

Name: Sanjana Singh

Registration No.: 19BCE10274

Mentor's Name: Uma Rani

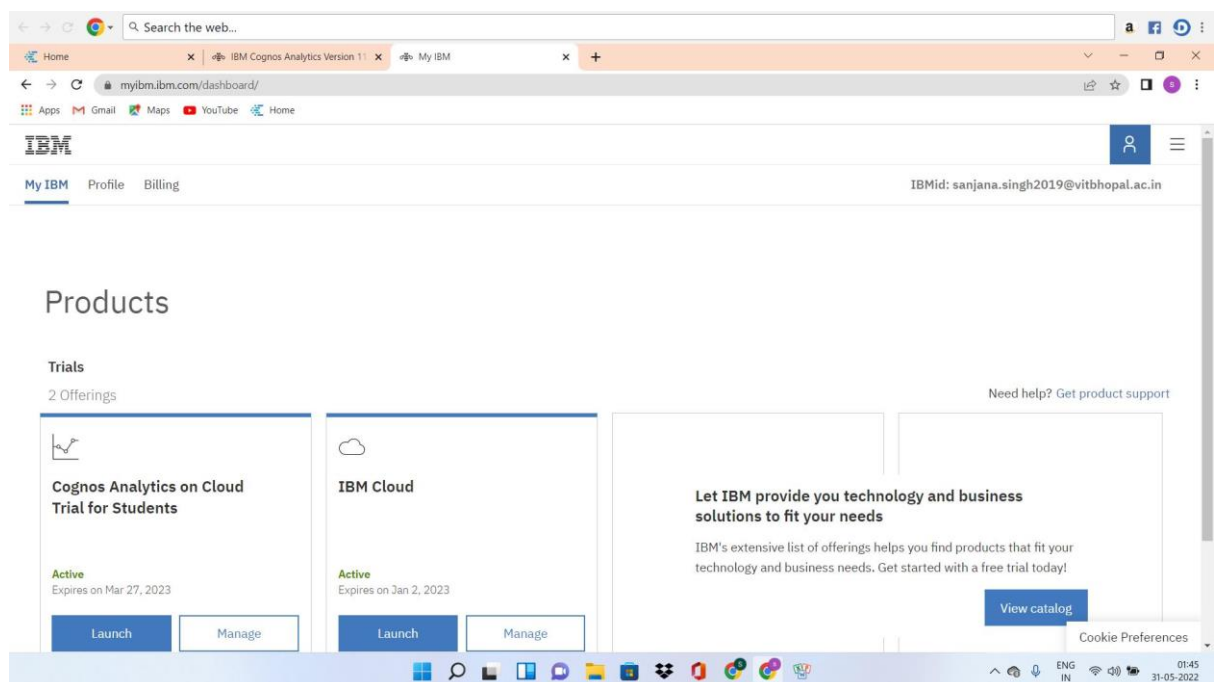
World Population Data Analytics using IBM Cognos

- ABOUT THE DATASET

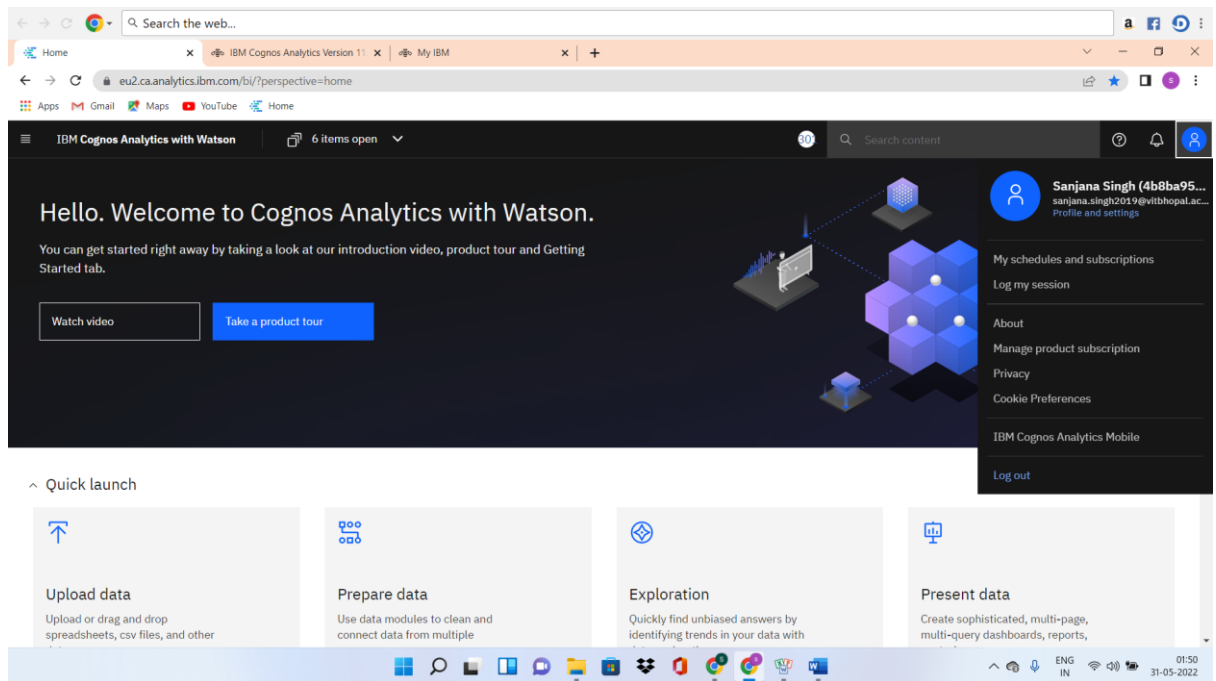
A description of the empirical data used and the methods applied in revising past estimates of population and components of demographic change (fertility, child, adult and overall mortality, international migration) is available here for each country or area for the period 1950-2020. For the countries with less than 90,000 inhabitants in 2019, only the data sources for total population are made available.

On 28 August 2019 a minor technical correction was made to the population projected after 2050 for selected countries and regions, and to the population interpolated by single year and single age for both sexes. Interactive Data, Excel and CSV files have been updated accordingly.

1. IBM Cloud Creation



2. IBM Cognos Analytics



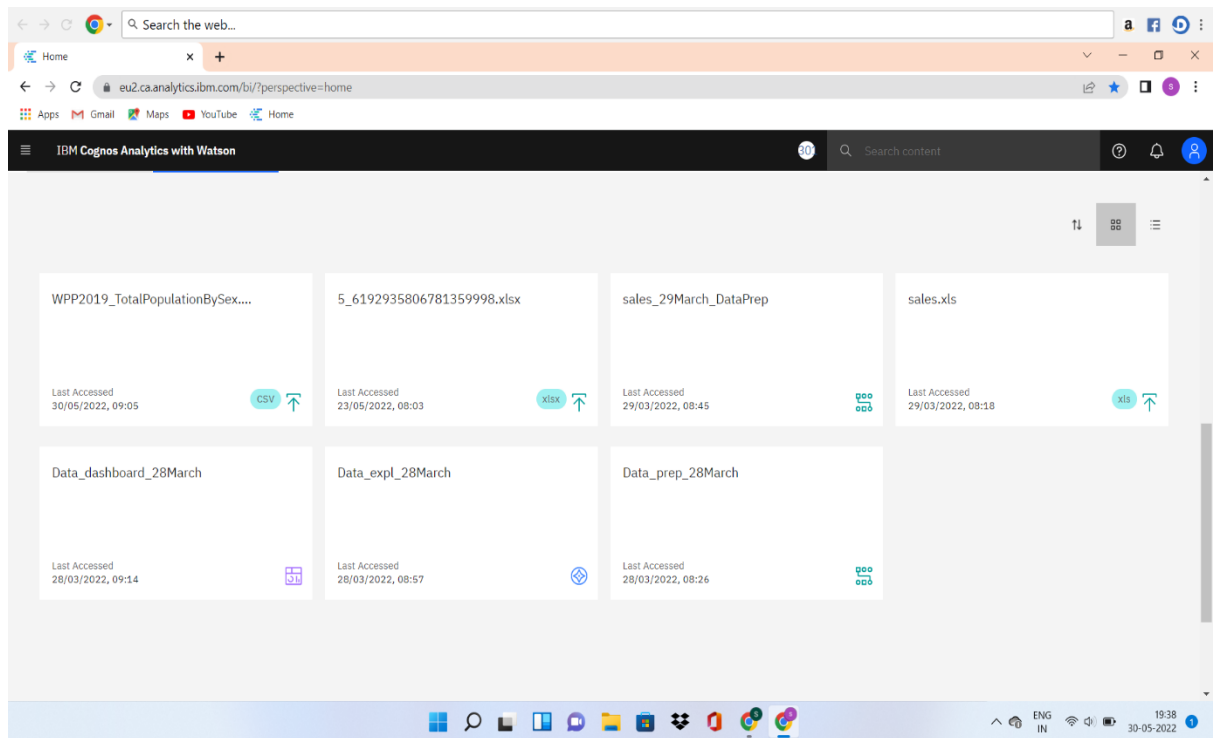
3. Working with the dataset

3.1 Understand the dataset

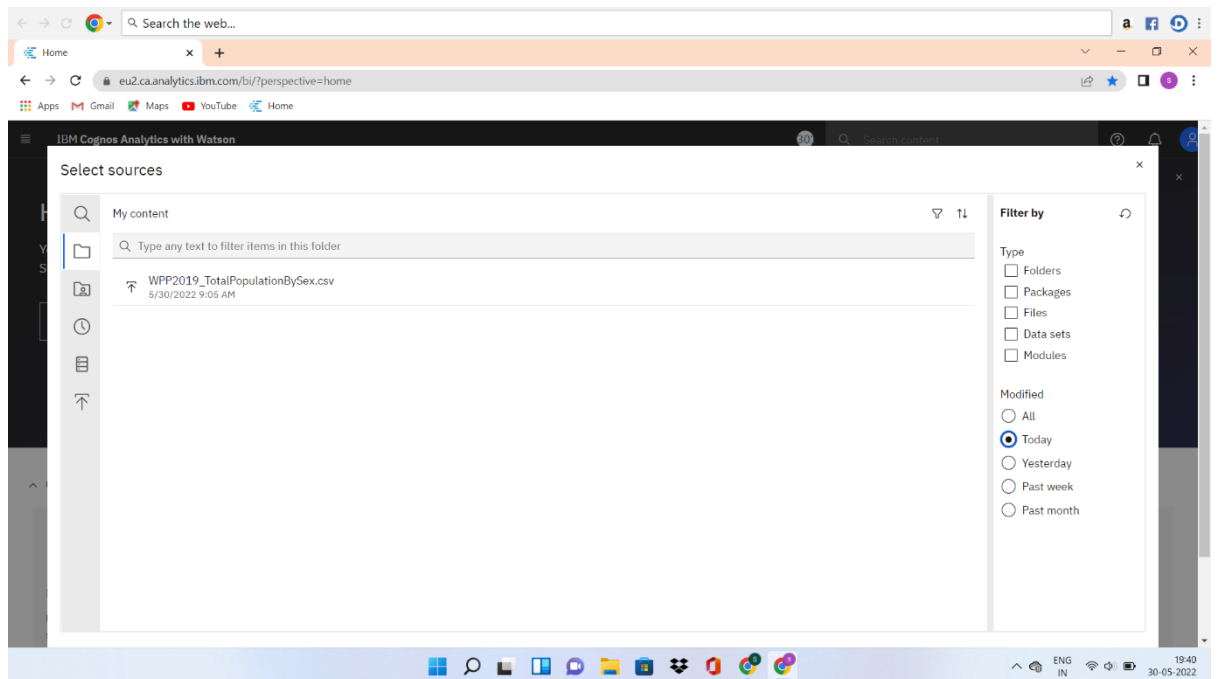
The screenshot shows a Microsoft Excel spreadsheet titled 'WPP2019_TotalPopulationBySex'. The spreadsheet contains data for Afghanistan, with columns for Location, VarID, Variant, Time, MidPeriod, PopMale, PopFemale, PopTotal, and PopDensity. The data is organized into rows, with the first row (row 1) containing the headers and subsequent rows (rows 2-23) containing the data for Afghanistan.

	Location	VarID	Variant	Time	MidPeriod	PopMale	PopFemale	PopTotal	PopDensity
1	4 Afghanistan	2	Medium	1950	1950.5	4099.243	3652.874	7752.117	11.874
2	4 Afghanistan	2	Medium	1951	1951.5	4134.756	3705.395	7840.151	12.009
3	4 Afghanistan	2	Medium	1952	1952.5	4174.45	3761.546	7935.996	12.156
4	4 Afghanistan	2	Medium	1953	1953.5	4218.336	3821.348	8039.684	12.315
5	4 Afghanistan	2	Medium	1954	1954.5	4266.484	3884.832	8151.316	12.486
6	4 Afghanistan	2	Medium	1955	1955.5	4318.945	3952.047	8270.992	12.669
7	4 Afghanistan	2	Medium	1956	1956.5	4375.8	4023.073	8398.873	12.865
8	4 Afghanistan	2	Medium	1957	1957.5	4437.157	4098	8535.157	13.073
9	4 Afghanistan	2	Medium	1958	1958.5	4503.156	4176.941	8680.097	13.295
10	4 Afghanistan	2	Medium	1959	1959.5	4573.914	4260.033	8833.947	13.531
11	4 Afghanistan	2	Medium	1960	1960.5	4649.573	4347.394	8996.967	13.781
12	4 Afghanistan	2	Medium	1961	1961.5	4730.25	4439.156	9169.406	14.045
13	4 Afghanistan	2	Medium	1962	1962.5	4816.05	4535.392	9351.442	14.324
14	4 Afghanistan	2	Medium	1963	1963.5	4907.03	4636.17	9543.2	14.618
15	4 Afghanistan	2	Medium	1964	1964.5	5003.245	4741.527	9744.772	14.926
16	4 Afghanistan	2	Medium	1965	1965.5	5104.765	4851.553	9956.318	15.25
17	4 Afghanistan	2	Medium	1966	1966.5	5210.122	4964.718	10174.84	15.585
18	4 Afghanistan	2	Medium	1967	1967.5	5319.123	5080.813	10399.94	15.93
19	4 Afghanistan	2	Medium	1968	1968.5	5434.458	5202.606	10637.06	16.293
20	4 Afghanistan	2	Medium	1969	1969.5	5559.836	5333.936	10893.77	16.686
21	4 Afghanistan	2	Medium	1970	1970.5	5697.024	5476.63	11173.65	17.115
22	4 Afghanistan	2	Medium	1971	1971.5	5845.351	5630.099	11475.45	17.577
23	4 Afghanistan	2	Medium	1972	1972.5	6000.805	5790.227	11791.03	18.061

3.2 Loading the dataset



3.3 Prepare the datasets



The screenshot shows the IBM Cognos Analytics with Watson interface. On the left, the 'Data module' sidebar is open, showing a search bar and a list of data modules. The 'New data module' module is selected, and its contents are listed: 'Navigation paths', 'WPP2019_T...BySex.csv', and several columns: '# Row Id', '# LocID', 'Location', '# VarID', 'Variant', 'Time', 'MidPeriod', 'PopMale', 'PopFemale', 'PopTotal', and 'PopDensity'. The main area displays a 'Preview data' message with a calendar icon and text: 'To preview data, select a table, a column in a table, or a folder that contains columns.'

The screenshot shows the IBM Cognos Analytics with Watson interface with the 'Grid' view selected. The 'Data module' sidebar on the left shows the 'WPP2019_T...BySex.csv' module selected. The main area displays a table with 12 rows and 8 columns. The columns are: 'Row Id', 'LocID', 'Location', 'VarID', 'Variant', 'Time', and 'MidPeriod'. The data shows population statistics for Afghanistan from 1950 to 1961.

Row Id	LocID	Location	VarID	Variant	Time	MidPeriod
1	4	Afghanistan	2	Medium	1950	1950.5
2	4	Afghanistan	2	Medium	1951	1951.5
3	4	Afghanistan	2	Medium	1952	1952.5
4	4	Afghanistan	2	Medium	1953	1953.5
5	4	Afghanistan	2	Medium	1954	1954.5
6	4	Afghanistan	2	Medium	1955	1955.5
7	4	Afghanistan	2	Medium	1956	1956.5
8	4	Afghanistan	2	Medium	1957	1957.5
9	4	Afghanistan	2	Medium	1958	1958.5
10	4	Afghanistan	2	Medium	1959	1959.5
11	4	Afghanistan	2	Medium	1960	1960.5
12	4	Afghanistan	2	Medium	1961	1961.5

- Removing the existing PopTotal column

The screenshot shows the IBM Cognos Analytics interface. A data table is displayed with the following columns: Variant, Time, MidPeriod, PopMale, PopFemale, PopTotal, and PopDensity. The 'PopTotal' column is highlighted, and a context menu is open over it, showing various actions. The table contains data for years 1955 to 1966.

Variant	Time	MidPeriod	PopMale	PopFemale	PopTotal	PopDensity
1955	1955.5	1955.5	4318.945	3952.047	8270.992	12.669
1956	1956.5	1956.5	4375.8	4023.073	8398.873	12.865
1957	1957.5	1957.5	4437.157	4098	8535.157	13.073
1958	1958.5	1958.5	4503.156	4176.941	8680.097	13.295
1959	1959.5	1959.5	4573.914	4260.033	8833.947	13.531
1960	1960.5	1960.5	4649.573	4347.394	8996.967	13.781
1961	1961.5	1961.5	4730.25	4439.156	9169.406	14.045
1962	1962.5	1962.5	4816.05	4535.392	9351.442	14.324
1963	1963.5	1963.5	4907.03	4636.17	9543.2	14.618
1964	1964.5	1964.5	5003.245	4741.527	9744.772	14.926
1965	1965.5	1965.5	5104.765	4851.553	9956.318	15.25
1966	1966.5	1966.5	5210.122	4964.718	10174.84	15.585

- Calculating the new PopTotal column. $\text{PopTotal} = \text{PopMale} + \text{PopFemale}$ and validate the expression to formate the data.

The screenshot shows the 'Create calculation' dialog in IBM Cognos Analytics. The 'Name' field is 'C_PopTotal_New'. The 'Expression' field contains '1 PopMale + PopFemale'. The 'Validation Results' section shows a green checkmark and the message 'The expression is valid.' The dialog has 'Cancel' and 'OK' buttons at the bottom.

- Creation of new PopTotal column

The screenshot shows the IBM Cognos Analytics interface. On the left, the 'Data module' pane lists the data source 'WPP2019_T...BySex.csv' and its columns: Row Id, LocID, Location, VarID, Variant, Time, MidPeriod, PopMale, PopFemale, and PopDensity. The 'C_PopTotal_New' column is selected. The main grid displays 12 rows of data for Afghanistan, with columns for Row Id, LocID, Location, VarID, Variant, and Time.

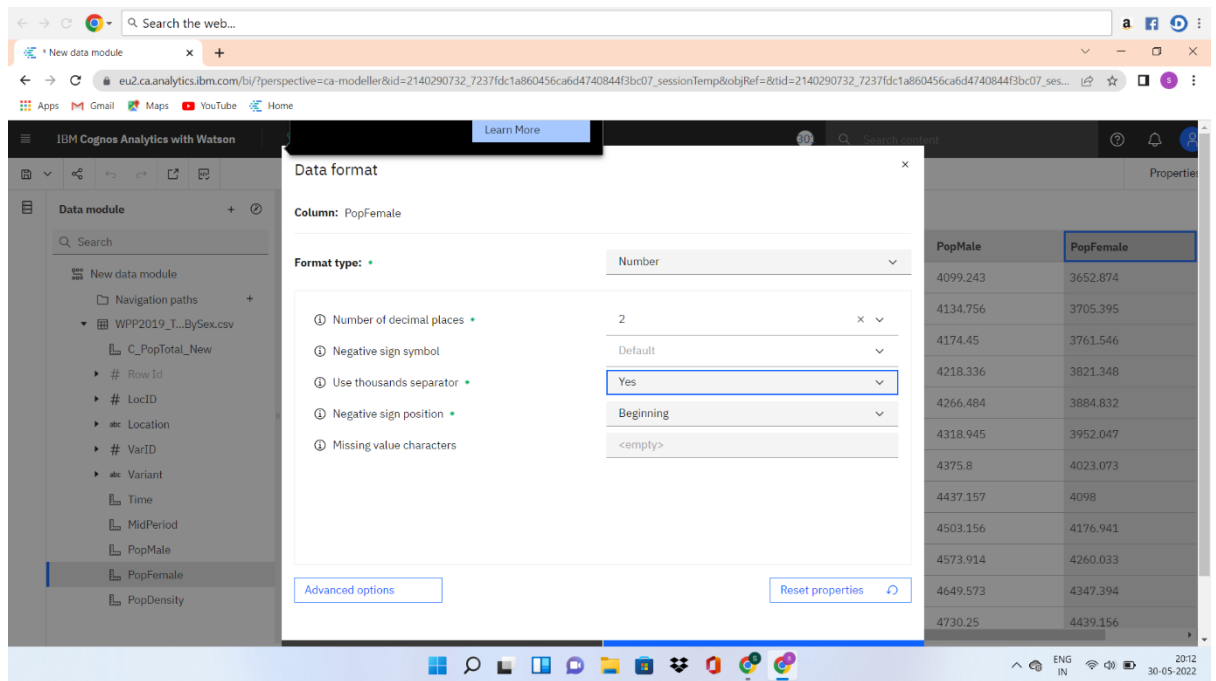
Row Id	LocID	Location	VarID	Variant	Time
1	4	Afghanistan	2	Medium	1950
2	4	Afghanistan	2	Medium	1951
3	4	Afghanistan	2	Medium	1952
4	4	Afghanistan	2	Medium	1953
5	4	Afghanistan	2	Medium	1954
6	4	Afghanistan	2	Medium	1955
7	4	Afghanistan	2	Medium	1956
8	4	Afghanistan	2	Medium	1957
9	4	Afghanistan	2	Medium	1958
10	4	Afghanistan	2	Medium	1959
11	4	Afghanistan	2	Medium	1960
12	4	Afghanistan	2	Medium	1961

- Formate the PopMale data type

The screenshot shows the 'Data format' dialog box for the 'PopMale' column. The 'Format type' is set to 'Number'. The 'Number of decimal places' is set to 2. The 'Negative sign symbol' is set to 'Default'. The 'Use thousands separator' is set to 'Yes'. The 'Negative sign position' is set to 'Beginning'. The 'Missing value characters' are set to '<empty>'. The background shows a data table with columns 'MidPeriod' and 'PopMale'.

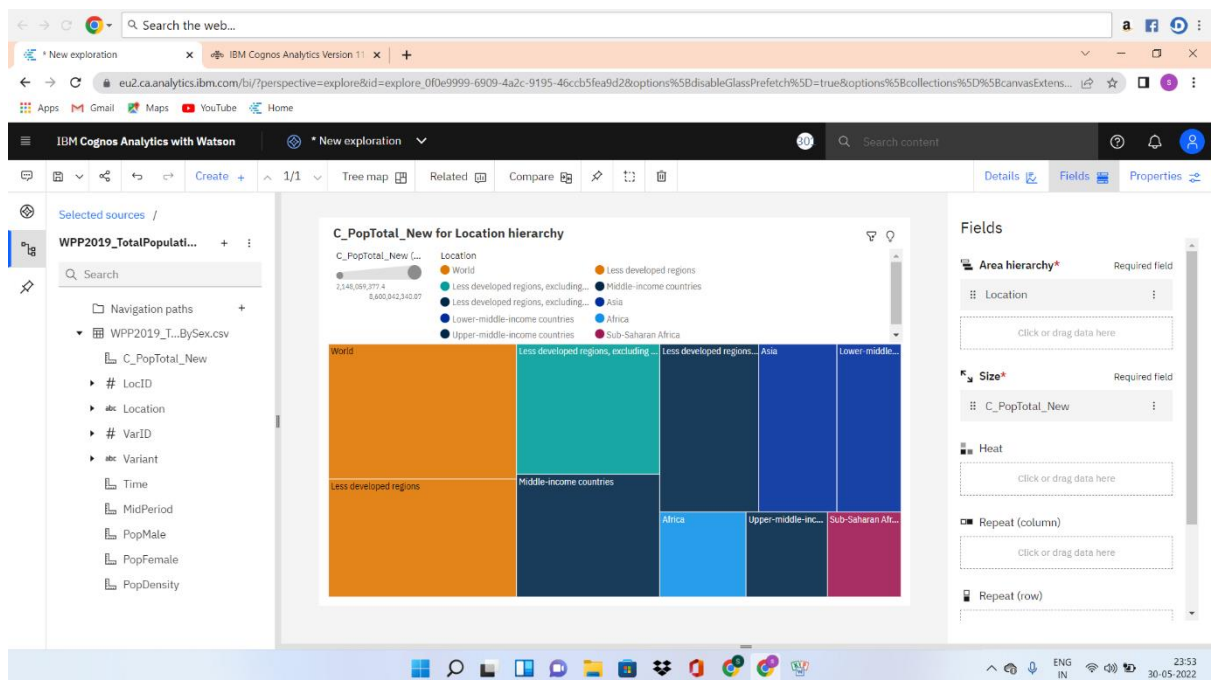
MidPeriod	PopMale
1950.5	4099.243
1951.5	4134.756
1952.5	4174.45
1953.5	4218.336
1954.5	4266.484
1955.5	4318.945
1956.5	4375.8
1957.5	4437.157
1958.5	4503.156
1959.5	4573.914
1960.5	4649.573
1961.5	4730.25

- Formate the PopFemale data type

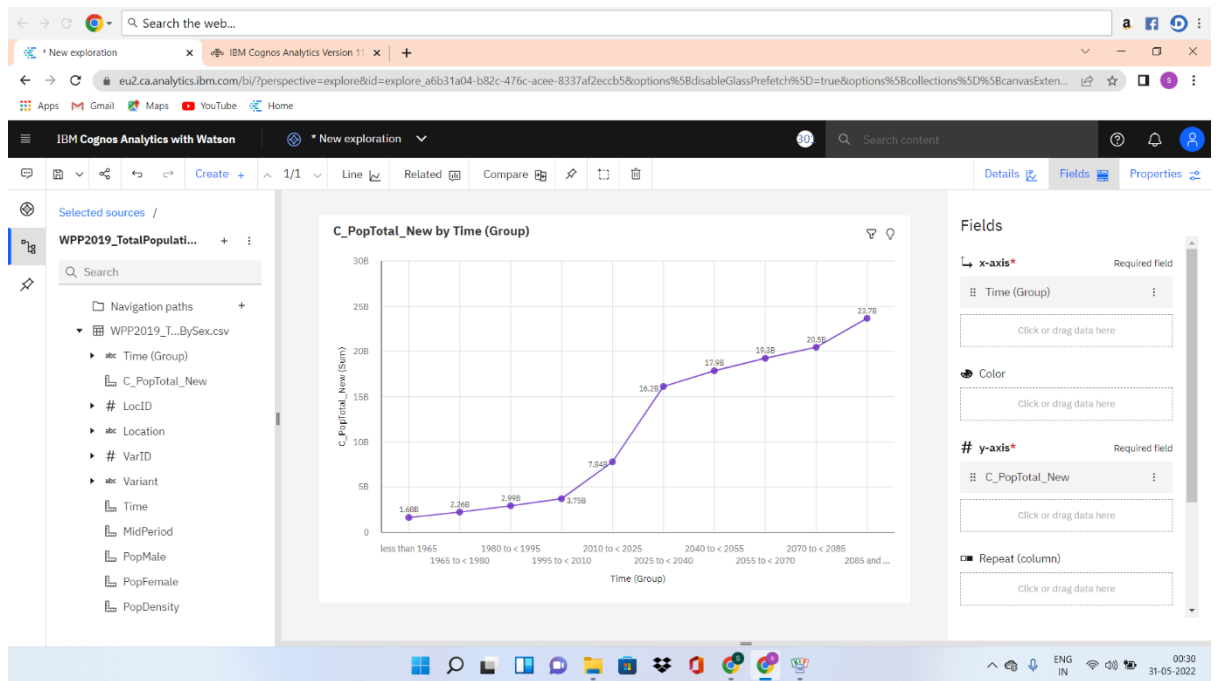


2. Data Visualization Chart

4.1 Top10 Pop Total by Location Using Tree Map

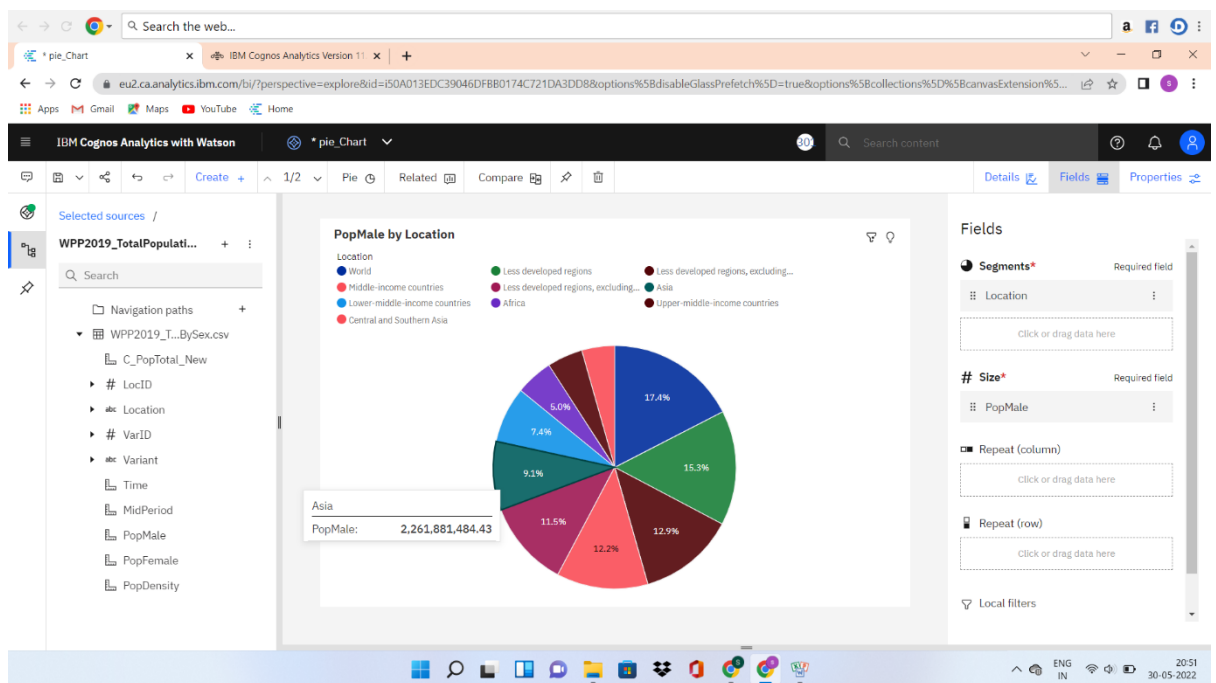


4.2 Pop Total by Time Using Line Chart

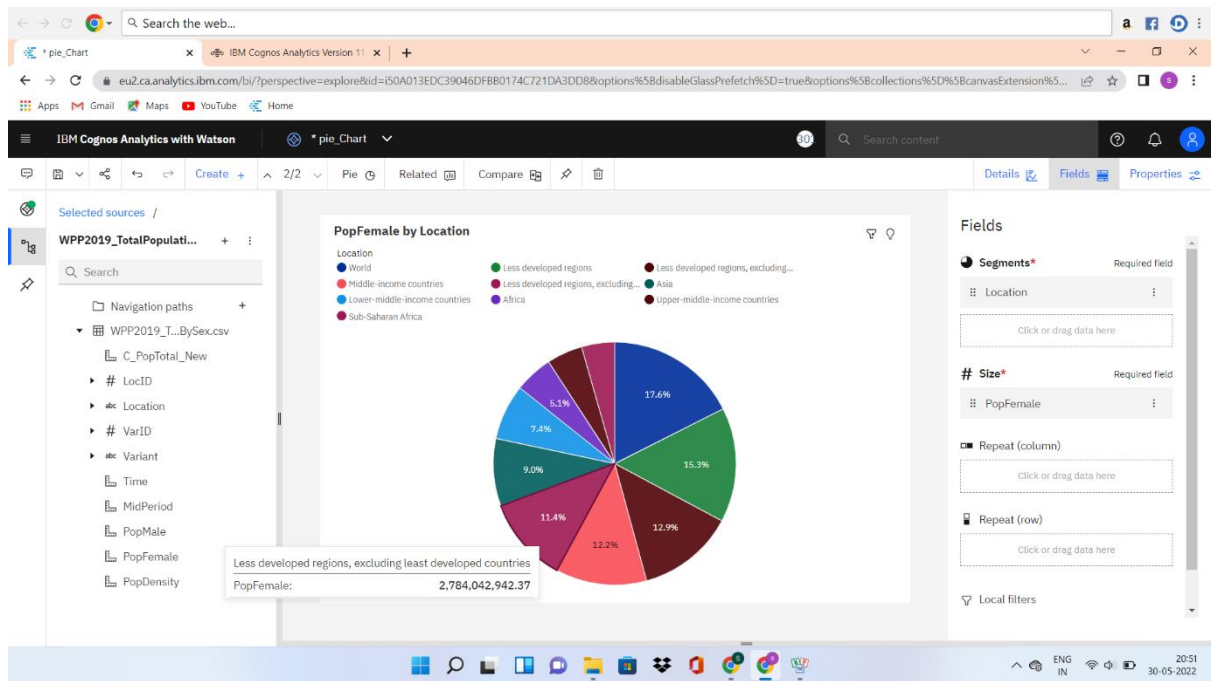


4.3 Pop Male by Location and Pop Female by Location using Pie Charts

- PopMale by Location

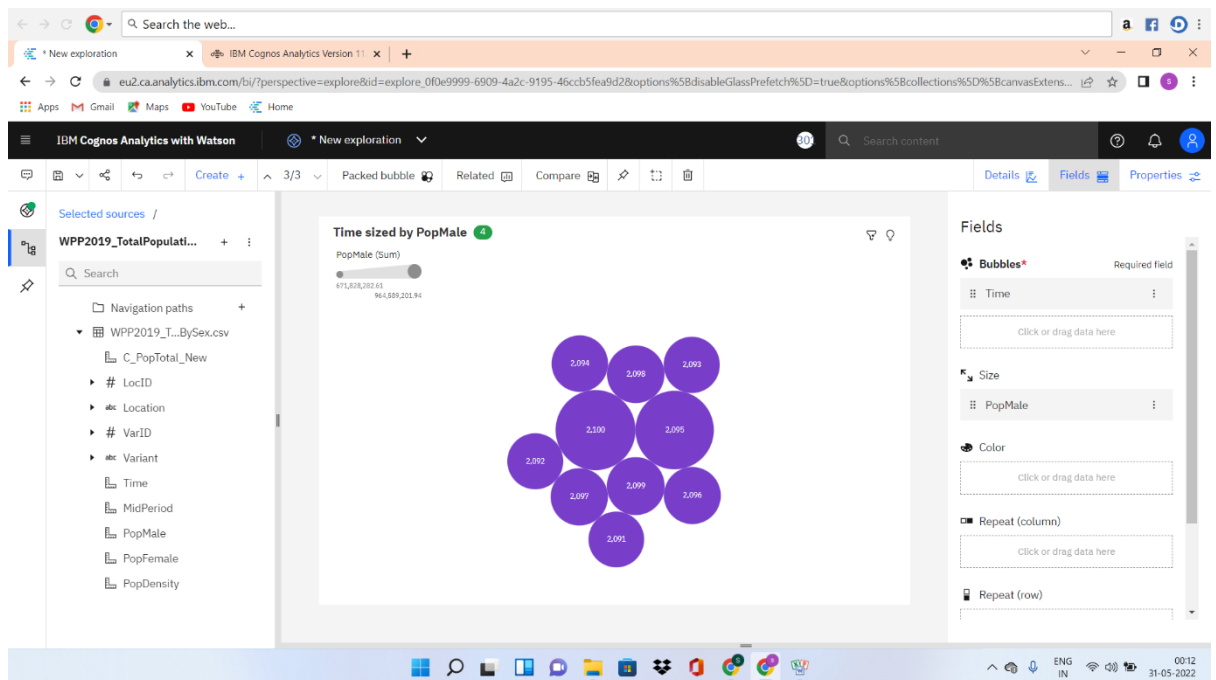


- PopFemale by Location

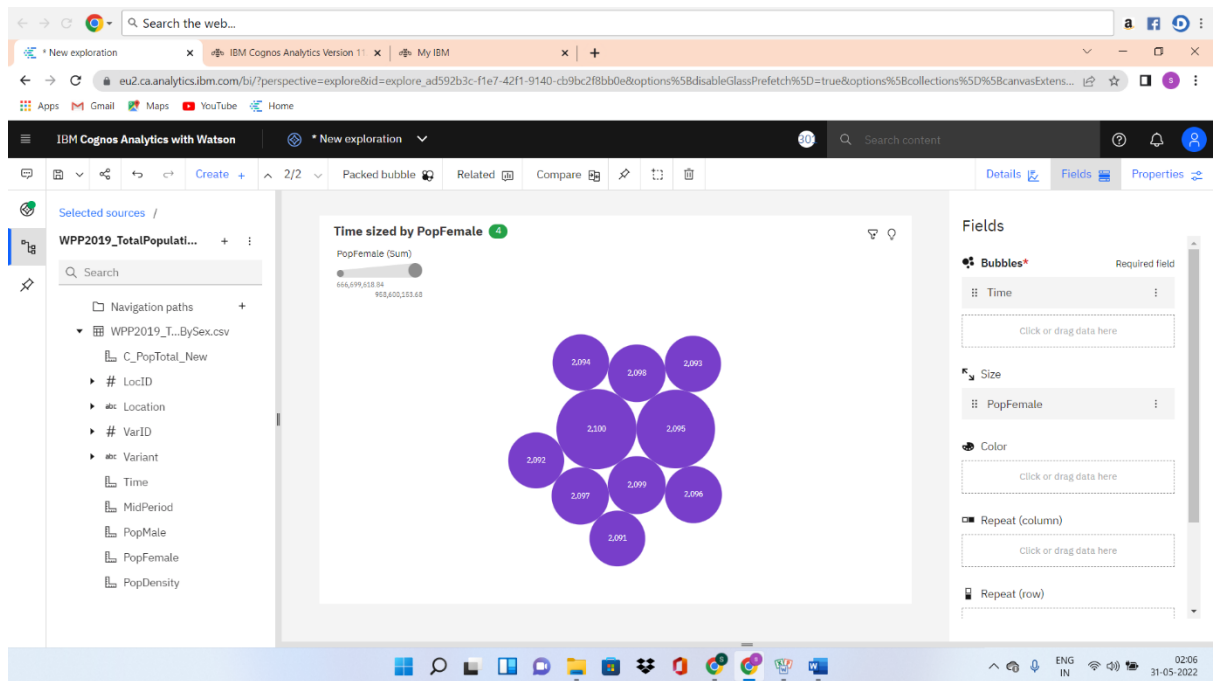


4.4 Pop Male by Time and Pop Female Using Packed Bubble Charts

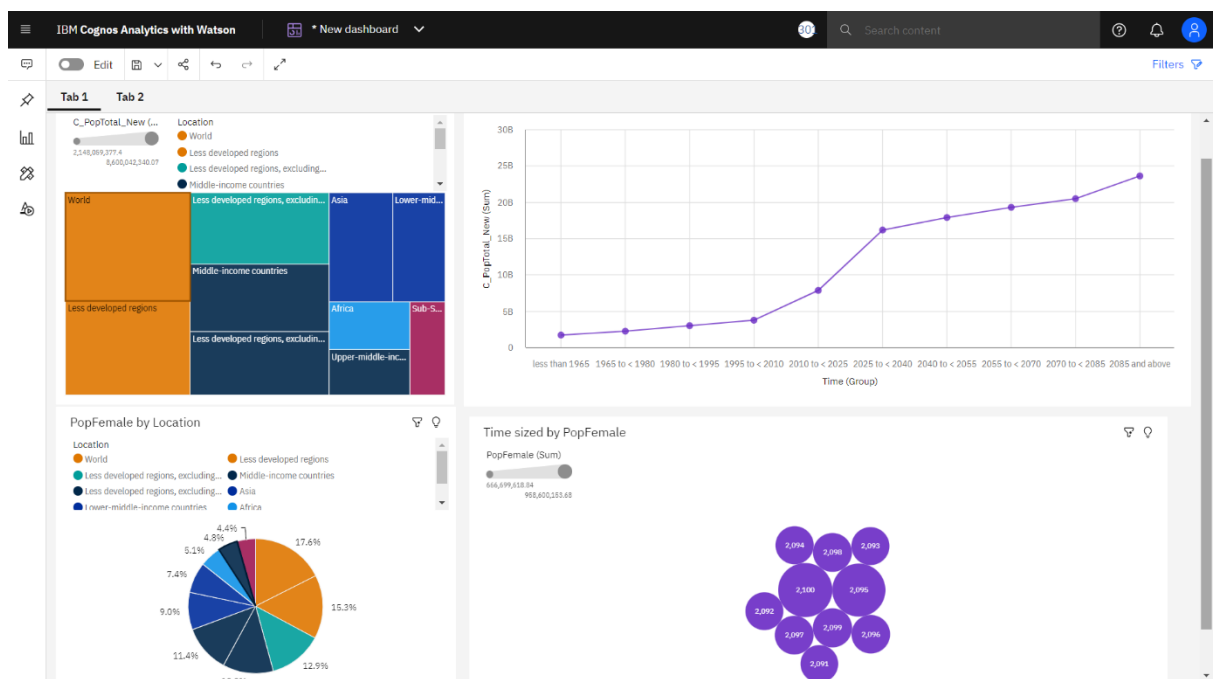
- Time sized by PopMale



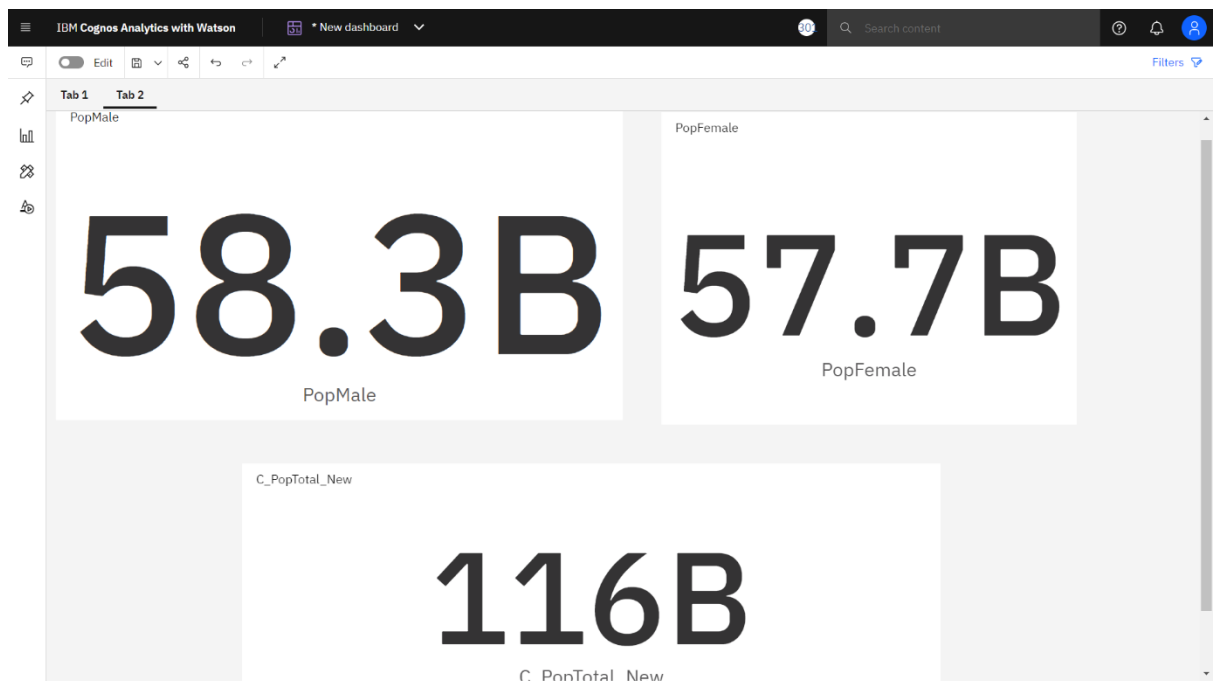
- Time sized by PopFemale



4.5 Building of Dashboard



4.6 Pop Male, Pop Female and Pop Total using Summary



- **Advantages and Disadvantages of the IBM Tool**

Pros

1. Lower costs—reduces maintenance due to complete report coverage and a zero-footprint environment.
2. Faster results—shortens reporting time due to seamless integration and adaptive authoring.
3. Improved decision making—reports and dashboards present data in easily-understood formats.
4. Adaptive authoring automatically adjusts report layout when objects are added, moved, or removed.
5. High performance data access across all sources.

Cons

1. Total Cost of Ownership (TCO) is more significant than other tools
2. Minimal forecast capabilities
3. Investment in Cognos R&D by IBM is declining
4. Won't work smoothly with large data sets having many parameters.
5. Cross-browser compatibility is often problematic

- **Conclusion**

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.

List of sources of empirical data used or considered and the methods applied in revising past estimates of population and components of demographic change (fertility, child, adult and overall mortality, international migration) are presented in tabular form for each demographic component and country or area for the period 1950-2020.

The goal of this project is to find and analyze United Nations population estimates and projections and present the given data in a visual format for better understanding. For the visualization we have used IBM Cognos Analytics. By the end of this Project we have gain a broad understanding of plotting different graphs and able to create meaningful dashboards