# How is your company using text data and much more?

By Ebenezer OBENG-NYARKOH

The emerging world of information technology (IT) is one in which data is critical for driving innovation and development. Leveraging related digital technologies such as analytics and cloud computing as well as machine learning algorithm on artificial intelligence platform (ML/AI) is significantly transforming businesses and improving new and innovative products and services. Data science is in the center of all these changes and has been very effective in offering different real-world solutions including development and deployment of autonomous automotive systems such as Global Navigation Satellite Systems (GNSS), Light Detection and Ranging (LiDAR), Inertia Navigation System (INS), sensors, cameras, radars and much more.

Until a few years ago, just a select few organizations were using data science for data-driven business decision. But now, increasing number of users and consumers cut across many different disciplines including government, public and private sectors. In a 2020 Global DataSphere report, the International Data Corporation (IDC) disclosed that the amount of worldwide data created over the next three years will be more than the data created over the past 30 years and more than 59 zettabytes (ZB) of data will be created, captured, copied and consumed. A related study by International Data Group (IDG) outlined that the technology initiatives that are expected to drive most IT investment in 2021 are data/business analytics and ML/AI.
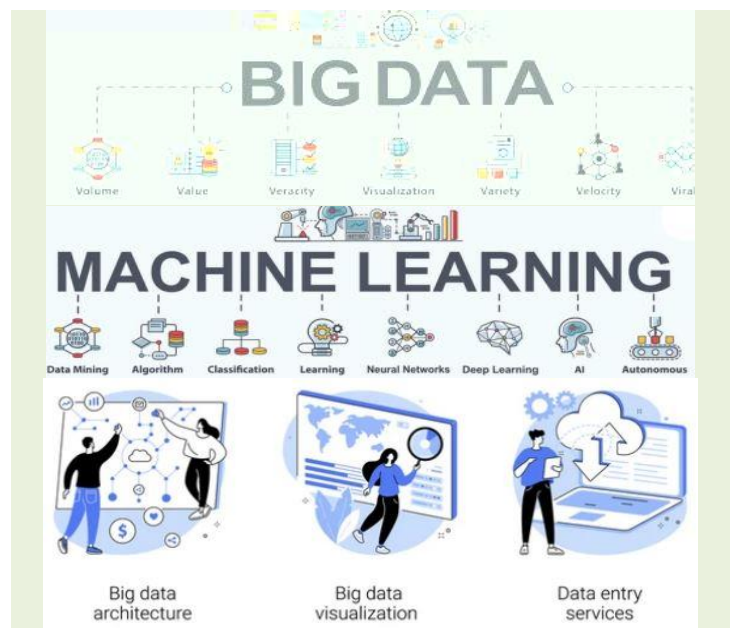

What does this mean for Ghana and Africa in general? Obviously greater investment is critical and the following key areas will successfully facilitate data storage and consumption locally

- Building strong foundation for knowledge and skillset transfer to deal with large volumes of data, work with real-time data, handle data in various ways including cloud computing, as well as heterogeneous data from structured and unstructured sources.
- Committing to long term capacity building and manpower development activities to strengthening and develop the knowledge, skills, understanding, values, attitude, motivation and capabilities necessary for data management and acumen
- Collaboration with local data science companies to locally develop products and services that will provide strategic solutions for local consumption
- Best practices and choice of practical tools to help get firm handle on wide variety of techniques including working with statistical models like regression and ML models and techniques

The knowledge and essential skills needed to keep redefining innovation of product and services requires understanding of complex data and advanced text analytics. Depending on business problem and availability of complex data, a data scientist would be able to create series of knowledge and communicate results. As a case study, the following word cloud with data sourced from various WhatsApp platforms is densely packed with visualizations beyond the commonly used graphs and charts. Evidently, there are more insights and allows consumers to compare and analyze which words are used most frequently. Our team however,

observed a pattern that can be used to make unprecedented personalized offers to customers. As expected, COVID was among the most frequently discussed issue on Mondays in January among all groups
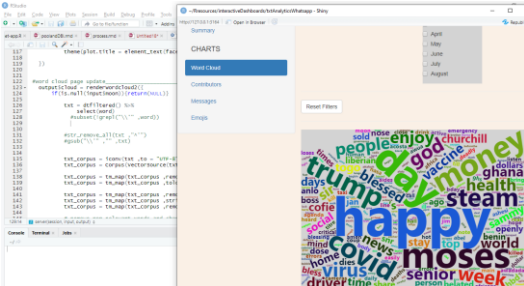
Other frequently used words in Group 1 included people, ghana, passport, country, world, canada, time, pay, and job
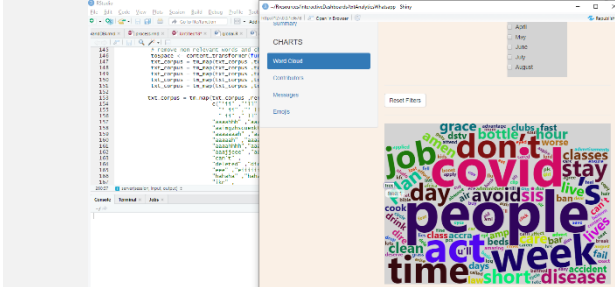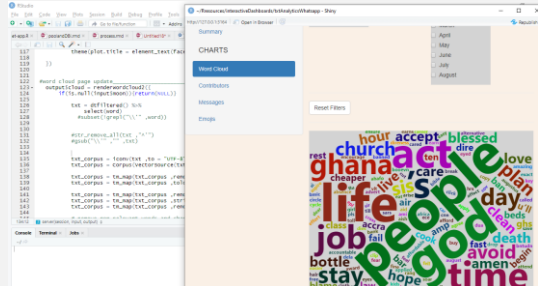

Group 2 on the other hand were interested in relationship, sex, women, money congratulations, germany, day, and god
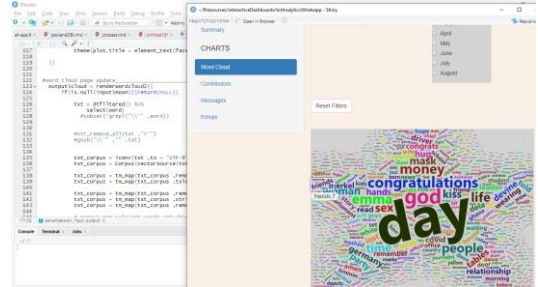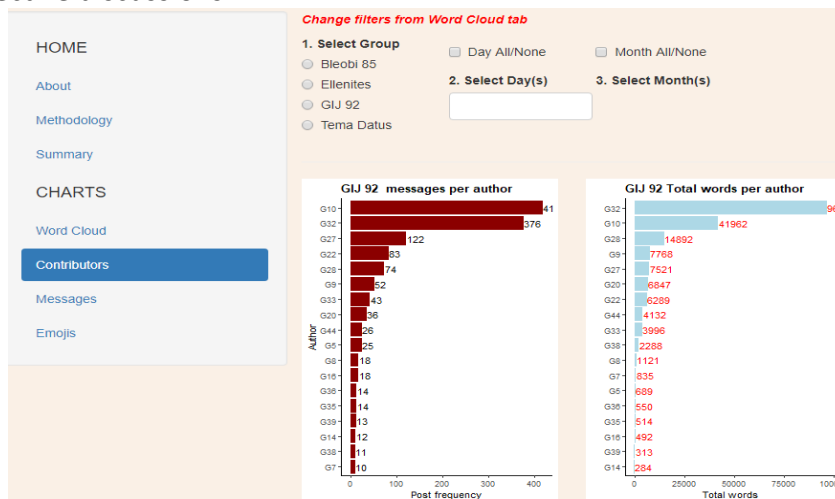

Group 3 was interested in happy, money, ghana, trump, moses, president, health, god, dollars, and day

Group 4 discussed people, god, job, time, life, ghana, bottle, church, death, and love more frequently

Similarly, it is remarkably clear to determine who are unique top contributors and how much they are contributing to respective discussions

A robust text analytics of WhatsApp chats designed by Data Insight Group to help understand and identify opinions from complex data in a text and sentiment analysis

Clearly, in this era of ever increasing volume, velocity, variety, and veracity of data, our team believe this is an exciting time for adoption of Word Cloud to help understand and identify opinions from text data. As an essential qualitative tool, the technique

- Offers powerful visualizations and present text data in a simple and clear format in which the size of the words depends on respective frequencies. The words are visually nice to look as well as easy and quick to understand.
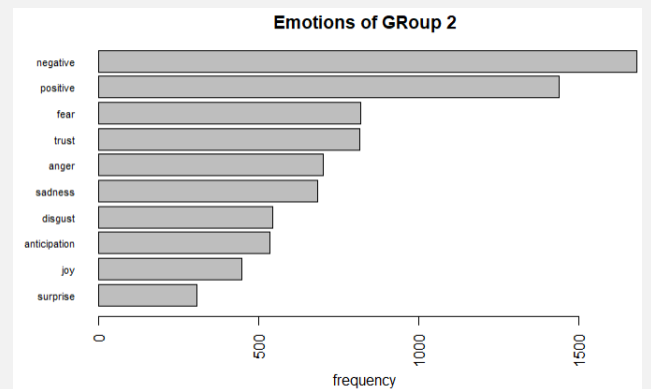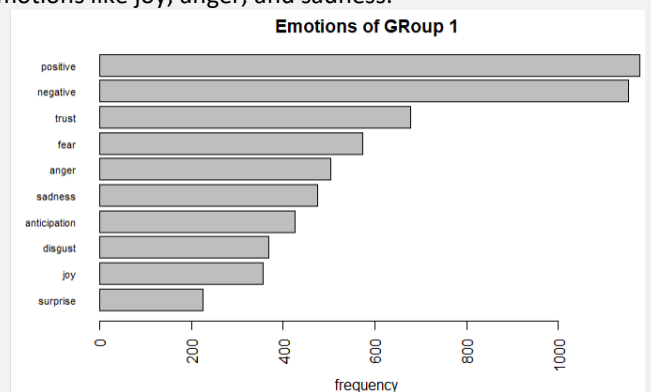
- Is insightful, powerful and great communication tool offering incredibly easy to consume information to users wishing to communicate descriptive insights on text data, voice or video from reports, social media or other customer reviews
- Continuously representing visually engaging words in a cloud that stimulates users to think and draw the best insight as well as allows flexibility in their interpretation to draw several insights quickly depending on business needs

Committing to building intensive data science capacity and manpower development activities offers an exclusive opportunity to develop both individual and institutional skillset. Undoubtedly, data science is ubiquitous and encompasses data analytics, data mining, deep learning, ML/AI and several other related disciplines such as math, statistics, programming, data management, and visualization. Potential training in say a real-time sentiment analysis which uses computational task to analytically determine and identify prevailing emotional opinion and feelings helps understand what a writer is expressing in text

Mixed feelings for positive, negative sentiment, and also emotions like joy, anger, and sadness.

Fundamentally, customer sentiments analysis or opinion mining has become a remarkable marketing strategy needed to keep redefining convenience and customer experience in a digital space. Modelling the technique essentially leverages machine learning algorithms to classify positive, negative, or neutral to uncover and gain actionable insights about how customers feel in many sources of text like surveys, customers or stakeholders' reviews, social media, articles on the web, and even in customer support queries.


In one such integrated training initiative to help transform digital products and services, our client interestingly extracted the following actionable insights per bar chart from many WhatsApp chats.

Incredible number of opportunities continue to exist in this era of big data and requires careful planning and particular skill set including experience, abilities, and knowledge to understand and harness actionable insight. All data are local and thinking critically, managing massive data locally should provide local solution in all related vertical market including food and agriculture, telecommunication, politics and government, NGO, healthcare, education, communication, transport as well as banking, financial services, and insurance (BFSI).

Generational diversity and technology is continuously disrupting both large and small businesses. For all businesses or agencies struggling to grow revenue or get the best return on investment, outsourcing or collaborating with local data science companies is more cost effective. Giving the sensitivity of
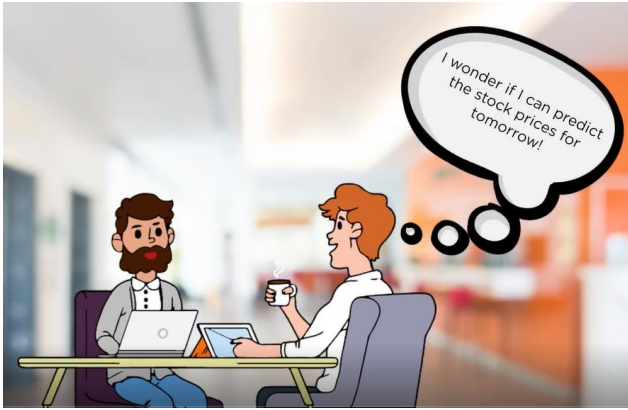
demographic bridge, using external local data science group can help develop deeper understanding of cultural differences, politics, go-to-market strategy and demographics segmentation including age, gender, generational and multi-cultural audience. This is key for preparing customer profiles and optimize campaign performance to target market.

Use of various forms of data and, access to live and historical datasets for advanced analytics in real-time is a critical game changer and key to identify new opportunities, optimize operations, overcome financial challenges, and make data-driven decisions. Several data collection and processing tools exist however, for optimal outcome, we recommend the following data mining and analytic tools basically designed for machine learning objectives and operations to handle large volume of data and human involvement for building models. Primarily, these include (but not limited to)

- Open source and professional software RStudio. An incredible tool for statistical analysis and predictive analytics
- Tableau to accelerate data visualization and analysis of statistical data
- SQL for data connection and, data extraction, transformation and loading (ETL) in a sandbox and staging area

Data analytic tools recommended for professional use
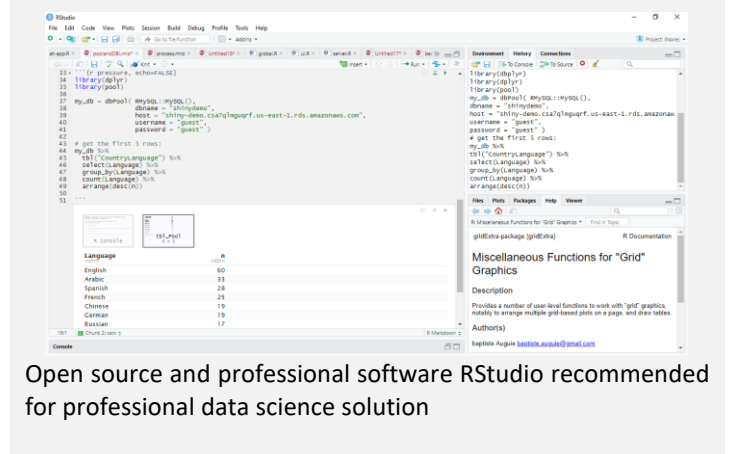
Open source and professional software RStudio recommended for professional data science solution

The impact of the digital economy on industries like health, education, energy, communication, construction, transportation, and natural & renewable resource among others is clearly generating opportunities. And depending on business problem and availability of complex data, projects may vary. However, this is the time for all sectors to harness the disruptive power of data and put data science into practice to drive better business decisions and optimize operational strategy.

Data Insight Group (DIG) is a data and research company helping other businesses to jumpstart, their analytical and data science initiatives to understand and identify actionable insights from complex data. Our data scientist, engineers and researchers have over 20 years' experience. The team is led by Ebenezer Obeng-Nyarkoh, expert in implementation of data science and statistical methodologies. He has unparalleled technology-driven business solutions and extensive experience in leading and leveraging data science solutions and analytics for valuable insights for top tier companies in the USA and Ghana.