Milestone 1:
Form a Team and Outline a Goal

Write down the names of the people in your group and your option choice.
The group members for team 8 are Ali Shair, Krish Patel, and Shivam Patel. We chose option B talking about the usage of Youtube as a content creator.

Describe
what data you plan to use, what insights can be derived from this data, and why you think the listed insights are important.
Our main goal is to really dive into YouTube data to make our channel better and attract more viewers. Since the online world is getting more crowded, using data to make smart decisions is key, especially on a big platform like YouTube.
To get started, we're looking at three important sets of data. At the top of the list is the "YouTube Channel and Influencer Analysis" (Data 1). This set will show us what makes a YouTube channel popular and what types of videos people like to watch. By understanding this, we can create videos that our viewers will love and want to watch again and again. Using this information, we can plan our videos more wisely. We'll know what our viewers like and can create more of that type of content. This way, our current viewers will stick around, and we'll also bring in new ones. By combining our creativity with what the data tells us, we can make sure our videos reach as many people as possible.

Describe how do you plan to design the database to store this data and explain your rationales?
The team has decided to work with a dataset we're calling Data 1. After checking out the information inside, we noticed that it could be split neatly into four main parts. So, we're going to organize it into four tables: one for videos, one for creators, one for channels, and one for interactions (like comments or likes).
When we looked closely at the dataset, it made sense to split the information this way because each category had its own set of details. For example, the video table would have things like video titles and lengths, while the creator table would have names and maybe contact info. Now, to make sure our database runs smoothly and everything is connected properly, we needed something unique to link everything together. We chose the Creator's name as this special link because no two creators can have the exact same name. This unique name will act like a passport, helping us connect details across the different tables. By using the Creator's name in this way, it's easier for us to find and organize all the related information when we need it.

Outline the tables that you plan to work with, and describe the column names, their corresponding data types and constraints, and the relationships between the tables.

For Data 1, the information that we would need would be split into 4 tables. The table names would be Video, Creator, Channel, and Interaction.

Video table:

- **Link(string)**
- **VideoViews(integer)**
- **Title(string)**
- **Language(string)**
- **Quality(Integer)**

Creator Table:

- **Creator Name(string)**
- **Creator Gender(string)**
- **Link(string)**

Channel Table:

- **SubCount(integer)**
- **ChannelViews(integer)**
- **NumVideos(integer)**
- **NumPlaylist(integer)**
- **Link(string)**

Interaction Table:

- **Numlikes(integer)**
- **NumComments(integer)**
- **CommunityEngagement(integer)**
- **Link(string)**

We believe that Creator's Name should be the Primary Key because you cannot have the same name as another creator. The creator table would be linked to the channel and video table. For the video table, the primary key will be the video link, because they are all unique. The video table will be connected to the interaction table.

Describe your teamwork: how did you make the decision about the project goal, list the contribution of each team member.

There was a unanimous decision for all of us to do option B. We all looked at the milestone prior to the meeting, and we all selected that option. Ali talked about the different datasets and the insights they may provide. Krish was in charge of outlining the tables that we plan to work with and describing the column names, the data types, and the constraints that

they may have. Shivam was in charge of explaining how we would design the database to store all of the data we have.