CDA 500 STDV – Homework

Sri Pavankrishna Yenugu

UB ID : 50539743

-> Take a screenshot of the table when you find “Erie County” on “1/1/1990”. Be sure to include the PercUnempr value in this screenshot.

Abc County Area	County Date	County Laborforce	County EMP	County Unemp	County Unemprate	Calculation PercUnempr
Erie County	1/1/1990	470,500	441,200	29,400	6.2000	0.062000
Essex County	1/1/1990	17,200	15,500	1,700	9.6000	0.096000
Franklin County	1/1/1990	20,500	17,700	2,800	13.7000	0.137000
Fulton County	1/1/1990	27,400	24,400	2,900	10.8000	0.108000
Genesee County	1/1/1990	31,400	29,200	2,200	7.1000	0.071000
Greene County	1/1/1990	21,500	20,300	1,100	5.3000	0.053000

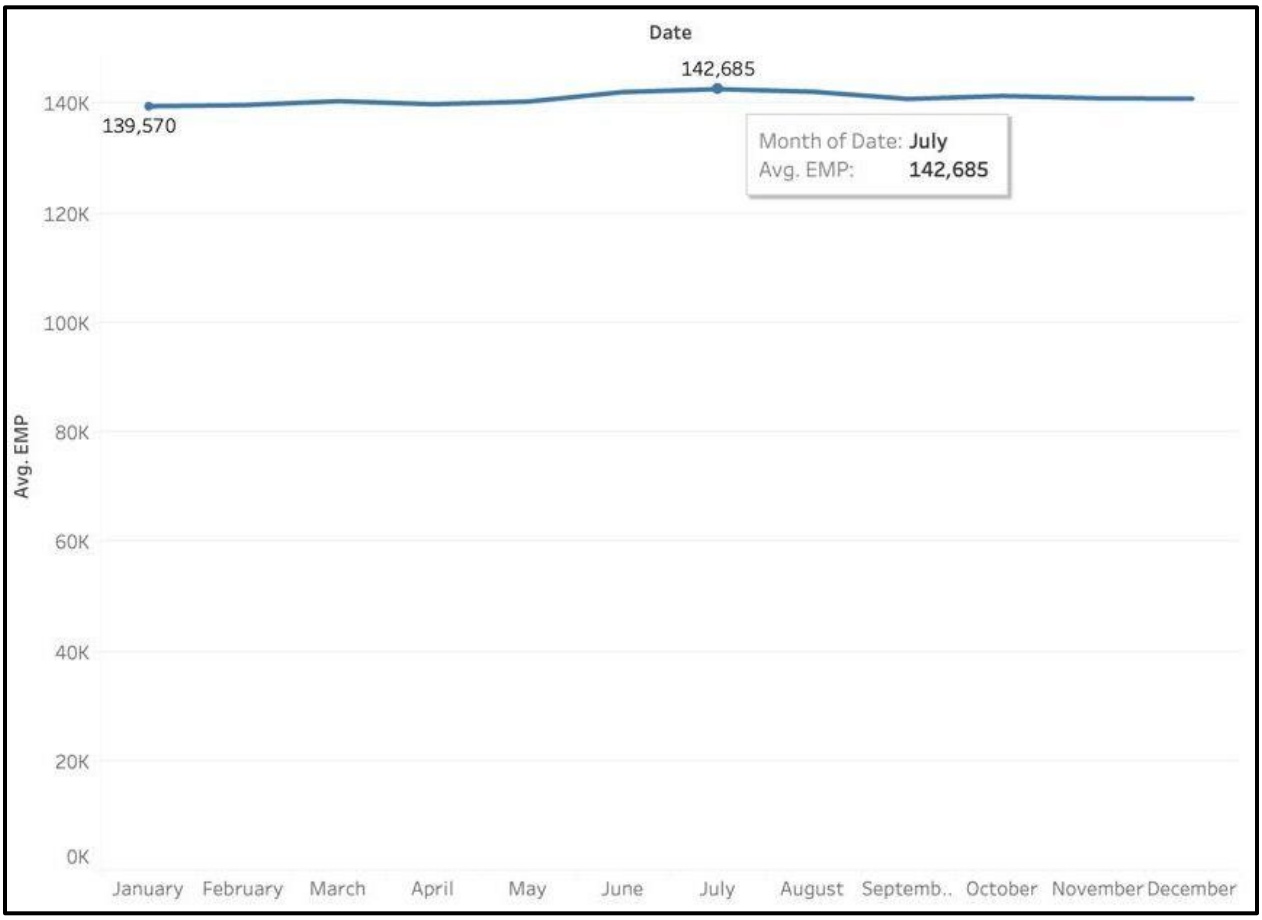
-> What is the resulting output? Why might this be bad for your data analysis?

Ans : PFB screenshot for the resulting output. It clearly shows that changing the datatype(number) from decimal to a whole number introduces inaccuracies, which can lead to errors during data analysis and visualisation. (e.g: 0.137 != 0)

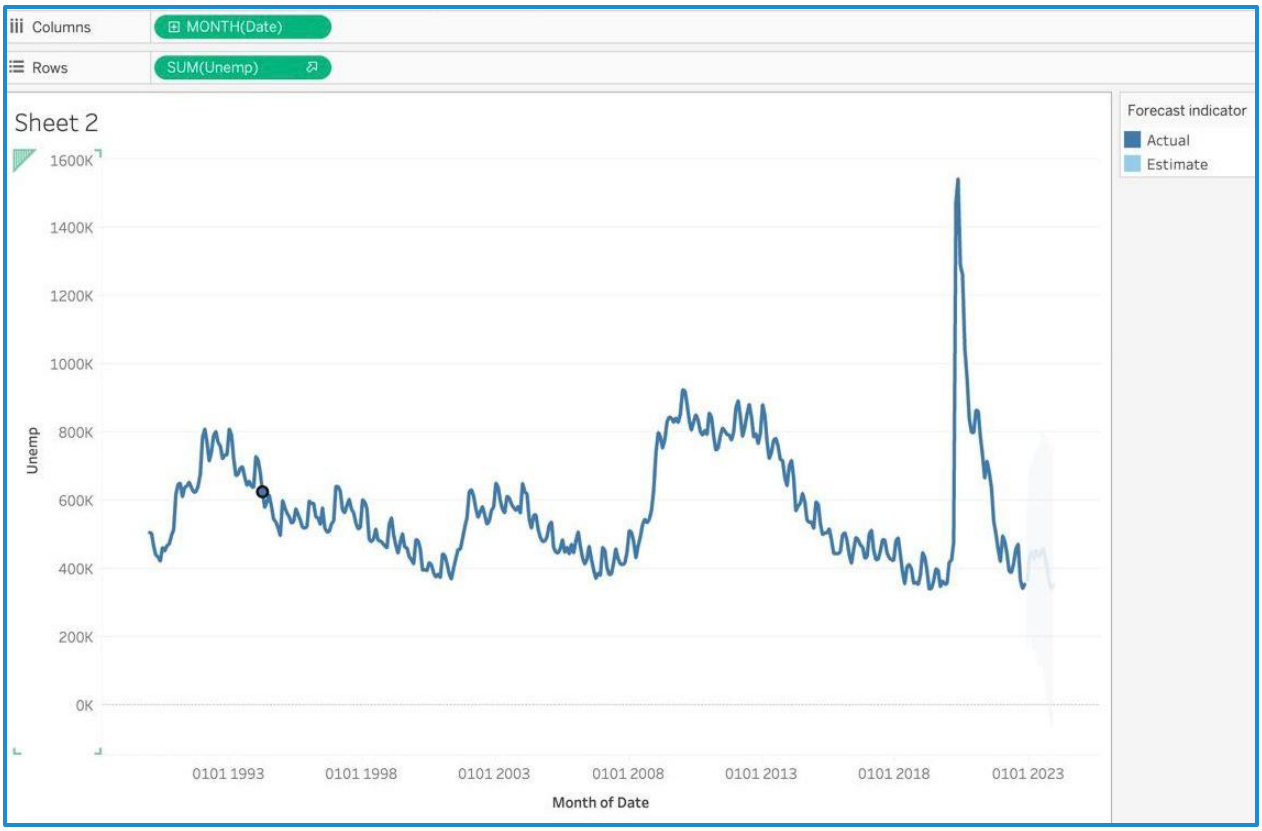
Abc County Area	County Date	County Laborforce	County EMP	County Unemp	County Unemprate	Calculation PercUnempr
Erie County	0101 1990	470,500	441,200	29,400	6.2000	0
Essex County	0101 1990	17,200	15,500	1,700	9.6000	0
Franklin County	0101 1990	20,500	17,700	2,800	13.7000	0
Fulton County	0101 1990	27,400	24,400	2,900	10.8000	0
Genesee County	0101 1990	31,400	29,200	2,200	7.1000	0
Greene County	0101 1990	21,500	20,300	1,100	5.3000	0

-> On average, which month had the highest employment numbers across the dataset?

Ans : Month of **July** had the highest average employment numbers.



-> You might be thinking “That forecast is too wide of a forecast!” and you would be right! Forecasts are directly driven by past data. If our past data has a high amount of variation, the forecasts for the future will be less predictable. What is happening in our current chart that you feel is contributing to the increase in “uncertainty” in our forecast?

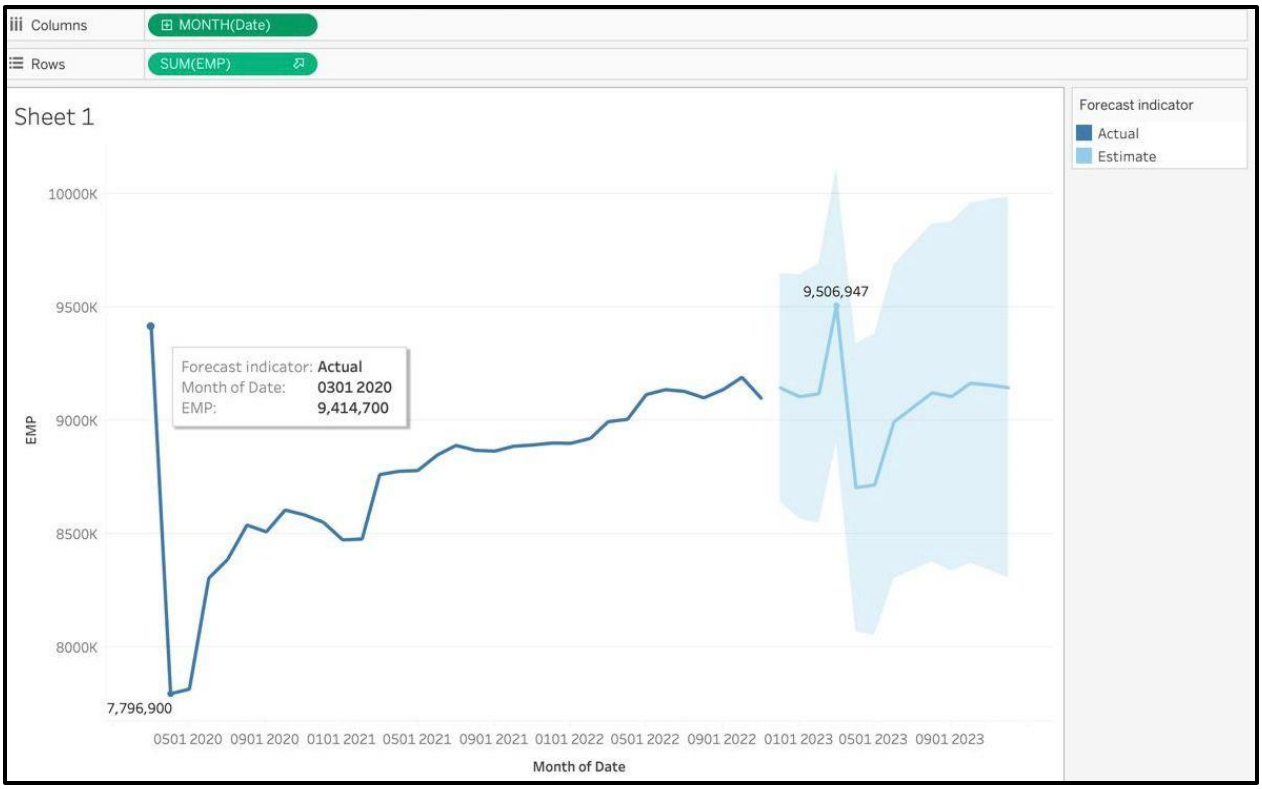


There are a couple of reasons why this might be happening based on the data presented.

High variability in historical data: The line representing the unemployment rate appears to be volatile, with sharp increases and decreases throughout the historical period. This volatility makes it difficult to predict future trends.

Limited historical data: The chart only shows a relatively short period of historical data, from 1993 to 2023. With a longer time series, forecasting models might be able to identify patterns and trends that are not apparent in this limited view and create a more accurate forecast.

-> Set your filter to happen after that event (September 2020). Does doing that increase the “certainty” of your forecast? Paste a screenshot of your forecast.

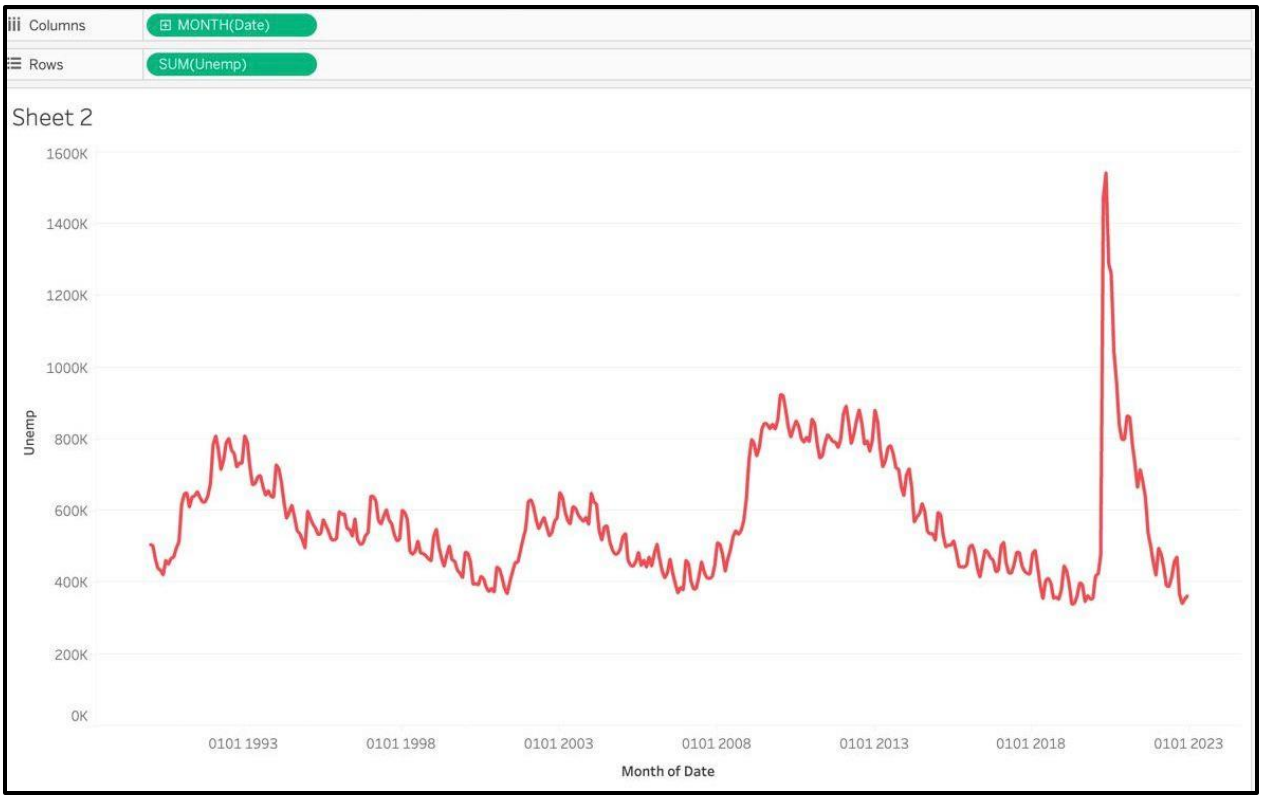


No, filtering the data after September 2020 would not necessarily increase the certainty of the forecast :

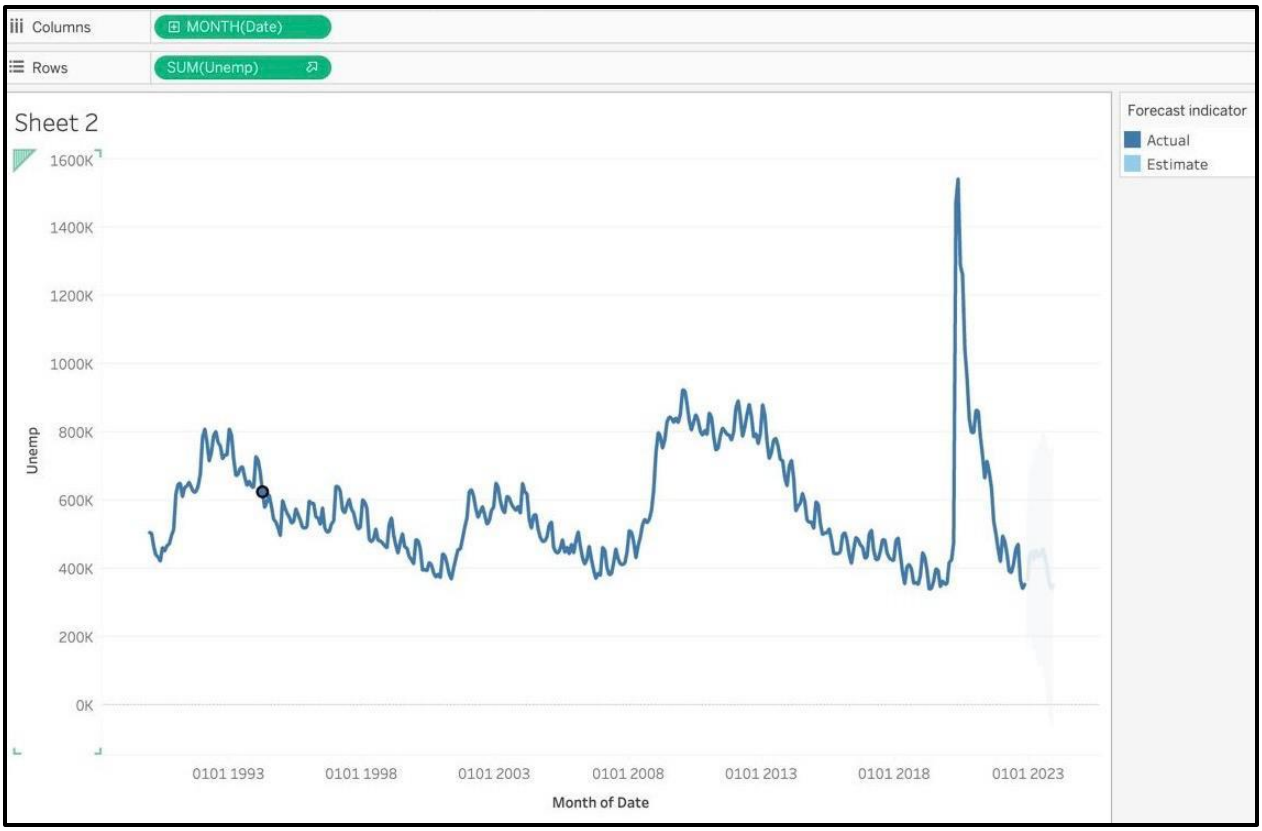
- **Limited impact on historical variability:** The image you sent shows the website traffic from January 2020 to September 2023. Even if you filter out the data after September 2020, the historical data still shows a high amount of variability. This variability makes it difficult to predict future website traffic trends regardless of the filtering.
- **Loss of valuable data:** Filtering out recent data removes information that could be useful for forecasting future trends. For instance, recent website traffic patterns might indicate seasonal changes or the impact of marketing campaigns that could influence future traffic.

In conclusion, filtering out data after a certain point might seem logical, but it can actually reduce the accuracy of your forecast by limiting the available historical data.

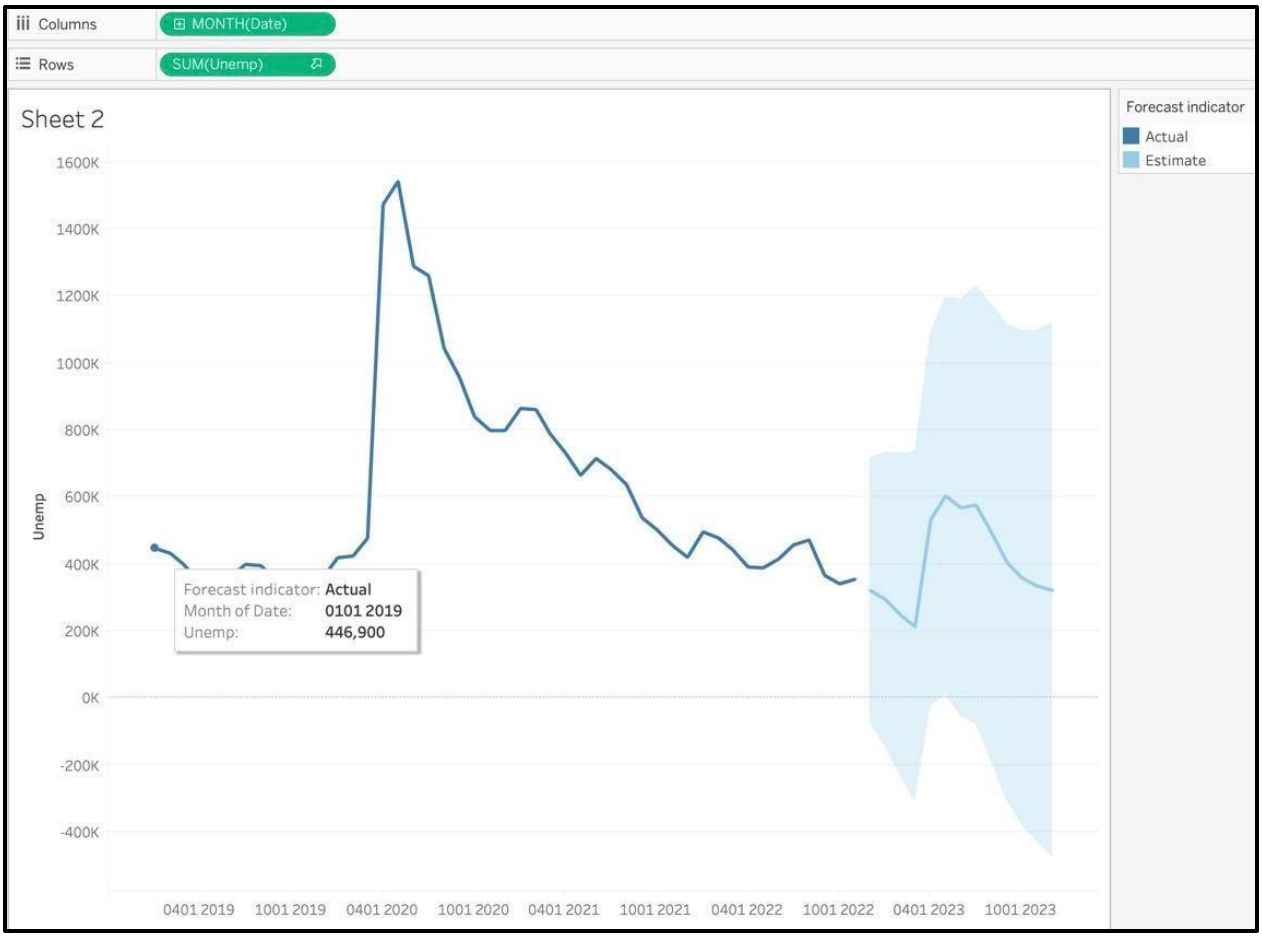
-> Now that you have the basics down, create a second sheet with a line chart using the “Unemp” field. HINT: Don’t forget to change the date hierarchy to match the time values from the previous section. Paste a picture of the resulting line chart. Change the color of the line chart to something other than blue. Paste a picture of the new line chart in your report.



-> Create a forecast, setting the model to custom and the season to additive. Paste a picture of the line chart in your report.



-> Change your line chart to filter for a Relative date showing items since 2019.

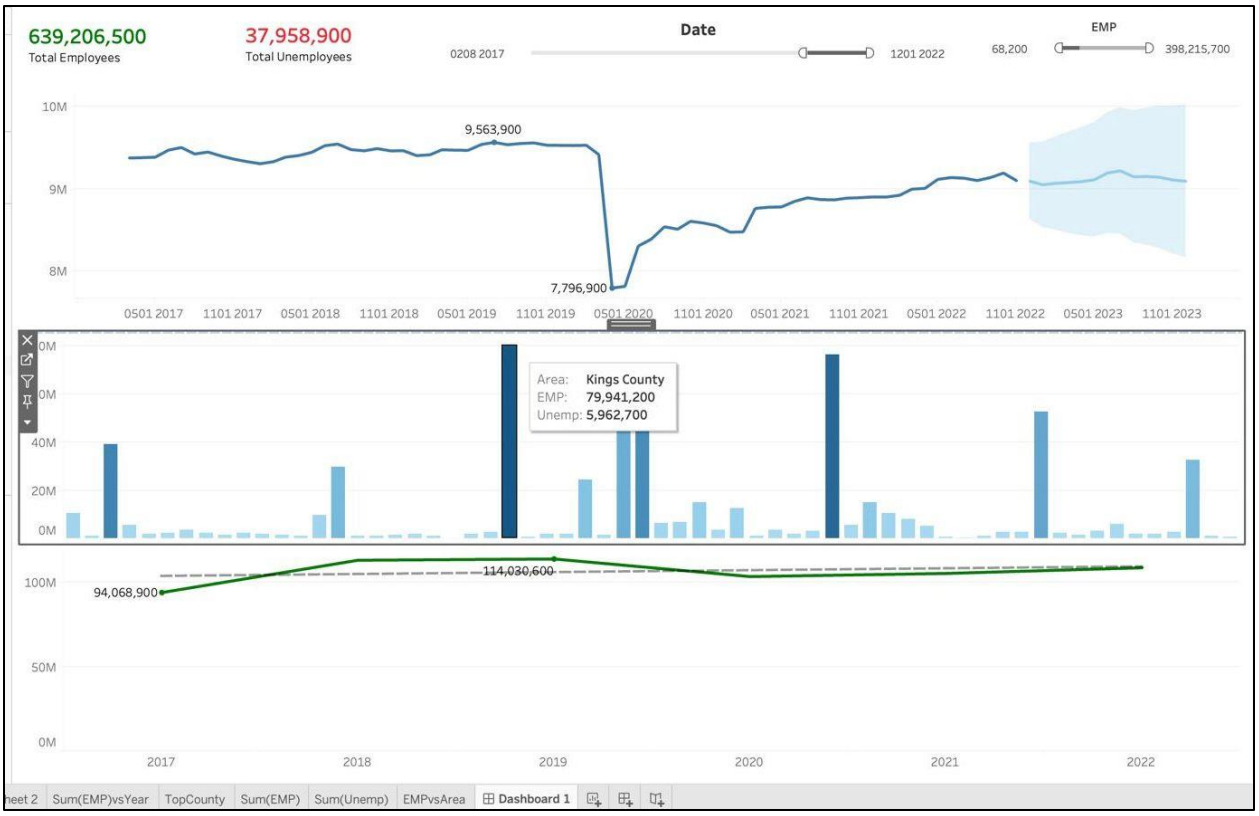


-> Create a Dashboard and add your two sheets to it in a configuration of your choosing. Paste a screenshot of this dashboard in your report. Tableau has a dashbard button on the bottom row of options, which allows you to combine multiple charts into one “dashboard” view.

The graphs reveal an interesting relationship between employment and unemployment trends in the analyzed area. The "Sum of Employees" line (decreasing sharply) suggests a decline in overall employment, while the "Sum of Unemployed" line (increasing) shows a rise in unemployment from 01/03/2020 to 01/04/2020. This indicates a potential job market shift for the given area(counties) over that particular period of time.

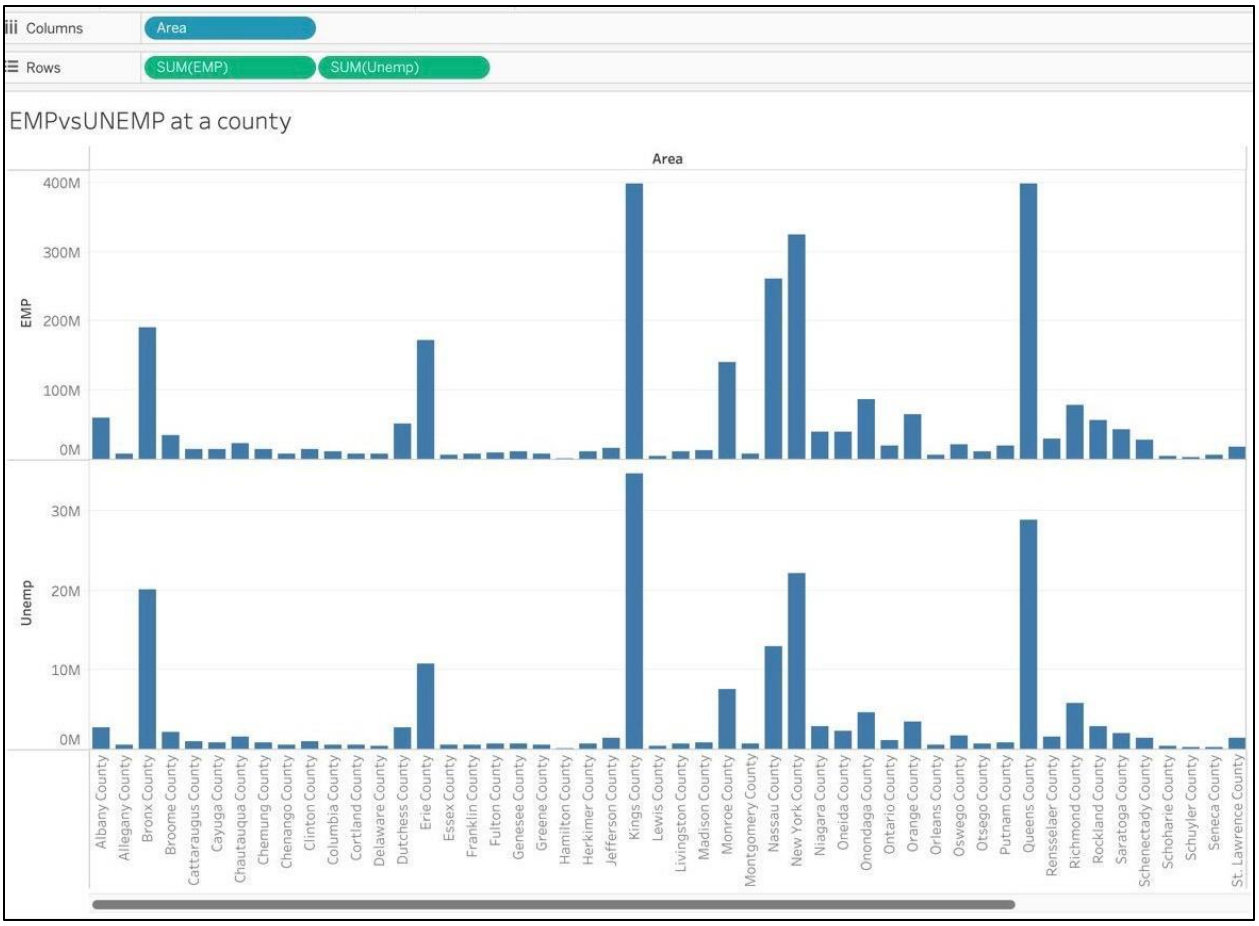
Also the lowest no.of employees (7,796,900) was observed on 01/04/2020 while the highest was recorded on 01/07/2019 with 9,563,900.

-> Using the information, you know about Tableau and the data source, create a 1-page dashboard in Tableau of the data. Your 1-page dashboard must include at least 4 distinct elements (e.g., three charts and a table).

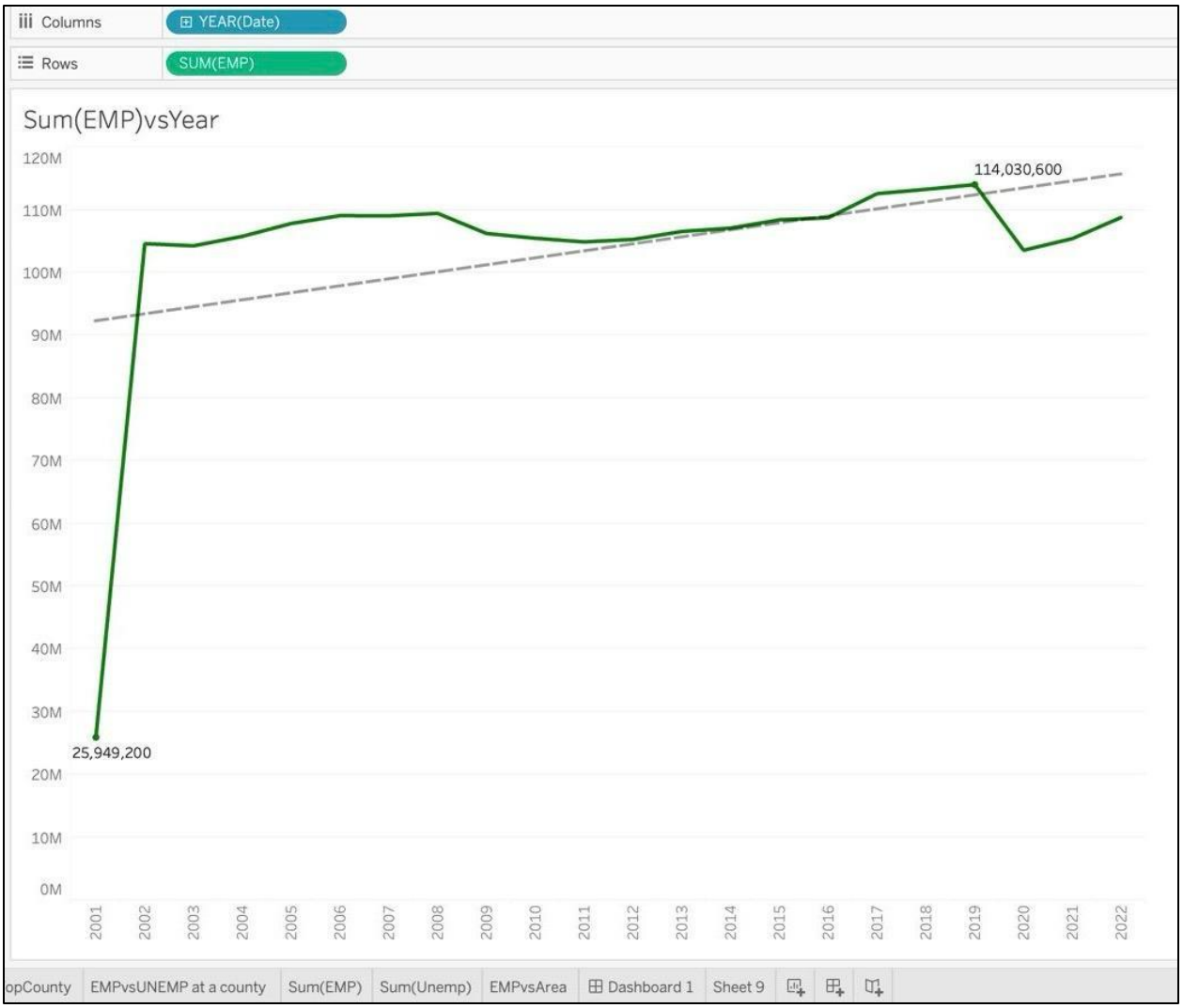


-> If you go through multiple iterations of charts, I want to see that progress! Put some screenshots of your journey in your final submission along with why you chose to change charts over time.

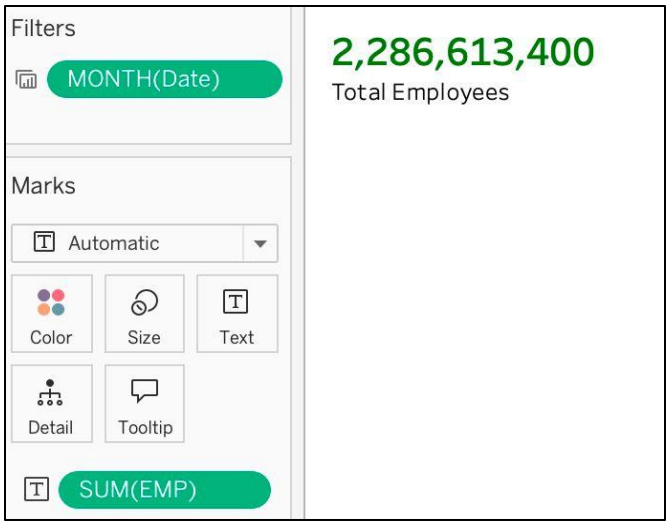
PFB Screenshots :

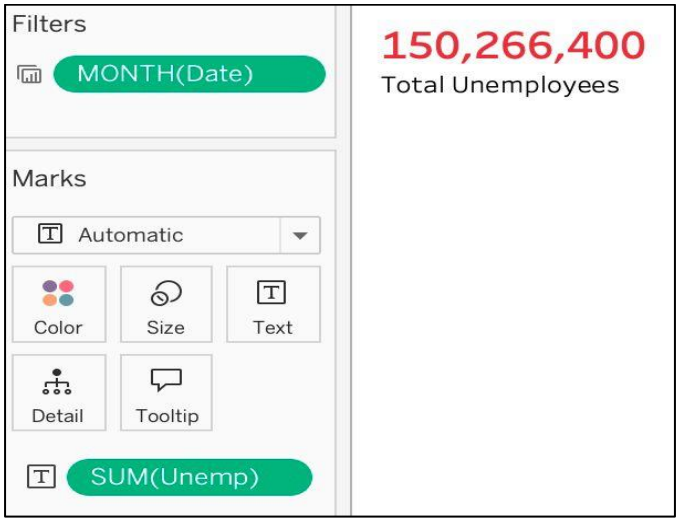

Columns
Area
Rows
SUM(EMP)
SUM(Unemp)
EMPvsUNEMP at a county
Area
EMP
400M
300M
200M
100M
0M
Unemp
30M
20M
10M
0M
Albany County
Allegany County
Bronx County
Broome County
Cattaraugus County
Cayuga County
Chautauqua County
Chemung County
Chenango County
Clinton County
Columbia County
Cortland County
Delaware County
Dutchess County
Erie County
Essex County
Franklin County
Fulton County
Genesee County
Greene County
Hamilton County
Herkimer County
Jefferson County
Kings County
Lewis County
Livingston County
Madison County
Monroe County
Montgomery County
Nassau County
New York County
Niagara County
Oneida County
Onondaga County
Ontario County
Orange County
Orleans County
Oswego County
Otsego County
Putnam County
Queens County
Rensselaer County
Richmond County
Rockland County
Saratoga County
Schenectady County
Schoharie County
Schuyler County
Seneca County
St. Lawrence County


Columns
Area
Rows
SUM(EMP)
EMPvsArea
280M
260M
240M
220M
200M
180M
160M
140M
120M
100M
80M
60M
40M
20M
0M
Area: Kings County
EMP: 273,597,300
Unemp: 22,000,100
SUM(Unemp)
44,700
22,000,100
SUM(Unemp)
44,700
22,000,100
TopCounty
EMPvsUNEMP at a county
Sum(EMP)
Sum(Unemp)
EMPvsArea
Dashboard 1
Sheet 9


Columns
YEAR(Date)
Rows
SUM(EMP)
Sum(EMP)vsYear
120M
110M
100M
90M
80M
70M
60M
50M
40M
30M
20M
10M
0M
114,030,600
25,949,200
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
opCounty
EMPvsUNEMP at a county
Sum(EMP)
Sum(Unemp)
EMPvsArea
Dashboard 1
Sheet 9


Filters
SUM(EMP)
Area
MONTH(Date)
Marks
Automatic
Color
Size
Text
Detail
Tooltip
Area
SUM(EMP)
Kings County
53,363,800
County with the most employees





Initially, I wanted to show the total number of employees versus unemployed individuals in a county for a given date. However, I struggled to implement this idea on the dashboard due to the large number of counties and extensive date samples, which made it difficult to maintain a clean and focused presentation.

Therefore, I decided to focus on showing the variance in the number of employees for a county on a given date. To achieve this, I added two single-value graphs to display the number of employees and unemployed individuals. I also incorporated a filter for the month and year, applying it to all relevant sheets to ensure consistency.

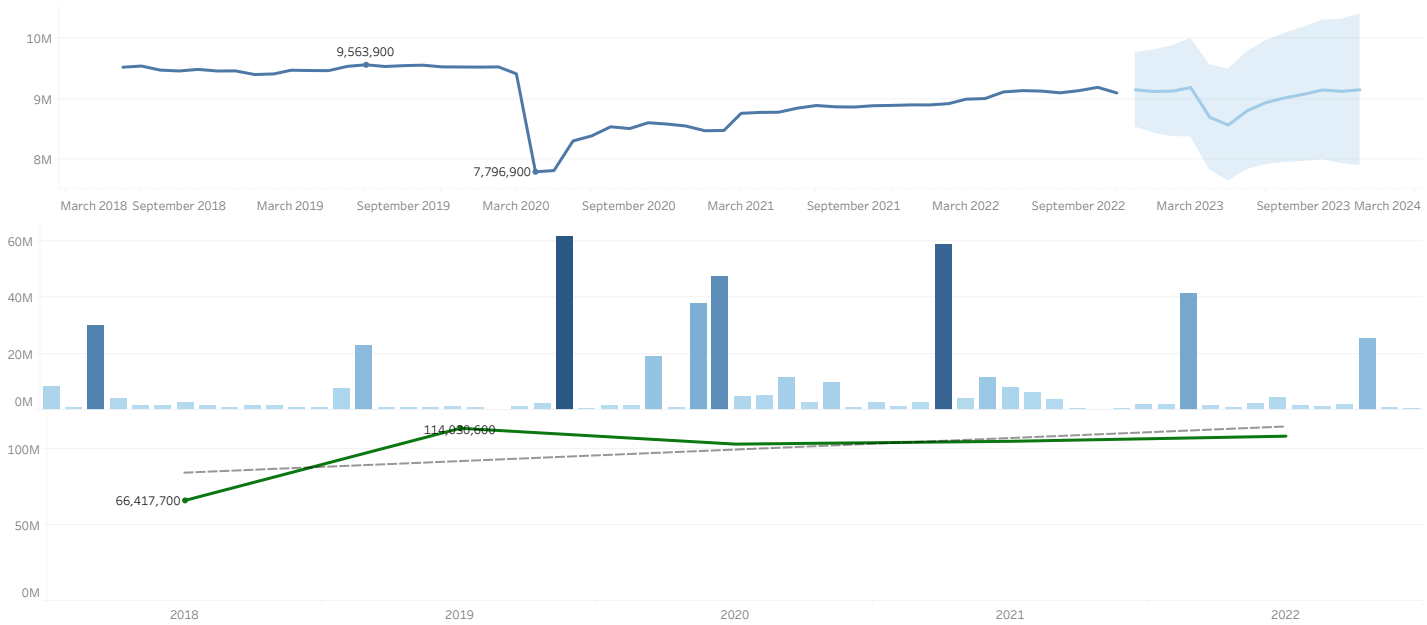
-> Paste a screenshot of your dashboard in your final report. You can also use the built in “Export to PDF” option in Tableau

498,270,800
Total Employees

31,348,700
Total Unemployees

Date
0514 2018 to 1201 2022

EMP
68,200 to 398,215,700



-> Write 1 page (double-spaced) describing your dashboard (tell your story!)

This dashboard, paints a picture of the job market in various Counties for a given period of time. The story unfolds through two primary visualizations: **line graphs**.

The first line graph, labeled "Sum(EMP)vsYear," utilizes the left Y-axis labeled "EMP" to depict the **total number of employees** in a County. The second line graph, labeled "Sum(Unemp)vsYear," relies on the right Y-axis labeled "Unemp" to showcase the **total number of unemployed people** in the county. Both graphs leverage the X-axis representing the date, with markers placed every six months, providing a clear timeline for analysis.

At the top left corner, the dashboard displays the most recent data points, likely offering a quick snapshot of the current employment situation. These values could be the **number of employees** and the **number of unemployed people** in Kings County as of May 2023. We can access those precise values in the tableau workbook by using the filters and by just clicking on the bars so that we can access the data with live visuals.

The line graphs themselves reveal a trend. "Sum(EMP)vsYear" exhibits a **general downward slope**, indicating a steady decline in the total number of employees in the Counties over the six-year period. The line starts high, possibly around 68,200 employees in May 2017, and dips to a concerningly lower value by May 2023. Conversely,"Sum(Unemp)vsYear" displays an **upward trend**, suggesting a concerning rise in unemployment. The line starts at a lower value in May 2017 and climbs to a significantly higher number by May 2023.

This stark contrast between declining employee numbers and rising unemployment paints a picture of a shifting employment landscape for the given area. It begs further investigation into the underlying causes. Are there industry changes driving this trend? Is there a broader economic slowdown impacting the job market?