The tribe

Youtube viewership

May 21,2020

Table of contents

Overview

Tribe Digital Agency is a marketing agency that has a significant following on YouTube. It seeks to leverage this platform as a marketing tool, by uploading videos. Its main objective is to identify when it's optimum to publish videos so as to increase viewership, brand awareness, social engagement and business development.

Project Plan

The project plan is highlighted below.

Tasks	Deadline	Assignee
Project planning	Day 1	Wambui Kuria
Data cleaning	Day 1	Wambui Kuria
Data analysis	Day 2	Wambui Kuria
Report writing	Day 3	Wambui Kuria
Visualisation	Day 3	Wambui Kuria
Power point	Day 3	Wambui Kuria

Cost and Benefit analysis

The benefits of this analysis were more than the costs as the members were working remotely and the datasets were readily available at Kaggle at no cost.

Business Understanding

Tribe Digital Agency herein "Tribe" is a marketing agency that uses digital platforms to promote its clients' products. It predominantly uses Youtube due to its nature of subscribers thus recording numerous video views . Tribe seeks to understand when it is best to upload videos based on when YouTube users are most active. This analysis will help the business to determine when their videos can attract the most number of views, likes and comments. This way, Tribe will maximize the number of views a video gets thus reaching out to a bigger audience and possibly increasing their brand awareness.

Business objective

To analyze the YouTube data in order to gain insights about the trending videos and the best days and time to upload videos in order to reach a wider audience and hence improve business development efforts.

Business success criteria

Research Specifications

- Personnel: Tribe members
- Data: Our dataset was from kaggle.Click here to view the data set.
- Notebook: The link to our notebook [link].
- Risks: There were no immediate risks involved.

Requirement, Assumptions and Constraints

Requirements

Every team member was aligned on working on the project

Assumption

The dataset is an extract of a real word dataset and there were no assumptions made.

Constraints

It was not possible to get Youtube views for the current year 2020.

Data Mining

To determine the:

- Relationship between views and likes, views and comments and title length and views
- Correlation between the number of likes and views and views and comments.
- Calculation of the rates of comments, views and likes.

Data Mining Success Criteria

Our data mining success criteria is to recognise the best day(s) to upload videos on YouTube.

Data Understanding

Our dataset is from kaggle and was collected using YouTube API. Click here to access the dataset.The dataset was collected between 2017 and 2018. It has 40949 rows and 16 columns. The fields are category_id, video title, channel title, publish time, tags, views, likes and dislikes, description and comment count.

The data was readily available in a csv file format and no data mining or scraping techniques were used to get the data.

We explored the data and verified it for our analysis.

Data Preparation

The steps taken in the preparation are as follows:

1. **Loading the dataset**

   The data was in csv format and was loaded using pandas for cleaning purposes.

2. **Data Cleaning**

   The data was cleaned using the following steps:

   - Checking for consistency: This involved checking for duplicates which were dropped
   - Accuracy: We created new columns for date and time using python date time object
   - Completeness : We filled the empty cells in the description column randomly
   - Outliers: Identifying outliers.

3. Exporting the dataset

We exported the cleaned dataset for further analysis and visualisation.

Data Analysis

Data analysis on this project involved:

1. Exploratory data analysis (EDA) which included univariate, bivariate and hypothesis testing.
2. Data visualizations using tableau.

3. Exploratory Analysis

Majority of the attributes are binomial hence permitting for generalized conclusions on the analysis. There were also numeric variables that allowed us to draw more insights on the dataset collected in 2017 and 2018. Some of the analysis done are as follows:

- The dataset was collected in 2017 and 2018 with 77% of it in 2018 and 23% in 2017.It shows that the most popular day to post a video is Friday.
- The number of views and likes are positively correlated, so as one increases, the other increases and vice versa.
- Views and likes are positively correlated(0.85)
- The title length has a normal distribution where most videos have a title length between 33 and 60 characters.

4. Hypothesis Testing

Tribe seeks to investigate a claim that videos released on the weekends have greater views on average than the ones released on weekdays. Our hypothesis are as follows:

- Ho = Videos released over the weekend have greater views on average than the ones released on weekdays
- Ha = There's no difference in viewership between videos released on weekdays and weekends.

Our hypothesis is important as it will help us understand and make informed decisions on which is the best day to place our ads. This will in turn inform our digital marketing strategy which will then be communicated to the digital marketing personnel.

i) Hypothesis Testing Procedure

- First, we will create a sample of the dataset. In this case, we chose *n = 200*.This is done using a simple random sampling method.
- The significance level, *a= 0.05*; confidence interval *c = 95%*.
- After calculating the sample n, we will calculate the test statistic using z-test which gives us the P-Value.
- Finally, we compare the P-Value with the significance level.

If:

P-value < 0.05, the null hypothesis will be **accepted.**

else:

P-Value > 0.05, the null hypothesis **rejected**

ii) Hypothesis Testing Results

The test returned a P Value of 1.57884404
Since alpha is 0.05, the confidence level will be 95%.
Hence, it is not statistically significant and indicates weak evidence against the null hypothesis.

We, therefore, neither reject the null hypothesis nor do we accept it.

Data visualisation

We did further analysis between attributes via data visualizations with Tableau,so as to make our analysis more informative.

Summary and Conclusions

Here are some of the conclusions we arrived at:

- We analyzed a dataset that contains information about YouTube trending videos. The dataset was collected in 2017 and 2018. It contains 40949 video entries.
- Trending videos that have most views have a title length between 33 and 60 characters approximately.
- The delimiters - and | were common in trending video titles.
- The words 'Official', 'Video', 'Trailer', 'How', and 'Video' were common also in trending video titles.
- There is a strong positive correlation between the number of views and the number of likes of trending videos: As one of them increases, the other increases.
- There is a moderate positive correlation also between the number of likes and the number of comments, and a slightly weaker one between the number of dislikes and the number of comments.
- The p-value is greater than the level of significance, so I neither reject the null hypothesis nor accept the alternative hypothesis, there's no difference in viewership between videos released on weekdays and weekends.

Recommendation

- Friday is the most suitable day of the week to publish a video. Periodic videos should be published on Friday.
- Most viewers are online at 1600hours onwards on average. Therefore, it is wise to upload the video between 1400hours and 1600hours
- Most viewers on youtube watch and view music videos. Therefore, we should recommend that our ads are tagged on music videos by youtube especially for our premium clients
- We should also seek to market our products using a musical theme.
- We should use SEOs that are common in trending videos like "official".

References

1. Notebook colab

   https://colab.research.google.com/drive/1rQ6lFZMAqdIPbKnrJWP0vZlzHiVdVpv5?usp=sharing

2. Tableau link

   https://public.tableau.com/profile/wambui.kuria#!/vizhome/Draftfour/Firstdraft