Student Id: 22k-5336

YouTube Data Analysis

Available Data: approx. 800 million videos among **37 million channels** on the platform

Problem Statement

People around the world from the last decade have shifted towards visual content from the last 100 years practice of text reading and audio listening. Business of all sectors are investing huge number of bucks for promoting brands and to increase revenue. Data analyst can provide insights by using the YouTube Data API which can help any business in ROI.

Following statistics highlights of YouTube platform is the evidence of people habits around the world.

- *2.6 billion active users worldwide. [World population: 7.888 billion]*
- *The social media platform with the highest penetration.*
- *Second most popular search engine after Google*
- *Localized in more than 100 countries and available in 80 languages.*
- *1 billion hours every day generate billions of views.*
- *62% of businesses use YouTube as a channel to post video content.*
- *90% of people say they discover new brands or products on YouTube.*
- *500 hours of video uploaded to YouTube every minute.*

https://www.oberlo.com/blog/youtube-statistics

Description of Data

I have selected the data from YouTube of 7 different channels of Pakistani Beverages

	Channel_name	Channel_Id	Subscribers	Views	Total_Videos	Playlist_Id
0	Fanta Pakistan	UC7-sLqTDHslac63eYnNSArQ	23100	20739352	32	UU7-sLqTDHslac63eYnNSArQ
1	Sprite Pakistan	UCX5Z2XnBvyZR-hnOvaqx6BQ	51200	95802845	124	UUX5Z2XnBvyZR-hnOvaqx6BQ
2	7up Pakistan	UCCnBhAAU9Acv0wQ-NbC41iw	18700	83150681	61	UUCnBhAAU9Acv0wQ-NbC41iw
3	Mountain Dew PK	UCs367hYQ8u91NOwF8kcYPKw	47300	68052354	121	UUs367hYQ8u91NOwF8kcYPKw
4	Pepsi Pakistan	UCQXd4P5-_QDdK9nnW7GT7_g	455000	253524945	556	UUQXd4P5-_QDdK9nnW7GT7_g
5	Coca-Cola Pakistan	UCSpu6fELJgTYyKuMblHtnig	153000	132763156	233	UUSpu6fELJgTYyKuMblHtnig
6	Mirinda Pakistan	UCs3OjUTj_mHdFHRKHlu_tDw	10000	21723452	3	UUs3OjUTj_mHdFHRKHlu_tDw

Data Pipeline

Data Collection: Data of channels collected using youtube data api which provide access on the youtube data using api key generated from https://console.cloud.google.com/apis. Channel Ids extracted using web scrapping tool available online https://commentpicker.com/

Data Preparation: Projection is used to prepare data collected from channels to filter features which are important for the scope in this project for analysis.

Data Visualization: Following python libraries are used for visualization of the data.

import seaborn as sns

import matplotlib.pyplot as plt

Insights

1. Pepsi Pakistan is more active in digital content than others.
2. Fanta Pakistan and Mirinda Pakistan both not active in digital content so far, and having less number of views on content uploaded on these channels
3. Coca Cola Pakistan is competing with Pepsi Pakistan to get number of subscribers and views more.
4. Sprite Pakisan, 7up Pakistan, and Mountain Dew Pakistan channels have very less difference in number of views on content uploaded on respective channels.
5. Mirinda has uploaded content in video category *People and Blogging,* category id 22
6. Mirinda Pakistan uploaded same content 3 times in the last 3 years while the channel created in 12 Oct 2017, and that's why only 10k subscribers.
7. Pepsi Pakistan is continuously creating content of Music category
8. Coca Cola Pakistan is creating content of Entertainment category
9. Mountain Dew Pakistan is more focused on gaming category and adventure while their video category is gaming and entertainment.
10. YouTube video categories are listed in cell 55 in notebook.
11. First video uploaded on YouTube with caption containing Pepsi

Refer cell 96 in notebook

U-RHMyjr4hY	jerry yan's pepsi commmercial	2006-01-08T19:25:30Z	UC44HNwWHmnO5jcGTpferrFw	f4ever

jerry yan's pepsi commmercial

12. First video uploaded on YouTube with caption containing Coca Cola

9AgwBiYEMp0	Coca Cola - Despedido	2005-12-28T00:50:46Z	UCx8scA19exFvDiitaJfHLcw	snork512

https://www.youtube.com/watch?v=9AgwBiYEMp0

13. First video uploaded on YouTube with caption containing 7up

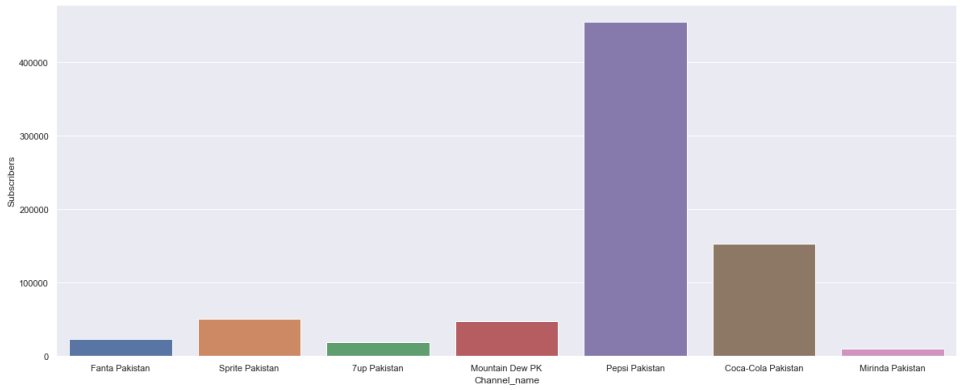
FxtHxfZEkaQ	1980's 7-Up Christmas Commercial w/Santa C...	2006-02-12T09:45:49Z	UCw24HDBD7ssZzZ0nUCZdfaA	Jenna Van Wyler

14. First video uploaded on YouTube with caption containing sprite

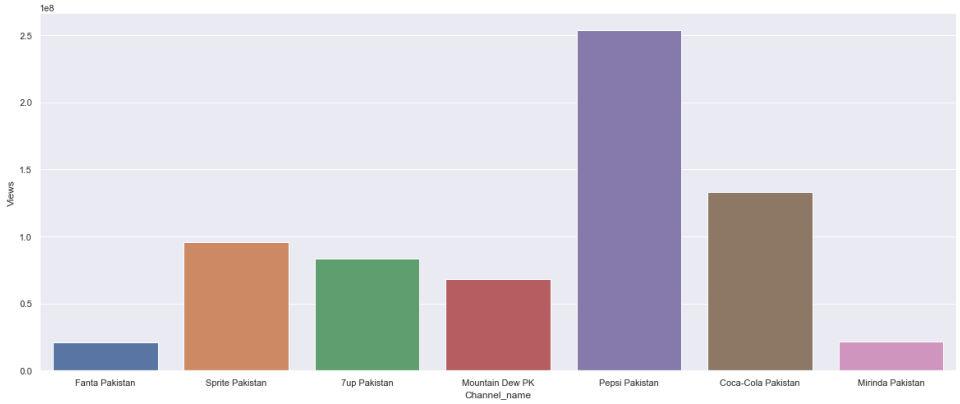
jE1LADaXsw0	sprite basketball court commercial	2006-03-23T21:09:23Z	UCqg8jtVOuvcO8QJpSzhrv8Q	Oscar Vega

EDA Graphs

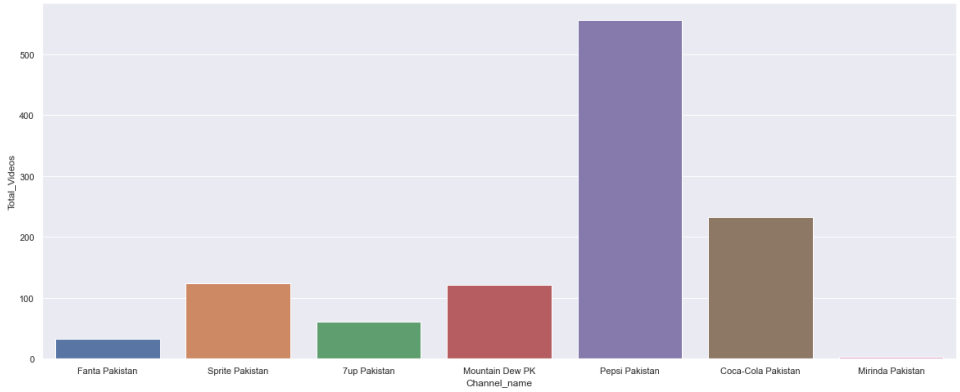
Subscribers



Views



Videos Count



Output of ML Algorithm

1. Collaborative Filtering Algorithm
2. Ethical Algorithm

Final Note:

- The code in this project is not limited to the channels selected in this report for analysis. I can perform analysis on any channel(s) by replacing channel Id in notebook cell 51
- I can also find further insights by analysis of most like/dislike video comments posted under comment section.

xxxx