VIRTUAL INTERNSHIP CERTIFICATE

accenture

Forage

Inspiring and empowering future professionals

Ahad Rehman

Data Analytics and Visualization Virtual Experience

Certificate of Completion

July 6th, 2023

Over the period of June 2023 to July 2023, Ahad Rehman has completed practical tasks in:

Project Understanding
Data Cleaning & Modeling
Data Visualization & Storytelling
Present to the Client

Caroline Dudley
Managing Director
North America
Recruiting

Tom Brunskill
CEO, Co-Founder of
Forage

Enrolment Verification Code CY7uZNAjmALQ5ap84 | User Verification Code L4F6Q5gpiTAPAcBDY | Issued by Forage

ABSTRACT

Visualisation and data analytics are essential steps in the data analysis process.

Data analytics is the study, purification, transformation, and modelling of data to find patterns, insights, and trends.

The process of visualising data involves using graphs, maps, charts, and other graphical elements to portray it. It facilitates the clear and natural transmission of insights and patterns.

Forage's publicly accessible Accenture Data Analytics Virtual Experience Programme was something I took part in. Where I worked as a data analyst to assist "Social Buzz" in analysing their data and understanding how to make use of their enormous amount of data effectively.

1. ABOUT ORGANIZATION

Accenture's services and solutions include data analysis as a key part of their offering. Accenture uses data analysis to assist businesses in deriving useful insights from their data, making wise decisions, and achieving business goals. Accenture's data analysis capabilities are supported by a combination of domain knowledge, technology know-how, and data science talents. In order to produce efficient data analysis solutions, they work closely with clients to comprehend their unique business requirements, make use of state-of-the-art analytics tools and processes, and implement best practises.

Here is a summary of Accenture's use of data analysis:

1.1 Data Governance and Strategy: Accenture assists clients in creating data strategies that complement their corporate objectives. For the purpose of ensuring the accuracy, consistency, and security of data for analysis, it is necessary to define data governance frameworks, data quality standards, and data management practises.

1.2 Data Preparation and Integration- Accenture helps with data preparation and integration, including the integration of both structured and unstructured data from a variety of sources. They use data integration techniques to make sure the data is consistent, clean the data to get rid of mistakes or duplicates, and convert the data to get it ready for analysis.

1.3 Advanced Analytics and Modelling: Accenture makes use of advanced analytics methods including statistical analysis, machine learning, and predictive modelling to find patterns, trends, and correlations in data. In order to forecast outcomes, pinpoint risks, improve workflows, and enhance decision-making, they create models.

1.4 Business Intelligence and Performance Management- Accenture helps businesses put business intelligence solutions in place so that data-driven decision-making is possible. Business intelligence and performance management. Based on data analysis, they create scorecards, KPI dashboards, and performance management frameworks to track and assess corporate performance.

1.5 Industry-specific Analytics Solutions: Accenture offers solutions for many industries, including banking, healthcare, retail, manufacturing, and telecommunications. These solutions are tailored to meet the demands of these industries. In order to provide useful insights, these solutions take into account particular problems and draw on professional knowledge of the sector.

Accenture's data analytics and visualisation training programme is created to meet the varied needs of staff members from a range of experience levels and business disciplines. Data gathering, data cleaning, data analysis methods, and insight interpretation are just a few of the topics covered in the curriculum. In order to successfully communicate findings through attractive visual representations, participants are also taught to various data visualisation tools and approaches. The training course employs a blended learning methodology that combines both theoretical and practical elements. While interactive workshops and hands-on activities provide participants the opportunity to put principles into practise, instructor-led sessions give them a firm foundation of knowledge. In order to build a problem-solving mindset and promote effective teamwork, collaborative team activities imitate the pressures and dynamics of the workplace.

2. INTRODUCTION

Social media platforms have developed into data troves in the current digital era that can offer priceless insights into customer behaviour, market trends, and public mood. Organisations seeking to glean useful insights from the massive volumes of available social media data must make use of the capabilities of data analytics and visualisation tools. Accenture has put in place a specialised training programme on data analytics and visualisation for its workers working on the "Social Buzz" project in recognition of the importance of these abilities.

The goal of the Social Buzz project is to better understand customer preferences, sentiment analysis, and brand perception by analysing and interpreting social media data. The goal is to give clients useful information that will help them make strategic decisions, optimise their marketing efforts, and improve the perception of their brand as a whole. To accomplish this, Accenture has created a thorough training programme that gives its staff the knowledge and abilities to successfully analyse social media data and visualise crucial discoveries.

The training course starts out by giving participants a foundation in data analytics and introducing them to key ideas like data gathering, data cleansing, and data transformation. Employees gain knowledge of numerous data analysis techniques, such as sentiment analysis, topic modelling, and social network analysis, through practical exercises and real-world case studies. These methods provide them the ability to glean insightful information from social media data, spot trends, and detect patterns.

Simultaneously, the training programme focuses on data visualisation, teaching staff how to turn large amounts of data into visually appealing and informative representations. The use of interactive dashboards, heat maps, and network diagrams are just a few examples of the numerous data visualisation tools and methods that are introduced to the participants. Through appealing visualisations that clearly and entertainingly present the story underlying the data, they learn how to successfully communicate findings.

3. LITERATURE SURVEY

Tukey's "Exploratory Data Analysis" (1977) [1], which promoted visualisation, exploratory modelling, and robust statistics, revolutionised data analysis. It continues to be a fundamental work that emphasises the significance of comprehending data through graphical methods and repeated investigation.

A key publication introducing the ggplot2 package in R is Wickham's "ggplot2 [2]: Elegant Graphics for Data Analysis" (2016). It is a helpful resource for academics and data analysts since it offers a thorough tutorial on how to make excellent, aesthetically pleasing visualisations for data exploration and analysis.

C. O. Wilke (2019).[3] Basics of Data Visualisation: A Guide to Creating Insightful and Powerful Figures. Media by O'Reilly. The key manual for producing powerful data visualisations is Wilke's book. It serves as an important resource for data analysts and scientists since it discusses the ideas, methods, and best practises for creating relevant and aesthetically pleasing figures.

E. R. Tufte (2001). Graphics Press, "The Visual Display of Quantitative Information."[4] Tufte's work, which is regarded as a classic in the field of data visualisation, examines the rules and methods for successfully visualising data. It is a useful tool for creating insightful and captivating data visualisations.

W. S. Cleveland (1993). Data Visualisation, Hobart Press.[5] Cleveland's book, which offers insights into methods for developing successful visual representations of data, is a foundational work in data visualisation. It encompasses ideas like the design of clear and informative visual displays and graphical perception.

Inselberg (2009).[6] Parallel Coordinates: Applications of Visual Multidimensional Geometry. Science & Business Springer. The idea of parallel coordinates is presented in Inselberg's book as a potent visualisation method for deciphering and analysing high-dimensional data. It is a useful tool for data analysts and researchers since it examines the theory, procedures, and applications of parallel coordinates.

J. J. Caban, D. Gotz, and T. Munzner (2017).[7] How to read a research report on visualisation. 23(1), 391-400, IEEE Transactions on Visualisation and Computer Graphics. This essay offers a thorough manual for reading and evaluating research articles on visualisation. It provides helpful advice and tips to assist readers understand the material and interpret it, advancing the area.

H. Wickham (2010). a visual language with layers. 19(1), 3-28.[8] Journal of Computational and Graphical Statistics. In this important study, a layered language of graphics is introduced, offering a foundation for making and comprehending visualisations. The definition of components, transformations, and mappings makes it possible to generate visuals that are flexible and cohesive.

(2009). Hastie, T., Tibshirani, R., and Friedman. Data Mining, Inference, and Prediction:[9] The Elements of Statistical Learning (2nd Edition). Springer. This well-known book offers a thorough review of statistical learning techniques and how they may be used for prediction and data mining. It is a useful tool for both scholars and industry professionals.

4. PROJECT OVERVIEW AND BUSINESS PROBLEM:

Client's Name: Social Buzz

Client's Industry: Social media & Content creation

4.1 Social Buzz (Client's Background)

Two former engineers from a significant social media corporation—one from London and the other from San Francisco—founded Social Buzz. In 2008 they created Social Buzz because they thought they might be able to create a new platform where content takes centre stage in order to expand on the foundation that their prior company laid. By keeping all users anonymous and solely measuring user reactions to each piece of content, Social Buzz puts an emphasis on the content. Users can respond to material in more than 100 different ways, going beyond the more conventional likes, dislikes, and comments.

This makes sure that popular content—rather than specific users—is at the top of user feeds.

Each month over the past five years, Social Buzz has reached more than 500 million active users.

4.2 Social Buzz Requirements:

They require the assistance of an advising firm to successfully manage their scaling process because they have scaled more quickly than they had anticipated.

The amount of data they produce, gather, and have to analyse is enormous because of their rapid expansion and the digital nature of their primary product. Over 100,000 different pieces of content, including text, photographs, videos, and GIFs, are posted every day. Since most of this data is highly unstructured, managing and maintaining it requires incredibly sophisticated and expensive technology. 200 of the 250 employees at Social Buzz are technical staff members who are responsible for maintaining this extremely complicated system.

They require the assistance of an advising firm to successfully manage their scaling process because they have scaled more quickly than they had anticipated.

The amount of data they produce, gather, and have to analyse is enormous because of their rapid expansion and the digital nature of their primary product. Over 100,000 different pieces of content, including text, photographs, videos, and GIFs, are posted every day. Since most of this data is highly unstructured, managing and maintaining it requires incredibly sophisticated and expensive technology. 200 of the 250 employees at Social Buzz are technical staff members who are responsible for maintaining this extremely complicated system.

They require the assistance of an advising firm to successfully manage their scaling process because they have scaled more quickly than they had anticipated

5. METHODOLOGY

Some common methodologies and approaches used in this training program:

5.1 Blended Learning: A blended learning strategy, including online and offline components, may be used in the training programme. Instructor-led classes, online courses, interactive workshops, and practical training are a few examples of this. Flexible scheduling, individualized instruction, and a combination of academic and practical training are all made possible through blended learning.

5.2 Theoretical Foundations: The training programme typically begins by giving participants a basic understanding of the theories underlying data analytics and visualisation. As part of this, ideas including data collecting, data cleaning, data transformation, statistical analysis, data modelling, and visualisation principles will be introduced.

5.3 Practical Application: To reinforce learning, the training programme places a strong emphasis on application in Real Life. Projects involving the analysis and visualisation of social media data may be industry-specific or based on real-world case studies for participants. Applying the concepts and methods acquired, they analyse and visualise social media data, interpret their findings, and generate practical recommendations.

5.4 Genuine Data Sets: The training programme may include interacting with real social media data sets relevant to the clients' sector or the Social Buzz project to replicate real-world situations. The ability to handle and transform with ease and truth can be learned via practical experience.

5.5 Assessment and Feedback: Feedback systems are crucial for assessing participant development and enhancing the training programme. To determine how well participants comprehend, are able to use, and feel confident using the tactics, a post-training assessment, quiz, or poll may be employed. Feedback is gathered to fill in any gaps or adapt to new needs and to improve the training programme regularly.

To start our engagement with Social Buzz we are launching a three-month pilot project to establish our credibility as the company they should choose. The following is what they anticipate:

- An audit of their big data procedures
- An examination of their content categories that focuses on the top 5 with the highest overall popularity.

5.6 Business challenge: The client has recently grown significantly and lacks the internal resources to handle it.

5.7 My job: I am the data analyst tasked with analysing the customer's sample data sets, producing insights, creating visualisations, and creating a client presentation.

5.8 Tasks (Provided by Accenture)

- Task 1- Project Understanding

As part of this role, I received a briefing on the client's business, their business problem, the project's needs, and the team we would be working with. The information about the client and the team is then tested on, to make sure I comprehend it.

- Task 2- Data Cleaning & Modelling

Gathering the information required for the analysis, cleaning the data, and properly modelling the data are the main objectives of this step. Finding out which content categories have the most overall popularity was what the customer was after. The knowledge of the data is also put to the test. This case study's fifth section outlines the steps for both data preparation and modelling.

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

In this section of the project, my objective is to make the data come to life. I created a presentation to provide my client that included a summary of everything I performed throughout the project. I created charts and visualizations in Tableau that may offer the client information and suggestions for acting.

6. IMPLEMENTATION OF WORK

6.1 Resources

For this project, the following resources were used:

Microsoft SQL Server was utilised to clean the datasets and produce significant insights into the datasets. When I conducted my study on **MS SQL Server**, I used **Microsoft Power BI** to visualise the conclusions and prepared the presentation slides for the client using **Microsoft Power Point**.

6.2 Microsoft SQL Server-related tasks.

- **Data-Importing-** Importing data into SQL Server from Excel spreadsheets, or other databases which were provided by the Social Buzz.
- **Data cleaning and transformation-** Cleaning and transforming the data is necessary to make sure it is accurate, consistent, and meets the criteria set by client for analysis. In doing so, it is necessary to eliminate duplicate records, deal with missing values, standardise formats, make calculations, or aggregate data.
- **Querying and Retrieval-** writing SQL queries to retrieve the necessary data from SQL Server databases. This entails choosing columns, filtering rows according to criteria, merging several tables, and using functions or aggregations to compute summary statistics.
- **Data aggregation and summarization-** combining and analysing data to draw conclusions. SUM, AVG, COUNT, MAX, MIN, and other functions offered by SQL Server were used to create aggregations based on particular standards or groups.
- **Data Analysis Functions**- Utilising the built-in analytical features offered by SQL Server, like RANK, ROW_NUMBER, NTILE, LAG, LEAD, and others. These functions make it possible for SQL queries to perform complex analysis and processing.
- **Reporting and Visualisation-** Using Microsoft Power BI, Tableau, or Excel to integrate SQL Server data with reporting and visualisation software to produce interactive dashboards, charts, and reports for data analysis and presentation.

6.3 Microsoft Power BI-related tasks

- **Data Connection:** Importing data into Power BI by connecting to multiple data sources, like as databases, Excel files, cloud-based services, or online APIs. To establish connections and retrieve data, Power BI supports a wide variety of data connectors.
- **Data modelling-** in this process establishing connections between various tables or data sources in the data model of Power BI is done. To create data linkages and make it possible to analyse data effectively, this entails defining primary and foreign key relationships.
- **Data Transformation**- Power Query Editor, a built-in data preparation tool in Power BI, is used for data transformation, which involves transforming and manipulating data. Data can be cleaned, duplicates can be removed, rows can be filtered, columns can be merged or divided, and transformations like calculations, pivoting, or aggregations can be used.
- **Data Visualization**- Designing interactive and visually appealing visualizations of data using the drag-and-drop interface of Power BI. In order to analyse and communicate

data insights effectively, users may construct a broad variety of visualizations, including charts, graphs, maps, tables, and customized visuals.

- **Dashboard Creation**- Building interactive dashboards that offer a consolidated view of important metrics and insights is known as dashboard creation. Power BI gives users the ability to integrate various visualizations, build interactive filters and slicers, and arrange pieces to make visually appealing dashboards.
- **Sharing and Collaboration**- Datasets, dashboards, and reports can all be shared with team members or stakeholders as part of a collaborative effort. Users can share and collaborate on reports by publishing them to the Power BI service or an on-premises Power BI Report Server using Power BI.

6.4 Microsoft Power Point-related tasks

1. **Making Data Analysis Presentations**- Compiling and organising the data analysis results in a visually appealing and well-organized manner using PowerPoint. Slides can be made to display graphs, tables, charts, and other visual representations of the data as well as to explain or discuss the findings**.**
2. **Visualizing Data-** visualisation is the process in which Microsoft PowerPoint's charting and graphing tools were used to produce visual representations of data. To effectively communicate the findings of a data analysis, you can design bar charts, line charts, pie charts, and other graphic features**.**
3. **Content Organisation and Structure**- Arrange the information on the slides of the presentation in a logical and coherent manner. it is important to be sure that the presentation has a unified and polished appearance for a consistent and professional look, make use of PowerPoint's slide layout settings.
4. **Adding Annotations and Explanatory Text**- Including text boxes, annotations, or callouts on presentations to offer clarifications, insights, or background data regarding the data analysis procedure and findings.
5. **Integrating External Visualisations-** is the process of incorporating graphics or charts made using specialized data analysis or visualization software into PowerPoint presentations. This enables users to utilize the strength of specialized data analysis tools while using PowerPoint for presenting needs.
6. **Presenting Insights-** Giving the intended audience a concise and organised presentation of the results of the data analysis is known as "presenting findings and insights." In doing so, it may be necessary to describe the analysis's methodology, present its important metrics, emphasise its patterns, and talk about its findings or suggestions.

   With the help of this experience, I have learned and used my data analytics abilities to:
   - Clean up the client's data, then use Microsoft SQL Server to perform data modelling.
   - Highlight the top 5 content categories for the customer with the highest overall popularity using data analysis.
   - Utilise Microsoft Power BI to create data visualisations and convey stories with the data.
   - Create a presentation for the customer.

7. ANALYSIS AND INSIGHTS

- **Fig.1** states that the top 5 content categories, out of a total of 16, are food, science, healthy eating, animals, and technology. Animal-related content topped the charts, scoring a total of around 75,000 points.

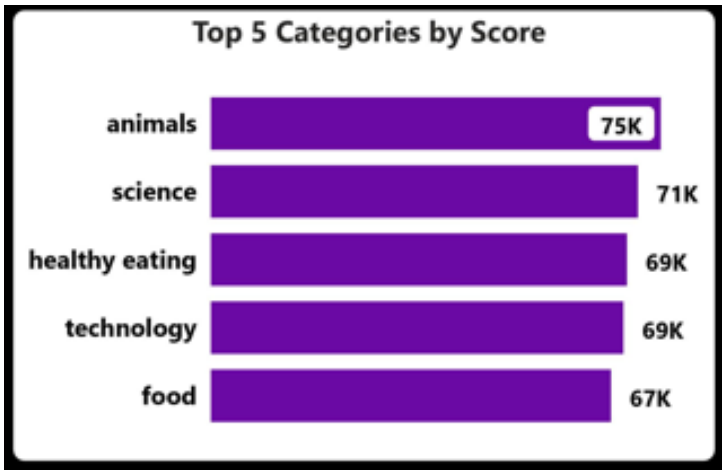


Fig. 1

- **Fig. 2** states that Animals has a popularity percentage share of 21.36%, science has a popularity percentage share of 20.28%, healthy eating is in third place and food is in fifth place, with popularity percentage shares of 19.76% and 19%, respectively, among the top 5 content categories.

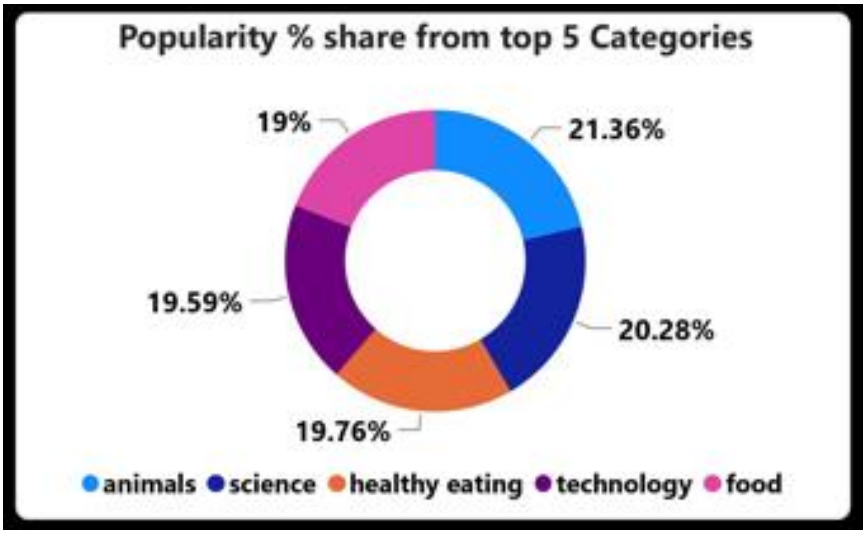


Fig. 2

- **Fig. 3** states the top five content categories are ordered sequentially in relation to the preceding two visuals when we look at the total number of posts by category. Science is second on the list, with 1,864 posts overall, while animals are first with 1,967 posts overall. Public speaking, which received a total of 1,266 posts, came in last.

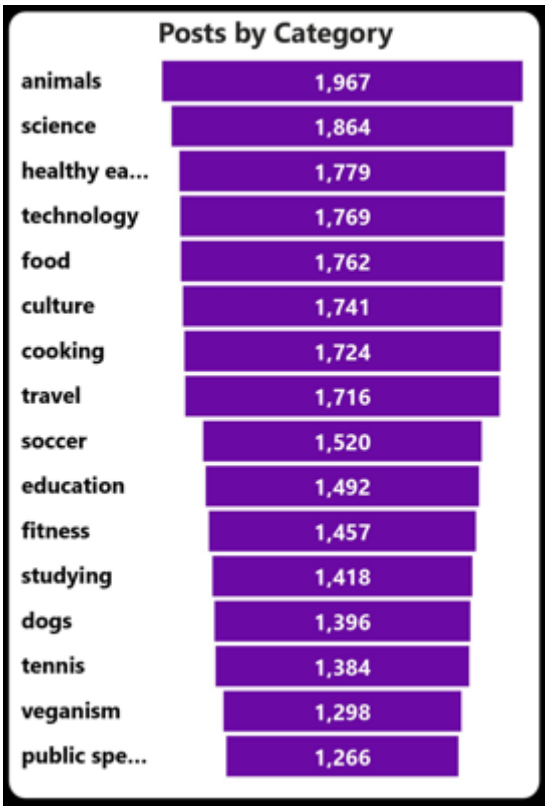


Fig. 3

- **Fig. 4** states that if we break down the total number of posts by content type, images have the most posts with a total of 6,847, followed by videos with a total of 6,499, GIFs in third with a total of 6,313 posts, and audio in fourth with a total of 5,894 posts.

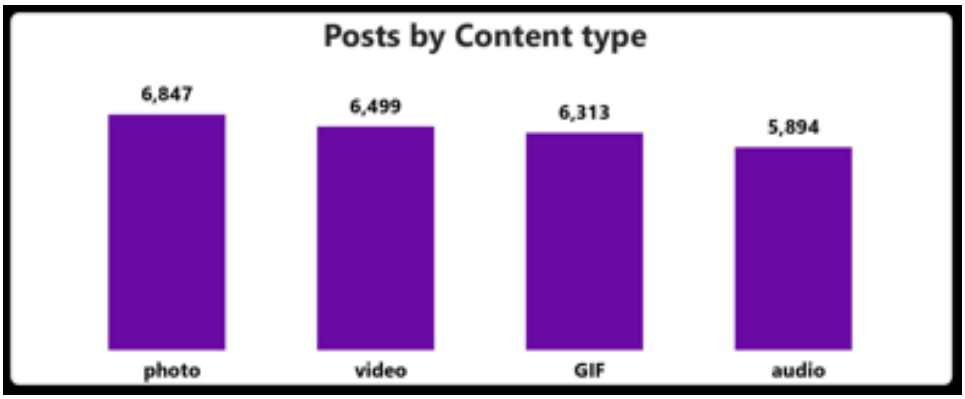


Fig. 4

- Fig. 5 states that with a total of 2,218 posts, January leads the pack in terms of post volume, indicating that this time of year is a busy one for posting. Another high. However, with only 1,980 posts in February, there has been a significant decrease in overall posting. This graph demonstrates that there are monthly undulations (rises and dips) in the number of posts.

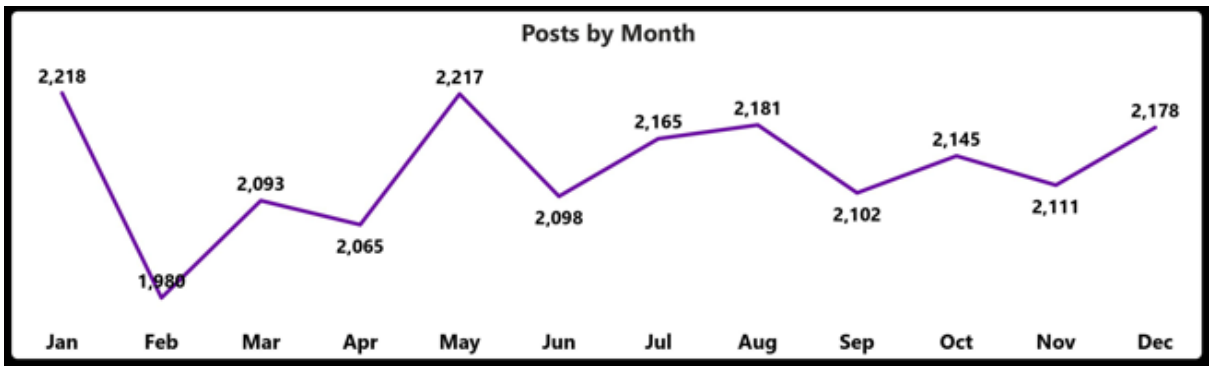
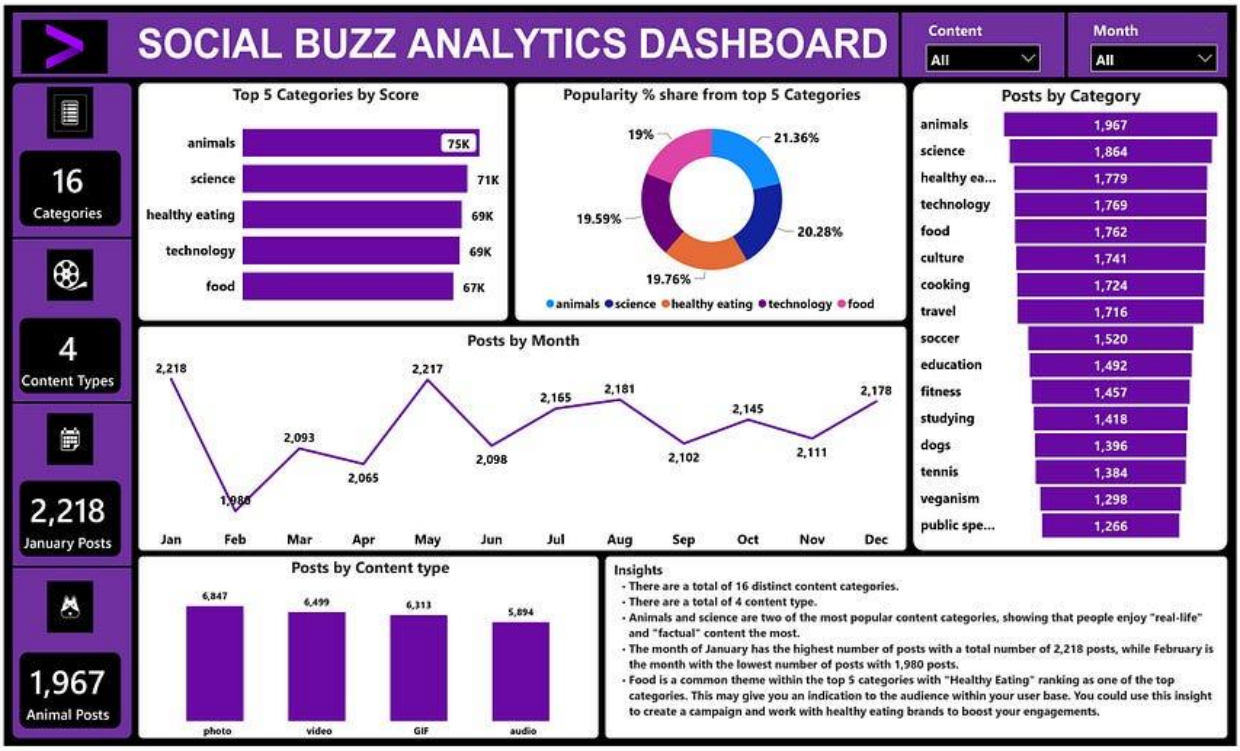


Fig. 5

- **Final Dashboard**



8. RECOMMENDATIONS & FUTURE SCOPE

- In terms of content categories, science and animals are two of the most popular, demonstrating that "real-life" and "factual" information are the most appealing to viewers. Considering this, we recommend you to keep producing content that falls under these two headings.

- The top 5 categories all have a food component, with "Healthy Eating" coming in at number one. The audience within your user base may get a clue from this. This information might be used to develop a campaign and collaborate with companies that sell healthy foods in order to increase user involvement.

- It should come as no surprise that technology-related material is among the top categories given the advancement of technology. It demonstrates that users appreciate your technological material. Working with some of the biggest IT companies would be a good idea, since user engagement would undoubtedly increase.

The platform can then make use of this data to reward writers of content in related categories in order to increase audience engagement. Additionally, make sure that the content being produced is balanced across the board to prevent instances where one content category appears to be receiving the majority of the content posted. To stop that, attention should also be paid to the bottom categories.

9. CONCLUSION

As a result, the customer has benefited greatly from the analytical and data visualisation work done for the Social Buzz project. I was able to examine the client's data and determine the top 5 content categories with the highest aggregate popularity by utilising Microsoft SQL Server for data cleansing and modelling and Microsoft Power BI for data visualisation.

These insights can assist the customer in determining which content categories resonate with their audience the most and help direct their content generation and marketing initiatives. They may increase their reach, engagement, and ultimately boost the success of their social media organisation by concentrating on these well-liked categories.

The client's understanding of the insights was further enhanced by the storytelling with data method and Microsoft PowerPoint presentation. The client may readily understand the consequences of the data analysis and make decisions based on the insights thanks to the clear and succinct presentation of the findings.

By completing this assignment successfully, I have shown that I am proficient in data analytics and that I have strong visually appealing abilities. This work not only addressed the client's operational issue with managing their enormous volume of data, but also gave them access to huge organization best practises and insights, assisting them in successfully managing the difficulties of big data.

REFERENCES

[1]. Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley.

[2]. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.

[3]. Wilke, C. O. (2019). Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. O'Reilly Media.

[4]. Tufte, E. R. (2001). The Visual Display of Quantitative Information. Graphics Press.

[5]. Cleveland, W. S. (1993). Visualizing Data. Hobart Press.

[6]. Inselberg, A. (2009). Parallel Coordinates: Visual Multidimensional Geometry and Its Applications. Springer Science & Business Media.

[7]. Caban, J. J., Gotz, D., & Munzner, T. (2017). How to read a visualization research paper. IEEE Transactions on Visualization and Computer Graphics, 23(1), 391-400.

[8]. Wickham, H. (2010). A layered grammar of graphics. Journal of Computational and Graphical Statistics, 19(1), 3-28.

[9]. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition). Springer.