Data Wrangling

1. Use general data wrangling method to choose the data in the primary data set that contain the following information: year = 2019, country, Youtuber, Subscribers, Video Views, Created\_year
2. Join the data with primary and secondary data set, use country as the link between two tables.
3. Use machine learning method to show the graph for the first question. Like this

A map of the world

Description automatically generated

1. Use a spread or gather operation (or equivalent) to deal with the data
2. Use transformation to deal with the primary data, such as to find the max() for each year

Use linear regression to find the relationship between subscriber and video views (personally I want to do one more this, which can help us analyze the data further)

Data Visualization:

1. Use different gems to represent the data that is chosen from the data wrangling
2. Use layer graph to show the data for both questions
3. decision tree to analyze the primary data set. Category —> Video views —> country