BIG DATA ECONOMY:
DRIVING THE ECONOMY THROUGH DATA SCIENCE

Keynote Remarks By

DR. YEMI KALE

Statistician-General of the Federation / Chief Executive Officer

National Bureau of Statistics

Delivered at the

1st National Summit on Big Data Economy /

2nd Data Science Bootcamp

Organised by

Data Science Nigeria

October 12, 2017

Introduction

I am honored to join you here today at the 1st National Summit on Big Data Economy and the 2nd Data Science Bootcamp being organized by Data Science Nigeria. I consider this event, its organization and it's objectives highly commendable, especially as it aims to address a pressing challenge in our society today, which is preparing today's youth for the workforce of tomorrow, in particular, by building up the skills gap among young professionals and students in the emerging field of data science, and specifically in machine learning, programming and data analytics.

Undoubtedly, one of the most engaging and increasingly important areas of discussion since the dawn of the 21st century has been on statistics and getting data right. Today we talk of open data, big data, and the right data. We hear debates about whether African data is poor or whether there is a statistical tragedy or renaissance in Africa. Attention to the quality of data has increased globally. Even in more developed countries like the UK, Canada and the USA, there are questions being raised about the quality of data, errors in data, wrong use of data and polarisation of data.

All around us, we observe the quantum leap in the type, size and scale of data that is being driven by rapid advances in the world of computing. Vast amounts of data are being generated every second of the day, across the world in various forms and across various sectors, what we call **'Big Data'**. (The term simply represents the increasing amount and the varied types of data that is now being collected.)

We often define or describe data as a collection of facts that has been translated into a form that provides information. When we hear data we often think of numbers and figures. However, data can be in the form of numbers, images, words, figures, facts or ideas. It is thought to be the lowest unit of information from which other measurements and analysis can be done. In other words, data (whether ordinary or what we call big data) is simply another word for information. The thing that differentiates Big Data from the "regular data" we were analyzing before is that the tools we use to collect, store and analyze it have had to change to accommodate the increase in size and complexity. With the latest tools on the market, we no longer have to rely on

sampling. Instead, we can process datasets in their entirety and gain a far more complete picture of the world around us.

However, data in itself cannot be understood and to get information from the data one must interpret it into meaningful information. Accordingly, the 'bigger' the data, the more complex and the more difficult it becomes to analyze it to get that much needed information we require to progress. All that data needs, however, is to be processed and interpreted by someone before it can be used for insights. No matter what kind of data you're talking about, that someone is usually a data scientist. Data scientists are arguably now one of the most sought-after positions. A former executive at Google even went so far as to call it the "sexiest job of the 21st century". So, congratulations to all of us in this room. We have arguably the sexiest job of the 21st century

With the vast amount of data being produced daily and its importance in providing information and given the importance of information in taking decisions for progress, whether as a business or as an economy or country, or as individual in our daily lives or as entrepreneurs the question that faces us in Nigeria today, therefore, becomes: ***How can we take such data and convert it into actionable knowledge to move our lives, businesses and economy forward?***

In my remarks today, I examine very briefly, this concept of big data economy and specifically, how the field of data science can contribute towards enhanced and meaningful information for greater business development, structural transformation and economic diversification.

But let me start with some comments on the importance of and growing demand for data.

2. The importance of data

The growing importance of data in the global environment is indisputable. Data is making an impact on every sector across every industry on a global scale. Whichever industry you work in, or whatever your interests, you will almost certainly have come across a story about how "data" is changing the face of our world. It might be helping to cure a disease, boost a company's revenue, make a building more efficient or be responsible for those often irritating but targeted advertisements you keep seeing on your 'smart TVs' or on your phones as text messages or in those free apps we use on a daily basis. It has therefore played and is playing a major role in

shaping almost every aspect of human life, from administration to astronomy, biology to business, housing to health, engineering to environment, commerce to community, marketing to management, infrastructure to industry, and policy to politics. In the olden days, the kings and other rulers used statistics based on population census, primarily to procure food for the people and prepare the army for security. Censuses have been used by governments across the world ever since for various purposes, including planning for socio-economic development: how many hospitals, schools, teachers, new roads etc are needed over the next 10 years? How many policemen or women will be required to safeguard a community? How many graduates and technical persons will join the labour force and how many jobs need to be produced for them in the next 20 years? Arriving at the answers to all of these questions, and many more require that we get our data right.

The value of data is becoming more and more apparent as we continue to move towards an information-driven economy where data has become the new currency. In oil producing countries like Nigeria, some say data is the new oil.

The volume of data and information which businesses now process on a daily basis is on an unfathomable scale. The advent of connected devices, smart infrastructure, complex networks and the constant availability of digital services mean businesses are sitting on a wealth of data and consequently information. Have you wondered how come Google, Facebook, Twitter, Instagram, Whatsapp, etc. offer applications and services free of charge but at the same time these companies are worth billions? Are they being foolish? Are we being smart using such important applications for free? Data is the reason Amazon knows which book you like to read before you do or why Apple"s Itunes can recommend music and videos to you before you ask. All you do is switch your phone on and Itunes, unsolicited, is making recommendations to you what they believe you may like. Those of you that use Netflix will open your account and there is a line showing movies you may be interested in watching based on what you may have watched previously. Those of us that have lived in more advanced countries will recall that based on your purchases in stores, you get loyalty vouchers to purchase items free or at a discount based on your history of purchases.

A rapidly growing universal truth in international business today is that all roads lead to data. In an increasingly complex and connected world, the ability of an organization to collect, manage

and analyze huge amounts of data effectively separates the winners from the runners-up or even losers.

A recent report from Capgemini which surveyed over 1000 businesses revealed that 65 percent of respondents indicated that their organizations are at risk of becoming uncompetitive due to the highly competitive data landscape. 36 percent said that, due to the strategic importance of big data, they have had to circumvent IT teams to carry out the necessary data analytics required to gain business insights. 64 percent indicated that big data is changing traditional business boundaries and enabling new providers to move into their industry. 54 percent reported that their big data investments over the next three years will outstrip past investments. 43 percent have already or are in the process of reorganizing in order to exploit new big data opportunities.

Data (and statistics) are therefore vital as they provide us with clear, objective, numerical evidence on all aspects of our lives and the state of our businesses or country, including the growth and characteristics of our population, economic performance, levels of health and wellbeing and the condition of our surrounding environment.

Data aids the decision-making process by enabling us to establish numerical benchmarks, monitor and evaluate the progress of policies or programmes, ensuring that our policy interventions are well designed, meeting initial aims and identifying any areas which require improvement. Accordingly, the significance of statistical information for making evidenced-based decisions that guide the implementation of new policy, monitoring of existing policy and evaluation of the effectiveness of policy decisions can therefore not be over-emphasized.

Without these, we cannot make well-informed decisions that will catalyse our socio-economic development and secure the lives of future generations. It is when we are able to collate, understand and interpret data correctly, as well as identify key areas in our society or our economy that require change, that the policy prescriptions and direction of our governments and businesses are more likely to respond to the real needs of our communities.

I am of the opinion that if we do the right things at the right time and for the right reasons, Nigeria can achieve its well-recognized potential as a great economy and country. Our inability to do what was right in the past for the right reasons and at the right time is one of the reasons we

have the challenges we have today as a country and getting our data and data analysis right is one of such right things we must do.

3. The growing demand for data

Let us digress slightly and consider for example, the arena of data gathering in Nigeria. The Nigerian Statistical System has a long evolutionary history. Many of us remember the highly unorthodox and disorganized statistical system during the days of the FOS where you will see stacks and stacks of files with papers supposedly containing some form of data in them. Until recently we tolerated and came to accept a statistical system that was less than optimal, weak, uncoordinated and largely ineffectual in meeting the needs of policymakers, business investors and citizens who needed accurate, reliable and timely data on key socio-economic indicators to make informed decisions. At some point, policy makers and other data users ignored official data and made up their own irrespective of the fact that it was largely "guesstimates". This is however gradually changing and in recent times the demand and supply of quality data has increased considerably. Of course, there is still a lot of work to do to develop the statistical system so that it meets that requirement of providing useable and valuable information from complex analysis of vast amount of data. We are even still largely at the level of producing and analyzing what I described as ordinary data although we have a National Strategy for the Development of Statistics which includes handling big data.

Nevertheless, there is undeniably, an increasing recognition of the importance of statistics in Nigeria. This data summit and boot camp for example will probably not have even been considered at all a few years ago. The growing demand for data driven information and decision making has led to an emerging resurgence in the supply of data and statistical information in Nigeria.

There are two factors driving the demand for data in Nigeria: *exogenous* factors and *endogenous* factors.

The *exogenous factors*, which are arguably more dominant, typically involve "external demand" as dictated by conditions occurring outside the country. The drive for data driven decisions in more advanced countries coupled with the increase in international investors seeking investment

havens with attention turning to emerging markets, with Nigeria being one of the preferred destination.

On the domestic front, or *endogenous factors,* data demand is being fuelled by growing insistence on accountability and good governance by citizens, as well as the desire by governments at all levels to demonstrate progress and democratic dividends in various sectors. The current economic challenges facing the country has further amplified the demand for accurate, reliable and timely data on virtually all sectors of the Nigerian economy.

Like never before, we are all living witnesses to the transformational role that increased data availability and quality national statistics have played, and continues to play, in our national life. We see it in how our government has stepped up the use of data in planning, implementing and targeting social programmes even though a lot still needs to be done. We see it in how businesses track and rely on official and nonofficial statistics to make critical investment decisions that ultimately impacts not only government operations, but the welfare of ordinary citizens. We see it in how ordinary Nigerians are able to better understand how they are affected by policy and political decisions, appreciate the state of the economy around them, and their own responsibilities as citizens. We see it in the number of requests by private organizations in Nigeria to collaborate with the NBS to generate data or in their increasing private sector investment data production and in producing more indicators to better inform their clients or for internal decision making.

As custodian of official statistics and coordinator of the National Statistical System, the National Bureau of Statistics, despite continued challenges, has redoubled its efforts towards ensuring high frequency and quality data is available for policymakers, business investors and citizens alike. In this regard we have partnered and continue to partner with various private sector outfits involved in data production and analytics. Today NBS is an integral part of various government committees and decision-making initiatives actively engaging in advocacy efforts to ensure that NBS data products are relevant to informing economic policymaking where and when required. The critical role of using data to inform policymaking bears out clearly in our recent economic experience. Today one of the first places potential investors visit is the NBS to get a sense of what the data is saying.

4. Driving the economy forward: role of data science

For our lives as citizens to improve, for our businesses to progress and for economy to be truly diversified and sustainable in the long run, we have to cultivate stronger technological capabilities and deepen technical innovation. Without a doubt, one of the most significant aspects of technological innovation in the last century has been the advent of computing and one would not be exaggerating in saying that our planet runs on computers these days. Look at this room and try and count the number of things that run on computing. Check your pockets or your hands and see the number of things that depend on computing. When you go home or back to your offices take a moment and admire the way computing has taken over our entire lives. We have almost taken a lot of these things for granted.

The rate of advancement in computing power, from the bits and bytes to big data analytics, particularly in recent decades has made it possible to retrieve data from a variety of sources or formats, process them and present them in ways that are more helpful for decision making. This is where the field of Data Science comes in. **I repeat the sexiest and most sought-after job in the 21st century.**

That being said, as a discipline, there remains controversy among scholars and practitioners over what exactly constitutes "Data Science". Some consider it as ***"an evolutionary step in interdisciplinary fields like business analysis that incorporate computer science, modeling, statistics, analytics, and mathematics"[1].*** Others see it as ***"a multidisciplinary blend of data inference, algorithm development, and technology in order to solve analytically complex problems"[2].*** That is how you know how important and esoteric something is when there is disagreement about what it actually is.

Either way, it is clear that data science brings concepts, tools and methods from various fields, particularly computer sciences and statistics, to process vast amounts of data into useful knowledge and information for decision-making.

Data, in and of itself, is useless if it only exists in a form that is not readily meaningful or useful for decision making. We still need to be able to pull together the vast microdata being collected,

[1] http://datascience.nyu.edu/what-is-data-science/
[2] https://datajobs.com/what-is-data-science

generated and produced across the economy into meaningful and actionable knowledge and information which can then be used by:

i. Policymakers for design, implementation, monitoring and evaluation e.g to provide population related services, predict likely emergence of natural disasters or track epidemics;

ii. Businesses to better determine consumer preferences and future demand, based on past demand trends, or to target their advertisements and products;

iii. Entrepreneurs to identify new business opportunities, or future markets based on household consumption profiles;

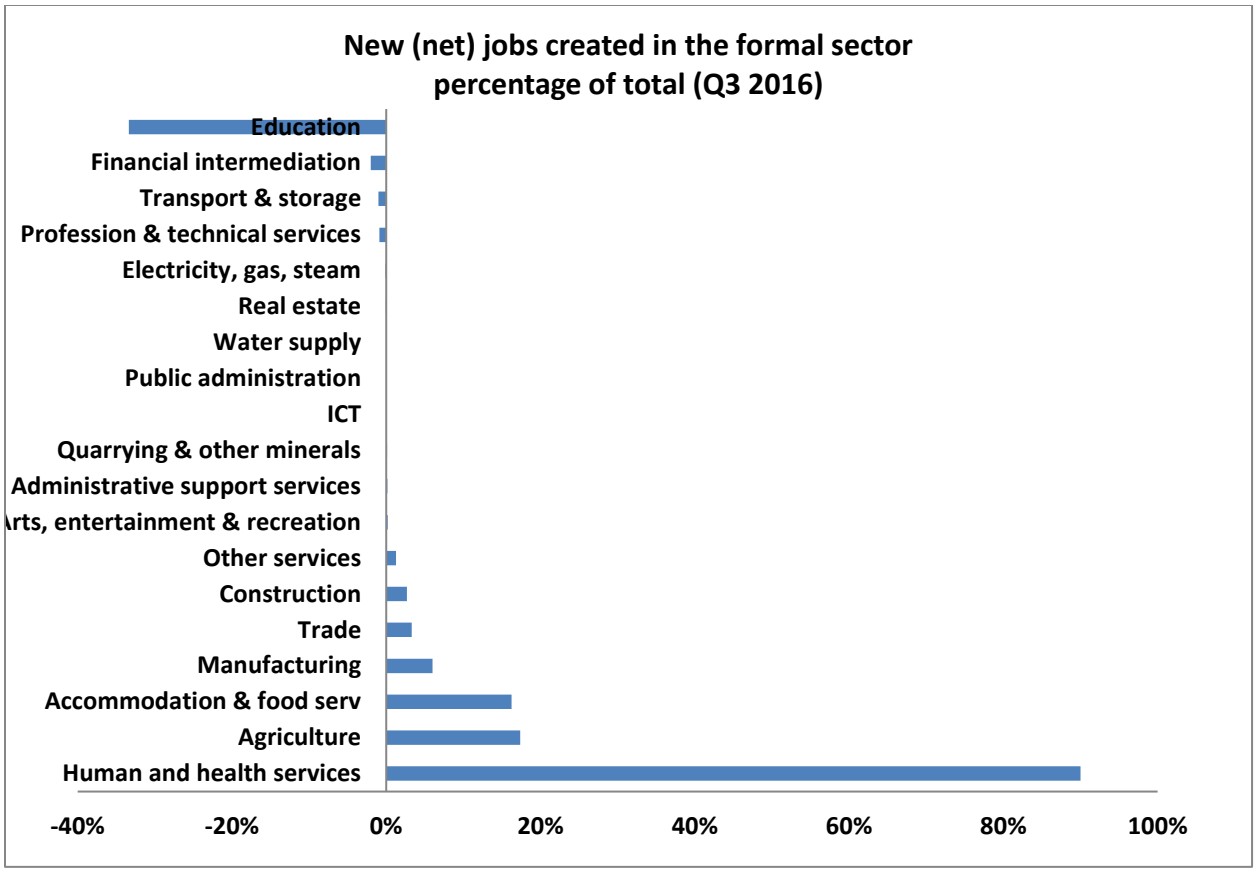
iv. Students to make decisions about future growth in job opportunities in various sectors etc.

The Economic Recovery and Growth Plan (ERGP) like all other plans we have had in our nation's history talks about the need for economic diversification

Can data and data science help to make such economic diversification possible?

Yes, it can as I showed earlier. By applying the tools and methods of data science to various types of microdata, the outcome can be seen in better information, higher productivity, greater efficiency, increased output and ultimately a higher welfare level for society.

Let us consider some sector focused applications and how the tools and methods of data science can foster improved outcomes for the economy as a whole. The sectors to be considered can be identified based on their job creation potentials in Q3 2016. Four sectors were selected from among those that recorded new job creations, while four sectors were selected from those that recorded job losses.



We identified 8 sectors as follows: Human and health services, agriculture, trade, arts, entertainment and recreation, public administration, electricity, gas & steam, transport and storage and financial intermediation (services).

a. **Agriculture**: the usefulness of data in understanding soil types and which are best for growing which crop or in predicting weather patterns and environmental events is a prominent area of application of data science. But in our context, there are a wider variety of other application areas. For example, relying on data analytics for each stage of the value chain could help producers make informed production decisions even before the first seeds are planted. In a practical sense, this would require combining weather information with data on seed/soil properties, transportation/logistic options, business conditions and household purchasing power analysis. The availability and use of such data can enable the modern agricultural entrepreneur (or groups of producers) to guard against both risks from nature as well

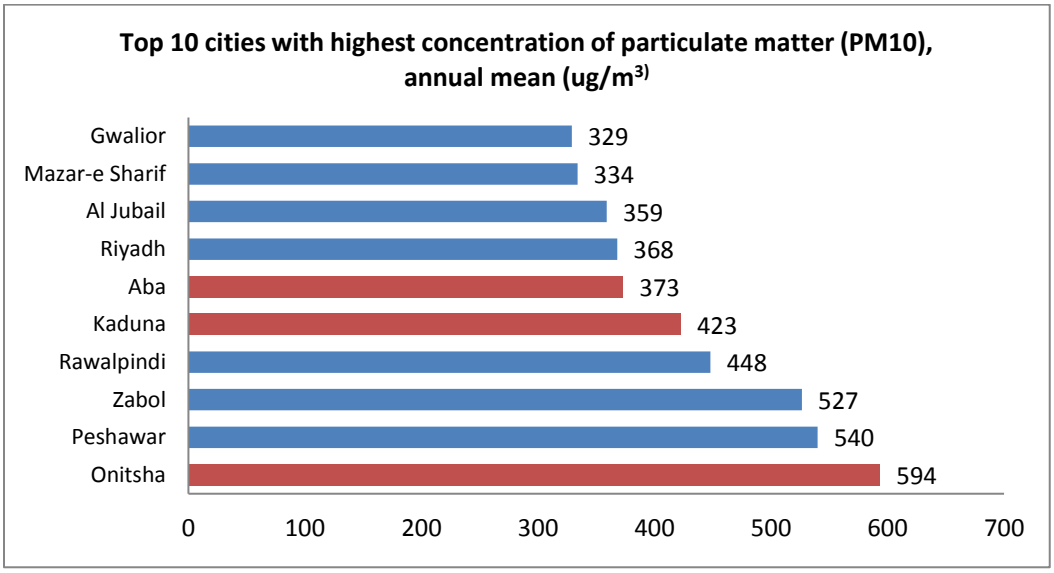
as business/market risks, essentially synchronizing final demand with planting decisions from the start, which contributes to greater efficiency, less wastage, higher output and economic growth.

b. **Health and human services**: the use of big data in modeling disease spread, real time identification of emergencies such as epidemics is already prevalent, following the trail of the Google's flu tracking (GFT) program a few years ago. Using algorithms developed for internet searches within the geographical area of concern, and greater precision, epidemic outbreaks can be more rapidly contained. Other health or environmental emergencies such as major accidents, earthquakes, floods etc can also be detected through analysis of phone and internet data potentially saving lives and property.

c. **Trade and commerce**: The tools of data science are quite widely applied especially in the retail and commercial trade sectors. For instance, customers' purchasing habits both offline and online are being used not only to predict *what* their future preferences, but also *when* their next purchases will be due, enabling businesses to target advertisements and recommendations to such customers. In addition, companies use customer information in the design of new products. Such sophisticated use of customer data improves the sales turnover for producers, enabling them to improve output, sustain employment and contribute to economic growth.

d. **Financial and insurance services**: In a 2013 NBS/SMEDAN collaborative survey, SMEs identified the major challenge they experience to be access to finance, ahead of poor infrastructure, inconsistency in government policies and multiple taxation. At the moment, financial institutions in Nigeria are increasingly hesitant to extend credit to the real sector and non-collateralized borrowers due to high risk and low credit worthiness. With better use of customer data, for example using past expenditures, income flows to analyze probabilities of risk and default, finance and insurance institutions can better provide targeted financial products to specific groups of clients, based on client profiles and purchasing power, thus reducing their own risk exposure while supporting economic activities and growth.

e. **Transportation and urban planning**: data generated by vehicular traffic on a daily basis can be utilized to (re)design transportation routes, and improve urban and city planning. Traffic congestions can be very costly. A 2014 study estimated that traffic congestion cost the US economy about $124billion a year.[3] Congestion also leads to increased concentration of pollutants (such as PM10) which are formed when gases emitted from vehicle exhaust react in the atmosphere. In 2016, the World Health Organisation lists three Nigerian cities among the top 10 cities with highest concentration of particulate matter (PM10) globally. Addressing traffic congestions in Nigeria's largest cities therefore contribute not only to better economic outcomes but also health and environmental improvements.

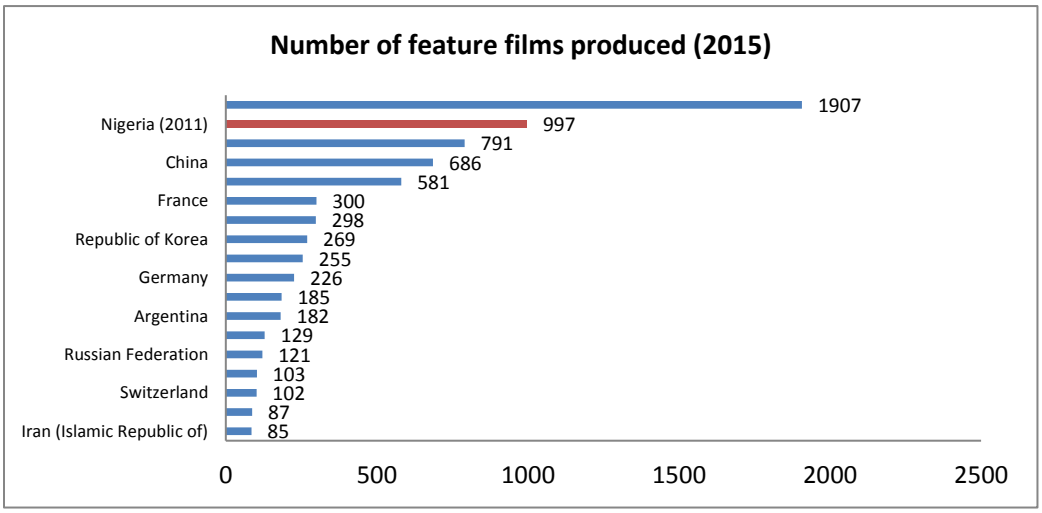


Source: WHO, 2016

f. **Entertainment**: Nigeria's entertainment industry has seen significant growth over the past two decades, and at the moment, the major growth factor points to the role of ICT as enabler of quicker time to market, improving quality, and more importantly access to a wider (international) market. In 2015, Nigeria would have ranked second on UNESCO's list of top feature film producing countries, behind India but ahead of the US, China, Japan and the UK. By leveraging on customer preferences, viewing habits, purchasing patterns and other properties of production, stakeholders can

[3] https://www.forbes.com/sites/federicoguerrini/2014/10/14/traffic-congestion-costs-americans-124-billion-a-year-report-says/#c09182cc107a

generate insights into future purchasing and viewing patterns, including personalized entertainment options, to the benefit of the industry.



Source: UNESCO Institute for Statistics database, 2017

g. **Public administration, security and governance**: Nigeria's security challenges today manifest in various forms including civil disturbances, low level crimes, economic sabotage, financial crimes and terrorism. The prevalence of crime raises the country risk profile, discourages foreign investment and threatens the socio-economic and political stability of the country. By introducing better identity management systems which are accessible to authorized law enforcement agencies at all levels of government, the prevalence of crime can be significantly curtailed. Similarly, the administration of public services such as taxes and other revenue collection, welfare transfers, citizen information services and other public services becomes more efficient and transparent since they are provided on the basis of actual rather than presumed recipients. Such population –related data would also be useful for the design of targeted interventions e.g to the lowest income earning households, families with sick children, members of specific occupational groups, unemployment insurance etc. An improvement in public service delivery and the quality of governance will positively impact on socio- political stability and economic growth in the long run.

h. **Electricity**: In 2015, about 60% of the Nigerian population had electricity access. According to the International Energy Agency, Nigeria has the second highest number of people without electricity access, next to India. One of the most critical constraints to economic growth concerns the electricity sector. Across industry, commercial businesses and households, the poor power supply situation severely limits production efficiency while increasing costs for many producers. Better energy sector planning can be achieved with better use of already available data as a significant quantum of data is already being generated and collected by electricity distribution companies regarding household consumption patterns across the country, at least for households with electricity meters. Such data, when matched with data on socio-economic demographics and urban planning can be useful for investment decision making by utility companies and government in the on-going efforts to expand the electricity grid.

In addition to the economic activities I have discussed earlier, it is also worth noting that the Federal Government's economic development agenda, the Economic Recovery and Growth Plan (ERGP) prioritizes the role of data analytics in addressing certain sectoral issues some of which include:

i. Solid minerals: Increase access to information by improving the archiving of geo-data, harmonizing their format, and promoting their dissemination
ii. Information and communications technology: Promote the use of e-governance and digitize Federal Government data
iii. Power: Implement a data-driven approach in power sector development planning
iv. Environmental protection: Establish a functional database on drought and desertification

5. Key skills for data scientists

Nevertheless, it is important to also keep in mind the human resource base required to achieve these desirable outcomes. As much as technological advancement can drive economic growth, skilled workers are also required to apply the science, operate these technologies and drive continued innovation. In the field of data science, the skills required cover three major areas:

i. Mathematics / statistical sciences
ii. Technology/computer sciences
iii. Applied fields (e.g communication, business, sports, engineering, religion, development etc)

To become an effective data scientist therefore, you need a solid foundation in computer science, modeling, statistics, analytics and math. What sets data scientists part from traditional job titles, therefore, is an understanding of economic and business processes and an ability to communicate findings to business and IT leaders as well as economic managers in a way that can influence how an organization, individual or business approaches a challenge or exploits an advantage.

While the field of data science is still emerging as a definitive field of practice, there are actions that can be taken by key stakeholders in ensuring that Nigeria continues to build a pipeline of talent with expertise in the above areas to prepare today's youths for tomorrow's jobs. One of such actions is what we are doing here today.

The key stakeholders identified in this regard include:

i. Government
ii. Employers
iii. Educational institutions and
iv. Professional associations

Some recommended strategies towards building the required human resource base include:

i. Strengthening educational curriculum from secondary level to increase analytical content of educational programmes
ii. Improving collaboration between institutions, employers and professional associations to ensure that industry requirements and job descriptions are matched with educational and training programmes of students prior to joining the workforce

iii. Increasing investment in on-the-job training for local hires to build local capacity rather than relying on foreign workers; and

iv. Step up demand and utilization of evidence based decision making at all levels of public governance to ensure that the supply of data scientists is matched by demand in both public and private sector.

6. Conclusion

I would like to conclude by noting that Nigeria can become a developed nation only if everyone contributes to the best of his or her ability and capacity. The task of government is to set a strategic vision for the nation and to design policies, plans and programmes to drive that vision. We all must then respond to that vision and to give that vision flesh and bones and to advance ideas to translate the vision into reality.

In many cases, however, we find that in Nigeria, it is the general public and the private sector that even sets the tone and dictates the direction. This is regrettably the case with dealing with big data and data analytics in Nigeria and this is why I find these engagements very useful. As Nigeria's Statistician General, I believe Data Society Nigeria is leading the way in this regard and should be commended for its efforts.

Accordingly, I am here not only to speak to you but also to pick your brains and hopefully learn a few things from the vast amount of knowledge we will all gain from the many experienced speakers and interaction from the audience.

As a nation, I think we face two overall choices: we can either drift passively into the future, or we can plan actively and deliberately by providing and using data and information that enables us to design policies plans and programmes for development. As participants in this programme, you are setting yourselves apart as building blocks upon which the emerging field of data science in Nigeria will be solidly established. It is my hope and genuine expectation that the resources, knowledge and network of relationships you have acquired over the past few months and in the course of the remaining part of this summit will remain with you long after you have left this place. I also invite you to engage with our data products on the NBS website and to make it your

playground. I urge you to also remember that the country is looking up to you to use your skills and knowledge to lift up the millions who would have loved to be here, but have not been as fortunate as you are.

I wish you all a fruitful and successful session. Thank you all for listening.