EXPLORATORY DATA ANALYSIS OF A YOUTUBE SHORTS DATASET
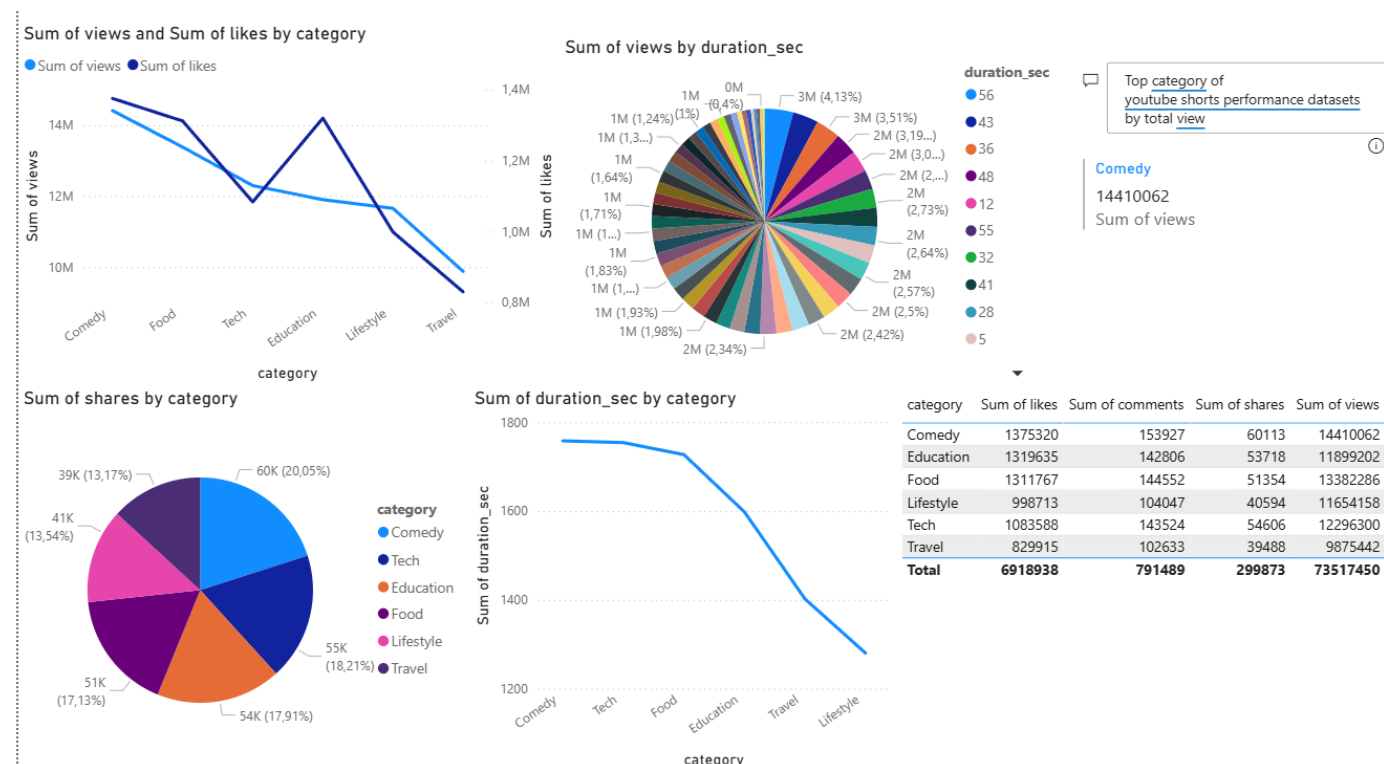
ASSIGNMENT:

The objective is to analyse a dataset from Kaggle showing different YouTube short contents. Explore and derive insights from the data that shows the engagement levels from different categories; to see how different categories perform in terms of engagement

Parameters that will be included:

- Sum of views: to see how many viewers the content reaches.

- Sun of duration: to determine if the duration of the short influences the engagement.

- Type of content: to discover the engagement of each content.

- Sum of likes: to see how each content is received

- Sum of shares: to determine the engagement of each content

Presenting data:

The project will make use of Power Query to transform and clean the data. Power BI will be used to visualize the data and create Dashboards showcasing the engagement of different categories.

category	Sum of likes	Sum of comments	Sum of shares	Sum of views
Comedy	1375320	153927	60113	14410062
Education	1319635	142806	53718	11899202
Food	1311767	144552	51354	13382286
Lifestyle	998713	104047	40594	11654158
Tech	1083588	143524	54606	12296300
Travel	829915	102633	39488	9875442
Total	**6918938**	**791489**	**299873**	**73517450**

Findings from the data:

- Discovered that Comedy has the highest number of likes, views and shares amongst the other categories

- Comedy videos have a longer duration of seconds and Lifestyle content which has the lowest

- Education content produces the most engagement amongst all the categories with a drastically large number of likes compared to the view count.

- Comedy and Tech content produce high engagement with higher likes compared to views.

- Lifestyle, and travel content have less likes than suggesting that these categories produce low engagement.