

Automated Census Population Projection and Data Management System

Izakpa Getty Ebere¹, Ofualagba Mamuyovwi Helen², Ekhaton Uyiosa Emmanuel³

¹Department of Computer Technology, Petroleum Training Institute, Effurun, Delta State, Nigeria

²Department of Computer Science, Delta State Polytechnic, Oghara-Otefe, Delta State, Nigeria

³Department of Computer Science, National Open University of Nigeria

izakpa_eg@pti.edu.ng¹, komekaro83@gmail.com², fantfab3@gmail.com³

ABSTRACT

Article Info

Volume 7, Issue 1

Page Number: 244-254

Publication Issue :

January-February-2021

Article History

Accepted : 10 Jan 2021

Published : 30 Jan 2021

Population census data has become more pervasive to lives which has demanded for increase in scope, completeness, accuracy and validity, and improve on their national value and international comparability. The accuracy of population projections has been attracting more attention, driven by concerns about the possible long-term effects of aging and other demographic trends. This paper therefore, attempt to improve on the manual way of managing census data by developing an automated census data management system. This system in addition, is capable of projecting population growth using exponential growth equation. The National Population Commission, Benin City was used as a case study.

Keywords : Population Census, Census Estimation and Projection, National Population Commission

I. INTRODUCTION

Population census is the total process of collecting compiling, analyzing, evaluating, publishing and disseminating demographic economic, social and housing data pertaining at a specified time to all persons and all building in a country or in a well delineated part (UN, 2008; Measure Evaluation, 2020).

It generates a several data, including numbers of people, spatial distribution, age and sex structure, as well as living conditions and other key socio-economic characteristics within a country (UNFPA, 2019). This information is critical and of great relevance to the economics, political and socio-cultural planning, good governance, policy formulation, development planning, risk reduction, crisis response, social welfare programmes and

monitoring of development goals of a country (Okeke and Ekechukwu, 2013; UNFPA, 2019).

Population census is among the most complex and massive peacetime exercises a nation can undertake. It requires careful planning, resourcing and implementation – from mapping an entire country, mobilizing and training large numbers of enumerators, and conducting major public awareness campaigns, to canvassing all households, carefully monitoring census activities, and analysing, disseminating and using the resulting data (UNFPA, 2019).

Fosu (2001) noted that information technology has expanded the possible use of population census data beyond the traditional model. This data has become more pervasive to lives which has demanded for

increase in scope, completeness, accuracy and validity, and improve on their national value and international comparability.

This paper therefore, attempt to improve on the manual way of managing census data by developing an automated census data management system. This system in addition is capable of projecting population growth using exponential growth equation. The National Population Commission, Benin City was used as a case study.

II. LITERATURE REVIEW

A. Census Data Collection and Processing

According to Adele (2009), Nigerian Government has over the years expended and exerted much in order to achieve a reliable census. The idea of population census is that, everybody is counted, but the fact remains that everybody cannot be counted; that is, a complete and accurate population census in any part of the world is impossible.

A population census, often combined with a housing census, is usually conducted by the national statistics office every 10 years. It consists of an individual enumeration of the entire population, with the objective of providing an accurate count of the population at a fixed point in time. It is the most logistically complicated and expensive data collection operation that a country can undertake (UN, 2014a; UNFPA, 2019).

The UN Statistics Division developed and maintains a compendium of standards on conducting a census (UN, 2014a). It includes recommendations for planning and operationalizing the population and household enumerations, and it presents important concepts and definitions that should be applied consistently in a country's integrated program for data collection and analysis. It also provides a detailed

set of tabulations that maximize information collected on core census topics.

Although the census enumeration is mandated by a national law and obliges every resident to respond, Principle 6 of the UN Fundamental Principles of Official Statistics resolves that personal data collected are to be kept strictly confidential and used for statistical purposes only (UN, 2014b).

B. Census Population Essential Features

The United Nations lists four essential features of a census (Measure Evaluation, 2020):

- i. Each individual is enumerated separately; the characteristics of each person within the household are recorded separately.
- ii. The census covers a precisely defined territory and includes every person present or residing within its scope. The housing census should include every type of building and living quarters.
- iii. Each person and each type of building and living quarters is enumerated with respect to a well-defined point of time.
- iv. The census is taken at regular defined intervals, usually every 10 years.

C. Census Population Projection

A projection is a calculation of the size of the population for a future date in time. Population information for past, present, and future conditions can be used to make a projection about the population (Measure Evaluation, 2020).

The accuracy of population projections has been attracting more attention, driven by concerns about the possible long-term effects of aging and other demographic trends. This enables government policymakers and planners around the world use population projections to gauge future demand for food, water, energy, and services, and to forecast future demographic characteristics. Population

projections can alert policymakers to major trends that may affect economic development and help policymakers craft policies that can be adapted for various projection scenarios (Population Reference Bureau, 2001).

D. Related Works

Okeke and Ekechukwu, (2013) noted that census analysis is often not critically analyzed to bring out some of the basic and important attributes of census data information to give geo-spatial distribution of population. This is due to non-availability of the required tools for carrying out such analysis. The authors suggested the use of data-mining technique (Decision tree algorithm technique) to extract hidden information from large census data warehouse and geographical information system (GIS) as an integrating technology that gives geo-spatial distribution of the population. It is believed that these are patterns buried within the mass of data in the various editions of population census figures in Nigeria. These are patterns that will be impossible for humans working with bare eyes and hands, to uncover without computer system to give geo-spatial distribution of population in that area. This pattern could help better-informed business decisions and provide government with the intelligence for strategic planning, tactical decision-making and better policy formulation.

Sondhi (2017) paper presented an application of data mining in the analysis of census and looks at the noticeable trends. The objective of the paper is to discover the relevant information of gender inequality in all spheres from the primary census of the Raigarh district in Maharashtra with the help of appropriate data mining methods like clustering, visualization using the Weka tool.

Oyebola and Adenike (2018) research analyzed Nigeria population census data in order to bring forth knowledge that can aid Government in social-

economic decision-making. Thus, k-means algorithm, which is an unsupervised learning technique, was implemented on MapReduce with the aim of discovering knowledge from Priority Table IX of Nigeria Census Data of 2005. MapReduce was used to aid k-means computational challenges such as Euclidean distance computation, minimum sum of square error (MSSE) computation and global objective computation effectively. The big data analytics revealed local government areas that need Government Intervention in terms of low cost housing and those local governments that need urban restructuring for good distribution of population.

According to Al-Eideh and Al-Omar (2019), one of the important functions of the demographers is to provide information on the future trend of population projection, which is important to plan for human activities. Today, demographers are interested in describing phenomena in theoretical models involving population structure by considering the stochastic analogs of classical differences and differential equations. The authors presented a description of a population projection model, derived using a growth rate following a birth and death diffusion process. The mean and the variance as well as the predicted and the simulated sample path of the population projection process are also obtained. Numerical example for the case of the birth and death diffusion growth rate as well as the case of the constant growth rate are considered. The objective of the research was to identify critical knowledge types required by demographers of population projection techniques through building a population projection model.

The U.S. Census Bureau developed a new software application, Demographic Analysis and Population Projections System (DAPPS). DAPPS applies the cohort component projections technique to produce single year population projections. Cohort component projection is a method widely used by

National Statistics Offices to produce their official national population projections. DAPPS has a modern graphical user interface that walks users step-by-step through the process of applying demographic analysis techniques and producing population projections. It allows users to enter a variety of inputs for population, fertility, mortality, and migration, and produces a range of outputs for users to evaluate. It serves as a user-friendly platform for inputting and analyzing the data, and producing and evaluating population projections (Mizoguchi *et al*, 2020).

III. MATERIALS AND METHODS

A. Existing System Analysis

The existing system is one in which the modus operandi is Manual; A system in which the method of storing Demographic data/information is such that records are written on Papers or Registers and kept in Files.

Critical analysis of the system reveals it is often prone to a lot of errors, hence resulting in its ineffectiveness. This always tends to make the office environment look unkempt. Also, files could be lost due to the inconsistency of the system.

A lot of problems are associated with the existing system. The system has proved defective in the area of: Data Redundancy, Time wasted in Searching/Sorting for Information, Poor Security and Protection, Misplacement and Mismanagement of Files, Inadequate Communication Link and Lack of Space to keep Files.

B. Facts Finding Methods

In order to identify the user requirement of the proposed system, this paper adopted the following fact finding methods:

Personal Interview: Some of the staff present at the commission, were interviewed to share their

thoughts and experiences about the manual system of storing demographic data/information. They stressed that the system has not helped them much.

Observation: A situation where files containing records were littered all over the place, made the office environment untidy. Also having observed that searching for record details was really hectic and time consuming, the ineffectiveness of the manual approach was thereby confirmed.

Net Surfing: Research from the internet also provided data/information concerning the draw backs in the use of the manual system.

C. Proposed System Analysis

The proposed system comprises of the following modules: Staff Registration, Department Creation, Census Data Creation, Report Generation and Population Projection.

Staff Registration is the component of the proposed system that would be used to manage records of every staff of the census department. Through this module, information about each staff and/or recently employed staff could be recorded, stored and retrieved when necessary.

Department Creation module manages records concerning the census department of the commission. Whenever a new department is created, it will be recorded and stored for retrieval at any moment. Modifications could also be done, in case there are alterations in the department.

The Census Data Creation module enables input of census data collected from the field officers into the system for storage and future analysis.

The Population Projection module help in projecting population growth over a period of time specified.

There are other sub modules which are not mentioned that are loosely coupled with the main modules. At the end of each activity, several reports can be generated from the system with the help of Report Generation module for decision and policy making.

D. Proposed System Design

I. Structure

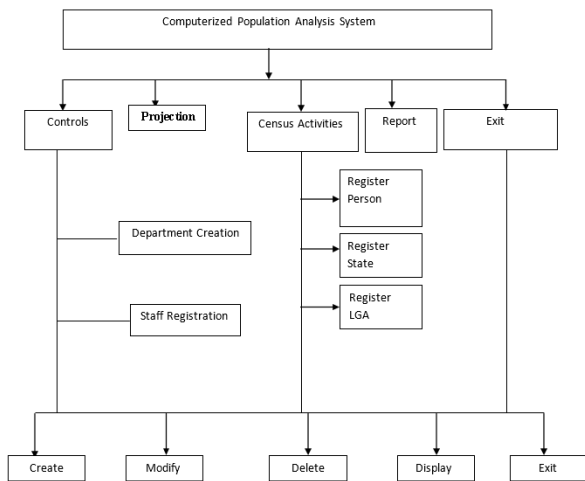


Fig. 1: Organizational Structure of the Proposed System

II. Data Flow Diagram (DFD)

Data Flow Diagram (DFD) is a graphical representation of the “flow” of data within an Information System. Below is the Data Flow Diagram for the Commission’s Proposed System.

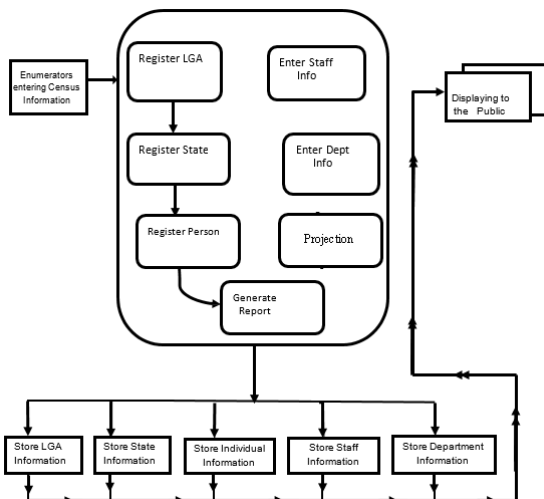


Fig. 2 : Data Flow Diagram (DFD) of the Proposed System

III. Entity Relationship Diagram (ERD)

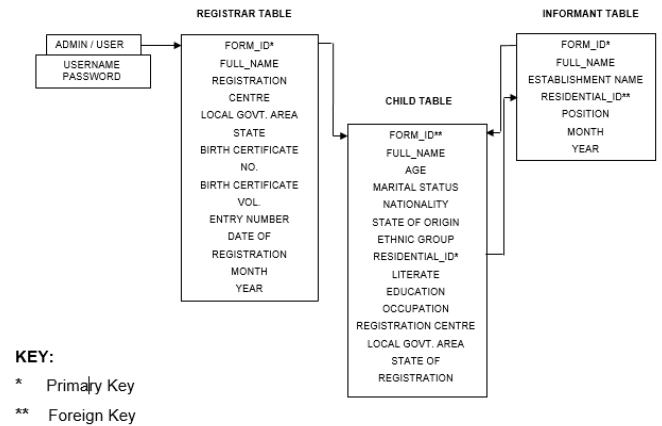


Fig. 3 : Entity Relationship Diagram (ERD) of the Proposed System

IV. Unified Modeling Language (UML)

UML is modelling language which provides a set of conventions that are used to describe the software system in terms of objects, showing different perspective views of the system parts. The Use Case and Sequence diagrams where the only UML diagrams used in this paper.

A. Use Case Diagram

The Use Case Diagram defines the Core Elements and Processes that make up a system. The Core Elements are termed “Actors”, while the Processes are termed “Use Cases”. It defines the system requirements being modelled and help write the scenario later used in testing.

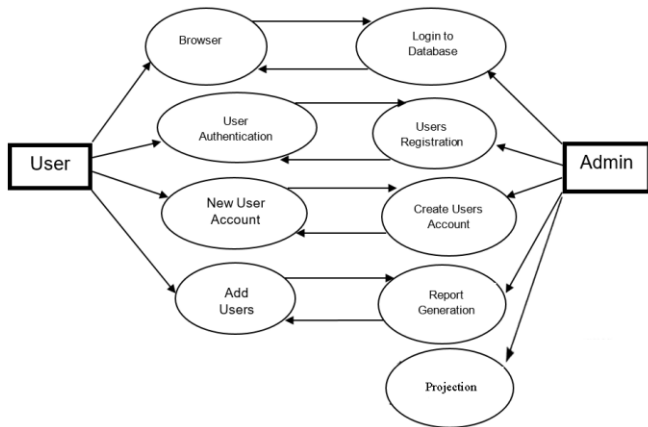


Fig. 4 : Use Case Diagram of the Proposed System

B. Sequence Diagram

A Sequence Diagram is an interaction diagram that shows how objects interact with one another and in what order. It is arranged in time sequence. They are often called Event Diagrams or Event Scenarios.

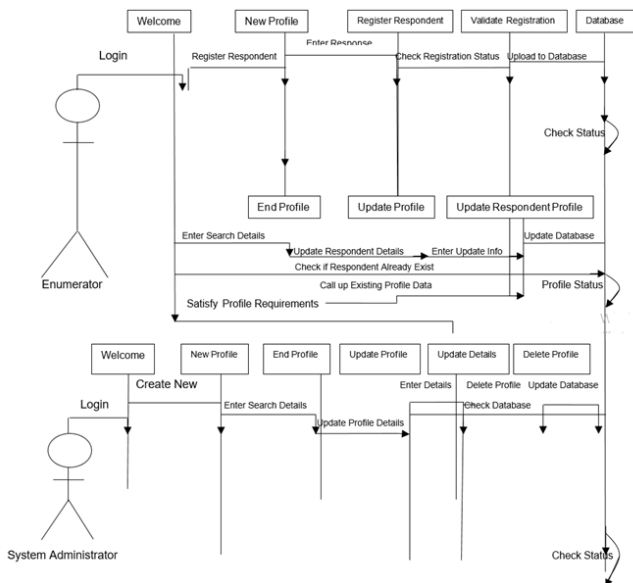


Fig. 5: A Sequence Diagram of the Proposed System

E. Research Data

Edo State Population Statistics given by the National Population Commission in 2006 were as follows:

Number of Males: 1633946

Number of Females: 1599420

Total Population: 3233366

The Total Population Statistics for Benin City given by National Population Commission, Benin in 2006 was 1147188. While the Statistics given in 2015 was 1495800.

Comparing both statistics, we can thus project the population in Benin City for subsequent years by using a mathematical model.

F. Mathematical Model

A mathematical model is a description of a system using mathematical concepts and language. A Model may help to explain a system, study the effects of different components, as well as to make predictions about its behaviour. Mathematical Models can take many forms, such as Dynamical Systems, Statistical Models, Differential Equations, Exponential Equations or Game Theoretic Models. This paper adopted Exponential Equation Model for its population projection.

G. Exponential Growth Model

To deduce the Exponential Growth Equation, we consider the differential equation below:

$$\frac{dp}{dt} = kp \dots\dots\dots (1)$$

$$\frac{dp}{p} = kdt \dots\dots\dots (2)$$

$$p(t) = e^{kt+c}, \text{ where } e^c = A \dots\dots\dots (3)$$

$$\text{At } t = 0, p(0) = Ae^{k(0)} \dots\dots\dots (4)$$

$A = p(0) \iff p_0$ is the Initial Population

Therefore substituting for A into (3), we have:

$$P(t) = P_0e^{kt} \dots\dots\dots (5)$$

Where,

$P(t)$ = Population at time (t) OR Population Growth.

P_0 = Population at time zero OR Number of Initial Population.

k = Exponential Growth Rate (Estimated Annually).

t = Elapsed time in years from time zero.

From the information earlier, we know that Benin City had an Initial Population (P_0) of 1147188 in 2006, while the Population $P(t)$ increased to 1495800 in

2015 within a period of 9 years (2015 – 2006). Hence, we can calculate the Exponential Growth Rate (k) within that period.

Given:

$$P(t) = 1495800$$

$$P_0 = 1147188$$

$$k = ?$$

$$t = 9$$

Substituting into the equation, we have:

$$1495800 = 1147188e^{9k}$$

By isolating the Exponential part, we have:

$$\frac{1495800}{1147188} = \frac{1147188e^{9k}}{1147188}$$

$$1.30 = e^{9k}$$

Since we have base (e), instead of taking the Common Log of both sides, we take the Natural Log of both sides. Hence, we have:

$$\ln 1.30 = \ln e^{9k}$$

Applying the Power Property of Logarithms, we have:

$$\ln 1.30 = 9k \cdot \ln e$$

Since $\ln e = 1$, we therefore have:

$$\ln 1.30 = 9k$$

To solve for k, we divide both sides by 9. Hence, we have:

$$\frac{\ln 1.30}{9} = k$$

$$K \approx 0.029152$$

The Exponential Growth Rate (k) is $\approx 0.029152 = (2.9152\%)$

IV. RESULT AND DISCUSSION

A. Benefits of the Proposed System

The proposed system when installed will be of benefit to the Commission in the following ways:

- The commission will have records of individuals stored in the database that enables direct access to data as reports on individuals would be generated, and stored for retrieval.

- The rate at which data could be accessed would be maximal.
- Creating, Updating and Modifying information details as opposed to the old Manual System will be facilitated.
- Collection of information which was always a tedious affair would now be an easy task.

B. Statistical Results

DATA 1

To project the population in 2017, we subtract the base year (2006) from 2017, which gives us 11 years.

Hence we have:

$$P(t) = ?$$

$$P_0 = 1147188$$

$$k = 0.029152$$

$$t = 11$$

Substituting into the equation, we have:

$$P(11) = 1147188e^{0.029152(11)}$$

$$P(11) = 1147188e^{0.320672}$$

$$P(11) = 1147188 \times 1.378053505$$

Rounding off to the nearest person, we have:

$$P(11) \approx 1580886. \text{ See Fig 7.}$$

To calculate the Annual Percentage Growth Rate for 2017, we use the formula -

$$\text{Annual \% GR (k)} = \frac{(P_t - P_0) \times 100\%}{\frac{P_0}{t}}$$

Where,

$$P_t = 1580886$$

$$P_0 = 1147188$$

$$t = 11, \text{ that is } (2017 - 2006)$$

Substituting into the equation, we have:

$$k = \frac{(1580886 - 1147188) \times 100\%}{\frac{1147188}{11}}$$

$$= \frac{433698 \times 100\%}{1147188}$$

$$11$$

$$k = \frac{0.378053 \times 100\%}{11} = \frac{37.8053}{11}$$

$$k = 3.436845\% = (0.03436845)$$

DATA 2

To project the population in 2020, we subtract the base year (2006) from 2020, which gives us 14years.

Hence we have:

$$P(t) = ?$$

$$P_0 = 1147188$$

$$k = 0.03436845$$

$$t = 14$$

Substituting into the equation, we have:

$$P(14) = 1147188e^{0.03436845(14)}$$

$$P(14) = 1147188e^{0.4811583}$$

$$P(14) = 1147188 \times 1.617947$$

Rounding off to the nearest person, we have:

$$P(14) \approx 1856089. \text{ See Fig 8.}$$

To calculate the Annual Percentage Growth Rate for 2020, we use the formula -

$$\text{Annual \% GR (k)} = \frac{(P_t - P_0) \times 100\%}{\frac{P_0}{t}}$$

Where,

$$P_t = 1856089$$

$$P_0 = 1147188$$

$$t = 14 (2020 - 2006)$$

Substituting into the equation, we have:

$$k = \frac{(1856089 - 1147188) \times 100\%}{\frac{1147188}{14}}$$

$$= \frac{708901 \times 100\%}{\frac{1147188}{14}}$$

$$k = \frac{0.617947 \times 100\%}{14} = \frac{61.7947}{14}$$

$$k = 4.413907\% = (0.04413907)$$

DATA 3:

To project the population in 2026, we subtract the base year (2006) from 2026, which gives us 20years.

Hence we have:

$$P(t) = ?$$

$$P_0 = 1147188$$

$$k = 0.04413907$$

$$t = 20$$

Substituting into the equation, we have:

$$P(20) = 1147188e^{0.04413907(20)}$$

$$P(20) = 1147188e^{0.882781}$$

$$P(20) = 1147188 \times 2.417615$$

Rounding off to the nearest person, we have:

$$P(20) \approx 2773459. \text{ See Fig 9.}$$

To calculate the Annual Percentage Growth Rate for 2026, we use the formula -

$$\text{Annual \% GR (k)} = \frac{(P_t - P_0) \times 100\%}{\frac{P_0}{t}}$$

Where,

$$P_t = 2773459$$

$$P_0 = 1147188$$

$$t = 20 (2026 - 2006)$$

Substituting into the equation, we have:

$$k = \frac{(2773459 - 1147188) \times 100\%}{\frac{1147188}{20}}$$

$$= \frac{\frac{1626271 \times 100\%}{1147188}}{20}$$

$$k = \frac{1.417615 \times 100\%}{20} = \frac{141.7615}{20}$$

$$k = 7.088075\% = (0.07088075)$$

DATA 4

To project the population in 2035, we subtract the base year (2006) from 2035, which gives us 29years.

Hence we have:

$$P(t) = ?$$

$$P_0 = 1147188$$

$$k = 0.07088075$$

$$t = 29$$

Substituting into the equation, we have:

$$P(29) = 1147188e^{0.07088075(29)}$$

$$P(29) = 1147188e^{2.05554175}$$

$$P(29) = 1147188 \times 7.811068$$

Rounding off to the nearest person, we have:

$$P(29) \approx 8960764. \text{ See Fig 10.}$$

C. System Interfaces

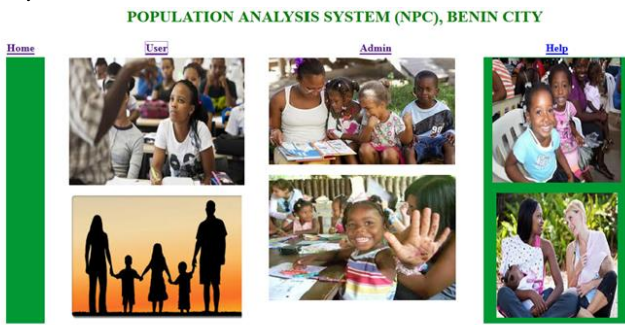


Fig 6. Homepage

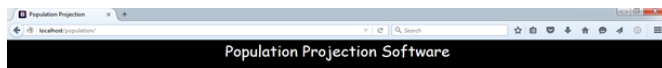


Fig 7. Screenshot of the Population Projection for B/C in 2017

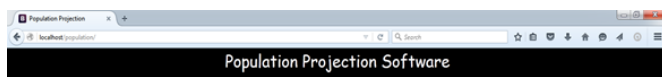


Fig 8: Screenshot of the Population Projection for B/C in 2020

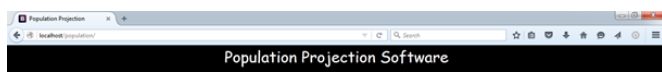


Fig 9. Screenshot of the Population Projection for B/C in 2026

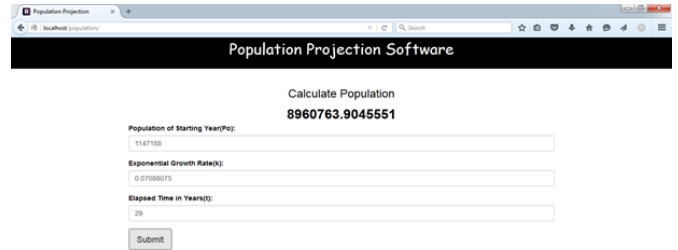


Fig 10. Screenshot of the Population Projection for B/C in 2035

Table 1 : Population Projection Statistics for Benin City

DATA	YEAR (t)	PROJECTED POPULATION [P(t)]
1	2017	1580886
2	2020	1856089
3	2026	2773459
4	2035	8960764

Graphical Representation of Population Projection Statistics for Benin City

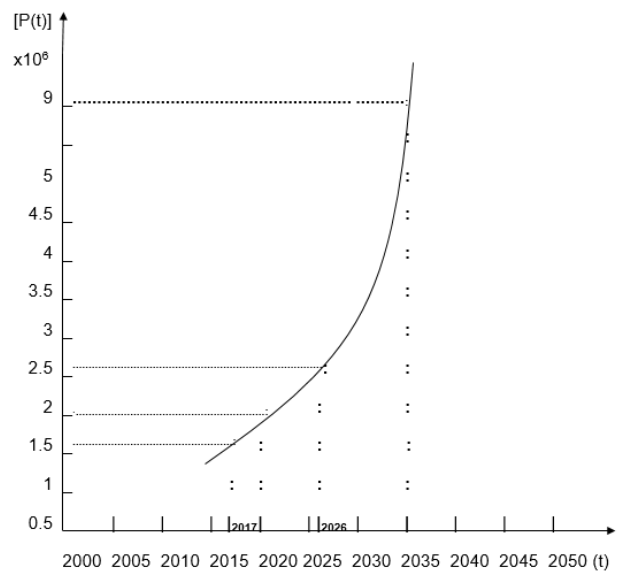


Fig. 11: Plot of Projected Population [P(t)] against Year (t)

D. Observation

From the graph, it is observed that the population of Benin City increased initially from 1147188 in 2006 to 1580886 in 2017 with an Exponential Growth Rate (k) of 2.9152% per annum. It later increased from

1580886 in 2017 to 1856089 in 2020 with an Exponential Growth Rate (k) of 3.4368% per annum. The statistics then increased from 1856089 in 2020 to 2773459 in 2026 with an Exponential Growth Rate (k) of 4.4139% per annum. Finally, there was a rapid growth in the population, as the statistics increased from 2773459 in 2026 to 8960764 in 2035 with the Exponential Growth Rate (k) rising much higher to about 7.0881% per annum. The Exponential Growth is depicted by the rapid rise in the Curve, as shown in the graph.

So comparing the projected statistics for 2017 (1580886) with that stated by the National Population Commission (1617579), it is observed that approximately equal statistics were generated in both cases, with a minimal difference of 36693.

Thus, highlighting its importance (most notably; Speed and Accuracy) as regards the generation of automated population statistics, the Software can be effectively recommended, operated and used by the National Population Commission, Benin City.

E. System Implementation

The proposed system was developed as a web based application using Hypertext Markup Language (HTML), Hypertext Preprocessor (PHP), JavaScript, and MySQL. The system can be deployed on Apache server running on any operating system.

V. CONCLUSION

The study of the manual system was done and the new system designed. The need for the computerization of the commission was highly emphasized, since computers have obvious advantages over human agents as regards the storage, update and retrieval of information, as well as the projection of population statistics. Computers could always process data and produce accurate and reliable

results when given correct data. The use of computers in census operations would solve problems encountered in the manual system. Hence, one could then conclude that the computerization of population census related activities is a welcomed development that must be undertaken as its advantages include, among others, the following:

- Increase in processing speed as well as the accuracy in statistical data / information.
- Improved storage facilities and easy retrieval – which plays an important role in National planning and policy making.
- Ease encountered during data movement via a computer network as opposed to the manual approach where vehicles are used for transporting data.

VI. RECOMMENDATION

The efficiency and effectiveness of using computers to handle population census related activities have already been identified during the course of the research. The proposed software is therefore recommended to the National Population Commission, Benin City and all other states requiring efficient, time-saving and security of data. Hence, in order to successfully implement the proposed system, the following recommendations have been made in the light of the above findings:

- The parallel implementation method should be adopted, as that would give room for the comparison of results before undergoing the total change in implementation.
- The staff of the commission should be thoroughly trained on the use/operation of the software, as it would enhance competence towards service delivery.

VII. REFERENCES

- [1] Okeke O.C and Ekechukwu B.C (2013). Using Data-Mining Technique for Census Analysis to Give Geo-Spatial Distribution of Nigeria. IOSR Journal of Computer Engineering (IOSR-JCE), 14(2), 1-5
- [2] Adele B.J (2009). Falsification of population census data in a heterogeneous Nigerian state: The fourth republic example. African Journal of Political Science and International Relations, 3(8), 311-319
- [3] United Nations. (2014a). Principles and recommendations for population and housing censuses (Revision 3). New York, NY: United Nations. Retrieved from http://unstats.un.org/unsd/demographic/meetings/egm/NewYork/2014/P&R_Revision3.pdf
- [4] United Nations. (2014b). Resolution adopted by the General Assembly on 29 January 2014. Fundamental principles of official statistics. Retrieved from <http://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf>
- [5] Measure Evaluation (2015) Health Information System Strengthening: Standards and Best Practices for Data Sources. University of North Carolina, USA.
- [6] Mizoguchi N., Fennell S. and Fischer O.P (2019). DAPPS: A New Tool for Demographic Analysis and Population Projections. Retrieved from <http://paa2019.populationassociation.org/abstracts/192618>
- [7] Fosu G.B (2001). Evaluation of population census data through demographic analysis. Symposium on Global Review of 2000 Round of Population and Housing Censuses, Department of Economic and Social Affairs, United Nations Secretariat, New York, 7-10 August 2001.
- [8] Al-Eideh B.M and Al-Omar H.O (2019). Population Projection Model using Exponential Growth Function with a Birth and Death Diffusion Growth Rate Processes. European Journal of Scientific Research, 151(3), 271-276
- [9] Oyebola A. and Adenike O. (2018). Big Data Analytic of Nigeria Population Census Data using MapReduce and K-Means Algorithm. International Journal of Computer Applications, 181(23), 14-21
- [10] Sondhi D. (2017) Application of Data Mining in Census Data Analysis using Weka. International Journal of Engineering Trends and Technology (IJETT), 53(3), 157-161
- [11] UNFPA (2019). Census. Retrieve from <https://www.unfpa.org/census>
- [12] UN (2008). Principles and Recommendations for Population and Housing Censuses. UN, New York, <https://doi.org/10.18356/8b8ce311-en>.
- [13] Population Reference Bureau (2001). Understanding and Using Population Projections. Retrieved from <https://www.prb.org/understandingandusingpopulationprojections>

Cite this article as :

Izakpa Getty Ebere, Ofualagba Mamuyovwi Helen, Ekhaton Uyiosa Emmanuel, "Automated Census Population Projection and Data Management System", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 1, pp. 244-254, January-February 2021. Available at doi : <https://doi.org/10.32628/CSEIT20666>
Journal URL : <https://ijsrcseit.com/CSEIT20666>