

19BCE2647

Aadarshbushan Singh
VIT, Vellore
Data Analytics

Question:

World Population Data Analytics Using IBM Cognos

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.

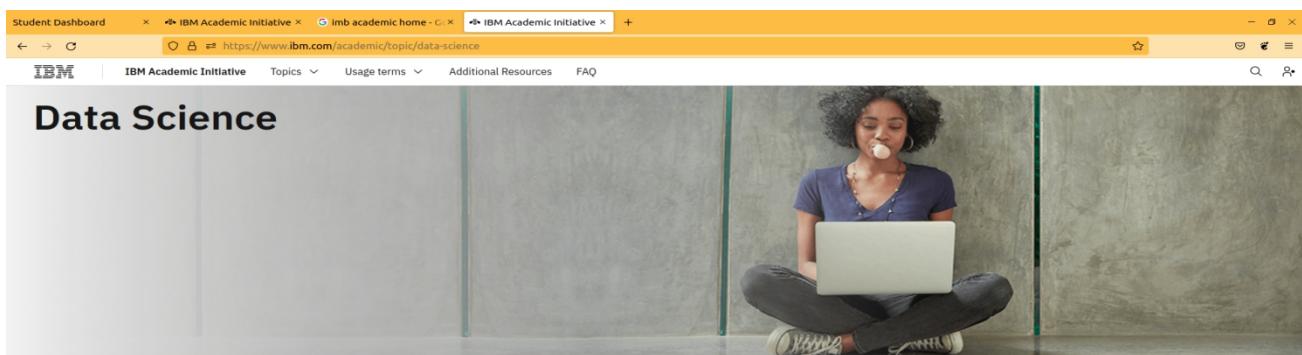
Goal of the case study:

To find and analyze United Nations population estimates and projections and present in a visual format for better understanding.

Dataset link:

<https://population.un.org/wpp/Download/Standard/Population/>

Steps are explained along with the images.



What is Data Science?

Today's businesses are awash in data with more streaming in every day. Simply having the data, however, doesn't guarantee that anything useful can be extracted from it. That's where data science comes in. Data science is the practice of leveraging a unique set of skills and tools to help businesses derive value from their data.

Fig1: IBM Cognos Landing Page

A screenshot of a web browser showing the IBM Cognos Analytics landing page. The URL in the address bar is https://www.ibm.com/academic/topic/data-science. On the left, a sidebar menu lists 'Courseware', 'Software', and 'Resources'. The main content area displays three software offerings: 'ILOG CPLEX Optimization Studio', 'SPSS Modeler Premium', and 'Cognos Analytics on Cloud'. Each item has a brief description and a blue 'View All' button. Below this, there is another section for 'Cognos Analytics' with a similar layout.

Fig: Cognos Analytics

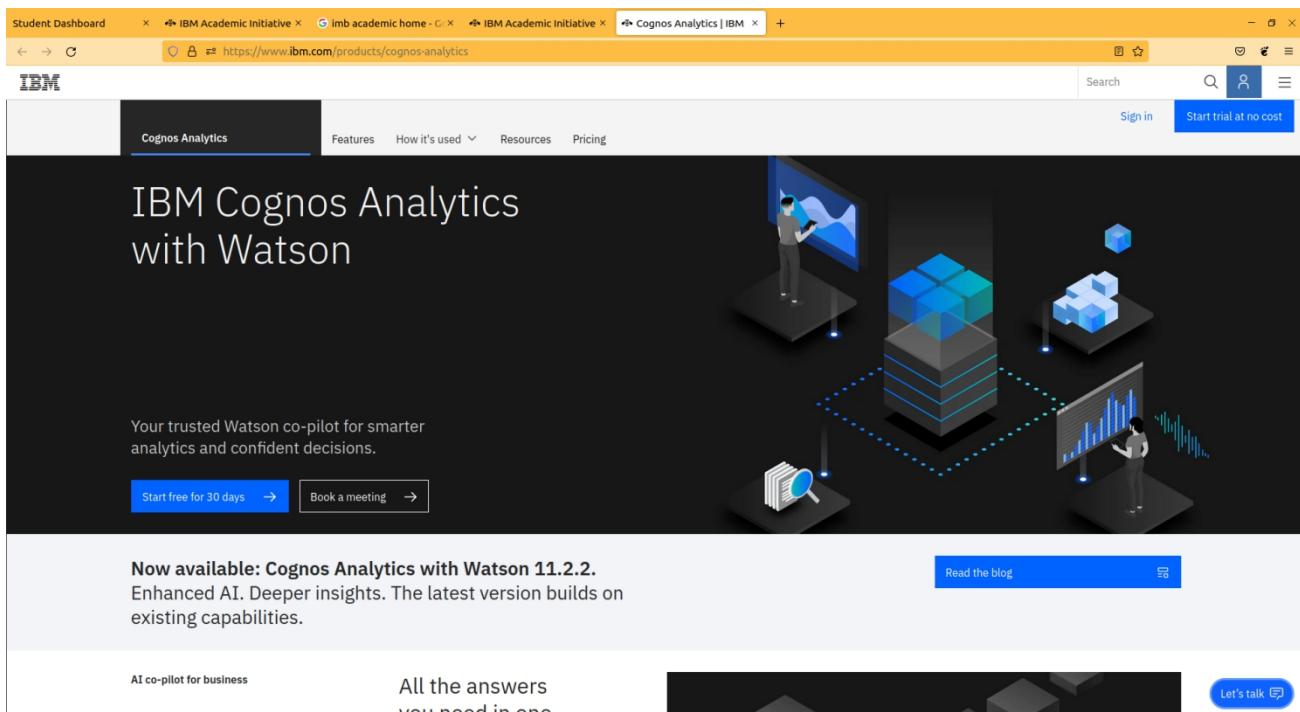


Fig: Cognos Analytics Lab

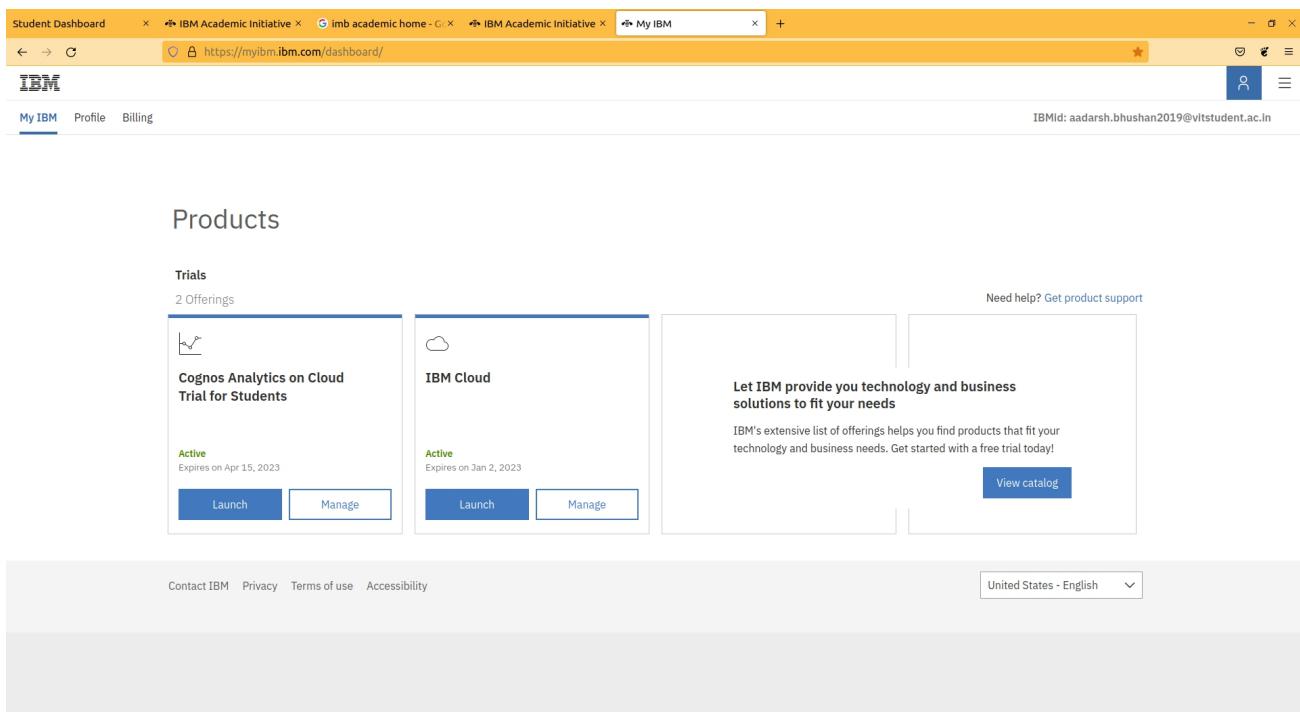


Fig: Launch Cognos Analytics Cloud

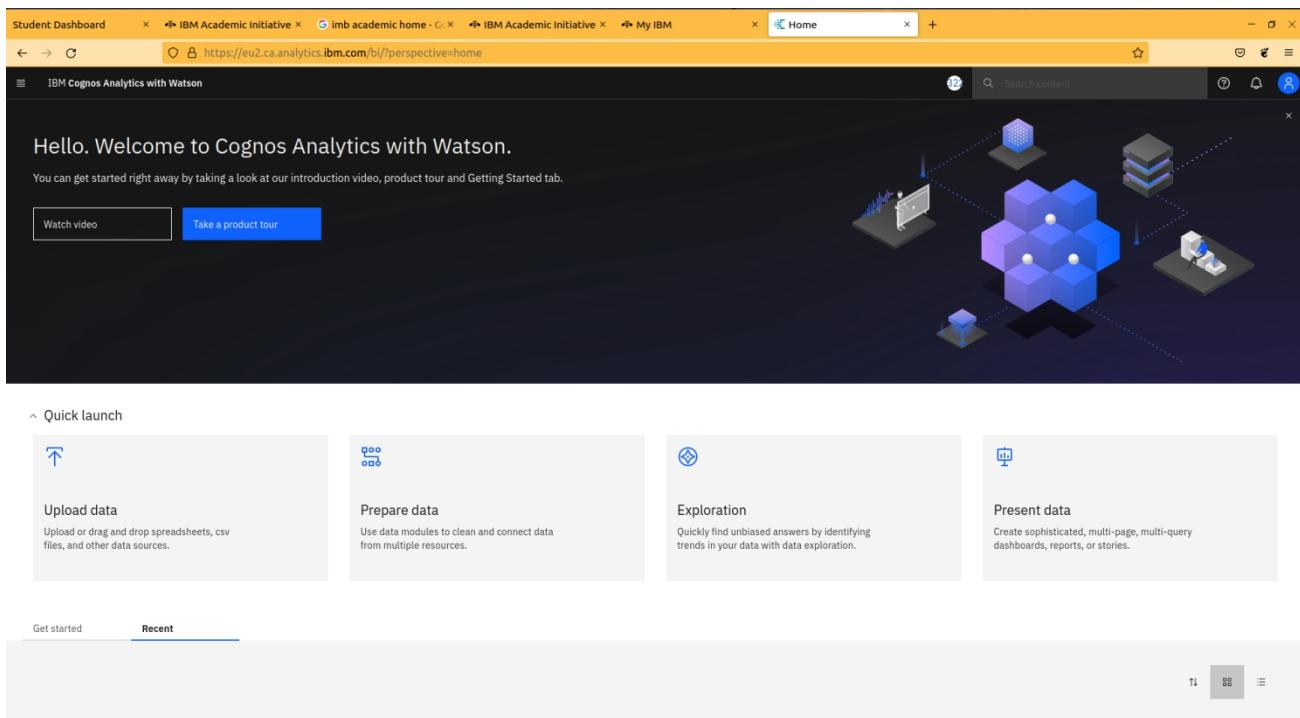


Fig: Cognos Analytics Landing Page

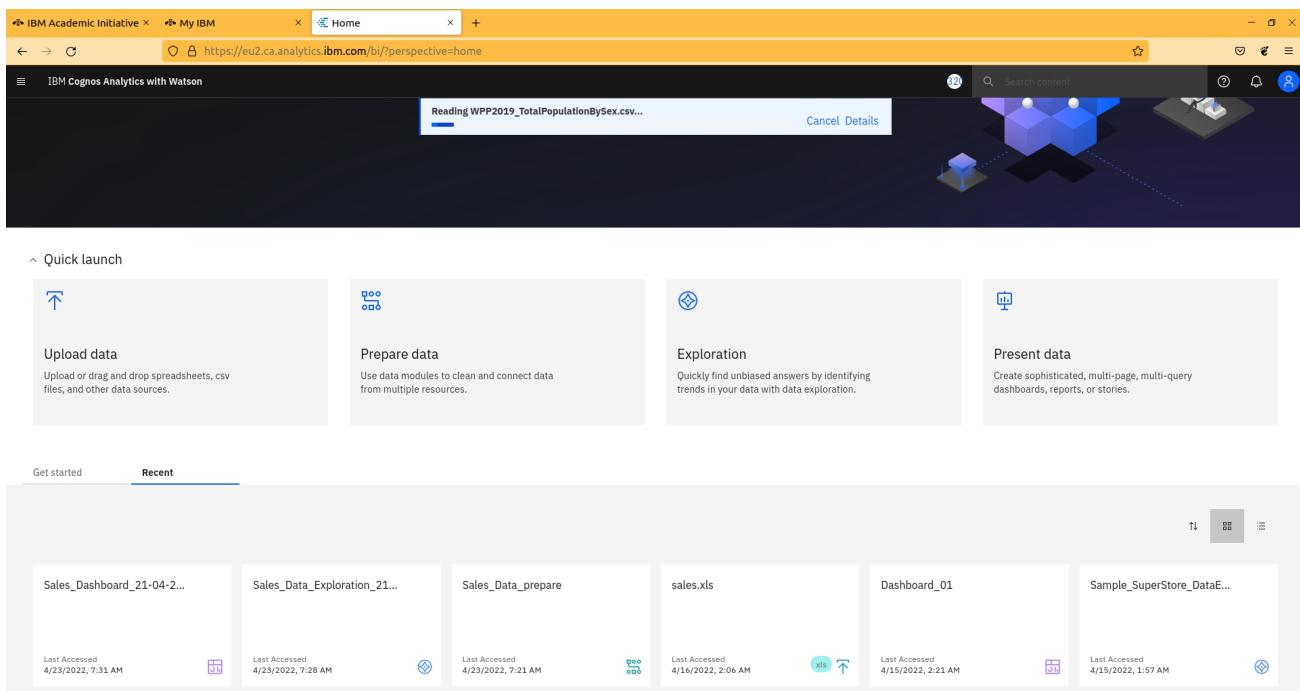


Fig: Uploading the dataset

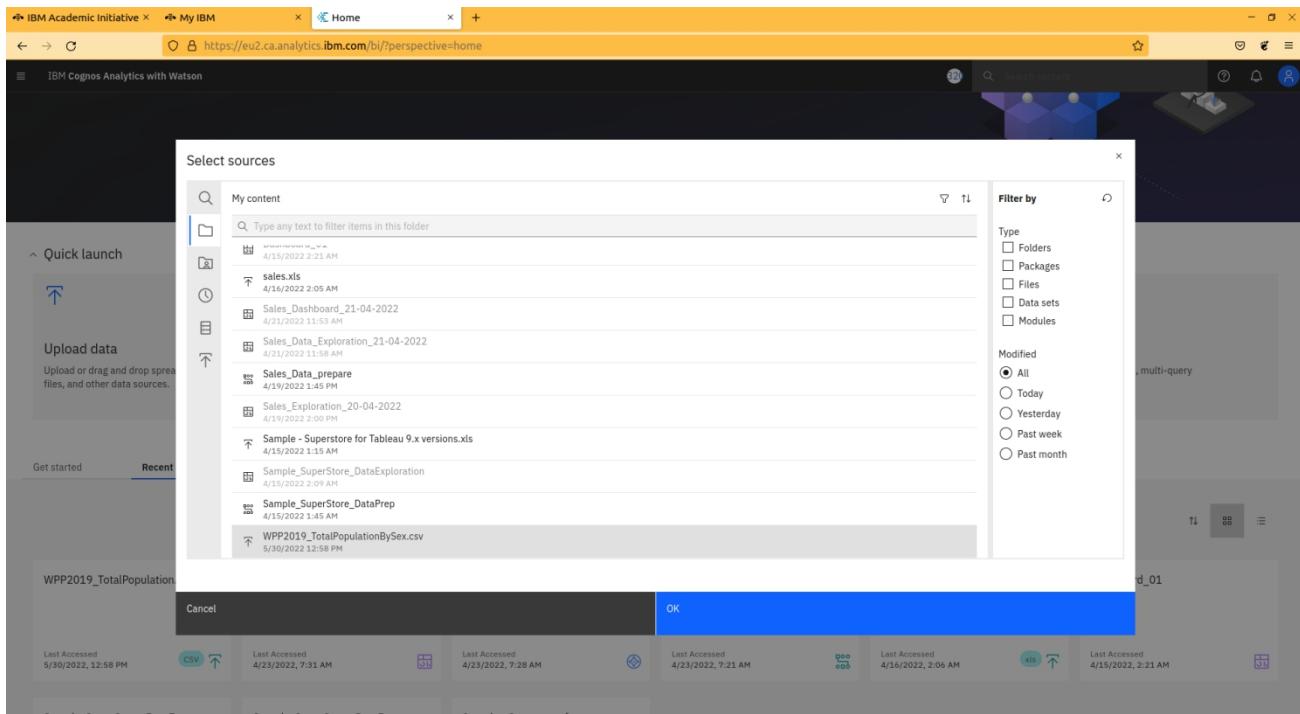


Fig: Selecting the dataset for data preparation

LocID	Location	VarID	Variant	Time	MidPeriod	PopMale	PopFemale	PopTotal
4	Afghanistan	2	Medium	1950	1950.5	4099.243	3652.874	7752.117
4	Afghanistan	2	Medium	1951	1951.5	4134.756	3705.395	7840.151
4	Afghanistan	2	Medium	1952	1952.5	4174.45	3761.546	7935.996
4	Afghanistan	2	Medium	1953	1953.5	4218.336	3821.348	8039.684
4	Afghanistan	2	Medium	1954	1954.5	4266.484	3884.832	8151.316
4	Afghanistan	2	Medium	1955	1955.5	4318.945	3952.047	8270.992
4	Afghanistan	2	Medium	1956	1956.5	4375.8	4023.073	8398.873
Afghanistan		2	Medium	1957	1957.5	4437.157	4098	8535.157
Afghanistan		2	Medium	1958	1958.5	4503.156	4176.941	8680.097
Afghanistan		2	Medium	1959	1959.5	4573.914	4260.033	8833.947
Afghanistan		2	Medium	1960	1960.5	4649.573	4347.394	8996.967
Afghanistan		2	Medium	1961	1961.5	4730.25	4439.156	9169.406
Afghanistan		2	Medium	1962	1962.5	4816.05	4535.392	9351.442
Afghanistan		2	Medium	1963	1963.5	4907.03	4636.17	9543.2
Afghanistan		2	Medium	1964	1964.5	5003.245	4741.527	9744.772
Afghanistan		2	Medium	1965	1965.5	5104.765	4851.553	9956.318
Afghanistan		2	Medium	1966	1966.5	5210.122	4964.718	10174.84
Afghanistan		2	Medium	1967	1967.5	5319.123	5080.813	10399.936
Afghanistan		2	Medium	1968	1968.5	5434.458	5202.606	10637.064

Fig: Removing Pop Total

The screenshot shows the IBM Cognos Analytics with Watson interface. A context menu is open over a data row in a grid. The menu items include:

- New
- Calculation... (highlighted with a blue box)
- Filter...
- Folder...
- Table...
- Show query information...
- Specify column dependencies
- Refresh members
- Hide from users
- Remove
- Refresh properties...
- Sort...
- Rename
- Cut
- Copy
- Manage filters
- Properties

The main grid displays data for Afghanistan from 1950 to 1968, with columns for Row Id, LocID, Location, VarID, Variant, Time, MidPeriod, PopMale, and PopFemale.

Fig: Creating New Calculation for Total Population

The screenshot shows the 'Create calculation' dialog in IBM Cognos Analytics. The 'Components' pane on the left lists the available fields: Row Id, LocID, Location, VarID, Variant, Time, MidPeriod, PopMale, PopFemale, and PopDensity. The 'Expression' pane on the right contains the expression `PopMale + PopFemale`. The 'Validation Results' pane at the bottom indicates that the expression is valid. At the bottom of the dialog, there is a checkbox labeled 'Calculate after aggregation' and two buttons: 'Cancel' and 'OK'.

Fig: Creating new calculation

IBM Academic Initiative > My IBM > * New data module > https://eu2.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=655199587_864098878e6a4f1d8de5901b690eb896_sessionTemp&objRef=&tid=655199587_864098878e6a4f1d8de5901b690eb896

IBM Cognos Analytics with Watson > * New data module > Properties

Data module: WPP2019_T...BySex.csv

Grid View:

TotalPop	Row Id	LocID	Location	VarID	Variant	Time	MidPeriod	PopMale
7752.117	1	4	Afghanistan	2	Medium	1950	1950.5	4099.243
7840.151	2	4	Afghanistan	2	Medium	1951	1951.5	4134.756
7935.995999999999	3	4	Afghanistan	2	Medium	1952	1952.5	4174.45
8039.684	4	4	Afghanistan	2	Medium	1953	1953.5	4218.336
8151.31600000001	5	4	Afghanistan	2	Medium	1954	1954.5	4266.484
8270.992	6	4	Afghanistan	2	Medium	1955	1955.5	4318.945
8398.873	7	4	Afghanistan	2	Medium	1956	1956.5	4375.8
8535.157	8	4	Afghanistan	2	Medium	1957	1957.5	4437.157
8680.097	9	4	Afghanistan	2	Medium	1958	1958.5	4503.156
8833.947	10	4	Afghanistan	2	Medium	1959	1959.5	4573.914
8996.967	11	4	Afghanistan	2	Medium	1960	1960.5	4649.573
9169.405999999999	12	4	Afghanistan	2	Medium	1961	1961.5	4730.25
9351.442	13	4	Afghanistan	2	Medium	1962	1962.5	4816.05
9543.2	14	4	Afghanistan	2	Medium	1963	1963.5	4907.03
9744.772	15	4	Afghanistan	2	Medium	1964	1964.5	5003.245
9956.318	16	4	Afghanistan	2	Medium	1965	1965.5	5104.765
10174.84	17	4	Afghanistan	2	Medium	1966	1966.5	5210.122
10399.936	18	4	Afghanistan	2	Medium	1967	1967.5	5319.123
10637.063999999998	19	4	Afghanistan	2	Medium	1968	1968.5	5434.458
10802.225	20	4	Afghanistan	2	Medium	1969	1969.5	5550.094

Fig: Total Population Created

IBM Academic Initiative > My IBM > * New data module > https://eu2.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=655199587_864098878e6a4f1d8de5901b690eb896_sessionTemp&objRef=&tid=655199587_864098878e6a4f1d8de5901b690eb896

IBM Cognos Analytics with Watson > * New data module > Properties

Data module: WPP2019_T...BySex.csv

Grid View:

TotalPop	Row Id	LocID	Location	VarID	Variant	Time	MidPeriod	PopMale
7752.117	1	4	Afghanistan	2	Medium	1950	1950.5	4099.243
7840.151	2	4	Afghanistan	2	Medium	1951	1951.5	4134.756
7935.995999999999	3	4	Afghanistan	2	Medium	1952	1952.5	4174.45
8039.684	4	4	Afghanistan	2	Medium	1953	1953.5	4218.336
8151.31600000001	5	4	Afghanistan	2	Medium	1954	1954.5	4266.484
8270.992	6	4	Afghanistan	2	Medium	1955	1955.5	4318.945
8398.873	7	4	Afghanistan	2	Medium	1956	1956.5	4375.8
8535.157	8	4	Afghanistan	2	Medium	1957	1957.5	4437.157
8680.097	9	4	Afghanistan	2	Medium	1958	1958.5	4503.156
8833.947	10	4	Afghanistan	2	Medium	1959	1959.5	4573.914
8996.967	11	4	Afghanistan	2	Medium	1960	1960.5	4649.573
9169.405999999999	12	4	Afghanistan	2	Medium	1961	1961.5	4730.25
9351.442	13	4	Afghanistan	2	Medium	1962	1962.5	4816.05
9543.2	14	4	Afghanistan	2	Medium	1963	1963.5	4907.03
9744.772	15	4	Afghanistan	2	Medium	1964	1964.5	5003.245
9956.318	16	4	Afghanistan	2	Medium	1965	1965.5	5104.765
10174.84	17	4	Afghanistan	2	Medium	1966	1966.5	5210.122
10399.936	18	4	Afghanistan	2	Medium	1967	1967.5	5319.123
10637.063999999998	19	4	Afghanistan	2	Medium	1968	1968.5	5434.458
10802.225	20	4	Afghanistan	2	Medium	1969	1969.5	5550.094

Right-click context menu for Row 7 (8039.684):

- Filter...
- Create calculation...
- Create data group...
- Hide from users
- Remove
- Refresh properties...
- Format data...
- Clean...
- Rename
- Cut
- Copy
- Paste
- Properties

Fig: Formatting data

Time	MidPeriod	PopMale
1950	1950.5	4099.243
1951	1951.5	4134.756
1952	1952.5	4174.45
1953	1953.5	4218.336
1954	1954.5	4266.484
1955	1955.5	4318.945
1956	1956.5	4375.8
1957	1957.5	4437.157
1958	1958.5	4503.156
1959	1959.5	4573.914
1960	1960.5	4649.573
1961	1961.5	4730.25
1962	1962.5	4816.05
1963	1963.5	4907.03
1964	1964.5	5003.245
1965	1965.5	5104.765
1966	1966.5	5210.122
1967	1967.5	5319.123
1968	1968.5	5434.458
1969	1969.5	5550.094

Fig: Data formatting for numbers

World_Data_Formatted	WPP2019_TotalPopulation...	Sales_Dashboard_21-04-2...	Sales_Data_Exploration_21...	Sales_Data_prepare	sales.xls
Last Accessed 5/30/2022, 1:20 PM	Last Accessed 5/30/2022, 12:58 PM	Last Accessed 4/23/2022, 7:31 AM	Last Accessed 4/23/2022, 7:28 AM	Last Accessed 4/23/2022, 7:21 AM	Last Accessed 4/16/2022, 2:06 AM
Dashboard_01	Sample_SuperStore_DataE...	Sample_SuperStore_DataP...	Sample - Superstore for Tableau 9.x versions.xls		xls
Last Accessed 4/15/2022, 2:21 AM	Last Accessed 4/15/2022, 1:57 AM	Last Accessed 4/15/2022, 1:45 AM	Last Accessed 4/15/2022, 1:15 AM		xls

Fig: Save the prepared data

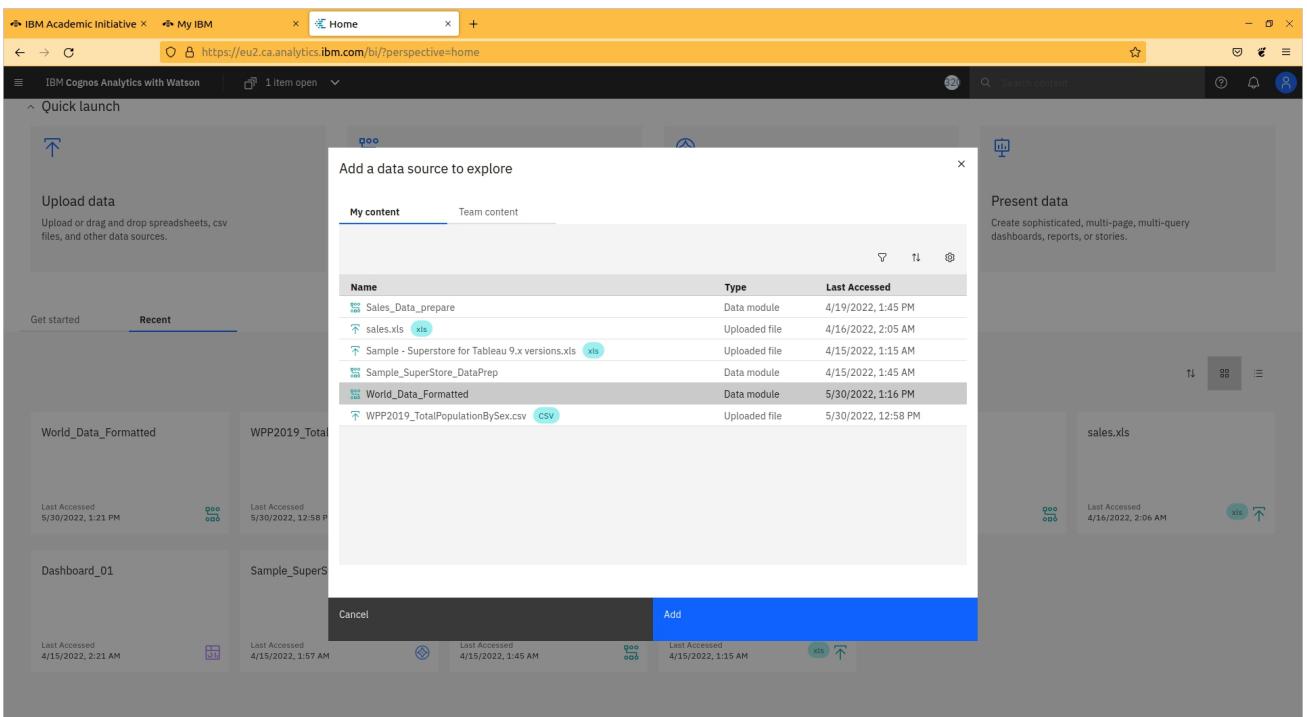


Fig: Importing formatted data for data exploration

Create a visualization part

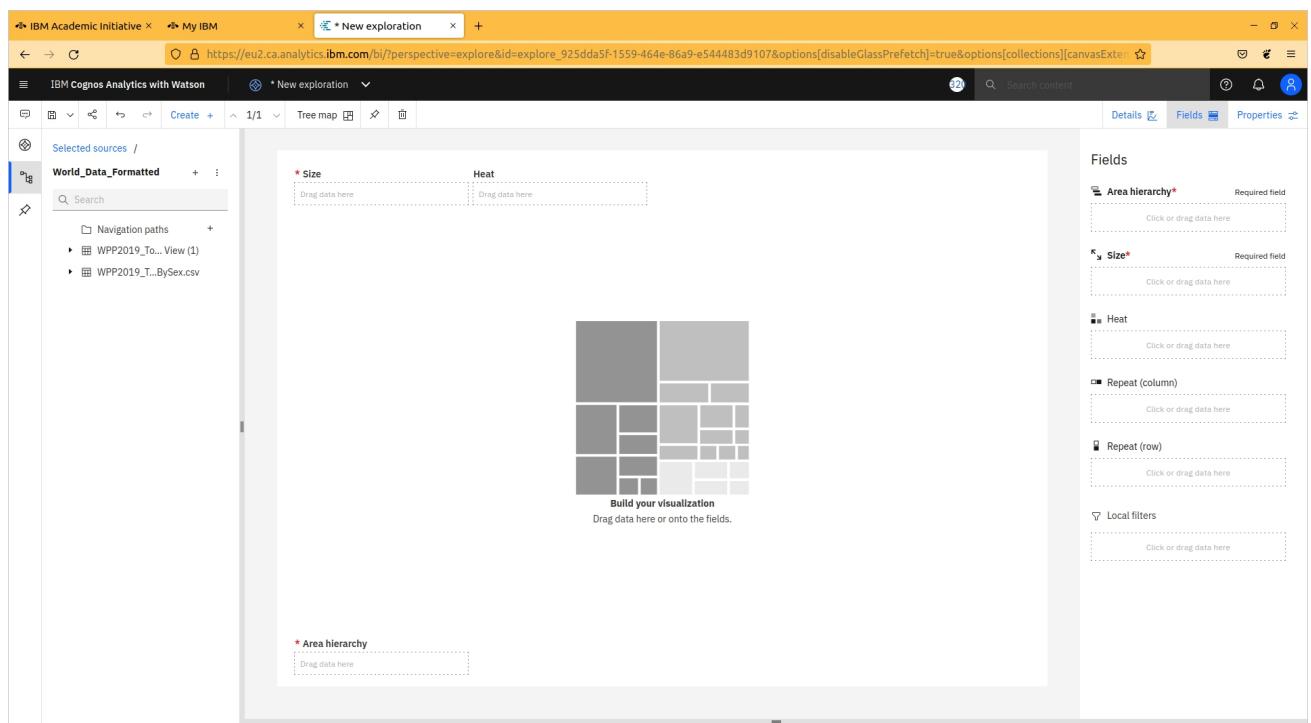


Fig: Select the tree map

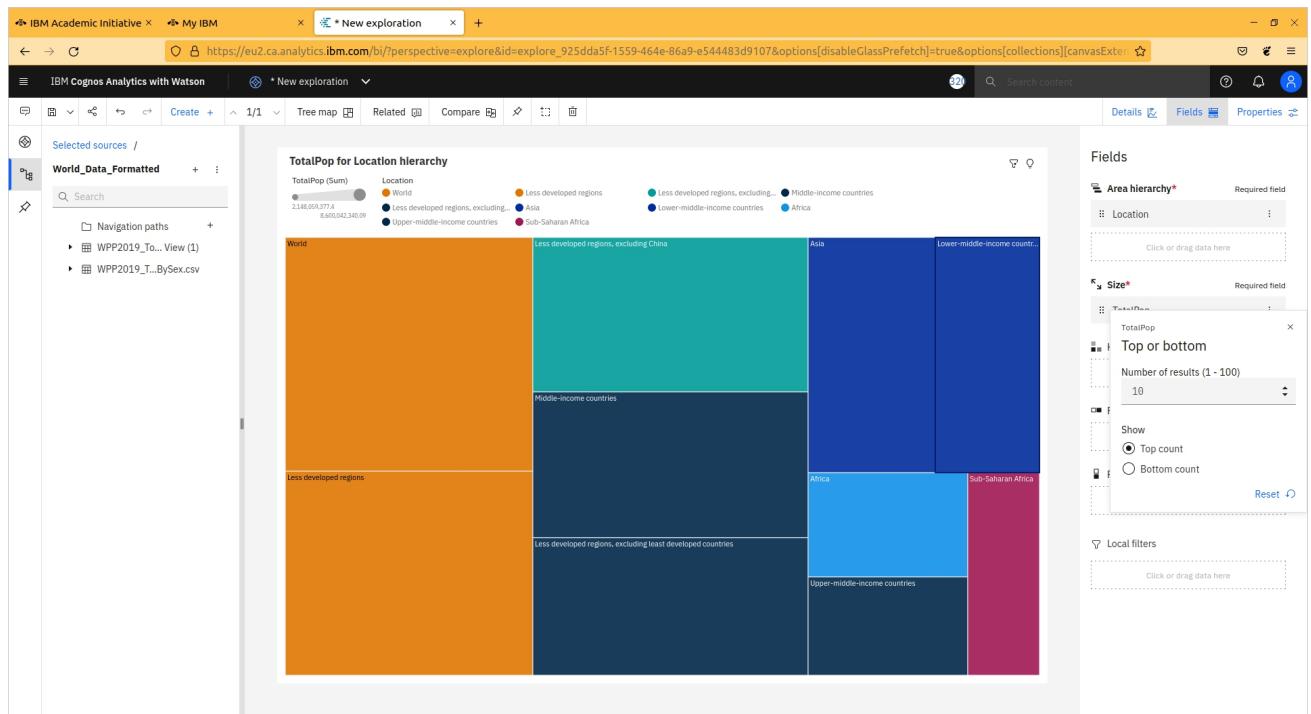


Fig: Selecting top 10 countries and showing it through tree map

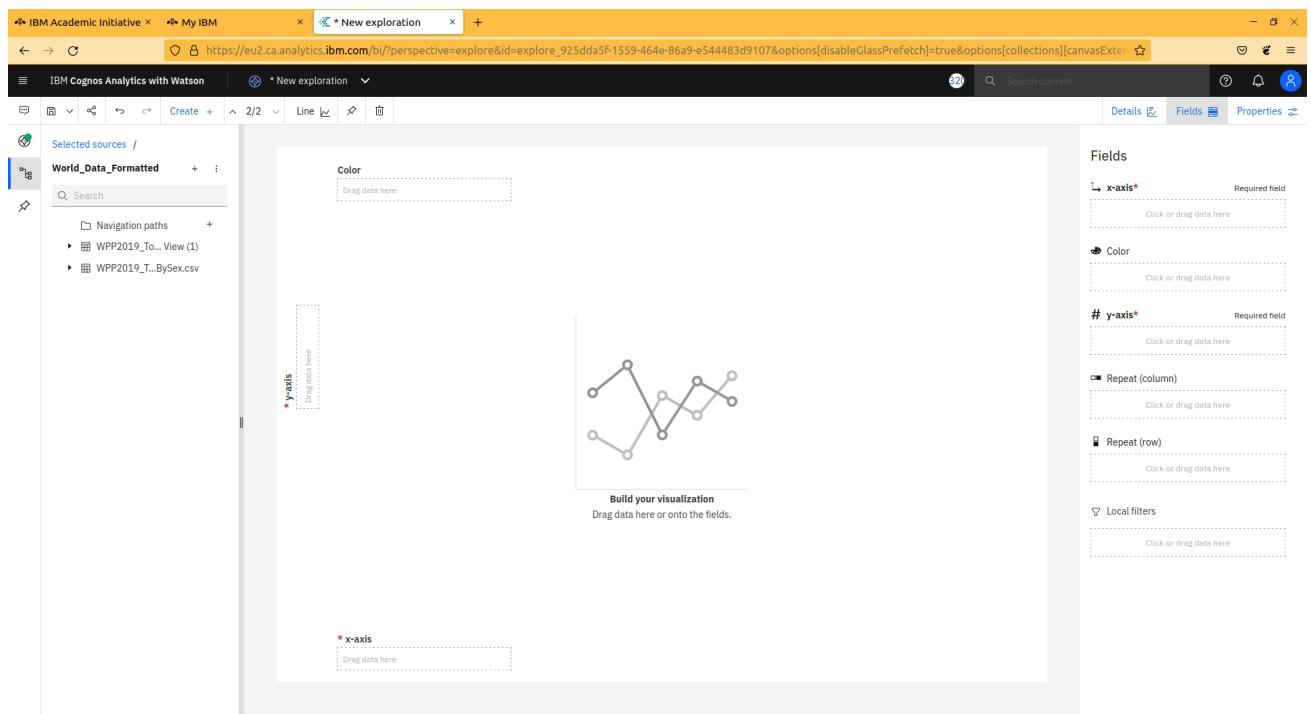


Fig: Selecting line chart

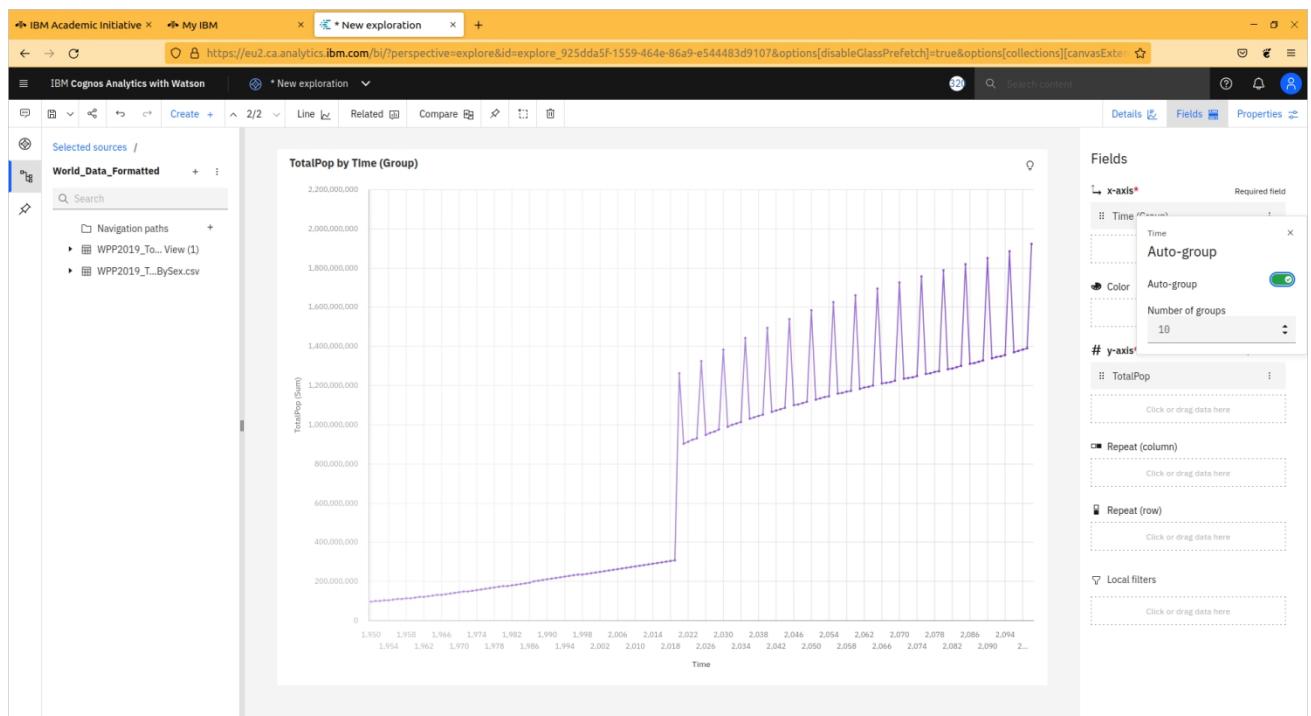


Fig: Selecting X-axis and Y-axis and doing auto group

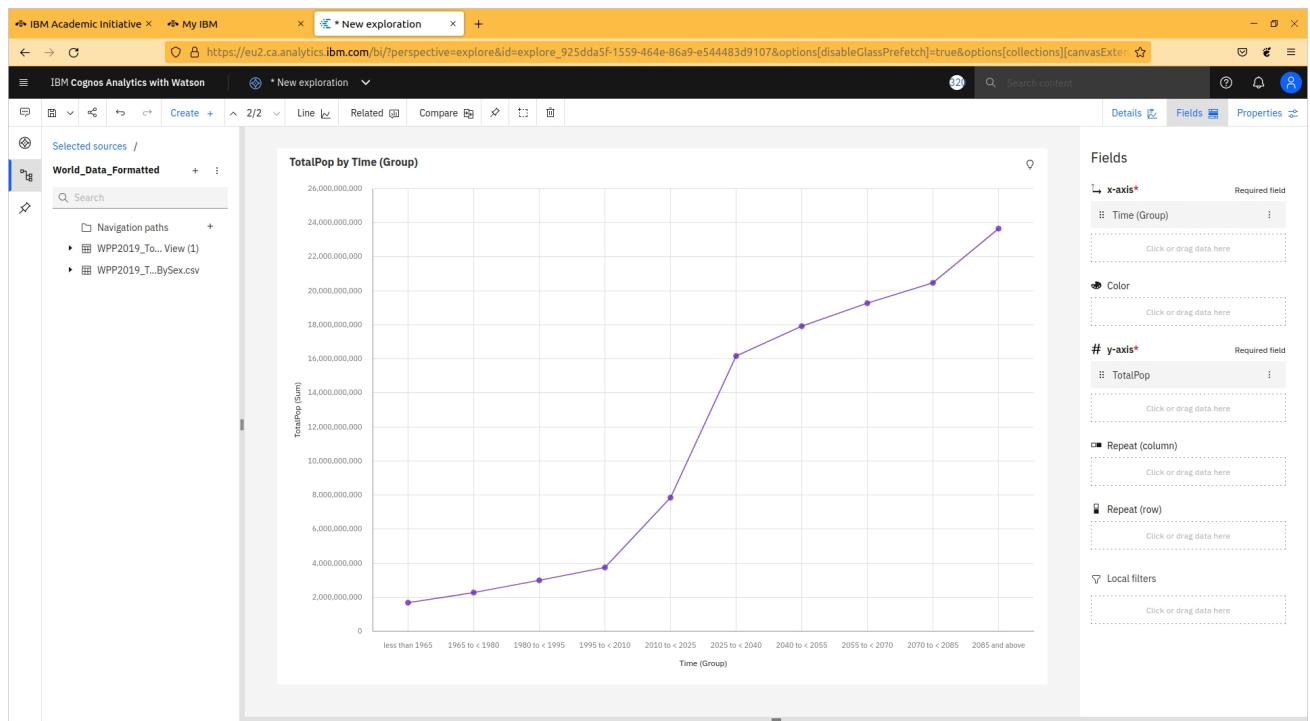


Fig: After doing auto-group

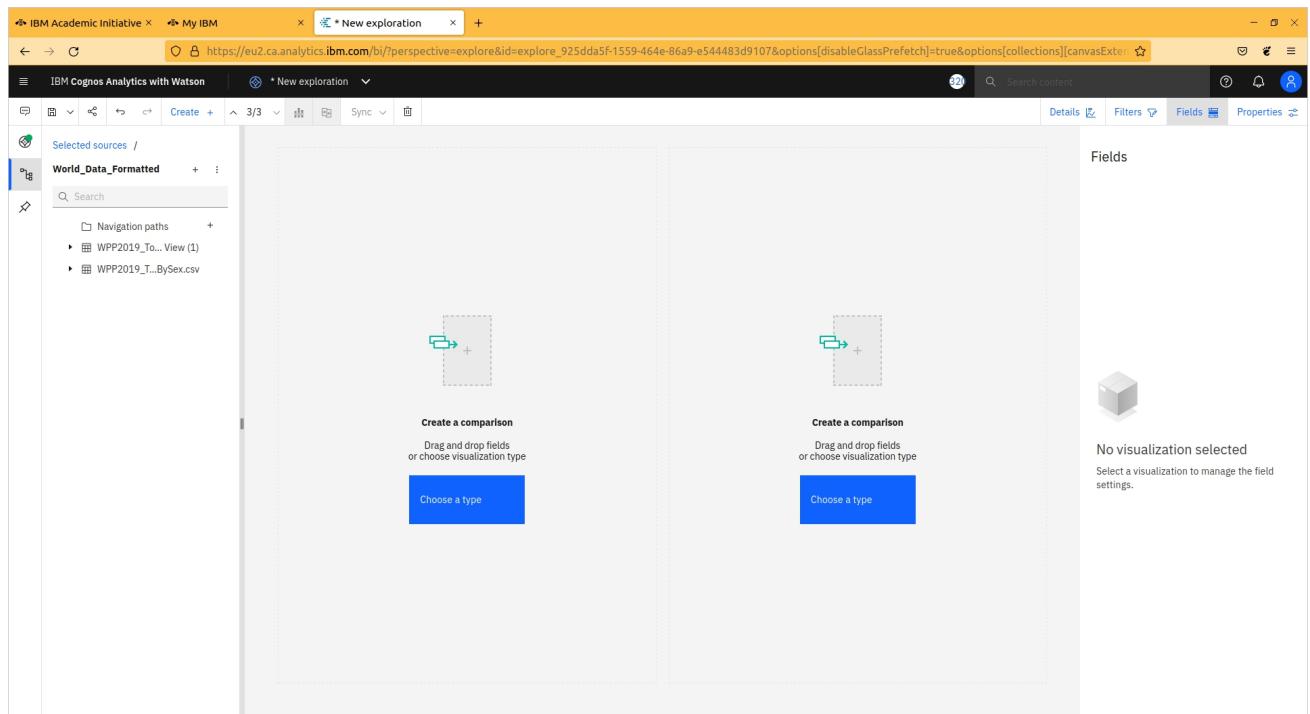


Fig: Selecting dual visualization board

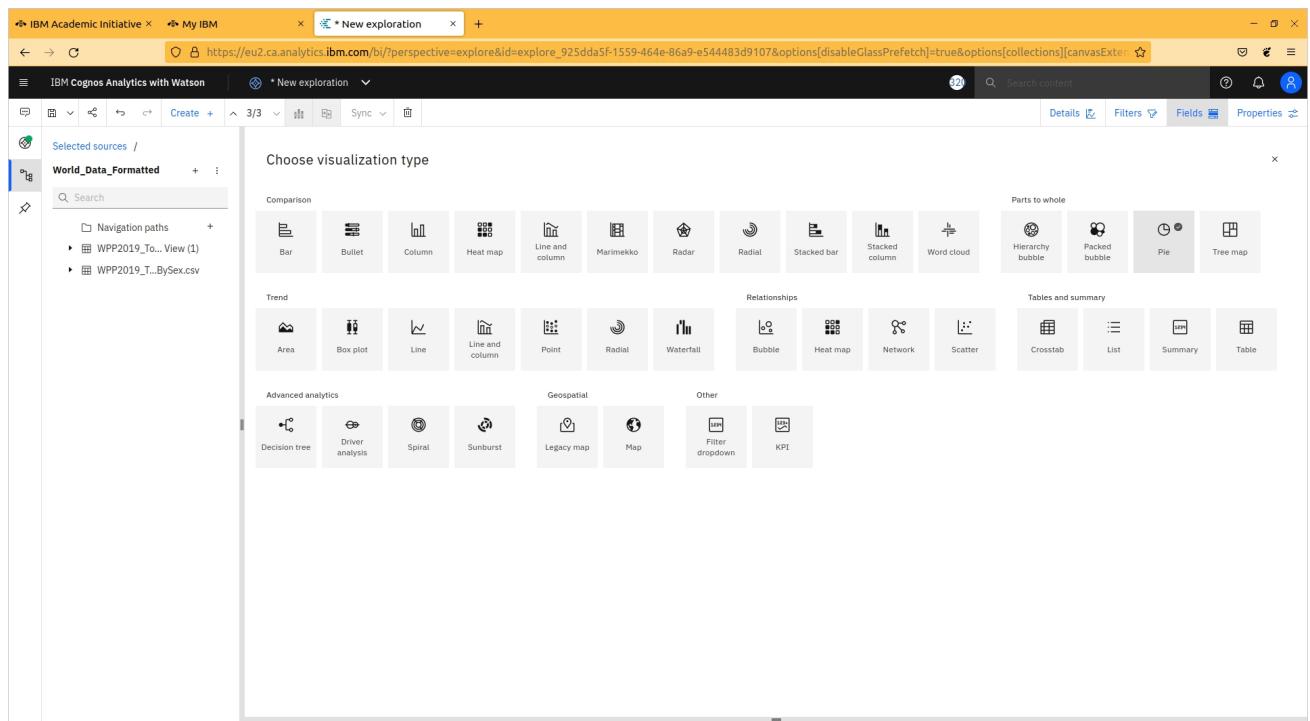


Fig: Selecting pie-chart

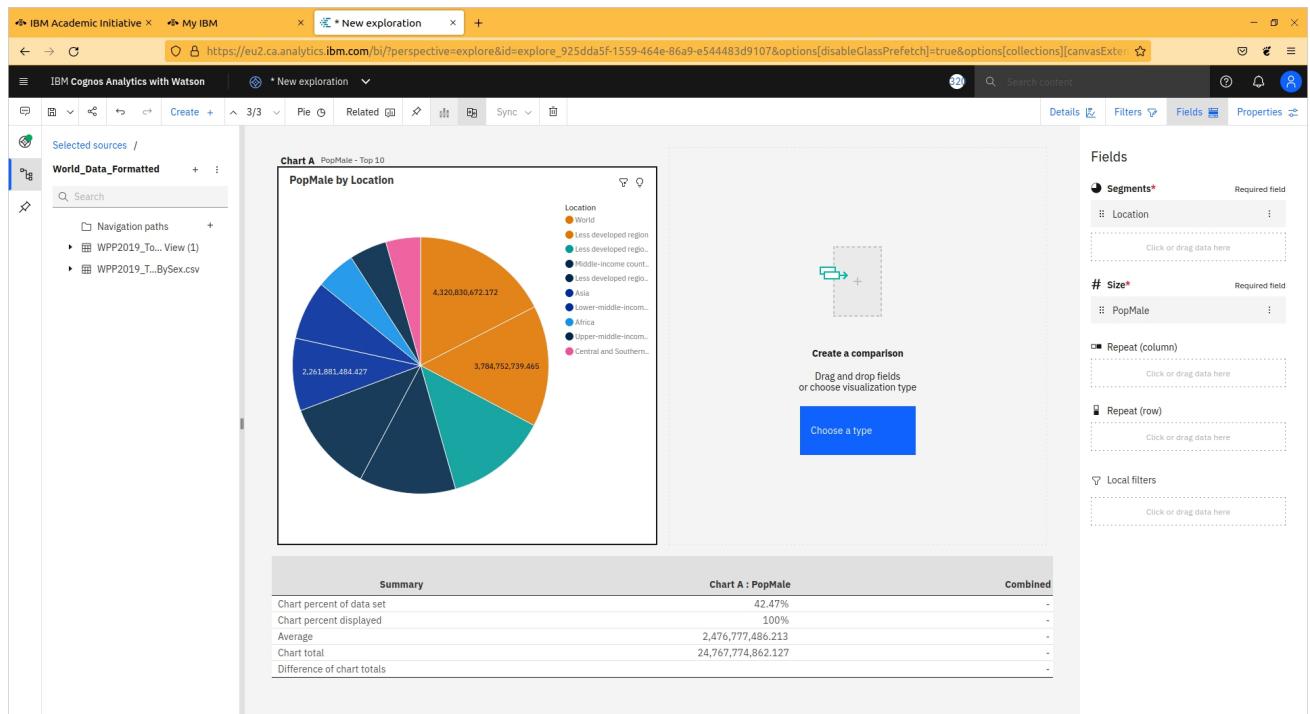


Fig: Doing pop male part

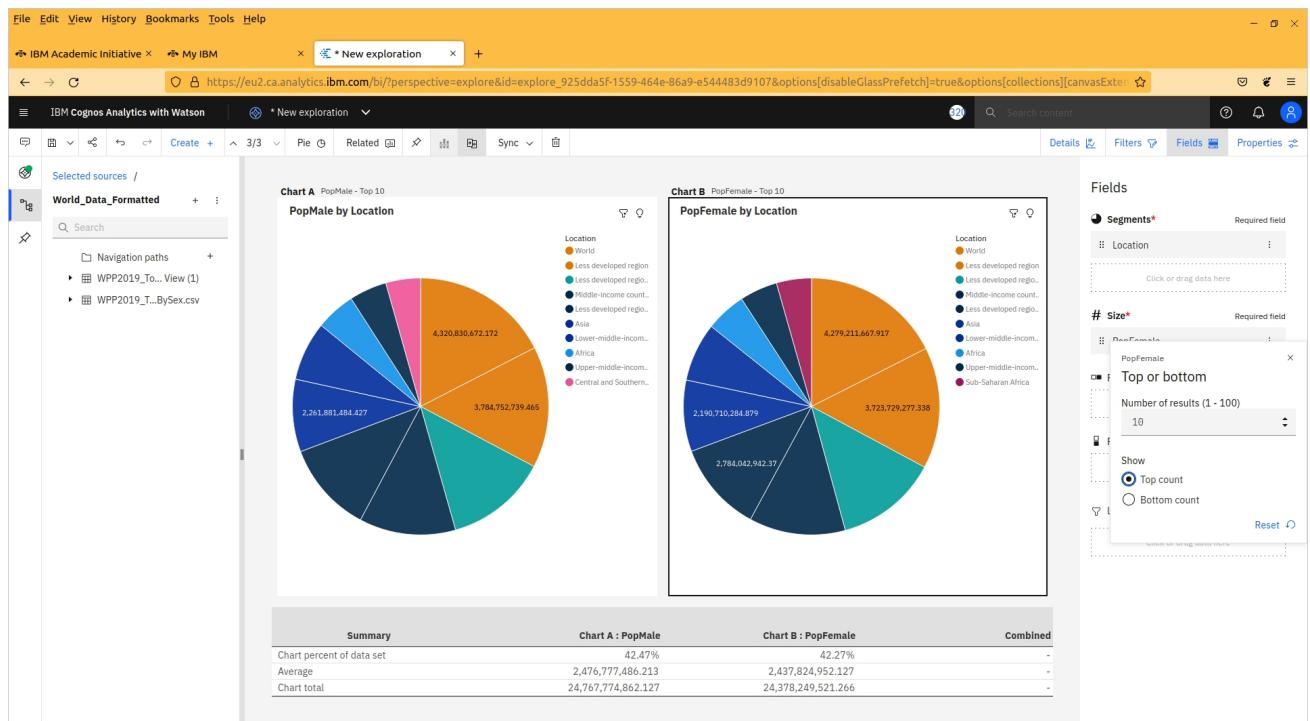


Fig: Selecting top of female part

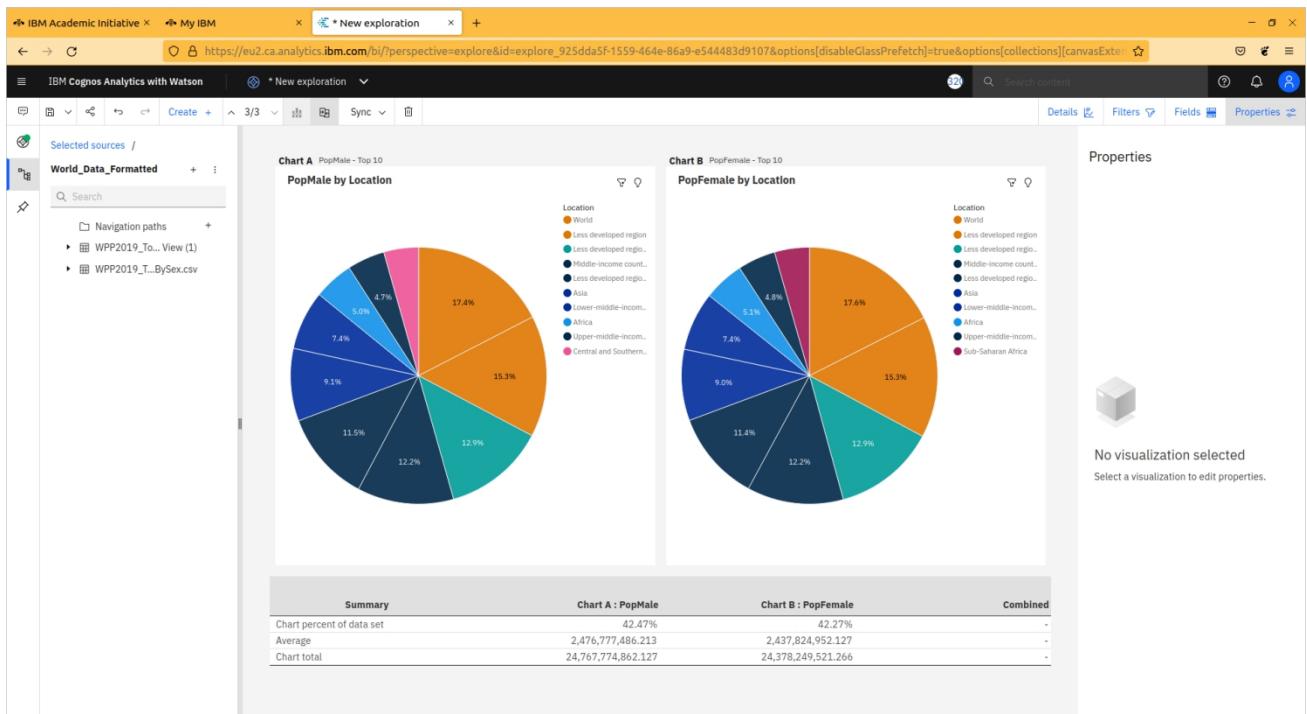


Fig: Converting both into percentage and showing

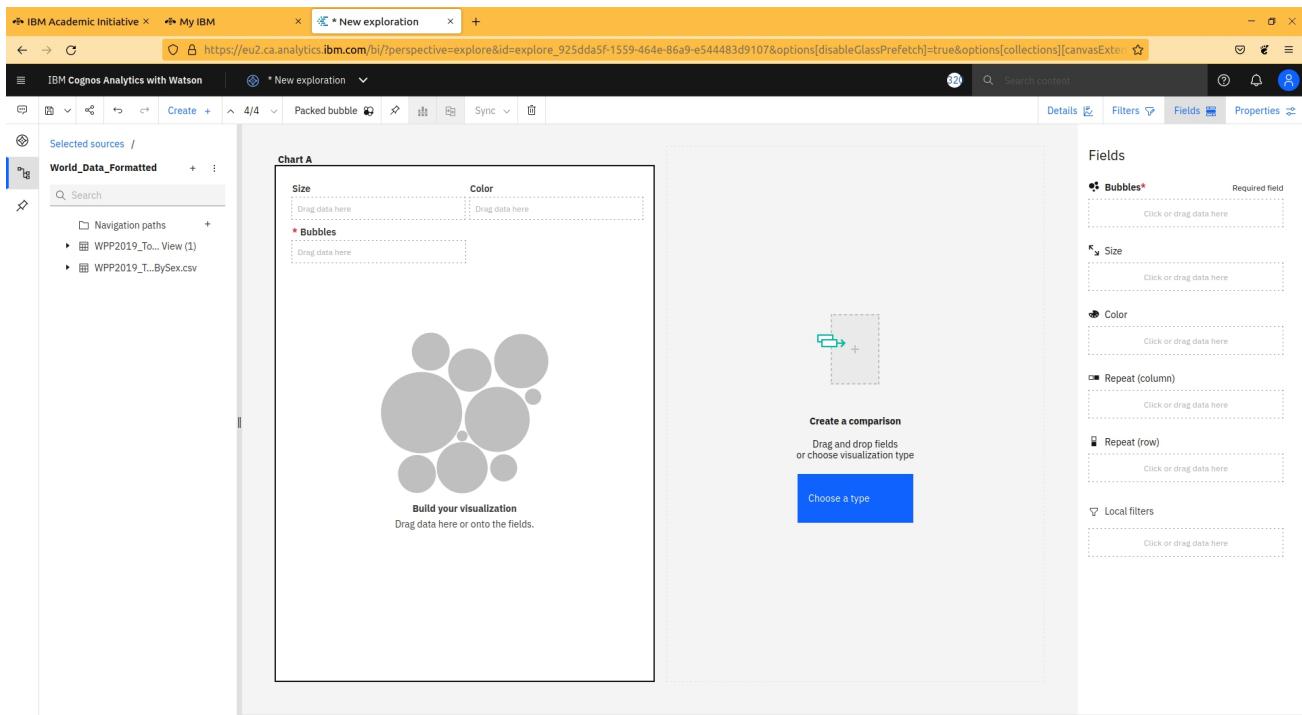


Fig: Selecting packet bubble

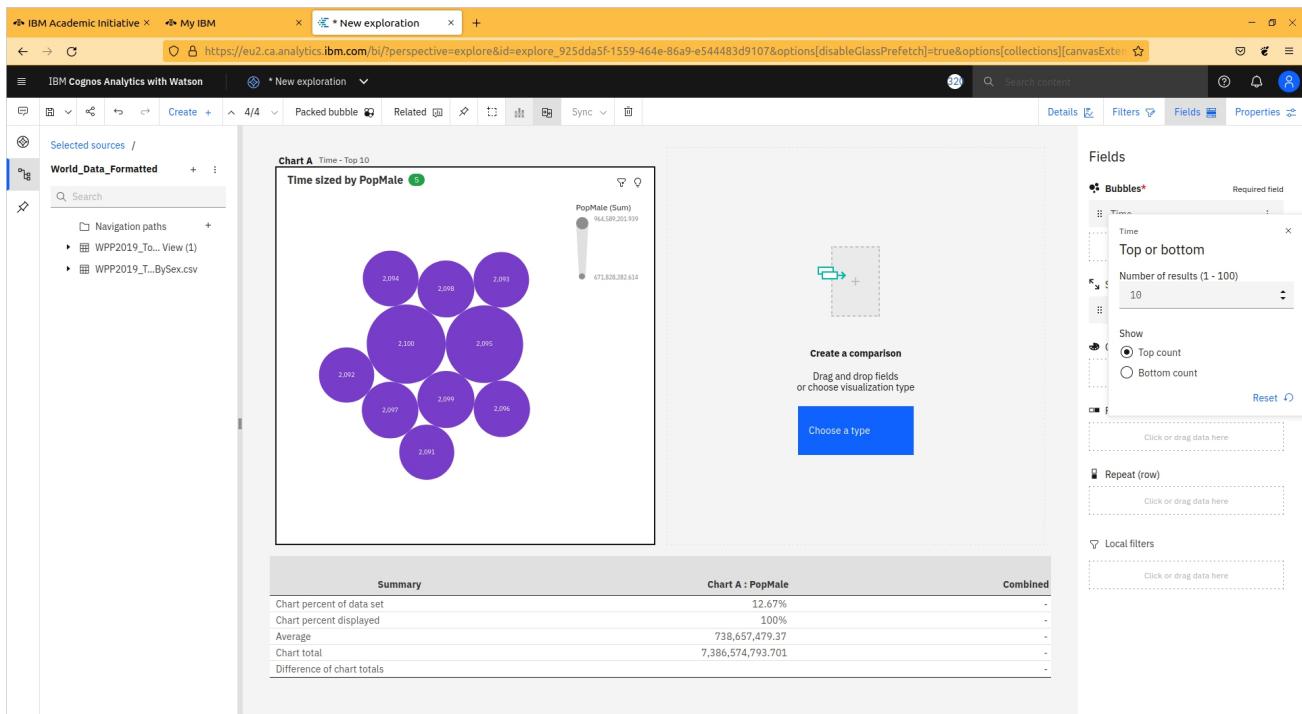


Fig: Selecting top count for male with count

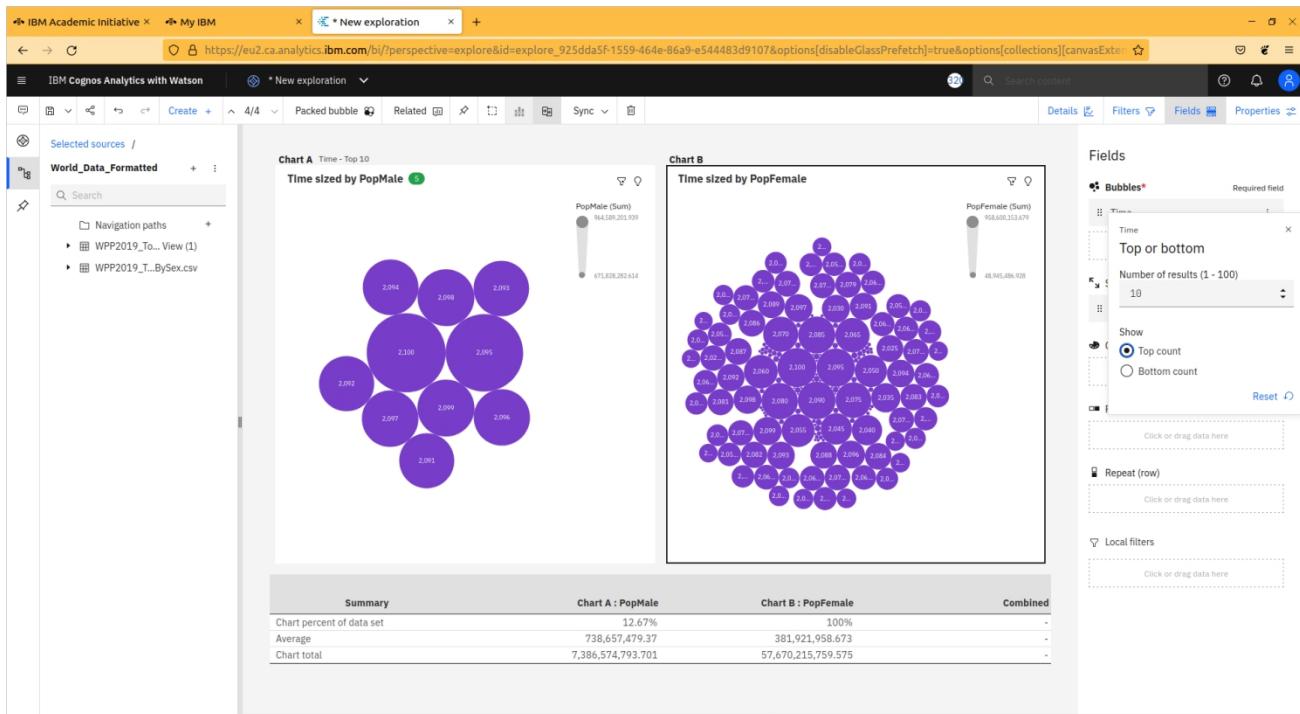


Fig Selecting female with time for bubble

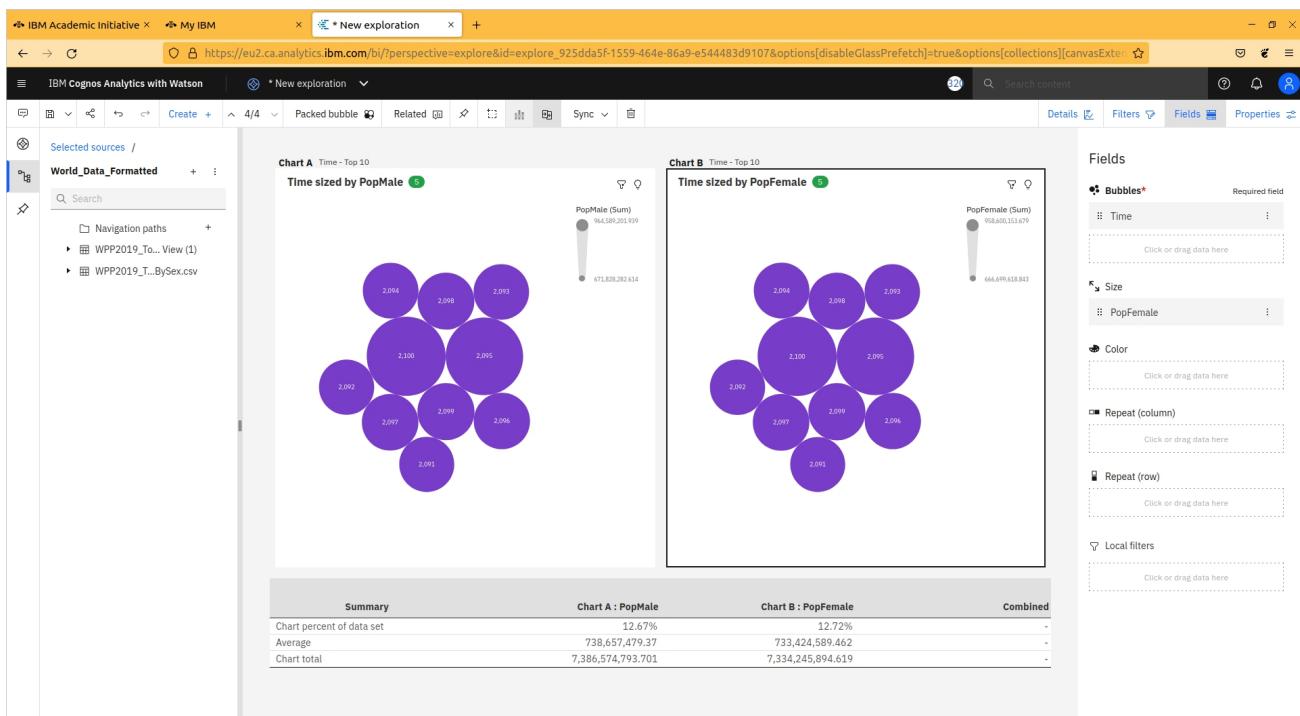


Fig: Selecting female top 10 for bubble

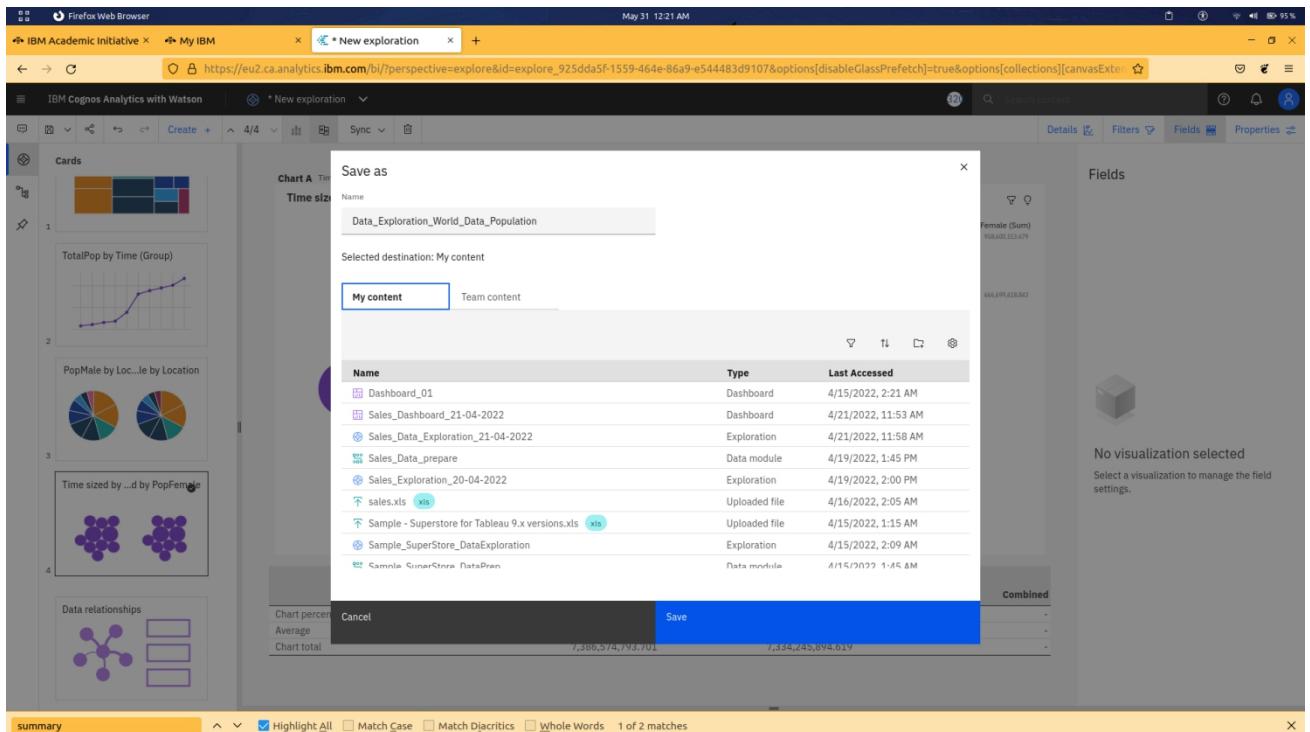


Fig: Saving file for data exploration

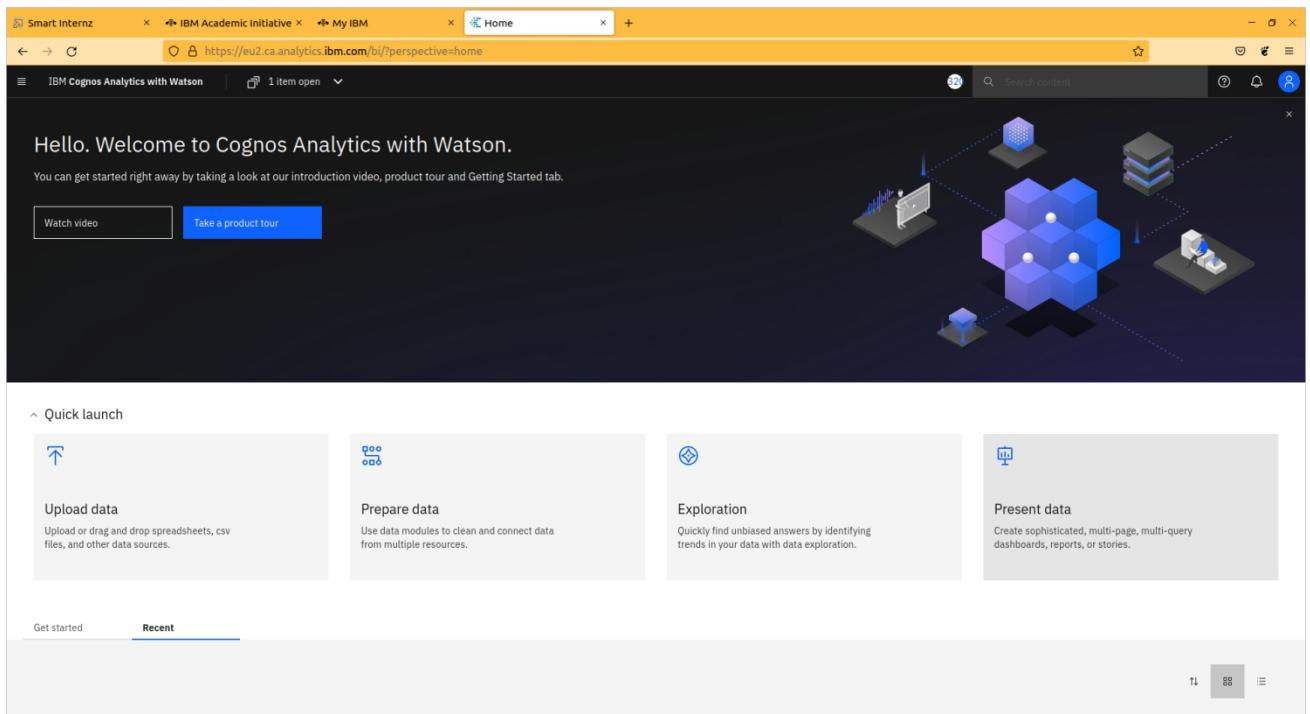


Fig: Going to dashboard for presenting data

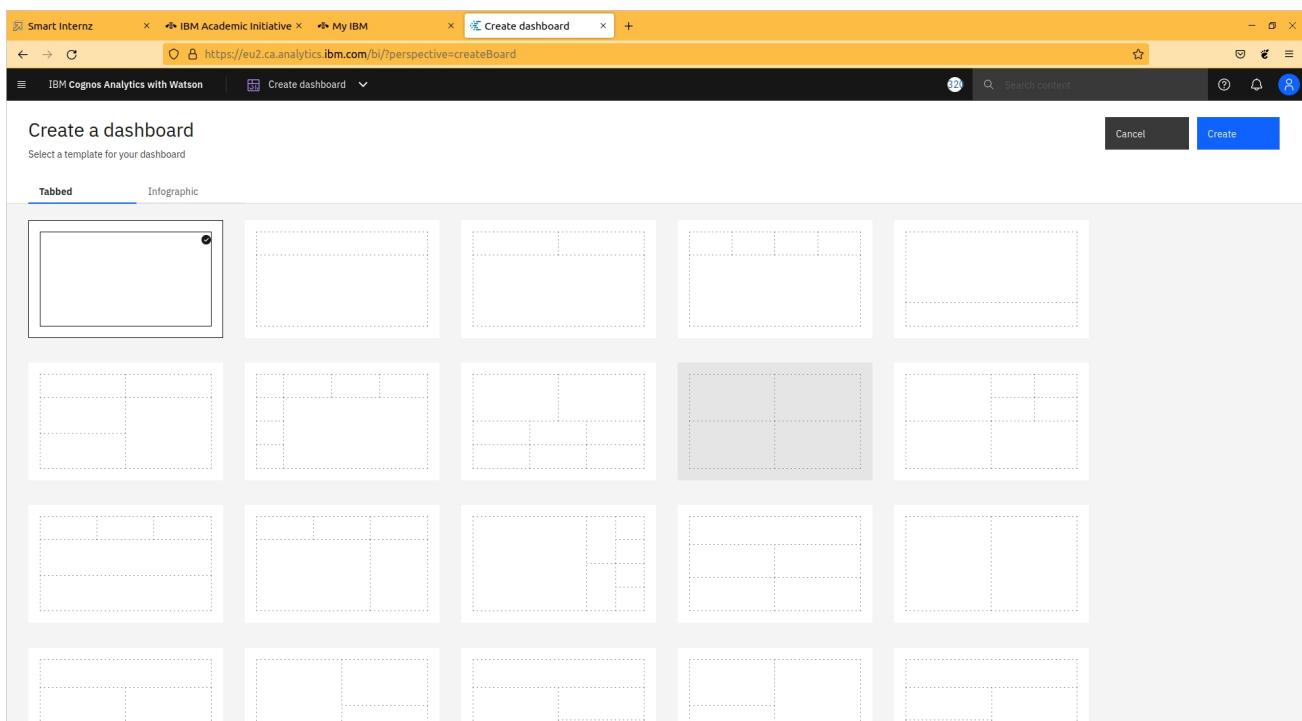


Fig: Selecting dashboard for presenting data

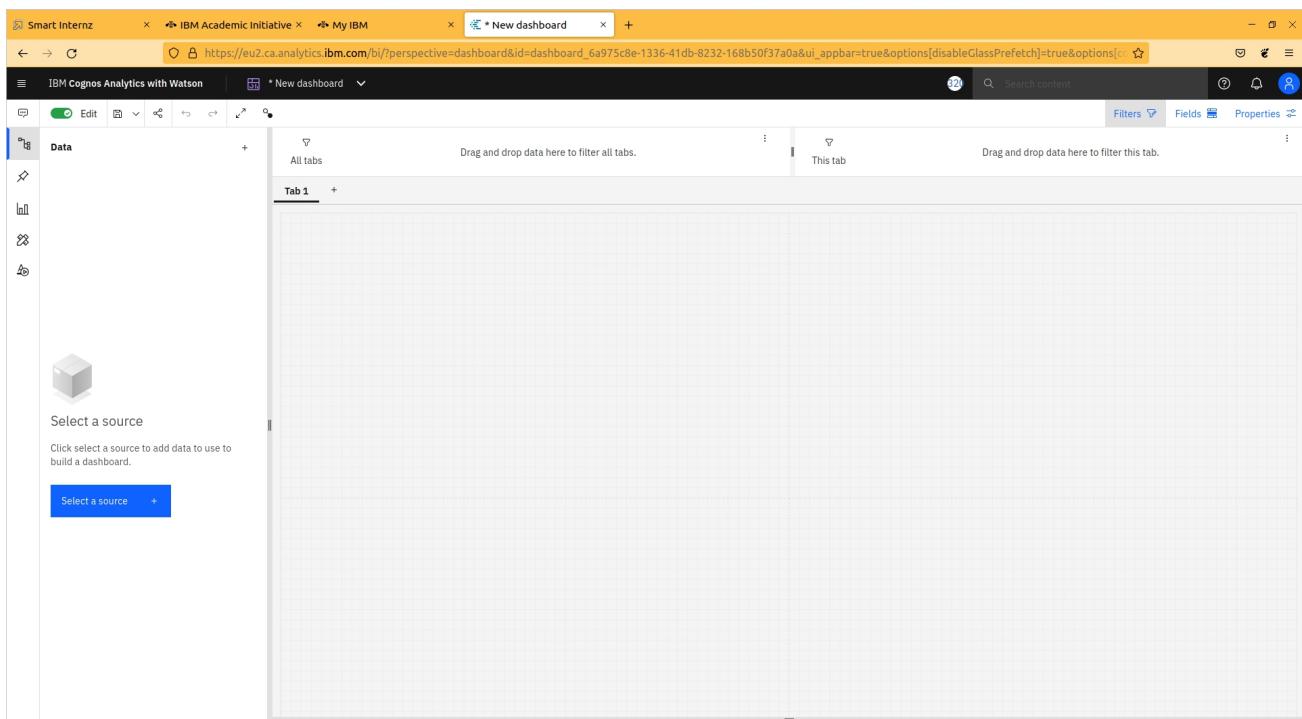


Fig: Raw Dashboard

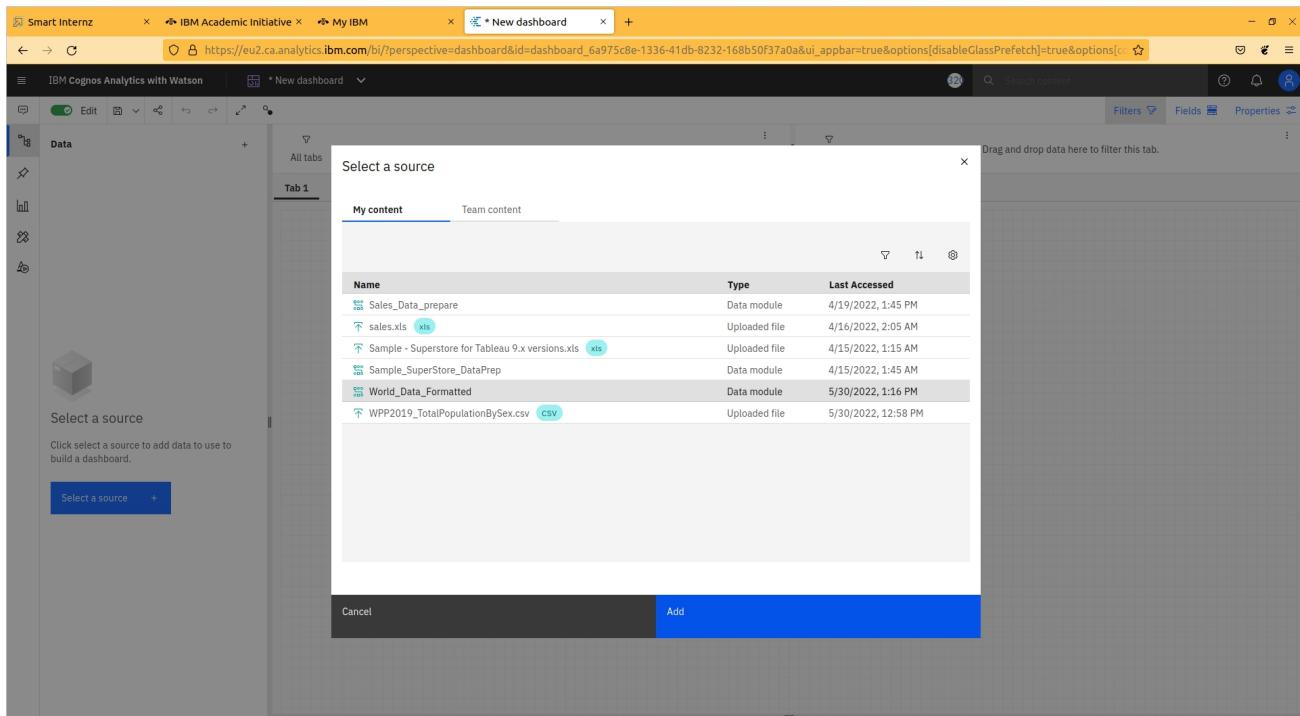


Fig: Selecting source for preparing dasboard

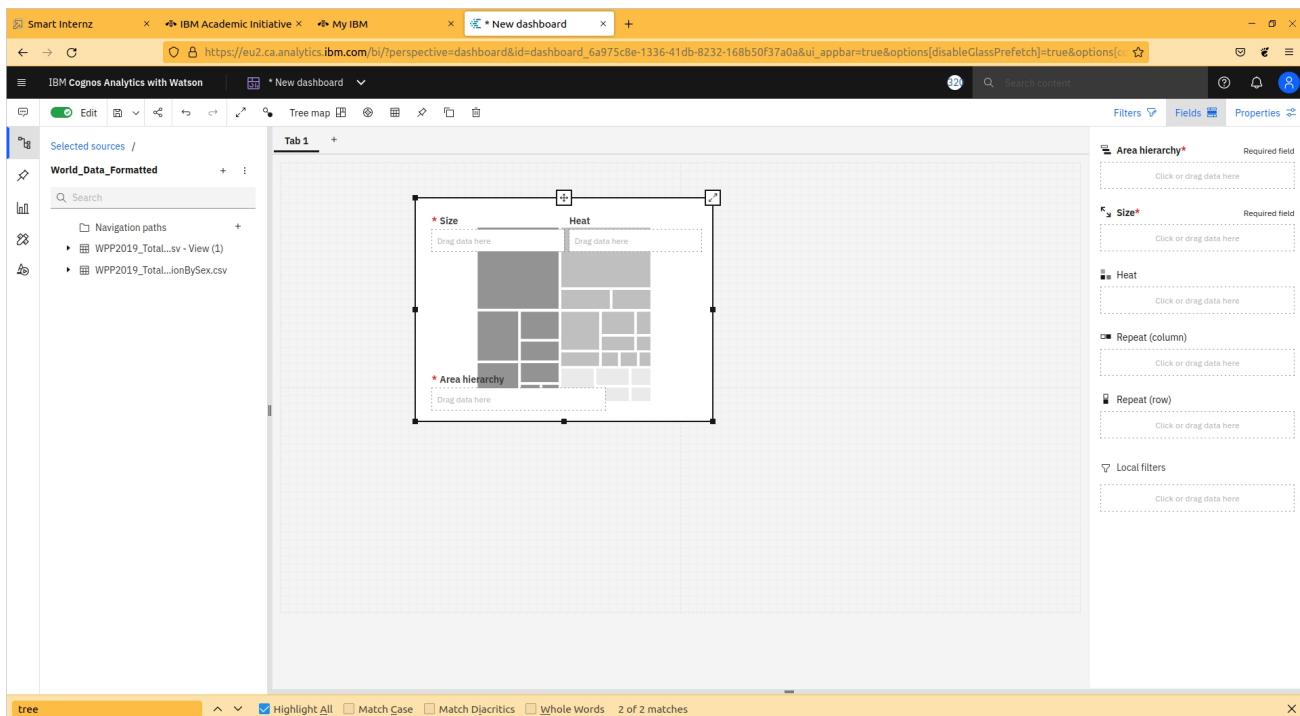


Fig: Dragging tree map for preparing visualization

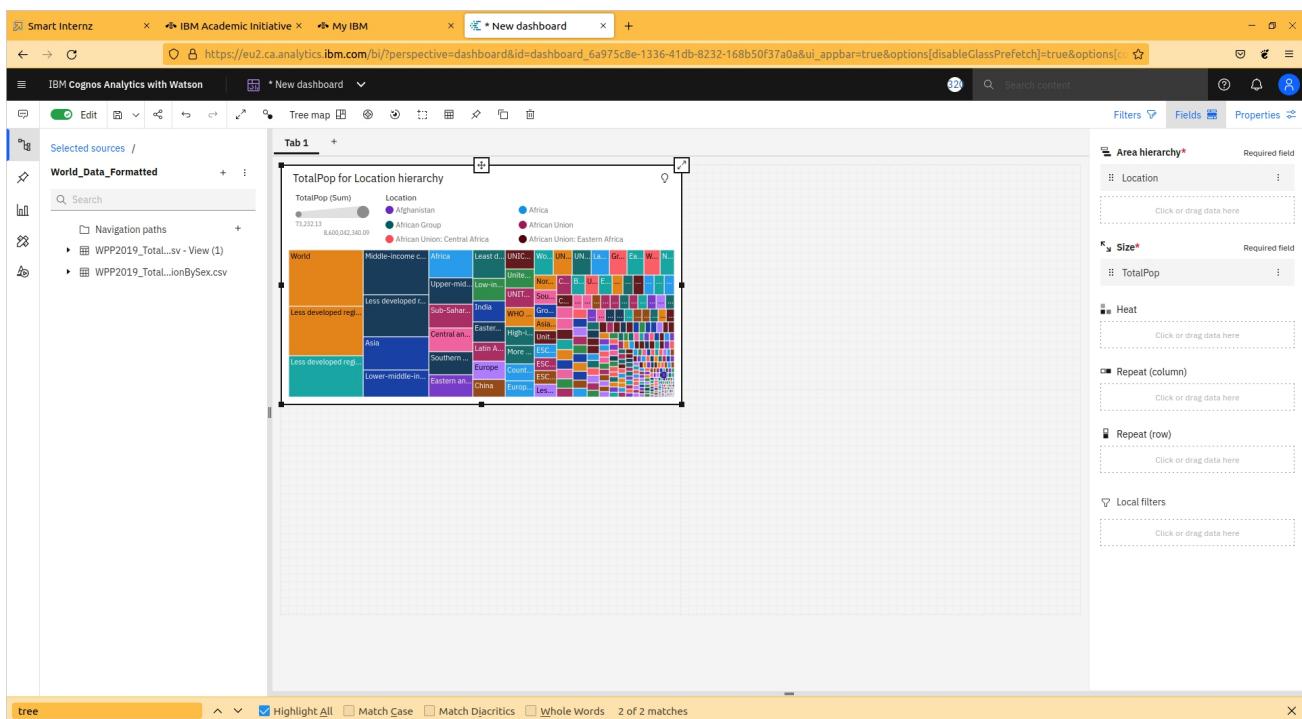


Fig: Raw visualization for tree map

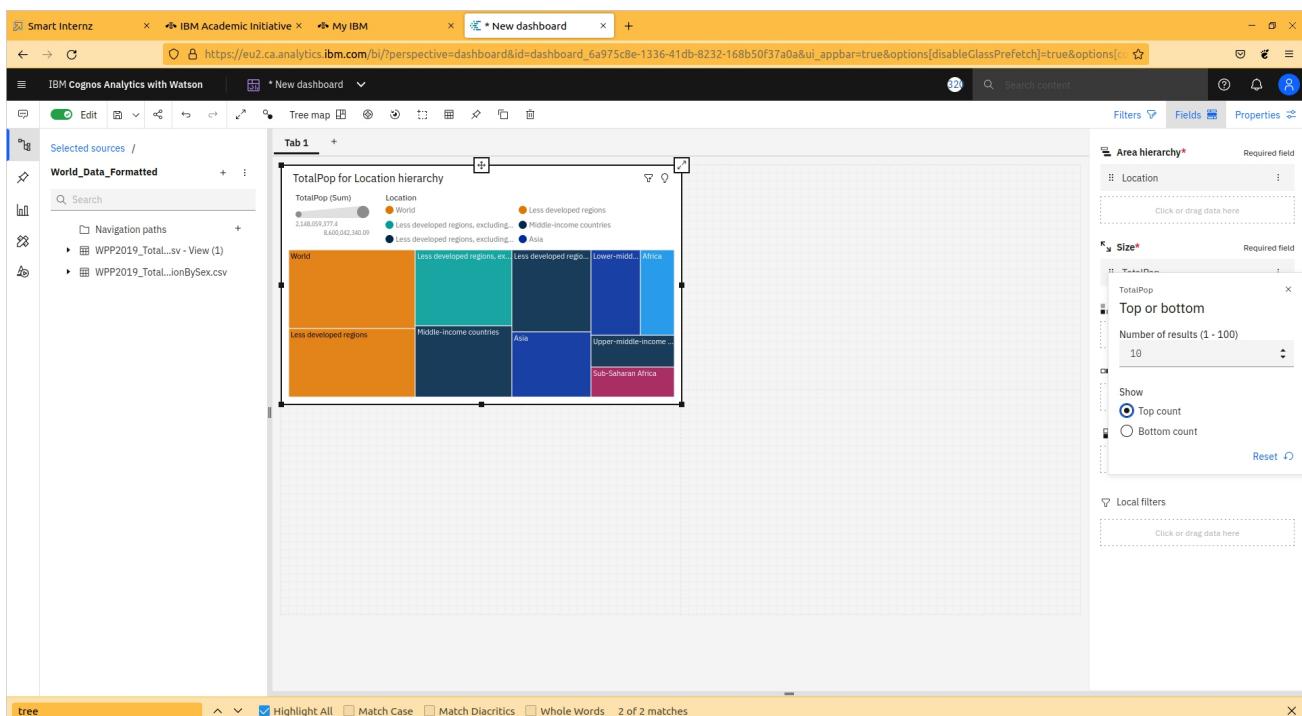


Fig: Selecting top count for total population

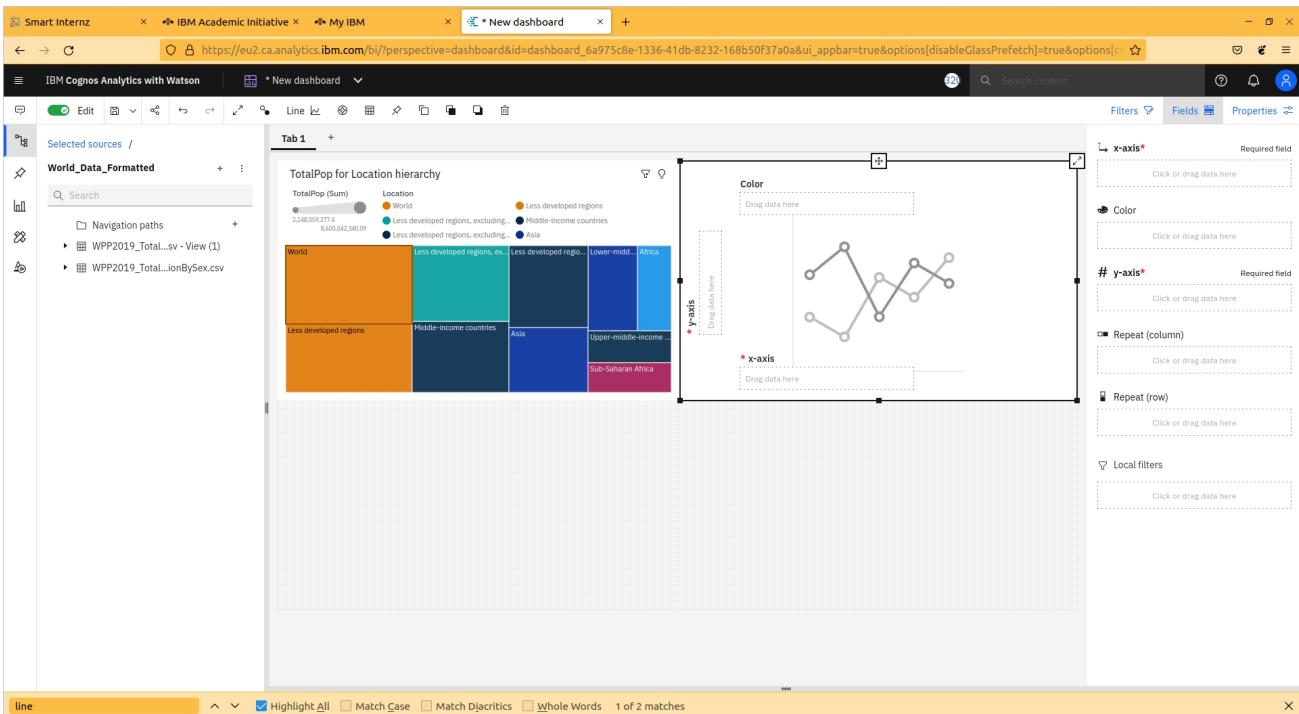


Fig: Selecting line graph

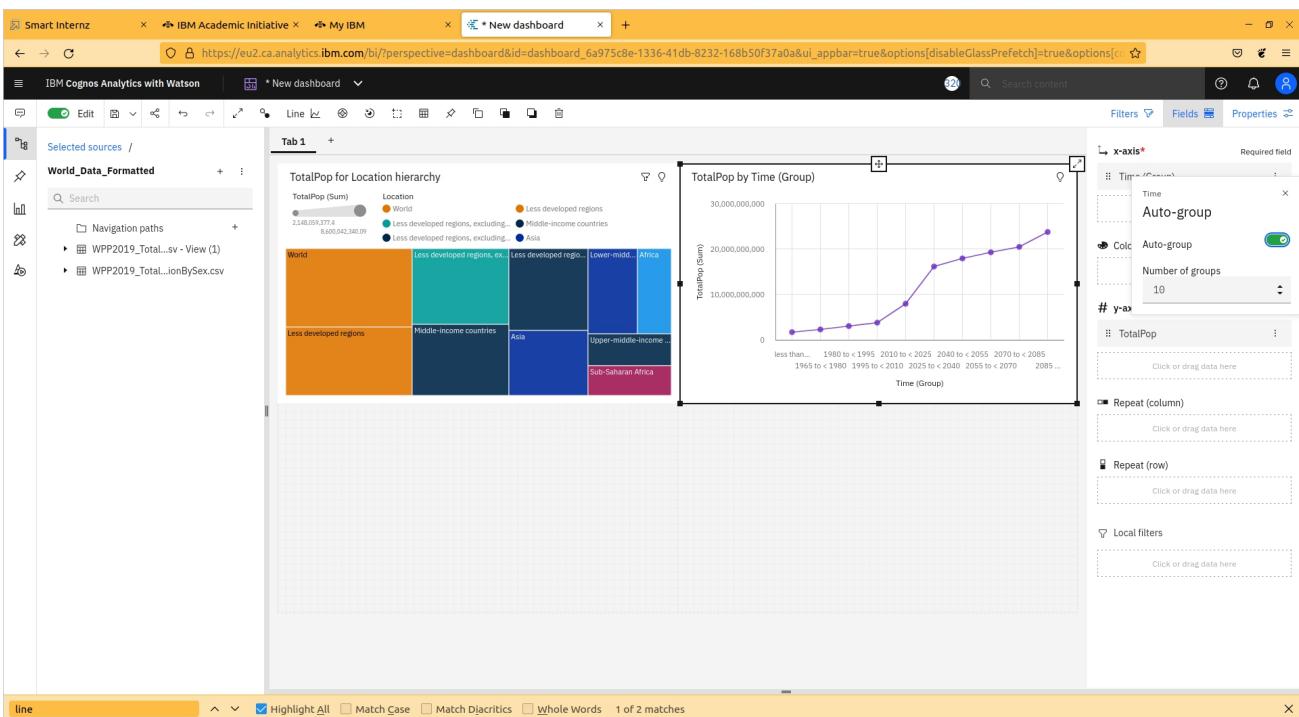


Fig: Selecting time and auto grouping

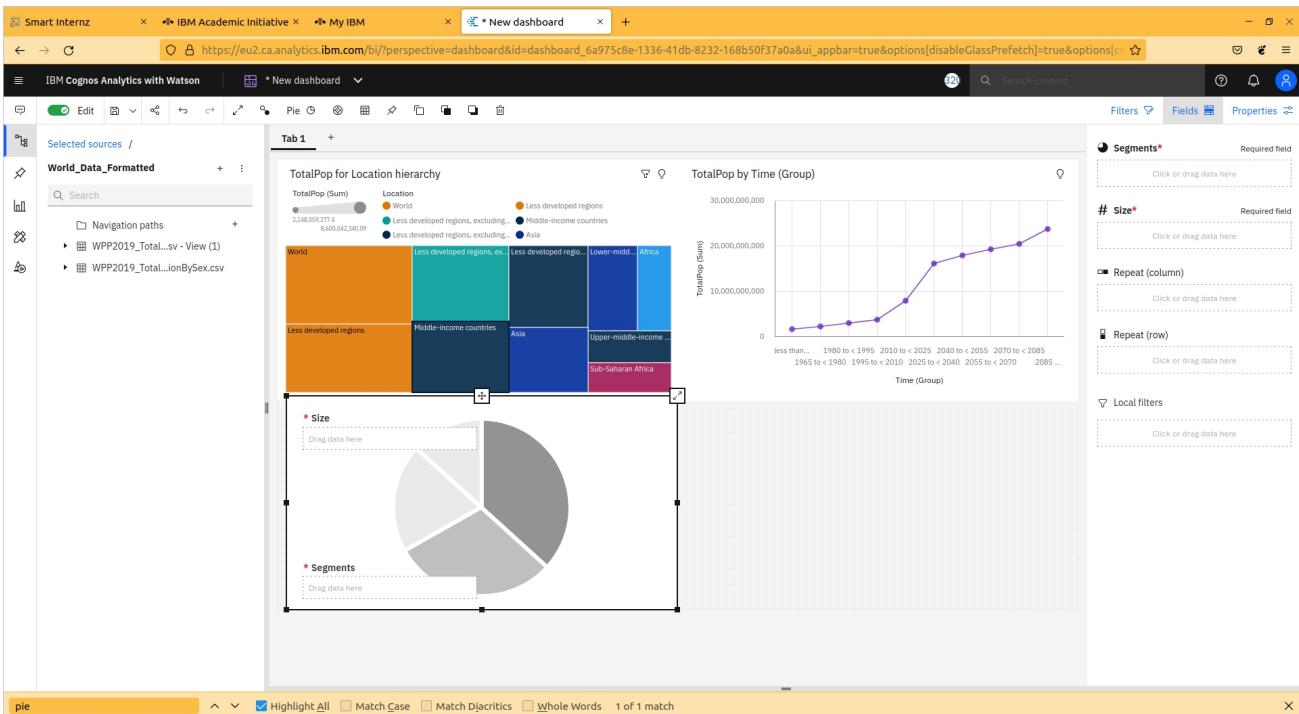


Fig: Selecting pie chart

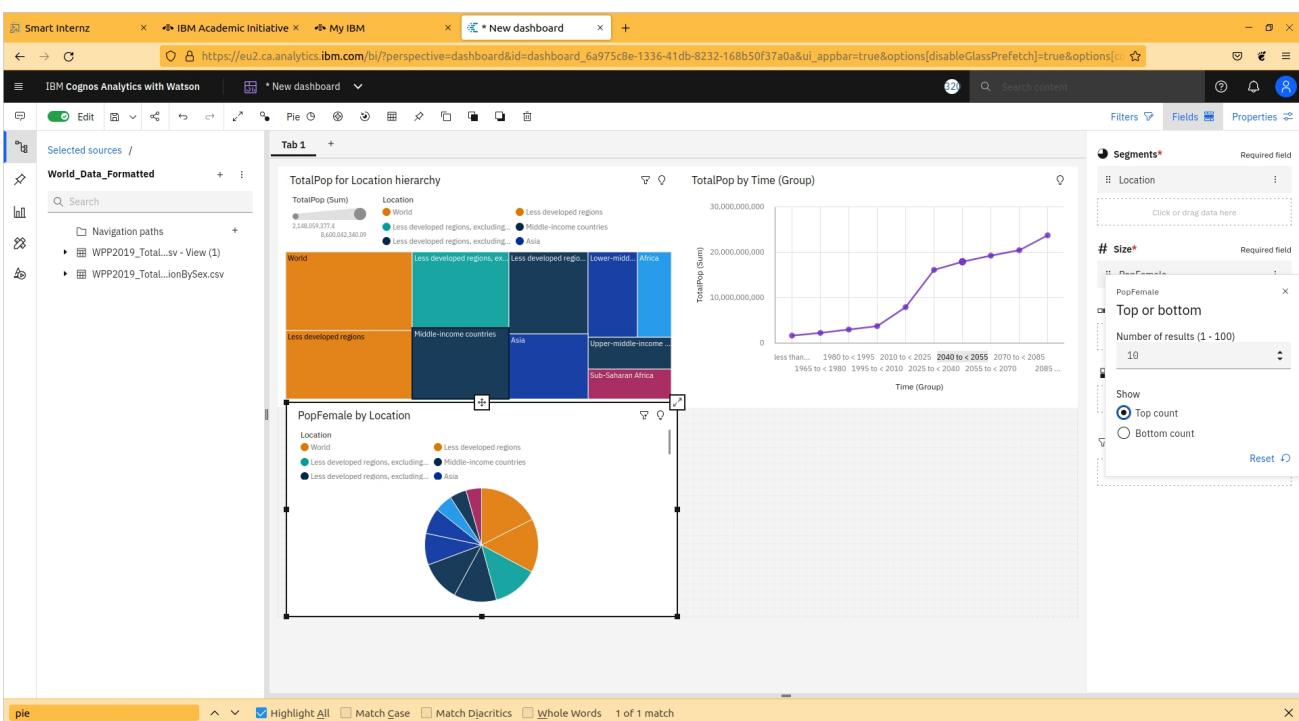


Fig: Selecting total female top 10 and presenting

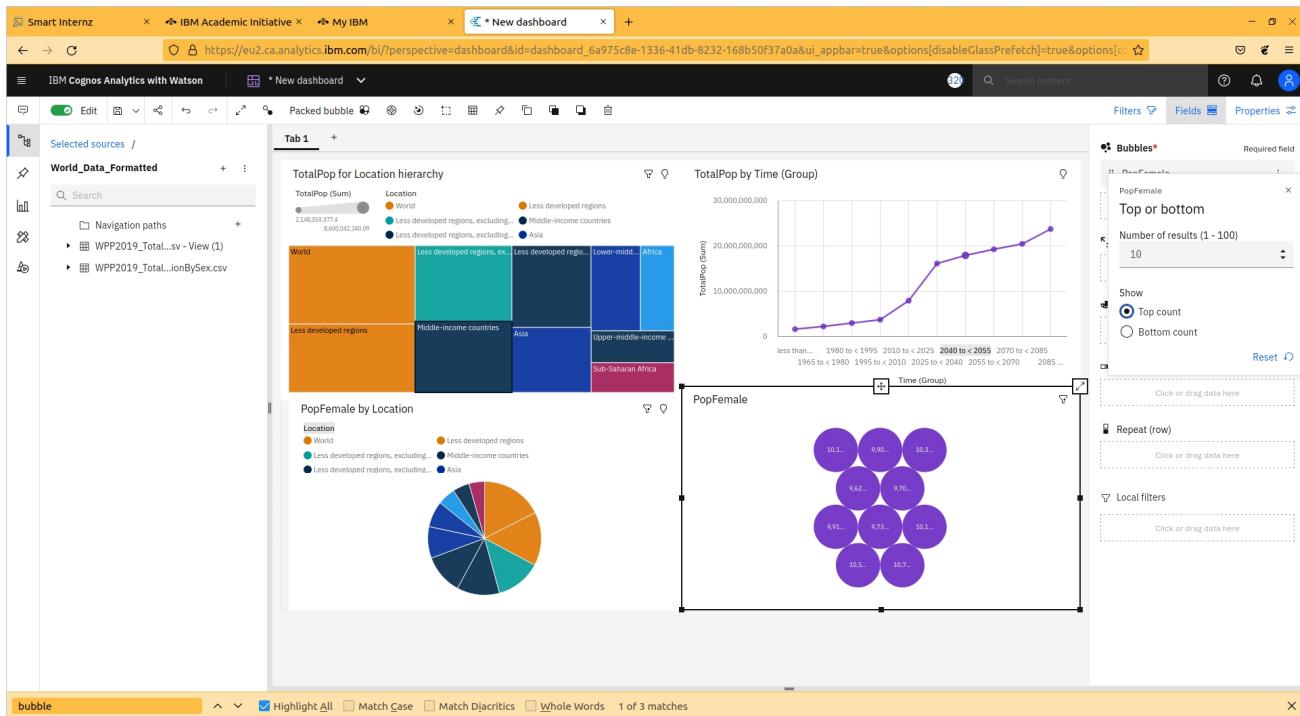


Fig: Selecting total size by female population and presenting in bubble

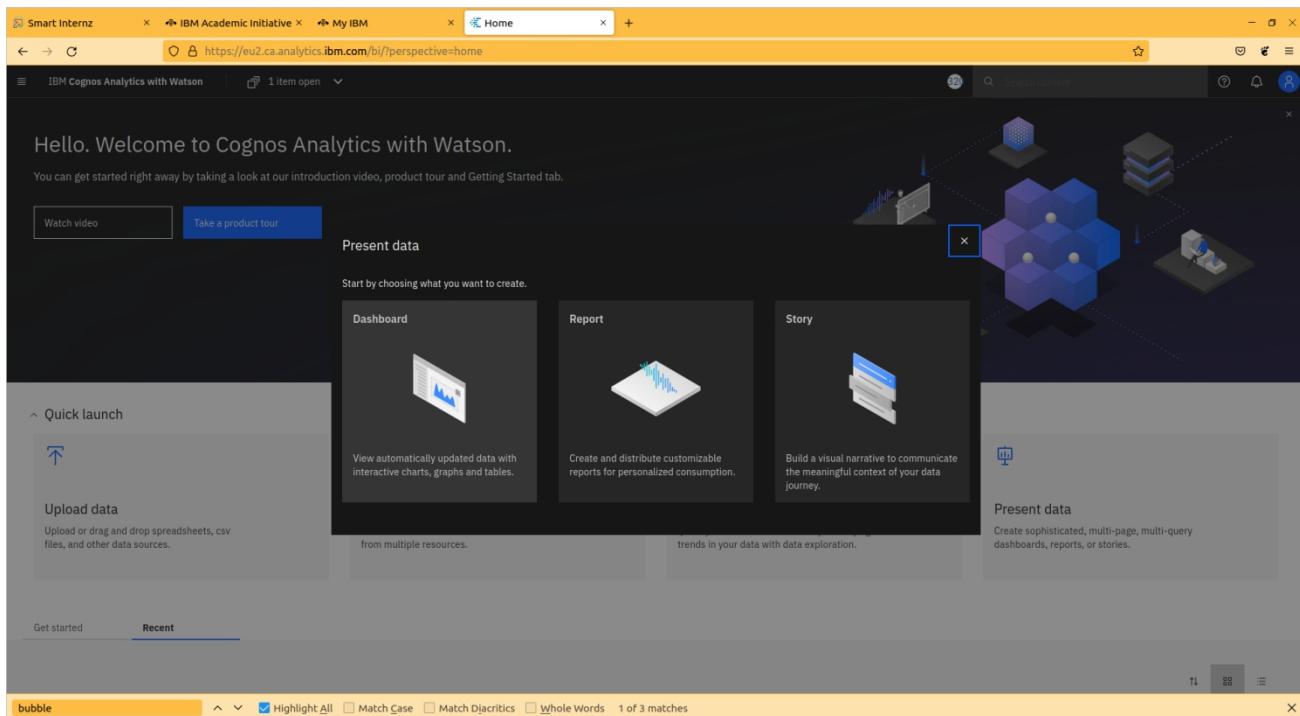


Fig: Selecting dashboard for another summary

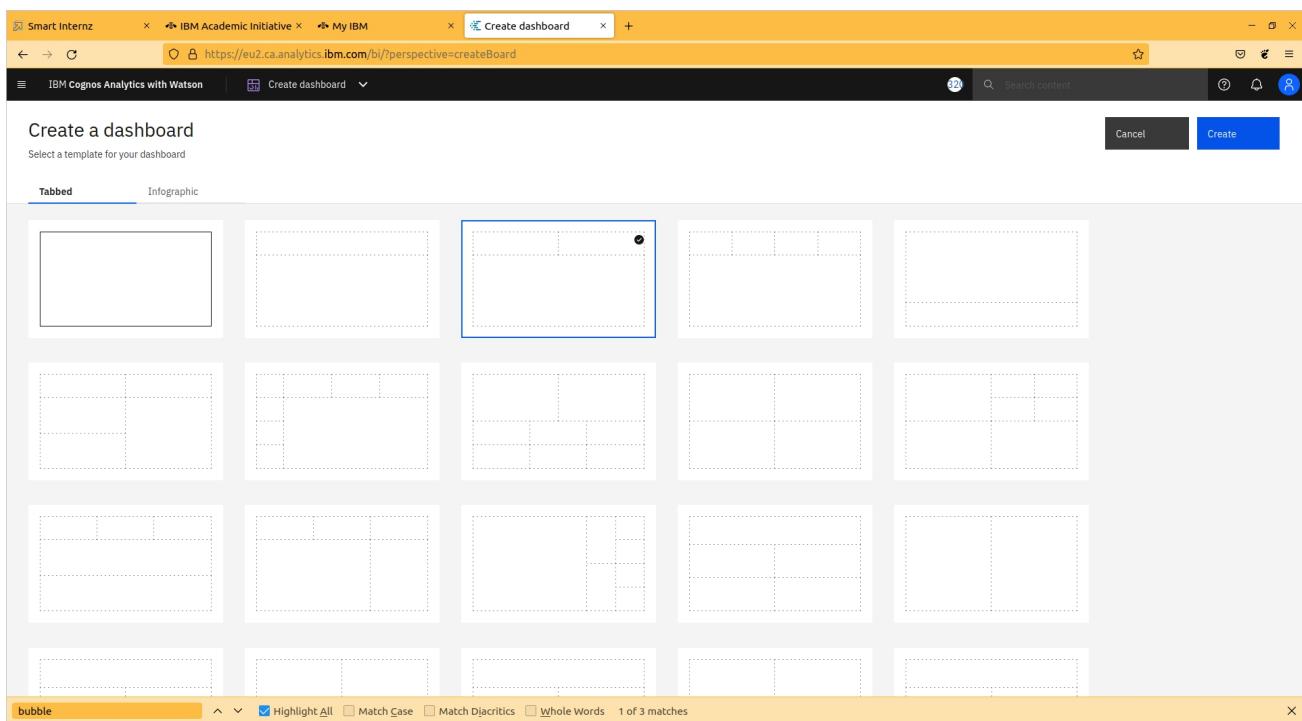


Fig: Selecting appropriate format for dashboard

Name	Type	Last Accessed
Sales_Data_prepare	Data module	4/19/2022, 1:45 PM
sales.xls	Uploaded file	4/16/2022, 2:05 AM
Sample - Superstore for Tableau 9.x versions.xls	Uploaded file	4/15/2022, 1:15 AM
Sample_SuperStore_DataPrep	Data module	4/15/2022, 1:45 AM
World_Data_Formatted	Data module	5/30/2022, 1:16 PM
WPP2019_TotalPopulationBySex.csv	Uploaded file	5/30/2022, 12:58 PM

Fig: Selecting source for summary

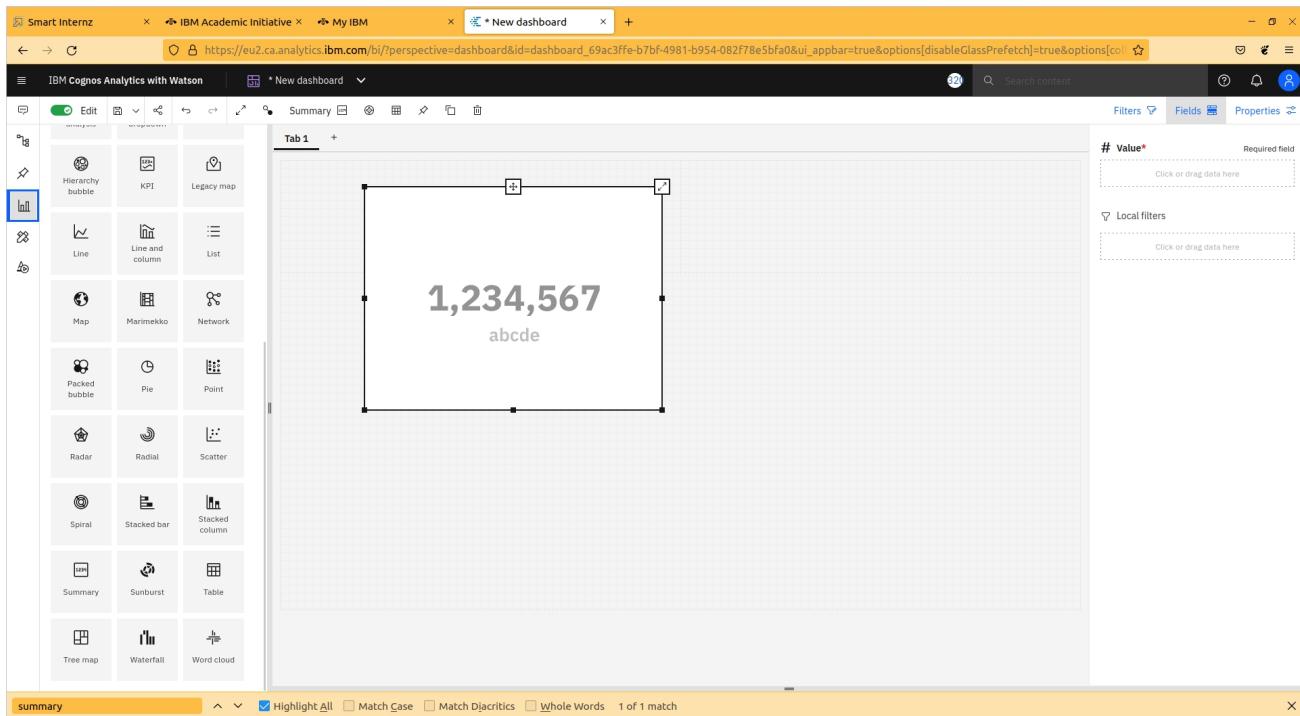


Fig: Adding summary to the dashboard

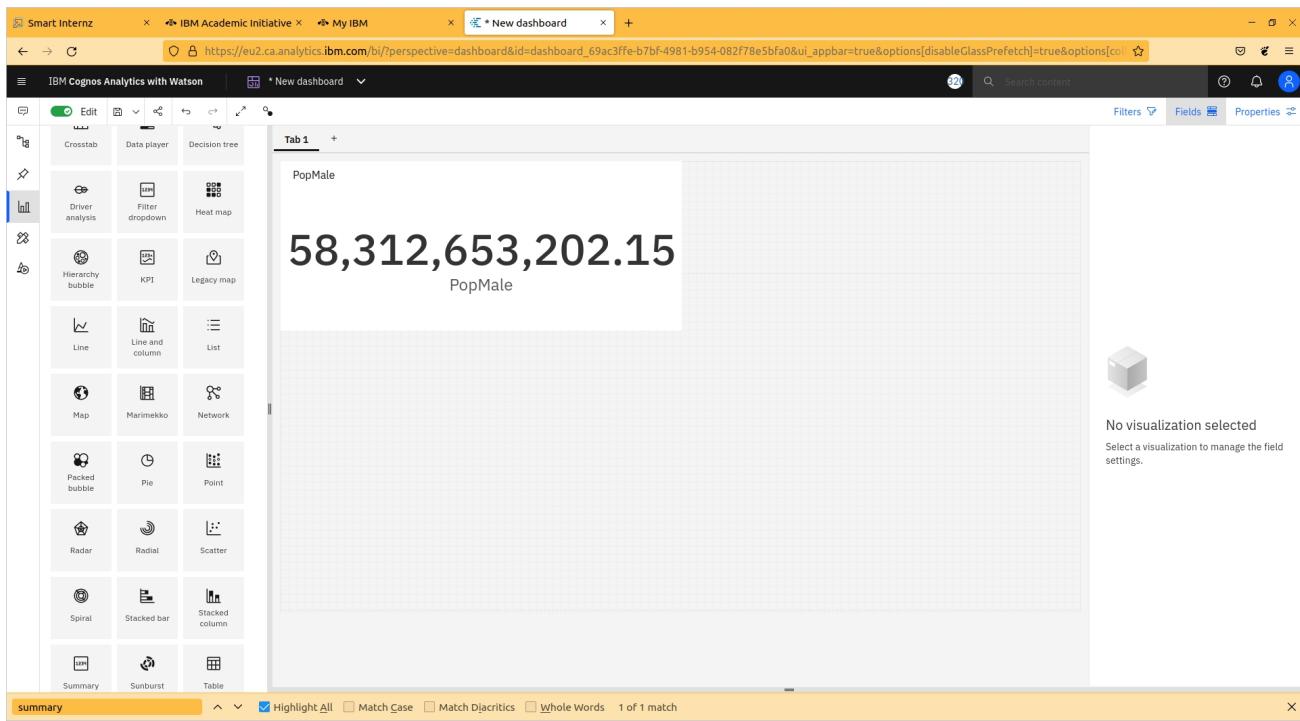


Fig: Selecting population Male

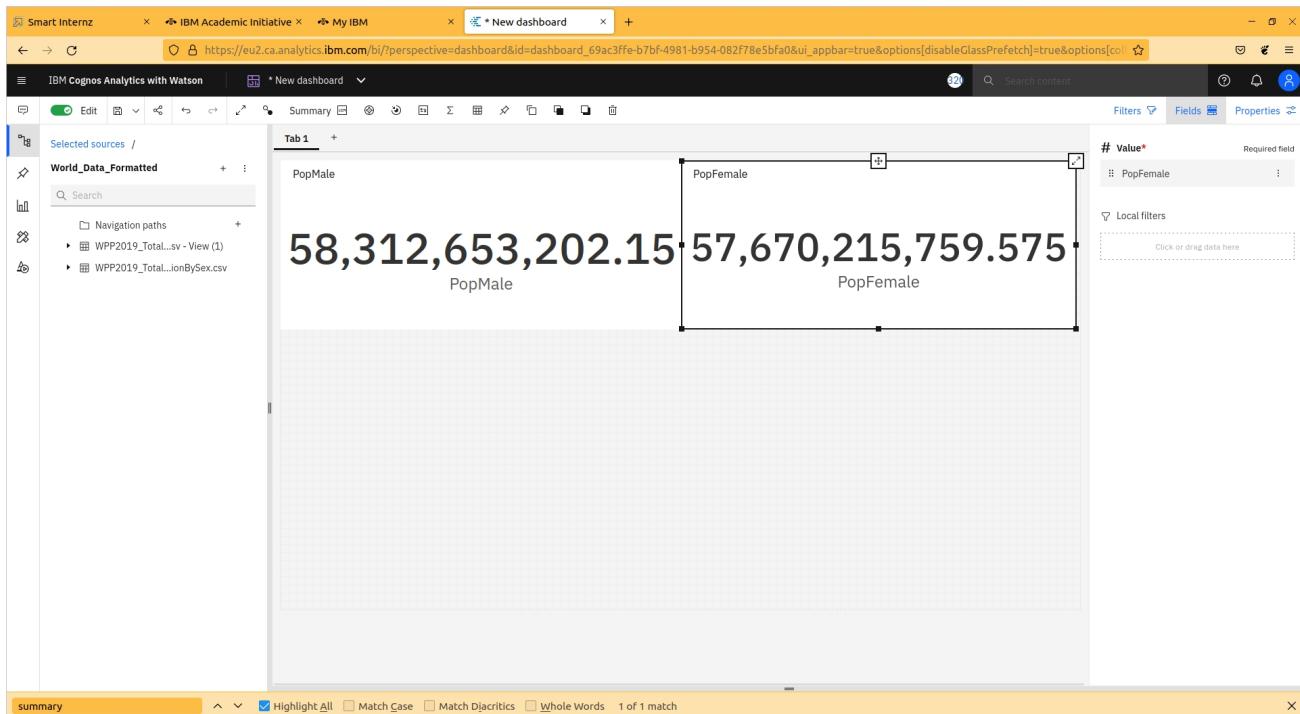


Fig: Selecting population female

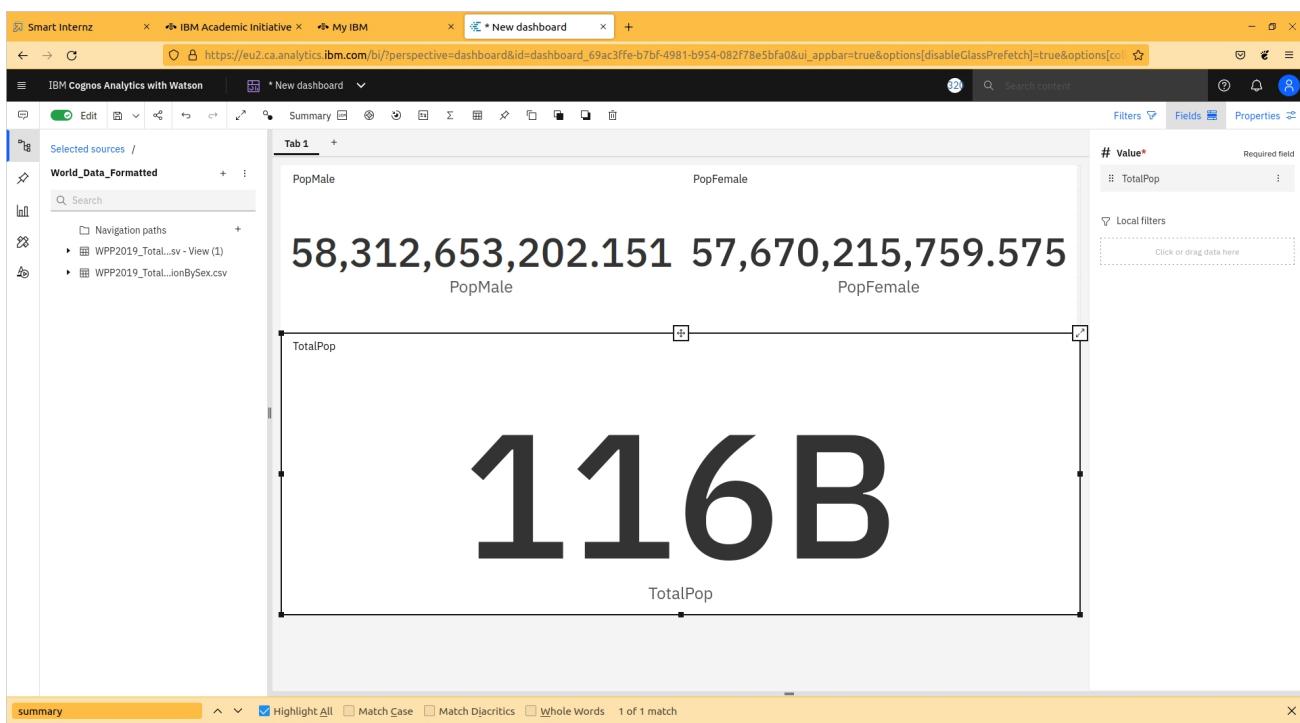


Fig: Selecting total population

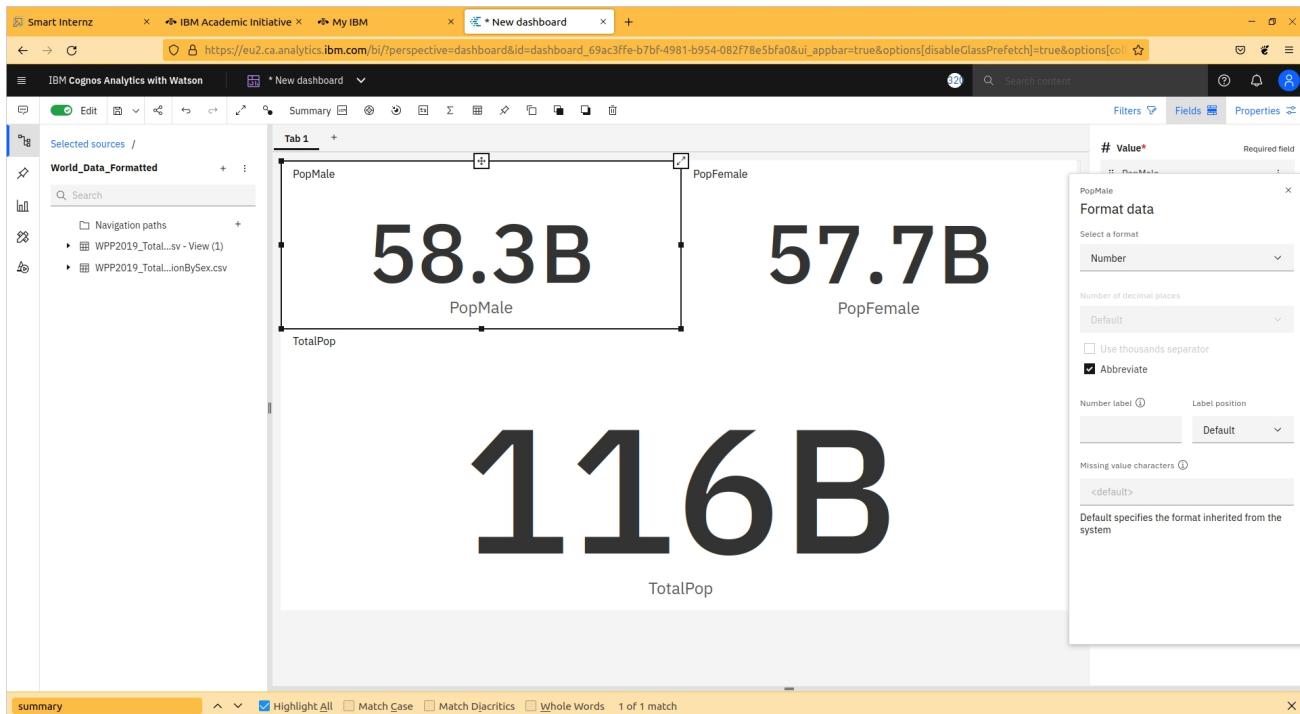


Fig: Formatting and selecting abbreviation