GLOBAL YOUTUBE STATISTICS

VISWA SUHAAS PENUGONDA

11/07/2023

—

FA23 INFO-I 535 Management, Access, and Use of Big and Complex Data

—

Prof. Inna Kouper

INTRODUCTION:

YouTube has become one of the most dominant platforms for video content and boasts over 2 billion monthly active users worldwide. This has created an ecosystem where creative individuals can potentially reach a massive audience and even monetize their channels into full-time businesses. However, despite its vast scale and impact, there has been little quantitative analysis done on the platform itself - especially across different regional markets.

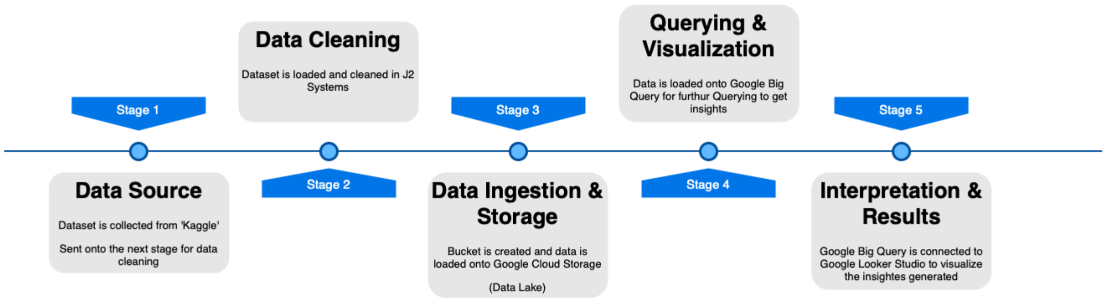
To fill this gap, this project aims to provide in-depth data-driven insights into YouTube's global landscape using a recently released dataset on the top 1,000 channels worldwide. By covering granular statistics like number of subscribers, total upload views, estimated earnings as well as geographic, category and channel type breakdowns, this dataset enables a multifaceted perspective into what content resonates best on YouTube and the dynamics behind a channel's success. Potential analyses include exploring the correlation between uploads and earnings, content preferences based on country demographics, analyzing breakout categories and more. These learnings can in turn be leveraged by existing and aspiring creators, media analysts as well as policy makers looking to support domestic talent. With online video projected to be a $335B industry by 2025, understanding the world's largest video platform is crucial. This project attempts to shed more light by deriving actionable, data-backed insights from the ground up based on real-world YouTube performance.

BACKGROUND:

Despite its vast scale and impact on internet culture and trends, relatively little structured analysis has been done on the dynamics behind what makes a YouTube channel successful. While individual influencers may share tips and tricks, there has been no holistic, data-driven study that quantitatively analyzes patterns across top channels on the platform. This project was conceptualized to fill that important gap by leveraging a recently released dataset on the top 1,000 YouTube channels worldwide. The dataset provides granular statistics on subscribers, views, estimated earnings as well as breakdowns by country, category, and channel type. Having such aggregated indicators across the very best channels on YouTube offers a unique opportunity to really understand what works and what doesn't for creators in different niches and geographies. Some examples of potential analyses include exploring correlations between number of uploads, video types, and earnings; analyzing differences in content preferences and creator personalities across regions; tracking breakout trends and categories over time.

These data backed insights can provide tremendous value to the public discourse as YouTube continues its relentless growth. Aspiring creators can optimize their strategy based on lessons from the top channels in their genre. Media analysts and journalists can leverage these findings to produce more informed commentary. Policy groups can track domestically relevant trends and calibrate support programs accordingly. Even for casual viewers, getting a pulse on what succeeds on the world's second largest search engine is powerful. With online video projected to be a $335B industry by 2025, decoding the YouTube phenomenon through data is hugely important - which is exactly what this project aims to do by deriving actionable, structured insights straight from aggregate statistics of real-world channel performance.

METHODOLOGY:



Data Source:

The foundational step of this project involved procuring the raw Global YouTube Statistics dataset from Kaggle, which contained approximately 1,000 rows of aggregated performance metrics for top channels on the platform. This real-world data established a n objective factual baseline for analysis. I subsequently imported this CSV resource into a Jupyter notebook leveraging Python within a J2 Virtual Machine environment for streamlined preprocessing and transformation.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 995 entries, 0 to 994
Data columns (total 17 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   rank                                995 non-null    int64
 1   youtuber                            995 non-null    object
 2   subscribers                         995 non-null    int64
 3   video_views                         995 non-null    int64
 4   uploads                             995 non-null    int64
 5   country                             995 non-null    object
 6   channel_type                        995 non-null    object
 7   country_rank                        995 non-null    int64
 8   channel_type_rank                   995 non-null    int64
 9   video_views_for_the_last_30_days    995 non-null    int64
 10  lowest_monthly_earnings             995 non-null    float64
 11  highest_monthly_earnings            995 non-null    float64
 12  lowest_yearly_earnings              995 non-null    float64
 13  highest_yearly_earnings             995 non-null    float64
 14  subscribers_for_last_30_days        995 non-null    int64
 15  population                          995 non-null    int64
 16  unemployment_rate                   995 non-null    float64
dtypes: float64(5), int64(9), object(3)
```

Information of the dataset

Data Cleaning:

Enabling analytics and visualization first required a series of data wrangling techniques to optimize integrity. I standardized the column identifiers to concatenate spaces. Then I programmatically scanned for and filled two missing values using mean/mode imputation of their respective attributes to mitigate distortion. As certain variables were irrelevant to answering our exploratory queries, I filtered the DataFrame to the analytically valuable subset. To avoid potential duplication biases, I scanned and dropped based on the unique name identifier. Finally, I iterated through the column data types using Pandas astype() to optimize software processing and querying downstream.

```
rank                                         0
youtuber                                     0
subscribers                                  0
video_views                                  0
category                                    46
title                                        0
uploads                                      0
country                                    122
abbreviation                               122
channel_type                                30
video_views_rank                             1
country_rank                               116
channel_type_rank                           33
video_views_for_the_last_30_days            56
lowest_monthly_earnings                      0
highest_monthly_earnings                     0
lowest_yearly_earnings                       0
highest_yearly_earnings                      0
subscribers_for_last_30_days               337
created_year                                 5
created_month                                5
created_date                                 5
gross_tertiary_education_enrollment_(%)    123
population                                 123
unemployment_rate                          123
urban_population                           123
latitude                                   123
longitude                                  123
dtype: int64
```

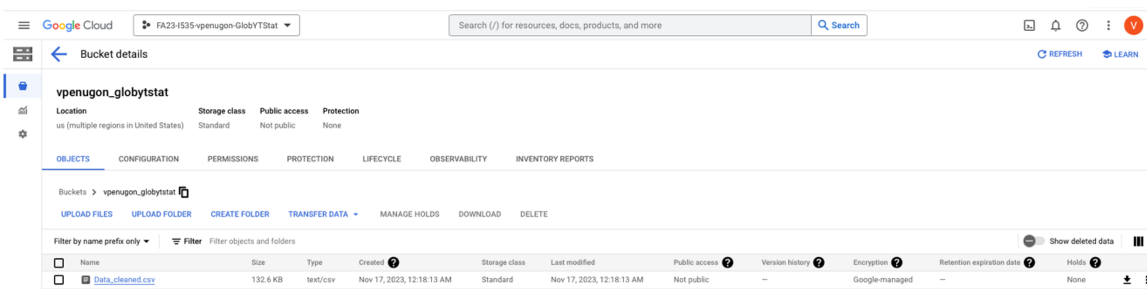
Raw Data

```
rank                                0
youtuber                            0
subscribers                         0
video_views                         0
uploads                             0
country                             0
channel_type                        0
country_rank                        0
channel_type_rank                   0
video_views_for_the_last_30_days    0
lowest_monthly_earnings             0
highest_monthly_earnings            0
lowest_yearly_earnings              0
highest_yearly_earnings             0
subscribers_for_last_30_days        0
population                          0
unemployment_rate                   0
dtype: int64
```

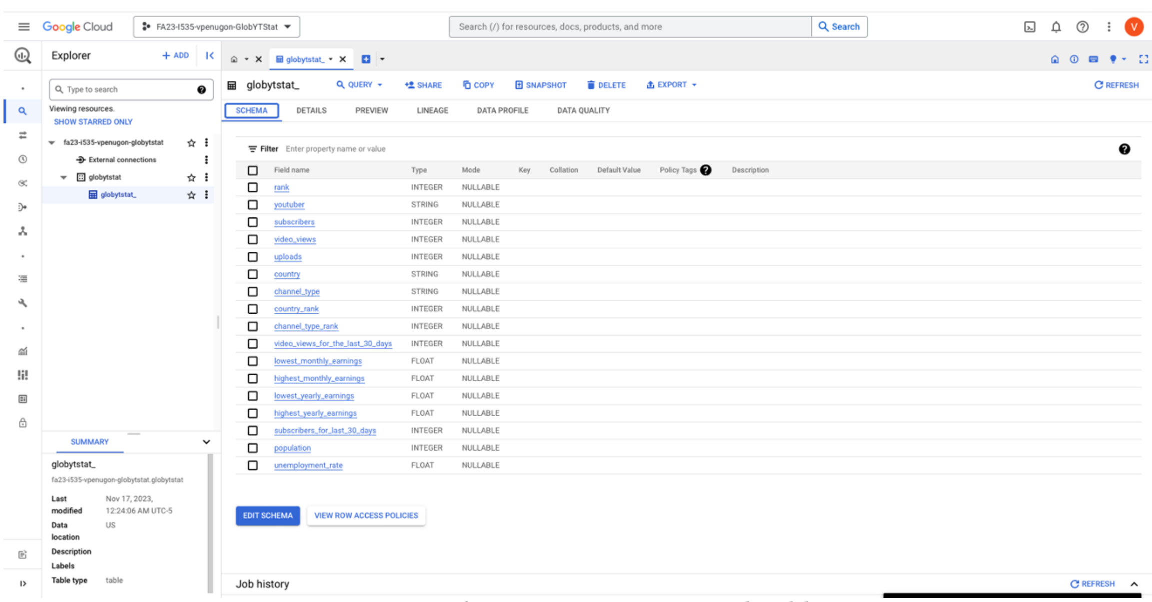
Clean Data

Data Ingestion & Storage:

Upon a clean, analytics-ready dataset, I constructed a cloud-based data architecture within Google Cloud Platform to centralize storage and access. Specifically, I created a Cloud Storage bucket and uploaded the cleaned CSV using the console UI. Then I used BigQuery to ingest this cloud-resident data into a managed SQL data warehouse under a custom YouTube schema. This established a data lake for hosting the resource.



Creation of a Bucket within Google Cloud Storage



Creation of a BigQuery Dataset and Table

Querying & Visualization:

To enable analysis and collaborative insights in Looker, I configured a persistent connection from Looker to the BigQuery schema I engineered. This allowed me to execute analytical SQL within BigQuery and visualize these queries with Looker while leveraging the speed and scalability benefits of BigQuery. I wrote SQL queries focused on aggregations, attribute filtering with the dataset. Charts, graphs provided dynamic visualizations like geographic maps and plots tailored to respond to my custom filtering and selections. The output was an interactive, visually rich analysis built on real-world YouTube channel performance statistics tuned to derive data-driven insights.

```sql
# Popular Categories and Upload Frequencies
SELECT channel_type, COUNT(*) as num_channels, AVG(uploads) as avg_uploads
FROM `fa23-i535-vpenugon-globytstat.globytstat.globytstat_`
GROUP BY channel_type;
```

Query results

JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTION D

Row	channel_type	num_channels	avg_uploads
1	Film	42	2729.857142857...
2	Entertainment	334	10016.85029940...
3	Music	216	1967.361111111...
4	Games	98	4076.255102040...
5	News	30	132971.6999999...
6	Comedy	51	1602.392156862...
7	Howto	36	2206.027777777...
8	Sports	13	12230.38461538...
9	People	101	1162.613861386...
10	Education	49	2986.836734693...
11	Tech	17	1987.058823529...
12	Nonprofit	2	102912.0
13	Autos	3	316.3333333333...
14	Animals	3	14289.33333333...

Query 1: Popular Categories and their Upload Frequencies

```sql
# Explore correlation between video views and highest monthly earnings by channel type
SELECT channel_type,
       SUM(video_views) AS total_views,
       AVG(highest_monthly_earnings) AS avg_highest_monthly_earnings
FROM `fa23-i535-vpenugon-globytstat.globytstat.globytstat_`
GROUP BY channel_type;
```

Query results

JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTION DETAILS

Row	channel_type	total_views	avg_highest_monthly
1	Film	459242749807	464607.1540476...
2	Entertainment	3619847727431	639672.0662874...
3	Music	3252736453212	571581.7175462...
4	Games	726607151373	352275.7192857...
5	News	311352817910	701639.9999999...
6	Comedy	453140574157	744474.7870588...
7	Howto	195650627890	235672.2222222...
8	Sports	181104971711	659715.3853846...
9	People	885919109792	675705.3651485...
10	Education	764805933171	754267.3469387...
11	Tech	59355536528	220358.8247058...
12	Nonprofit	10862911785	390400.0
13	Autos	19577546410	466670.0699999...
14	Animals	44135255601	2833333.333333...

Query 2: Relation between video views & monthly earnings

```sql
# Correlation among countries with highest unemployement rate and youtuber count
SELECT
    country,
    AVG(unemployment_rate) AS average_unemployment_rate,
    COUNT(youtuber) AS number_of_youtubers
FROM
    `fa23-i535-vpenugon-globytstat.globytstat.globytstat_`
GROUP BY
    country
ORDER BY
    average_unemployment_rate DESC;
```

Query results

JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTION DETAILS

Row	country	average_unemploym	number_of_youtuber
1	Jordan	14.72	3
2	Spain	13.96	22
3	Turkey	13.49	4
4	United States	13.17970542022...	435
5	Iraq	12.82	2
6	Brazil	12.08	62
7	Afghanistan	11.12	1
8	Egypt	10.76	2
9	Barbados	10.33	1
10	Italy	9.89	2
11	Argentina	9.79	13
12	Colombia	9.71	11
13	Andorra	9.279277522935...	1
14	Morocco	9.02	1

Query 3: Countries with high unemployment rate & their youtuber count

Untitled 5 RUN SAVE DOWNLOAD SHARE SCHEDULE MORE

```sql
# Channels with the most growth in subscribers in the last 30 days from each country along with their channel type
WITH ranked_channels AS (
    SELECT
        youtuber,
        country,
        channel_type,
        subscribers_for_last_30_days,
        ROW_NUMBER() OVER(PARTITION BY country ORDER BY subscribers_for_last_30_days DESC) as rank
    FROM
        `fa23-i535-vpenugon-globytstat.globytstat.globytstat_`
)
SELECT
    youtuber,
    country,
    channel_type,
    subscribers_for_last_30_days
FROM
    ranked_channels
WHERE
    rank = 1;
```

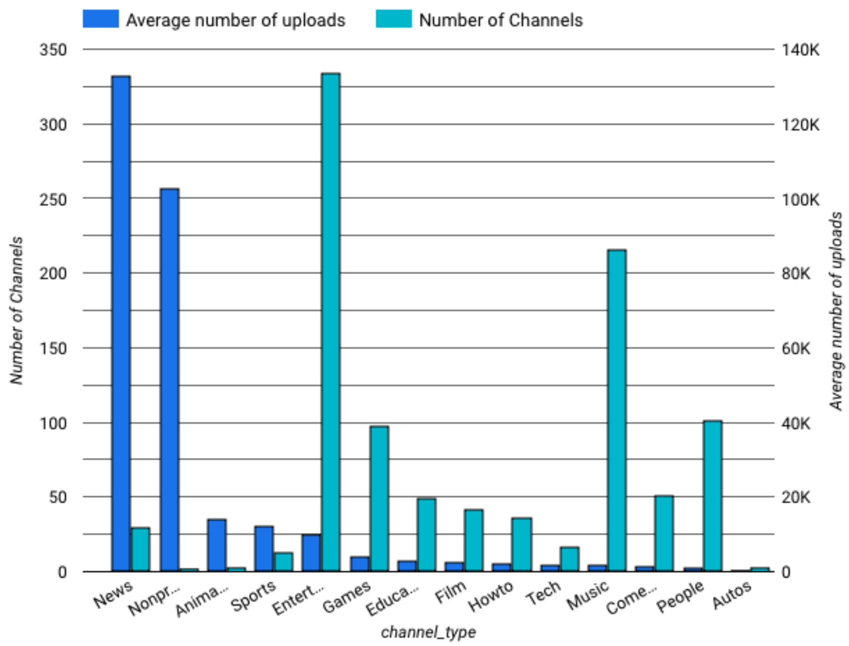
Query results

JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTION DETAILS EXECUTION GRAPH

Row	youtuber	country	channel_type	subscribers_for_last_30_days
1	LEGO	Singapore	Entertainment	200000
2	PANDA BOI	Italy	Entertainment	2000000
3	GMA Network	Philippines	Entertainment	500000
4	RaptorGamer	Ecuador	Games	300000
5	Alan Walker	Thailand	Film	349079
6	YOLO AVENTURAS	Venezuela	Comedy	100
7	Fernanfloo	El Salvador	Games	200000
8	MoreAliA	United Kingdom	Games	600000
9	Musas	Andorra	People	349079
10	A4	Cuba	People	10
11	[illegible]	Peru	Entertainment	349079

Query 4: Channels with the most growth in subscribers in the last 30 days from each country along with their channel type.

RESULTS:



Bar Plot for Popular Categories and their Upload Frequencies

The graph visually represents two sets of data: the average number of uploads and the number of channels for each category.

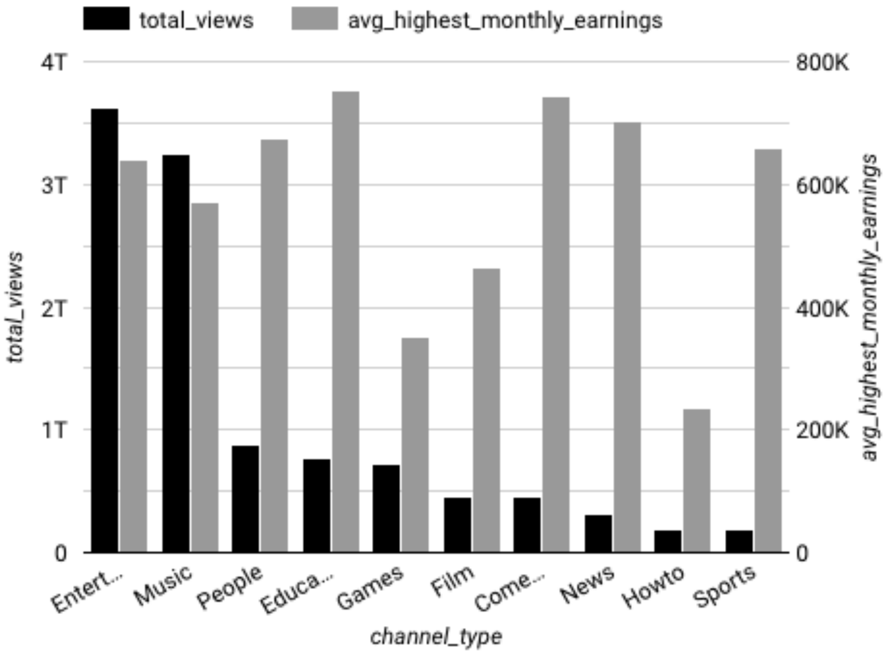
The Entertainment category has the highest number of channels, with 334 channels, shows a lower average upload frequency, indicating a larger but less active group of channels.

Music channels, numbering 216, has a low upload frequency. This might reflect the slow production of new music content. The Games category has a substantial number of channels (98) with a moderate upload frequency, indicating a steady content flow.

News channels, although numerous (30), have a relatively high average upload count, likely due to the ongoing nature of news reporting. In comparison, the How-to category, with a similar number of channels (36), has a significantly lower upload average, which might suggest content in this category requires more time to produce.

The Autos category, with 3 channels, has lower upload frequencies, suggesting niche content with less frequent updates.

Overall, the data suggests that categories with fewer channels tend to have higher upload frequencies, possibly due to specialization and dedicated audiences, while categories with more channels might face more competition, leading to varied upload frequencies.



Bar Plot to show the relation between Video Views & Monthly Earnings

Entertainment channels lead in total views, suggesting they are the most watched, but their average highest monthly earnings are not the highest, indicating that high viewership does not necessarily translate to the highest earnings. Music channels also have a high viewership, yet their earnings are relatively moderate compared to their views.

On the other hand, Gaming channels show a balance between views and earnings, suggesting a potentially more engaged or monetizable audience. The News category, while having fewer views compared to Entertainment and Music, shows higher average earnings, which could be due to more dedicated viewers or a better monetization strategy.

The How-to category shows a lower view count but maintains moderate earnings, hinting at niche content that could be highly valued by a specific audience. Sports channels, despite having a substantially lower number of views, reflect high earnings, which could be due to a variety of factors such as the type of content, monetization policies, or audience demographics.

In summary, the Entertainment category leads in both total views and earnings, followed closely by Music and People in views but with Music channels earning significantly more on average. This suggests that while Entertainment channels may have broader reach, Music channels likely have higher engagement or more lucrative monetization strategies. The Film category, although not the most viewed, commands substantial average earnings, indicating a potential for high revenue generation per view. The data underscore the potential profitability of content creation in certain genres on YouTube. The data also suggests that there is not a direct correlation between the number of views and earnings; some categories with fewer views have higher average earnings, and vice versa. While the viewer engagement and content type are significant factors in channel earnings, rather than just the total number of views.

	country	unemployment_rate ▼	youtuber
1.	Jordan	14.72	3
2.	Spain	13.96	22
3.	Turkey	13.49	4
4.	United States	13.18	435
5.	Iraq	12.82	2
6.	Brazil	12.08	62
7.	Afghanistan	11.12	1
8.	Egypt	10.76	2
9.	Barbados	10.33	1
10.	Italy	9.89	2
11.	Argentina	9.79	13
12.	Colombia	9.71	11
13.	Andorra	9.28	1
14.	Morocco	9.02	1
15.	Ukraine	8.88	8

1 - 49 / 49

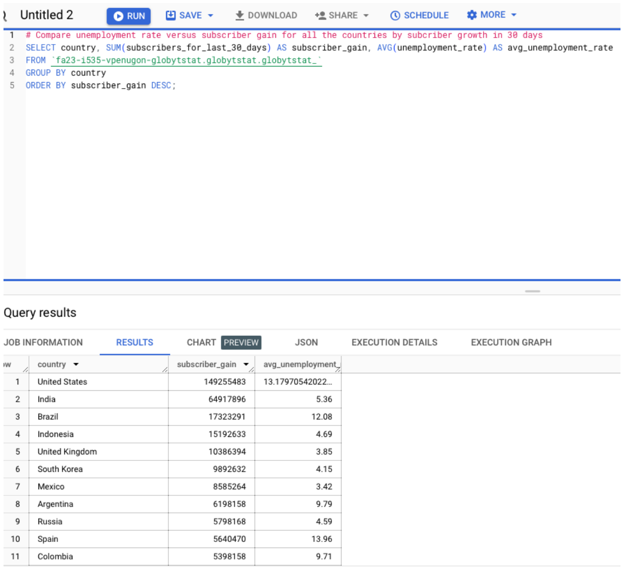
Table with a heatmap to show Countries with high unemployment rate & their youtuber count

The table with a heatmap illustrate the relationship between a country's unemployment rate and the number of YouTubers in that country. Notably, there is a visible diversity in the distribution, showing that a high unemployment rate does not necessarily equate to a higher number of YouTubers.

For instance, Jordan, with the highest unemployment rate listed (14.72%), has only a small number of YouTubers (3). Similarly, Spain, with a significant unemployment rate of 13.96%, has a comparatively modest count of YouTubers (22), suggesting a potential turn to content creation as an alternative income source amidst job scarcity. However, the United States has a moderately higher unemployment rate despite a significantly higher number of YouTubers, indicating a possibly more established or appealing market for content creators irrespective of the employment landscape, additionally factors other than unemployment may influence the prevalence of YouTubers, such as internet penetration, cultural trends, and economic factors that support content creation as a viable occupation.

The case of Turkey, with an unemployment rate close to that of Spain but far fewer YouTubers (4). Conversely, Brazil shows a lower unemployment rate of 12.08% but has a substantially higher number of YouTubers (62), indicating that a variety of socio-economic factors influence the decision or opportunity to become a content creator.

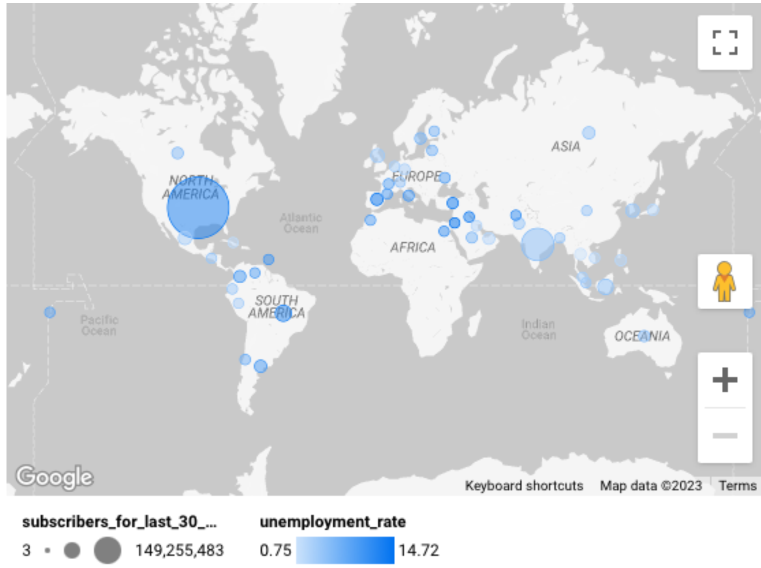
Overall, the data indicates that while there may be some correlation between unemployment rates and the number of YouTubers in certain contexts, this is not a consistent trend across all countries, and local conditions likely play a significant role.



```
# Compare unemployment rate versus subscriber gain for all the countries by subcriber growth in 30 days
SELECT country, SUM(subscribers_for_last_30_days) AS subscriber_gain, AVG(unemployment_rate) AS avg_unemployment_rate
FROM `fa23-i535-vpenugon-globytstat.globytstat.globytstat_`
GROUP BY country
ORDER BY subscriber_gain DESC;
```

Query results

ow	country	subscriber_gain	avg_unemployment_
1	United States	149255483	13.17970542022...
2	India	64917896	5.36
3	Brazil	17323291	12.08
4	Indonesia	15192633	4.69
5	United Kingdom	10386394	3.85
6	South Korea	9892632	4.15
7	Mexico	8585264	3.42
8	Argentina	6198158	9.79
9	Russia	5798168	4.59
10	Spain	5640470	13.96
11	Colombia	5398158	9.71



Query & Visual map to compare unemployment rate versus subscriber gain for all the countries by total subscribers

The map uses bubble size to represent the number of subscribers gain, with larger bubbles indicating a count of higher subscriber gain, and color intensity to indicate unemployment rates, with darker shades representing higher rates.

The United States stands out with a substantial subscriber gain and a relatively high unemployment rate, indicating a strong presence on YouTube possibly related to a mature digital market and diverse content creation ecosystem while ensuring access to internet for most of the unemployed citizens. India follows with a significant number of subscriber gain, despite a moderate unemployment rate, suggesting a high level of engagement with digital content and potential for growth in online viewership.

Indonesia and Brazil also show noteworthy subscriber gains, with unemployment rates varying from moderate to low, which may point to digital platforms being a viable alternative for income in the face of job scarcity. The United Kingdom, with a lower unemployment rate, maintains a strong subscriber base, reflecting its established media industry's adaptation to the digital space.

The data imply that while there's a trend of higher YouTube engagement in countries with elevated unemployment rates, suggesting a turn to content creation in tough economic times, there are also strong subscriber gains in countries with healthy employment statistics, indicating that the platform's reach is broad and not solely dependent on economic hardship.

DISCUSSION:

Interpretation of the Results:

This comprehensive investigation of the top 1,000 channels on YouTube yielded revealing insights into audience preferences, content strategies, and monetization trends on the world's dominant online video platform.

Starting with viewership patterns, entertainment-focused channels unsurprisingly led total video views, reflecting the mass appeal of funny, irreverent short-form content. However, despite accruing the most eyeballs, entertainment channels' earnings trailed more niche categories like music. This indicates entertainment drives volume while music and gaming channels cultivate intensely loyal, engaged subscribers. Specifically, while music commanded comparable aggregate views to entertainment, it generated markedly higher average monthly incomes - likely a result of dedicated fans willing to directly fund artists. Similarly, though lower in total viewership, gaming channels balanced healthy watch time with elevated earnings suggestive of an involved community.

Collectively, these trends underscore a dichotomy - wider reach through widely consumed entertainment versus deeper monetization payoffs in passionate verticals like music or gaming. This means creators should tailor content strategies and monetization approaches to their target audience - either broader, casual viewers or specialist superfans. Moreover, the data indicates that simply amassing views is not enough; converting visibility into subscriber buy-in is crucial for income.

Beyond viewership and income correlations, the analysis also highlighted motivating factors for channel creation across different countries. There was a weak correlation observed between unemployment rates and number of resident YouTubers. For instance, higher joblessness in Spain (13.96%) yielded marginally more YouTubers (22) than the vast American market. However, counterexamples like Brazil disproved this link. Despite lower unemployment figures (12.08%), Brazil boasted significantly more content creators (62), proving viability for such careers extends beyond macroeconomics. Irrespective of local conditions, engaged digital culture and infrastructure enable and incentivize creation.

Interpreting subscriber gains also conveyed heterogeneous trends. India and the US accrued the highest growth, pointing to maturing digital media ecosystems with vast reach. Meanwhile smaller nations like Indonesia and Brazil nevertheless charted steadier adoption, reflecting increasing internet proliferation lifting baseline viewership. This indicates that while economic factors can motivate initial conditions, continued technology access and shifting consumption habits reinforce participation at scale.

In summary, this revealing quantitative profile highlights crucial optimization strategies for YouTube creators seeking the winning formula - either broader entertainment appeal or intensely engaged community focus. It also demonstrates complex, evolving interplay amongst economic variables underpinning viewer habits. As online video continues unprecedented expansion towards a projected $335B value, updated tracking of channel performance dynamics can direct creators, advertisers, and policymakers alike.

Skills/Technologies Employed from INFO-I 535:

The analysis leveraged a robust technology stack spanning data cleaning, storage, processing and visualization to transform raw YouTube statistics into actionable insights. Python within a Jupyter notebook hosted on a J2 Virtual Machine provided the initial data wrangling environment. Python's extensive data science-oriented libraries like Pandas and NumPy enabled efficient standardization, imputation, filtering and type casting on the 1,000 row YouTube CSV dataset. Jupyter's interactive, document-oriented structure accelerated iterative data prep while the preconfigured J2 VM instance simplified setup complexity.

For durable storage and pipeline staging, Google Cloud Storage provided an infinitely scalable and secure data lake for persisting the cleaned CSV. Integration with BigQuery made this a seamless source for the analytics data.

BigQuery itself offered a fast, scalable SQL data warehouse for flexible analysis. As Google's serverless analytics engine, BigQuery scales to any data volume while optimizing costs by triggering compute per-query. This made it fast, efficient, and convenient for ad-hoc analysis.

Finally, Looker delivered the visualization layer with its unified semantic abstraction for rapidly building graphs, filters, and dashboards. As a dedicated BI tool, Looker provided white-glove optimization for analytics applications. The integrated stack offered end-to-end data preparation, storage, processing, and visualization tuned exactly to the demands of the quantitative YouTube channel analysis.

Together these technologies permitted efficient transformation of raw YouTube data at scale into actionable.

UCGS Data Lifecycle:

I acquired a raw dataset of global YouTube channel statistics from Kaggle containing performance metrics for the top 1,000 channels.

In the process phase, I cleaned this dataset in a Jupyter notebook using Python - standardizing identifiers, imputing missing values, subsetting relevant columns, deduplicating, and optimizing data types.

I then ingested the processed CSV into Google Cloud Platform, loading it into a BigQuery data warehouse under a custom YouTube schema. This established a centralized storage and querying capabilities. I subsequently configured an interface between BigQuery and Looker to visualize analytical SQL queries focused on aggregations and attribute filtering. The analysis yielded insights into audience preferences, content strategies, and monetization trends - such as entertainment driving views while music channels cultivated loyalty for earnings. Interactive charts conveyed geographic breakdowns in creator motivations as well as differences between channel visibility and income.

I interpreted correlations linking number of uploads, video types, and earnings.

Overall, the workflow followed the USGS data lifecycle approach - acquiring an authoritative dataset, processing it for integrity, establishing reliable storage and access, analyzing for trends, and deriving actionable conclusions to help optimize YouTube creator strategies.

The project output was an interactive, visually rich examination of real-world YouTube statistics tuned to provide data-backed insights.

Challenges Faced: Barriers and Failures Encountered During Insight Generation:

The key challenges faced in this YouTube data analysis project centered around data cleaning and transformation. First, missing values in the raw dataset had to be carefully handled through substitution to avoid skewing aggregations. Deduplication was also critical to prevent statistical distortion from repeat observations. Iterative scanning and type casting using Python and Pandas was required to optimize the CSV for analysis.

Establishing reliable ingestion into BigQuery and appropriately structuring the schema for SQL convenience involved certain troubleshooting. Configuring interactive Looker visualizations on top of the BigQuery data warehouse also required careful tuning. Most profoundly, formulating meaningful questions and translating them into analytical SQL queries that uncovered insights was an iterative process requiring exploratory mindset. While technologies like BigQuery and Looker accelerated flexibility, deriving value from the YouTube dataset ultimately required creativity and perseverance to clean, structure, query, interpret, and communicate findings effectively.

CONCLUSION:

This comprehensive investigation of real-world statistics for the top 1,000 YouTube channels worldwide yielded data-driven insights into audience preferences, content strategies, and monetization trends on the platform. By leveraging aggregated performance indicators across the most successful YouTube creators globally, the analysis highlighted crucial optimization approaches for existing and aspiring video producers.

One key finding was that broadly entertaining content tended to drive the most views, yet more niche categories like music and gaming fostered intensely loyal fans that translated into higher earnings. This indicates creators should tailor approaches based on either mass appeal or specialized community building. The investigation also revealed only moderate correlation between unemployment and YouTube participation across countries, proving viability transcends macroeconomics alone. Ongoing internet proliferation and shifting media habits enable creator ecosystems irrespective of local conditions.

Deriving these actionable conclusions required surmounting data cleaning and structuring challenges around missing values, duplication, ingestion, and query formulation. Ultimately creativity, flexibility and perseverance were essential to transform raw YouTube statistics into interactive visual insights.

This project exemplifies the power of data mining to decode online platform dynamics. Applying a rigorous quantitative approach elicited tangible lessons for optimizing video strategy and illuminated complex relationships between economic factors and user behavior. Updating the analysis over time could reveal evolving trends as the industry matures.

Overall, this project established a methodological template for extracting value from digital datasets. The data-driven insights uncovered around YouTube's global landscape demonstrate how statistical learning techniques can yield advantageous intelligence even from readily available public data. Extending this analytical approach across further internet platforms and bench marking against longitudinal shifts represent fruitful avenues for future investigation.

REFERENCES:

1. Dataset Link: https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023/data
2. Data Cleaning- https://www.analyticsvidhya.com/blog/2021/06/data-cleaning-using-pandas/
3. Google Cloud Platform BigQuery Documentation- https://cloud.google.com/bigquery/docs
4. Google Cloud Platform Cloud Storage Documentation- https://cloud.google.com/storage/docs
5. Google Looker Studio/Data Studio Documentation- https://cloud.google.com/looker/docs
6. UCGS Data Lifecycle- https://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf