

Executive Summary

After gathering information from the business, designing a data warehouse, and creating an ETL pipeline for the Avacados2Go, we finally proceeded with the visualization and analytics i.e., reviewing the data model and relationships and then building reports and dashboards. The whole purpose of doing the same is to get a better picture of the entire process and demonstrate the final outcomes to businesses and those who require it for business purposes.

While working on the guided assignment in the Employment Numbers table it was instructed to change the decimal values in the Employed column to a whole number by changing the data type. And then multiply them by 1000 to get the true employed number. But what we did is we multiplied the decimal value in the employed column with 1000 and got the true employed value and then made that data type of that column to a whole number.

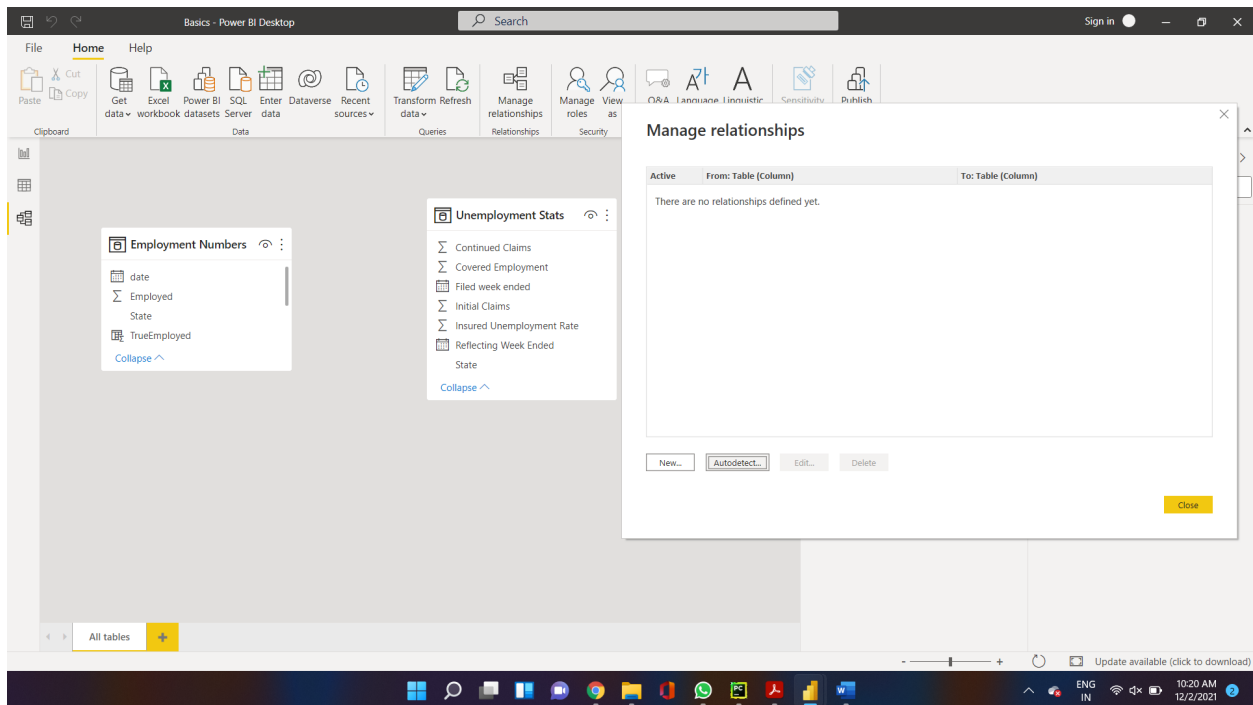
After completing the guided experiment. We started working on the Avacados2Go database. We connected to our SQL Server Database and imported all the dimensions and fact tables to the Power BI. There is a wide range of methods for determining the variation in sales and the objective of the project is to create a sales forecast based on the historical data provided to us. We created graphs showing sale price by year and category, sales for the top three products, sale price by city, and sale price by year. The process of the creation of the dashboards has been documented along with the snapshots of each of the steps in the process. The tools used for this purpose were Microsoft Power BI and SSMS i.e., SQL Server Management Studio 18.

We then used the features of Power BI to create an interactive dashboard. We used the model view in the quick view panel to understand the relationship between the tables. And the data view provided us a view of the contents in the tables. And we created the dashboard using the report option in the quick view panel. We used the analytics option in the Power BI to forecast the sales for the future years using the historical data.

All this work with the help of Power Bi seemed smooth but in comparison with Tableau, Tableau was more user friendly, had a better user interface and features, and methods to create calculated fields are less cumbersome in Tableau.

APPENDIX: -

1. 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?



There is no relationship between the data. Because there is no foreign key relationship between tables. Also, there are no common fields between the two tables for joining except "State" field which can't be taken as primary key as it is not unique.

2. Paste a screenshot of the resulting screen in your report.

Table: Employment Numbers (90 rows) Column: TrueEmployed (87 distinct values)

State	Year	date	Employed	TrueEmployed
NY	2021	Tuesday, June 1, 2021	8716.9	8716900
NY	2021	Saturday, May 1, 2021	8627.9	8627900
NY	2021	Thursday, April 1, 2021	8661.1	8661100
NY	2021	Monday, March 1, 2021	8639.5	8639500
NY	2021	Monday, February 1, 2021	8325.8	8325800
NY	2021	Friday, January 1, 2021	8321.7	8321700
NY	2020	Tuesday, December 1, 2020	8446.2	8446200
NY	2020	Sunday, November 1, 2020	8468.2	8468200
NY	2020	Thursday, October 1, 2020	8433.9	8433900
NY	2020	Tuesday, September 1, 2020	8387	8387000
NY	2020	Saturday, August 1, 2020	8380.9	8380900
NY	2020	Wednesday, July 1, 2020	8062.3	8062300
NY	2020	Monday, June 1, 2020	7955	7955000
NY	2020	Friday, May 1, 2020	7517.6	7517600
NY	2020	Wednesday, April 1, 2020	7372.8	7372800
NY	2020	Sunday, March 1, 2020	9043.6	9043600
NY	2020	Saturday, February 1, 2020	9130	9130000
NY	2020	Wednesday, January 1, 2020	9134.7	9134700
NY	2019	Sunday, December 1, 2019	9136.3	9136300
NY	2019	Friday, November 1, 2019	9159	9159000
NY	2019	Tuesday, October 1, 2019	9185	9185000
NY	2019	Sunday, September 1, 2019	9181.6	9181600
NY	2019	Thursday, August 1, 2019	9168.4	9168400
NY	2019	Monday, July 1, 2019	9208.8	9208800
NY	2019	Saturday, June 1, 2019	9178.3	9178300
NY	2019	Wednesday, May 1, 2019	9120.3	9120300
NY	2019	Monday, April 1, 2019	9119.3	9119300
NY	2019	Friday, March 1, 2019	9136.7	9136700

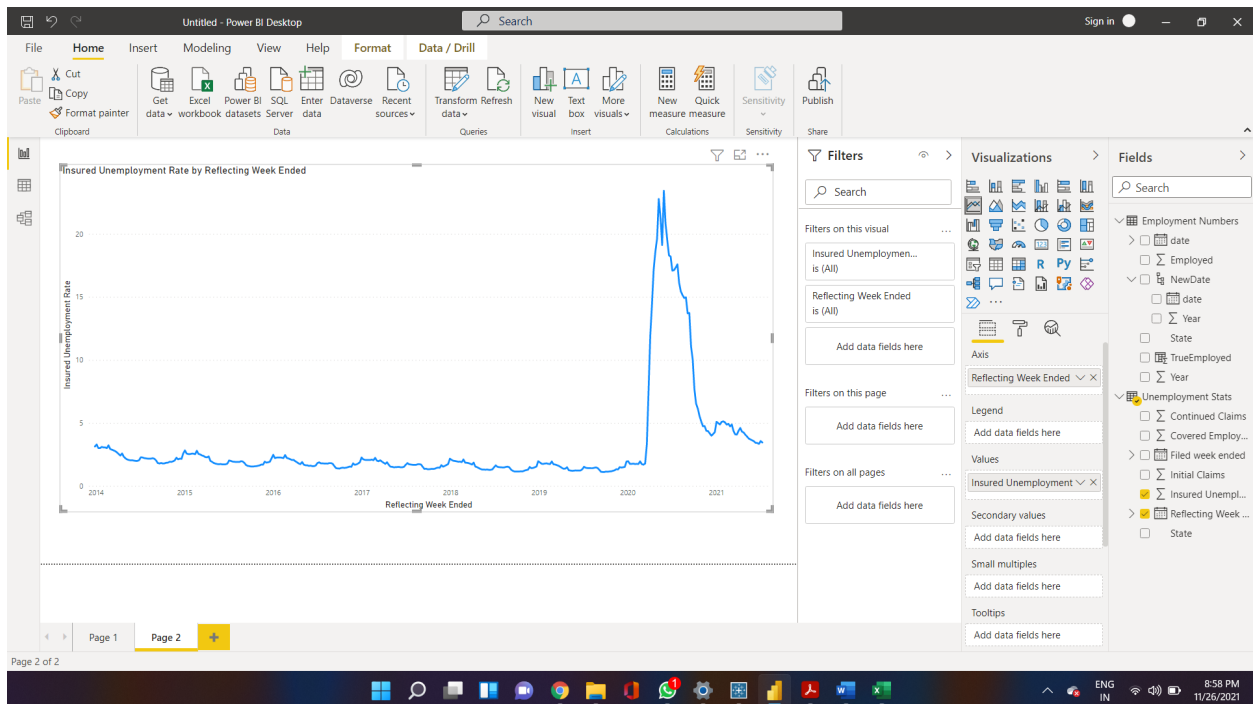
What was the TrueEmployed value for Saturday, February 1, 2014? **8817700**

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

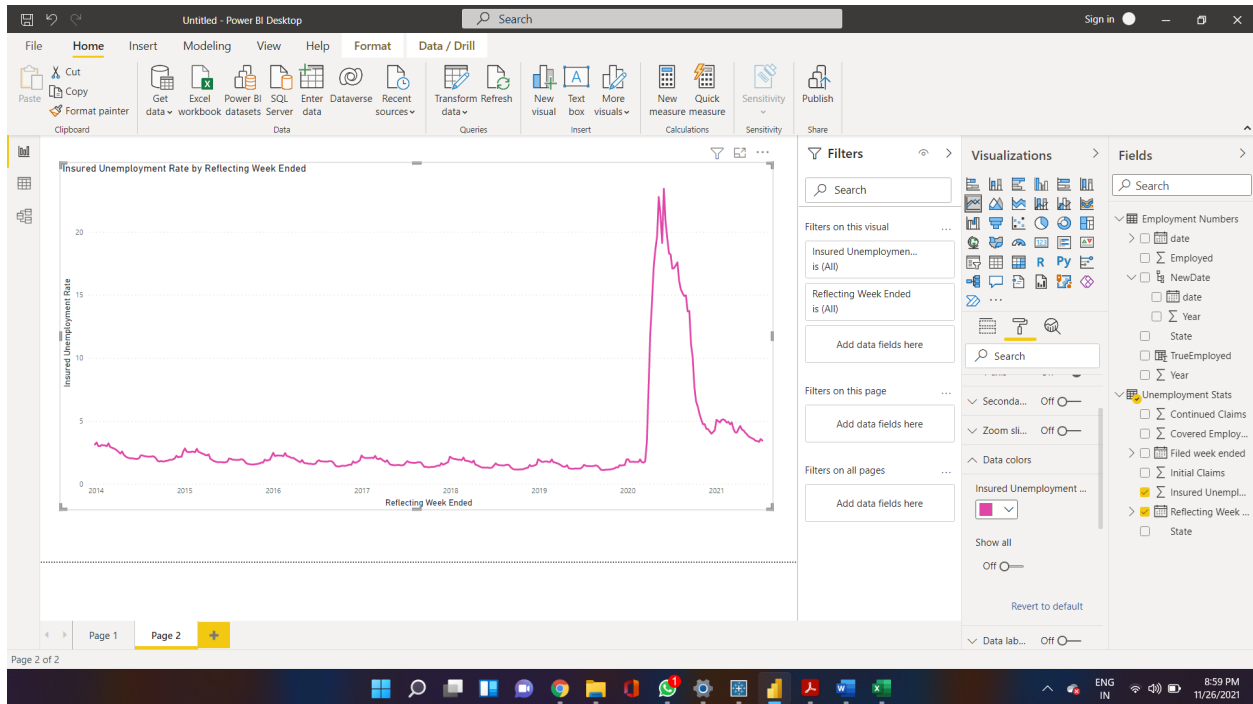
- a. Reflecting Week Ended
- b. Insured Unemployment Rate

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.

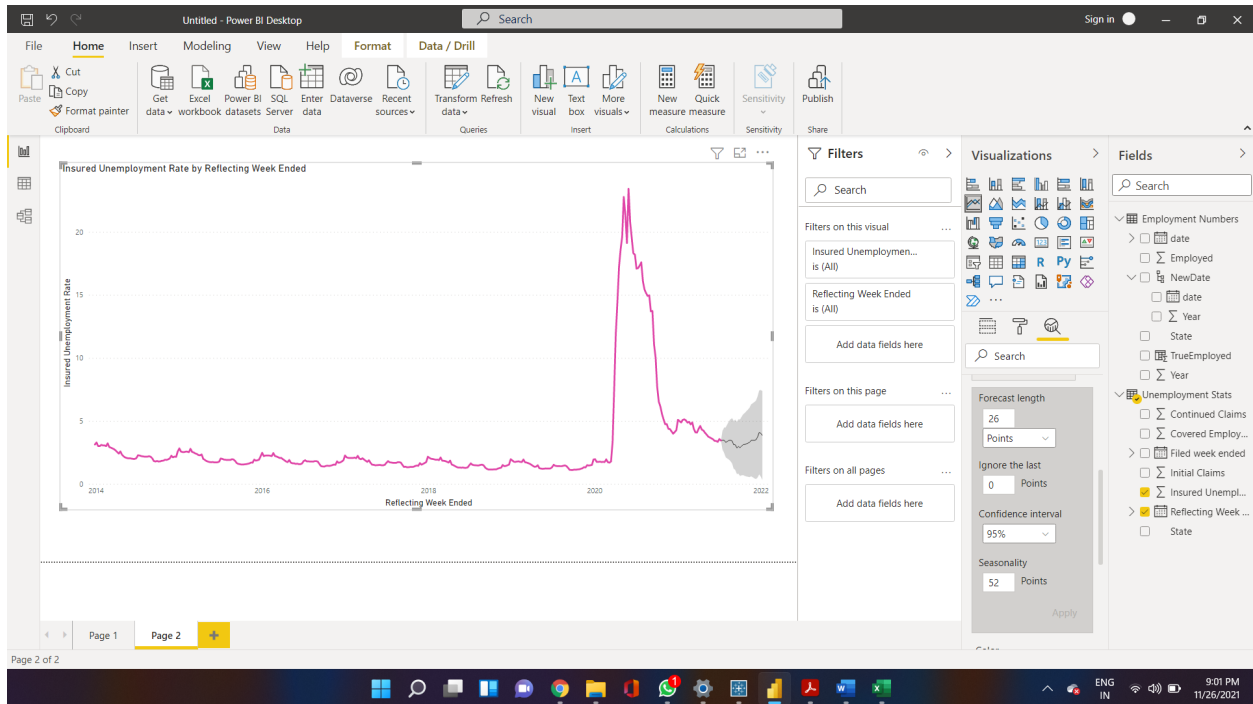
Paste a picture of the resulting line chart.



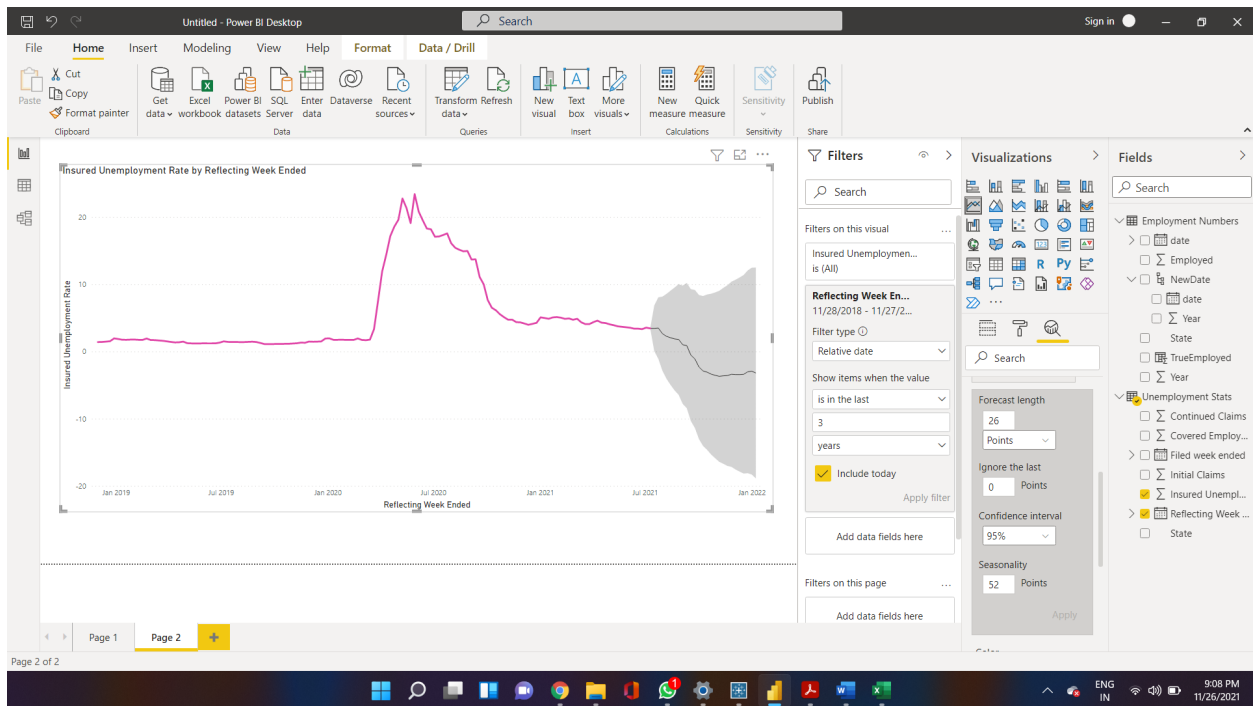
4. Change the color of the line chart to something other than blue. Paste a picture of the new line chart in your report.



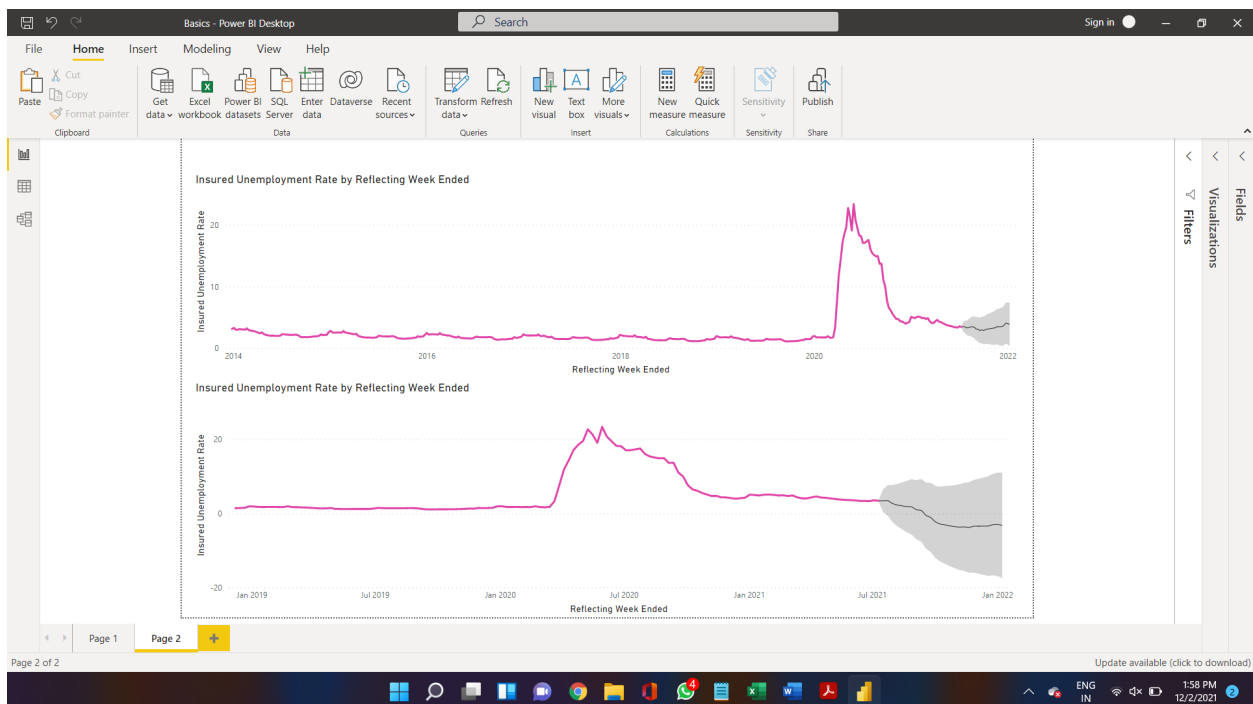
5. Create a forecast, setting the forecast length to 26 and the seasonality to 52 (this matches the data seasonality). Paste a picture of the line chart in your report.



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



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



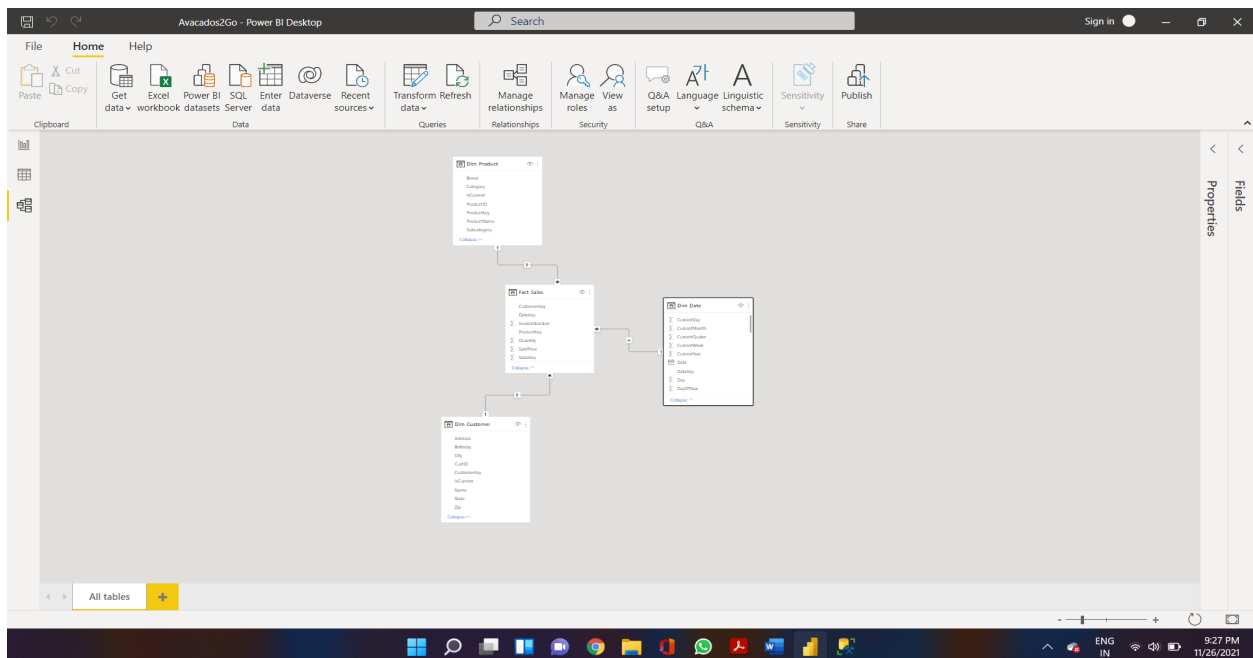
8. Write 2-4 sentences in your report detailing any insights you can draw from these two line charts.

From the above two line charts, we can infer that the employment rate started dropping from mid-March 2020 and on the first day of April there was a maximum drop in unemployment.

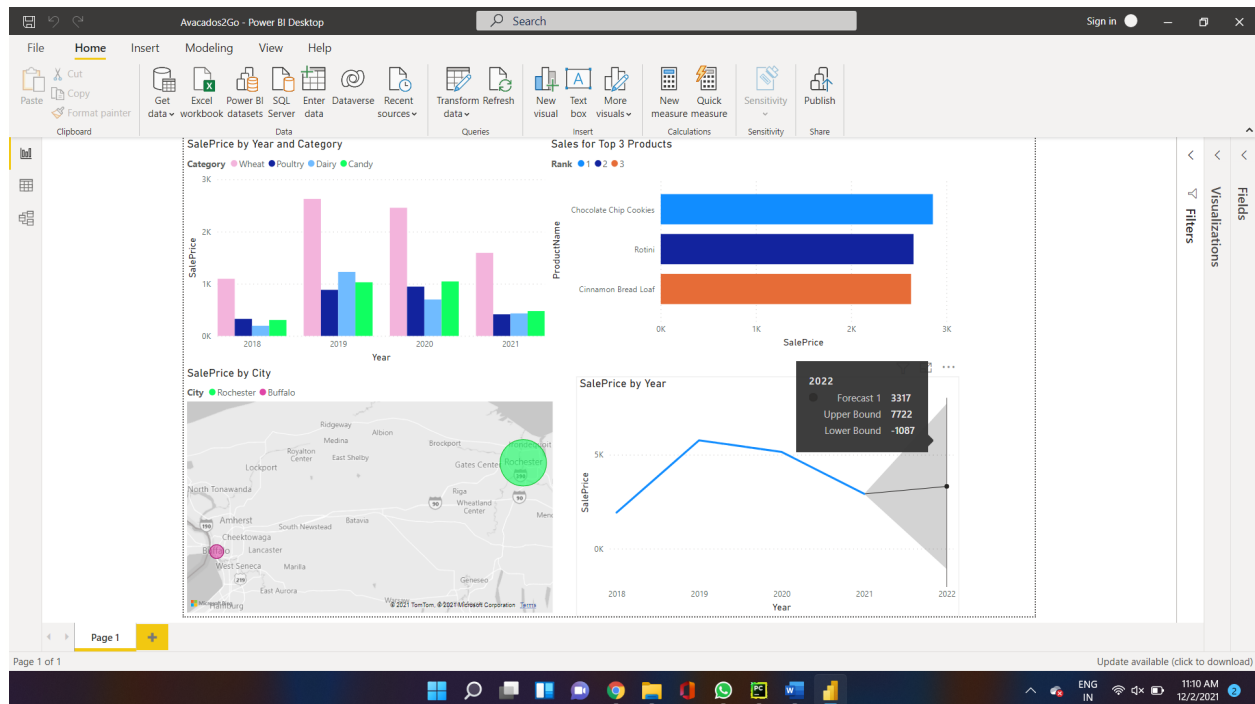
However, the market started recovering back towards the average rate after that. And it can be forecasted that the number of true employed people would increase further. Looking at the trend, there is a possibility that a drop in employment again in March 2022 can occur. As per the insured unemployment line chart, it can be inferred that the unemployment rate is directly proportional to the insured unemployment rate i.e. if the former increase, the latter also increases and vice versa.

Importing warehouse data:

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



Dashboard:



To visualize the top 3 products sold based on the sale price we used the below code to add the SumSales as a new column in the Product table for all products.

```
SumSales = CALCULATE (SUM ( Fact_Sales[SalePrice] ),'Fact_Sales'[ProductKey])
```

And using the Rank function ranked the sales and filtered the top three ranks and created a visualization for it.

```
Rank = RANKX('Dim_Product', Dim_Product[SumSales])
```

Dashboard Story:

This story-driven dashboard shows a clear pulse of the company's sales. By tracking the product performance over the years, sales by city, and forecast sales it's easy to keep track of the performance.

The category wheat has the highest sales and the top three selling products have been Cinnamon Bread Loaf, Rotini, and Chocolate Chip Cookies. Avacodes2go has witnessed the highest sales in the year 2019. The forecasted sales price for 2022 is 3317. With the historical and predictive insights from the dashboard, the company can focus on ways to increase sales and profit.