

The Potential Role of Artificial Intelligence in Capturing Census Data

Student Name and Number: Warren Mushonga 14020012

Supervisor: Dr Hennie Pelser

Module Name and Code: Research Methodology RESM8419p

Qualification: Bachelor of Commerce Honours in Management

I hereby declare that the research report submitted for the Bachelor of Commerce Honours in Management degree to the Independent Institute of Education is my own work and has not previously been submitted to another University or Higher Education Institution for degree purposes.

Word count: 9298

Table of Contents

Abstract	4
List of Key Terms and Abbreviations	5
Acknowledgements	6
Chapter 1	7
Introduction.....	7
Background.....	7
Rationale.....	7
Problem Statement	8
Research Goal	8
Research Questions	8
Research Objectives	8
Summary.....	9
Chapter 2	9
Literature Review	9
Introduction.....	9
Theory of Counting	10
Theory of Measure	10
The Importance of Census Accuracy	10
Forms of Artificial Intelligence in Census	11
Factors Leading to Undercounts.....	12
Advantages of Using Artificial Intelligence	13
The Impact of Inaccurate Census Data	14
Summary.....	15
Chapter 3	16
Methodology	16
Introduction.....	16
Research Paradigm	16
Research Design	16
Research Plan	17
Population	17
Unit of Analysis.....	17
Sampling	18
Sampling Method and Size	18
Data Collection Methods.....	18

Data Analysis	18
Chapter 4	20
Findings and Interpretation of Findings	20
Theme 1: Effectiveness and Efficiency	20
Theme 2: Lack of Cooperation.....	21
Theme 3: Misinformation.....	22
Theme 4: New Data Collection Methods.....	22
Theme 5: The Future	23
Link Findings to Prior Literature	24
Link Findings to Prior Theory	25
Ensuring quality of this research	26
Trustworthiness.....	26
Summary.....	27
Chapter five	28
Conclusion	28
Indication of how the research questions were answered.	28
Indication of how the research problem was solved	29
Indication of how the research goal was reached.....	30
Implications of findings for future practices.	30
Heuristic value of the study.....	30
Ethical Consideration.....	30
Limitations	31
Conclusion	31
References.....	32
ANNEXURE A: Consent Forms	35
ANNEXURE B: Research Questionnaire	38
ANNEXURE C: Permission letter	39
ANNEXURE D: Ethics Clearance	40
ANNEXURE E: Originality Report	42
ANNEXURE F: Research Report Document Table.....	43

Abstract

“For too long, African leaders have undermined the growth of Africa. They lost the plot of measurement because they could not count numbers” (Lehohla, 2011). Census plays a critical role in the development of a country. The government needs to use census data for planning purposes for example to provide grant payments and service delivery. Since the world is gearing up for the Fourth Industrial Revolution, it is crucial for Statistics South Africa to enter into a transitioning phase of using artificial intelligence in capturing census data. The focus of this research is that the traditional census methods used by Statistics South Africa to capture census data have resulted in undercounts. This research used an interpretivist, qualitative approach to promote a deeper understanding on the potential role of artificial intelligence in capturing census data. This research used in-depth semi structured interviews as a data collection method, which allowed the researcher to generate the research findings. The most significant finding of this research is that artificial intelligence has the potential to produce accurate census data compared to traditional census methods. Lastly, this research can enlighten readers on the importance of accurate census data and the use of artificial intelligence.

List of Key Terms and Abbreviations

Census – an official count of a population.

Artificial Intelligence – is the ability of a computer system to think and learn.

Under counts – a count of a population that is inaccurately low.

Over counts – a count of a population that is higher than the actual figure.

Accuracy – the state of being precise.

Stats SA – Statistics South Africa

SASSA – South African Social Security Agency

SARS – South African Revenue Service

RDP – Reconstruction and Development Programme

AI – Artificial Intelligence

CAPI – Computer Assisted Personal Interview

CAWI – Computer Assisted Website Interview

CATI – Computer Assisted Telephone Interview

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Hennie Pelsers for his astonishing support and guidance he provided throughout the past 18 months. He has made it possible for me to complete my research in time despite the fact that there was a major setback due to the Corona virus. Secondly, I would like to thank Statistics South Africa for allowing me to conduct interviews with their employees, for I have gained knowledge on the importance of census. Thirdly, I would like to thank my class colleagues for the teamwork they displayed which has motivated me to finish this research. Lastly, I would like to thank everyone at the Independent Institute of Education who have contributed in making this research possible.

Chapter 1

Introduction

“There is an urgent need to develop a far reaching view of how technology is affecting our lives and reshaping our political, social, economic and cultural environments” (Klaus, 2018, pp. 1-2). In this chapter, this research will give a background on what the study is about and provide a problem statement in order to explain the importance of this research topic. Questions and objectives of this research will give a better understanding on what the researcher is looking for.

Background

Prior to the advent of democracy, there was a lack of reliable information about South Africa as a whole (Lehohla, 2011). The old Statistics Department did not have the total population figures and this acted as a barrier for the newly elected government to plan. Knowing the actual population figures is important because it assists the government in creating employment and providing service delivery. Since the first national census of 1996, Statistics South Africa has been struggling to capture accurate census data. Largely, this is due to the traditional methods that the department had been using. The traditional methods used by Statistics South Africa have relied mostly on human labour.

There are growing tensions between the public and the government that have resulted from poor service delivery, lack of educational opportunities and a slow distribution of RDP houses. The public has been conducting strikes and demonstrations across the country with intentions to make their voices and concerns heard by the government. Statistics South Africa plays a critical role in the growth and development of the South African economy. Traditional methods that Statistics South Africa has used for the past two decades have produced inaccurate population data. Since the world has entered into the Fourth Industrial Revolution, it is time to shift from manual census methods to modern census methods that use artificial intelligence. This research used different techniques to support this study, which include literature review, research findings and research methodology.

Rationale

This research is relevant because it is identifying whether artificial intelligence can assist Statistics South Africa in capturing accurate census data. Accurate census data enables the government to plan on how it can provide service delivery and create employment. Accurate census data can help the government to set an actual budget that would help resolve

economic issues such as unemployment and lack of service delivery that are currently affecting the majority of the South African population. Traditional census methods previously used by Statistics South Africa have been failing to produce reliable census data. This research is trying to explore if artificial intelligence can address some of the challenges faced by Statistics South Africa.

Problem Statement

Over the past two decades since the first national census, Statistics South Africa has relied on traditional methods to capture census data. These traditional methods have led Statistics South Africa to record undercounts. Inaccurate population figures have affected the planning process of the government resulting in poor service delivery and lack of grant payments. Can the introduction of artificial intelligence help Statistics South Africa to capture census data that is accurate and reliable? Lehohla (2011) the former Statistician General of South Africa mentioned that in order for Statistics South Africa to capture accurate census data, the department needs to shift from traditional census methods to modern census methods. Despite the advancement in artificial intelligence and the need for accurate census data, the value of artificial intelligence as part of the future census purpose has yet to be determined

Research Goal

The main goal of this research is to determine whether artificial intelligence can enable Statistics South Africa to capture and produce accurate census data.

Research Questions

1. Can artificial intelligence help Statistics South Africa to obtain accurate census data?
2. Can artificial intelligence lower the costs of conducting census since most of the work will be executed using technology?
3. Can the introduction of artificial intelligence help Statistics South Africa to solve problems that occur during capturing census data?

Research Objectives

1. To identify whether artificial intelligence can help Statistics South Africa in capturing accurate census data.
2. To investigate whether AI has the potential to address the deficiencies associated with traditional census methods.
3. To identify if the problems faced by the statistics department during census can be solved by the introduction of AI

Summary

Traditional census methods used by Statistics South Africa need to change because they are not capturing reliable data that the government can work with. The core objective of Statistics South Africa is to capture accurate census data. The department needs to consider alternative methods that will enable them to achieve their objectives. In the next chapter, this research will include the literature review in order to provide a foundation of knowledge for this research topic.

Chapter 2

Literature Review

Introduction

What is the Fourth Industrial Revolution and why is it important? Devon (2018) describes it as the fusion of advances in artificial intelligence, 3D printing, genetic engineering and robotics. The Fourth Industrial Revolution is important because it improves productivity and eliminates the use of manual processes. This literature review is exploring the potential role that artificial intelligence has on capturing census data. There are five literature sources used below that enabled this research to conclude the extent to which artificial intelligence has on capturing census data. The five researchers from the sources are Yvonne (2011) , Mark (2014), William (2019), Bernard (2012) and Catherine (2018).

The Fourth Industrial Revolution represents an institutional change in the way we work, live and interact with each other. It is a new chapter in human development enabled by remarkable technology. The speed of this revolution is forcing humans to rethink how organisations should generate profits and how countries should develop. Countries such as China, South Korea and Singapore have become world powers in a fraction of time because they have managed to adapt to the new chapter much faster than most countries (Klaus, 2018). The impact of the Fourth Industrial Revolution has led organisations to cut millions of jobs worldwide because artificial intelligence has replaced humans in most employment sectors. The Fourth Industrial Revolution has faced many criticisms in recent times for example; Rebecca (2020) suggested that the Covid-19 virus originated from 5G network which was recently invented by a Chinese telecommunication company, Huawei.

“Technologies that are emerging today will shape the world tomorrow and into the future. Since we have entered the fourth industrial revolution, it is important that we discuss and ensure that humanity is served by the remarkable technologies invented so that we can continue to prosper” (Mariette, 2019, p. 2).

Theory of Counting

Kimberlee (2015) defined theory of counting as a counting principle that seeks to find the number of outcomes in a probability problem. It is common to face problems when using manual methods to conduct census. It is important for an organisation to identify problems before partaking any form of work (Catherine, 2018). When conducting a census, errors can have a negative impact on the accuracy of the captured data. There are different forms of artificial intelligence that can be used in census to predict uncertainties for example the GUM tree calculator and the Metro-data GmbH GUM Workbench (Blair, 2020).

Theory of counting is important for Statistics South Africa because it identifies and minimize possible errors that might affect the accuracy of the census data. Yvonne (2011) explained that data captured by Statistics South Africa needs to be accurate because the public and private sectors use it for planning and decision making purposes. However, this theory can be a vital tool that Statistics South Africa can use when conducting census with an aim to identify the problems that they might encounter.

Theory of Measure

Eric (2020) defined theory of measure as a mathematical and statistical theory that generalizes on aspects such as length, area and volume. It is important for enumerators to know the topography of the area they are working on. However, Yvonne (2011) stressed a notion on the importance of proper planning and measuring before conducting census. Yvonne (2011) explained that one of the main challenge faced by Statistics South Africa is penetrating terrains that are difficult to navigate. She concluded that Statistics South Africa needs to shift to modern census methods that use advanced technology to determine areas in which enumerators should work in.

The Importance of Census Accuracy

Judging from a larger perspective, both Yvonne (2011) and William (2019) strongly emphasized the importance of census accuracy. Yvonne (2011) mentioned that a census should record a 100% population count in order for the government to set a budget for economic activities such as service delivery and educational grants, while William (2019)

states that there are countless decisions made in the private and public sectors that depend on data collected from census. William (2019) added that census is the backbone of a government's democratic system whereby the state relies on the census data to distribute state funding each year to different provinces.

Yvonne (2011) mentioned that some of the challenges faced by the South African government are caused by the lack of accurate and reliable census data. Yvonne (2011) explained that poor service delivery experienced in the country are a result of shortage of reliable information to plan properly. Inaccurate census data has affected districts such as the Capricorn of the Limpopo province whereby the government has delivered services to a handful of people instead of supplying to all the people (Yvonne, 2011). Traditional methods of capturing census have a lesser probability of capturing accurate census data in townships because of overpopulation.

These challenges are the very reason why countries across the globe have opted for alternative means of collecting and capturing data. One of these alternative means is the use of artificial intelligence.

Forms of Artificial Intelligence in Census

Bernard (2012) conducted a research of how the United Kingdom, Malaysia and Netherlands have managed to shift from traditional census methods to modern methods. Bernard (2012), explained that Netherlands invented a system called the Virtual Census. The Virtual Census consolidates records from administrative and statistical operations such as hospitals, schools, passport/ID centres and employment sectors. The Virtual Census has allowed Netherlands to have census data with timeliness which has enabled the Dutch government to plan. Since the introduction of the Virtual Census, Netherlands moved away from relying on enumerators to conduct door to door census since the system combines different registers across the country and produces data which the public and private sectors can work with.

Moreover, Bernard (2012) explained how the United Kingdom shifted from traditional census to modern census. Previously, the United Kingdom relied on enumerators to conduct door to door census in order to get participants to complete questionnaires. In 2011, the United Kingdom introduced a smart method of capturing census data. The new method was carried out by the census organisation posting questionnaires to every house hold in Britain. Participants were required to send back the completed questionnaires online. In situations

where participants failed to send the questionnaires back, the organisation would electronically track the questionnaires in order to inform participants to return back the completed questionnaire. Since the United Kingdom introduced AI into their census, they have managed to produce reliable and accurate data.

Furthermore, Bernard (2012) explained how Malaysia shifted to modern census. In 2010 Malaysia conducted a national census that enumerated individuals on a de jure basis instead of de facto basis which was previously used. Malaysia introduced a multi modal data collection method. One of the methods allowed participants to complete questionnaires electronically. In situations where participants could not use the online platform, enumerators were sent to assist them to complete the questionnaires electronically.

In relation to Bernard (2012), Catherine (2018) explained how the United States of America shifted from traditional census to modern census. Catherine (2018) explained that United States of America created a system known as the linkable 1940 census microdata file which comprises of administrative records from federal agencies and surveys. Similar to the Virtual Census of Netherlands, the 1940 microdata file consolidates records from different registers and it has enabled the government of the United States of America to distribute federal funds across the country fairly.

Factors Leading to Undercounts

Previous census data recorded by Statistics South Africa indicate that they have been undercounts in rural and farming areas. Mark (2014) and Yvonne (2011) both indicated factors that lead to undercounts in census. Mark (2014) explained that it is difficult to navigate and access most locations in rural and farming areas across South Africa. Enumerators end up asking local residents for directions and in many cases the residents were not cooperative.

Yvonne (2011) argued that using traditional methods to capture census data in rural areas is very difficult because of the housing patterns. Yvonne (2011) added that in most rural locations, there is a scattered housing pattern whereby houses are far away from each other and enumerators end up missing them.

In addition, Mark (2014) and Yvonne (2011) both explained that since the first national census of 1996, Statistics South Africa has been estimating the number of people living in communal areas. Mark (2014) added that they have been undercounts recorded in

provinces such as Eastern Cape, Kwazulu-Natal, Limpopo and Mpumalanga. He also indicated that the reason for undercounts is that during census, many people move around and this makes it difficult for enumerators to track them down and get them to complete the questionnaire. Mark (2014) estimated that 60% of undercounts recorded by Statistics South Africa come from communal areas

In relation to what Mark (2014) explained, Yvonne (2011) mentioned that in Limpopo, rural people do not like to disclose information about their life because of certain cultural beliefs. This challenge makes it difficult for Statistics South Africa to capture data from every household. It is clear from the comparison between Mark (2014) and Yvonne (2011) that Statistics South Africa has to change its traditional approach and adopt a new modern method that uses artificial intelligence.

From under counts to poor service delivery, technology will play a critical role in finding solutions for many challenges the world faces today (Jurgens, 2019).

Advantages of Using Artificial Intelligence

The coming of the fourth industrial revolution has enabled organisations around the world to stimulate the growth of their businesses and profits due to the fact that many jobs are being replaced by machines. The introduction of the Virtual Census of Netherlands resulted in many jobs being lost whereby the system replaced most of the labour done by enumerators and people who previously counted census (Bernard, 2012). It is without a doubt that Netherlands saved millions of revenue since the introduction of artificial intelligence into their census process. Moreover, Bernard (2012) mentioned that Malaysia and the United Kingdom introduced census methods that allowed participants to self-enumerate using online platforms. These census methods are less costly to operate since participants complete questionnaires on their own.

Artificial intelligence is timeliness. Machines are much faster than human beings and work is conducted effectively and efficiently. Bernard (2012) and Catherine (2018) explained different forms of AI used in Netherlands and America which consolidate data from different registers across the country and produce information within a fraction of time. The Virtual Census and the 1940 Linkable Micro-Data File of Netherlands and America respectively, do not require human labour to gather data instead the system does it on its own thus saving time.

Artificial intelligence is accurate. Machines have a lesser chance of making errors than human beings. Bernard (2012) and Catherine (2018) explained how countries such as Netherlands, Malaysia, United Kingdom and America have managed to capture census that is accurate since the introduction of artificial intelligence into their census processes. The Virtual Census consolidates records on its own and it produces data that is much reliable which the Dutch government has management to work with (Bernard, 2012). The 1940 linkable Micro-data of America has enabled the American government to distribute government funds across the country fairly (Catherine, 2018).

In summary, artificial intelligence has a higher probability of producing accurate census data than manual methods. It is crucial for Statistics South Africa to shift from traditional methods of capturing census data to modern methods that use artificial intelligence. Artificial intelligence has the capability of solving many challenges that the government of South Africa is currently facing for example lack of service delivery, lack of government funds and poor distribution of RDP houses.

The Impact of Inaccurate Census Data

Yvonne (2011) and Mark (2014) both agree that institutions like municipalities, government departments and tertiary institutions are experiencing challenges mostly because of the incompetency by Statistics South Africa. Yvonne (2011) explained that Statistics South Africa has created difficulties for the government to allocate grants to students and SASSA payments to the elderly. The difficulties are due to the lack of actual census data in which the government can plan and allocate a budget that will be sufficient. Mark (2014) stressed a notion that many municipalities for example Msunduzi and Ethekewini fail to offer service delivery to every location because they work with an unreliable data captured by Statistics South Africa.

Moreover, Yvonne (2011) and Mark (2014) pointed out that there is a large number of people that are not included in the unemployment rate because of the unjust system. There is an urgent need of shifting from the current traditional methods used by the Statistics South Africa to more modern methods that rely on artificial intelligence.

Yvonne (2011) and Mark (2014) both indicated that there is a compelling need of an alternative system that will replace the traditional method in order to tackle serious economic and social challenges faced by South Africa. Yvonne (2011) explained that the reason why the government is failing to build enough RDP houses for every poor citizen is that they are

working on an estimated and inaccurate census data captured by Statistics South Africa. Yvonne (2011) added that in rural areas there is a shortage of clean water because the government drilled few boreholes that cannot supply the whole communities. In relation to Yvonne (2011), Mark (2014) explained that the South African government has been failing to measure the actual extent of poverty in the country due to the lack of reliable information. The incompetency of Statistics South Africa leaves masses of South Africans vulnerable to poverty and unemployment whereby the government sets a budget that will only help a handful instead of the entire community.

Bernard (2012), Yvonne (2011), Catherine (2018), William (2019) and Mark (2014) all agreed that census data is the backbone that supports any government to make decisions. Inaccurate census data not only does it affect the decision-making process but stunts the growth of the economy. Statistics South Africa needs to shift to a modern census approach if it needs to save the South African economy (Yvonne, 2011). Mark (2014) added that recessions that have occurred in South African are due to the lack of appropriate information and this has forced the government to give fluctuating budget figures every year for economic activities like grant payments, RDP housing and SASSA payments. Yvonne (2011) mentioned that the South African national budget fluctuates every year because Statistics South Africa gives them an estimated figure that cannot meet the demand.

Summary

The five sources confirm that using artificial intelligence to capture census data is the only way Statistics South African can obtain reliable census data for the whole country. Geographical locations such as Kwazulu-Natal and Eastern Cape have topographies with landscapes that are hard to navigate and the only solution is to introduce systems from countries like Netherlands whereby enumerators no longer required to conduct door-to-door census but instead the census organisation uses a system that gathers data from different registers. South Africa needs to adopt to systems introduced by other countries because they produce reliable information and there are less costly.

Nevertheless, it is unknown to measure the extent to which if artificial intelligence can capture census data that is 100% accurate. Basing from the five sources used, artificial intelligence produces a figure that is more reliable in which governments can easily work with. Unlike traditional methods, artificial intelligence has a lesser chance of making errors.

In the next chapter, this research will focus on the methodology to evaluate the overall trustworthiness of this study.

Chapter 3

Methodology

Introduction

Wilkinson (2000) defined research methodology as a technique used to identify, select, and analyse information about a specific research topic. This research used Microsoft Teams as a data collection tool for interviews. A thematic analysis process was used to identify, examine and record themes generated from the interviews that were undertaken. Other techniques used include purposive sampling, accessible and target population, unit of analyses, interpretivism paradigm and grounded theory as a research design.

Research Paradigm

A research paradigm as a set of assumptions and views shared by a research community (Corne & Rose-Marie, 2013). Paradigms are important because they provide researchers with an idea to select research designs and methods. In order to confirm the relevant paradigm for this study, this research considered ontology and epistemology. Ontology refers to the nature of our beliefs about reality. Researchers have implicit assumptions about reality on how it exists therefore it is an ontological question that directs a researcher to examine the kind of reality that exists (Corne & Rose-Marie, 2013). Epistemology is the philosophical theory of knowledge (Corne & Rose-Marie, 2013). Epistemology is concerned with ways of studying and knowing the social reality.

However, this research used an Interpretivism paradigm. Interpretivism paradigm stresses that reality is complex and a single phenomenon can have more than one interpretations (Eric, 2020). Moreover, interpretivism paradigm uses naturalistic approaches of collecting data such as interviews. Interpretivism paradigm are underpinned by interpretations and observations, thus to interpret is to create sense of information by drawing inferences, while to observe involves collecting valid information from events (Corne & Rose-Marie, 2013).

Research Design

Corne (2013) defined research design as a set of processes used in gathering and analysing measures of variables mentioned in the research problem. This research used a grounded

theory design. Merriam (2011) defined grounded theory as a design that gathers and analyses data. Moreover, grounded theory involves interviewing participants with open-ended questions. The importance of using this design is it allows the researcher to use focus groups and fieldwork when researching. Furthermore, grounded theory falls under qualitative research.

Grounded theory design enabled the researcher to further the theoretical reach of this research and to make explicit processes and meanings tacit. Grounded theory made the ethnography to be more analytic and content analysis more focused. Since grounded theory focused on data collection, it enabled the research interviews to be more in-depth.

Research Plan

A research plan allows a principal investigator to discuss the proposed research (Plooy-Cilliers, 2013). The research plan for this study consisted of unit of analysis, population, sampling, data collection and data analysis. These techniques identified selected, processed and analysed the information from participants.

Population

Target and Accessible Population

Du Plooy-Cilliers (2013) defines target population as a collective group of people from which a sample might be drawn. Accessible population is population that a researcher can study and draw a conclusion (Corne, 2013). Nevertheless, the research target population were managers from Statistics South Africa in Kwazulu-Natal.

Population Parameter

A population parameter is a number that encapsulate various aspects of the whole population (Paul, 2008). The geographic location of this research was based in Pietermaritzburg which is a city found in the province of KwaZulu-Natal. This research focused on the managers of Statistics South Africa in Pietermaritzburg. The Population of this research were people aged between 40-60 years of age. The Gender of the population was both male and female.

Unit of Analysis

Corne and Rose-Marie (2013) defined unit of analysis as the entity that frames the investigation of the research. This Research Unit of analysis where managers from Statistics South Africa in Pietermaritzburg.

Sampling

Helen (2009) defined sampling as a process that involves collecting samples of something for the purpose of analysis.

Sampling Method and Size

Purposive Sampling

Foley (2018) defined it as a non-probability sampling method in which researchers select population members based on their own judgement. However, this research used purposive sampling whereby the researcher selectively chose managers from Statistics South Africa. The advantage of this method is that it is straight forward thus saving time and money during sampling. This research-sample size was three managers from Statistics South Africa.

Data Collection Methods

Data collection is a procedure of gathering information from various sources with an aim to find answers to a research problem (Bhat, 2018). This Research used in depth semi-structured interviews and a set of open-ended questions were questioned to participants from Statistics South Africa.

This research used a social network platform called Microsoft Teams to conduct interviews. Microsoft Teams enabled the researcher to observe the physical reaction of participants. Moreover, Microsoft teams allowed participants to communicate with features like memes, GIFs and emoticons that enabled the researcher to have a better understanding of the participant's feelings. However, all the recordings from the interviews were stored in a password protected hard drive which only the researcher has access to.

Data Analysis

Data analysis is the process of transforming data with an objective to identify useful information (Corne & Rose-Marie, 2013). This research used labels, themes and categories to code data accumulated from interviews in order to present empirical findings. Lastly, the researcher used existing theoretical frameworks as a way of confirming and supplementing empirical findings. This research used thematic analysis with a six-step process prescribed by (Braun & Clarke, 2006).

Step one – Familiarising yourself with your data

The first step involved the researcher transcribing the interactions. After transcribing the interactions, the researcher read the gathered transcripts several times (open coding) and listened to the recordings from the interviews in order for him to get a sense of what it

contained. Moreover, the researcher familiarised himself with all aspects of the data and had a comprehensive understanding of the content. This step was very vital for it provided a foundation of the subsequent analysis.

Step two - Generating initial codes

After reviewing and exploring the transcripts, the researcher began to identify preliminary codes (axial coding). Axial coding is a process of identifying relationships among open codes (Bezuidenhout, 2014). The codes identified were specific than themes and they provided an idea of the context of the conversation generated during interviews.

Step three – Searching for themes

This process involved the researcher examining collated codes in-order to identify themes that were relevant to the study. The researcher combined relevant data extracts according to overarching themes.

Step four – Reviewing themes

In this step, the researcher ensured that his themes were accurate and relevant representations of data. The researcher had a checklist to see if he had missed anything, if his themes were present in the data and to see what he needed to change for his themes to work better.

Step five – Defining and naming themes

After obtaining a final list of themes, the researcher named and defined each theme. Braun and Clarke (2006) explained that naming themes involves coming up with a concise and understandable theme names.

Step six – Producing the report

In this final step, the researcher transformed his analysis into a piece of writing that was interpretable by using examples that related to the themes, literature and research questions.

Thematic analysis allowed the researcher to approach large sets of data more easily by sorting them into broad themes. The next chapter will discuss the findings of this research.

Chapter 4

Findings and Interpretation of Findings

The information from the research findings came from three participants whose identities remain protected in adherence to the ethical policies from the researcher's College. All participants were ensured that their personal information will be protected.

The findings of this research were generated from three interviewed participants from Statistics South Africa. Participant A is an African man in his mid-fifties who holds an MBA degree. He is part of the top management at Statistics South Africa with more than 10 years in the Statistics field. The Second participant was an African woman who is also part of the managerial department at Stats SA. She holds a Master's Degree in Demography and population studies with more than 14 years of experience in the Statistics field. The third participant was an African woman in her mid-forties who holds an MBA degree with 10 years of work experience. However, the researcher conducted a six-step thematic analysis process after the interviews concluded and five themes emerged.

Discussion and Analysing of Themes

Theme 1: Effectiveness and Efficiency

Question 1: What are your thoughts about the potential role of artificial intelligence and machine learning in addressing some of the challenges faced by Statistics South Africa? The most prominent theme that emerged was that machines are effective and efficient when conducting work. All the participants agreed that artificial intelligence has the potential to solve the challenges faced by the organisation. They all agreed that artificial intelligence is effective in the sense that it can minimise chances of making errors during the census process and efficient whereby it is cost saving. The following is an illustration of the response from the two participants:

Participant A: *When I think about artificial intelligence, two things come into my mind, efficiency and effectiveness. Artificial intelligence is the solution to current challenges that we face as an organisation due to the following reasons:*

- *Machines are faster and time saving.*
- *Artificial intelligence enables the data collection process to be cheaper.*
- *Artificial intelligence is much safer because since data is collected remotely.*

Participant B: *It is without a doubt that machines are much effective and efficient than human labour. As an organisation, we have introduced new methodology tests that rely on artificial intelligence which were on pre-trial for the upcoming 2021 national census. It is safe to say that these three data collection methods have proved to be effective and efficient.*

Participant C: *Our mission as Statistics South Africa is to lead in statistical systems and this is only achievable by using census methods that rely on artificial intelligence since they are effective and efficient. In the past, enumerators used to give false data to the organisation whereby they will not go to the areas which they were assigned to. Thanks to artificial intelligence, we now have a system that can divide areas in which enumerators can work in and also it is capable of tracking their movement.*

In Summary, it is clear that Statistics South Africa acknowledges that AI has the capability of solving some of the challenges they face.

Theme 2: Lack of Cooperation

Question 2: What are the challenges that you face when capturing census data? Both participants agreed that lack of cooperation from the public is a major challenge they encounter during capturing census data. Both participants mentioned that the public does not take census seriously and this affects the reliability of the data captured.

Participant A: *Although census is mandatory, the public still takes it lightly whereby there are few individuals that do not cooperate with us when capturing census data.*

Participant B: *Our major challenge as an organisation is lack of cooperation from the public. Many undocumented foreign nationals in South Africa avoid the questionnaires because they are afraid that Home Affairs might get hold of their information and deport them.*

Participant C: *Our greatest challenge is to make the entire population of South Africa take part in census. During census, many people have a habit of moving up and down and they end up missing the questionnaire. Lack of cooperation from the public makes it impossible to capture accurate data which the government can work with.*

In summary, Statistics South Africa has done a lot to ensure that the public knows of the importance of census taking. The challenge is convincing people to take part in the census process. Artificial intelligence can be the antidote to their major challenge whereby they can adopt a system similar to the virtual census system of Netherlands that gathers data from

different registers. With such a system, you do not need to ask for permission to fill in the questionnaire but allow the system to gather data on its own since census is mandatory.

Theme 3: Misinformation

Question 3: What are the main reasons that cause some of the challenges you encounter when capturing census data? From the three different responses of the participants, the theme that emerged was misinformation. All the participants agreed that there are few misinformed individuals in South Africa concerning the census process. All the participants agreed that as an organisation, they conduct awareness campaigns in order to enlighten the public about importance of census. The following is an illustration of the responses:

Participant C: *A minority of individuals in South Africa are misinformed about the reasons why we conduct census. Many people owe organisations such as SARS money and they believe that by giving us their information Statistics South Africa will then report them to SARS.*

Participant B: *A minority of people in South African believe that Statistics South Africa sells information to other organisations such as SARS. Such misinformation has a negative impact on us as an organisation.*

Participant A: *The reason to our challenges has to be misinformation. Many immigrants from countries like Zimbabwe, Malawi and Zambia do not feel safe to participate in census since they are in the country illegally.*

In summary, misinformation about why Statistics South Africa conducts census hinders chances of capturing accurate census data. The organisation needs to come up with better strategies to ensure that every individual in South Africa participates in the census process.

Theme 4: New Data Collection Methods

Question 4: What measures have you put in place to tackle these challenges? The three participants mentioned that the organisation implemented new data collection methods that will tackle some of the challenges they face. These include: the computer assisted website interview, computer assisted telephone interview and the computer assisted personal interview.

Participant A: *As an organisation, we have recently introduced three data collection methods that we will be using in the upcoming 2021 national census. These three data collection*

methods were on trial for the 2021 national census and they have proven to be accurate and reliable.

Participant B: The organisation introduced three data collection methods in order to tackle some of the challenges the organisation faced in the previous censuses. The new data collection methods are namely: the computer assisted telephone interview, the computer assisted website interview and the computer assisted personal interview.

Participant C: *We have introduced new data collection methods that will enhance the effectiveness of the census process. Some of the challenges that we faced will be solved by these new data collecting methods for example if a registered participant fails to fill in the questionnaire from their places, we have a system called the Computer assisted telephone interview whereby the organisation will call the participant and fill the questionnaire through the phone. There is a greater possibility of obtaining accurate census data in census 2021.*

In summary, Statistics South Africa took an important step and strategy of shifting to modern methods that use artificial intelligence. However, the organisation might need to rely more on artificial intelligence whereby they need to move away from using enumerators to conduct door to door census.

Theme 5: The Future

Question 5: How do you see the role of AI in capturing census data? All the participants agreed that artificial intelligence is the future of census taking. They all mentioned that traditional census methods are no longer relevant in this day of age because the population figures are growing rapidly in South Africa. They also agreed that artificial intelligence is the solution to current challenges faced by Statistics South Africa. Below is an illustration of what they said:

Participant B: *It is with a doubt that artificial intelligence is the future of census taking. If you compare the population figures of census 1996 and census 2011, you can clearly see that the South African population is growing rapidly. Since the population is now bigger, the traditional methods will not be able to produce an accurate census count.*

Participant A: *The Fourth Industrial Revolution came at a right time for the South African nation. Our plans for the future as an organisation have changed because of this new phase that we have entered. As an organisation, we are planning to introduce advanced AI systems like the virtual census of Netherlands whereby everything electronically operated. Artificial*

intelligence is the future in capturing census data not only in South Africa, but also for the rest of the African continent.

Participant C: Artificial intelligence will solve many challenges that we have been facing from the previous years. It will make the census process much faster and precise. It is the new era of census taking.

In summary, it is clear that all the participants acknowledge that artificial intelligence is the way and the future of capturing census data. They all agreed that traditional methods no longer apply in this modern day. The population of South Africa is increasing and it is now impossible for manual methods to successfully carry out a national census.

Link Findings to Prior Literature

Theme 1 from the research findings talks on how artificial intelligence is effective and efficient in capturing census data compared to traditional methods. The interviewed participants acknowledged that artificial intelligence has the capability to produce accurate census data whereby machines are accurate, faster, cost and time saving. However, linking to the prior literature, Bernard (2012) explained that forms of artificial intelligence such as the Virtual Census System has enabled Netherlands to conduct census in an effective and efficient manner whereby the system produces data that is accurate and reliable while operating at a lower cost since most of the labour is replaced by advanced technology. The link between theme 1 and the literature is that both findings agree that artificial intelligence is effective and efficient compared to traditional census methods.

Theme 2 from the research findings discusses the lack of cooperation. The interviewed participants explained that the public fails to cooperate with Statistics South Africa during the census period. This challenge makes it impossible for Statistics South Africa to capture accurate and reliable data. Yvonne (2011) explained that in Limpopo, rural people do not like to disclose personal information because of certain cultural beliefs. The link between the findings and the literature is that both findings agree that the public does not cooperate with Statistics South Africa when it is capturing census data.

Theme 3 discussed on misinformation as the main reason to some of the challenges faced by Statistics South Africa. The interviewed participants explained that there are people in South Africa who are misinformed about the reason why Statistics South Africa capture census data. The participants said that people in South Africa believe that the police and

SARS retrieve the collected information by Statistics South Africa to catch criminals and therefore the public avoids participating in the census process. Yvonne (2011) confirmed this by stating that some people in South Africa do not feel safe to participate in census due to certain cultural beliefs, which misinform individuals thus preventing them to participate in the census process. In summary, the link between the two findings is that they both agree that there are a few individuals who are misinformed about census.

Theme 4 discussed about the new data collection methods whereby Statistics South Africa introduced new methods of collecting data, which use artificial intelligence. The new data collection methods are a measure to tackle some of the challenges they previously faced in the last 2011 census. However, the link with the literature is that Mark (2014) and Yvonne (2011) both agreed that in order for Statistics South Africa to achieve an accurate census, it should shift from a traditional method to modern. The link between the research findings and the literature was that both findings acknowledge that artificial intelligence is the solution needed to tackle challenges faced by Statistics South Africa when capturing census data.

Theme 5 talked about the future whereby the interviewed participants agreed that using artificial intelligence in census is the future and the way to obtain accurate census data. The literature agreed with the interview findings whereby Catherine (2018) and Bernard (2012) explained that in order for census organisations to capture accurate census data, they should adopt new census methods such as the virtual census and the 1940 linkable system. In summary, the link between the findings and the literature is that both findings agree that artificial intelligence is the future and the way for census capturing.

Link Findings to Prior Theory

Theme 4 from the interview findings discussed on how Statistics South Africa implemented three data collection methods that were on trials to see whether if they were capable of producing accurate census data and also to identify possible errors that might occur in the upcoming 2021 census. However, in connection with the prior theory, this research used theory of counting which is defined as a counting principle that seeks to find the number of outcomes in a probability problem (Kimberlee, 2015). The theory emphasized the importance of identifying possible uncertainties. In summary, the link between the two findings is that they stress on the importance of identifying possible errors before capturing census data.

Participant C from theme 1 mentioned that Statistics South Africa introduced a system that is capable of dividing the areas in which enumerators can work in. This system has enabled Statistics South Africa to allocate and track enumerators during their 2021 census trials. However, in connection to the prior theory, this research used a second theory called the Theory of Measure which is defined as a statistical theory that generalizes on aspects such as length, area and volume (Eric, 2020). This theory emphasizes on the importance of measuring areas before conducting any form of field work. In summary the link between the interview findings and the theory is that both findings highlight the importance of measuring areas which enumerators will work in before conducting census.

Ensuring quality of this research

Trustworthiness

From a quantitative perspective, trustworthiness incorporates validity and reliability. However, this research is a qualitative study and it uses the term trustworthiness. The core purpose of trustworthiness is to support the argument that the research findings are worthy of attention (Lincoln & Guba, 1985). In order to ensure quality and trustworthiness of this research, the following factors were used namely, credibility, transferability, confirmability and dependability.

Credibility

James (2016) defined credibility as the extent to which a qualitative researcher is confident in his/her research findings. In connection with this research, the data collection method used enabled the findings of this research to be truthful and accurate. In depth semi structured interviews was the data collection method used for this research via a social network called Microsoft Teams. Online Interviews through Microsoft Teams gave the researcher an opportunity to obtain detailed information about views, personal feelings and facts from the participants. The interviews were between 45 minutes to an hour long. Another measure used to ensure credibility was frequent briefings between the researcher and the supervisor that occurred on a daily and weekly basis through online pods and WhatsApp messages.

Transferability

For a researcher to establish transferability he has to provide evidence to readers that his research findings produce similar results if applied to other contexts (James, 2016). Lincoln and Guba (1985) explained that for a researcher to provide evidence that his research is

applicable, he has to provide a thick description of the phenomenon. This research used thick description as a tool to ensure that this study is transferable. The researcher ensured that the participant's voices, feelings, actions and meanings during the online interviews were accurately represented in the findings. Microsoft teams enabled the researcher to pay attention to contextual detail.

Confirmability

James (2020) defined confirmability as the extent to which the data collected agrees with the research findings and interpretations. There are various ways to determine confirmability. This research used an audit trail to provide evidence of the sequence of activities that took place throughout the course of completing this research.

Dependability

Dependability refers to the consistency and reliability of research findings (James, 2016). This research used inquiry audit to establish dependability. The Research supervisor examined the research and data analysis process in-order to ensure that the research findings were consistent and reliable.

Summary

Online interviews via Microsoft teams made it possible for this research to generate relevant findings. The data collection method used for this research enabled the researcher to obtain detailed information about perceptions and personal feelings of the participants. The thematic analysis method used to analyse the data enabled the researcher to produce themes that were relevant. The next chapter will discuss the conclusion of this research.

Chapter five

Conclusion

Indication of how the research questions were answered.

Research Questions:

1. Can artificial intelligence help Statistics South Africa to capture accurate census data?

Literature Findings

The findings from the literature review proved that artificial intelligence is capable of producing accurate census data that is more reliable and dependable. Catherine (2018) and Bernard (2012) researched on how America, Netherlands and United Kingdom have managed to capture accurate census data from using artificial intelligence.

Interview Findings

The findings from the interviewed participants showed that Statistics South Africa has implemented three new data collection methods that operate electronically. The new methods were on trial to test if they were dependable. However, the three data collection methods passed the trial since they produced accurate census data. In summary, according to the information gathered from the interviews and literature, artificial intelligence can help Statistics South Africa to capture census data. In summary, both findings proved that artificial intelligence could help Statistics South Africa to capture accurate census data.

2. Can artificial intelligence lower the costs of conducting census since most of the work will be executed using technology?

Literature Findings

According to the findings from the literature review, AI does reduce the costs of conducting census in that countries that have introduced AI into their census process have saved a lot of money due to the reduction of employment since many jobs were replaced by AI. Bernard (2012) stipulated that Netherlands has saved millions of revenue since the launch of the Virtual Census since most of the work was replaced by the system.

Interview Findings

Interviewed participants from Statistics South Africa, all pointed out that it was too early to conclude that AI can reduce the census costs since they recently shifted to using AI in

capturing census data. In summary, the literature findings proved that artificial intelligence could lower the costs of conducting census.

3. Can the introduction of artificial intelligence help Statistics South Africa to solve problems that occur during capturing census data?

Literature Findings

One of the major problem faced by Statistics South Africa when capturing census data is lack of cooperation from the public. Bernard (2012) pointed out that the Netherlands uses a system that consolidates data from different registers across the country. Such a system ensures that it has data for every household and individual in Netherlands. In summary, yes artificial intelligence can solve the major challenge that Statistics South Africa faces during census.

Interview Findings

According to the findings from the interviews, Statistics South Africa has recently introduced three new data collection methods that where on a trial in preparation for the upcoming 2021 national census. These three methods have successfully proved that artificial intelligence can solve their challenges whereby if a registered participant missed a chance to fill in a questionnaire, he/she would complete the questionnaire via online or telephonically. The use of these three new data collection methods in the trials has enabled Statistics South Africa reach almost every registered participant so far thus solving undercounts. In summary, the introduction of artificial intelligence can help Statistics South Africa to solve problems that occur during capturing census data.

Indication of how the research problem was solved

Research Problem: the research problem was that the traditional methods that Statistics South Africa has been using over the past two decades have resulted in undercounts. Challenges such as poor service delivery and lack of grant payments have resulted from inaccurate population figures from Statistics South Africa. The department needed to shift from using traditional census methods previously used in census 2011 in order to produce accurate census data. However, the research problem was solved by the findings from the literature. Bernard (2012) explained how artificial intelligence has enabled countries such as Netherlands, United Kingdom and Malaysia to produce accurate and reliable census data that has made their governments to plan since shifting from traditional methods.

Indication of how the research goal was reached

Research Goal: the main goal of this research was to determine if artificial intelligence would enable Statistics South Africa to capture and produce accurate census data. The findings of this study made it possible to achieve this research goal. According to the findings from the interviews, Statistics South Africa recently introduced three new data collection methods that rely on AI. These three methods were on trial for the upcoming 2021 national census. However, the new data collection methods have enabled Statistics South Africa to capture accurate census data. Moreover, according to the literature findings, Catherine (2018) and Bernard (2012) proved that AI has enabled countries such as Netherlands and America to capture accurate census data by using artificial intelligence such as the Virtual Census and the 1940 linkable data system.

Implications of findings for future practices.

There are few scholarly articles about census and artificial intelligence. This research can assist future scholars who need to conduct a study that is in line with this research topic. This research can give a better and understandable insight on the importance of census. Moreover, this research can educate readers on the impact of the Fourth Industrial Revolution and how it has managed to change the way countries conduct census. This research can enlighten readers on how inaccurate census data can affect the government in the planning process.

Heuristic value of the study

This research can provide knowledge to readers on why census is important for a country. There are ongoing strikes in South Africa for poor service delivery and this research provides reasons why the government has been battling to provide service delivery to all South African citizens. Moreover, this research provides solutions to most of the major problems affecting Statistics South Africa in capturing census data for example instead of relying on enumerators to conduct door-to-door census, they can use systems like the virtual census that automatically counts the number of people in the entire country using different electronic registers.

Ethical Consideration

Ethics are moral principles that guide a person to behave in a particular manner (Corne & Rose-Marie, 2013). Research ethics is important because it supports the values required for collective work for example fairness and mutual respect. However, the participants of this research had a clear understanding of the type of questions they were going to respond to

before the interviews because they were given questionnaires as shown in annexure B. This allowed participants to feel secure and comfortable for they knew the questions meant no harm to their personal life. Consent forms were given to the participants for them to sign in order to ensure that their personal information was to remain private after the interview as shown in annexure A. Lastly, this research did not deceive the participants by lying to them that the research was about other reasons none other than the potential role of AI in capturing census data.

Limitations

Due to the Corona virus outbreak, a limited number of participants were available to participate in the interviews. The cost of conducting online interviews limited the time to engage with different participants from Stats SA in-order to get in-depth answers. Moreover, it was difficult to get hold of hard copy journals and books to expand my research field because there was a national lockdown. I relied mostly on online journals and other academic sources and this limited my field of research.

Another limitation was that it was hard to get hold of participants from Statistics South Africa due to the corona virus. In addition, it was difficult to convince the participants to use online platforms such as Microsoft Teams, zoom and skype because it meant that they had to purchase mobile data to connect with me in which I became a liability. Lastly, the field of this research is relatively new, so there were few sources that focused directly on the potential role of artificial intelligence in capturing census data.

Conclusion

“Once a new technology rolls over you, if you are not part of the steamroller, you are part of the road “ (Stewart, 2017, p. 3). The South African population is growing rapidly and traditional census methods are no longer capable of producing accurate census data. The demand for better service delivery and a better living in South Africa is growing rapidly. Statistics South Africa needs to rely on artificial intelligence in order for them to produce accurate census data that can enable the government to plan.

References

- Ben, F., 2018. *Surveygizmo, Purposive Sampling 101*. [Online] Available at: <https://www.google.co.za/amp/s/www.surveygizmo.com/resources/blog/purposive-sampling-101/amp/> [Accessed 7 May 2020].
- Bernard, B., 2012. *Modern Census*, Sydney: Australian National University.
- Bezuidenhout, 2014. *Research Matters*. 2 ed. Cape Town: Juta.
- Bhat, A., 2018. *What is data collection*, United Kingdom: QuestionPro,
- Blair, H., 2020. *The Gum Tree Calculator for Python*, San Francisco: GitHub.
- Braun, V. & Clarke, V., 2006. *Six steps of thematic analysis*. [Online] Available at: <https://jvrafricagroup.co.za/blog/six-simple-steps-to-conduct-a-thematic-analysis> [Accessed 18 May 2020].
- Campuslabs, 2018. How to analyse qualitative data. *campuslabs*, pp. 1-3.
- Catherine, M., 2018. *Linking 1940 census with the modern*, New York: Taylor and Francis.
- Corne, D. & Rose-Marie, B., 2013. *Research Matters*. 1st ed. Cape Town: JUTA.
- Eric, W., 2020. *Measure theory*, New York: MathWorld,
- Helen, B., 2009. *Health Knowledge, Non probability Sampling*. [Online] Available at: <https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/methods-of-sampling-population> [Accessed 6 March 2020].
- James, L., 2016. *What is credibility in qualitative research*. [Online] Available at: <https://www.statisticssolutions.com/what-is-credibility-in-qualitative-research-and-how-do-we-establish-it/> [Accessed 12 July 2020].
- Jurgens, J., 2019. *Fourth Industrial Revolution*, Davos: World Economic Forum,

Katie, M. & Deborah, B., 2014. *A guide to ontology, epistemology and philosophical perspectives for interdisciplinary researchers*. [Online]
Available at: <https://www.google.co.za/amp/s/i2insights.org/2017/05/02/philosophy-for-interdisciplinarity/amp/>
[Accessed 2 May 2017].

Kimberlee, D., 2015. *Fundamental Counting Principle; Defination and Examples*, Chicago: Study.com,

Klaus, S., 2018. *The Fourth Industrial Revolution: what it means, how to respond*, World Economic Forum. [Online]
Available at: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
[Accessed 5 July 2020].

Lani, J., 2020. What is confirmability in qualitative reseach. *Statistics Solutions*, Volume 1, pp. 1-4.

Lauren, L., 2020. Advantages of Video Interviews You Do Not Know About. *Spark Hire*, pp. 1-6.

Lehohla, P., 2011. *Statistics South Africa*. [Online]
Available at: www.stassa.gov.za/?page_id=4021
[Accessed 15 july 2020].

Lincoln, Y. & Guba, E., 1985. Trustworthiness. *Sagepub Journals*, pp. 1-8.

Lorelli, S. J. M., 2017. *Thematic Analysis: Striving to meet the Criteria of Trustworthiness*. [Online]
Available at: <https://journals.sagepub.com/doi/full/10.1177/1609406917733847>
[Accessed 23 April 2020].

MacInnes, C., 1959. *Absolute Beginners*, London: Hart-Davis.

Mariette, D., 2019. *10 Emering Technologies 2019*, Davos: World Economic Forum.

Mark, O., 2014. *Calculating the undercount*, Pretoria: Statistics South Africa,

Massey, C., 2018. *Linking 1940 census with the modern*, New York: Taylor and Francis.

McGinnis, D., 2018. *What is the forth industrial revolution*. [Online] Available at: <https://www.salesforce.com/blog/2018/12/what-is-the-fourth-industrial-revolution-4IR.html>

[Accessed 9 April 2020].

Merriam, C., 2016. *Ontology and Epistemology*, Massachusetts: Merriam Webster Inc.

Merriam, G., 2011. *The Merriam Webster Dictionary*. Massachusetts: Merriam Webster Inc.

Moon, K., 2016. Qualitative Research elements. *A guideline to improve qualitative social science publishing in ecology and conservation journals*, 11(22), pp. 20-34.

Moran, M., 2020. Transferability. *Trustworthiness in Qualitative Research*, Volume 1, pp. 1-2.

Paul, J., 2008. *Population Parameter*, New York: Sage,

Plooy-Cilliers, D., 2013. *Research Matters*. 1st ed. Cape Town: JUTA.

Rebecca, H., 2020. *How the 5G coronavirus conspiracy theory went from fringe to main stream*. [Online]

Available at: <https://www.vox.com/recode/2020/4/24/21231085/coronavirus-5g-conspiracy-theory-covid-facebook-youtube>

[Accessed 15 May 2020].

Salkind, N., 2010. Triangulation. *Sage Research Methods*.

Stewart, B., 2017. *Technology quates to inspire and Motivate*, Nevada: Criteriaforsuccess.

Webster, M., 2015. *Paradigm*, Cape Town: Juta.

Whitting, B., 2010. *Data Analysis*. 1 ed. Cape Town: Juta.

Wilkinson, D., 2000. *The Researcher's Toolkit*. 1st ed. London: Routledge.

William, O., 2019. *The Importance of census accuracy*, Denver: Springer,

Yvonne, B., 2011. Population Count South Africa. *Challenges facing Statistics South Africa in conducting population census*, Volume 1, pp. 1-168.

ANNEXURE A: Consent Forms



CONSENT FORM TEMPLATE

Explanatory information sheet and consent form for participants
To whom it may concern, My name is Warren Mushonga and I am a student at Varsity College. I am currently conducting research under the supervision of Dr Hennie Pelser about the Potential role of artificial intelligence in capturing census data. I hope that this research will enhance our understanding of the importance of artificial intelligence in capturing census data I would like to invite you to participate in my study. In order to explain to you what your participation in my study will involve, I have formulated questions that I will try to fully answer so that you can make an informed decision about whether or not to participate. If you have any additional questions that you feel are not addressed or explained in this information sheet, please do not hesitate to ask me for more information. Once you have read and understood all the information contained in this sheet and are willing to participate, please complete and sign the consent form below.
What will I be doing if I participate in your study?
I would like to invite you to participate in this research because (insert reason). If you decide to participate in this research, I would like to (explain exactly what participation will involve). You can decide whether or not to participate in this research. If you decide to participate, you can choose to withdraw at any time or to decide not to answer particular interview questions.
Are there any risks/ or discomforts involved in participating in this study?
Whether or not you decide to participate in this research, there will be no negative impact on you. There are no direct risks or benefits to you if you participate in this study. You might, however, indirectly find that it is helpful to talk about your (insert what you are examining). If you find at any stage that you are not comfortable with the line of questioning, you may withdraw or refrain from participating.
Do I have to participate in the study?
• Your inclusion in this study is completely voluntary; • If you do not wish to participate in this study, you have every right not to do so;

- Even if you agree to participate in this study, you may withdraw at any time without having to provide an explanation for your decision.

Will my identity be protected?

I promise to protect your identity. I will not use your name in any research summaries to come out of this research and I will also make sure that any other details are disguised so that nobody will be able to identify you. I would like to ask your permission to record the interviews, but only my supervisor, I and possibly a professional transcriber (who will sign a confidentiality agreement) will have access to these recordings. Nobody else, including anybody at VARSITY COLLEGE, will have access to your interview information. I would like to use quotes when I discuss the findings of the research, but I will not use any recognisable information in these quotes that can be linked to you.

What will happen to the information that participants provide?

Once I have finished all interviews, I will write summaries to be included in my research report, which is a requirement to complete my BCOM HONOURS IN MANANEGEMENT. You may ask me to send you a summary of the research if you are interested in the final outcome of the study.

What happens if I have more questions about the study?

Please feel free to contact me or my supervisor should you have any questions or concerns about this research, or if there is anything you need to know before you decide whether or not to participate.

You should not agree to participate unless you are completely comfortable with the procedures followed.

My contact details are as follows:
Warren Mushonga

The contact details of my supervisor are as follows:
Dr Hennie Pelzer

Consent form for participants

I, _____, agree to participate in the research conducted by Warren Mushonga about the potential role of artificial intelligence in capturing census data.

This research has been explained to me and I understand what participation in this research will involve. I understand that:

• I agree to be interviewed for this research.	
• My confidentiality will be ensured. My name and personal details will be kept private.	
• My participation in this research is voluntary and I have the right to withdraw from the research at any time. There will be no repercussions should I choose to withdraw from the research.	
• I may choose not to answer any of the questions that are asked during the research interview.	
• I may be quoted directly when the research is published, but my identity will be protected.	
Signature	Date

**THE INDEPENDENT
INSTITUTE OF
EDUCATION**

CONSENT FORM FOR

AUDIO OR VIDEO RECORDING

Consent form for participants	
I, _____, agree to allow Warren Mushonga to audio record my interviews as part of the research about the potential role of artificial intelligence in capturing census data. This research has been explained to me and I understand what participation in this research will involve. I understand that:	
• My confidentiality will be ensured. My name and personal details will be kept private.	
• The recordings will be stored in a password-protected file on the researcher's computer.	
• Only the researcher, the researcher's supervisor and possibly a transcriber (who will sign a confidentiality agreement) will have access to these recordings.	
Signature	Date

ANNEXURE B: Research Questionnaire

Research questionnaire

Postgraduate honours degree in Management

Varsity College - Pietermaritzburg

Student Name: Warren Mushonga

Supervisor: Dr Hennie Pelser

INTRODUCTION OF RESEARCH STUDY

The objective of Statistics South Africa is to render an accurate census count. Lack of accurate census disables the government to meet the actual demand of people who need service delivery. My research is exploring the potential role of artificial intelligence in capturing census data. The main goal of this research is to identify if artificial intelligence can assist Statistics South Africa in capturing census data.

About the researcher

I am a part time student at Varsity College in Pietermaritzburg. This research is not politically motivated instead it is pursuing a fulfilment of an honours degree in Management. Information provided will remain confidential after this interview.

Questions

1. What are your thoughts about the potential role of artificial intelligence and machine learning in addressing some of the challenges faced by Statistics South Africa?
2. What are the challenges that you face when capturing census data?
3. What are the main reasons that cause some of the challenges you encounter when capturing census data?
4. What measures have you put in place to tackle some of these challenges?
5. How do you see the role of AI in capturing census data?

ANNEXURE C: Permission letter

ANNEXURE D: Ethics Clearance



17 July 2020

Student name: Warren Mushonga

Student number: 14020012

Campus: Varsity College: Pietermaritzburg

Re: Approval of BCom Honours in Management Proposal and Ethics Clearance

HONOURS ETHICAL CLEARANCE LETTER

Your research proposal and the ethical implications of your proposed research topic were reviewed by your supervisor and the campus research panel, a subcommittee of The Independent Institute of Education's Research and Postgraduate Studies Committee.

Your research proposal posed no significant ethical concerns and your supporting documents and instruments are in order to proceed. We hereby provide you with permission to proceed with your research.

In the event of you deciding to change your research methodology in any way, kindly consult your supervisor to ensure all ethical considerations are adhered to and pose no risk to any participant or party involved. A revised ethical clearance letter will be issued.

We wish you all the best with your research!

GENERAL CONDITIONS TO BE FULFILLED IN RELATION TO RESEARCH

Permission is granted to proceed with the above study subject to the conditions listed below being met and may be withdrawn should any of these conditions be flouted.

Please note: The panel has not considered the merits, accuracy or ethical soundness of the research. The only merits examined are the use of The IIE as a sample.

Permission is granted subject to the following conditions:

1. The researcher will need to obtain informed consent in writing from all the participants in their sample if the study is not anonymous.
2. The researcher may only use the data collected for research purposes and in no other way.
3. Photographs of human subjects may only be taken if relevant to the research, informed consent was obtained, and even with informed consent, the photographs may not be published on any online platforms.
4. The researcher is responsible for supplying and utilising their own research resources, such as stationery, photocopies, transport, faxes and telephones and should not depend on the goodwill of the institutions and/or the offices visited for supplying such resources.
5. No names or identifying information of participants may be used within the research and the research must be voluntary.
6. Please make it clear that the information will not be used punitively in any way and participants may in no way be counselled/advised based on this.

Supervisor: Dr HJ Pelser

Campus Postgraduate Coordinator (CPC): Dr HJ Pelser

ANNEXURE E: Originality Report

Blackboard

SafeAssign Originality Report

RESM8419p_VCPM1 • RESM8419p Research Report [View Originality Report - Old Design](#)

Warren Mushonga

Submission UUID: e878ffb6-1013-291f-c2fa-cd7058576876

Total Score:  Low risk 10 %

Total Numb...	Highest Ma...	Average Ma...	Submitted ...	Average Wo...
1	10 % RESEARCH REPO...	10 %	20/08... 18:14 GMT+2	10,349 Highest: RESEAR...

ANNEXURE F: Research Report Document Table

Research Purpose/ Objective	Primary Research Question	Research Rationale	Seminal Authors/ Sources	Literature Review – Conceptual Framework	Paradigm	Approach	Data Collection Method(s)	Ethics	Anticipated Findings	References
<i>To identify whether artificial intelligence can help Statistics South Africa in capturing accurate census data.</i>	<i>Can artificial intelligence help Statistics South Africa to capture accurate census data?</i>	This research is relevant because it is identifying if artificial intelligence can assist Statistics South Africa in capturing census data, which is accurate.	Yvonne Billa Mark Orkin William O'Hare Bernard Baffour Catherine Massey	Theme 1: <i>Theory of Counting</i> Theme 2: <i>Theory of Measure</i> Theme 3: <i>Forms of artificial intelligence in census</i> Theme 4: <i>Factors leading to undercounts</i> Theme 5: <i>Advantages of using artificial intelligence</i> Theme 6: <i>The impact of inaccurate census data</i>	<i>Paradigm: Interpretivism</i> <i>Epistemology: To understand the potential impact of AI in census</i> <i>Ontology: Use of theories and existing sources to get a better understanding of the potential role of AI in capturing census</i> <i>Axiology: Axiology focuses on what a researcher values in a research study. The core value of this research is to explore if AI can replace traditional methods in capturing census.</i>	<i>Qualitative</i> <i>Population</i> <i>Target population - Stats SA KZN</i> <i>Accessible population Stats SA PMB</i>	<i>Online interviews via Microsoft Teams</i> <i>Open ended and in-depth semi structured questions</i>	<i>This research had an informed consent to all the questionnaire participants. Participants signed a written declaration stating the disclosure of their personal information.</i>	<i>This research provides solutions to most of the major problems affecting Statistics South Africa during census for example instead of relying on enumerators to conduct door-to-door census, they can use systems like the virtual census that automatically counts the number of people in the entire country using different electronic registers..</i>	Whitting, B., 2010. <i>Data Analysis</i> . 1 ed. Cape Town: Juta. Orkin, D. F., 2012. <i>Calculating the undercount</i> . DataFirst, pp. 1-20. O'Hare, W., 2018. <i>The Importance of census accuracy</i> . SpringerLink, pp. 1-20. Bezuidenhout, 2014. <i>Research Matters</i> . 2 ed. Cape Town: Juta Massey, C., 2018. <i>Linking 1940 census with the modern</i> , New York: s.n.
Research Problem	Secondary Questions/ Hypotheses/ Objectives	Key Concepts	Key Theories			Sampling	Data Analysis Method(s)	Limitations	Anticipated Contribution	
<i>. Over the last two decades, Statistics South Africa has relied on traditional methods to capture census data. These traditional methods have led the department to record under-counts.</i>	<i>Can artificial intelligence lower the costs of conducting census since most of the work will be executed using machines?</i>	1. <i>Artificial intelligence</i> 2. <i>Census</i> 3. <i>Undercounts</i> 4. <i>Accuracy</i>	<i>Theory of Measure</i> <i>Theory of Counting</i>			<i>Non-probability sampling</i> <i>Purposive sampling</i> <i>3 participants</i>	<i>This research used labels, themes and categories to code the data accumulated from interviews in-order to present the empirical findings.</i> <i>-Thematic Analysis</i>	<i>. The field of this research was small whereby they were few sources that focused directly on the potential role of artificial intelligence in census recording. Due to the Corona virus outbreak, a few participants were available to participate in the interviews.</i>	<i>This research can assist Statistics South Africa to adopt a strategy used by the United Kingdom that gathers data for all rural and urban settlement.</i>	