Project Report

*On*

**YouTube Data Visualization Using Power BI and Python**

***Submitted in partial fulfillment of the requirement for degree of***

Bachelor of Technology

*In*

Computer Science and Engineering (Data Science)



Under the guidance of

**Dr. Narendra Kumar Kamila**

**HOD of CSE (DS)**

**Bhubaneswar**

Submitted By

Bibek Kumar Dalabehera,

Dinesh Kumar Sahu

Registration No.: 2201287579, 2201287588

4th Semester

*Department of Computer Science Engineering(DS)*

**GITA Autonomous College, Bhubaneswar**

**(Affiliated to Biju Pattanaik University of Technology, Odisha, Rourkela**

**CERTIFICATE**

This is certified that the entitled “**YouTube Data Visualization Using Power BI and Python**” submitted by Bibek Kumar Dalabehera and Dinesh Kumar Sahu bearing Registration number 2201287579 and 2201287588, 4th semester is approved in partial fulfillment of the requirement for completion of Bachelor of Technology in computer Science and Engineering (Data Science).

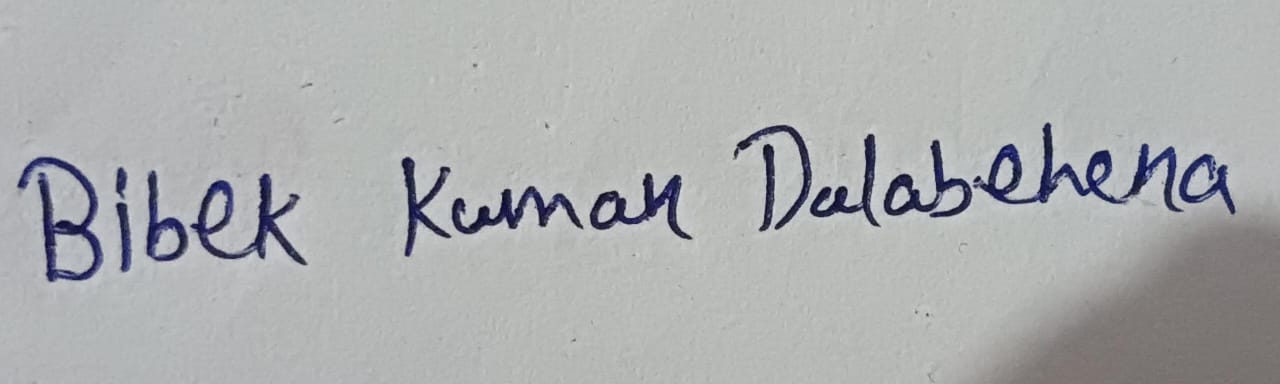
Dr. Narendra Kumar Kamila

(Guide)

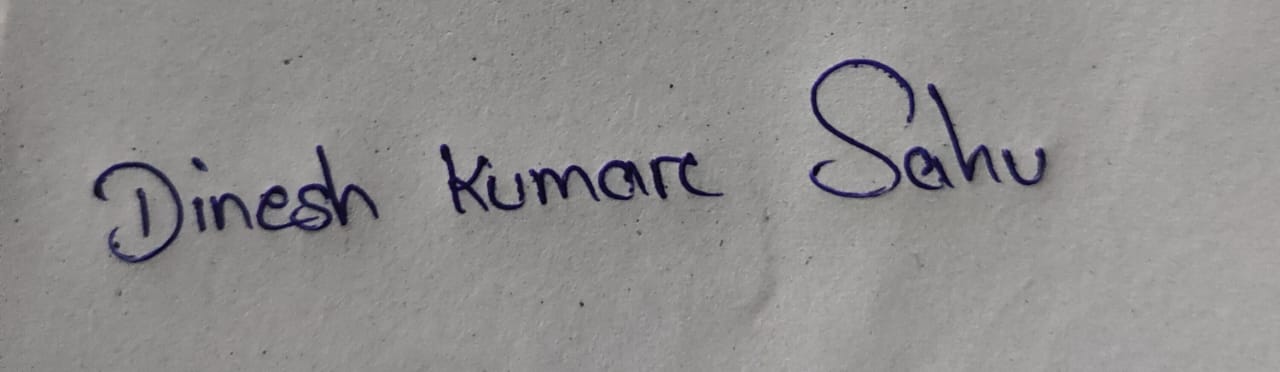
**Acknowledgement**

We would like to express our sincere gratitude to Dr. Narendra Kumar Kamila (HOD of CSE-DS) for their invaluable support and guidance throughout the duration of this project. Their expertise and encouragement were instrumental in shaping our ideas and refining our approach to YouTube data visualization. We are grateful for their mentorship, which significantly contributed to the successful completion of this endeavor. Lastly, we acknowledge each other, for our collaborative efforts and dedication to this project. Together, we navigated through the intricacies of data visualization and successfully translated our vision into reality.

Bibek Kumar Dalabehera



Dinesh Kumar Sahu



**ABSTRACT**

This project aims to perform a comprehensive data analysis and

visualization on YouTube data, focusing on various aspects such as total number of subscriber, views, likes, dislikes, view databases, and the popularity of YouTubers on world viewers. The dataset used for this project is derived from YouTube sources from 2010 to 2023.The ultimate goal of this project is to transform raw data into insightful visualizations that can aid in decision-making processes and provide a compelling narrative about the YouTube content creator. The dashboard integrates various visualizations, including pie charts, line charts, bar charts, slicers, and cards, to provide a comprehensive overview of engagement metrics, content performance trends, audience demographics, and geographic distribution of subscribers. Through dynamic filtering options and interactive features, stakeholders can explore the data and gain actionable insights to optimize content strategies, enhance audience engagement, and drive channel growth

TABLE OF CONTENTS

Chapter No. Topics Page No.

Chapter 1: Introduction 01

* PROJECT REVIEW
* OVERVIEW
* OBJECTIVE

Chapter 2: Details of project 02-03

* NEED FOR THE PROJECT
* ADVANTAGES FOR THE PROJECT
* SOFTWARE REQUIREMENT
* PROJECT SUMMARY

Chapter 3: Result 04

Chapter 4: CONCLUSION & REFERNCES 05

**Contents**

**CHAPTER 1: INTRODUCTION**

* **Project Review :**

In this project we build up the YouTube Data Visualization using Power Bi and Data cleaning using Python (Pandas&Numpy).Pandas is a python library used for data manipulation and analysis and Numpy is also a Python library is a fundamental package for numerical computing in Python, also used in pandas during manipulation and Power Bi is a software used for visualizing and analyzing data from various sources. With it help we visualize the data of YouTube and represent it insight of a dashboard.

* **Overview**

Data visualization is the practice of presenting data in a graphical or visual format, making complex information more accessible and understandable. Our data visualization project aimed to analysis focuses on the performance of YouTube content creators across different regions, examining metrics such as likes, dislikes, subscribers, comments, and views to understand audience engagement and content reception, leveraging the capabilities of Power BI for visualization and Pandas for data cleaning and preprocessing.

* **Objective**

The primary objectives of our project were:

1. Understand the performance and popularity of individual You Tubers in different regions.
2. Identify trends and patterns in audience engagement and content reception across various countries.
3. Assess the effectiveness of Assess the effectiveness of content strategies and audience targeting by comparing performance metrics among different YouTubers and countries audience targeting by comparing performance metrics among different YouTubers and countries.
4. Inform decision-making processes for content creators, marketers, and platform administrators to optimize content creation, distribution, and audience engagement.

**CHAPTER 2: DEATAILS OF PROJECT**

* **Need for this project:**

1. Analyzing metrics helps comprehend how viewers interact with content, aiding in tailoring future videos to audience preferences.
2. By examining data across creators and regions, trends are identified, successful strategies recognized, and areas for improvement pinpointed.
3. Insights empower creators to refine their content approach, adapting formats or topics to better resonate with their target audience.
4. Data-driven decisions on content promotion and localization enable creators to expand their audience globally, leveraging regional preferences for growth.

* **Advantages for this project:**

1. This project enables informed decisions by providing detailed performance insights per YouTuber and country.
2. Analysis helps creators tailor content to audience preferences, enhancing engagement and viewership.
3. Understanding regional trends allows creators to expand their reach and connect with diverse audiences worldwide effectively.
4. Analyzing performance metrics uncovers opportunities for improvement, providing a competitive edge in content creation and distribution strategies.

* **Software Requirement:**

1. Power BI is used for creating interactive visualizations and dashboards to analyze the data
2. Specifically, the Pandas library is used for data cleaning and manipulation.
3. dropna(), fillna, astype(),replace(), rename(), isnull() are some pandas function used for data cleaning.
4. np.isnan(),np.nan\_to\_num(), np.isin() are some methods are used for data manipulation.
5. And many other method are used for manipulation.
6. Jupyter Notebook is used for write and execute Python scripts for data preprocessing and analysis.
7. Access of dataset containing metrics such as likes, dislikes, comments, subscribers, views, YouTuber names, and country information.

* **Project Summary:**

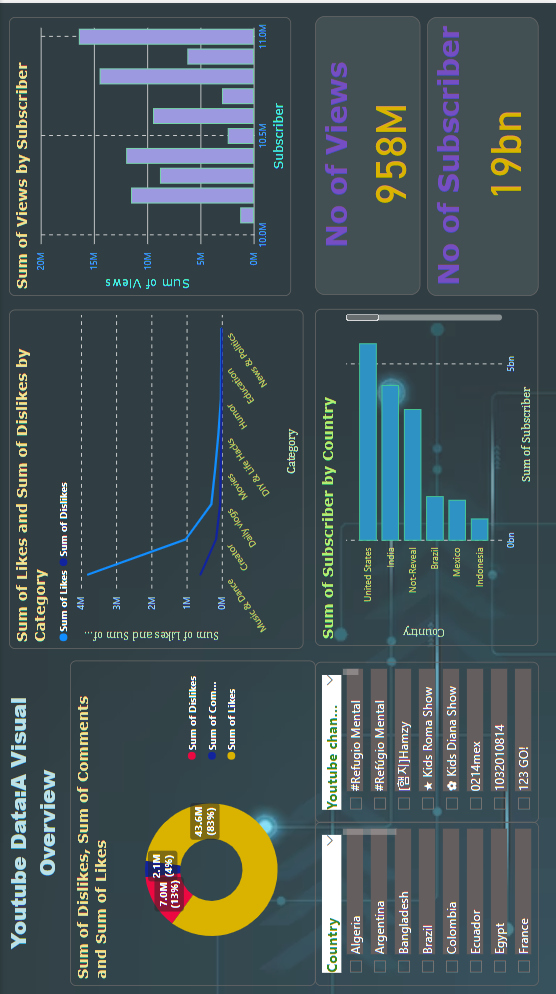
Our project describe these following features:

1. A pie chart which contain sum of likes, dislikes, comments displays the overall engagement metrics, providing a quick overview of audience interaction with the content.
2. The line chart which have sum of like and dislikes by categories visualizes the trend of likes and dislikes over time or across different categories, allowing for deeper analysis of audience preferences and content performance.
3. A bar chart of subscriber and view illustrates the relationship between subscribers and views, helping to understand the level of engagement relative to the size of the audience.
4. A horizontal bar chart of sum od subscriber by country provides insights into the distribution of subscribers across different countries, enabling targeted marketing strategies and content localization efforts.
5. The two cards which are Number of view and number of subscriber presents key metrics such as total views and subscribers in a concise format, offering immediate visibility into the overall performance of the channel or content.
6. Two Slicers country and Youtuber names allows users to filter the data dynamically based on specific countries or individual YouTuber names, facilitating focused analysis and comparison.

By combining these visualizations and interactive elements, the dashboard provides a comprehensive overview of YouTube channel performance, audience engagement, and channel growth.

**CHAPTER 3: RESULT**

The final result of our project is:



**CHAPTER 4: CONCLUSION**

In conclusion, we have learned the data visualization and analysis skills, optimized content strategies, and gained insights into audience preferences and geographic trends and regional preferences enhancing our proficiency in Power BI and interdisciplinary collaboration.

The project could summarize the key findings and insights derived from the analysis presented in the Power BI dashboard. It could also discuss the significance of these findings and their implications for content creators, marketers, and other stakeholders. The conclusion may discuss how effective the dashboard was in providing useful insights and its potential for guiding future decisions.

**REFERENCES:**

1. Google – [www.google.com/](http://www.google.com/)
2. Official Kaggle website - <https://www.kaggle.com/>
3. Python Pandas Docs - <https://pandas.pydata.org/docs/>
4. Python Numpy Docs - <https://numpy.org/doc/>