PROJECT NAME: TRENDING VIDEOS ON YOUTUBE

As a final project for Data Visualization with Tableau Specialization project, I would like to examine trends on YouTube by analyzing views, likes, dislikes and number of comments for 10 countries. For the sake of project, let's assume that I work for a marketing company as marketing analyst and we are currently working with a client which is an online language learning company. They would like to expand their company over new countries and consider about advertising their product on YouTube. The company would like to get key insights regarding YouTube trends according to countries and categories. To this purpose, I am supposed to prepare a presentation which reveals the trends on YouTube by countries and by categories. The audience of the presentation are the market growth team of the client and the project team from my company. There will be four people and almost all of them familiar with data and analytics.

PROJECT PROPOSAL

Why am I working on this project? What are the business questions?

I would like to examine the data to make meaningful insights regarding YouTube trends. Today, it is obvious that the social media channels are more attractive for marketing and advertising goods and services. The mainstream media has been less preferred by advertisers. In that sense, understanding the trends in social media is crucial to keep track the world's trends. With this dataset, I would like to reveal the common pattern on YouTube in each country and present the results in a most effective way. The main objectives of this project are:

- What are the trends (KPIs: views, likes, dislikes, number of comments) by categories for each country?
- How the most viewed, liked, disliked videos vary by countries and by categories?
- Is the publishing time important for the viewing numbers?
- Is there a common pattern between the most trending videos within the country and across the countries?

Who will be the audience?

Let's say, I am working for a marketing company focused on online marketing on YouTube. We are working with a client which is an online language learning platform and they would like to understand the dynamics on YouTube for advertisement of their product. My audience will be the market growth team of client company (growth lead, junior and senior specialists) and the manager of my team. There will be four people and they are familiar with the data and the market conditions. This presentation will be preliminary draft (in a more explanatory way). A final presentation will be framed based on the feedbacks during this presentation (The final presentation is not subjected to this project).

Why have I selected this dataset? What are the features of the dataset?

The dataset is selected since the YouTube data is attractive for not only marketing issues but also many other topics like politics, trendsetting, sociology, etc. Social media and its impacts on our lives especially on millennials is quite significant in many manners. For this project, I will handle this data to understand dynamics behind what makes a video on YouTube trending?

The "YouTube Trending Videos Statistics" dataset is open and available on Kaggle website. The data available for 10 countries in separate files. Countries included are United States of America, United

Kingdom, Canada, France, Germany, India, Mexico, Japan, South Korea, and Russia. There are 16 columns in dataset however, I am going to utilize ten of them namely: *video_id, trending_date, title, channel_title, category_id, publish_time, views, likes, dislikes,* and *comment_count*. I will create new fields like category_name and country_name. I also need to unify all files in a single file to be able to make comparison between countries. To this purpose, I need to standardize time accordingly the time zones of each countries.

The data is daily basis. It includes seven months, from 14.11.2017 to 14.06.2018.

How will I present the outcomes?

The presentation is supposed to be a preliminary and explanatory version of a final presentation. Therefore data, results and all elements in the presentation will be open to discuss. I am going to make an interactive story point presentation to be able to show the all outcomes easily. In addition to it, I am going to prepare the single-page summary handouts that summarize the findings regarding my analysis. It will be more visualized.

Which problems/challenges might I probably encounter during the project?

I think that compilation and standardization of data will be the main challenge of this project. There are 10 different big csv files needed to be compound and standardized. Identifying KPIs is not a big challenge for this dataset since it already has KPIs such as views, likes, dislikes, etc. Creating an interactive dashboard for the final presentation without overwhelmed the audience with numbers could be also a challenging part of this project. Moreover, since the audience consists of various segments and personas, it might be difficult to make them be on the same page.

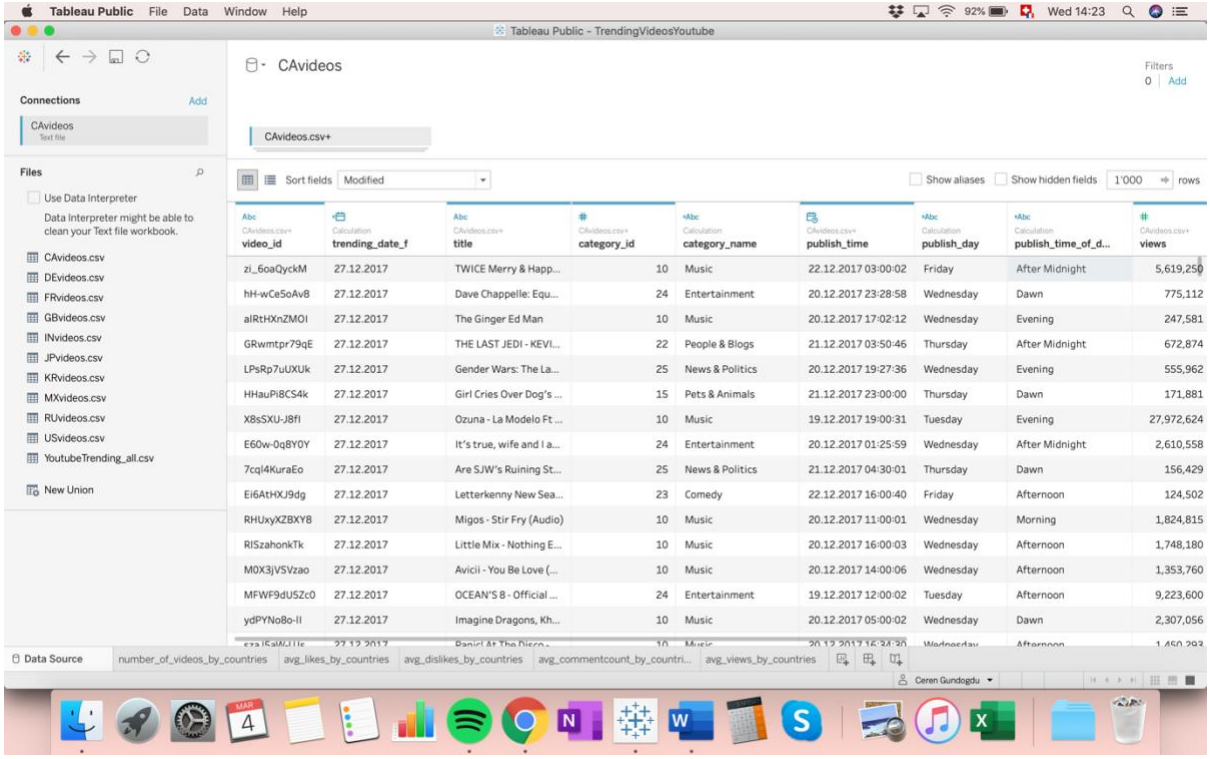
INFORMATION ON DATASET AND PREPARATION OF DATA FOR THE ANALYSIS

Write your narrative describing any data cleanliness issues found and your process for correcting those issues in the space provided.

There are ten different files regarding trending videos on Youtube. However, since using all of them is too big and saving the notebook takes a very long time, I decided to analyze just 5 countries' data. So, I included this project Canada, US, UK, France, and Germany and make them a table by using 'union'. I especially focused on the countries which have relatively high English-speaking rate (Germany:56%, France: 39%; the official language is already English in US, UK, Canada) to be able to make meaningful comparisons between the same trending videos in different countries.

I created five new fields. One of them shows the name of countries, and the second one is for pointing out the category names. Regarding the dates, I created 'publish_day' column to see the weekday of publishing videos and 'publish_time_of_day' column to get publishing time of the day such as morning, evening, afternoon, etc. I divided the day into seven parts to be able to make an analysis regarding the publishing time of the video. Moreover, 'trending_date' was in string format. I converted it to date formatting.

Besides those, the dataset is clean and no need to clean further.



EXPLORATORY ANALYSIS

Why are these KPIs key to answering your business question or achieving your project goals? What specific insight did these KPIs or worksheets provide you?

The KPIs for the business question of this project are the number of views and the interaction with the video such as likes, dislikes, and the number of comments. I analyzed KPIs based on categories and countries. Moreover, I have prepared a dashboard to reveal the number of views by weekdays and the time of the day with harmonization with each country's time zones. Based on the results of the analysis, it can be said that 'music' videos are the most viewed videos overall. "Entertainment" and "People&Blog" videos follow "Music" videos. Although the "Music" videos have higher views, the unique number of these videos are much less than "Entertainment" and "People&Blog" videos.

How does your dashboard or worksheet(s) exemplify use of at least one pre-attentive attribute?

In all three dashboards, I have utilized coloring and sizing as pre-attentive attributes. I also utilized filtering actions in all dashboards to link and communicate all graphs with each other. I avoided putting too much information, but I put annotation when it is necessary. Since it will be an exploratory analysis, I tried to have a high user control with high interactivity.

How does your dashboard or worksheet(s) exemplify use of at least one of the following KPI principles?

Based on our learning through the specialization, there are three components of a KPI: being crucial, being measurable and having some calculation. The KPIs of this project are unique values of views (i.e. how many times a video is watched), unique values of interactions such as likes, dislikes, and comments. There are crucial indicators for YouTube videos and they are measurable. I needed some calculations to be able to get unique values of indicators. Since the views increase cumulatively day by day and one video can stay trending more than one day, I needed to eliminate duplicated records by selecting the max views or likes, etc. I have created new calculated fields by using this formula ({ FIXED [video_id_unique] : max([views_notunique])} for each indicator.

Do these worksheets meet the needs of your intended audience as-is? How might they need to be changed before ultimately being presented?

I arranged all worksheets and created three interactive dashboards. Dashboards are easy to understand and allows user control with high interactivity. Since the audience has the ability to play with data, those interactive dashboards meet their need. Before, ultimately presenting it, I might add an introduction and conclusion parts and organize all of them as story point format.

Also, include discuss your thoughts on the exploratory process in general. How long did you spend exploring your data? What did you learn about the process? How many visualizations did you end up having to create before honing in on the best one to represent your project goal? What surprised you and what did you find difficult?

The exploratory analysis took more than 8 hours because figuring out how to create KPIs and selecting the max of the indicator to eliminate duplicated data were the challenging parts. I also spent considerable time designing dashboards. Moreover, since the dataset is pretty big, the saving process always took so much time and even sometimes it couldn't be saved and I had to shut down the computer without saving. And I spent time performing all I did previously once again.

PRESENTATION CHOICE

Submit your written explanation for why you chose to make a multi-frame data story or a single-frame visualization/infographic over the other option.

My audience will be the market growth team of the client company (growth lead, junior and senior specialists) and the manager of my team. There will be four people and they are familiar with the data and the market conditions. This presentation will be the preliminary draft (in a more explanatory way). A final presentation will be framed based on the feedback during this presentation (The final presentation is not subjected to this project). Therefore, data, results and all elements in the presentation will be open to discussing. That's why, I made an interactive story point presentation to be able to show all outcomes easily.

REFLECTION

I prefer to present my work as a multi-frame data story points since the analysis is comprehensive and the audience are familiar with the numbers. The purpose of the presentation is to explore data together with audience with my guidance. That's why I put high interactivity to my dashboards to be able to give control to users. In each dashboard, I firstly give the big pictures to make easier to

see general trends in data. Then I zoomed a certain part of analysis and filtered the data for getting more detailed insights. So, I applied the Shneiderman's Mantra: "Overview first, zoom and filter, then details-on-demand." I also utilized colors and sizing as pre-attentive attributes.

Three specific design choices I made during this project can be summarized as follows:

- Coloring and sizing as pre-attentive attributes
- High level of interactivity in compliance with the audience and their needs
- Multi-frame story points by applying Shneiderman's Mantra to each dashboard

In the project proposal that I prepared at the beginning of the course, my intention was analyze 10 countries' data. However, since using all of them is too big and saving the notebook takes a very long time, I decided to analyze just 5 countries' data. So, I included this project Canada, US, UK, France, and Germany and make them a table by using 'union'. I especially focused on the countries which have relatively high English-speaking rate (Germany:56%, France: 39%; the official language is already English in US, UK, Canada) to be able to make meaningful comparisons between the same trending videos in different countries.

Moreover, I stated that I will prepare both single-frame handouts and multi-frame story points for the presentation. However, I've seen that it's too difficult and meaningless at this point to make a static visualization to this analysis. That's why, I present the results as multi-framed story-points.

Persona 1: Micheal

Role: Growth Lead

Male, 38 years old, having a MBA

- **Goals:** Micheal is responsible for growth of acquiring new customers in new markets in the most successful and fearless way. To this purpose, as a team, they are trying to new approaches and marketing channels.
- **Challenges:** Micheal is a result-oriented person. He is familiar with data however he focused on more the strategies and outcomes as the nature of his role.
- **Context:** His goal is to find the most effective way to expand their company by examining different approaches on marketing their product in different markets. For this purpose, he would like to see the common trends in different countries on trending Youtube videos and map their marketing strategies accordingly such as deciding what kind of YouTuber they will work with.

Persona 2: Carrie

Role: Marketing Lead

- **Goals:** Carrie is responsible for digital marketing department of the company. She has good insights regarding marketing trends and she knows how to use them effectively. She has good relations with clients and always want to understand to their point clearly and serve them devotedly.
- **Challenges:** Carrie has an excellent insights regarding digital marketing with her more than 10 years experience in this area however she is not an expert of numbers and technical details behind data.
- **Context:** She is supposed to help me during the preparation period of this project.

Female, 43 years old, passionate about online marketing

Persona 3: Tony and Sara

Role: Junior and Senior Analysts

Demographics: Male and Female, 25 and 32 years old, holding Bachelor degrees in business

- **Goals:** Sara and Tony working in Growth Team at the client company. They are responsible to make market research and fond out new ways to expand the company in an efficient way.
- **Challenges&Needs:** They are very good with data and analytics however they haven't expertise in business insights.
- **Context:** They would like to make meaningful inferences from this presentation to be able to contribute the process of creating the growth strategy.