Project Proposal

Team: Mohammed Abuelwafa – Mina Ashraf – Marwa Sulaiman – Sherif Gabr

a) Trending YouTube videos dataset

Proposed Business Problem:

- Estimating YouTube and creator profit and video popularity by estimating the number of likes, comments, and views based on the video characteristics.

Problem type: Regression Problem.

We have a dataset that has statistics about YouTube videos for different countries. The dataset has 16 columns which is a mix of categorical values and numerical values. Moreover, the avg number of samples for each country is around 25K. hence, we want to predict the number of likes, comments, and views for around two countries. The problem of trying to estimate the profit and know if a video will become trendy and cover the production costs is an important problem for video creators. Hence, we are addressing a data-drive optimization problem where we predict the mentioned response variables and decide whether the video should be produced or not given the client's budged in addition to trying to come up with the perfect video recipe.

Dataset link: https://www.kaggle.com/datasnaek/youtube-new?select=CAvideos.csv

b) Global Commodity Dataset

Proposed Business Problem:

- Predict Exports and exports for a specific country (Egypt for example)

Problem type: Time series analysis Problem.

In the mentioned dataset, we have time data and location data about commodity. Hence, given historical data, we can predict the imports and exports Trade in USD value of the country. Hence, we can estimate the financial situation of the country and plan accordingly.

Dataset link: https://www.kaggle.com/unitednations/global-commodity-trade-statistics