PGP-DSBA PROJECT REPORT

PM – Coded Project

BY

ISHAAN SHAKTI JAYARAMAN

PGPDSBA.O.JULY24.A

Contents

LIST OF FIGURES

EXPLORATORY DATA ANALYSIS

1.1 Introduction

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.

1.2 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 spends, content timing clashes, weekends and holidays, etc. We have received the data of the current content in their platform, and we need to analyze the data and come up with a linear regression model to determine the driving factors for first-day viewership.

1.3 Data Overview

1.3.1 Problem Definition

ShowTime is an over-the-top (OTT) service provider that offers a diverse array of content, including movies and web shows. Recently, the platform has observed a decline in first-day viewership of its new releases. Understanding the factors that influence initial viewership is crucial for the platform to implement effective strategies to enhance audience engagement and improve viewership metrics.

The primary objective of this analysis is to identify and quantify the key driver variables that significantly impact first-day viewership of content on ShowTime. By constructing a linear regression model, we aim to predict first-day viewership based on various influencing factors and provide actionable insights to improve content performance.

1.3.2 Key Questions

These are the key questions to be answered during the exploratory data analysis

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?

1.3.3 Data Contents

The dataset (ottdata.csv) consists of data related to an OTT (Over the Top) service provider

- There are 1000 unique observations in the dataset.
- There are 8 columns with various information.
- There are 5 object data types and 3 numerical data types.
- major_sports_event column represents whether a sports event has occurred that day (0 = no, 1 = yes) This can be considered as categorical data.
- There are no null or missing values.
- There are no duplicate entries.

1.3.4 Data Dictionary

The data contains the different factors to analyze for the content. The detailed data dictionary is given below.

- 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 (0 = 'No', 1 = 'Yes')
- 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

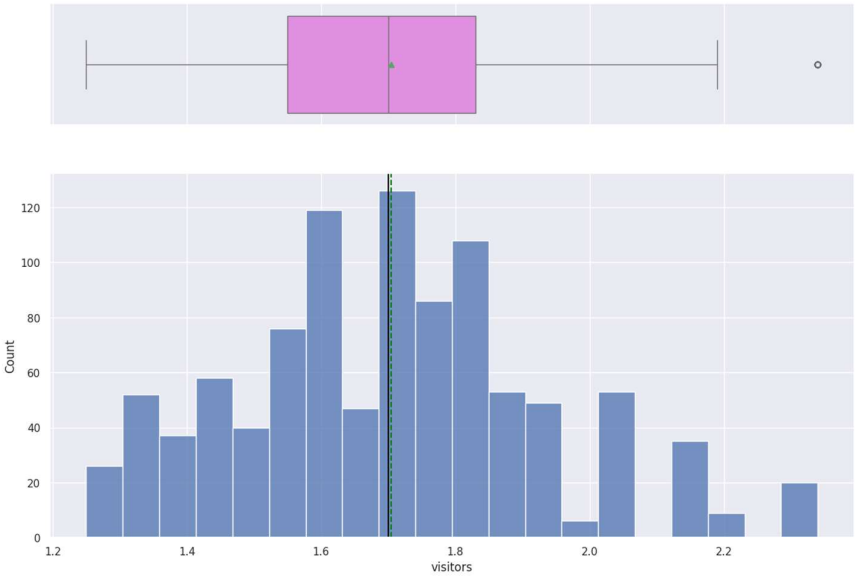
1.4 EDA

1.4.1 Univariate Analysis

A univariate analysis explores all variables in a dataset separately so that we may identify any patterns. In this dataset we will explore the numerical and categorical variables individually.

Visitors

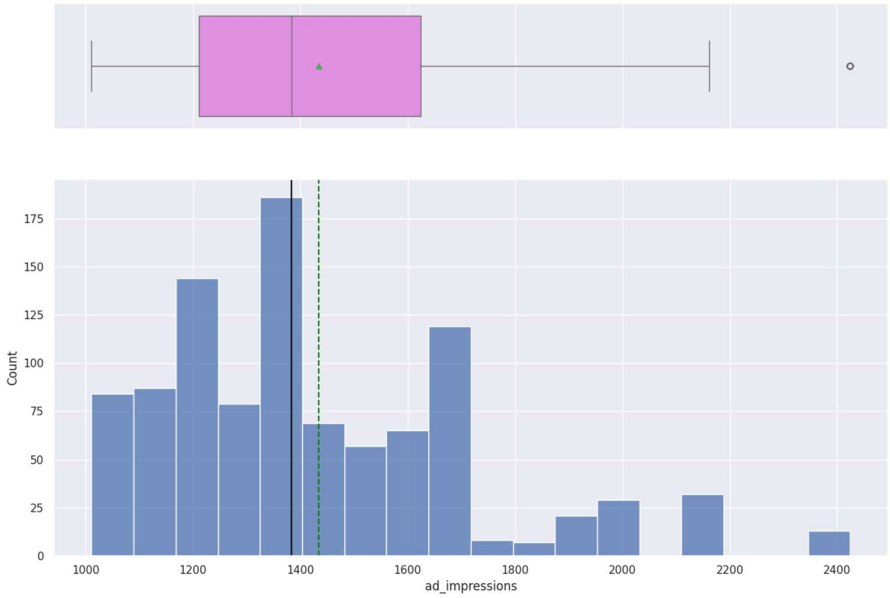
Figure 1 - Distribution of 'visitors'



- This shows the distribution is almost normal with a slight right skew
- On average there are approximately 1.7 million visitors

Ad Impressions

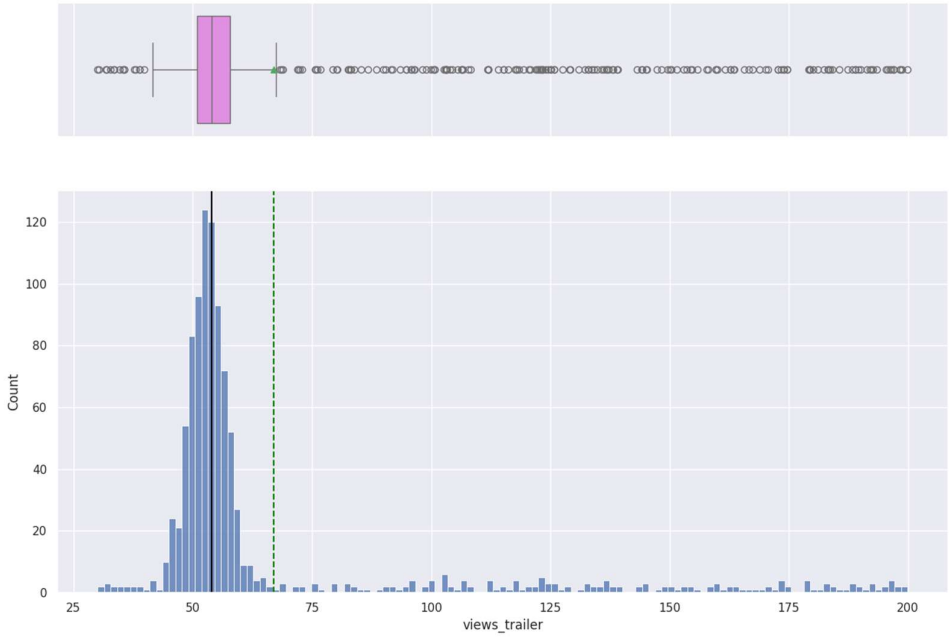
Figure 2 - Distribution of 'ad_impressions'



- The distribution is right skewed with median ad impressions less than 1400
- There are few occasions where the ads were displayed more than 2000 million times

Trailer Views

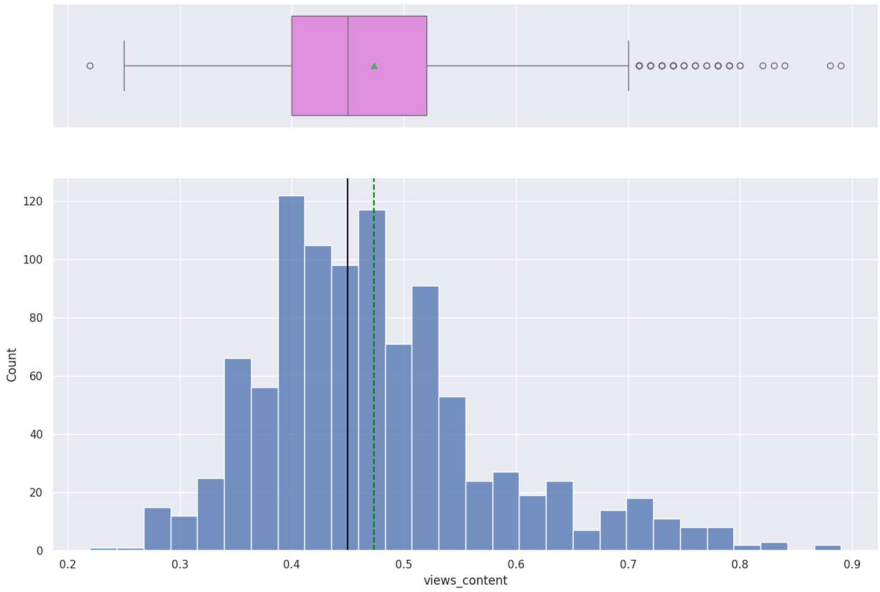
Figure 3 - Distribution of 'views_trailer'



- The distribution is right skewed with median views around 55 million
- There are a significant amount of outliers present

Content Views

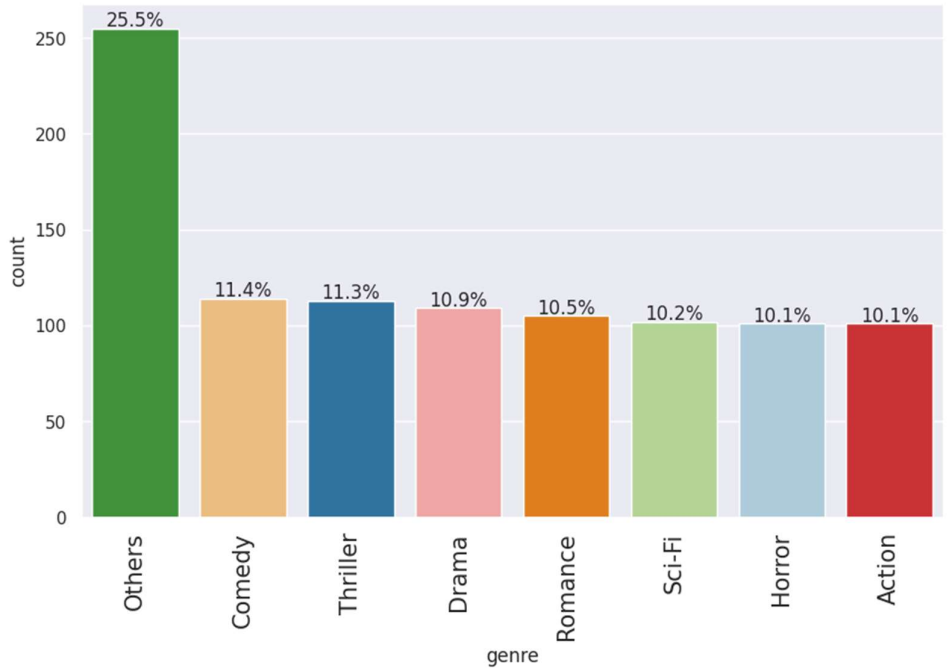
Figure 4 - Distribution of 'views_content'



- The distribution is almost normal with a slight right skew
- The median of content views is approximately 0.45 million

Genre

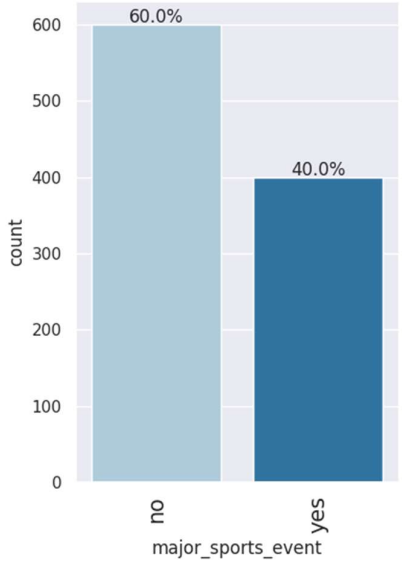
Figure 5 - Bar plot of 'genre'



- 25.5% of the data are under 'Others' genre category
- There is an almost uniform distribution with the rest of genres, with 'Comedy' and 'Thriller' being observed more

Major Sports Event

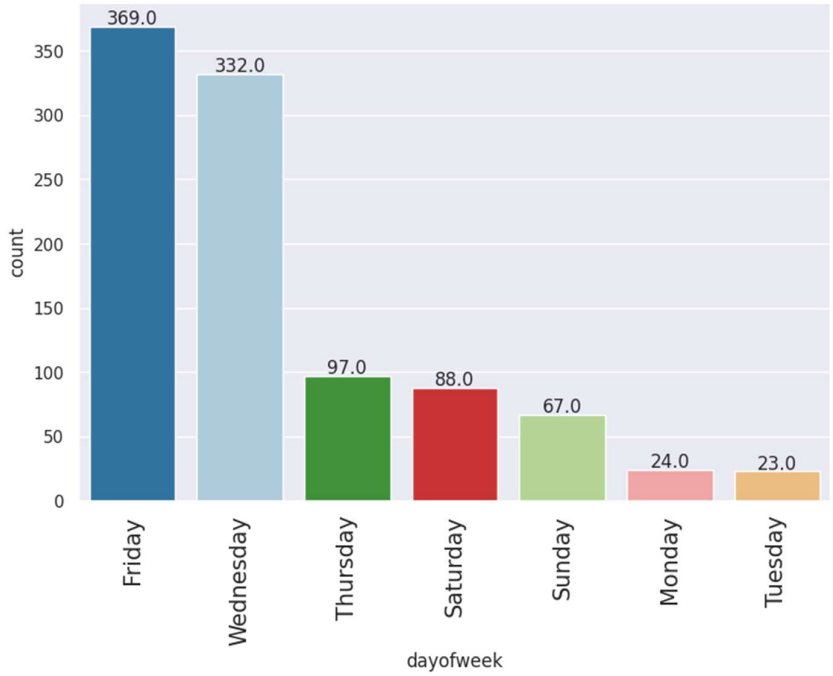
Figure 6 - Bar plot of 'major_sports_event'



- 60% of the observations show that no major sports event occurred on the day
- 40% of the observations show that there were major sports event on the day

Days of the Week

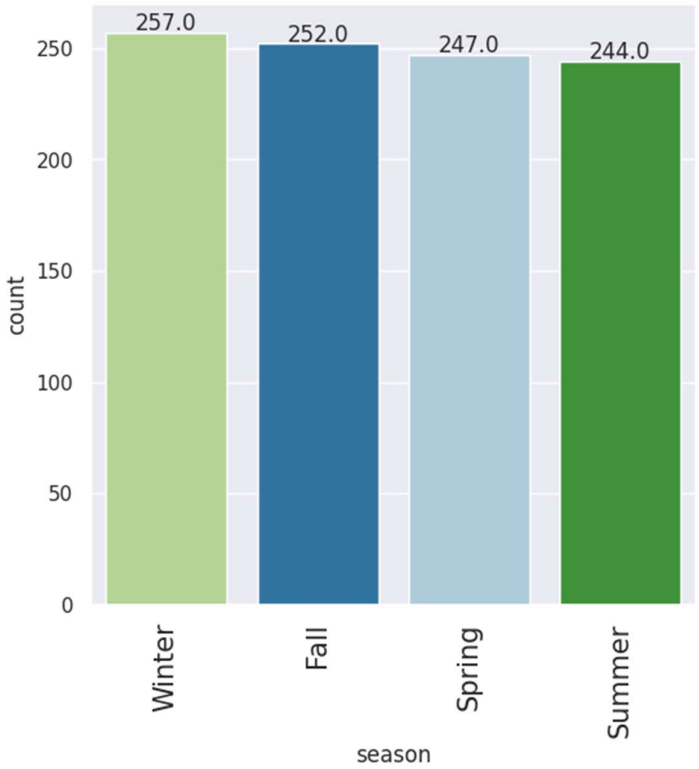
Figure 7 - Bar plot of 'daysofweek'



- The chart shows that ‘Friday’ and ‘Wednesday’ are the days where the observations are the higest
- ‘Monday’ and ‘Tuesday’ have the lowest observations

Season

Figure 8 - Bar plot of 'season'



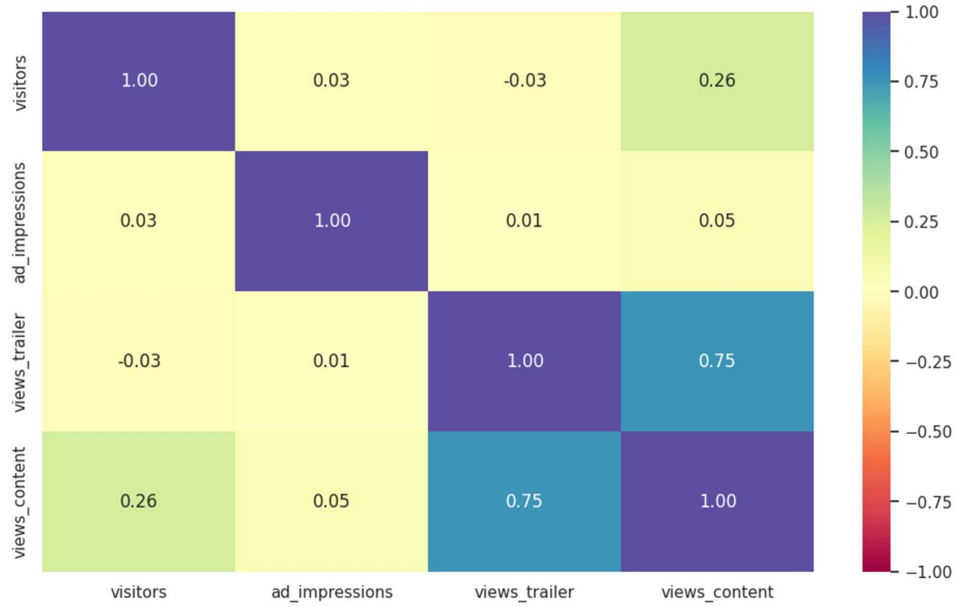
- This chart shows that ‘Winter’ is the most observations of content in the platform

1.4.2 Bivariate Analysis

A bivariate analysis explores the relationship between two or more variables in a dataset so that we may gain deeper insights. We will explore the relationship and correlation between all numerical variables as well as the relationship between numerical and categorical variables.

Heatmap

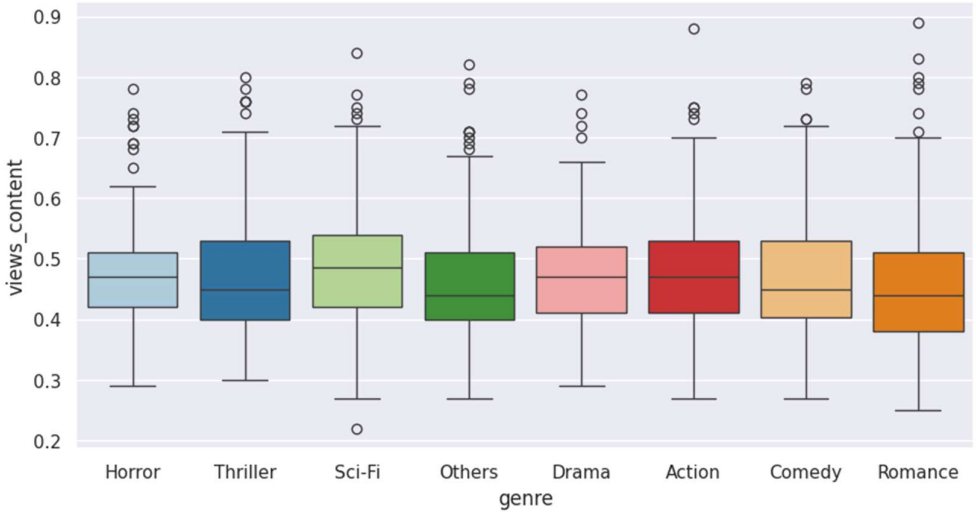
Figure 9 - Heatmap of all numerical variables



- The heatmap shows that trailer views and content views have the highest correlation with each other at 0.75
- Trailer views and visitors have the lowest correlation at -0.03

Genre against Views Content

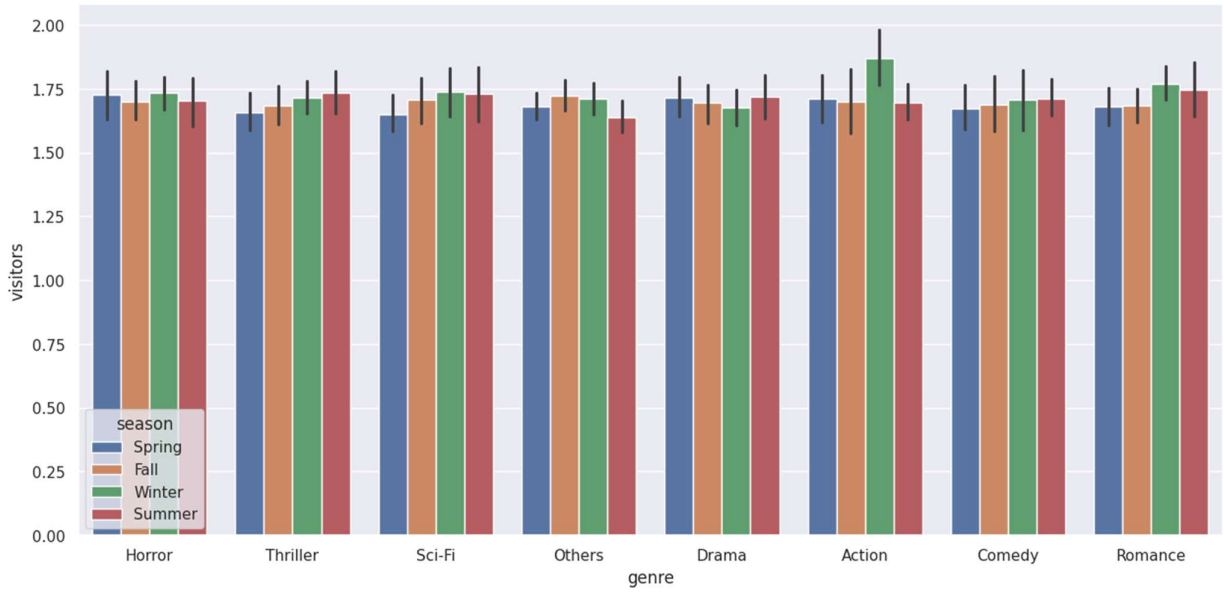
Figure 10 - Boxplot of 'genre' against 'views_content'



- ‘Thriller’ and ‘Romance’ genres have the highest number of views
- ‘Sci-Fi’ genre and ‘Action’ genre are the next most viewed content

Genre & Season against platform visitors

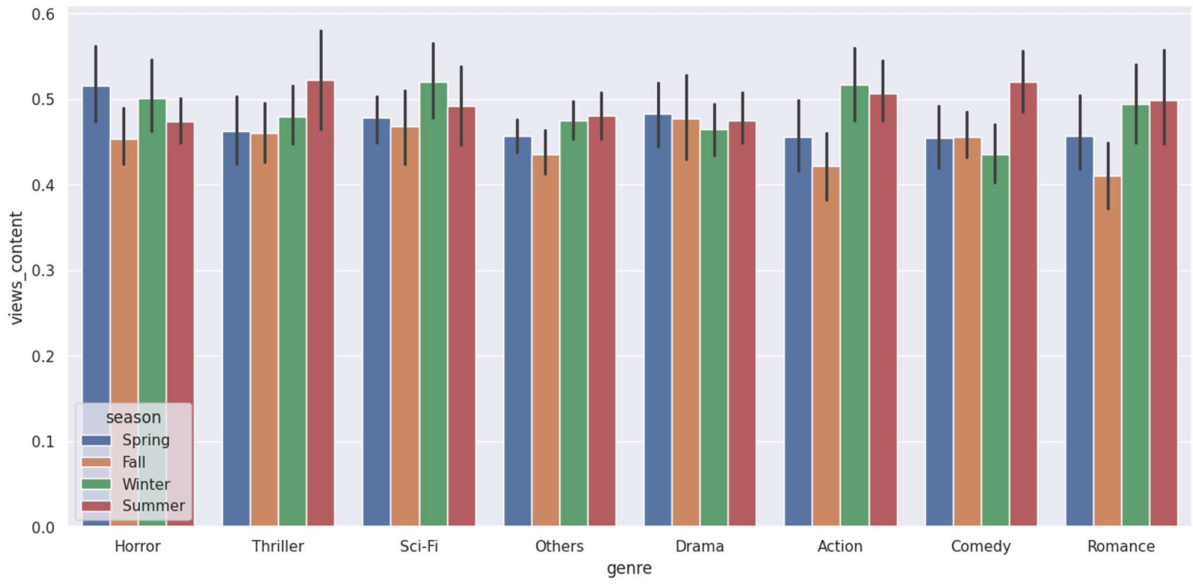
Figure 11 - Barplot of 'genre' & 'season' against 'visitors'



- There is a spike in visitors to the platform during Winter season due to interest in the 'Action' genre.

Genre & Season against content views

Figure 12 - Barplot of 'genre' & 'season' against 'views_content'

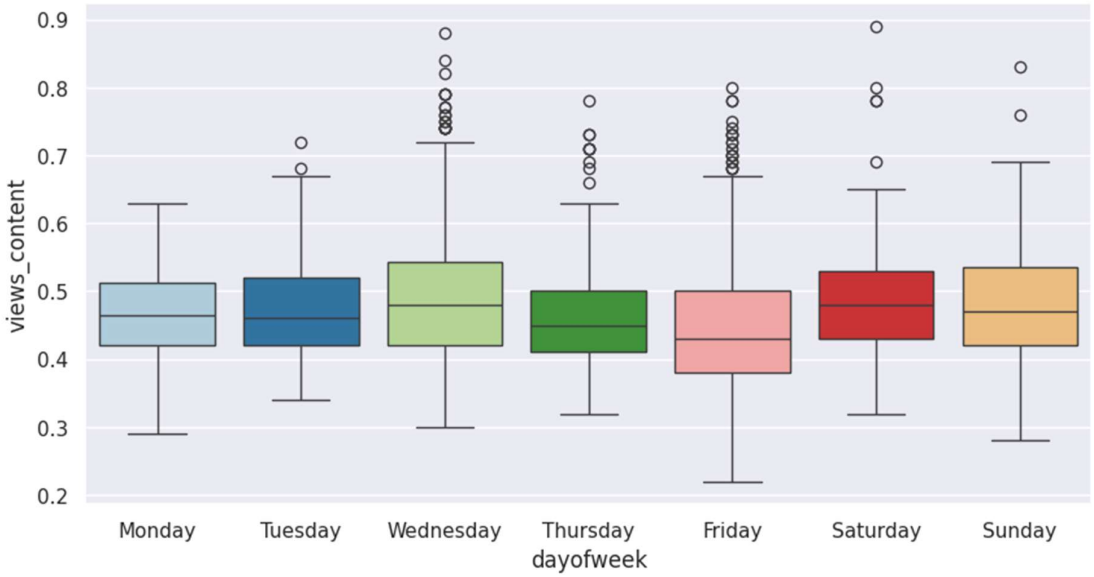


- During Summer, 'Thriller', 'Action' and 'Comedy' generates a lot of first day views.
- 'Horror' genre has a lot of first day views during Spring season

- Fall season has the lowest first day views among all genres

Days of the week against Content views

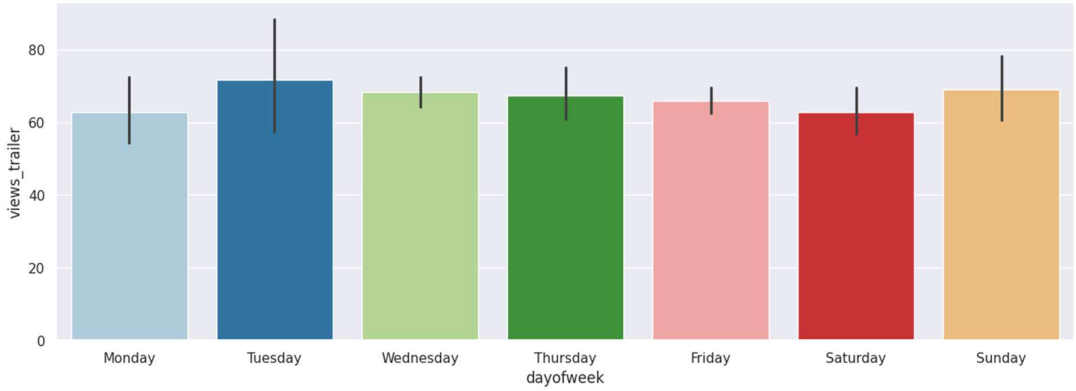
Figure 13 - Boxplot of 'dayofweek' against 'views_content'



- Based on the boxplot, Wednesday and Saturday has the highest number of first day views of the content

Day of the week against trailer views

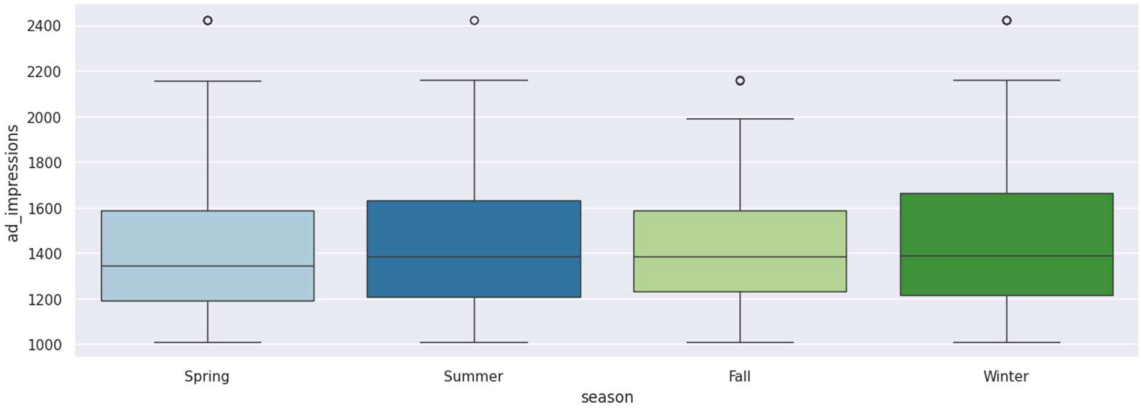
Figure 14 - Barplot of 'dayofweek' against 'views_trailer'



- Based on the bar plot we can see that Tuesday generates the highest amount of trailers views in the platform
- Sunday is the second highest contributor to trailer views

Season against Ad impressions

Figure 15 - Boxplot of 'season' against 'ad_impressions'

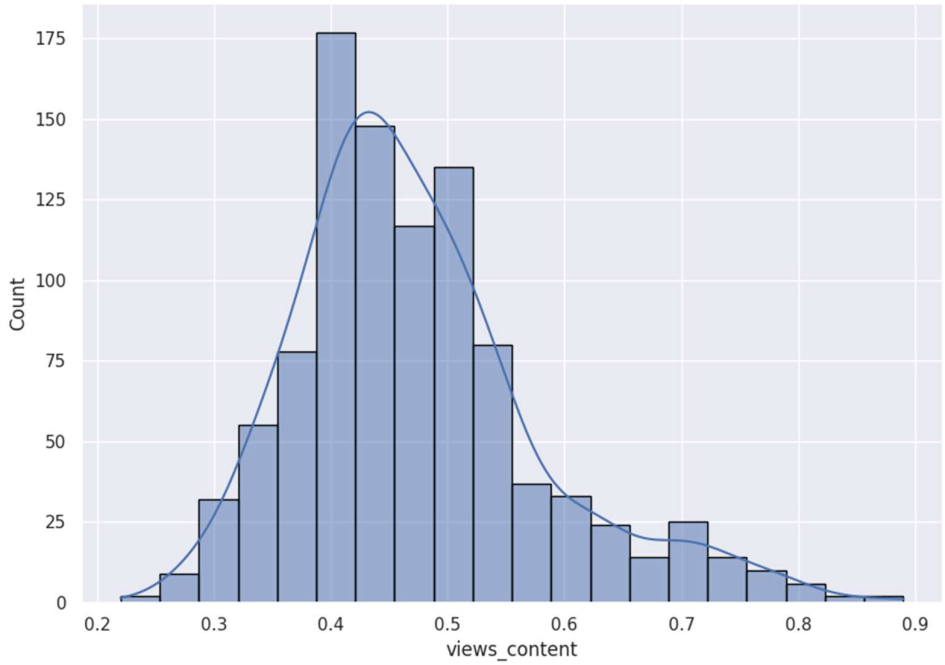


- We can see that number of ad impressions across all content are similar for Spring, Summer and Winter season
- Fall has the least impact on ad impressions

1.5 Answers to the Key Questions

1.5.1 What does the distribution of content views look like?

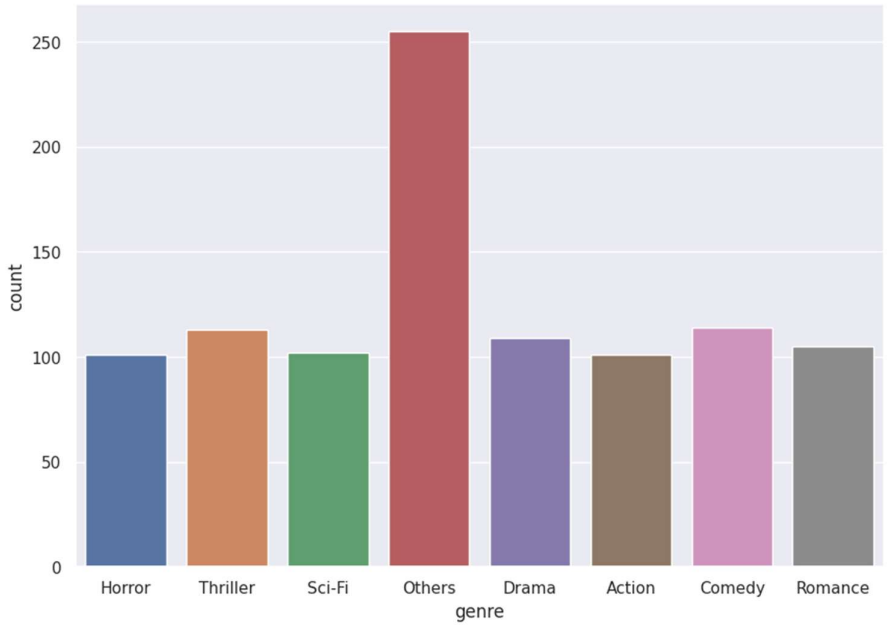
Figure 16 - Histogram of 'views_content'



- The Distribution of content views is almost normal with a slight right skew

1.5.2 What does the distribution of genres look like?

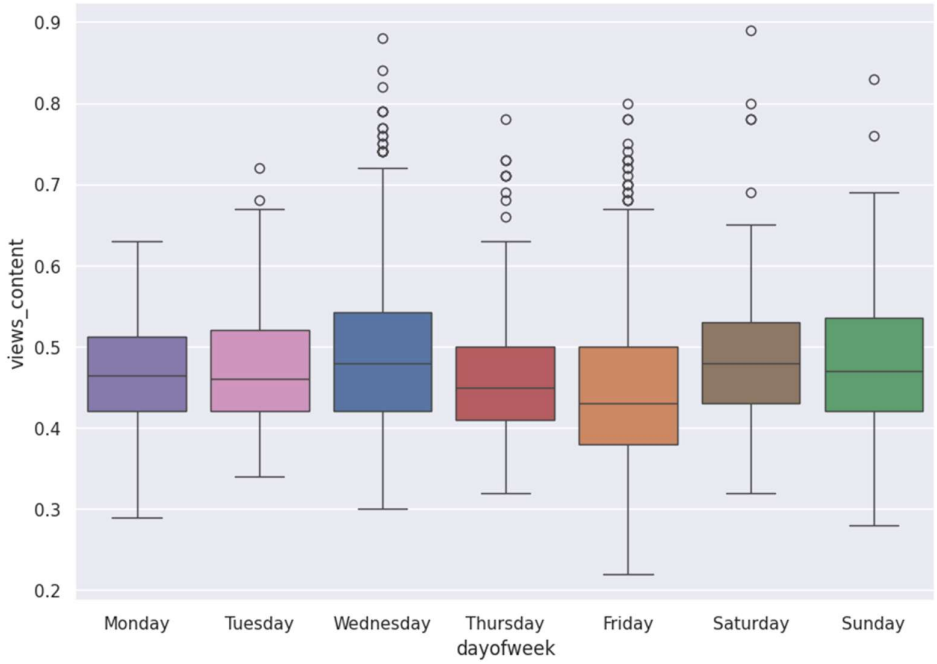
Figure 17 - Countplot of 'genre'



- Comedy' and 'Thriller' are the most released content in the platform followed by Drama and Romance

1.5.3 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?

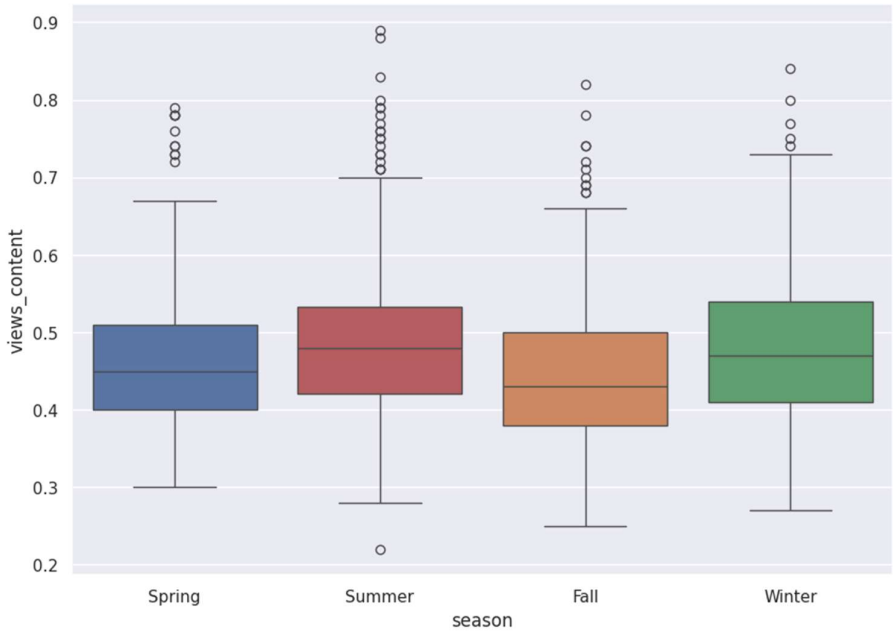
Figure 18 - Boxplot of 'dayofweek' against 'views_content'



- The box plot shows that ‘Wednesday’ and ‘Saturday’ are the most popular days for viewership
- Median views for ‘Monday’, ‘Tuesday’, ‘Thursday’ and ‘Sunday’ are almost similar

1.5.4 How does the viewership vary with the season of release?

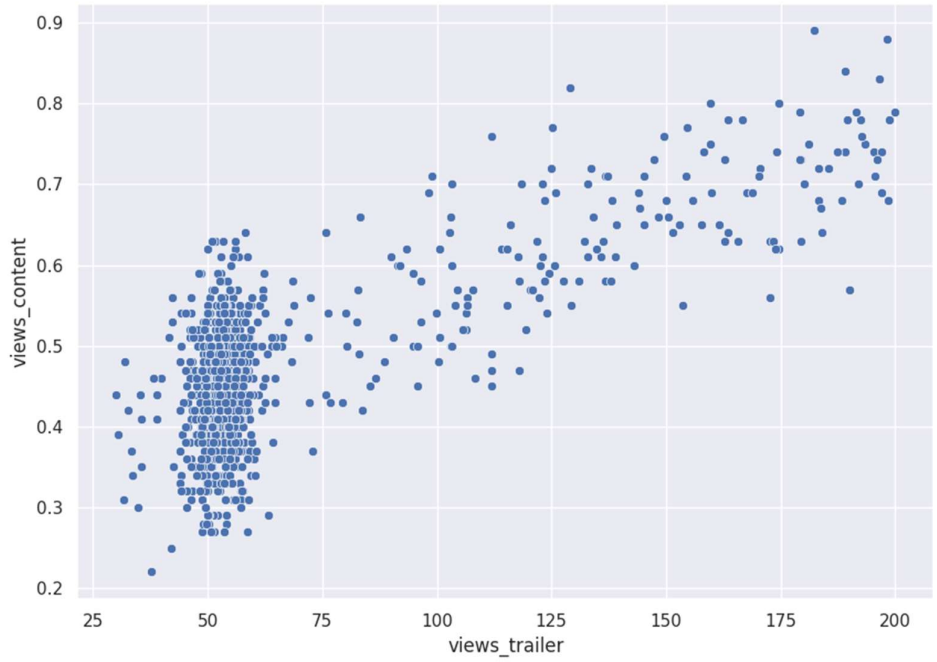
Figure 19 - Boxplot of 'season' against 'views_content'



- ‘Summer is the most popular season for first day content views in the platform
- ‘Winter’ season also generates a lot of first day views for the platform
- ‘Fall’ generates lower amount of first day content views

1.5.5 What is the correlation between trailer views and content views?

Figure 20 - Scatterplot of 'views_trailer'



- The scatter plot shows the 'views_content' and 'views_trailer' are positively correlated with each other, indicating that there are more first day views if there are more trailer views.

1.6 Insights Based on EDA

- Releasing content during summer season has the highest impact for increasing first day views. Summer holidays may contribute large number of views in the platform.
- When trailers are viewed more this in turn increases the number of content views in the platform.
- Wednesday and Saturday are the days which generate the highest number of views for content on the platform.
- Viewership for new content in the platform is more when there is no major sports event during the day of release.
- During Winter season, Action and Sci-Fi genre are popular for first day views in the platform
- During Summer season Thriller and Comedy are the most popular for views

DATA PREPROCESSING

2.1 Duplicate Values Treatment

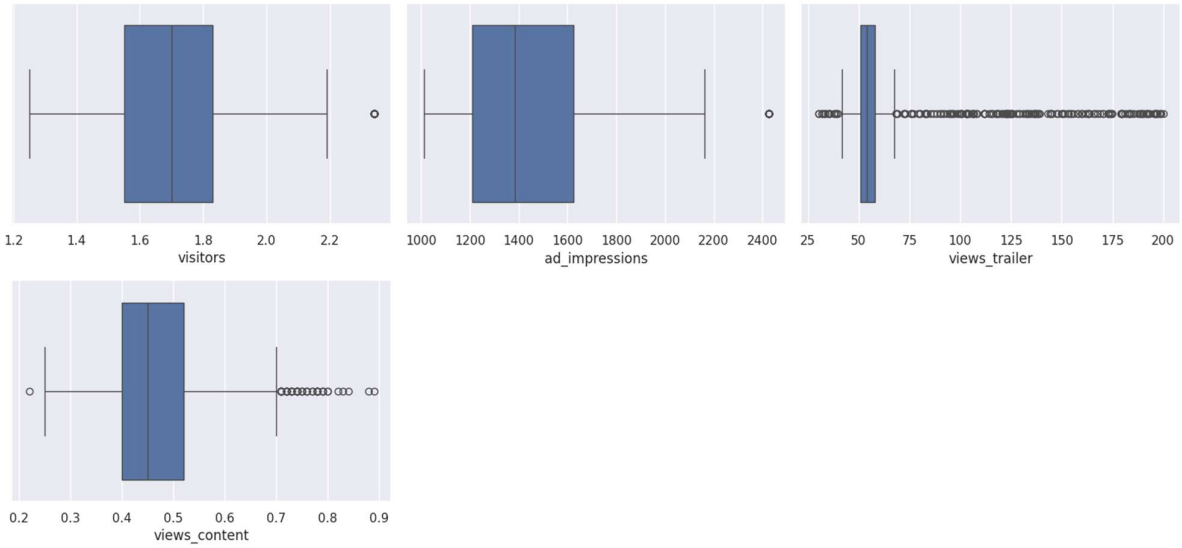
There are no duplicate values in the dataset therefore no treatment is required

2.2 Missing Values Treatment

There are no missing values in the dataset therefore no treatment is required.

2.3 Outlier Treatment

Figure 21 - Boxplot of all numerical variables



There are few outliers present in ‘visitors’, ‘ad_impressions’, ‘views_trailer’ and ‘views_content’. These outliers are genuine values and so we will not do any treatment for these outliers and will be considered as such.

2.4 Feature Engineering

After a thorough analysis of the dataset, currently there are no opportunities for new features that can be added which would affect the model’s performance significantly. Existing features in the dataset will be considered

The ‘major_sports_event’ column contained data with ‘0’ and ‘1’. The values have been replaced with ‘no’ and ‘yes’ respectively automatically converting to a categorical variable.

2.5 Data Preparation for Modeling

We want to determine the driving factors for first-day viewership. Before building the model all categorical features will be encoded. The data will split into training and test data. The model performance will be built on the training data and evaluated on the test data.

We will split the data in 70:30 ratio. 70% of the data will be used as training data, 30% of the data will be used as testing data.

MODEL BUILDING – LINEAR REGRESSION

Figure 22 – First OLS Regression Summary

OLS Regression Results

Dep. Variable:	views_content	R-squared:	0.792
Model:	OLS	Adj. R-squared:	0.785
Method:	Least Squares	F-statistic:	129.0
Date:	Sat, 05 Oct 2024	Prob (F-statistic):	1.32e-215
Time:	18:29:58	Log-Likelihood:	1124.6
No. Observations:	700	AIC:	-2207.
Df Residuals:	679	BIC:	-2112.
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0602	0.019	3.235	0.001	0.024	0.097
visitors	0.1295	0.008	16.398	0.000	0.114	0.145
ad_impressions	3.623e-06	6.58e-06	0.551	0.582	-9.3e-06	1.65e-05
views_trailer	0.0023	5.52e-05	42.193	0.000	0.002	0.002
major_sports_event_yes	-0.0603	0.004	-15.284	0.000	-0.068	-0.053
genre_Comedy	0.0094	0.008	1.172	0.241	-0.006	0.025
genre_Drama	0.0126	0.008	1.554	0.121	-0.003	0.029
genre_Horror	0.0099	0.008	1.207	0.228	-0.006	0.026
genre_Others	0.0063	0.007	0.897	0.370	-0.008	0.020
genre_Romance	0.0006	0.008	0.065	0.948	-0.016	0.017
genre_Sci-Fi	0.0131	0.008	1.599	0.110	-0.003	0.029
genre_Thriller	0.0087	0.008	1.079	0.281	-0.007	0.025
dayofweek_Monday	0.0337	0.012	2.848	0.005	0.010	0.057
dayofweek_Saturday	0.0579	0.007	8.094	0.000	0.044	0.072
dayofweek_Sunday	0.0363	0.008	4.639	0.000	0.021	0.052
dayofweek_Thursday	0.0173	0.007	2.558	0.011	0.004	0.031
dayofweek_Tuesday	0.0228	0.014	1.665	0.096	-0.004	0.050
dayofweek_Wednesday	0.0474	0.004	10.549	0.000	0.039	0.056
season_Spring	0.0226	0.005	4.224	0.000	0.012	0.033
season_Summer	0.0442	0.005	8.111	0.000	0.034	0.055
season_Winter	0.0272	0.005	5.096	0.000	0.017	0.038

Omnibus:	3.850	Durbin-Watson:	2.004
Prob(Omnibus):	0.146	Jarque-Bera (JB):	3.722
Skew:	0.143	Prob(JB):	0.156
Kurtosis:	3.215	Cond. No.	1.67e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.67e+04. This might indicate that there are
strong multicollinearity or other numerical problems.

Interpreting the Regression Results:

Adjusted. R-squared:

Adjusted R-squared reflects the fit of the model, The higher value generally indicates a better fit, assuming certain conditions are met. In this case, the value for adj. R-squared is **0.785**, which is good.

Const coefficient:

It is the Y-intercept. It means that if all the predictor variable coefficients are zero, then the expected output would be equal to the *const* coefficient. In this case, the value for const coefficient is **0.0602**

3.1 Model Performance Check

Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.04853	0.038197	0.791616	0.785162	8.55644

Test Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.050603	0.040782	0.766447	0.748804	9.030464

Observations:

- The training R Squared is 0.79, so the model is not underfitting
- The train and test RMSE and MAE are comparable, so the model is not overfitting either
- MAE suggests the model can predict first day viewership within a mean error of 0.040 on the test data
- MAPE of 9.03 on the test data means that we are able to predict within 9.03% of the first day views.

TESTING THE ASSUMPTIONS OF LINEAR REGRESSION MODEL

We will be checking the following Linear Regression assumptions:

1. No Multicollinearity
2. Linearity of variables
3. Independence of error terms
4. Normality of error terms
5. No Heteroscedasticity

4.1 Test for Multicollinearity

Multicollinearity occurs when predictor variables in a regression model are correlated. This correlation is a problem because predictor variables should be independent. If the correlation between variables is high, it can cause problems when we fit the model and interpret the results. When we have multicollinearity in the linear model, the coefficients that the model suggests are unreliable.

The VIF (Variance Inflation Factor) is a method used to test for multicollinearity. VIF measures the inflation in the variances of the regression parameter estimates due to collinearities that exist among the predictors. It is a measure of how much the variance of the estimated regression coefficient βk is "inflated" by the existence of correlation among the predictor variables in the model.

If VIF value is between 1 and 5, then there is low multicollinearity. if the VIF value is between 5 and 10, there is moderate multicollinearity. If the VIF value is exceeding 10 then there is high multicollinearity

Figure 23 - VIF values of all variables

	feature	VIF
0	const	99.679317
1	visitors	1.027837
2	ad_impressions	1.029390
3	views_trailer	1.023551
4	major_sports_event_yes	1.065689
5	genre_Comedy	1.917635
6	genre_Drama	1.926699
7	genre_Horror	1.904460
8	genre_Others	2.573779
9	genre_Romance	1.753525
10	genre_Sci-Fi	1.863473
11	genre_Thriller	1.921001
12	dayofweek_Monday	1.063551
13	dayofweek_Saturday	1.155744
14	dayofweek_Sunday	1.150409
15	dayofweek_Thursday	1.169870
16	dayofweek_Tuesday	1.062793
17	dayofweek_Wednesday	1.315231
18	season_Spring	1.541591
19	season_Summer	1.568240
20	season_Winter	1.570338

Observations:

- All the variables have a VIF value below 5 so we can conclude that there is no multicollinearity.

Dealing with high p-value variables:

The dummy variables in the data which have p-value > 0.05 can be considered as not significant and can be dropped. A loop will be used to ensure the process is efficient and there is no loss in the model efficiency.

The model will be rebuilt with the dropped columns removed.

Figure 24 - Updated OLS Regression Summary

OLS Regression Results

Dep. Variable:	views_content	R-squared:	0.789
Model:	OLS	Adj. R-squared:	0.786
Method:	Least Squares	F-statistic:	233.8
Date:	Sat, 05 Oct 2024	Prob (F-statistic):	7.03e-224
Time:	18:30:01	Log-Likelihood:	1120.2
No. Observations:	700	AIC:	-2216.
Df Residuals:	688	BIC:	-2162.
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0747	0.015	5.110	0.000	0.046	0.103
visitors	0.1291	0.008	16.440	0.000	0.114	0.145
views_trailer	0.0023	5.5e-05	42.414	0.000	0.002	0.002
major_sports_event_yes	-0.0606	0.004	-15.611	0.000	-0.068	-0.053
dayofweek_Monday	0.0321	0.012	2.731	0.006	0.009	0.055
dayofweek_Saturday	0.0570	0.007	8.042	0.000	0.043	0.071
dayofweek_Sunday	0.0344	0.008	4.456	0.000	0.019	0.050
dayofweek_Thursday	0.0154	0.007	2.307	0.021	0.002	0.029
dayofweek_Wednesday	0.0465	0.004	10.532	0.000	0.038	0.055
season_Spring	0.0226	0.005	4.259	0.000	0.012	0.033
season_Summer	0.0434	0.005	8.112	0.000	0.033	0.054
season_Winter	0.0282	0.005	5.362	0.000	0.018	0.039

Omnibus:	3.254	Durbin-Watson:	1.996
Prob(Omnibus):	0.196	Jarque-Bera (JB):	3.077
Skew:	0.139	Prob(JB):	0.215
Kurtosis:	3.168	Cond. No.	662.

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

Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.048841	0.038385	0.788937	0.785251	8.595246

Test Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.051109	0.041299	0.761753	0.751792	9.177097

Observations:

- The adjusted R-Squared is 0.786 which will explain approximately 78.8% of the variance.

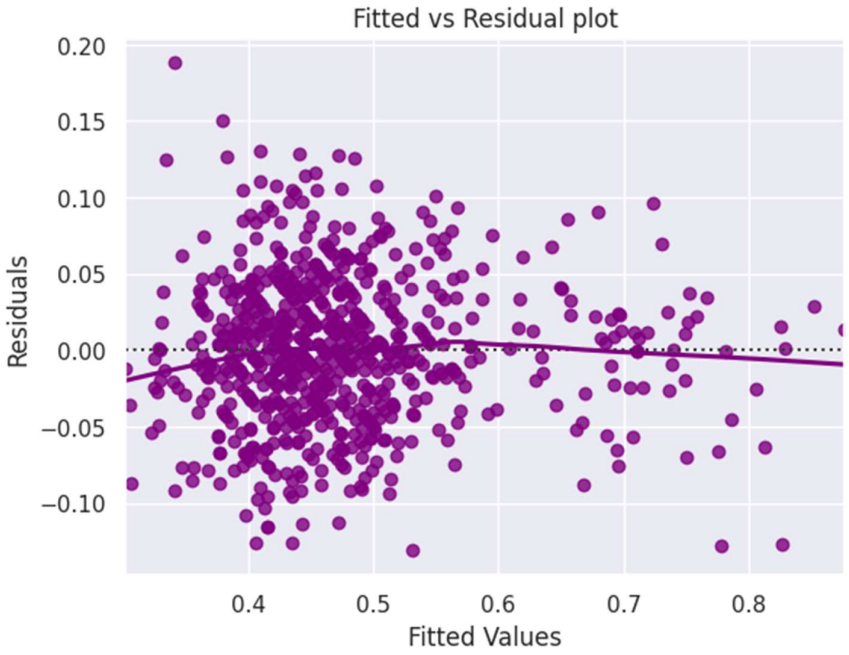
- No feature has a p-value > 0.05 so x_train1 will be considered as the final set of predictor variables and olsmod2 as the final model.
- The previous model's adjusted R-Squared value 0.785, showing that there has been very minimal impact on the model after dropping the variables
- The MAE and RMSE values of the training and testing data are comparable to each other. This shows the model is not overfitting.

4.2 Test for Linearity & Independence

The test for linearity is to show that the predictor variables must have a linear relationship with the dependent variable. The test for independence is needed to show that the variables do not provide information to the other variables i.e all variables are independent of each other.

To test for linearity and independence we use the scatter plot of fitted values vs residual values and check if the chart displays any patterns.

Figure 25 - Scatter plot of Fitted vs Residuals



Observations:

- There is no pattern in the above plot. We can conclude that the assumptions of linearity and independence are satisfied.

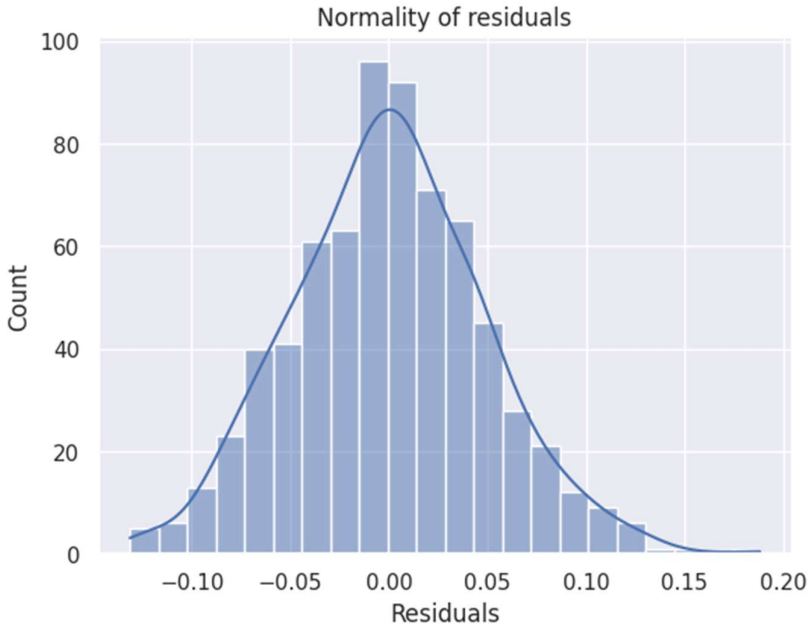
4.3 Test for Normality

The test for normality is to ensure the data is normally distributed so that the confidence interval of the coefficient estimates is not too wide or narrow. We use a histogram as well as a Q-Q plot to test for normality. The Shapiro-Wilk test will also be used by assuming the null and alternate hypothesis.

Null hypothesis: Residuals are normally distributed

Alternate hypothesis: Residuals are not normally distributed

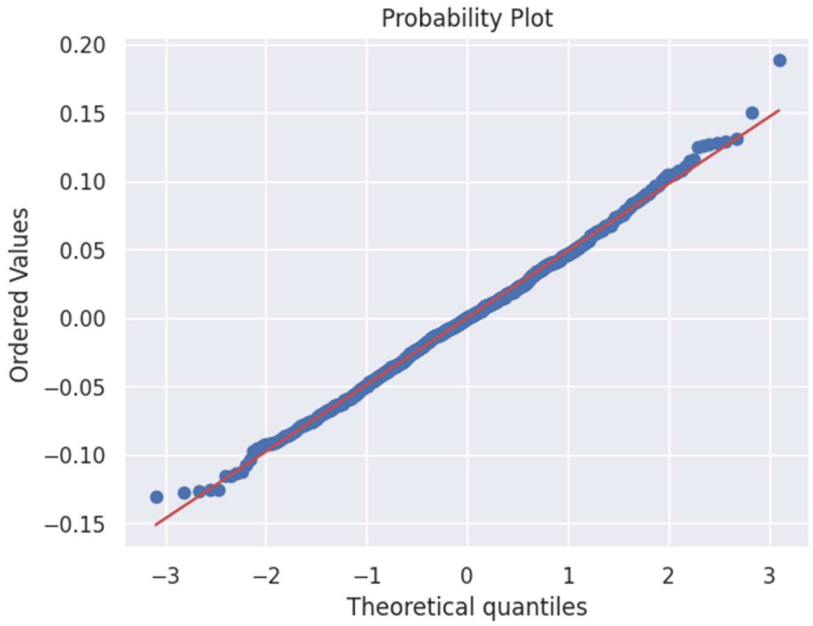
Figure 26 - Histogram of Residuals



Observations:

- There is a bell-shape in the histogram of the residuals.

Figure 27 - Q-Q plot of Residuals



Obeservations:

- We can see that the residuals follow a straight line, except for very few points in the tails

On performing the Shapiro-Wilk test, we can see the p-value is 0.310 which is greater than 0.05. We fail to reject the null hypothesis of the Shapiro Wilk's Test and can conclude that the residuals are normally distributed.

4.4 Test for Homoscedasticity

The test for homoscedasticity is to show that the variance of the residuals is symmetrically distributed across the regression line otherwise it is heteroscedastic. We will use the goldfeldquandt test to check for homoscedasticity and assume the null and alternate hypothesis

Null hypothesis: Residuals are homoscedastic

Alternate hypothesis: Residuals have heteroscedasticity

On performing the test, we receive a p-value of 0.128 which is greater than 0.05 so we have failed to reject the null hypothesis. We can conclude that the residuals are homoscedastic.

4.5 Predictions on Test data

As all the assumptions have been satisfied, we can predict on the test data and observe the result

Figure 28 - Predicted vs Actual on Test Data

	Actual	Predicted
983	0.43	0.434802
194	0.51	0.500314
314	0.48	0.430257
429	0.41	0.492544
267	0.41	0.487034
746	0.68	0.680000
186	0.62	0.595078
964	0.48	0.503909
676	0.42	0.490313
320	0.58	0.560155

Observations:

- We can see that the actual and predicted values are comparable to each other, These results are satisfactory.

MODEL PERFORMANCE EVALUATION

The model performance will be checked as all assumptions of the linear regression model have been satisfied.

Figure 29 - Final OLS Regression Summary

OLS Regression Results

Dep. Variable:	views_content	R-squared:	0.789
Model:	OLS	Adj. R-squared:	0.786
Method:	Least Squares	F-statistic:	233.8
Date:	Sat, 05 Oct 2024	Prob (F-statistic):	7.03e-224
Time:	18:30:03	Log-Likelihood:	1120.2
No. Observations:	700	AIC:	-2216.
Df Residuals:	688	BIC:	-2162.
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0747	0.015	5.110	0.000	0.046	0.103
visitors	0.1291	0.008	16.440	0.000	0.114	0.145
views_trailer	0.0023	5.5e-05	42.414	0.000	0.002	0.002
major_sports_event_yes	-0.0606	0.004	-15.611	0.000	-0.068	-0.053
dayofweek_Monday	0.0321	0.012	2.731	0.006	0.009	0.055
dayofweek_Saturday	0.0570	0.007	8.042	0.000	0.043	0.071
dayofweek_Sunday	0.0344	0.008	4.456	0.000	0.019	0.050
dayofweek_Thursday	0.0154	0.007	2.307	0.021	0.002	0.029
dayofweek_Wednesday	0.0465	0.004	10.532	0.000	0.038	0.055
season_Spring	0.0226	0.005	4.259	0.000	0.012	0.033
season_Summer	0.0434	0.005	8.112	0.000	0.033	0.054
season_Winter	0.0282	0.005	5.362	0.000	0.018	0.039

Omnibus:	3.254	Durbin-Watson:	1.996
Prob(Omnibus):	0.196	Jarque-Bera (JB):	3.077
Skew:	0.139	Prob(JB):	0.215
Kurtosis:	3.168	Cond. No.	662.

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

Training Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.048841	0.038385	0.788937	0.785251	8.595246

Test Performance

	RMSE	MAE	R-squared	Adj. R-squared	MAPE
0	0.051109	0.041299	0.761753	0.751792	9.177097

Observations:

- The model will able to explain approximately 78% of the variation in the data
- The MAPE of the test data suggests we can predict within 9.1% of the first day viewers
- The RMSE and MAE for the training and test data are low and comparable, this shows there is no overfitting.
- The final model (olsmodel_final) can be considered as good for prediction.

ACTIONABLE INSIGHTS & RECOMMENDATIONS

- For each additional visitor, content views increase by approximately 0.129. This suggests that increasing traffic can lead to more views, so focus on strategies to boost visitor numbers.
- Each additional view of the trailer results in about 0.0023 additional content views. This implies that promoting trailers effectively could increase first day views.
- During major sports events, content views decrease by approximately 0.0606. It is preferred to avoid timing content releases around major sports events.
- All seasons have a positive impact, with summer having the strongest impact.
- Releasing content on all days have a positive impact, with the biggest impact being on Saturdays and Wednesday. Prioritizing these days to release content will lead to more views.
- New content should be targeted for a ‘Summer’ release as it has the strongest impact on content views in the platform.