

# Data Management and Data Analytics Capstone Topic Approval Form

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of the following areas so you will have a complete and realistic overview of your project. Your course instructor cannot approve your project topic without this information.

**Student Name:** [REDACTED]

**Student ID:** [REDACTED]

**Capstone Project Name:** YouTube Advertisement Project

**Project Topic:** Analyzing YouTube data to decide what channels to target for promoting your advertisements.

**Research Question:.** What YouTube channels worldwide would be the most beneficial to target for selling the marketing companies advertisements and what factors have a positive effect on the amount of subscribers and views that a channel generates?

**Hypothesis:** Null Hypothesis: There is no correlation between the number of views a channel generates and the target audience of that channel. Alternative Hypothesis: There is a correlation between target audience and the number of views and subscribers a channel generates.

**Context:** This situation would benefit from data analysis because the marketing company looking to sell their ads to youtubers and as a seller of a product it is important to look for the most efficient way generate the most revenue. Data analysis will help the seller be able to know which channels would be best to target based on the popularity, niche and location of the YouTube channels content. The analysis will help the marketing company match the target audience of their advertisements with the audience of specific YouTube channels. As a marketing company it is important to have the largest audience seeing the ads you provide. If the viewers see an advertisement that is similar to their interests, they will be more inclined to click on the ad which in turn is the main goal of the marketing company.

**Data:** I will need to collect global YouTube data with fields such as channel name, channel niche or category, number of subscribers, number of views and channel rank to successfully answer my business question.

I will be using a dataset called Global YouTube Statistics 2023 that is provided by Kaggle.com. This data set consists of 27 columns and 996 rows that contains the necessary fields to answer my business question. The dataset can be found here: <https://www.kaggle.com/datasets/nelgiryewithana/global-youtube-statistics-2023>

This dataset is provided by Kaggle and is accessible as unrestricted data to Kaggle users. It is downloadable as a zip file directly from the site page.



**Data Gathering:** The Data I will be using is accessible to me from Kaggle by download. I plan to use Jupyter notebooks to complete the data gathering step in this project. I will use python and Pandas library to import the dataset as a CSV file using a Jupyter notebook where I can view, clean, and analyze the data to answer my business question.

**Data Analytics Tools and Techniques:** I plan to use Jupyter notebooks to gather, clean and analyze the dataset. I will be using python as the language utilizing Pandas, Matplotlib, Numpy, Seaborn Libraries and Tableau to compose statistical analysis as well as create plots that validate that analysis.

**Justification of Tools/Techniques:** *Explain why the data-analysis technique you chose is an appropriate technique to analyze the data collected.* I will be using these Jupyter notebooks utilizing these Python libraries for varying reasons. Pandas and Numpy libraries will be used to import my dataset into my Jupyter notebook and allow me to clean the dataset as well as perform statistical analysis of the dataset that will be required to answer my business question. The Matplotlib will be used to create visualizations that help me answer my business question and validate my statistical analysis. Tableau will be used to compare fields from my dataset to create visually appealing visualizations that help me look at the data fields in different ways and create a final presentation that is appealing and easy to understand.

**Application Type, if applicable (select one):**

- ☐ Mobile
- ☒ Web
- ☐ Stand-alone

**Programming/Development Language(s), if applicable:** Python

**Operating System(s)/Platform(s), if applicable:** Windows 10

**Database Management System, if applicable:** NA

**Project Outcomes:** I anticipate learning that the YouTube channels with the most subscribers and views are correlated with some fields more than others. I believe that the niche or category of a YouTube video is going to have an impact on the amount of views/subscribers. I also believe that we may gain some interesting results from this data about the success of the YouTube channel in relation with the geographical location of the YouTuber.

**Projected Project End Date:** [REDACTED]/2023

**Sources:** Nidula Elgiriye withana, 2023, Global Youtube Statistics 2023,  
<https://www.kaggle.com/datasets/nelgiriye withana/global-youtube-statistics-2023>



**Human Subjects or Proprietary Information**

Does your project involve the potential use of human subjects? (Y/N): N

Does your project involve the potential use of proprietary company information? (Y/N): N

---

**STUDENT SIGNATURE**

\_\_\_\_\_

**By signing and submitting this form, you acknowledge** that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

---

**TO BE FILLED BY A COURSE INSTRUCTOR**

**The capstone topic is approved by a course instructor.**

**COURSE INSTRUCTOR SIGNATURE:**



Jim Ashe, Ph.D. Mathematics

**COURSE INSTRUCTOR APPROVAL DATE:**

Monday, \_\_\_\_\_, 2023

**Project Compliance with IRB (Y/N): Y**

