

Youtube Data US

1. Insight no.1

- **Link -** [Youtube Data US | Tableau Public](#)

- **Summary -**

From the visualization, it can be immediately observed that the Music category trended the most and recorded the highest number of likes in the year 2018. Also the Entertainment category recorded the highest number of dislikes in the year 2014.

The ratio of Likes to Dislikes was calculated to derive a single measure to compare the popularity of the categories over a range of years.

- **Design -**

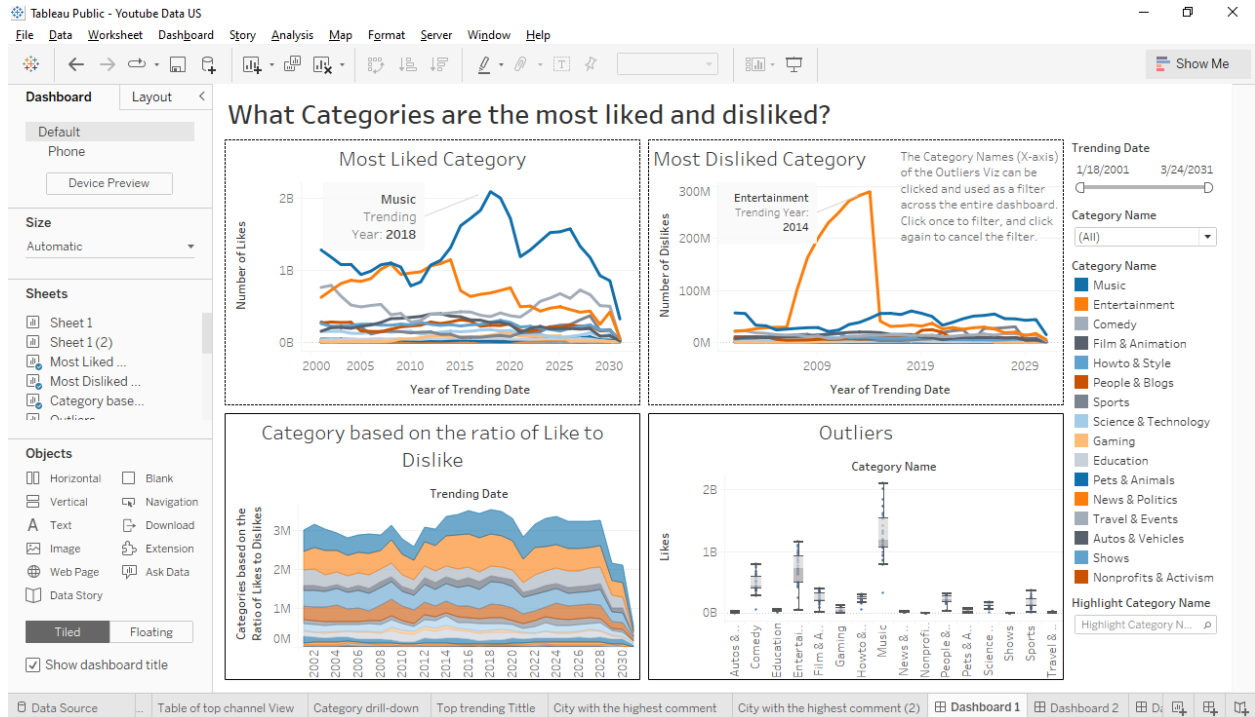
Firstly I chose a color palette suitable for the color blind. I opted for a continuous line graph to better represent the continuous variable (years). The area chart effectively shows the rise and fall of the measure over time and finally the box plot easily visualizes outliers.

I also included additional information (e.g. ratio of likes to dislikes, number of likes and dislikes) where necessary using the Tooltip. The X-axis of the Outlier viz is ideal for interaction to be used as a filter, the instruction is also included as a display text object.

- **Resources**

- <https://www.kaggle.com/datasnaek/youtube-new/data>
- https://www.researchgate.net/publication/356833178_BIG_DATA_ANALYSIS_ON_YOUTUBE_WITH_TABLEAU
-

Screenshot of the dashboard incase of discrepancies



2. Insight no.2

- **Link - [Youtube Data US | Tableau Public](#)**
- **Summary**

The visualization focuses on the popularity of channels at any selected year period. From Insight no. 1, it was observed that the music category was the most in the year 2018. Therefore, in this visualization **jypentertainment** recorded the highest number of views in the year 2018 from the state of **Florida**.

- **Design**

The choice of bar chart and bubble chart is ideal for representing categorical data. I also opted for a color palette suitable for color blind. I also infused a worksheet into the tooltip of the bubble visualization.

- **Resources -**

N/A

3. Insight no.3

- **Link - [Youtube Data US | Tableau Public](#)**
- **Summary**

This story viz focuses on all three insights, the **category**, **channel** and **title** trends. From the Category trend we can observe that the **Music category** trended the most the highest **number of likes** in the year 2018 while the **Entertainment category** trended negatively with the highest **number of dislikes** recorded in the year 2014.

Furthermore, knowing that the **Music category** trended the most with the highest number of likes in the year **2018**. The **Channel trend** shows us the channel with the highest **number of views**. It can be observed that **jypentertainment channel** in the **Music category** recorded the highest number of views in the year 2018 from the state of **Florida**.

Lastly, the **Title trends** tells us which titles recorded the most engagement and trends in any particular year. It can be observed that the title **GOT7 Look M/V** recorded the highest **number of comments** majorly from the city of **Pensacola** in the state of **Florida** in the year 2018.

In conclusion, this youtube Title **GOT7 Look M/V** published by **jypentertainment channel** under the **Music category** had the most engagement in the year 2018.

- **Design -**

The story visualization was made up of three different dashboards, all colors displayed were from the color blind palette. The visualization used in the Title trends were **horizontal bar chart**, **circle view** and an **area map**. **Area map** was ideal for displaying the states within the US where comments originated from. The circle view requires just a dimension (city) and a measure (sum of comment counts), while the horizontal bar chart is naturally ideal for representing categorical variables (in this case titles).

The **packed bubbles** chart requires at least a dimension (Category and State) and a measure (no of views). The **vertical bar chart** requires a minimum of 1 measure (no of views).


The **line chart** requires at least a measure (no of likes or dislikes) and since I was plotting it against time, it was my ideal choice of graph. The **area chart** requires at least a measure (ratio of likes to dislikes) and is ideal for showing the

rise and fall of a measure over time. Lastly, the **box-and-whisker** plots requires at least a measure (sum of likes) and is ideal for identifying outliers.

- **Resources -**

N/A

Reference

1.  Tips and tricks - Tableau project
2. <https://www.kaggle.com/datasets/usdot/flight-delays>
3. <https://www.youtube.com/watch?v=9xqHA732LMA>
4. <https://review.udacity.com/#!/rubrics/1060/view>
5. https://www.researchgate.net/publication/356833178_BIG_DATA_ANALYSIS_ON_YOUTUBE_WITH_TABLEAU