

Final Project Milestone #4

CIS 9440 - Data Warehousing for Analytics

Final Project Milestone 4

Group Number - 12

Student(s) – Chau Hoang & Tanay Mukherjee

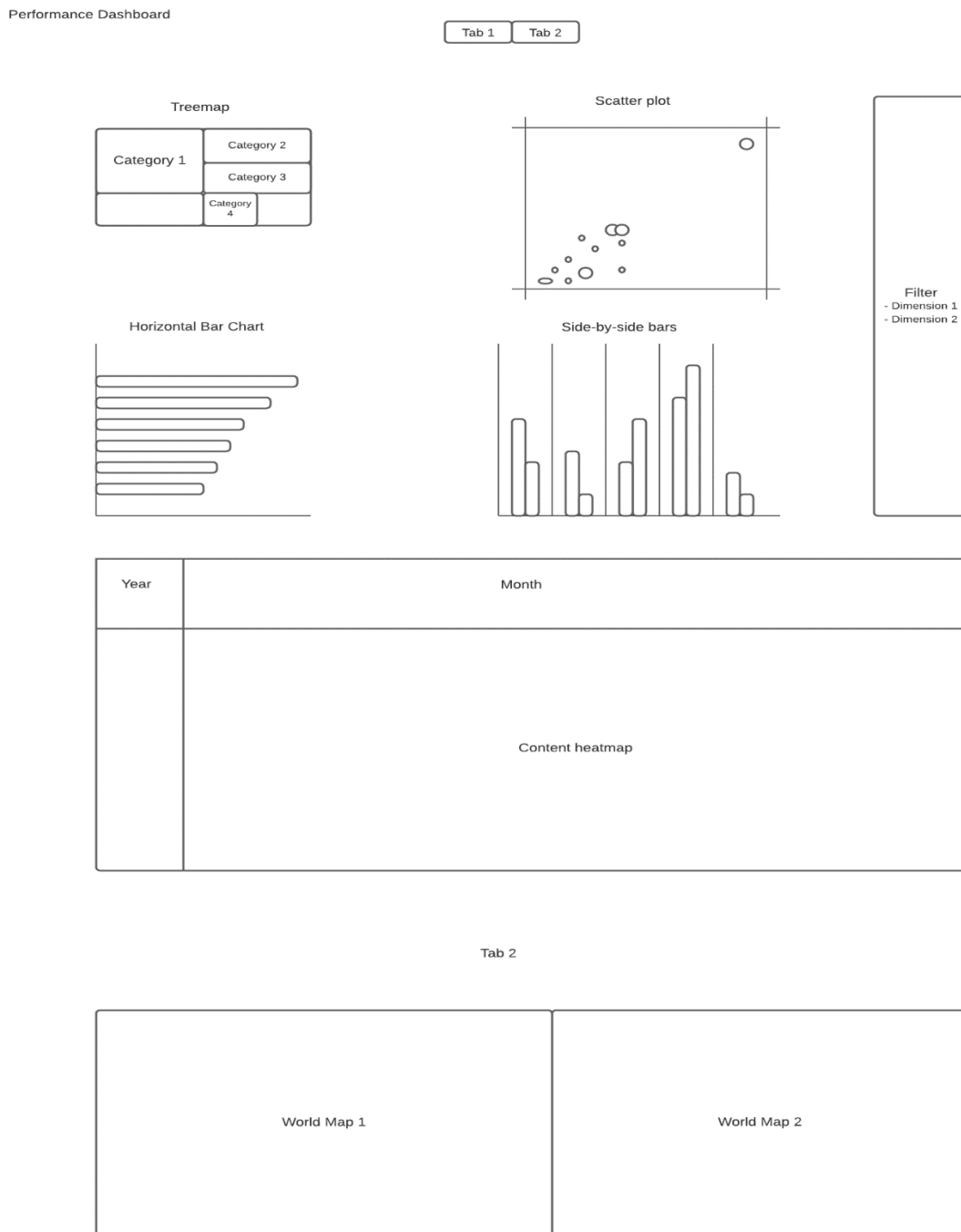
1) Our final set of KPIs:

- Total views by category
- Top channels with most trending videos
 - With filters for categories
- Interactions by video views
 - Likes and dislikes per view
 - Grouped by category
- Total Likes and Total Dislikes by Geo
- YoY Comparison [2017 vs 2018] of key metrics by category
 - Comment count
 - Total Likes
 - Total Dislikes
 - Total Views
- Total Likes and Total Dislikes by Year & Month
- Content heatmap for Netflix for by Year & Month
- Total records in Netflix by Geo

2) Description of Visualization we used for each KPIs and why we chose a particular visualization:

- a) Total views by category – **Tree map**: This viz. helps us not only measure the share of views for each category but also how big the share is with respect to overall data sample.
- b) Top channels with most trending videos – **Bar chart**: This viz. help us compare the top 10 channels for any category.
- c) Interactions by video views – **Bubble Chart (Scatter plot)**: This viz is useful as we can measure our dimensions against multiple metrics by size of the bubble, color of the bubble and positioning of the bubble across multiple x-axes and y-axis as it is in this case.
- d) Total Likes and Total Dislikes by Geo – **Map View**: This is to use the inbuilt world map from Tableau and plot the metrics against each country for YouTube.
- e) YoY Comparison of key metrics by category – **Stacked bar chart**: Just to compare various performance metrics for periods. In this case 2017 vs 2018.
- f) Total Likes and Total Dislikes by Year & Month – **Multiple Bar chart**: Comparison amongst bar charts for multiple metrics and how they stand against each other.
- g) Content heatmap for Netflix for by Year & Month – **Heatmap**: It will explain the spread of total content records across months and years and how the spread is higher or lower to their neighbours.
- h) Total records in Netflix by Geo – **Map View**: This is to use the inbuilt world map from Tableau and plot the metrics against each country for Netflix.

3) BI Application Wireframe



4) Link to our Tableau Public dashboard

https://public.tableau.com/profile/tanay.mukherjee#!/vizhome/CIS_9440_Project_Group_12/PerformanceDashboard

5) Additional Notes:

We followed the process as shared for each milestone and when we reach this final stage, we integrated Google Big Query with Tableau. Here, we called in the different fact and dimension tables we created in Phase 3 and did necessary joins. See below:

i) YouTube Database

Connections: BigQuery (Google BigQuery)

Billing Project: cis9440-project-group12

Project: cis9440-project-group12

Dataset: phase_3

Table: content_yt_dim, content_yt_dim, country_dim, date_dim, nfix_fact, New Custom SQL, New Union, Use Legacy SQL

YouTube_DB

Connection: Live (selected), Extract

Filters: 0 | Add

Sort fields: Data source order

Show aliases: ☐ Show hidden fields: ☐ 1,000 rows

Video Id (Content Yt Dim)	Title	Channel Title	Thumbnail Link	Tags	Description	Category	Comments Disabled
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False
hWuY48zvi	Sinclair's script for stations	D	https://ytimg.com/vi/hWuY48zvi	[none]	Other	People & Blogs	False

ii) Netflix Database

Connections: BigQuery (Google BigQuery)

Billing Project: cis9440-project-group12

Project: cis9440-project-group12

Dataset: phase_3

Table: content_nfix_dim, content_yt_dim, country_dim, date_dim, nfix_fact, New Custom SQL, New Union, Use Legacy SQL

Netflix_DB

Connection: Live (selected), Extract

Filters: 0 | Add

Sort fields: Data source order

Show aliases: ☐ Show hidden fields: ☐ 1,000 rows

Show Id (Content Nfix Dim)	Type	Title	Director	Category	Description	Country Id	Country (Country Dim)
s5	Movie	21	Robert Luketic	Dramas	A brilliant group of studen...	1003	United States
s5	Movie	21	Robert Luketic	Dramas	A brilliant group of studen...	1075	United States
s8	Movie	187	Kevin Reynolds	Dramas	After one of his high schoo...	1003	United States
s8	Movie	187	Kevin Reynolds	Dramas	After one of his high schoo...	1075	United States
s233	Movie	A Stoning in Fulham County	Larry Elikann	Dramas	After reckless teens kill an...	1003	United States
s233	Movie	A Stoning in Fulham County	Larry Elikann	Dramas	After reckless teens kill an...	1075	United States
s296	Movie	Across Grace Alley	Ralph Macchio	Dramas	A young boy, upset by his ...	1003	United States
s296	Movie	Across Grace Alley	Ralph Macchio	Dramas	A young boy, upset by his ...	1075	United States
s309	Movie	Adam: His Song Continues	Robert Markowicz	Dramas	After their child was abdu...	1003	United States

After this we were able to plot multiple KPIs mentioned above as individual sheets which were later put together in a dashboard before merging all the work in a story of 2 pages. We followed the wireframe above and split that into two dashboards. See below:

NOTE: The axis has been updated in a way that it reflects big numbers accordingly like suffix T is for trillions, B for Billions, K for thousand etc. However, when we publish the dashboard on public cloud it starts to show up the original number and thus gets overlapped. The screenshot below shows that it is working fine on local machine.

