ACCENTURE VIRTUAL EXPERIENCE PROGRAM BY FORAGE

1. Introduction

This is a virtual internship program case study with the company, Accenture. This program is hosted through the site Forage and enabled me to leverage my skills and tools as a Data Analyst in a real-world setting.

2. Internship Company

Accenture is a global professional services firm that provides consulting, technology, and outsourcing services to a wide range of industries. Its services are focused on helping clients to improve their performance and achieve their goals by leveraging technology and innovation, with expertise in digital transformation, technology integration, and management consulting. Accenture also provides a range of outsourcing services such as application services, infrastructure services and business process services.

3. Scenario Company

The client company is Social Buzz. Social Buzz was founded by two former engineers from a large social media conglomerate, one from London and the other from San Francisco. Social Buzz emphasizes content by keeping all users anonymous, only tracking user reactions on every piece of content. There are over 100 ways that users can react to content, spanning beyond the traditional reactions of likes, dislikes, and comments.

They have scaled quicker than anticipated and need the help of an advisory firm to oversee their scaling process effectively. Due to their rapid growth and digital nature of their core product, the amount of data that they create, collect, and must analyses is huge. Every day over 100,000 pieces of content, ranging from text, images, videos, and GIFs are posted. All this data is highly

unstructured and requires extremely sophisticated and expensive technology to manage and maintain.

To start our engagement with Social Buzz, Accenture is running a 3-month initial project to prove to them that we are the best firm to work with. They are expecting the following:

- An audit of their big data practice
- Recommendations for a successful IPO
- An analysis of their content categories that highlights the top 5 categories with the largest aggregate popularity.

4. Dataset Summary

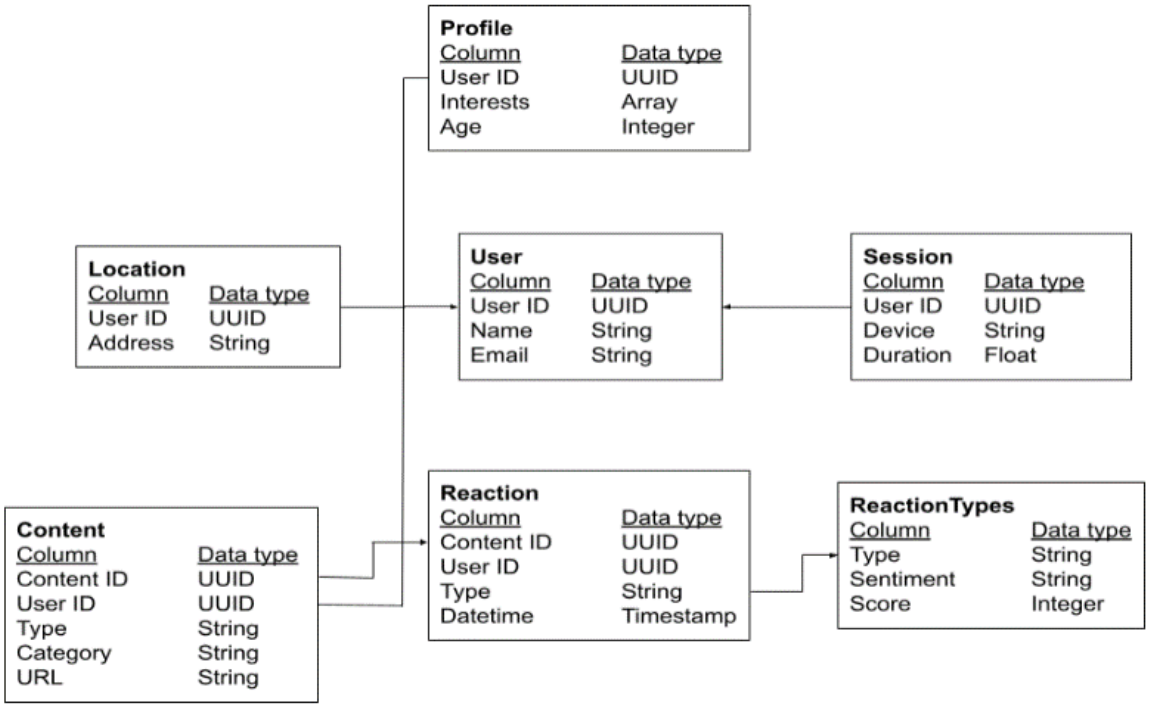
The client provided 3 datasets:

- **Content**
- **Reactions**
- **Reaction Types**

Here is the summary of the contents and columns from the datasets:

- **Content**
    - ID: Unique ID of the content that was uploaded (automatically generated)
    - User ID: Unique ID of a user that exists in the User table.
    - Type: A string detailing the type of content that was uploaded
    - Category: A string detailing the category that this content is relevant to
    - URL: Link to the location where this content is stored.
- **Reaction**
    - Content ID: Unique ID of a piece of content that was uploaded.
    - User ID: Unique ID of a user that exists in the User table who reacted to this piece of content.
    - Type: A string detailing the type of reaction this user gave
    - Datetime: The date and time of this reaction
- **Reaction Types**
    - Type: A string detailing the type of reaction this user gave
    - Sentiment: A string detailing whether this type of reaction is considered as positive, negative, or neutral
    - Score: This is a number calculated by Social Buzz that quantifies how “popular” each reaction is. A reaction type with a higher score should be considered as a more popular reaction.

They also provided us with the data model:



5. Data Preparation

We need to ensure that the data is cleaned before the analysis to provide a high-quality analysis that has less errors. We used Excel to clean the datasets with the following steps:

- Removed rows with missing values.
- Assigned the right data types to columns.
- Removed unnecessary columns that is not needed in the analysis.
- Combined the datasets using VLOOKUP formula.

6. Tasks

6.1 Task 1: Project Understanding

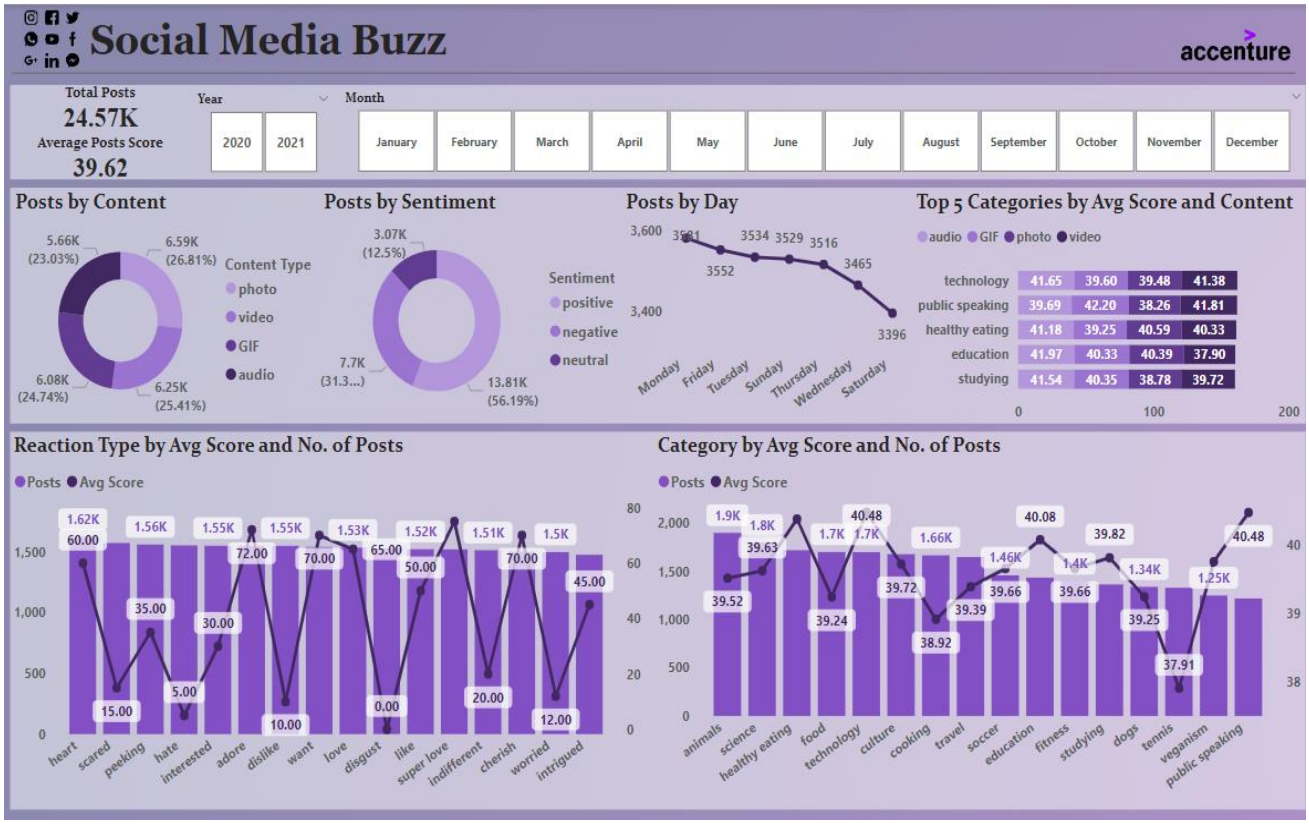
In this task, we are briefed about the client's company and talked about their business problem, project requirements and introduced the team that we're going to work with. We are then tested to make sure we understand the information about the client and the team.

6.2 Task 2: Data Cleaning & Modelling

In this step, we are focused on gathering the requirements needed for the analysis, cleaning those data, and doing proper modelling of the data. The client wanted to figure out which of the content categories has the highest aggregated popularity. We are also tested on information about the data. The data preparation and modelling steps are provided on the fifth part of this case study.

6.3 Task 3: Data Visualization & Storytelling and Task 4: Present to the Client

We are tasked with bringing the data to life in this part of the project. We summarized what we did in the project from start to finish to put in a presentation to show our client. Power BI is what I used to make charts and visualizations that can provide insights and recommendations for the client to act upon. Here are the insights we discovered during the exploratory phase of the analysis:



7. Summary and Recommendations

We know that the client wants to understand the top 5 content categories. They are Animals, Science, Healthy Eating, Technology and Food. Among the top content categories, Animals and Science are notable. The users like content related on Food/Healthy Eating and Science/Technology. The platform can then use this information to incentivize content creators with the same categories to boost audience engagement. Also make sure that the content produced is balanced all throughout the platform for when there is one content category appearing with most of the content posted. The bottom categories should also be focused to prevent that.