## Report on Social Media Data Analysis

**Overview:**

This report presents an in-depth analysis of social media data collected from multiple platforms, including Facebook, Instagram, LinkedIn and Twitter. The objective of the analysis is to extract valuable and meaningful insights from the data to inform strategic decision-making, marketing strategies, improve engagement and enhance brand presence. The report outlines the analysis **objectives, methodology, key findings and actionable recommendations** that the client can leverage to improve the results and achieve their goal faster.

**Objectives:**

The analysis aimed to achieve the following key objectives:

1. Audience Understanding: The primary goal was to gain a profound understanding of the audience's engagements and behaviors across various social media platforms.  This included deciphering the content that resonated most with them.

2. Content Optimization: Another critical objective was to identify the most effective content types, ideal posting times and engagement strategies. By analyzing the data,  contents that generated the highest levels of interaction and discern the optimal moments are checked. This optimization process are intended to bolster the brand's online presence.

**Methodology:**

The analysis was conducted using a combination of data preprocessing, exploratory data analysis (EDA), and data modeling techniques. The following steps were undertaken in the analysis:

**Data Collection:** The social media data was downloaded from the hackathon website(diceytech). The csv files of the twitter, facebook, instagram and linkedin were each downloaded.

**Data Cleaning:** The collected data underwent extensive cleaning and preprocessing to handle missing values, remove duplicates and ensure data consistency.

**Exploratory Data Analysis (EDA):** EDA techniques were applied to gain a deeper understanding of the data. This included summary statistics, data visualization, and identifying patterns and trends over time.

**Feature Engineering:** New features were created, including sentiment scores, engagement metrics, and user segmentation, to enhance the analysis.

**Machine Learning**: Machine learning models such as classification and regression were utilized to predict outcomes and uncover correlations within the data. This step is basic and could be advanced based on more information from the organizers of the competition. Their KPIs or targets need to be determined in order to create a good prediction model. An interview with the organizers can help to get more information.

**Key Findings:**

This sections shows the result of the steps in the methodology section above.

**Data Information :**

Table 1 : Table showing the information about the download data

| Social Media Platform | Rows | Columns | Content Types |
|---|---|---|---|
| Facebook | 9803 | 147 | Video, Photo, Text and Link |
| Twitter | 8529 | 147 | Video, Photo, Text and Link |
| Instagram | 10000 | 147 | Carousel, Photo, Video |
| Linkedin | 7760 | 147 | Photo, Text, Video, Poll, Document, Link |

From the Table 1 above, it can be inferred that all datasets collected for all platforms have 147 unique data or columns. All the platforms have sufficient data of more than 7500 entries.

The data for the platforms were individually analysed. I will share the insights for each platform.

**Facebook Data Analysis**
- The Facebook data was cleaned and explored. The data had 133 numeric columns and 14 non-numeric columns. About 46 of those columns were completely empty and they were removed so we can have a more valuable analysis. Also, some columns in the dataset had more than 50% missing values and as a general rule of thumb in data science, it is essential to drop such columns because it could misrepresent the data recorded. These columns were dropped. Some more columns had to be cleaned too and the data type were corrected.
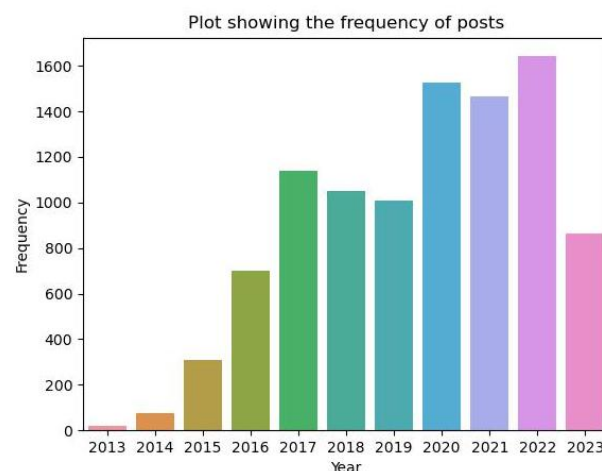


Figure 1 : Plot showing the frequency of Facebook posts per year

The Figure 1 above shows that more data were collected since 2013, the company has been posting contents since that moment. We have a low in 2023 because this hackathon was conducted in 2023 and the whole data for 2023 is yet to be collected.
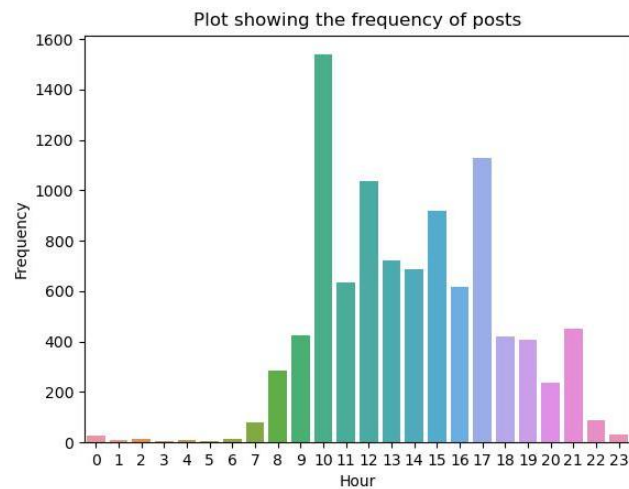
Figure 2 : Plot showing the frequency of Facebook posts per hour

Figure 2 shows that most of the posts on the Facebook platform were between 7 am to 21pm(9pm). This is expected because at past 7, people are getting up and getting engaged. Also there is a spike around 10am, this is probably a very good time to post contents. From 5pm, the number of posts starts decreasing and stays quiet all night
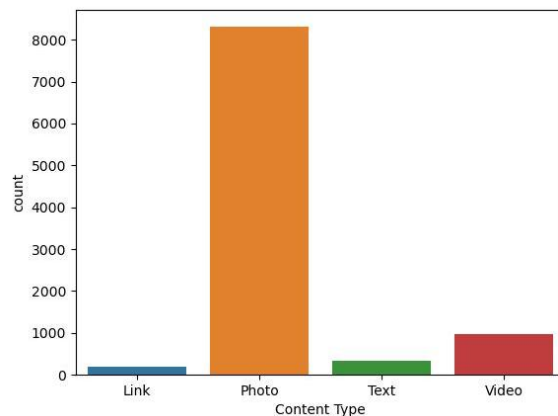and midnight till dawn.


Figure 3 : Plot showing the number of posts for each content type

Figure 3 is showing that most facebook posts are of the photo type, pictures say a thousand words, hence a big reason why it is mostly used. Next is video, text and link.
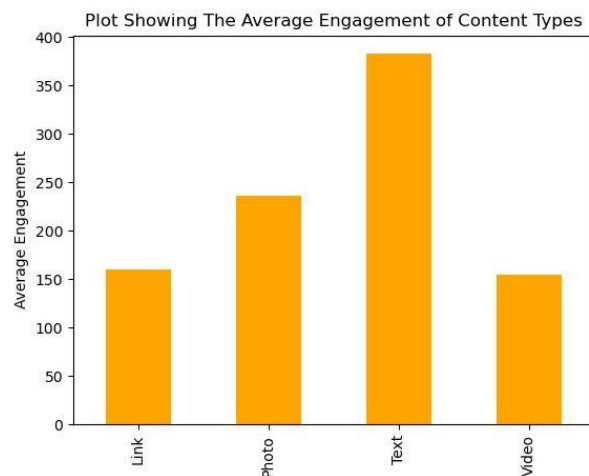

Figure 4 : Plot  showing the average engagement of content types

The figure 4 above shows that more people are likely to engage with the text content type than the other content types.
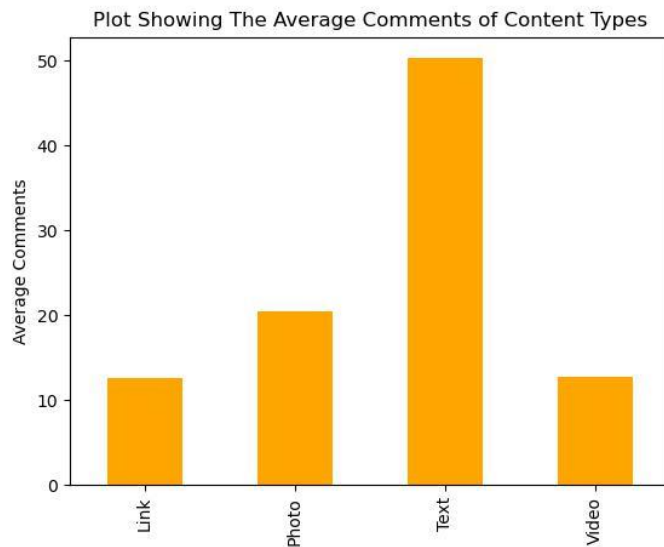


Figure 5 : Plot showing the average comments of content types

Looking at the figure 5 above, text content type has the highest number of average comments just as it was in figure 4 too. It means that for every post shared, the text is likely to get feedbacks. Photos are likey to get more CTR and shares. We will explore that too.
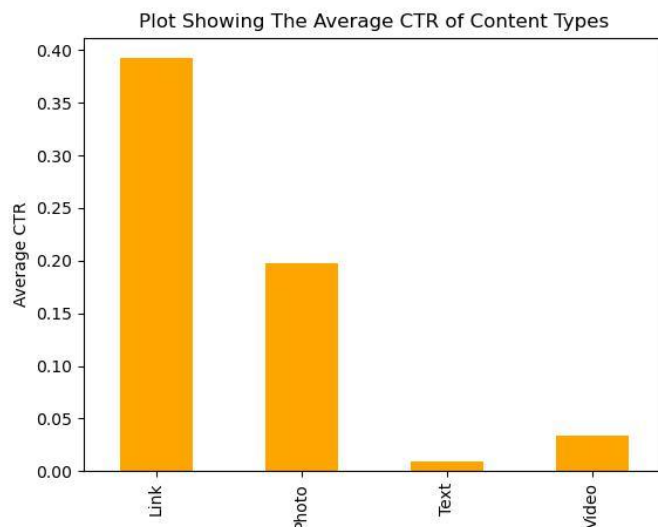


Figure 6 : Plot showing the average CTR for content types

Click-Through-Rate(CTR) is a very good metric for measuring user or customer response or activity. The figure 6 above shows that links always get a high CTR and next is photo and a few CTR for video and text. If the Play House Communications team need to use CTR as a metric, then they should use links in posts more often so that users would click and be engaged.
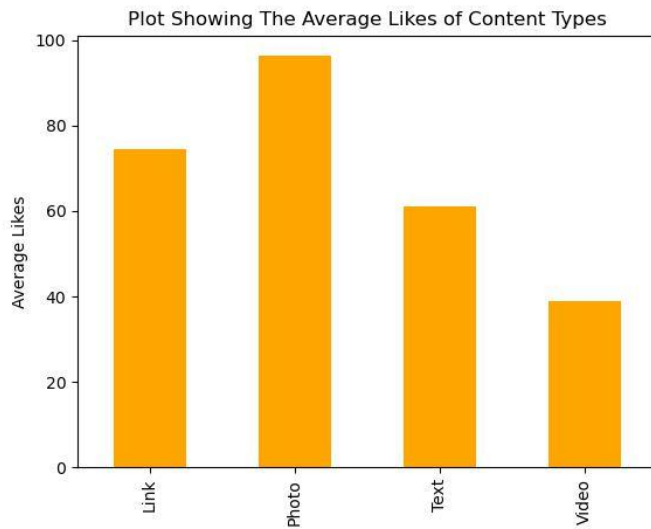
Figure 7 : Plot Showing the average likes of content types

In figure 7, we can see that on average for every photo post, they might be an average of 100 likes and for links  there might be about 75 likes. It simply means that links and photos together is likely to keep users engaged and call for their likes.
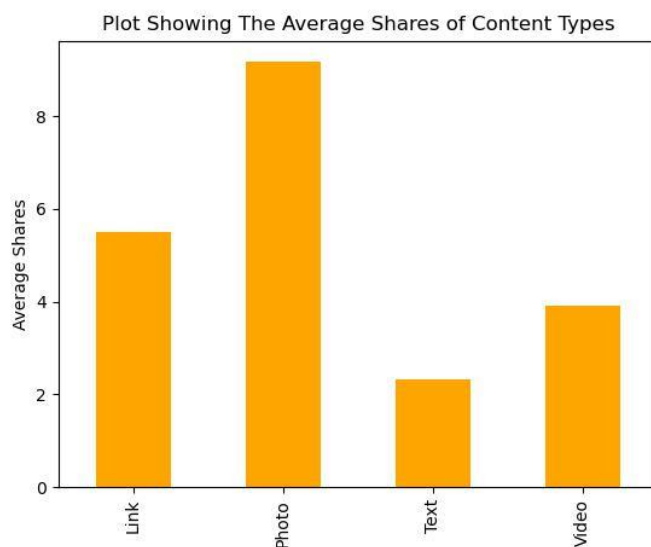


Figure 8 : Plot showing the average number of shares for posts

Figure 8 also supports the claim that photos are likely to have more shares, likes, comments etc. So the company might want to invest more in hiring a very competent graphics designer who would help capture the audience with appealing designs.

Figure 9 : Plot showing the post length of facebook posts

The above plot shows that most of the facebook plots are usually around 1 to 400 words per post. It might be interesting to see the multivariate correlation between post length and engagement. I suggest that the company should use smaller post length if it could convey all the meaning.

Now lets take a look at the word cloud which shows all the words used in the posts. It is important to do this so you can see which word is used more. The frequently used words are bigger than the others. It is shown in the figure below.



Figure 10 : Word Cloud Showing the Most Frequent Words in Facebook Posts

It appears that the words stanbic ibtc, business, wealth, time, send, today, mtual fund,  nigeria, started are all keywords in many of the facebook posts. I realized that we have 'it can be' also in

the plot above, it is not irregular, it is like a motto or hashtag that stanbic ibtc uses in her facebook posts.

**Twitter Data Analysis**

- The Twitter data was cleaned and explored. The data had 135 numeric columns and 12 non-numeric columns. About 97 of those columns were completely empty and they were removed so we can have a more valuable analysis. Also, some columns in the dataset had more than 50% missing values and as a general rule of thumb in data science, it is essential to drop such columns because it could misrepresent the data recorded. These columns were dropped. Some more columns had to be cleaned too and the data type were corrected.
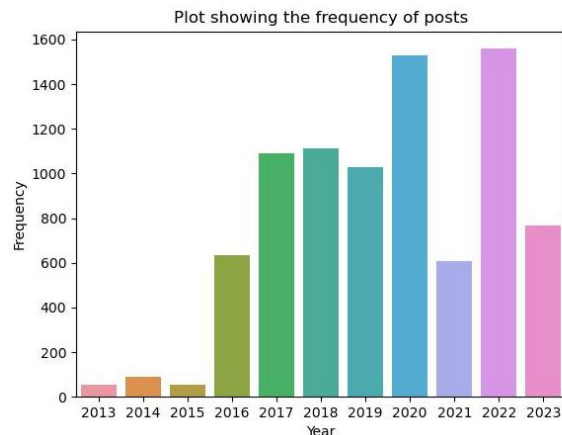


Figure 11 : Plot showing the frequency of Twitter posts per year

The Figure 1 above shows that more data were collected since 2013. There is a spike in 2020 which must have been as a result of the COVID-19 virus stay-at-home policy. More posts were pushed out at that time. Also in 2021, the twitter saw a low number of tweets, this must have been because of the ban of Twitter in Nigeria in that year. In 2022, the ban was lifted and twitter posts increased.
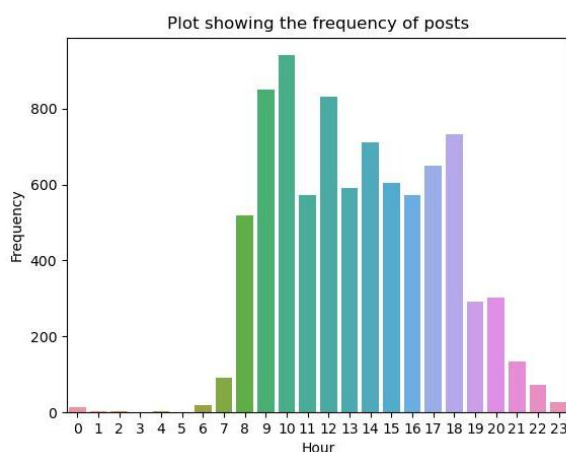


Figure 12 : Plot showing the frequency of twitter posts per hour

Figure 2 shows that most of the posts on the Twitter platform were between 7 am to 21pm(9pm). This is expected because at past 7, people are getting up and getting engaged. Also there is a spike around 10am, this is probably a very good time to post contents. From 5pm, the number of posts starts decreasing and stays quiet all night
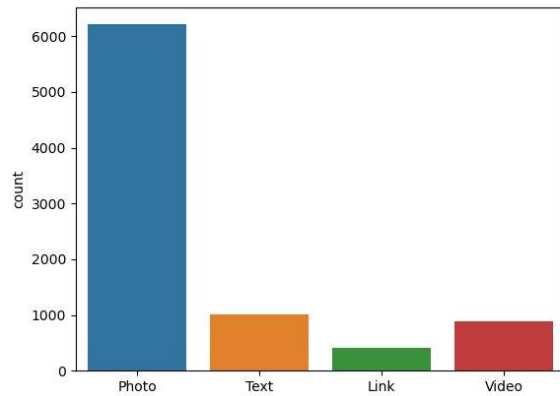and midnight till dawn.

Figure 13 : Plot showing the number of posts for each content type

Figure 3 is showing that most twitter posts are of the photo type, users of social media platforms like visually engaging contents hence a big reason why photos is mostly used. Next is text...
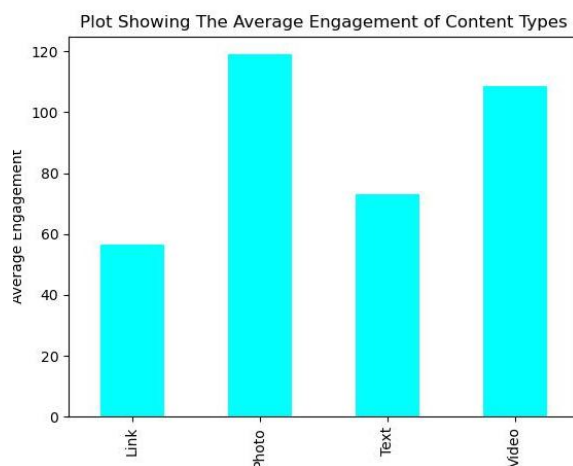


Figure 4 : Plot  showing the average engagement of content types

The figure 14 above shows that more people are likely to engage with the photos and video content type than the other content types. Text and Link also are good options but the could be regarded as second options.
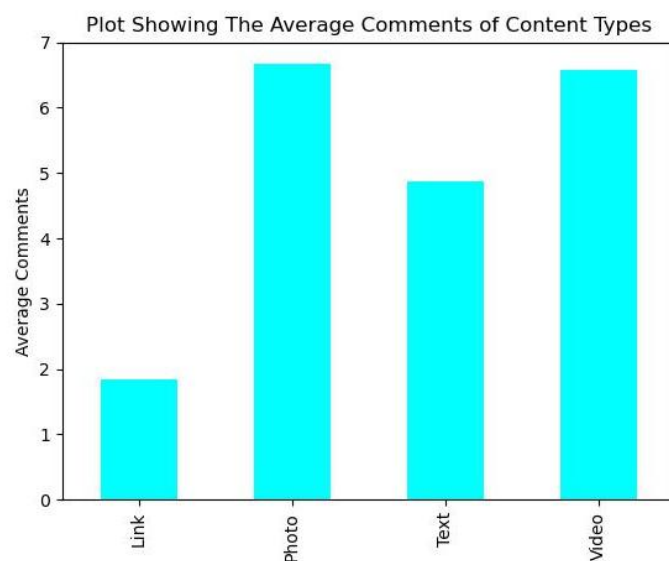
Figure 15 : Plot showing the average comments of content types

Looking at the figure 5 above, photo and videos content type has the highest number of average comments. It just implies that users are sensitive to photos and videos on this platform. On the twitter/x platform, they are quick to comment on the post.
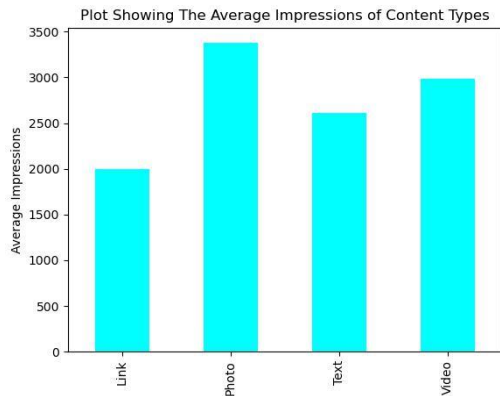


Figure 16 : Plot showing the average CTR for content types

Impressions is also a very good metric for measuring user or customer response or activity. Figure 6 is showing that photos get a lot of impressions and videos, its hard to tell the exact diference but photos are always going to have higher impressions.
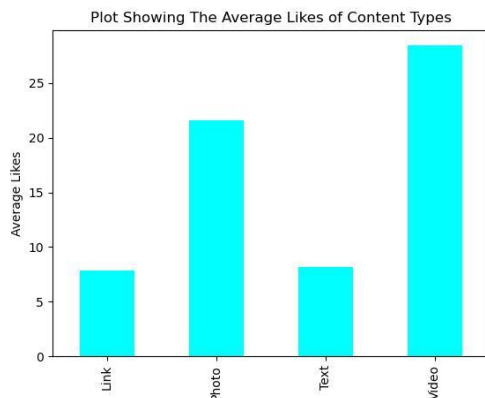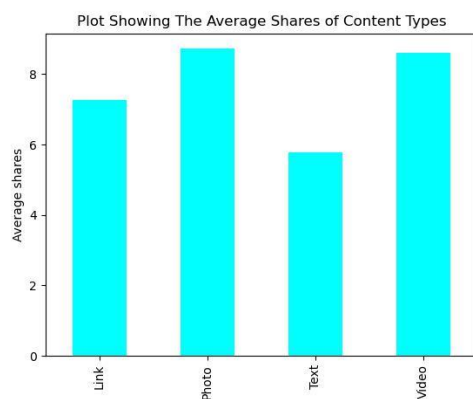


Figure 17 : Plot Showing the average likes of content types

In figure 7, videos and photo content types are likely to get an average of 25 likes. This is quite low and it depends on the quality of the post.



a

Figure 18 : Plot showing the average number of shares for posts

Figure 8 also supports the claim that photos and videos are likely to have more shares. So it is very important to create compelling and captivating posts to get more shares and marketing turnin
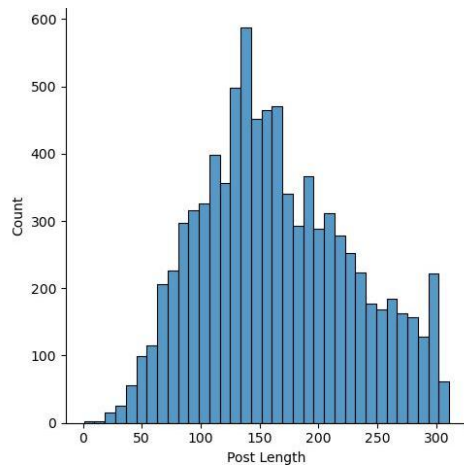


Figure 19 : Plot showing the post length of twitter posts

The above plot shows that most of the twitter post length are usually around 1 to 300 words per post. This is probably because twitter only allows a maximum of 300 characters.

Now lets take a look at the word cloud which shows all the words used in the posts. It is important to do this so you can see which word is used more. The frequently used words are bigger than the others. It is shown in the figure below.
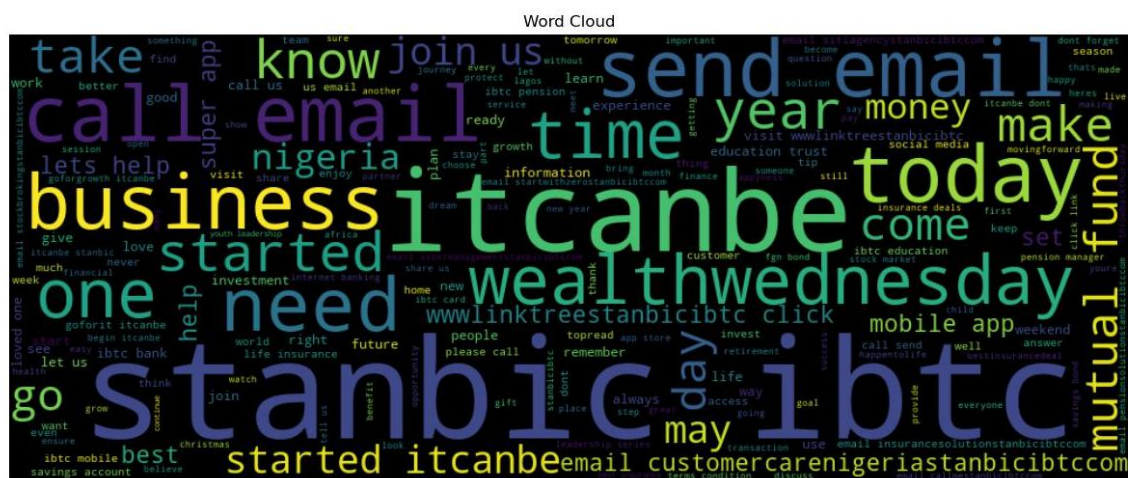


Figure 20 : Word Cloud Showing the Most Frequent Words in Twitter Posts

It appears that the words stanbic ibtc, business, wealth, time, send, today, mtual fund, nigeria, started are all keywords in many of the twitter data. I realized that we have 'it can be' also in the plot above, it is not irregular, it is like a motto or hashtag that stanbic ibtc uses while posting..

**LinkedIn Data Analysis**
- The Linkedin data was cleaned and explored. The data had 135 numeric columns and 12 non-numeric columns. About 120 of those columns were completely empty and they were removed so we can have a more valuable analysis. Also, some columns in the dataset had more than 50%

missing values and as a general rule of thumb in data science, it is essential to drop such columns because it could misrepresent the data recorded. These columns were dropped. Some more columns had to be cleaned too and the data type were corrected.
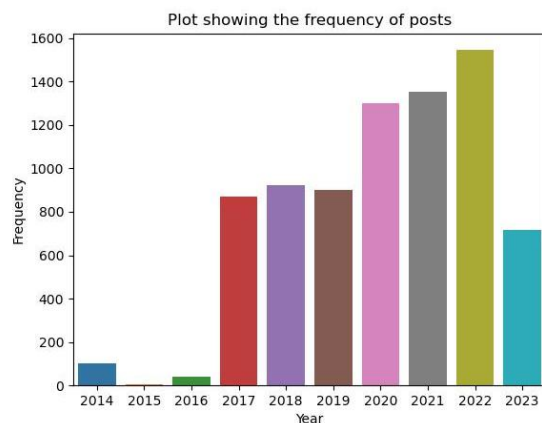


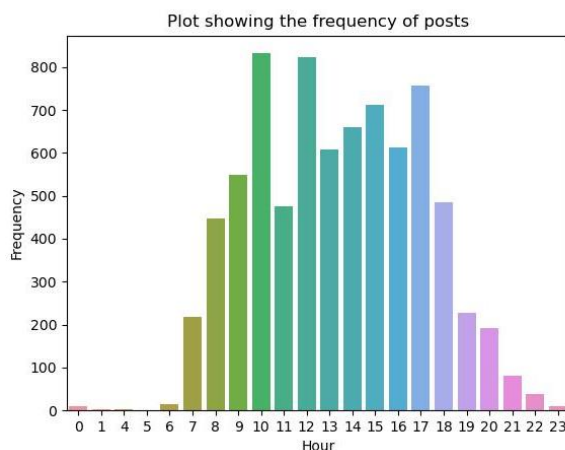Figure 21 : Plot showing the frequency of Linkedin posts per year



Figure 22 : Plot showing the frequency of twitter posts per hour

Figure 2 and 1 shows that most of the posts on the Twitter platform were between 7 am to 21pm(9pm) and was in 2022. This is expected because at past 7, people are getting up and getting engaged. Also there is a spike around 10am, this is probably a very good time to post contents. From 5pm, the number of posts starts decreasing and stays quiet all night and midnight till dawn.
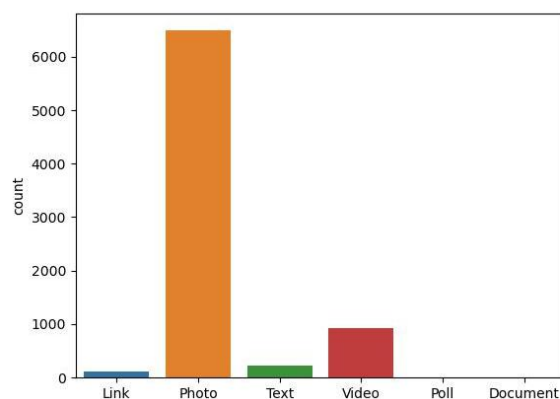


Figure 23 : Plot showing the number of posts for each content type

Figure 3 is showing that most linkedin posts are of the photo type. Photos are more engaging on Linkedin, when people see the images, they are quick to clicking on them. Videos too are engaging in this case.
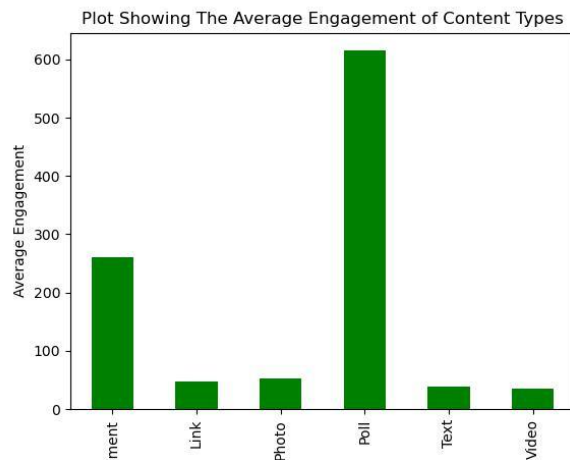


Figure 24 : Plot  showing the average engagement of content types

The figure 4 above shows that more people are likely to engage with the Poll and Document content type than the other content types. Others also are good options but the linkedin platform puts Polls first, it means that for the company to have a very significant engagement result, they should send the post/contents as a Poll or Photo and attach a document. Users on Linkedin are more official people and Documents can be a good way to communicate
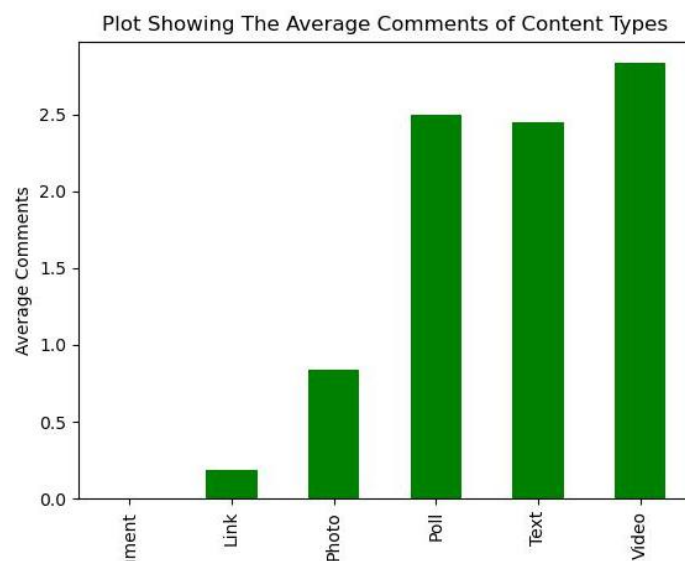


Figure 25 : Plot showing the average comments of content types

In figure 5 above, video, tax and poll content type has the highest number of average comments. So, the client/company should expect that when they post a text, video or poll, people are going to drop comments and at an average of 2 comments.
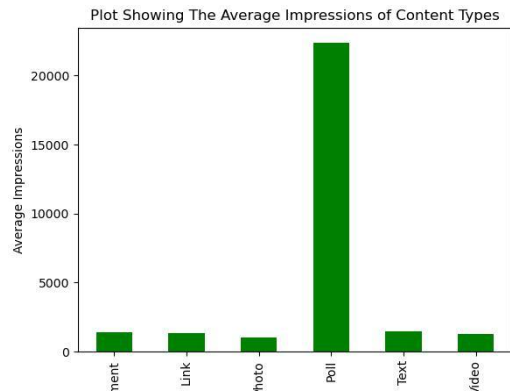
Figure 26 : Plot showing the average Impressions for content types

Impressions is also a very good metric for measuring user or customer response or activity. Figure 6 is showing that Poll get a lot of impressions. When users see Polls, they get moved to participate. Its just like a game but an informative game.
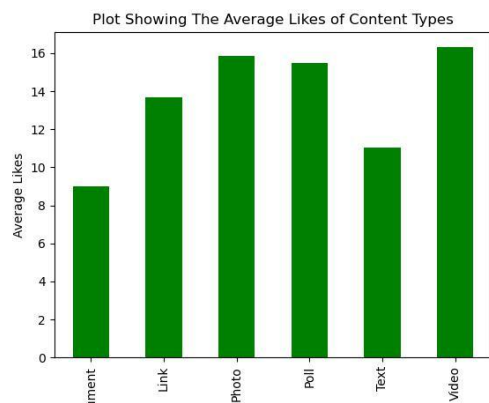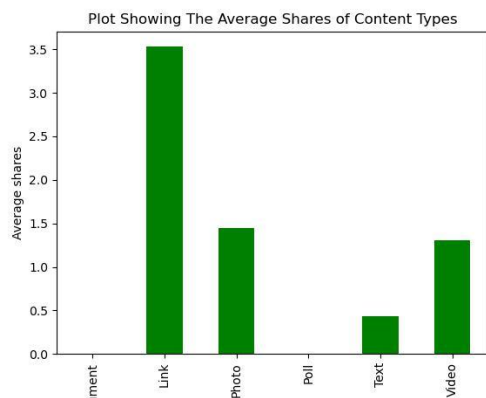


Figure 27 : Plot Showing the average likes of content types

In figure 7 above, photos,polls and videos are expected to have about an average of 15 likes when the content is shared on the Linkedin Platform. The company might want to investigate further and do some A/B testing in order to extract more juicy information.



a

Figure 28 : Plot showing the average number of shares for posts

Figure 8 says that once there is a link in a post, it is likely to be shared. The suggestion would be that when the posts are sent out, they should be accompanied with links.

Figure 29 : Plot showing the post length of Linkedin posts
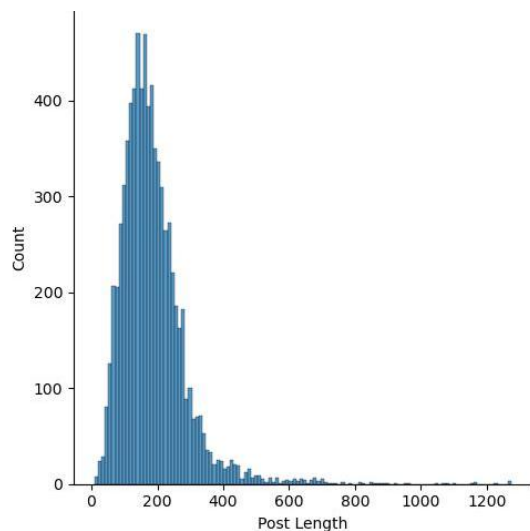
The above plot shows that most of the Linkedin post length are usually around 1 to 600 words per post. Linkedin doesn't exactly have a maximum number of characters for a post but it is better to keep the number of words low. In further analysis, I would love the company to investigate how the length of a post can affect users engagement.

Now lets take a look at the word cloud which shows all the words used in the posts. It is important to do this so you can see which word is used more. The frequently used words are bigger than the others. It is shown in the figure below.
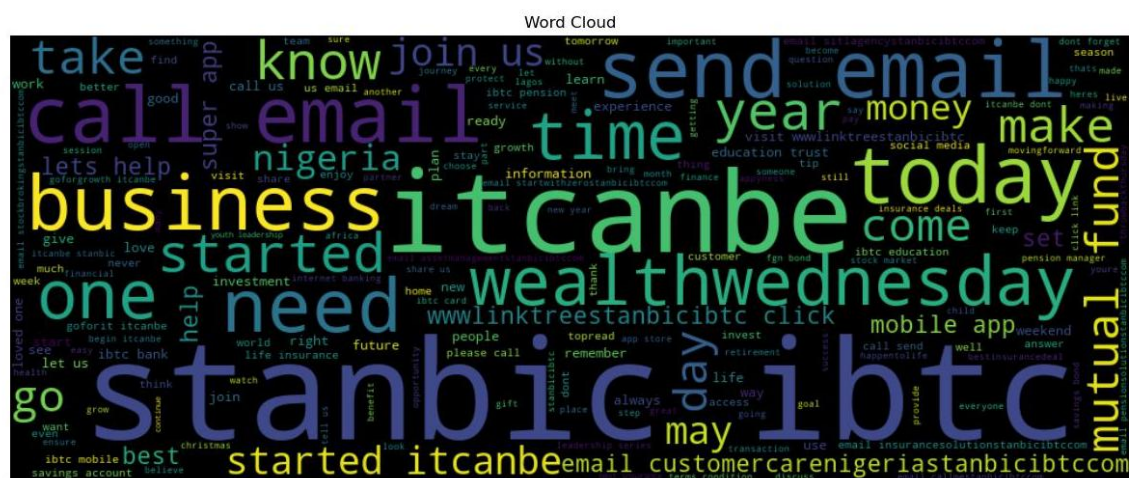


Figure 30 : Word Cloud Showing the Most Frequent Words in Linkedin Posts

Again, It appears that the words stanbic ibtc, business, wealth, time, send, today, mtual fund, nigeria, started are all keywords in many of the twitter data. I realized that we have 'it can be' also in the plot above, it is not irregular, it is like a motto or hashtag that stanbic ibtc uses while posting..

**Instagram Data Analysis**
- The Instagram data was cleaned and explored. The data had 136 numeric columns and 11 non-numeric columns. About 112 of those columns were completely empty and they were removed

so we can have a more valuable analysis. Also, some columns in the dataset had more than 50% missing values and as a general rule of thumb in data science, it is essential to drop such columns because it could misrepresent the data recorded. These columns were dropped. Some more columns had to be cleaned too and the data type were corrected.
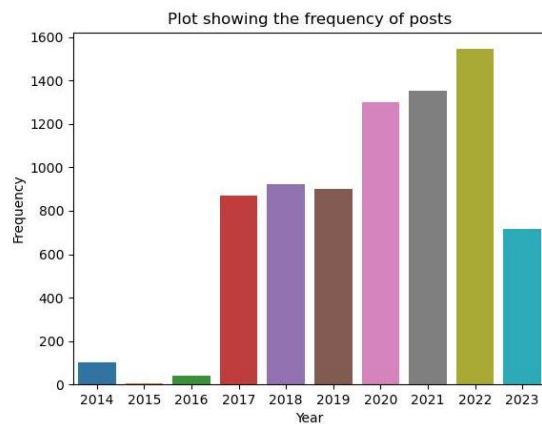


Figure 31 : Plot showing the frequency of Instagram posts per year
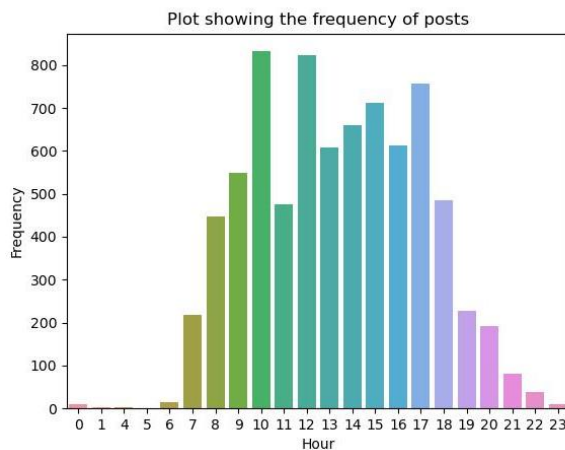


Figure 32 : Plot showing the frequency of Instagram posts per hour

Figure 2 and 1 shows that most of the posts on the Instagram platform were between 7 am to 21pm(9pm) and was in 2022. This is expected because at past 7, people are getting up and getting engaged. Also there is a spike around 10am, this is probably a very good time to post contents. From 5pm, the number of posts starts decreasing and stays quiet all night and midnight till dawn.
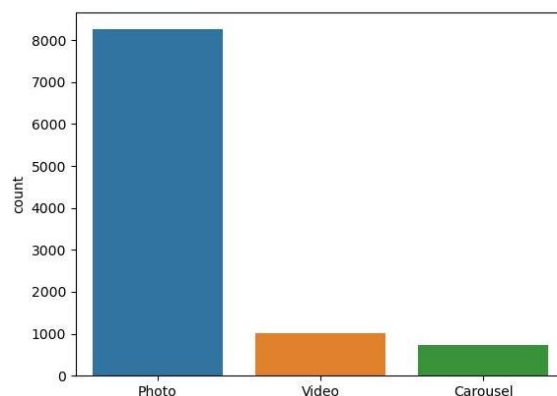


Figure 33 : Plot showing the number of posts for each content type

Figure 3 is showing that most Instagram posts are of the photo type. Photos are more engaging on Instagram. Lots of images are being dumped on the Instagram media space and most contents are photo or video types.
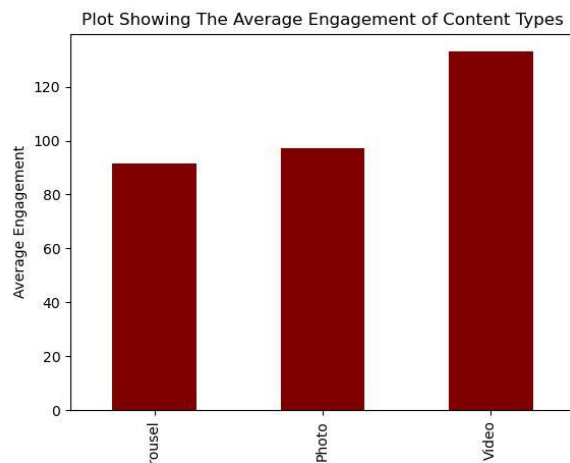


Figure 34 : Plot  showing the average engagement of content types

The figure 4 above shows that more people are likely to engage with the
Videos content type on instagram than the other content types. Though there is a close relationship betiween photos and videos but videos get more engagement from users. Instagram in Nigeria is a place where people go to in order to please their eyes I.e getting visually engaged. Therefore, almost every post is like to keep users engaged. The engagement level is higher than that of other platforms.
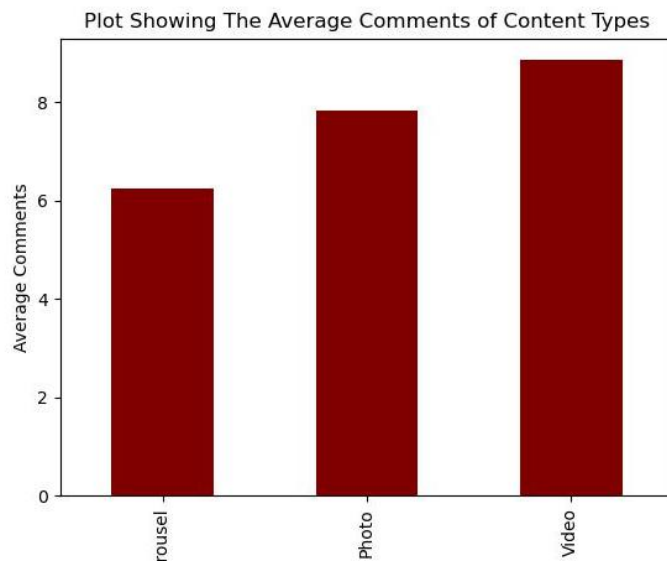


Figure 35 : Plot showing the average comments of content types

In figure 5 above, video content type has the highest number of average comments. So it means for every instagram video post sent, there is usually an average comment of 9.
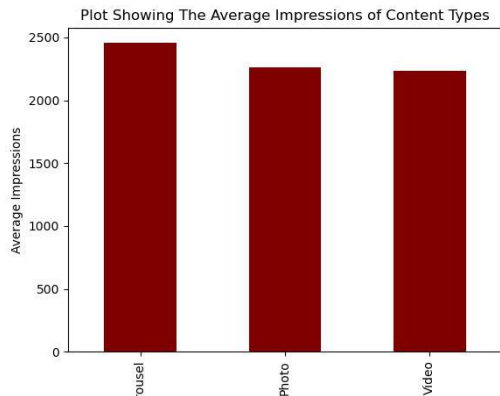
Figure 36 : Plot showing the average Impressions for content types

Impressions is also a very good metric for measuring user or customer response or activity. Figure 6 isimplies that all the content types seem to have just around the same impressions.
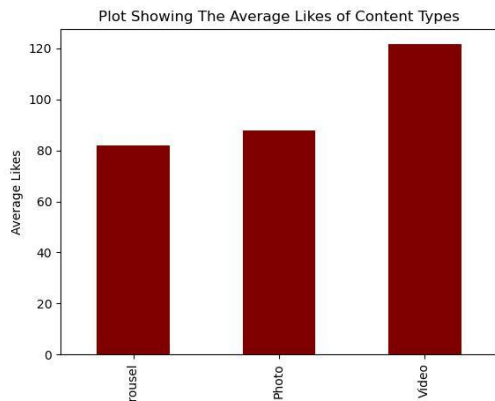


Figure 37 : Plot Showing the average likes of content types

In figure 7 above, videos are expected to have about an average of 120 likes when the content is shared on the Instagram Platform.

## Recommendations:

Based on the analysis, the following actionable recommendations are proposed:

**Content Optimization** : Most of the posts that were engaging are photos and videos. I believe that the company should invest in ensuring that quality photos and videos are put out. They should be engaging contents that can improve the content optimization. Emphasis should be focused on visual content and interactive posts for increased engagement.

**Video Content Optimization**: Invest in video marketing strategies, including live streaming, interactive videos, and short-form content, to capture the attention of the growing video-centric audience.

**Interactive Content Experiences**: Develop interactive content formats such as quizzes, polls, and surveys to boost engagement and collect valuable customer feedback.

**Influencer Marketing:** Collaborating with influencers who align with the brand's values and target audience. Many of these influencers can help promote a wider recognition across the country.

**Machine Learning Models:** Development of predictive models for audience segmentation, engagement prediction and content recommendation. It is important to understand customer feedback..

**A/B Testing:** Design and analyze multiple A/B or A/B/C tests to optimize the the method of selecting the best models.

**Content Recommendation Engine:** Build recommendation systems to suggest relevant content, products, or services to users based on their preferences and behaviors.

**Customer Journey Mapping:** Develop customer journey maps based on data insights to visualize the user experience and identify touchpoints for improvement.

**Localized Marketing Campaigns:** Develop location-specific marketing campaigns tailored to regional preferences and cultural nuances, ensuring relevance to local audiences.

## Conclusion:

In conclusion, the analysis of social media data from Facebook, Instagram, LinkedIn, and Twitter provides valuable insights for optimizing marketing strategies. By understanding the audience, content preferences, and competition, the brand can enhance its online presence and engagement. The recommendations presented in this report serve as a foundation for data-driven decision-making and future marketing initiatives.