



VIT-AP UNIVERSITY

World Population Data Analytics using IBM Cognos

Project Report

Submitted by:

Tamanampudi Naveen Sai

19BCD7003

email: naveen.19bcd7003@vitap.ac.in

Submitted to:

Ms. Uma Rani

SmartInternz

as a part of Guided Project under the Externships: Data Analytics

Repository Link:

<https://github.com/smartinternz02/SI-GuidedProject-50058-1653039670>

Table of Contents

S No	TOPIC	PAGE No
1	Introduction	
	● 1.1 Overview	3
	● 1.2 Purpose	3-4
2	Literature Survey	
	● 2.1 Existing Problems	5
	● 2.2 Proposed Solutions	5
3	Theoretical Analysis	
	● 3.1 Block Diagram	6
	● 3.2 Hardware/Software Designing	6-7
4	Experimental Investigations	7-8
5	Flowchart	9
6	Results	9-13
7	Advantages and Disadvantages	13-14
8	Applications	14
9	Conclusion	15
10	Future Scope	15

1. Introduction

1.1 Overview

The 2019 Revision of World Population Prospects is the twenty-sixth round of official United Nations population estimates and projections that have been prepared by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.

The main results are presented in a series of Excel files displaying key demographic indicators for each UN development group, World Bank income group, geographic region, Sustainable Development Goals (SDGs) region, sub region and country or area for selected periods or dates within 1950-2100.

The aim of this project is to find and analyze United Nations population estimates, projections using the WPP2019_TotalPopulationBySex.csv dataset and present in a visual format for better understanding.

1.2 Purpose

Data are individual facts, statistics, or items of information, often numeric. In a more technical sense, data are a set of values of qualitative or quantitative variables about one or more persons or objects.

In this world, there is lots of data, but very less information. Without gaining information and knowledge, the potential of the data will be

wasted. Also, using data to take decisions is a tedious task. So, In this project, we are using IBM Cognos Analytics to create data visualizations like:

Top 10 Pop Total by location: To plot a tree map to show the top 10 most populous locations and their total population.

Pop Total by Time: To plot a line chart showing the variations in the total population as the time(years) passed.

Pop Male by Location and Pop Female by Location: To plot a pie chart showing Male population by location (for top 10 populous locations) and Female population by location(for top 10 populous locations).

Pop Male by Time and Pop Female by Time: To plot a packed bubble chart with male population by time and female population by time to show the size of increase of population with respective genders by year (top 10).

Pop Male, Pop Female, Pop Total using summary: To create a summary card to show total male population, female population and total population as numbers.

Finally, we will build a dashboard using the above visualizations. We can interact with the dashboard and can observe how the population changes, by selecting any location or time from any one of the charts.

2. Literature Survey

2.1 Existing Problems:

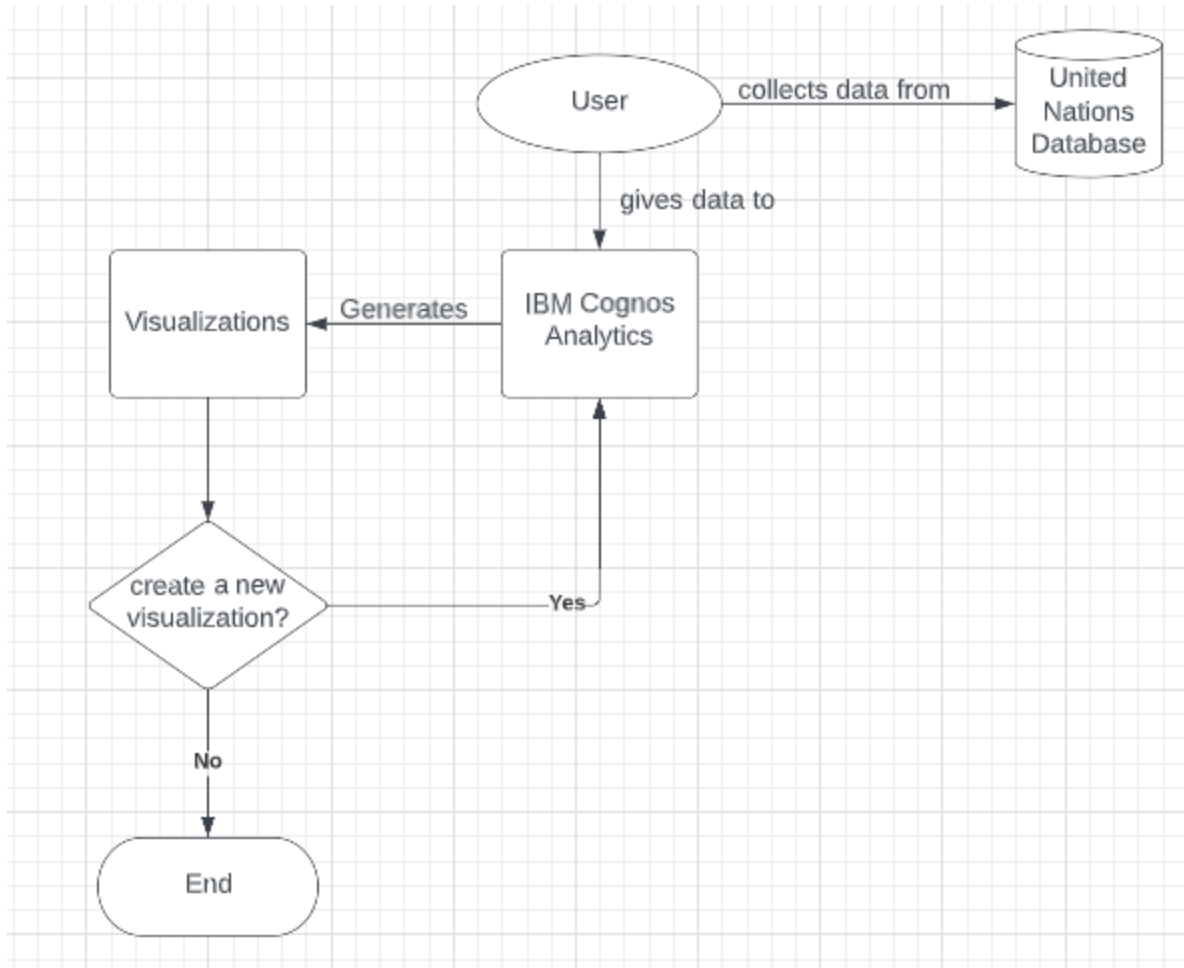
- There is lots of data all around the world. But, many people don't know how to use that data to make data driven decisions (D3).
- Having data is only useful when it's accurate. Low quality data not only serves no purpose, but it also uses unnecessary storage and can harm the ability to gather insights from clean data.
- Obtaining information from data without BI Tools is too tedious and required skilled workers.
- Lack of Understanding of the data creates false insights which creates chaos when put into practice.
- Information is useless when we don't know where to use it.

2.2 Proposed Solutions:

- We can use IBM Cognos Analytics to perform data preparation to clean and obtain accurate data.
- We can use IBM Cognos Analytics to create various charts to get good understanding about the data as graphs and charts are easy to understand than tabular data.
- We can create various dashboard and reports to represent the interests of the company in a single page.
- IBM Cognos Analytics provides automatic insights which can be used in decision making.

3. Theoretical Analysis

3.1 Block Diagram



3.2 Hardware/Software Designing:

Hardware Requirements:

Processor: Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99 GHz

Installed RAM: 8.00 GB (7.89 GB usable)

System type: 64-bit operating system, x64-based processor

Proper Internet

Software Requirements:

IBM Cognos Analytics: IBM Cognos Analytics is a web-based integrated business intelligence suite by IBM. It provides a tool set for reporting, analytics, score carding, and monitoring of events and metrics. The software consists of several components designed to meet the different information requirements in a company.

Chrome Browser: Google Chrome is a cross-platform web browser developed by Google. Its very important to use updated version of the chrome browser to support all the functionalities of the IBM Cognos Analytics.

4. Experimental Investigation

Dataset Name: WPP2019_TotalPopulationBySex.csv

Dataset Source:

[https://population.un.org/wpp/Download/Files/1_Indicators%20\(Standard\)/CSV_FILES/WPP2019_TotalPopulationBySex.csv](https://population.un.org/wpp/Download/Files/1_Indicators%20(Standard)/CSV_FILES/WPP2019_TotalPopulationBySex.csv)

Number of Columns: 10

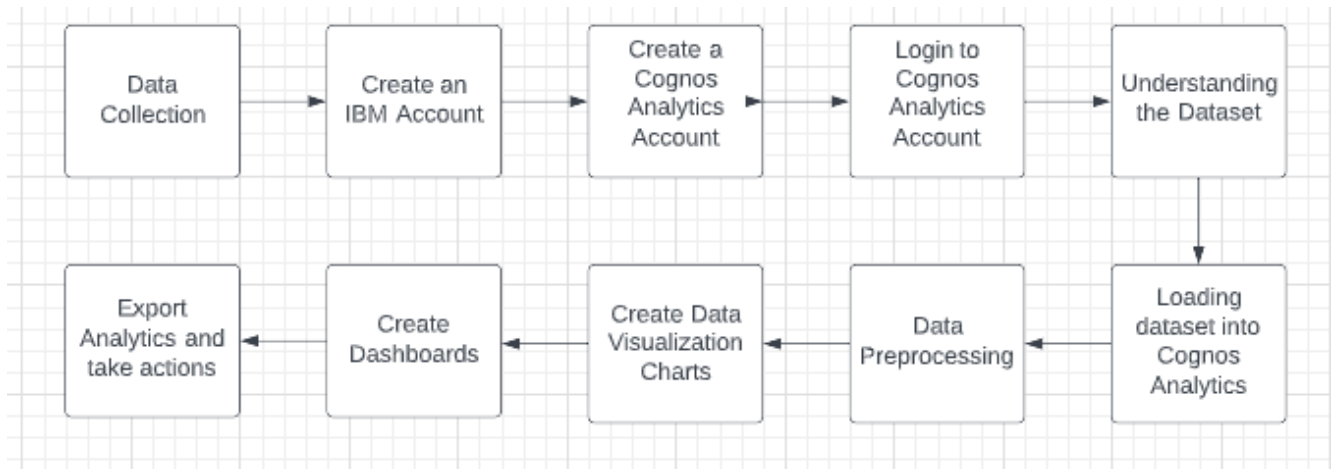
Number of Rows: 280932

Description of the Dataset:

The data set contains the following attributes:

- **LocID (numeric):** numeric code for the location; for countries and areas, it follows the ISO 3166-1 numeric standard
- **Location (string):** name of the region, sub-region, country or area
- **VarID (numeric):** numeric code for the variant
- **Variant (string):** projection variant name (Medium is the most used)
- **Time (string):** label identifying the single year (e.g. 1950) or the period of the data (e.g. 1950-1955)
- **MidPeriod (numeric):** numeric value identifying the mid period of the data, with the decimal representing the month (e.g. 1950.5 for July 1950)
- **PopMale (numeric):** Male Population in thousands
- **PoFemale (numeric):** Female Population in thousands
- **PopTotal (numeric):** Total Population in thousands
- **PopDensity (numeric):** Population Density in thousands

5. Flow Chart



6. Results

Data Preprocessing/Preparation

- Remove the existing column Pop total.
- Create A New Calculated Field Pop Total = Pop Male + Pop Female and Format the data.
- Format Pop Male and Pop Female field Data types.
- Here, the PopMale, PopFemale and PopTotal are in double datatype. But, population never exists as fractions or double values. So, converting their data types to Integer.

Student Dashboard | smartinternz02/SI-GuidedProject | Guided Project Dataprep

eu2.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=iA82E7471125E4CE1ACEE0665E1FC410A&objRef=iA82E7471125E4CE1ACEE0665E1FC410A&tid=21402...

IBM Cognos Analytics with Watson | Guided Project Dataprep

Search content

Properties

Data module

Search

Guided Project Dataprep

Navigation paths

WPP2019_T...BySex.csv

Pop Total

Row Id

LocID

Location

VarID

Variant

Time

MidPeriod

PopMale

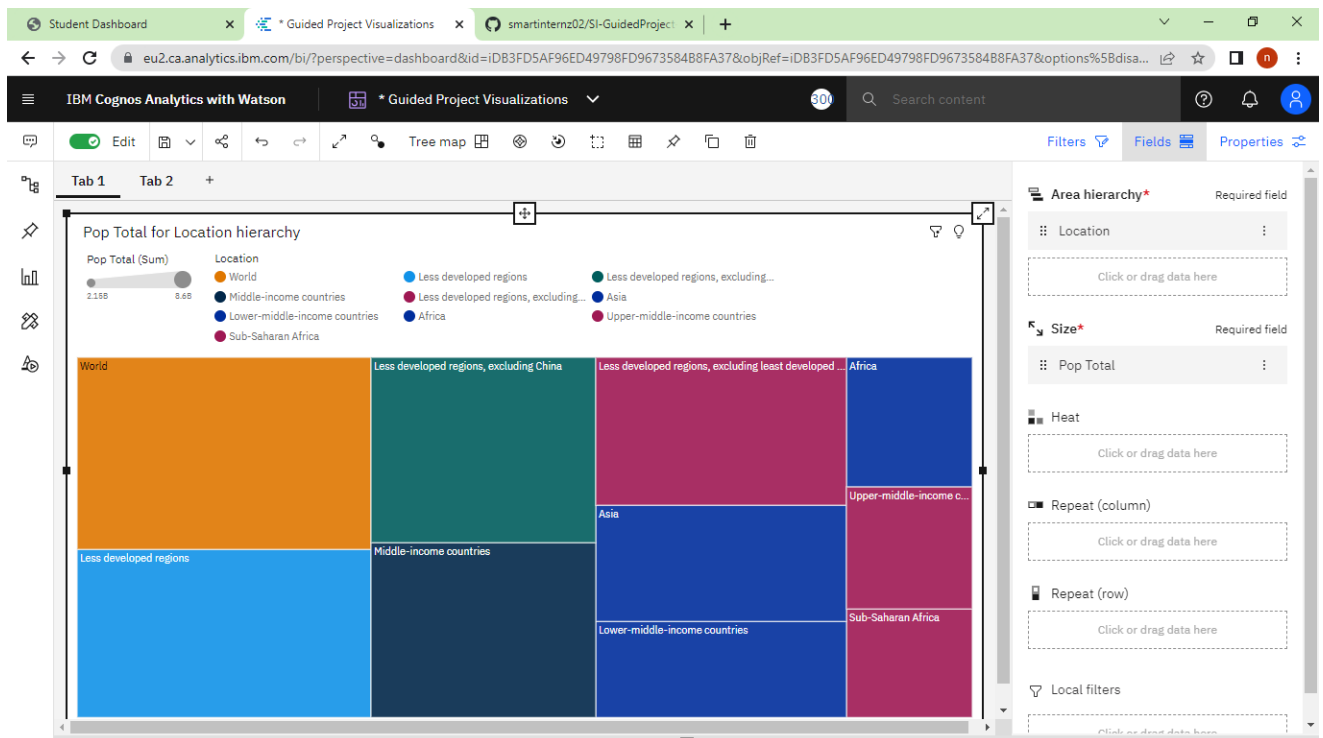
PopFemale

PopDensity

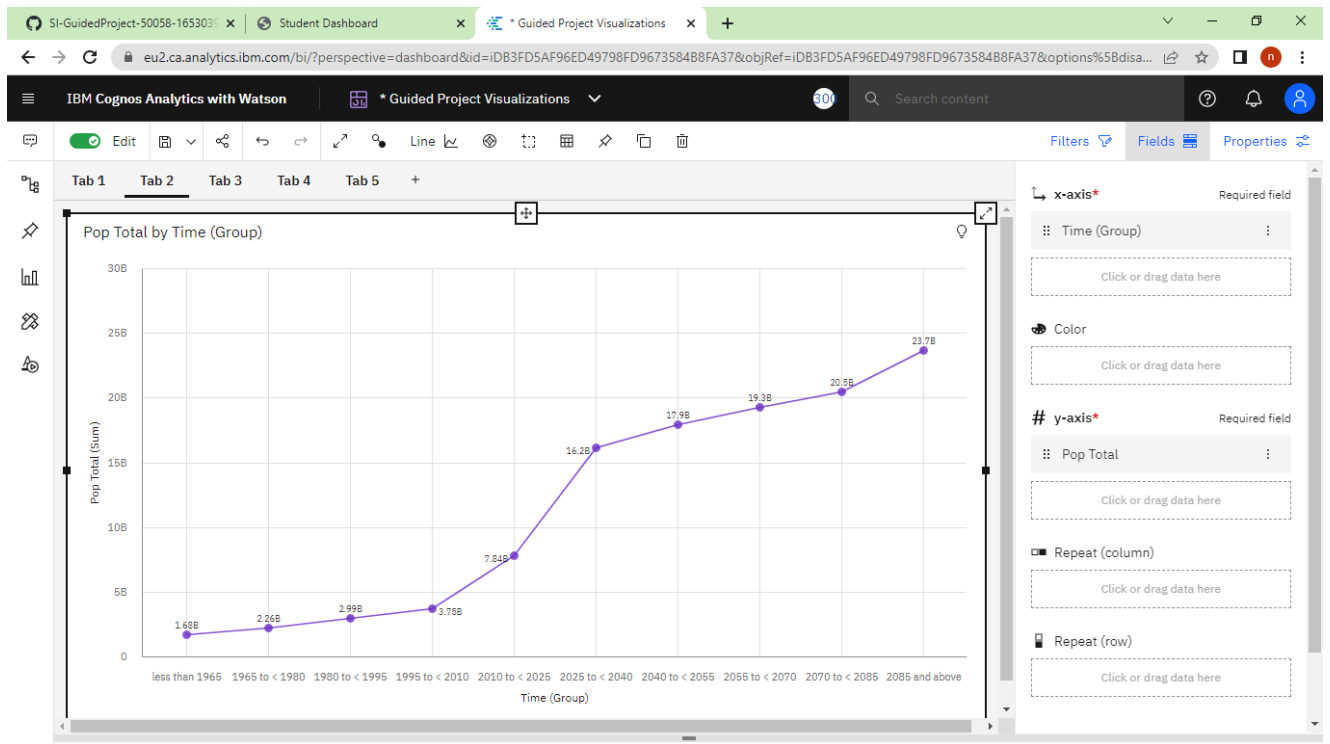
ID	Variant	Time	MidPeriod	PopMale	PopFemale
	Medium	1950	1950.5	4,099	3,653
	Medium	1951	1951.5	4,135	3,705
	Medium	1952	1952.5	4,174	3,762
	Medium	1953	1953.5	4,218	3,821
	Medium	1954	1954.5	4,266	3,885
	Medium	1955	1955.5	4,319	3,952
	Medium	1956	1956.5	4,376	4,023
	Medium	1957	1957.5	4,437	4,098
	Medium	1958	1958.5	4,503	4,177
	Medium	1959	1959.5	4,574	4,260
	Medium	1960	1960.5	4,650	4,347
	Medium	1961	1961.5	4,730	4,439

Data Visualization Charts:

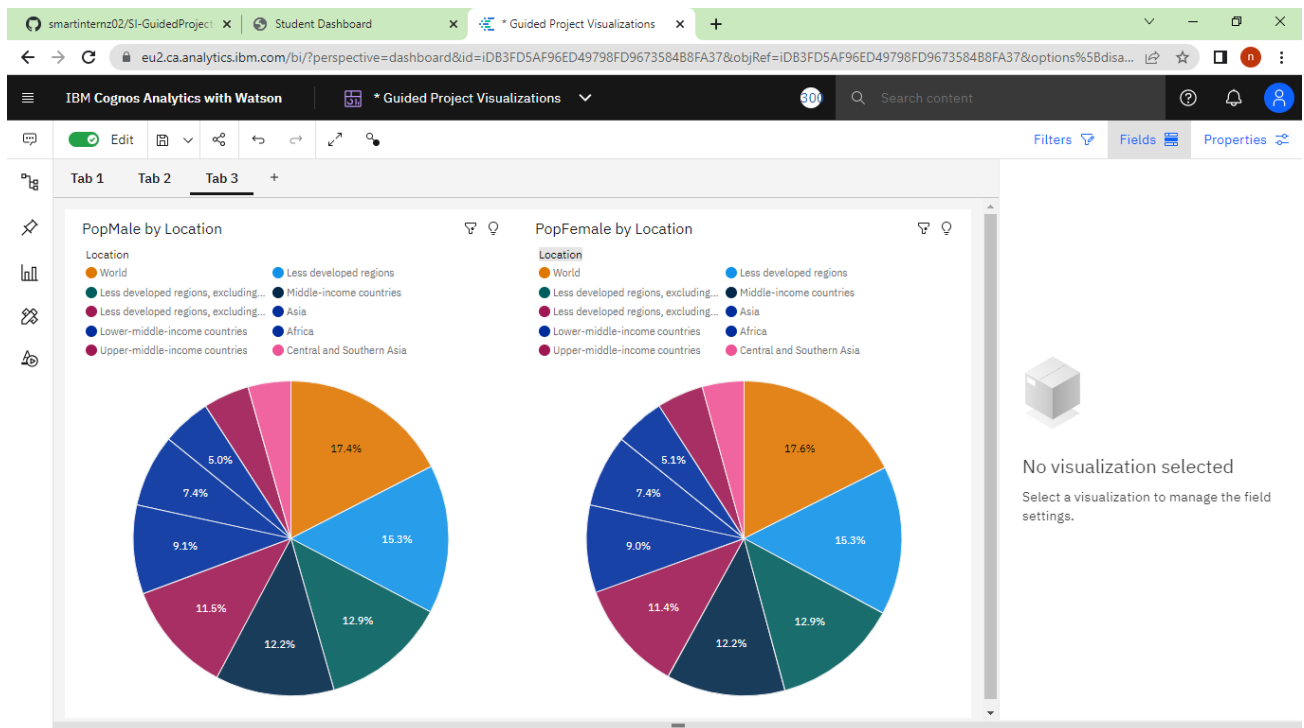
1. Top10 Pop Total By Location Using Tree Map



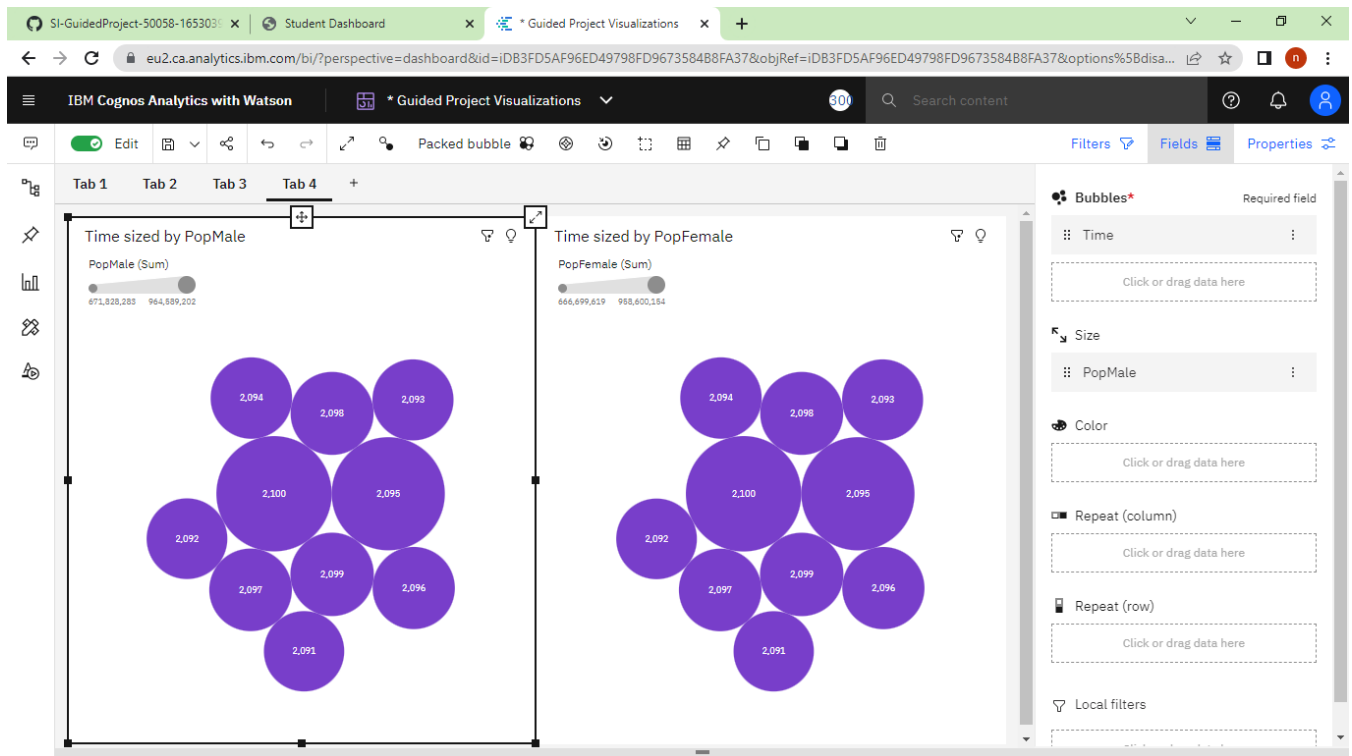
2. Pop Total By Time Using Line Chart



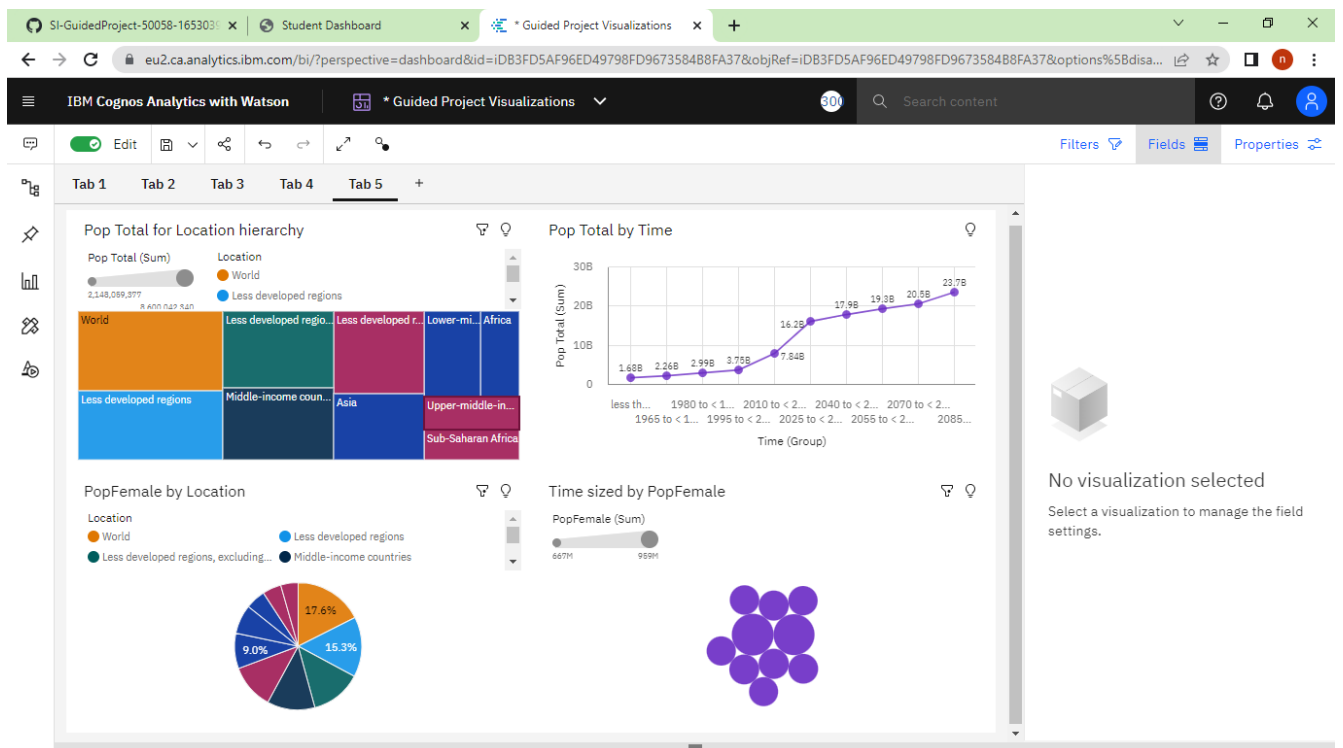
3. Pop Male By Location And Pop Female By Location Using Pie Charts



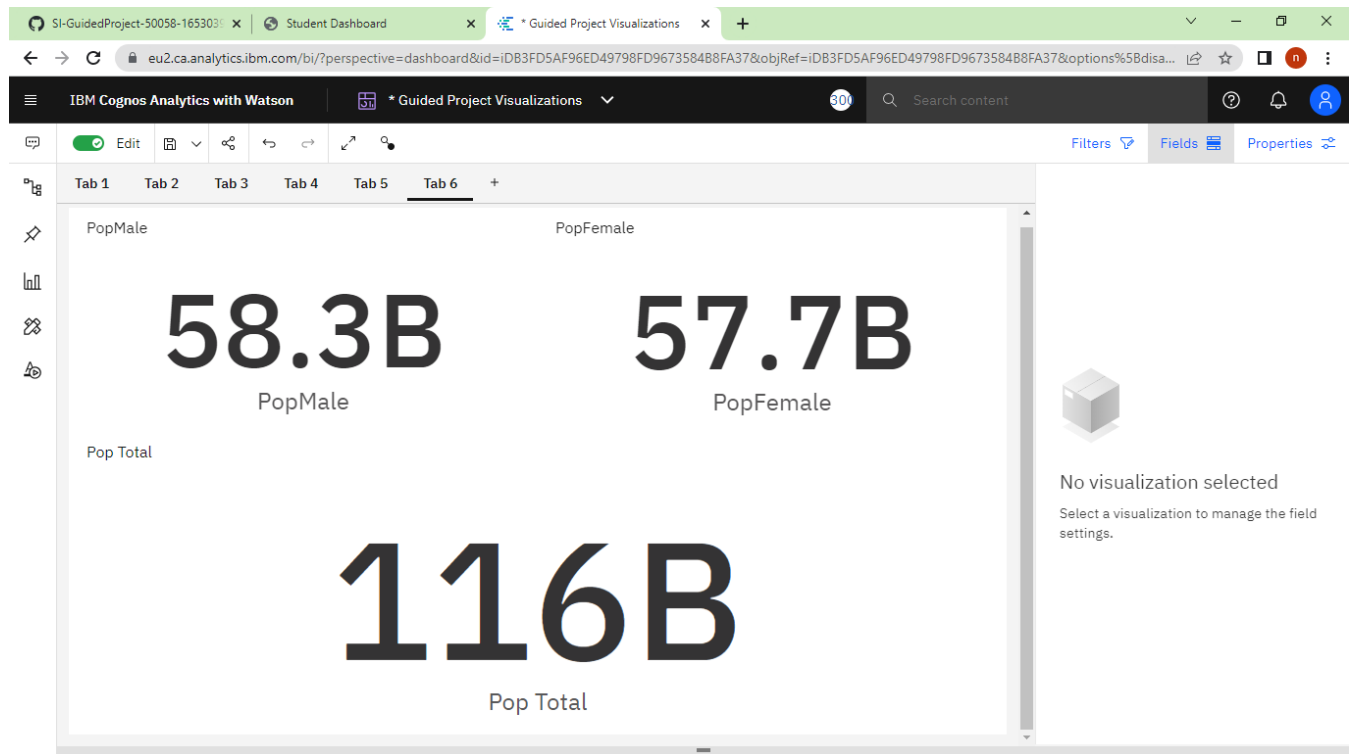
4. Pop Male By Time And Pop Female Using Packed Bubble Charts



5. Building Of Dashboard



6. Pop Male, Pop Female And Pop Total Using Summary



7. Advantages and Disadvantages

Advantages:

1. We can get important insights and patterns from the dataset very easily when compared to naive approaches.
2. The obtained information and insights can be applied onto various fields based on the domain which the dataset belongs to.
3. From this project, we can observe the most populous locations and which locations has highest population density to take necessary actions for the sustainability of the people over there.
4. We can also see the increase in population as the time passes.
5. Dashboards can be created which enables us to put all the

necessary visualizations in a single page.

6. All these features enhances our decision making capability i.e. Data Driven Decision (D3) Making.

Disadvantages:

1. The software used here i.e. IBM Cognos Analytics is not free of cost. We need to pay to get their services. This might add additional overhead to the organization.
2. This project requires large datasets to validate that the insights obtained are true and useful. If the dataset is small, then small changes in the dataset results in false insights.
3. We need skilled workers to work with this IBM Cognos Analytics.
4. If the organization doesnt have template for dashboards, then it may create additional confusion as the design of dashboard chosen depends on the worker.

8. Applications

- This can be utilized by the United Nations to observe the rate of increase in population in a particular location to take necessary actions.
- They can also obtain the names of most populous regions and most population dense locations to check if the amenities at those locations are up to the mark.
- This can also be utilized by International Welfare Organizations for the same purposes mentioned above.

9. Conclusion

In this project, we have successfully collected data from United Nations website, understood the dataset and performed data preparation to obtain accurate data using IBM Cognos Analytics. Then we have created data visualizations and dashboards. These dashboards provided various insights to the user which can be applied in Decision Making. These insights on world population dataset can be utilized by various organizations for the welfare of the people.

10. Future Scope

- This project can be fine grained to local level from the global level by collecting data from local regions. This data can be used by the local authorities to gain insights and take necessary actions. This will create even more impact as minor issues will also come into light. When global level is considered, these minor issues might be overshadowed by the other major issues.
- Voice Assistance can be added to IBM Cognos Analytics platform to ease the process of creating visualizations.