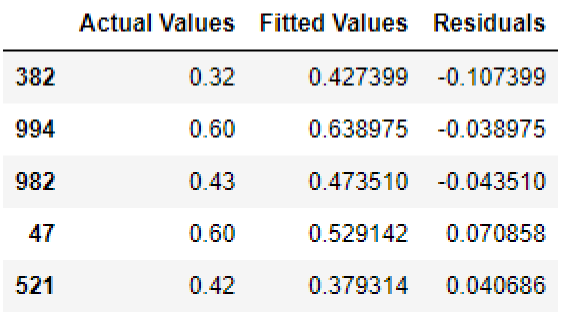
- Linearity describes a straight-line relationship between two variables, predictor variables must have a linear relation with the dependent variable.
- The independence of the error terms (or residuals) is important. If the residuals are not independent, then the confidence intervals of the coefficient estimates will be narrower and make us incorrectly conclude a parameter to be statistically significant.

How to check linearity and independence?

- Make a plot of fitted values vs residuals.
- If they don't follow any pattern, then we say the model is linear and residuals are independent.
- Otherwise, the model is showing signs of non-linearity and residuals are not independent.

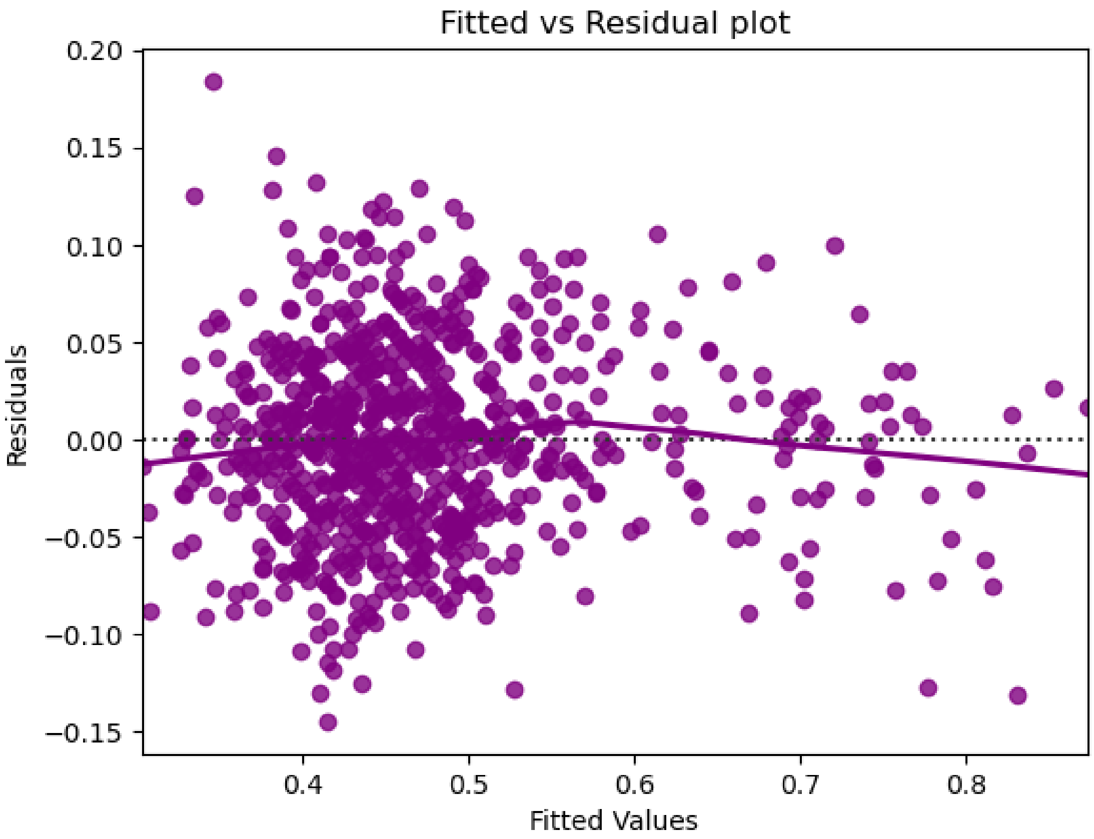
How to fix if this assumption is not followed?

- We can try to transform the variables and make the relationships linear.



	Actual Values	Fitted Values	Residuals
382	0.32	0.427399	-0.107399
994	0.60	0.638975	-0.038975
982	0.43	0.473510	-0.043510
47	0.60	0.529142	0.070858
521	0.42	0.379314	0.040686

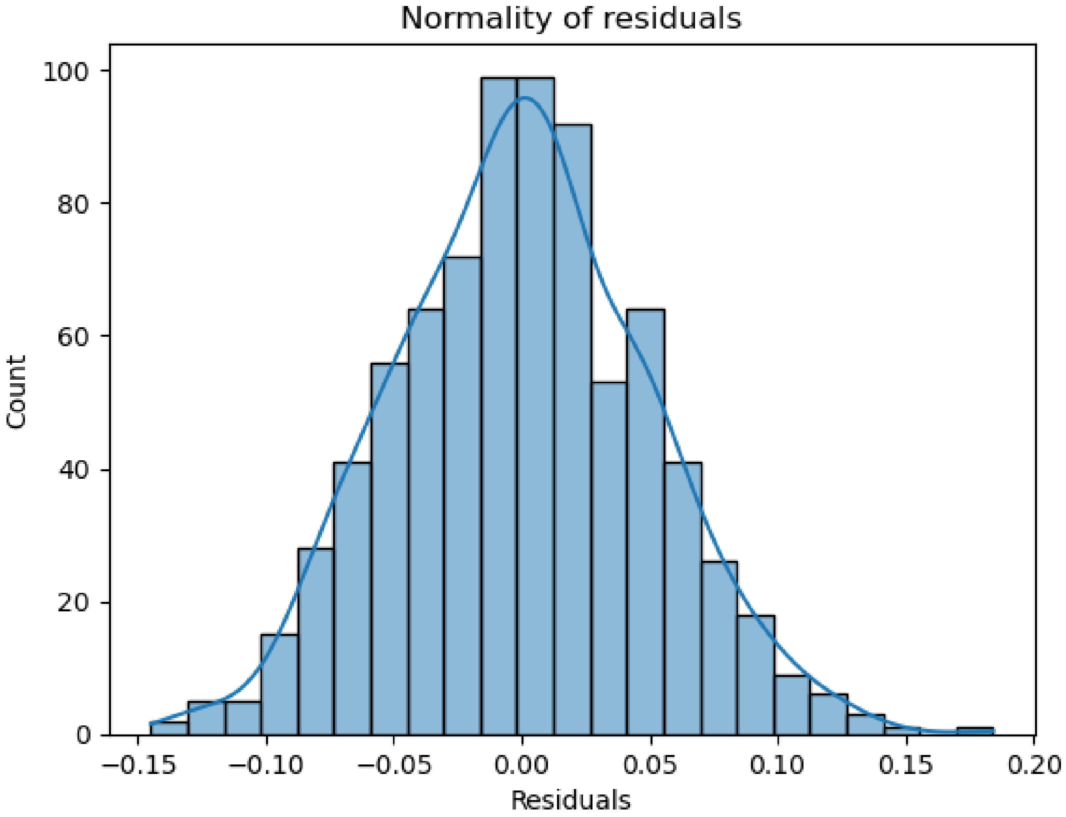
- We are going to plot a Scatter plot from the above table which consist of Actual values, Fitted values and residuals.



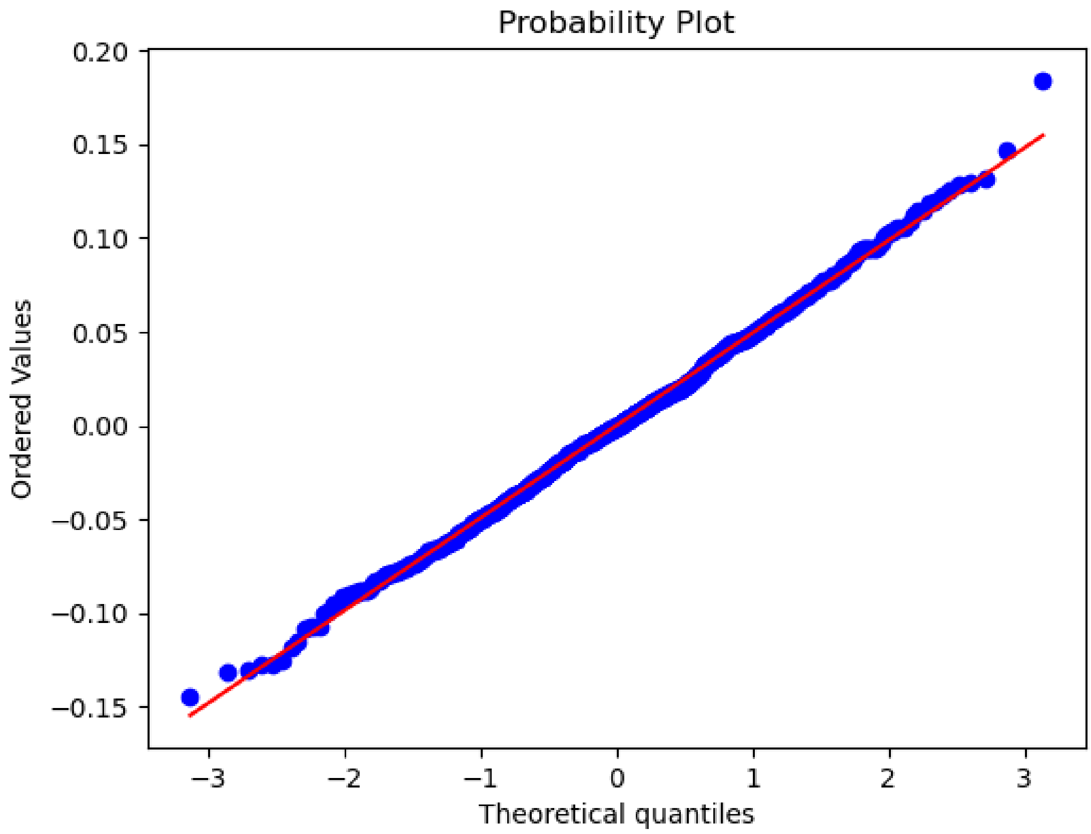
- The scatter plot shows the distribution of residuals (errors) vs fitted values (predicted values).
- If there exist any pattern in this plot, we consider it as signs of non-linearity in the data and a pattern means that the model doesn't capture non-linear effects.
- **We see no pattern in the plot above. Hence, the assumptions of linearity and independence are satisfied.**
-

The shape of the histogram of residuals can give an initial idea about the normality.

- The shape of the histogram of residuals can give an initial idea about the normality.
- It can also be checked via a Q-Q plot of residuals. If the residuals follow a normal distribution, they will make a straight line plot, otherwise not.
- Other tests to check for normality includes the Shapiro-Wilk test.
    - Null hypothesis: Residuals are normally distributed
    - Alternate hypothesis: Residuals are not normally distributed

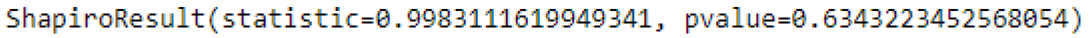


- The histogram of residuals does have a bell shape.
- Let's check the Q-Q plot.



- The residuals more or less follow a straight line except for the tails.
- Let's check the results of the Shapiro-Wilk test.

Shapiro's wilk test:



```
ShapiroResult(statistic=0.9983111619949341, pvalue=0.6343223452568054)
```

- Since p-value > 0.05, the residuals are normal as per the Shapiro-Wilk test.

- Strictly speaking, the residuals are normal.

- we can accept this distribution as close to being normal.

- So, the assumption is satisfied.

Test for Homoscedascity

Homoscedascity: If the variance of the residuals is symmetrically distributed across the regression line, then the data is said to be homoscedastic.

Heteroscedascity: If the variance is unequal for the residuals across the regression line, then the data is said to be heteroscedastic.

Why the test?

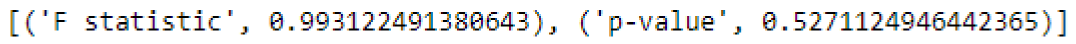
- The presence of non-constant variance in the error terms results in heteroscedasticity. Generally, non-constant variance arises in presence of outliers.

How to check for homoscedasticity?

- The residual vs fitted values plot can be looked at to check for homoscedasticity. In the case of heteroscedasticity, the residuals can form an arrow shape or any other non-symmetrical shape.
- The goldfeldquandt test can also be used. If we get a p-value > 0.05 we can say that the residuals are homoscedastic. Otherwise, they are heteroscedastic.
    - Null hypothesis: Residuals are homoscedastic
    - Alternate hypothesis: Residuals have heteroscedasticity

How to fix if this assumption is not followed?

- Heteroscedasticity can be fixed by adding other important features or making transformations.

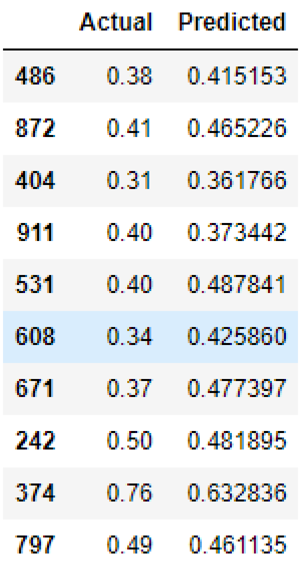


```
[('F statistic', 0.993122491380643), ('p-value', 0.5271124946442365)]
```

- Since p-value > 0.05, we can say that the residuals are homoscedastic. So, this assumption is satisfied.

Prediction on Test Data:

Now that we have checked all the assumptions of linear regression and they are satisfied, let's go ahead with prediction.

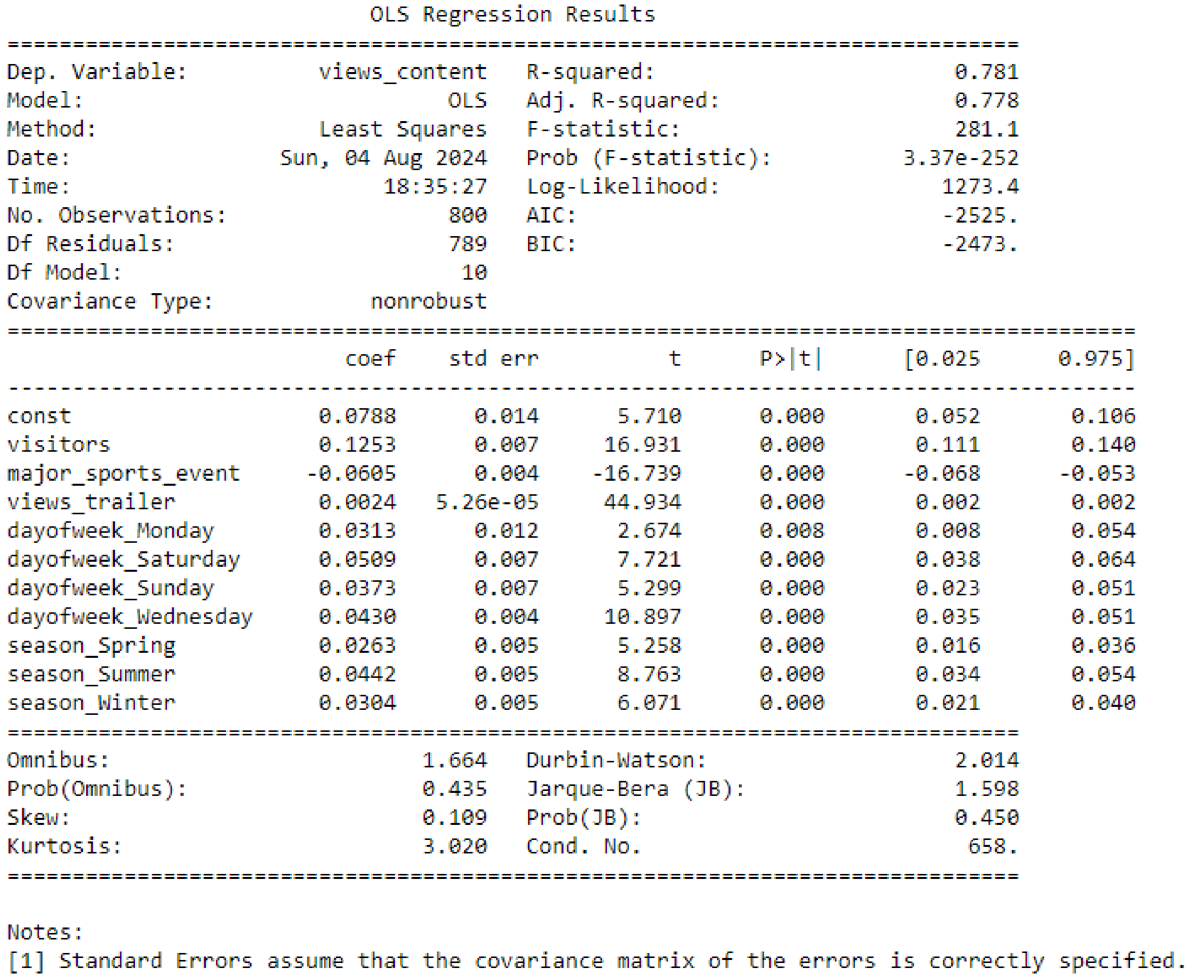


	Actual	Predicted
486	0.38	0.415153
872	0.41	0.465226
404	0.31	0.361766
911	0.40	0.373442
531	0.40	0.487841
608	0.34	0.425860
671	0.37	0.477397
242	0.50	0.481895
374	0.76	0.632836
797	0.49	0.461135

- We can observe here that our model has returned pretty good prediction results, and the actual and predicted values are comparable.

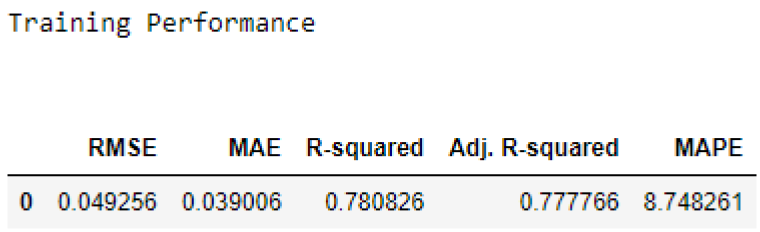
5. Model performance evaluation
Final modal:

- Let's recreate the final model and print it's summary to gain insights.



```
                            OLS Regression Results
==============================================================================
Dep. Variable:          views_content   R-squared:                       0.781
Model:                            OLS   Adj. R-squared:                  0.778
Method:                 Least Squares   F-statistic:                     281.1
Date:                Sun, 04 Aug 2024   Prob (F-statistic):         3.37e-252
Time:                        18:35:27   Log-Likelihood:                 1273.4
No. Observations:                 800   AIC:                            -2525.
Df Residuals:                     789   BIC:                            -2473.
Df Model:                          10
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
const                     0.0788      0.014      5.710      0.000       0.052       0.106
visitors                  0.1253      0.007     16.931      0.000       0.111       0.140
major_sports_event       -0.0605      0.004    -16.739      0.000      -0.068      -0.053
views_trailer             0.0024   5.26e-05     44.934      0.000       0.002       0.002
dayofweek_Monday          0.0313      0.012      2.674      0.008       0.008       0.054
dayofweek_Saturday        0.0509      0.007      7.721      0.000       0.038       0.064
dayofweek_Sunday          0.0373      0.007      5.299      0.000       0.023       0.051
dayofweek_Wednesday       0.0430      0.004     10.897      0.000       0.035       0.051
season_Spring             0.0263      0.005      5.258      0.000       0.016       0.036
season_Summer             0.0442      0.005      8.763      0.000       0.034       0.054
season_Winter             0.0304      0.005      6.071      0.000       0.021       0.040
==============================================================================
Omnibus:                        1.664   Durbin-Watson:                   2.014
Prob(Omnibus):                  0.435   Jarque-Bera (JB):                1.598
Skew:                           0.109   Prob(JB):                        0.450
Kurtosis:                       3.020   Cond. No.                         658.
==============================================================================

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

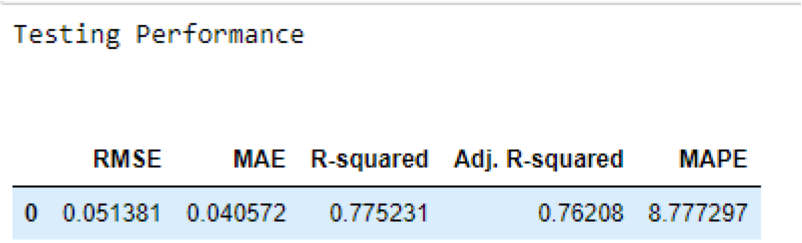
Performance of Training Dataset:



Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.049256	0.039006	0.780826	0.777766	8.748261

Performance of Testing Dataset:



Testing Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.051381	0.040572	0.775231	0.76208	8.777297

- The model is able to explain ~78% of the variation in the data

- The train and test RMSE and MAE are low and comparable. So, our model is not suffering from overfitting

- The MAPE on the test set suggests we can predict within 8.8% of the anime ratings

- Hence, we can conclude the model olsmodel_final is good for prediction as well as inference purposes

Final Conclusion:

- The model is able to explain ~78% of the variation in the data and within 8.8% of the content views on the test data, which is good.

- This indicates that the model is good for prediction as well as inference purposes

- If the visitors increases by one unit, then its content views increases by 0.1253 units.

- If the number of trailer views increases by one unit, then its content views increases by 0.0024 units.

- If there is a major sports events happening, then content views decreases by 0.0605 units.

6. Actionable Insights & Recommendations

- The movies which are released on a day which does not have a major sports event tends to have more number of views so it is important to plan movie release on a non major sports event day
- The movie with a more number of views in its trailer tends to have more number of views in the content so it is important point to note and to make the trailer more impressive and attractive to cover the audience.
- The movie release planned on Summer and winter have more views so summer movies have more views which is a good insight to note.