Youtube Analysis

12.13.2023

Vaishakhi Shah

New York University New York, NY 10036

Overview

Company Overview:

YouTube, founded on February 14, 2005, in San Bruno, California, is a global social networking and video-sharing platform that has revolutionized how we create and consume content. With billions of viewers and creators worldwide, YouTube serves as a hub for diverse content, fostering communities, providing income opportunities, and continually innovating in online video. It's not just a platform; it's a dynamic global force for creativity, education, and economic impact. This dataset showcases YouTube's immense reach, impact, and role as a cornerstone of the digital content landscape.

Project Overview:

Inviting to the enthralling world of YouTube celebrity, where this painstakingly curated dataset reveals the statistics of the most subscribed YouTube channels. This dataset, composed of YouTube titans, provides an ideal opportunity to analyze and gain valuable insights from the platform's luminaries. This treasure trove of information is a must-explore for aspiring content creators, data enthusiasts, and anyone interested in the ever-changing online content landscape, with comprehensive details on top creators' subscriber counts, video views, upload frequency, country of origin, earnings, and more. With this extraordinary dataset, you can immerse yourself in the world of YouTube success and gain access to a wealth of knowledge.

Goals

Geographical Insights:

Latitude and longitude coordinates for potential geospatial analyses.
Understanding global reach and distribution of channels.

Revenue Generation:

Estimated earnings offer insights into monetization potential and revenue generation.

Historical Analysis:

Examination of channel growth patterns via creation date information.
Real-time snapshot of activity through recent engagement levels (recent video views, subscriber growth).

Demographic View:

Population figures, educational enrollment rates for demographic insights into viewers. Spatial analysis possible with geospatial coordinates to reveal distribution patterns.

Comparative Analysis:

Comparative indicators across countries provide socio-economic context to channel performance.

Utility:

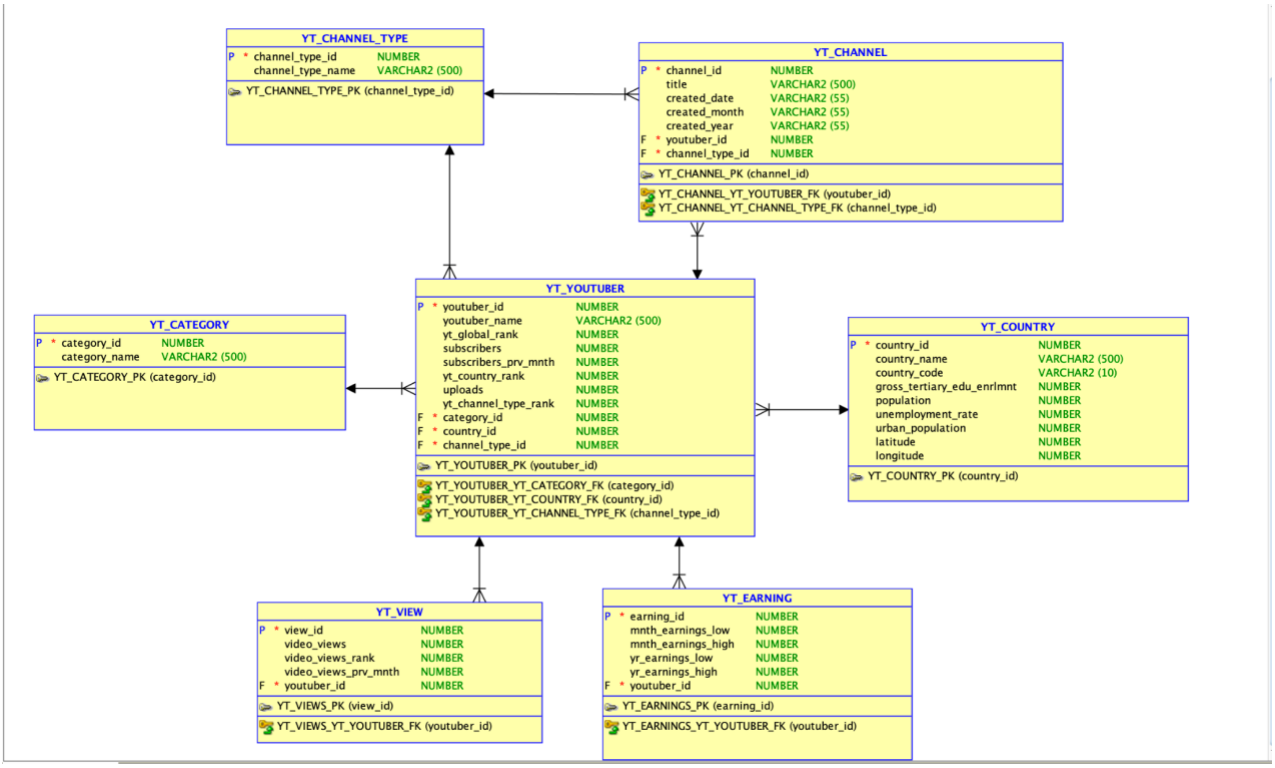
Valuable resource aiding content creators and YouTube community stakeholders. Informs decision-making based on channel dynamics and performance trends.

In essence, this dataset serves as a comprehensive resource for understanding YouTube channel dynamics, enabling informed decision-making and analysis for content creators and stakeholders within the YouTube community.

Deliverables

1. Introduction of Data Model and Data Preparation.
2. Data Analysis: Insights on key performance metrics of YouTube channels and analyzing the trends in channel growth, geographical distribution, and revenue potential.
3. Comprehensive and Comparative Analysis on different Metrics.
4. Visualizations: Visual representations of potential earnings, categories, demographics, and geographic representation.
5. Limitations and Summary about the project.

Data Model



Data Preparation

Dataset Overview:

- Comprehensive information about various YouTube channels.
- Each entry represents a channel with specific details: subscribers, video views, category, etc.
- Crucial metrics for assessing a channel's popularity and impact are included.

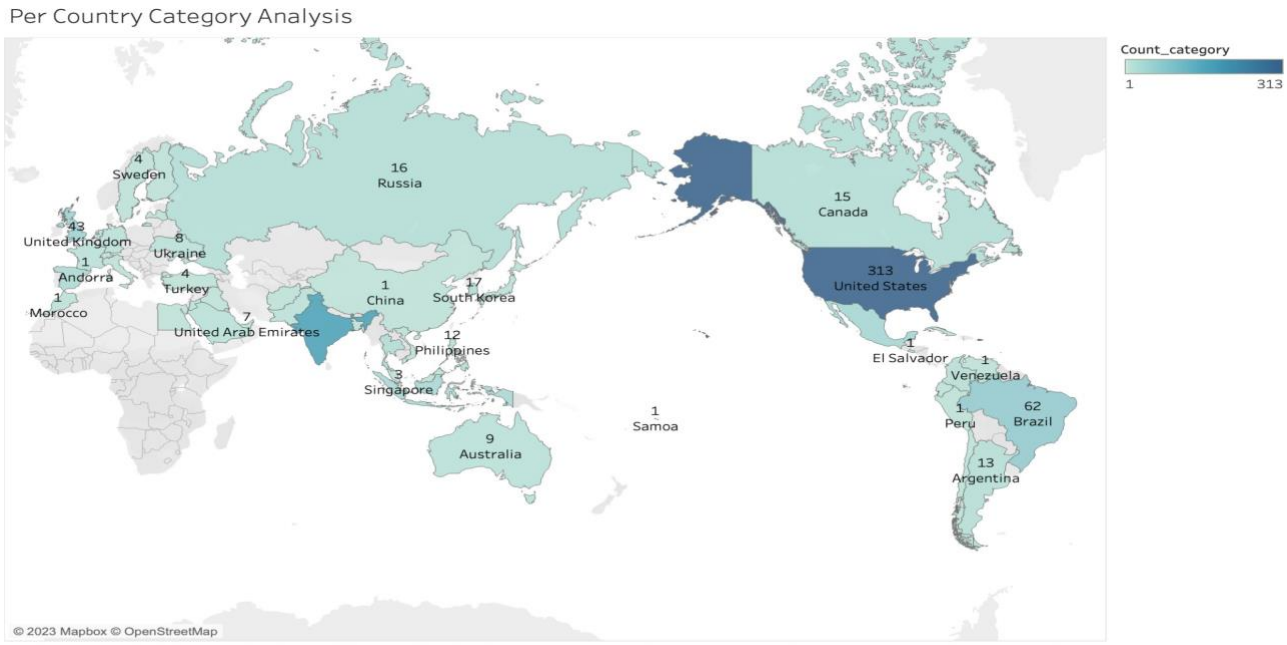
Key Metrics:

- Subscribers, video views, category, and country of origin provided.
- Estimated earnings for insights into revenue potential influenced by content type and engagement.
- Creation dates enable trend analysis and growth assessment.

Analysis Part I :

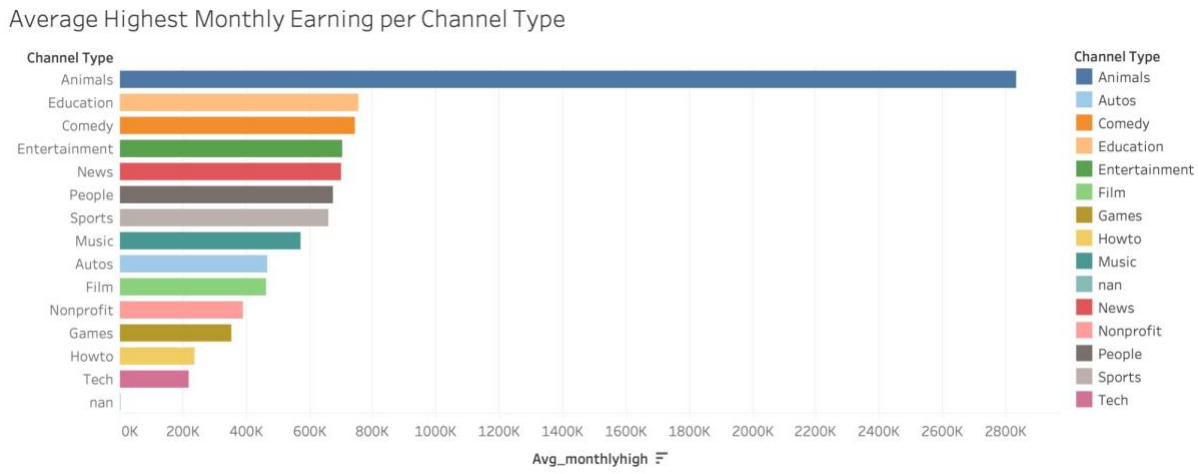
Global Insights: Analyzing Geographic, Monthly, and Subscriber - Based Earnings Trends

I. Geographic Analysis: Per Country Category Analysis



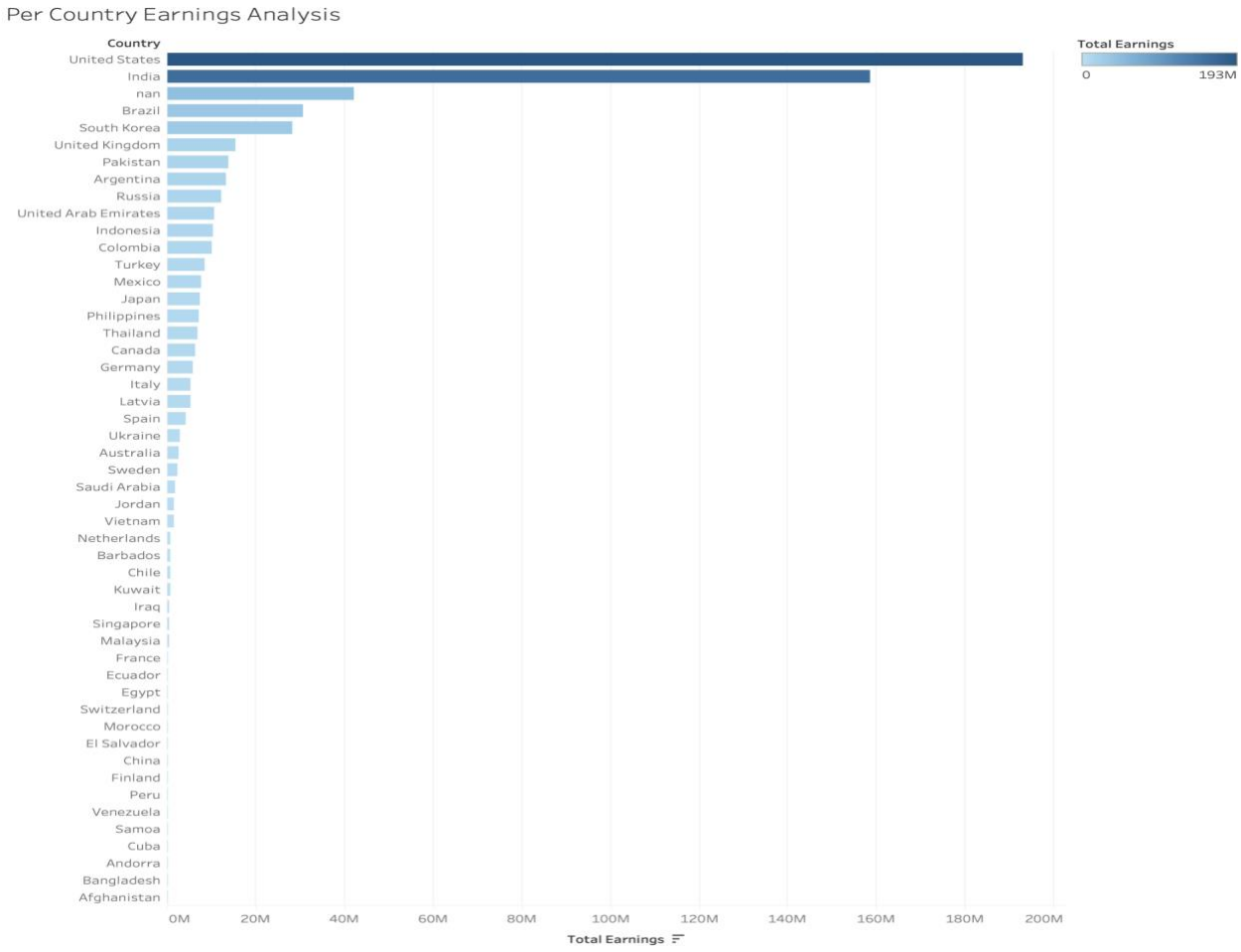
Map based on Longitude (generated) and Latitude (generated). Color shows Count_category. The marks are labeled by Count_category and Country. Details are shown for Country.

II. Average Highest Monthly Earnings Per Channel Type



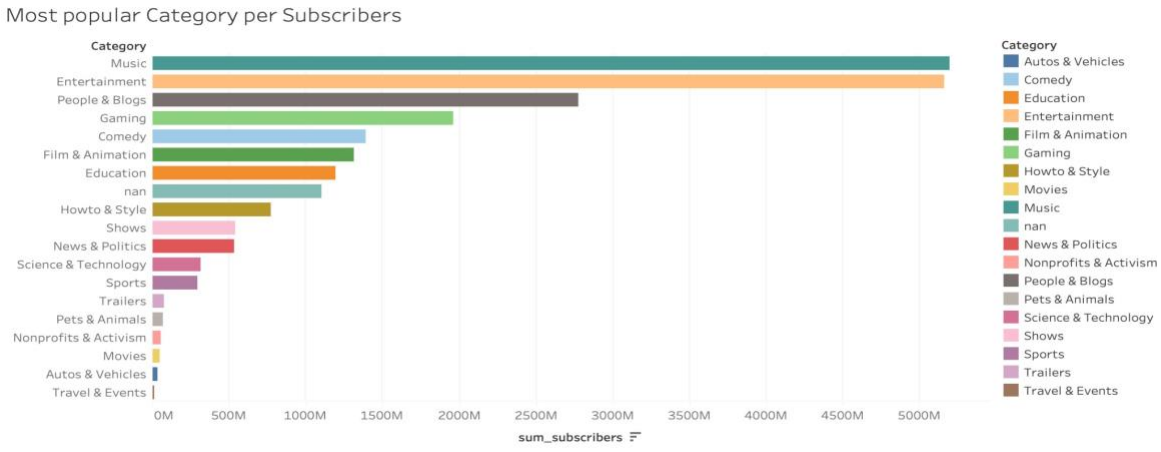
Avg_monthlyhigh for each Channel Type. Color shows details about Channel Type.

III. Per Country Earnings Analysis



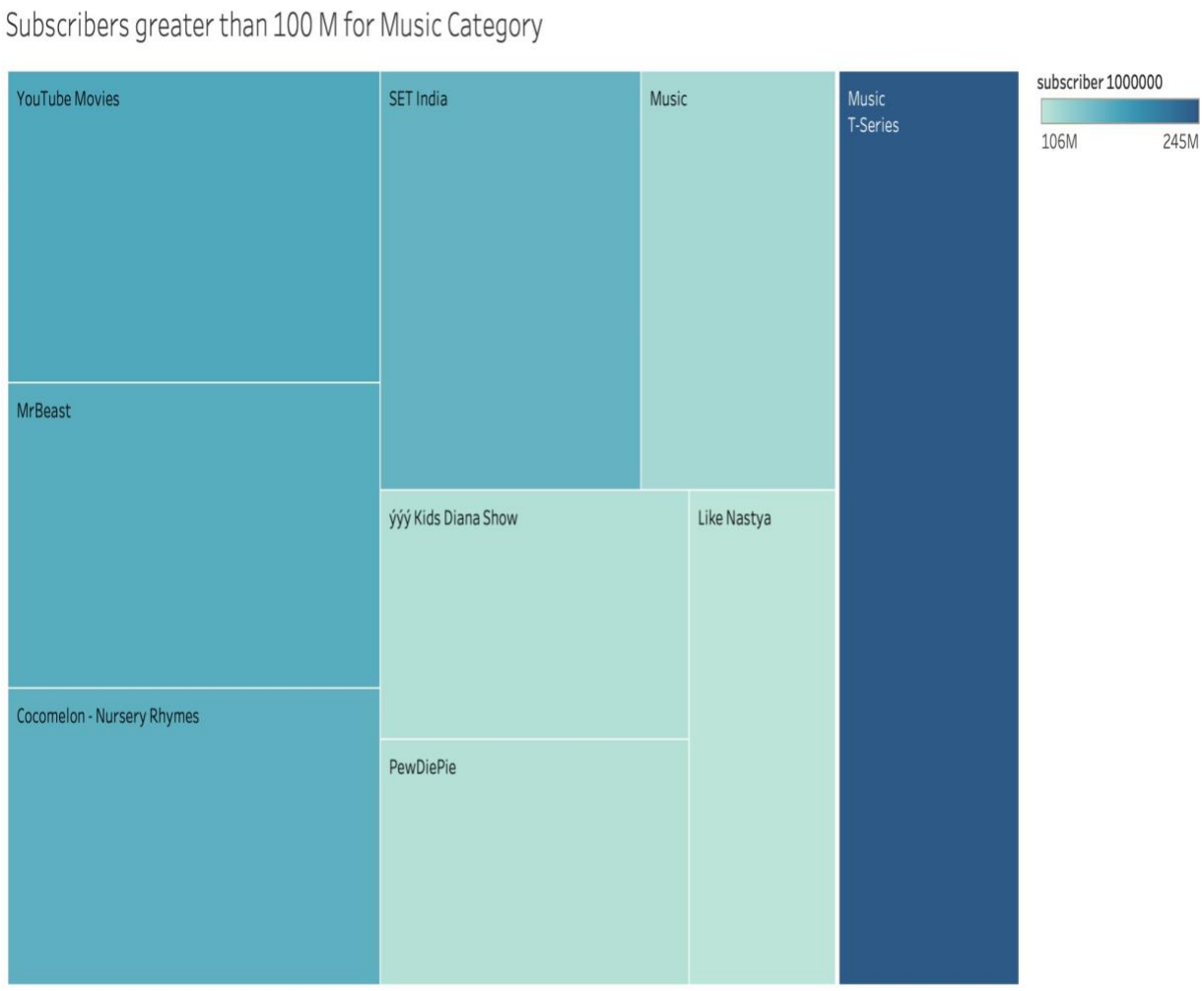
Total Earnings for each Country. Color shows Total Earnings.

IV. Most Popular Category per Subscribers



Sum_subscribers for each Category. Color shows details about Category.

V. Subscribers greater than 100 M for Music Category

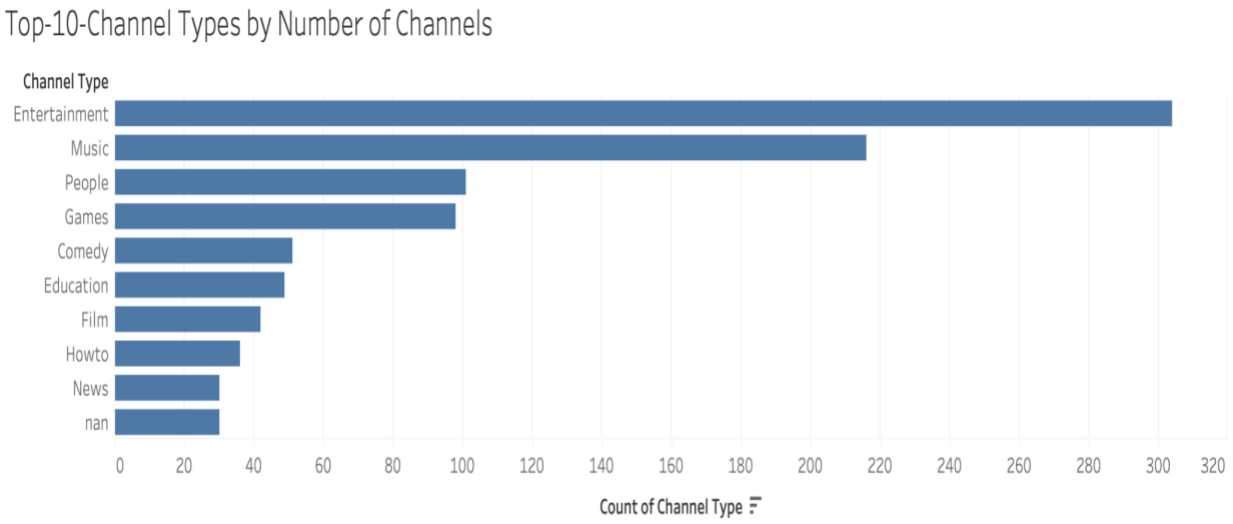


Music category and Youtuber. Color shows sum of subscriber 1000000. Size shows sum of subscriber 1000000. The marks are labeled by Music category and Youtuber.

Analysis Part II :

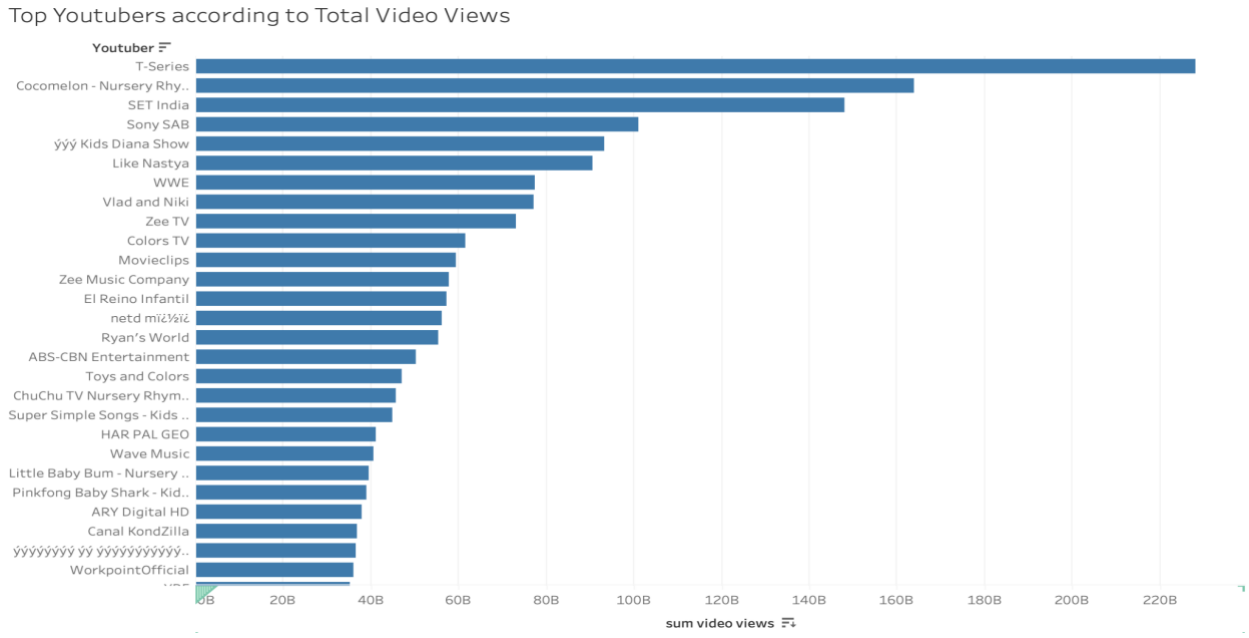
YouTube Dynamics: A Comprehensive Analysis of Top Channels and Categories

I. Top 10 Channel Types by Number of Channels

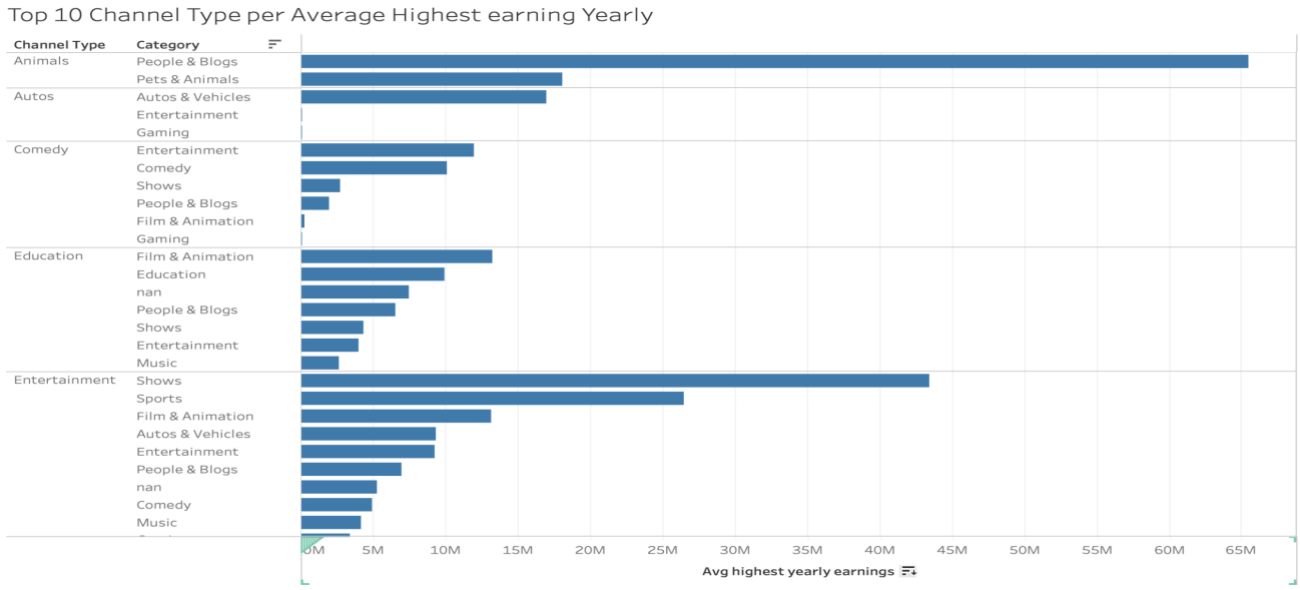


Count of Channel Type for each Channel Type. The view is filtered on Channel Type, which excludes Animals, Autos, Nonprofit, Sports and Tech.

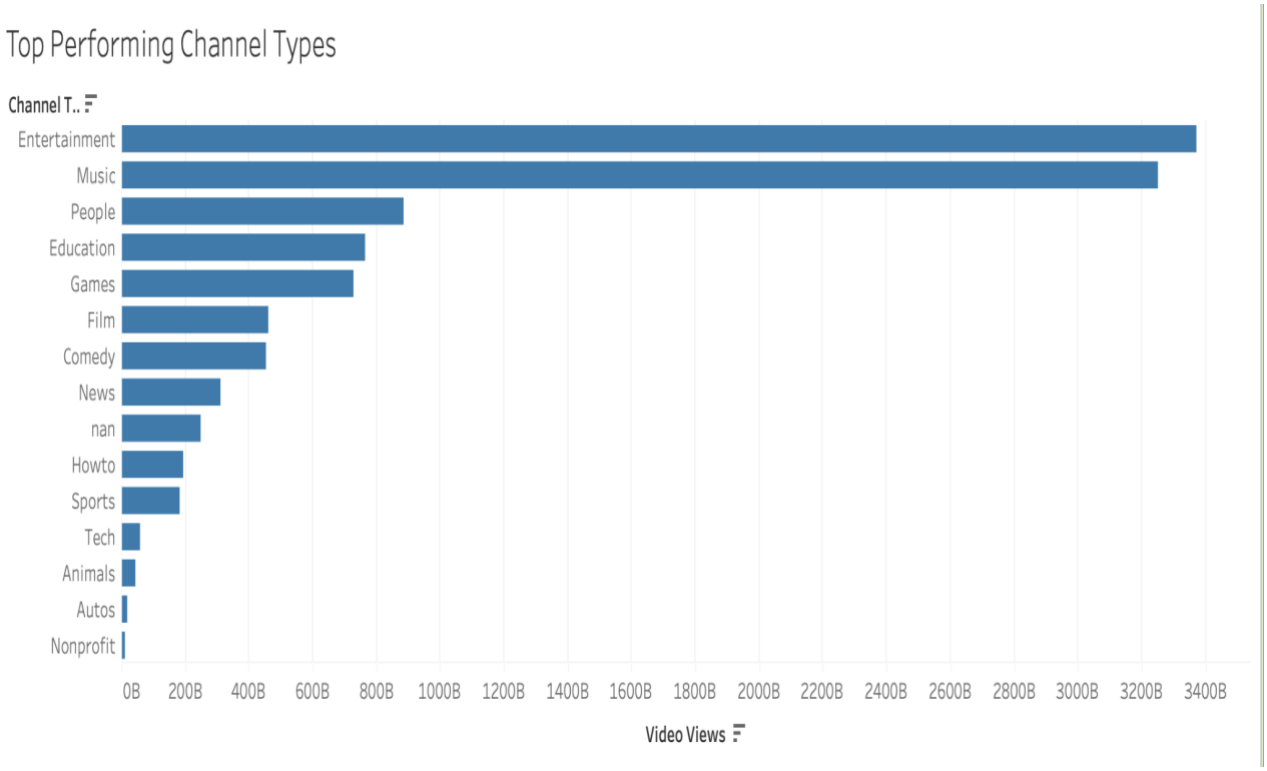
II. Top Youtubers Video Views Analysis



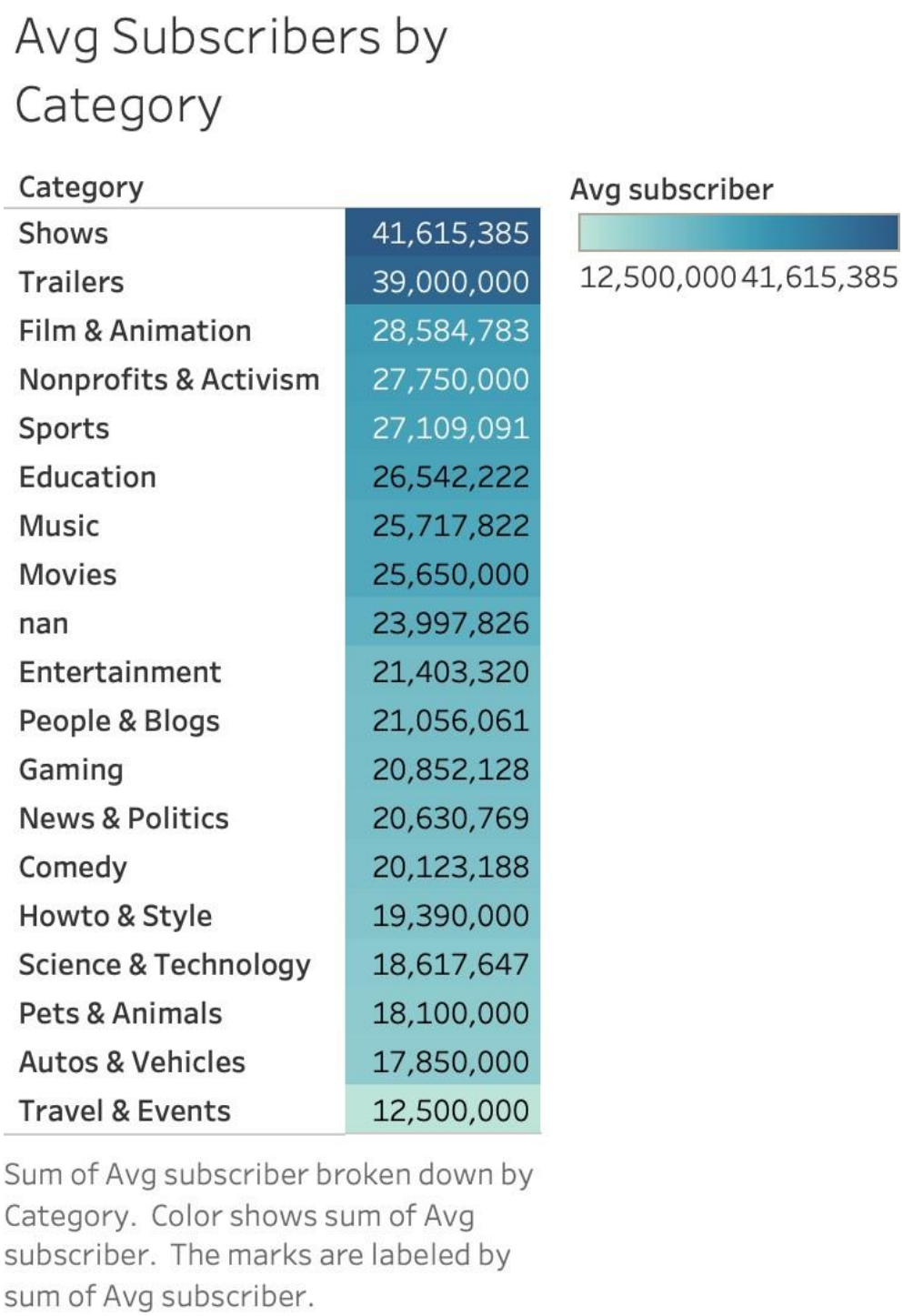
III. Top 10 Channel Type Per Average Highest Earning Yearly



IV. Top Performing Channel Types

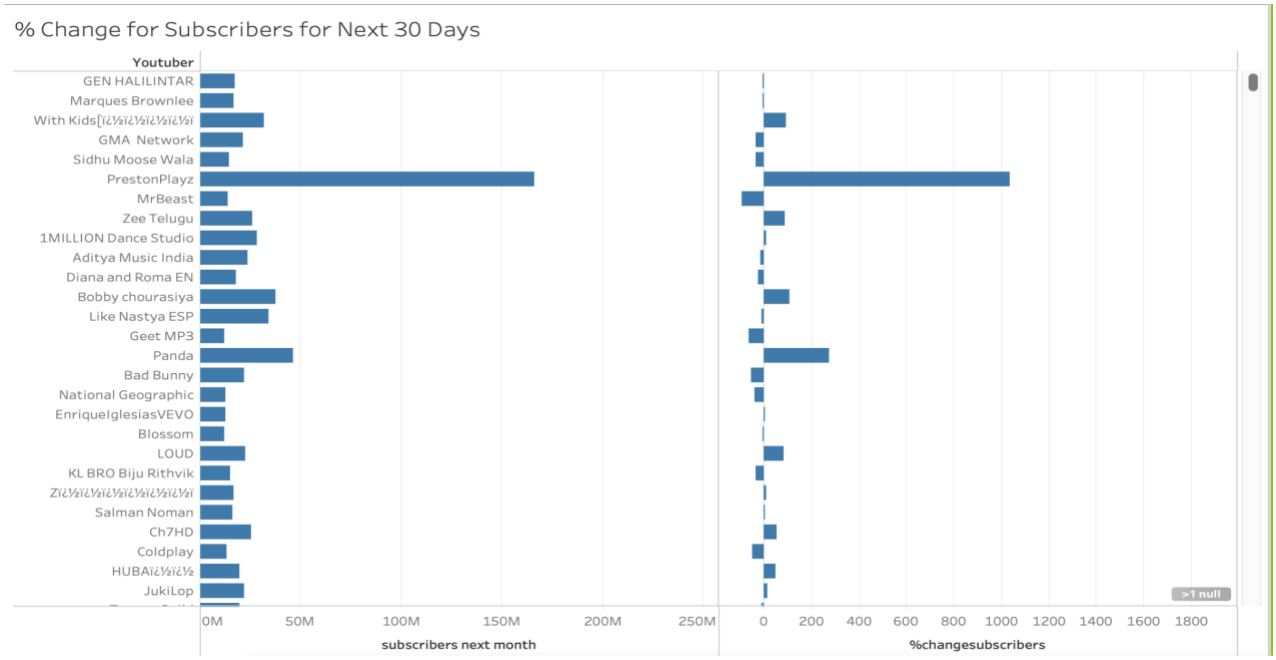


V. Average Subscribers by Category

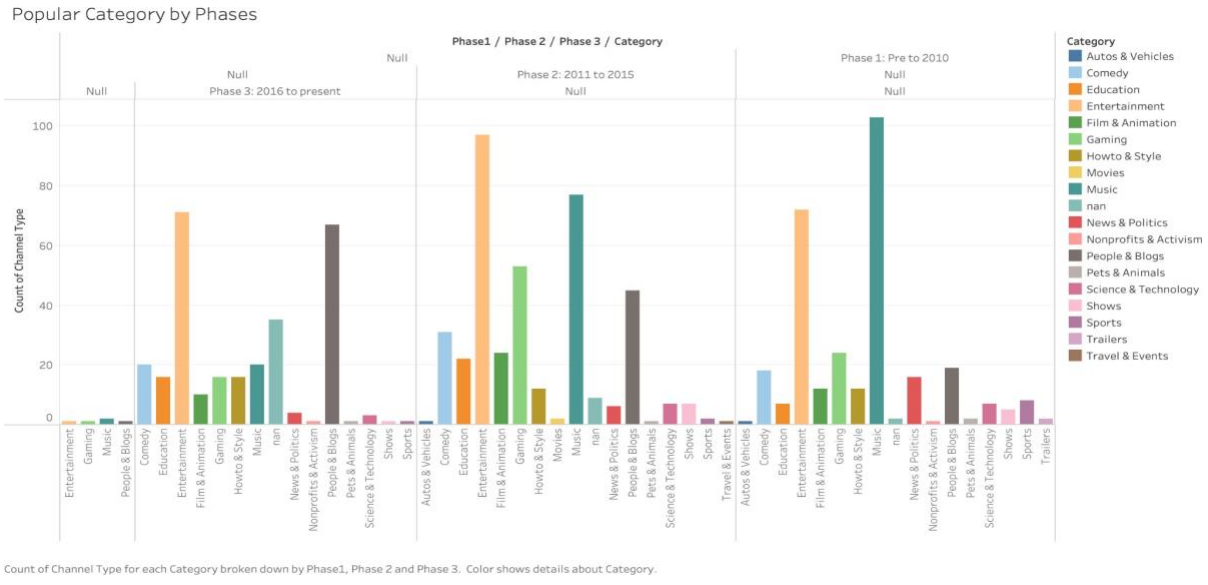


Analysis Part III :
YouTube Metrics Snapshot: Subscribers, Popular Categories, Correlations, Conversion Rates, and Trends

I. % Change Analysis for Subscribers per Youtuber

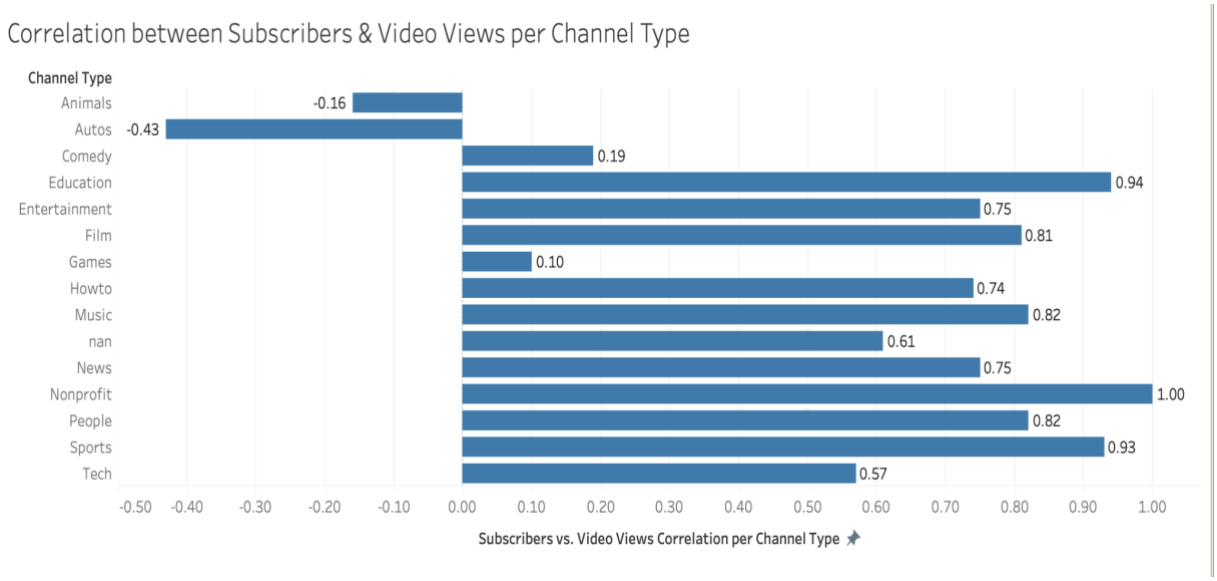


II. Popular Category by Phases

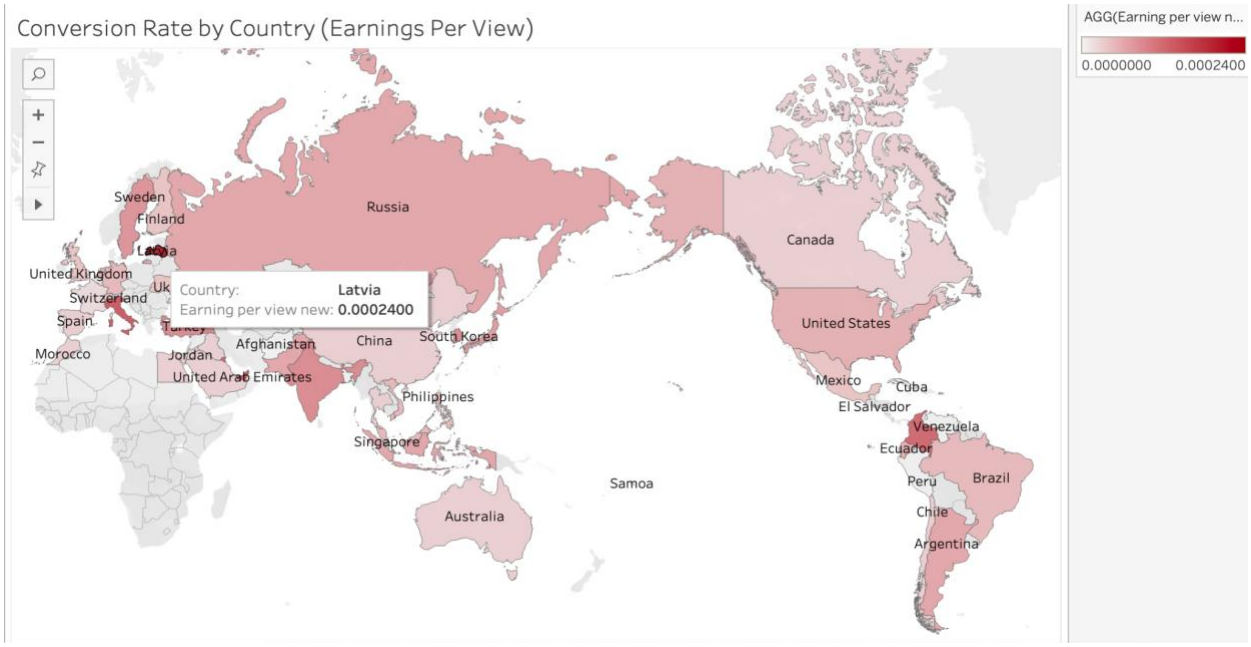


Count of Channel Type for each Category broken down by Phase1, Phase 2 and Phase 3. Color shows details about Category.

III. Correlation between Subscribers & Video Views per Channel Type



IV. Conversion Rate by Country for Earning per View



V. Subscribers Trend Analysis

Limitations

Temporal Scope:

The dataset might cover a specific period in 2023, which limits insights into long-term trends or seasonal variations across the platform.

Missing Variables:

It may lack certain critical variables that could significantly contribute to understanding YouTube dynamics, such as user engagement metrics, content specifics, or algorithmic changes.

Geographical Representation:

Limited geographic representation or uneven distribution of channels across countries might hinder comprehensive global analyses.

Summary of your project

Our project revolves around leveraging a YouTube dataset to create a series of SQL queries. The dataset encompasses various attributes, including Youtuber information, subscribers, video views, categories, titles, and more. By using these data points, we have devised a comprehensive set of SQL queries that allow us to extract meaningful insights and answers to specific questions.

These SQL queries serve as an essential tool for honing our SQL skills, encompassing various SQL concepts and operations such as filtering, sorting, aggregating, and joining tables. Our project aims to provide a platform for SQL enthusiasts and learners to practice their skills and gain proficiency in handling real-world data scenarios. Whether it's ranking YouTubers by subscribers, calculating averages, or dissecting data by country, category, or specific attributes, our SQL queries are designed to explore diverse dimensions of the YouTube dataset.

By combining data from YouTube with the power of SQL, our project facilitates a hands-on, practical learning experience, allowing users to gain proficiency in querying, analyzing, and visualizing data efficiently. Moreover, it encourages users to experiment, adapt, and extend these queries to suit their specific needs and dataset characteristics.

Appendix -- SQL

```
APPENDIX - SQL

-----PART 1----
------I. Geographic Analysis:  Per Country Category Analysis --------
create view yt_popular_categories_countries_v
as
SELECT DISTINCT yt_category.category_name,
        yt_country.country_name,
        count(yt_category.category_name) AS Category_Count
from
 yt_youtuber
JOIN yt_category
    ON (yt_category.category_id = yt_youtuber.category_id)
JOIN yt_country
    ON (yt_country.country_id = yt_youtuber.country_id)
    GROUP BY
 yt_country.country_name,yt_category.category_name
ORDER BY Category_Count desc


-----II. Average Highest Monthly Earnings Per Channel Type------------
create view yt_avg_highest_earning_v
as
SELECT
 yt_channel_type.channel_type_name,
 yt_category.category_name,
 avg(yt_earning.mnth_earnings_high) AS avg_highest_monthly_earnings
FROM
 yt_youtuber
JOIN
 yt_earning
    ON (yt_earning.youtuber_id = yt_youtuber.youtuber_id)
JOIN
 yt_category
    ON (yt_category.category_id = yt_youtuber.category_id)
JOIN
 yt_channel_type
    ON (yt_channel_type.channel_type_id = yt_youtuber.channel_type_id)
GROUP BY
 yt_channel_type.channel_type_id,
 yt_channel_type.channel_type_name,
 yt_category.category_name
HAVING
 category_name != 'NaN'
ORDER BY avg_highest_monthly_earnings desc
```

```sql
-------III. Per Country Earnings Analysis-------------------------
WITH CountryEarnings AS (
 SELECT
 c.country_id,
 c.country_name,
 SUM(e.mnth_earnings_low + e.mnth_earnings_high) AS total_earnings
    FROM yt_country c
LEFT JOIN yt_youtuber y
    ON c.country_id = y.country_id
LEFT JOIN yt_earning e
    ON y.youtuber_id = e.youtuber_id
    GROUP BY c.country_id,
        c.country_name
)
SELECT *
    FROM COUNTRY_EARNINGS_ANALYSIS
ORDER BY TOTAL_EARNINGS DESC

-----IV. Most Popular Category per Subscribers---------------------
create view yt_most_popular_category_v
as
SELECT
 category_name,
 sum(subscribers) AS total_subscribers
FROM
 yt_youtuber
JOIN yt_category
    ON (yt_youtuber.category_id = yt_category.category_id)
group by
 category_name
order by
 total_subscribers desc;
SELECT *
FROM yt_most_popular_category_v;


------V. Subscribers greater than 100 M for Music Category---------
create view Youtubers_100MSubscribersMusic_V
as SELECT Youtuber.youtuber_name AS Youtuber_Name,
        Youtuber.subscribers,
        Views.video_views AS Video_Views,
        Category.category_name AS Category_Name
    FROM yt_youtuber Youtuber
JOIN yt_view Views
    ON Youtuber.youtuber_id = Views.youtuber_id
JOIN yt_category Category
    ON Youtuber.category_id = Category.category_id
    WHERE Youtuber.subscribers > 100000000
        AND Category.category_name = 'Music'
ORDER BY Views.video_views DESC
```

```
----PART 2----

------I. Top Youtubers Video Views Analysis-------------------------
create view yt_top_youtuber_by_views
as
WITH
 top_youtubers AS (
SELECT youtuber_id,
        youtuber_name,
        yt_global_rank
    FROM yt_youtuber
 ),
 video AS (
 SELECT
 youtuber_id,
 SUM(video_views) AS Total_views,
 video_views_rank
    FROM yt_view
    GROUP BY youtuber_id, video_views_rank
 )


------II.Top 10 Channel Type Per Average  Highest Earning Yearly-----
create view yt_avg_highest_yr_earning_v
as
SELECT
 yt_channel_type.channel_type_name,
 yt_category.category_name,
 avg(yt_earning.yr_earnings_high) AS avg_highest_yearly_earnings
FROM
 yt_youtuber
JOIN
 yt_earning
    ON (yt_earning.youtuber_id = yt_youtuber.youtuber_id)
JOIN
 yt_category
    ON (yt_category.category_id = yt_youtuber.category_id)
JOIN
 yt_channel_type
    ON (yt_channel_type.channel_type_id = yt_youtuber.channel_type_id)
GROUP BY
 yt_channel_type.channel_type_id,
 yt_channel_type.channel_type_name,
 yt_category.category_name
HAVING
 category_name != 'NaN'
ORDER BY avg_highest_yearly_earnings desc
 FETCH FIRST 10 ROWS ONLY;
```

```sql
----III. Top Performing Channel Type-----------------------------
create view Top_Performing_channelTypes_v
as
SELECT Channel_type.channel_type_name,
        SUM(Views.video_views) AS total_views,
        SUM(Earning.mnth_earnings_low) AS total_monthly_earnings_low,
        SUM(Earning.mnth_earnings_high) AS total_monthly_earnings_high,
        SUM(Earning.yr_earnings_low) AS total_yearly_earnings_low,
        SUM(Earning.yr_earnings_high) AS total_yearly_earnings_high
    FROM yt_channel_type Channel_type
LEFT JOIN yt_youtuber Youtuber
    ON Channel_type.channel_type_id = Youtuber.channel_type_id
LEFT JOIN yt_earning Earning
    ON Youtuber.youtuber_id = Earning.youtuber_id
LEFT JOIN yt_view Views
    ON Youtuber.youtuber_id = Views.youtuber_id
    GROUP BY Channel_type.channel_type_name
ORDER BY total_views DESC;

----IV.-Top-10-Channel Types by Number of Channels--------------------
create view yt_avg_highest_yr_earning_v
as
SELECT
 yt_channel_type.channel_type_name,
 yt_category.category_name,
 avg(yt_earning.yr_earnings_high) AS avg_highest_yearly_earnings
FROM
 yt_youtuber
JOIN
 yt_earning
    ON (yt_earning.youtuber_id = yt_youtuber.youtuber_id)
JOIN
 yt_category
    ON (yt_category.category_id = yt_youtuber.category_id)
JOIN
 yt_channel_type
    ON (yt_channel_type.channel_type_id = yt_youtuber.channel_type_id)
GROUP BY
 yt_channel_type.channel_type_id,
 yt_channel_type.channel_type_name,
 yt_category.category_name
HAVING
 category_name != 'NaN'
ORDER BY avg_highest_yearly_earnings desc
 FETCH FIRST 10 ROWS ONLY;
```

```
-----V. Average Subscribers by Category-
create view yt_category_avg_subscribers_v
as
WITH category_avg_subscribers AS
 (select
 yt_category.category_name AS category,
 round(avg(subscribers),2) AS avg_subscribers
FROM
 yt_youtuber
JOIN yt_category
    ON (yt_youtuber.category_id = yt_category.category_id)
    GROUP BY
 yt_category.category_name)
select
 category,
 avg_subscribers
from
 category_avg_subscribers;


SELECT *
    FROM yt_category_avg_subscribers_v
ORDER BY avg_subscribers desc;



----PART 3----


----I. Conversion Rate by Country for Earning per View----------------
create view yt_country_conversion_rate_v
AS
SELECT country_name AS country,
 round(avg (mnth_earnings_high / video_views),5) AS earnings_per_view
FROM
 yt_youtuber
JOIN yt_view
ON (yt_youtuber.youtuber_id=yt_view.youtuber_id)
JOIN yt_earning
ON (yt_youtuber.youtuber_id=yt_earning.youtuber_id)
JOIN yt_country
ON (yt_youtuber.country_id=yt_country.country_id)
WHERE
 video_views > 0
        AND mnth_earnings_high > 0
GROUP BY
 country_name
order by
```

```
 earnings_per_view desc
fetch next 1 row only;

SELECT *
    FROM yt_country_conversion_rate_v;


----II. % Change Analysis for Subscribers per Youtuber-----------------
create view yt_youtuber_subscribers_change_v
as
SELECT
 youtuber_id,

 youtuber_name,

 MAX(subscribers_next_month) AS subscribers_next_month,

  ((MAX(subscribers_next_month) - MAX(subscribers)) / MAX(subscribers)) * 100
AS percent_change
FROM (
 SELECT
 youtuber_id,

 youtuber_name,

 LEAD(subscribers)
    OVER (ORDER BY youtuber_id) AS subscribers_next_month,
 Subscribers
FROM
 yt_youtuber
)
GROUP BY
 youtuber_id, youtuber_name;


----III. Popular Category by Phases-Popular Category By Phases-----------------
create view yt_most_popular_category_phases_v
AS
SELECT
'Phase 1: Pre to 2010' AS phase,
 category_name,
 count(*) AS channel_count
FROM
 yt_youtuber
JOIN yt_category
    ON (yt_youtuber.category_id = yt_category.category_id)
JOIN yt_channel
    ON (yt_youtuber.youtuber_id=yt_channel.youtuber_id)
WHERE
 created_year <= 2010
group by
```

```
 category_name
UNION ALL
SELECT
 'Phase 2: 2011 to 2015' AS phase,
 category_name,
 count(*) AS channel_count
FROM
 yt_youtuber
JOIN yt_category
    ON (yt_youtuber.category_id = yt_category.category_id)
JOIN yt_channel
    ON (yt_youtuber.youtuber_id=yt_channel.youtuber_id)
WHERE
 created_year
    BETWEEN 2011
        AND 2015
GROUP BY
 category_name
UNION ALL
SELECT
 'Phase 3: 2016 to present' AS phase,
 category_name,
 count(*) AS channel_count
FROM
 yt_youtuber
JOIN yt_category
    ON (yt_youtuber.category_id = yt_category.category_id)
JOIN yt_channel
    ON (yt_youtuber.youtuber_id=yt_channel.youtuber_id)
WHERE
 created_year >= 2016
group by
 category_name
ORDER BY
 phase,
        channel_count desc;
SELECT *
    FROM yt_most_popular_category_phases_v;


----IV.Correlation between Subscribers & Video Views per Channel Type--
create view yt_corr_videoviews_subscribers_v
as
SELECT
 channel_type_name,
 max(subscribers) AS max_subscribers,
 max(video_views) AS max_video_views,
 round(corr(subscribers,
        video_views),
       2) AS max_subscribers_video_views_correlation
FROM yt_youtuber
JOIN yt_view
ON (yt_youtuber.youtuber_id=yt_view.youtuber_id)
JOIN yt_channel_type
```

```
ON (yt_youtuber.channel_type_id=yt_channel_type.channel_type_id)
group by
 channel_type_name;

SELECT round(avg (max_subscribers_video_views_correlation),2)
AS videoviews_corr_subscribers
    FROM yt_corr_videoviews_subscribers_v;
```

```
-----V. Subscribers Trend Analysis -
create view yt_subscribers_trend_analysis_v
as
SELECT
 CORR(
 subscribers,
 (mnth_earnings_low+mnth_earnings_high)/2
 ) AS correlation
FROM
 yt_youtuber
JOIN yt_earning
    ON (yt_youtuber.youtuber_id=yt_earning.youtuber_id);
SELECT *
FROM
 yt_subscribers_trend_analysis_v;
```

References

https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023