Integrating Warehouses with BI Tools

• Write a 1-2 page executive summary summarizing your assignment activities

As we move onto the next stage of the Data warehousing for Avacados2Go, we would now be integrating the warehouse to Business Intelligence tools that help fully leverage the usage of data warehouses and help investigate and retrieve more actionable insights for the business.

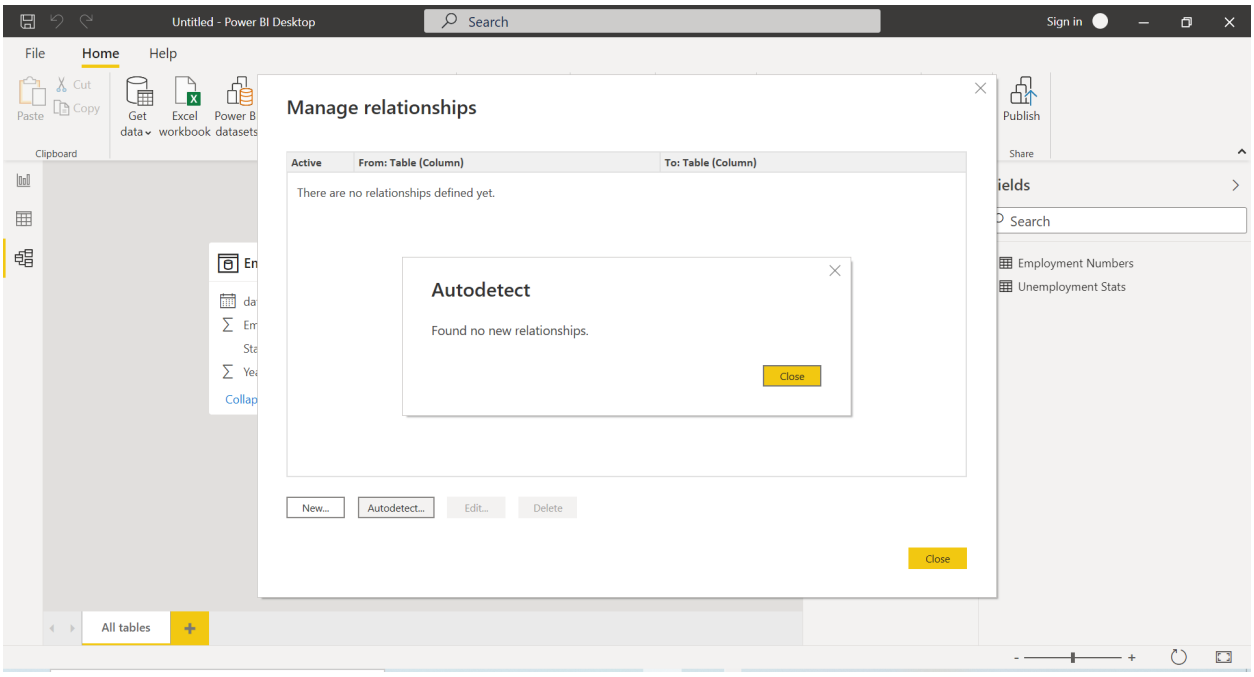
For these purposes, the BI Tool Power BI Desktop was chosen as this was free and readily available on the internet. The installation was also self-explanatory and simple.

Once installed, for initial testing purposes, we connected to a sample employment statistic worksheet that encapsulates the employment numbers over the years. Once we connected and imported the dataset into Power BI, we noticed that employed numbers were abbreviated in thousands (like 1.5K) and therefore needed cleaning. We understood that such inconsistencies need to be addressed and cleaned at the ETL stages itself.

We then created a sample line chart visualization for the worksheet using the automatically created Date Hierarchy and employment numbers. After playing around with the features, we also forecasted for the current trend of employment numbers and adjusted it to factor in the seasonality constraints. A similar chart was created for unemployment rates as well.

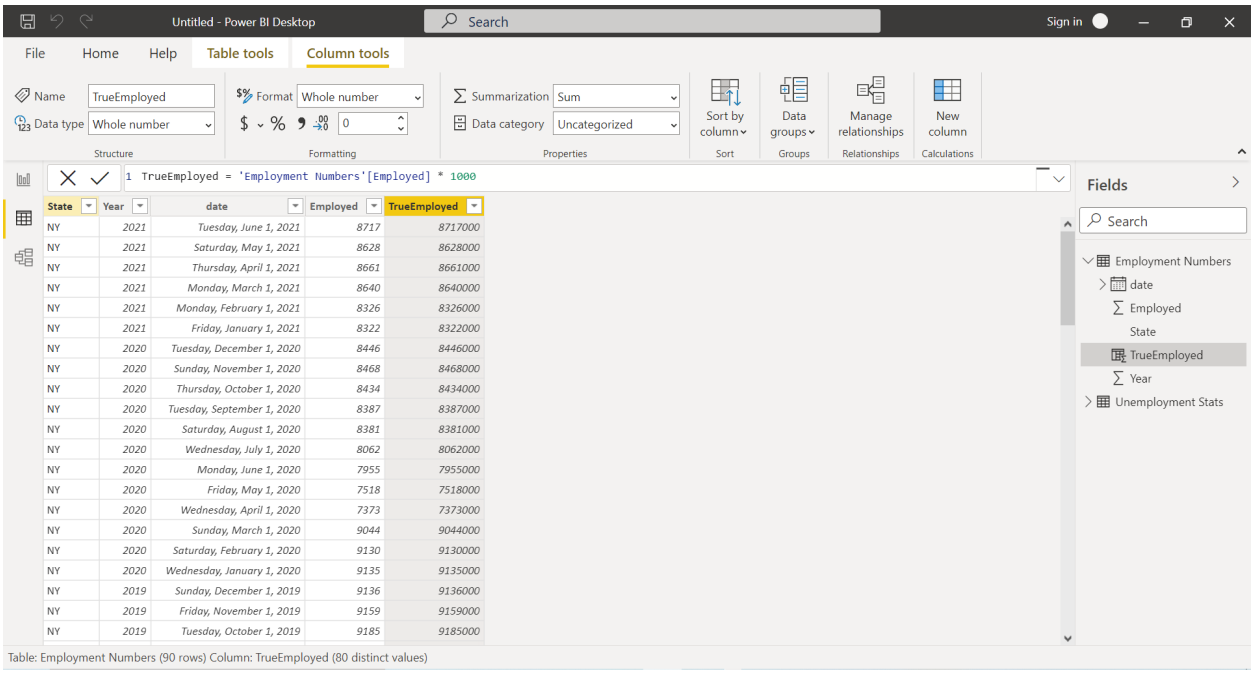
Moving forward, we successfully connected the data warehouse to the BI Tool and created a few derived attributes as columns in the Data pane for better visualizations. We then used these fields to create a Dashboard that showcases the sales over the years using the historical data, forecast over the following years and provides actionable insights to the business.

o Click on the “Manage relationships” button in the top toolbar. This will open the relationship manager. To see if there are any relationships between the data, you can select “Autodetect…” from the bottom of the management window. Click the button and paste a screenshot into your report. Answer the following questions: Did autodetect find any relationships between the data? Why do you think this was the case? You can also manually create relationships if you feel there should be a relationship the tool did not find. Close the Manage relationships window.



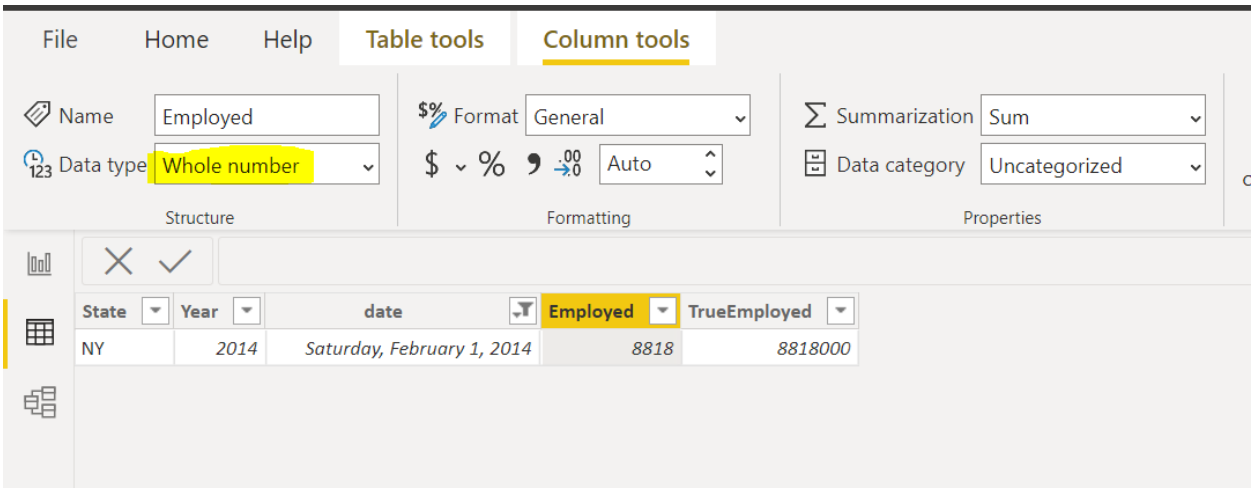
Autodetect did not find any relationships because there is no primary-foreign key relationship between the two tables to uniquely identify a row in both. Moreover, the only similar field viz., State, does not qualify to be used as a primary key (is not unique).

o Type the following formula into the equation: TrueEmployed = 'Employment Numbers'[Employed] * 1000 And hit enter. Paste a screenshot of the resulting screen in your report. Along with it, answer the following question: What was the TrueEmployed value for Saturday, February 1, 2014?



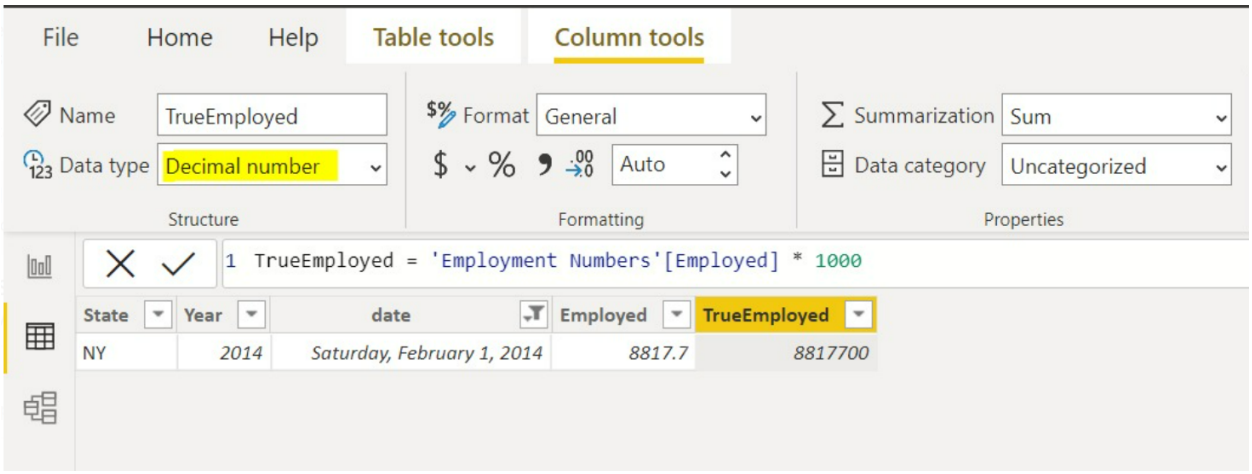
State	Year	date	Employed	TrueEmployed
NY	2021	Tuesday, June 1, 2021	8717	8717000
NY	2021	Saturday, May 1, 2021	8628	8628000
NY	2021	Thursday, April 1, 2021	8661	8661000
NY	2021	Monday, March 1, 2021	8640	8640000
NY	2021	Monday, February 1, 2021	8326	8326000
NY	2021	Friday, January 1, 2021	8322	8322000
NY	2020	Tuesday, December 1, 2020	8446	8446000
NY	2020	Sunday, November 1, 2020	8468	8468000
NY	2020	Thursday, October 1, 2020	8434	8434000
NY	2020	Tuesday, September 1, 2020	8387	8387000
NY	2020	Saturday, August 1, 2020	8381	8381000
NY	2020	Wednesday, July 1, 2020	8062	8062000
NY	2020	Monday, June 1, 2020	7955	7955000
NY	2020	Friday, May 1, 2020	7518	7518000
NY	2020	Wednesday, April 1, 2020	7373	7373000
NY	2020	Sunday, March 1, 2020	9044	9044000
NY	2020	Saturday, February 1, 2020	9130	9130000
NY	2020	Wednesday, January 1, 2020	9135	9135000
NY	2019	Sunday, December 1, 2019	9136	9136000
NY	2019	Friday, November 1, 2019	9159	9159000
NY	2019	Tuesday, October 1, 2019	9185	9185000

Whole number data type:



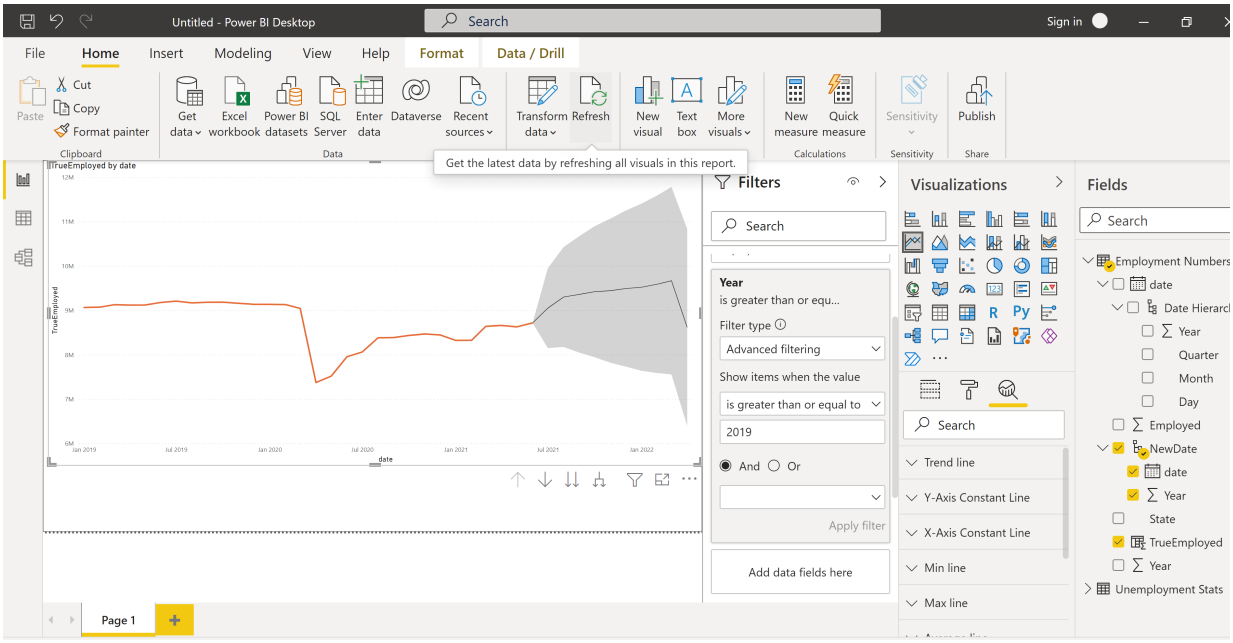
State	Year	date	Employed	TrueEmployed
NY	2014	Saturday, February 1, 2014	8818	8818000

Decimal number data type:



TrueEmployed value for Saturday, February 1, 2014, is 8818000 for whole number data type and 8817700 for decimal data type

Screenshot of the result set after the mentioned steps were followed:



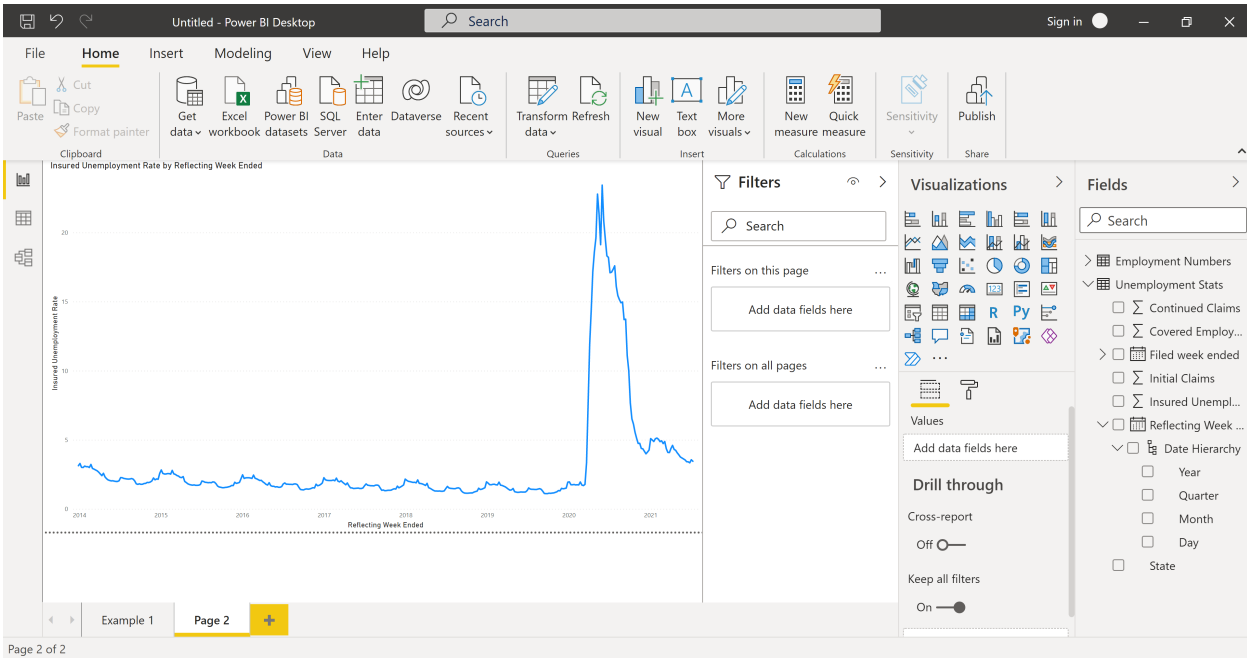
o Now that you have the basics down, create a second line chart using the "Unemployment Stats" worksheet. It should leverage the following fields:

▪ Reflecting Week Ended

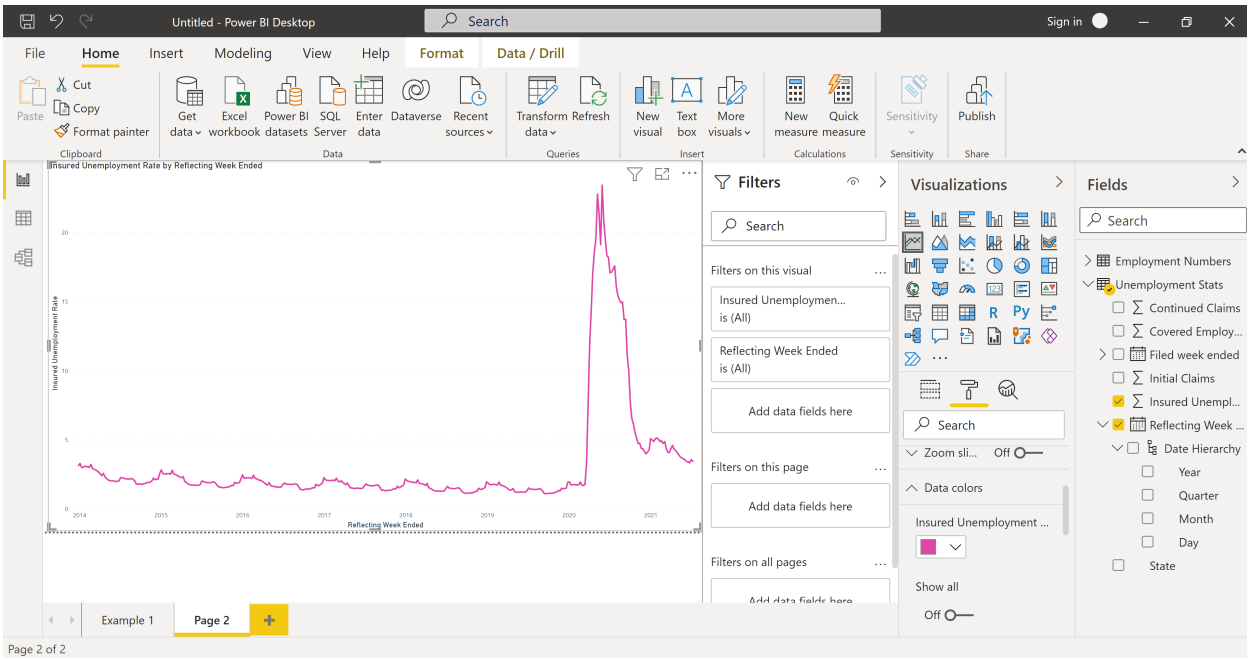
▪ Insured Unemployment Rate

HINT: The line chart will look “smoothed” due to the date hierarchy. You’ll need to change the Axis settings for Reflecting Week Ended from “Date Hierarchy” to “Reflected Week Ended” to fix this.

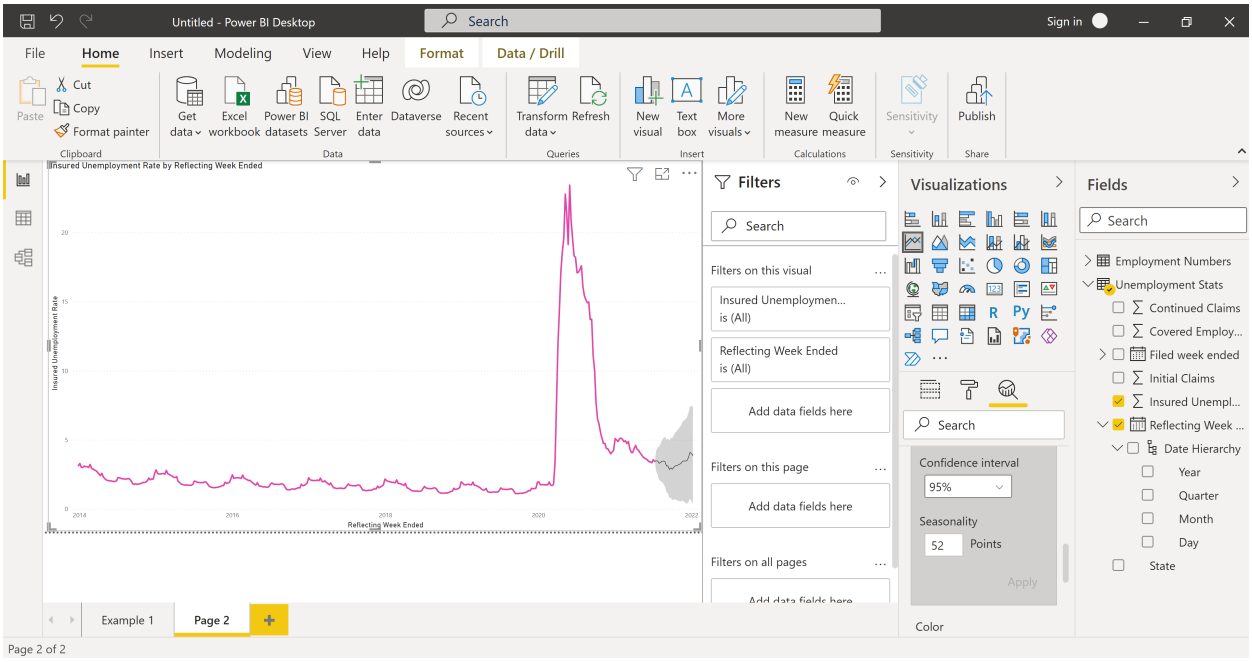
Paste a picture of the resulting line chart.



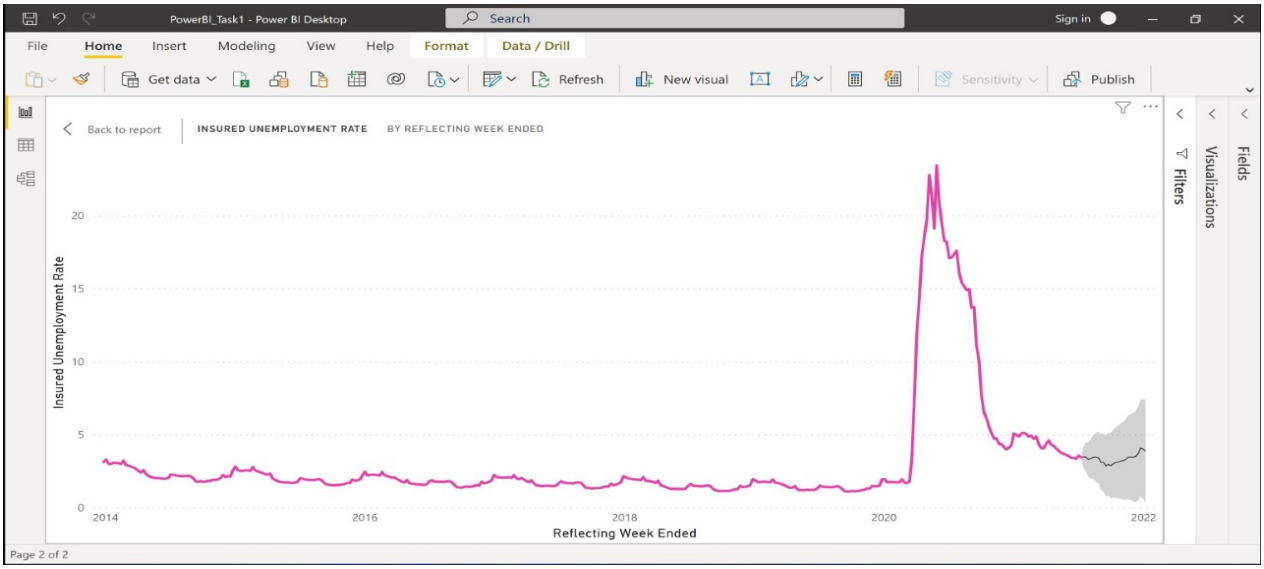
o Change the color of the line chart to something other than blue. Paste a picture of the new line chart in your report.



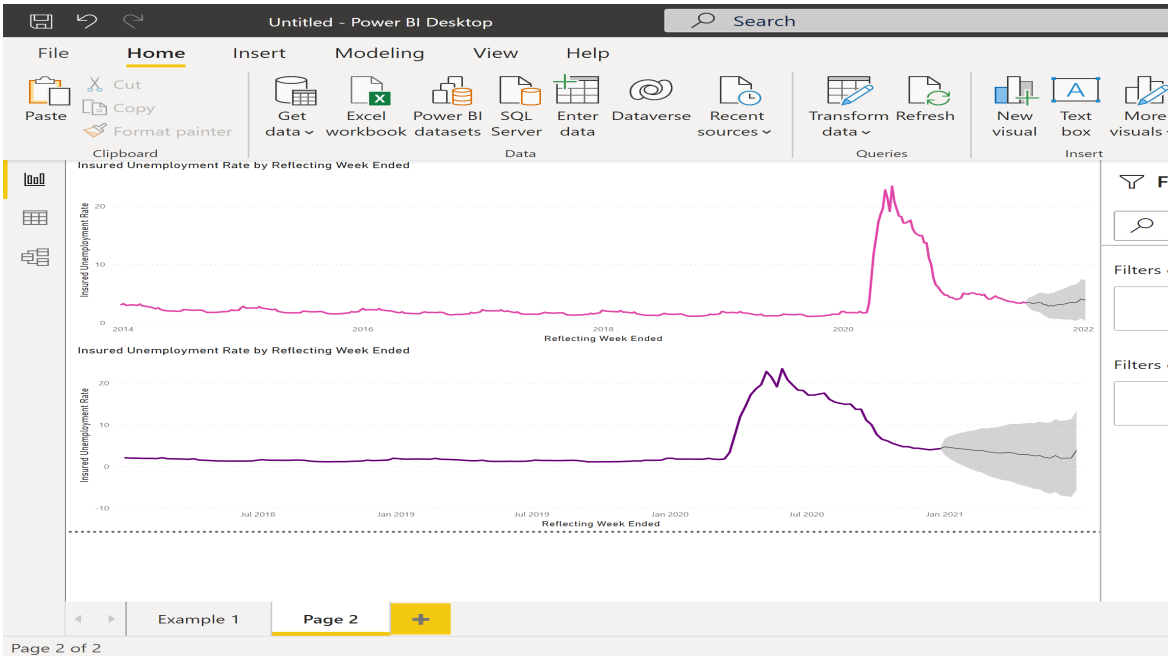
o Create a forecast, setting the forecast length to 26 and the seasonality to 52 (this matches the data seasonality). Paste a picture of the line chart in your report.



o Change your line chart to filter for a Relative date showing items in the last 3 calendar years.



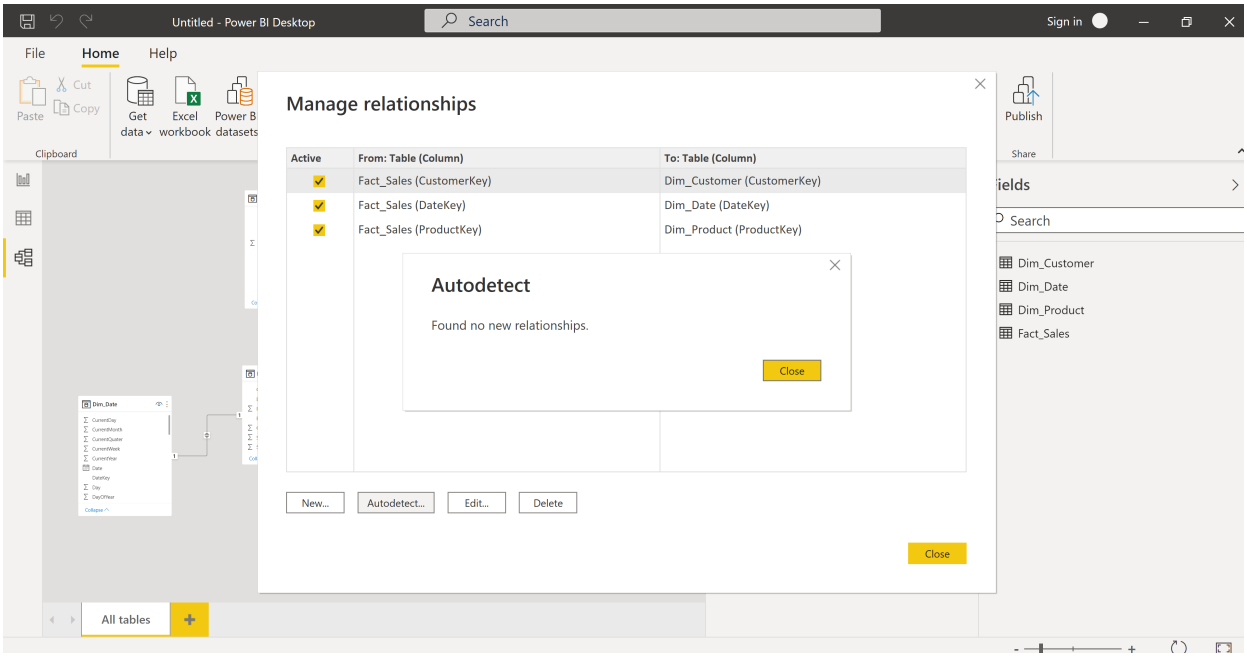
o Paste a picture of the full page (with both line charts) in your report.



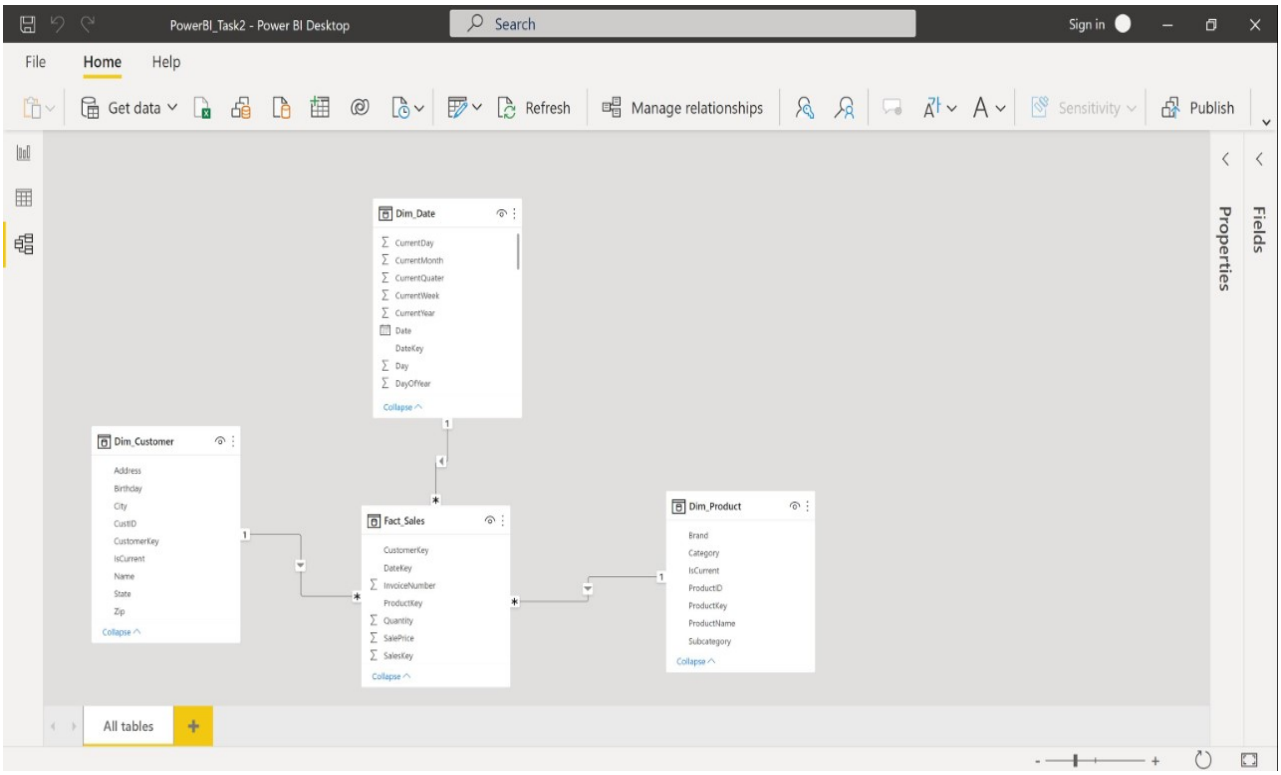
o Write 2-4 sentences in your report detailing any insights you can draw from these two-line charts.

The two line charts helps investigate the decline in unemployment rate by giving both a holistic and detailed view of the decline. We can infer that the unemployment rate started decreasing around August 2020 and started steadying around January 2021. It is forecasted that the unemployment rate will further decrease over time through Jan 2022.

o Open the “Model” workspace from the left panel. You should see all the imported tables and relationships. If you do not, use the autodetect function to add relationships between the tables (hint: review the earlier sections of the assignment if you don’t know how).

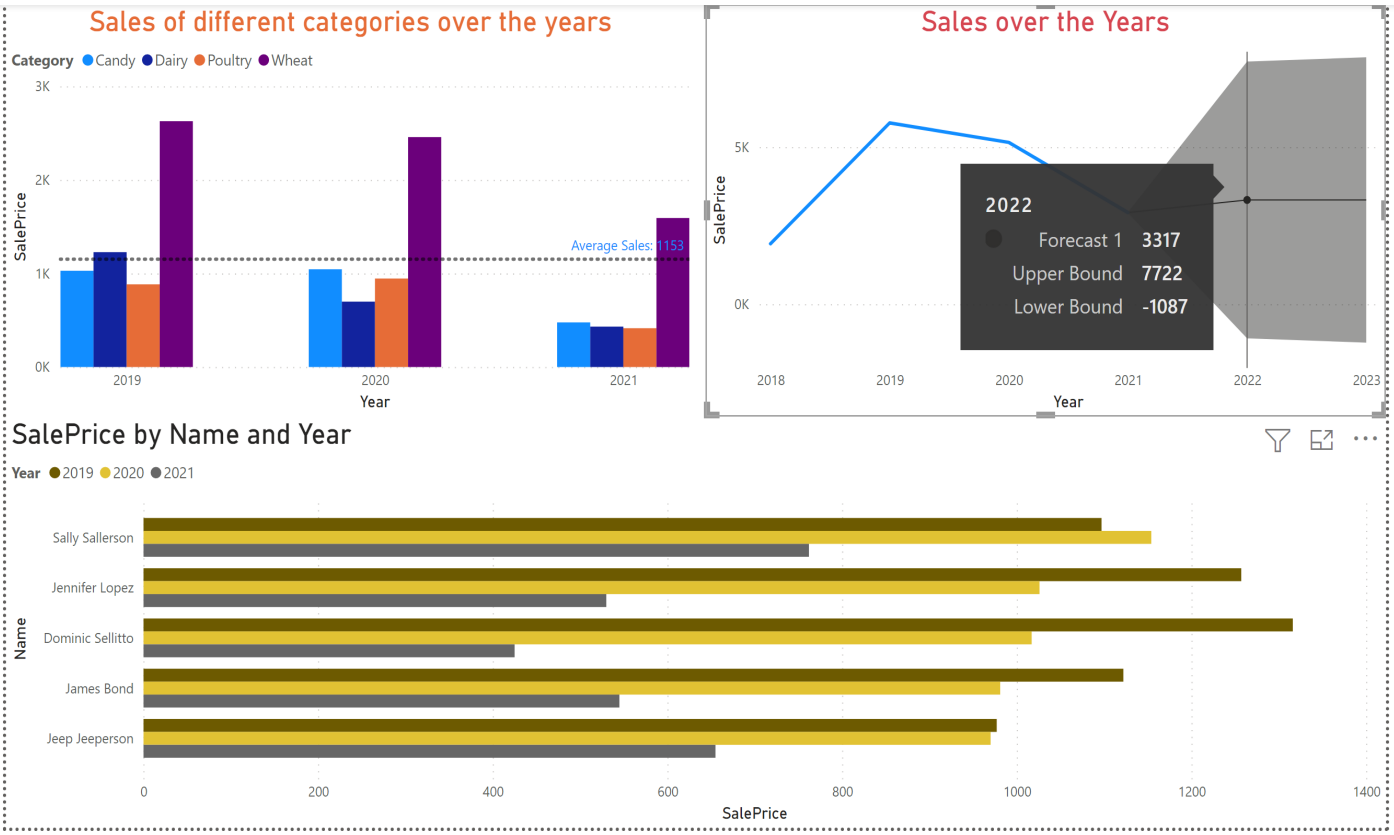


o Take a screenshot of the resulting relationship diagram and attach it to your report.



o Using the information you know about Power BI, the data source, and the business requirements for the warehouse, create a 1-page dashboard in Power BI of the data. Your 1-page dashboard can include any visuals you would like but must include at least one chart with a forecast.

▪ Paste a screenshot of your dashboard in your final report. You can also use the built in "Export to PDF" option in Power BI.



o Write 4-6 sentences describing your dashboard (tell your story!)

The dashboard showcases the decline in Sales through the 'Sales over the Years' graph. Moreover, we have forecasted the observed Sales trend to identify that, in the year 2022, there would be an increase in Sales followed by a steady Sales in 2023.

The Dashboard also helps investigate the decline by offering:

➔ A Product Category deep-dive that shows which product category contributed most to the decline.

➔ It also views Sales over the years through a Customer Lens through the ‘Sales by Name’ graph thus pointing out that the customer, Dominic, exhibits the biggest slump in sales.