Lab 1

CDA 500

Basic Business Intelligence with Tableau

Overview: Most of our insights from data are communicated through business intelligence platform. Your client has determined that their initial BI-layer tool will be Tableau. You have been tasked with presenting a proof-of-concept dashboard for the company, which addresses the NYS labor market.

Resources:

- https://www.tableau.com/academic/students#form

Your job:

- Install Tableau
- Learn how to connect Tableau to a dataset
- Create dashboard reports of current data
- Forecast values based on historical data

Requirements:

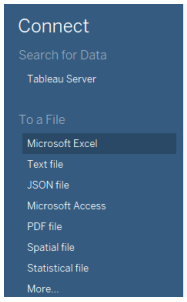
- Write a 1-2 page executive summary summarizing your assignment activities
- Provide the required responses highlighted in yellow throughout the lab in an appendix

Installing Tableau

1. Tableau is free to use for students, and can be downloaded directly from Tableau (link: https://www.tableau.com/academic/students#form)
2. Follow the standard installation prompts.

Connecting a data source

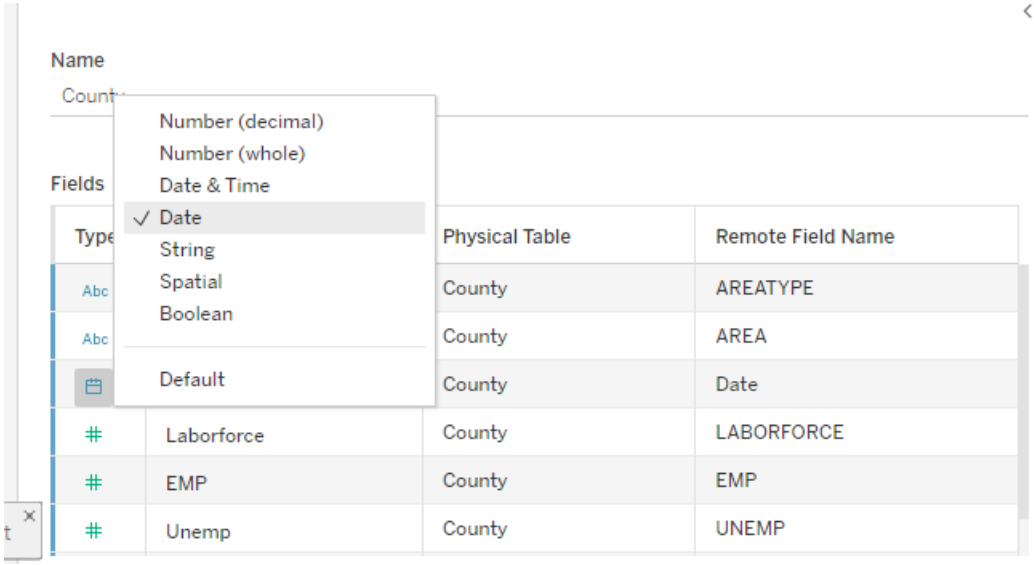
1. Open Tableau. On the left, you'll see "Connect". Under the "To a File" section, select Microsoft Excel.



2. Find and select the NYLaborStats file you downloaded from UBBox.
3. You'll notice that there are multiple sheets with the connection. County, Labor Market Area, Metropolitan Area. Drag only the County table to the main screen where it says "Drag tables here". Note that If you have multiple related sheets, Tableau will automatically create a relationship between tables.

Reviewing and modifying the data source

1. On the bottom of the screen, you have a data source view. You can use this view to examine your data source and modify it, if needed.
2. Take a look at "Date" under "Fields" on the left side of the panel. You'll notice that Type is set to Abc. This isn't what we want, as it means that it's treating "Date" like a text field, and not a "Date" field. This will become problematic if we want to plot time series data later, so we'll need to convert the data type. Click on the "Abc" button under Type for the Field Name "Date" and, from the resulting dropdown, select "Date" (not Date & Time). Tableau will automatically convert the data preview (on the right side) to show the new values.

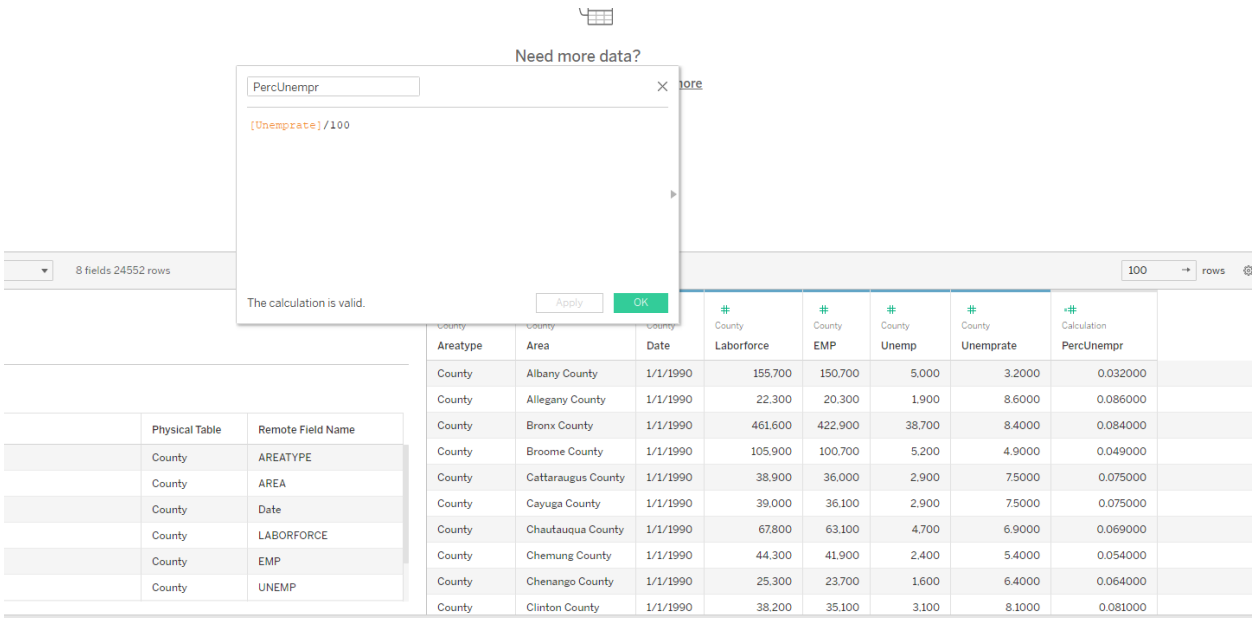


3. This still isn't quite correct though. You'll notice it changed the data format to "1/1/1990". Open the Excel spreadsheet where we're getting this data from and you'll see that it only has month and year, not day. Since we don't have the actual day in our data, we'll need to change this, but we will leave that for a little bit later.

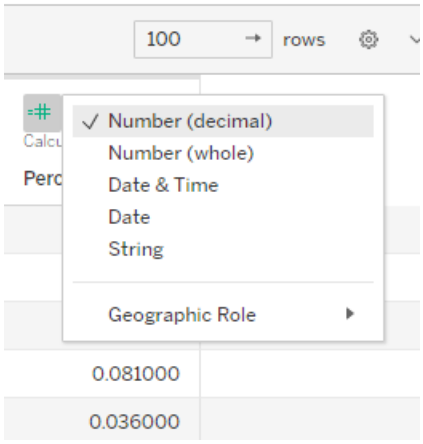
Areatype	Area	Date	Laborforce	EMP	Unemp	Unemprate
County	Albany County	1/1/1990	155,700	150,700	5,000	3.2000
County	Allegany County	1/1/1990	22,300	20,300	1,900	8.6000
County	Bronx County	1/1/1990	461,600	422,900	38,700	8.4000
County	Broome County	1/1/1990	105,900	100,700	5,200	4.9000
County	Cattaraugus County	1/1/1990	38,900	36,000	2,900	7.5000
County	Cayuga County	1/1/1990	39,000	36,100	2,900	7.5000
County	Chautauqua County	1/1/1990	67,800	63,100	4,700	6.9000
County	Chemung County	1/1/1990	44,300	41,900	2,400	5.4000
County	Chenango County	1/1/1990	25,300	23,700	1,600	6.4000
County	Clinton County	1/1/1990	38,200	35,100	3,100	8.1000

4. One other thing you'll notice is that the field Unemprate, which is supposed to be a percentage, is actually being coded as a whole number. If you add a percentage sign to this, the first value in the sheet will show up as "320%", which is (obviously) incorrect. We need to turn Unemprate into a decimal, so we can properly use it as a percentage. To do that, we can create a "Calculated Field" in Tableau. Right-click on the field in the data preview table and you will see "Create Calculated Field…" as an option. Select it.

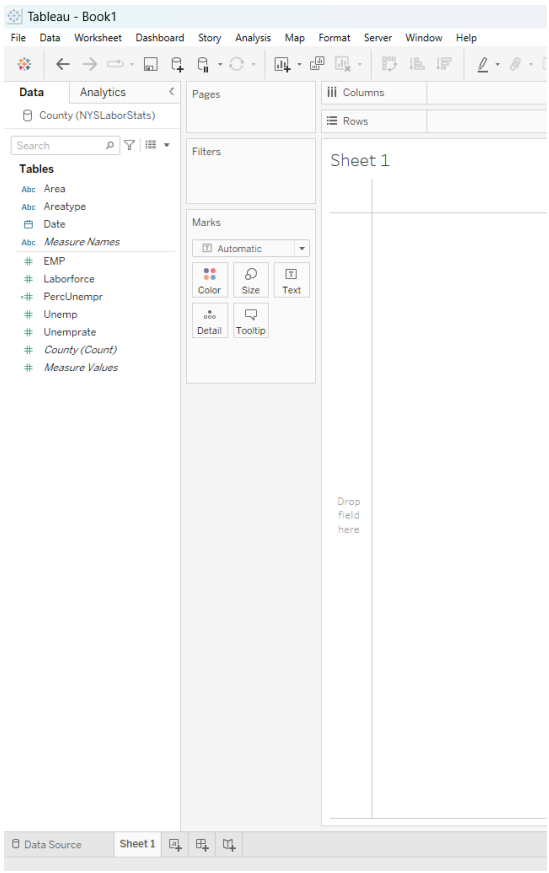
5. A new dialog box will pop up. To convert the whole number to a decimal, we just need to divide it by 100. In the box, you can change "Calculation1" to the new name of our field, which will be "PercUnempr". The calculation you will type is simply "[Unemprate]/100". Click "Apply" then "OK".



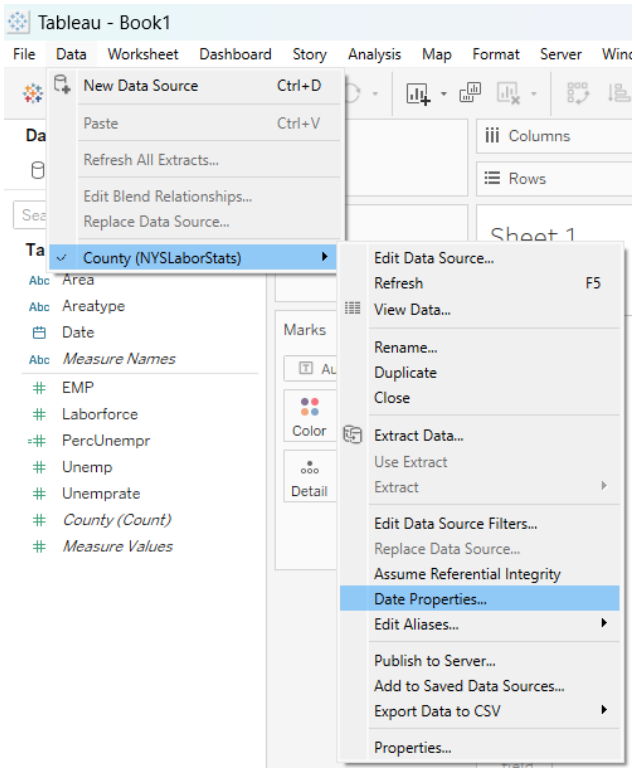
6. Scroll down the data preview table, and take a screenshot of the table when you find “Erie County” on “1/1/1990”. Be sure to include the PercUnempr value in this screenshot.
7. Excellent! We have to be careful when transforming data, though, as a small change could have drastic consequences for our data. Under PercUnempr, click on the “#” symbol and change the data type to “Number (whole)”.



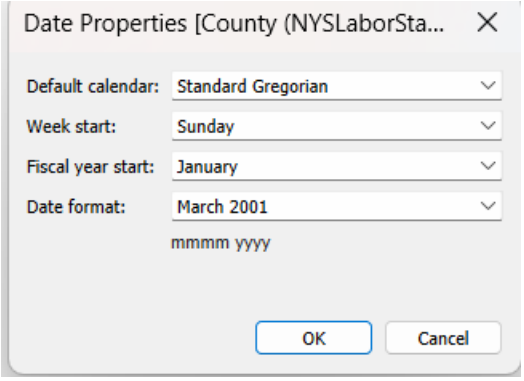
8. What is the resulting output? Why might this be bad for your data analysis?
9. To undo this change, you can once again click on the “#” button and select “Number (decimal) and you values will return to their original state.
10. We still have to fix our Date field. In order to do that, select “Sheet 1” from the bottom corner of your screen to exit the data source view and enter into the “dashboarding” view.



11. On the top toolbar, select Data->County(NYSLaborStats)->Date Properties…



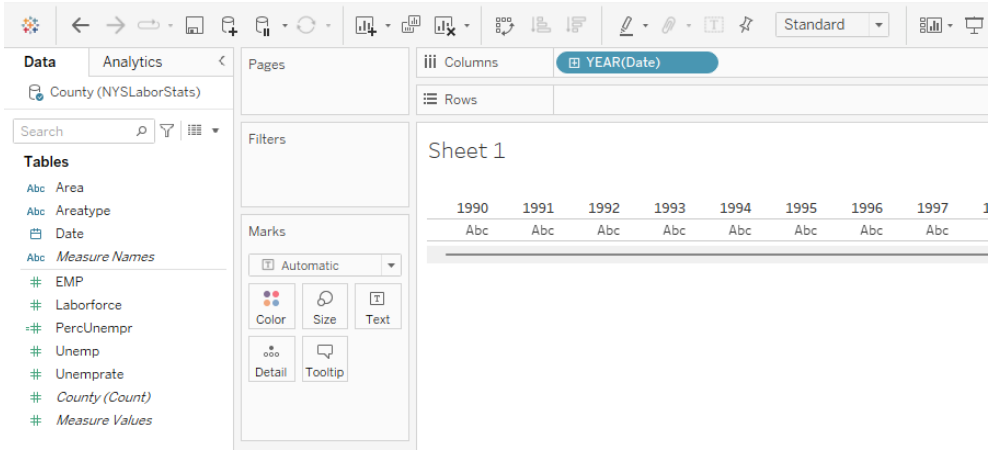
12. From here, we can see a variety of options, including important indicators like fiscal year start. For now, we want to leave everything the same except for “Date format” which we will change to the mmmm yyyy option (the example is “March 2001”).



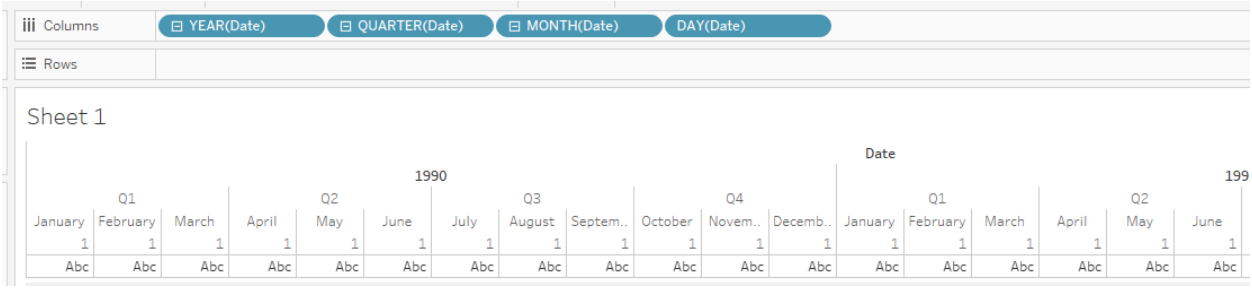
13. Click OK. Navigate back to the “Data Source” tab on the bottom of the screen. You will see that the dates have all been changed in the Data Preview to be Month Year. If not, revisit the above steps to try again.

Building reports and dashboards

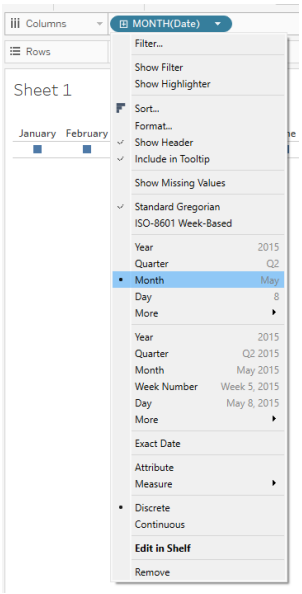
1. Open your dashboard worksheet view again by clicking on “Sheet 1” in the bottom corner of your screen.
2. Tableau is reminiscent of Excel in this way. You can also add more pages (or rename existing pages) as needed.
3. Examining the “Fields” section on the left, you’ll see each of the imported sheets from the data source and their accompanying fields.
4. Let’s see how hierarchies are denoted. One great thing about Tableau is that it automates a lot of things. For example, the common hierarchy **date** is usually Year->Quarter->Month->Day. Tableau automatically creates this hierarchy when we have a field denoted as a date. Let’s see how this works. Under Tables, double click on the field “Date”. Tableau will automatically add it to your visualization.



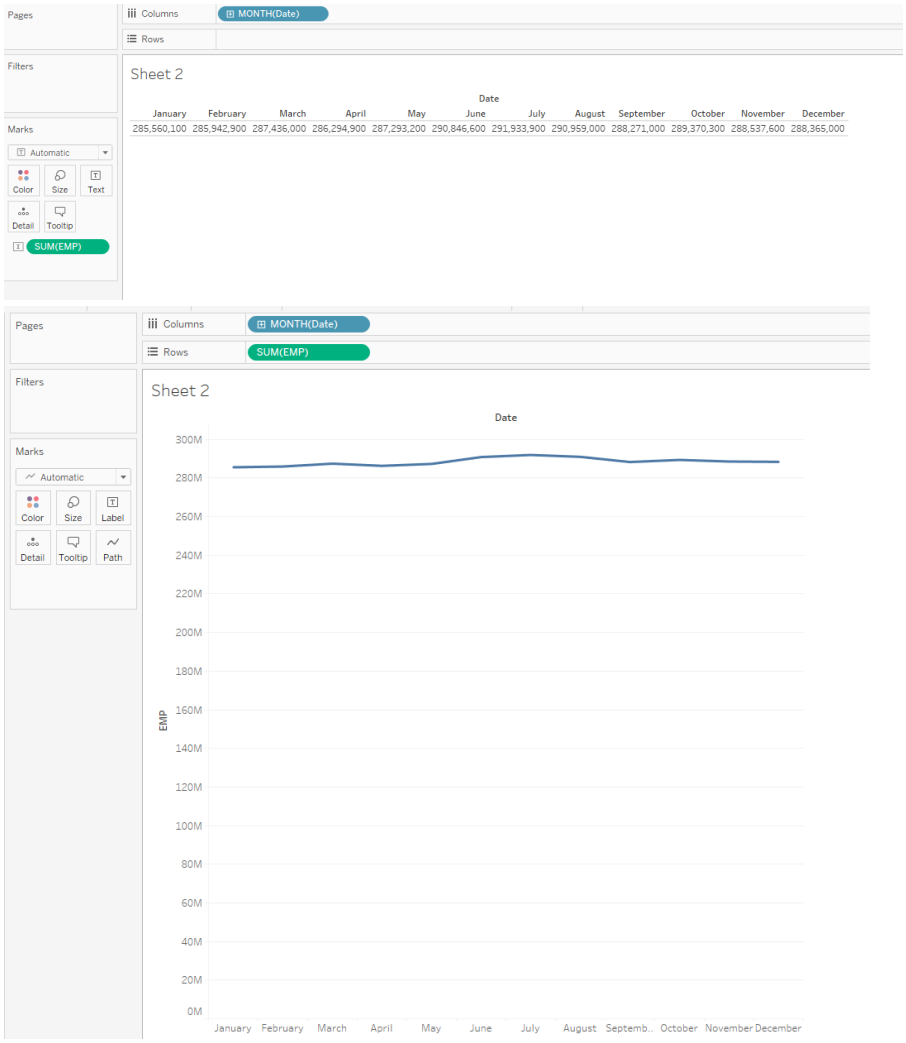
5. You'll notice, though, that Tableau adds "YEAR" as the date value. We can expand this out by clicking the + symbol next to YEAR(Date). Go ahead and click it, and you will see the next level on the hierarchy. If you keep doing this, you'll get the entire hierarchy.



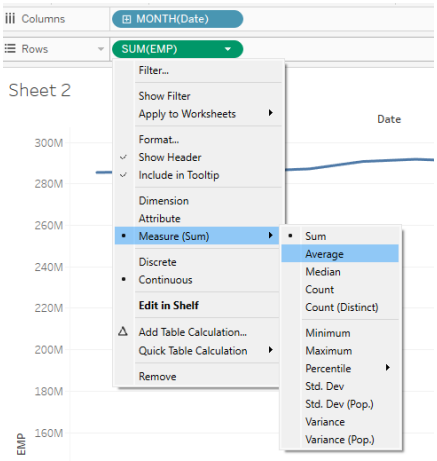
6. Let's go ahead and remove every one except for YEAR(Date). You can click on them and remove them or CTRL-Z your way back. If you click the arrow on the YEAR(Date) item, you will get a dropdown menu where you can change the properties and format of this item. Select the first Month option from this list.



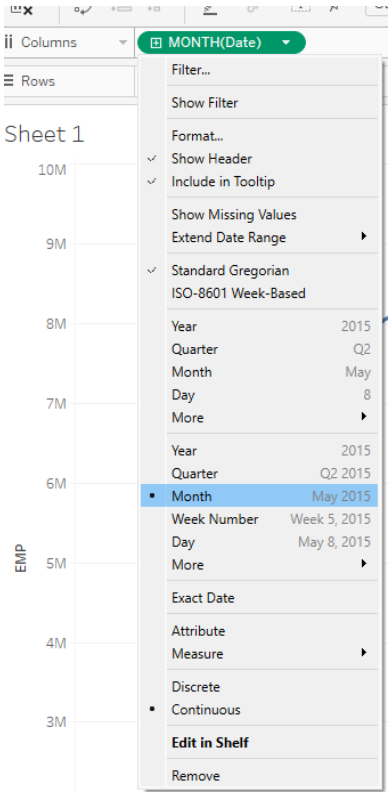
7. Let's create a chart. Double click on "EMP" under fields on the left side, and it will add that into your workspace. Depending on your Tableau version, it may create a table instead of adding the EMP field to the "rows" section (see screenshot below). If this happens, just undo what you just did (CTRL-Z or CMD-Z on Mac) and click and drag the EMP field over to "Rows".



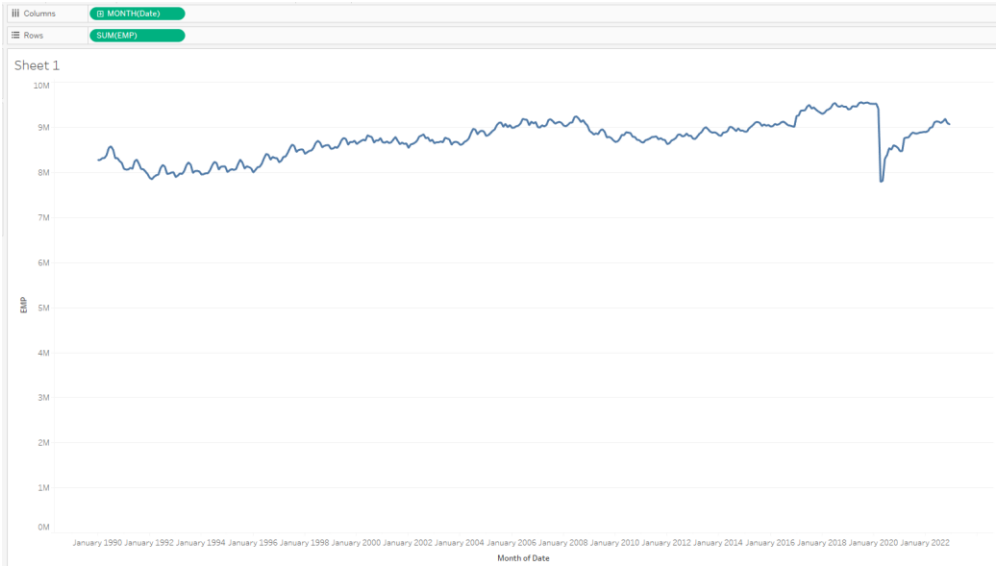
8. This resulting chart is showing you the sum of all employment numbers across the entire dataset, broken down by month. We can change the “SUM(EMP)” value to average. Hover over it in the “Rows” field of your workspace and click the arrow on the right-hand side of the field. Under Measure select Average.



9. We'll learn how to make these charts more aesthetically pleasing in a bit, but for now, hover over the different points on the chart. On average, which month had the highest employment numbers across the dataset?
10. Let's go ahead and change AVG(EMP) back to sum (same as above, click the arrow, measure, then change it to Sum). Then, we want to change the "Month(Date)" value to be "Month Year". Click the dropdown arrow on Month and select the *second* month in the list this time (it will say May 2015 as the example).

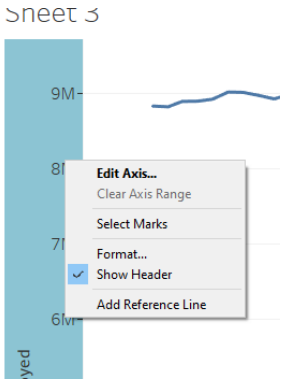


11. Tableau should automatically create a line chart of the results.

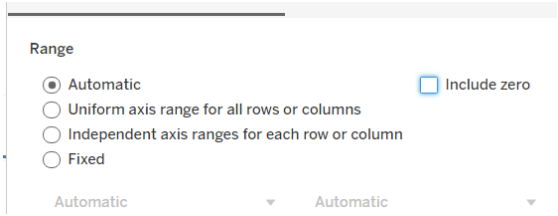


12. If your chart does not look like this, check your columns and rows (seen in green above). Particularly, if the Rows section does not show SUM(EMP) you'll need to change that by using the dropdown arrow (like we did with the year-to-month change above). If your chart is not a line chart, you can change the chart format by clicking the "Show Me" button in the upper right corner and selecting another chart type.

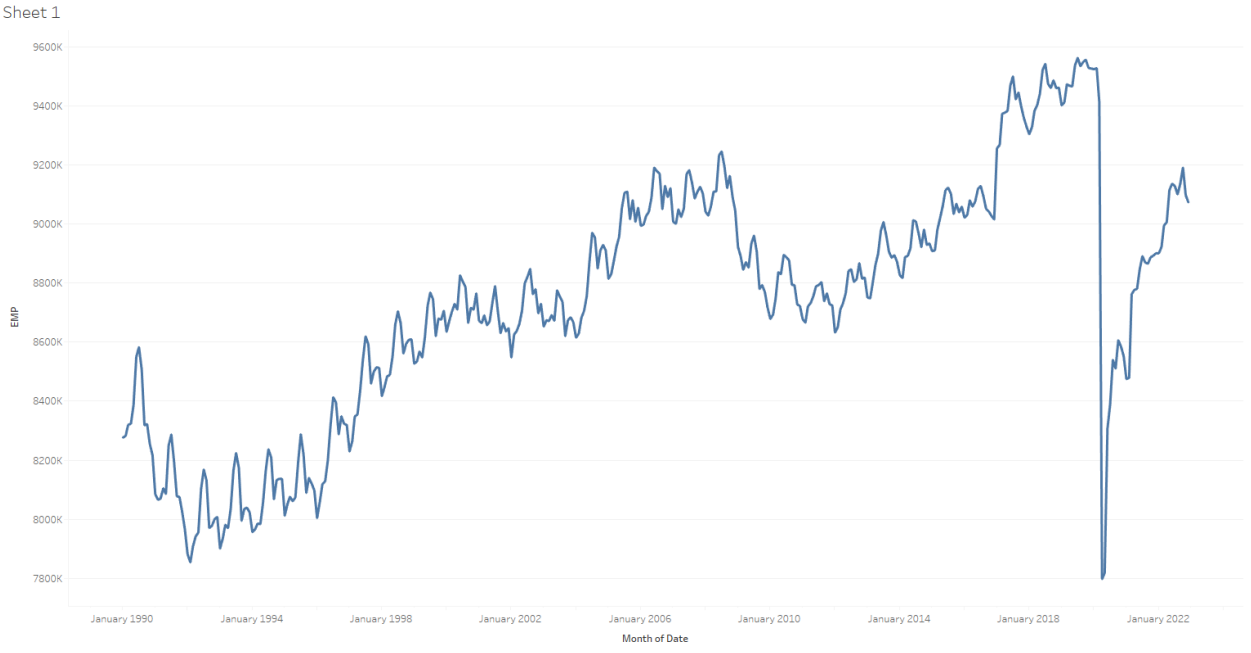
13. You'll notice that this line chart doesn't dip below 7 million. This is fine, but sometimes we may wish to realign our axes to better define the range of values present in the data. Right click on your Y-axis on the chart and select "Edit Axis".



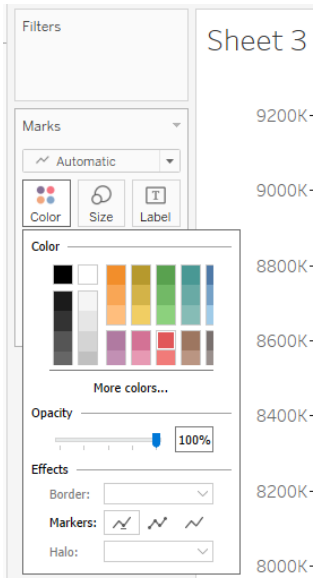
14. Uncheck the box that says "Include zero".



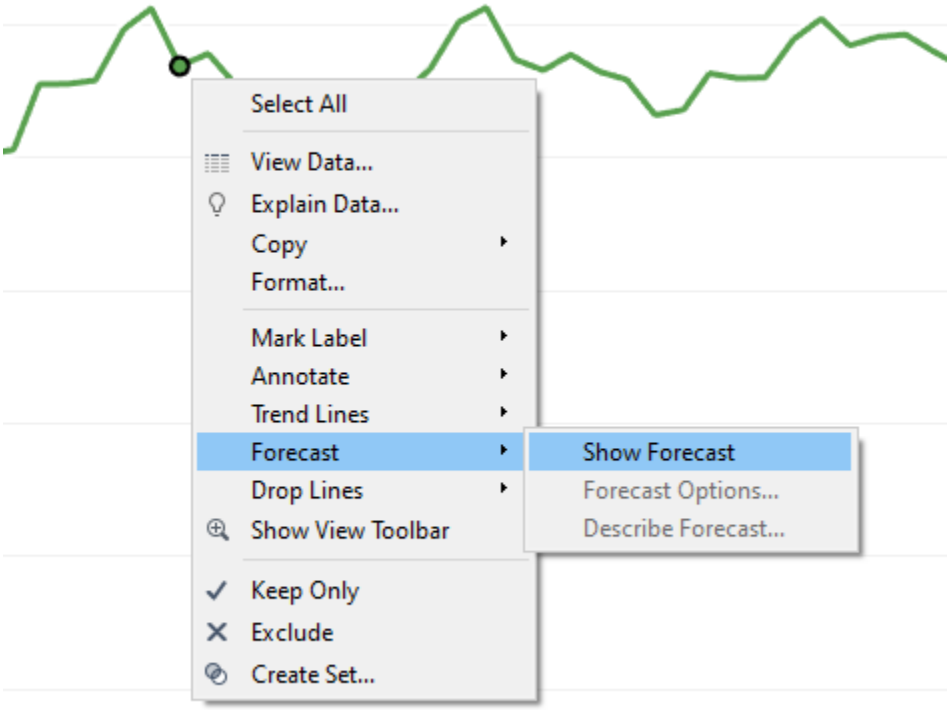
15. Your chart is now normalized for the specific range of values present in the source data. Be careful with this, as having different axis ranges can cause confusion or mislead viewers when displaying multiple charts in a dashboard.



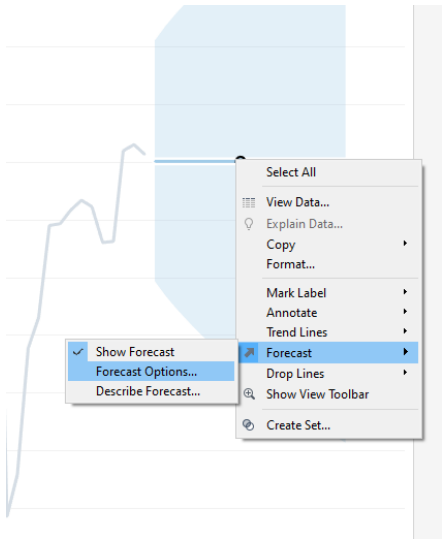
16. Maybe we don't want our line to be blue. Let's go ahead and change that. To the left of your line chart, click the "Color" button. Select any color you wish.



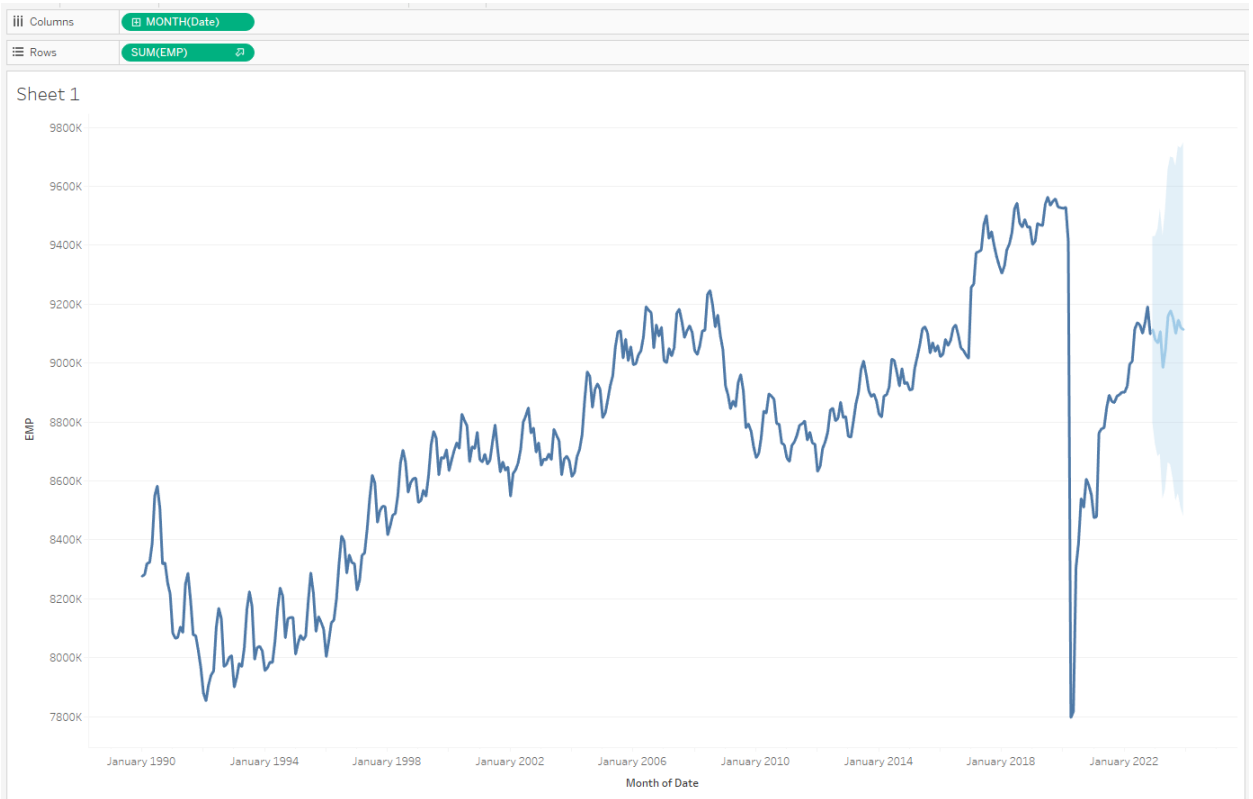
17. What if we want to ***forecast*** out values for the future. Right click on your line, and under forecast select “Show Forecast”.



18. This forecast is boring, and that’s because our model is not really fit for the data. Let’s go ahead and change that. Right-click on the forecast line and under “Forecast” select “Forecast Options”.

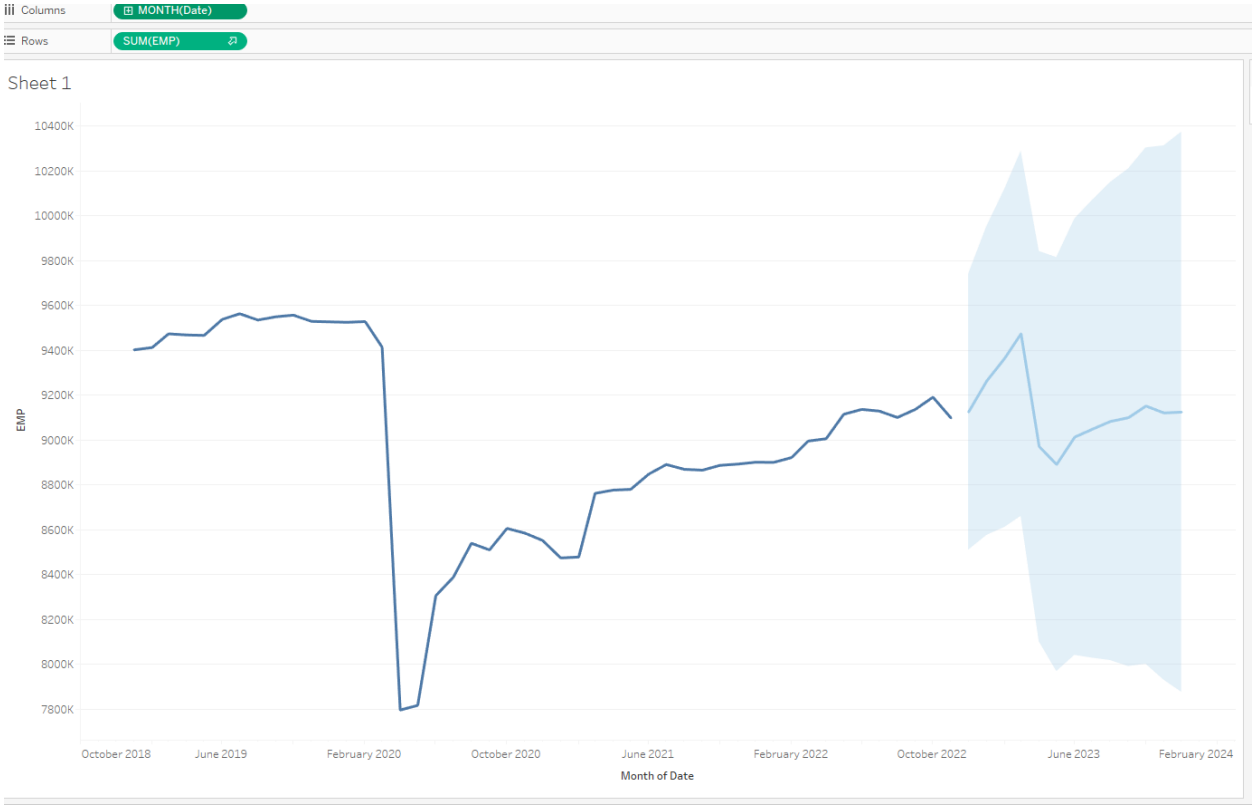


19. Under Forecast Model, let's change the Automatic forecast to a "Custom" one. Change "Trend" to None and Season to "Additive" and click OK. **Read about forecasting models here:** https://www.artofvisualization.com/blog/pitfalls-and-secrets-of-tableau-forecasting

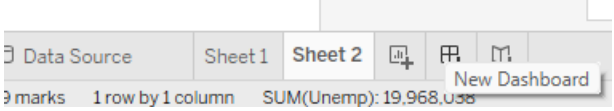


20. Your line should now reflect a more accurate forecast that is more aligned with the data.

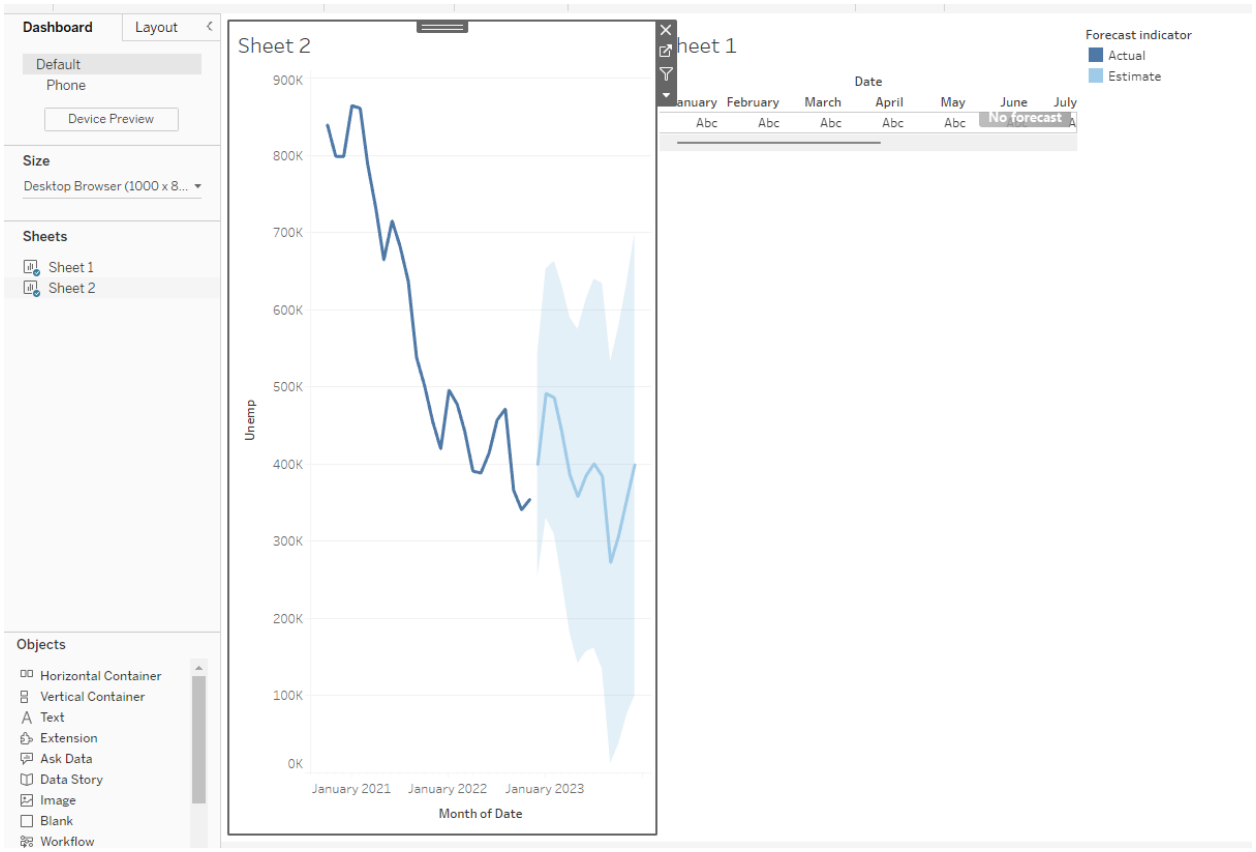
21. Right click on MONTH(Date) and select “Filter…”. Let’s filter the line chart to only focus on 2019 onward. Change the start year to reflect “1/1/2017” and click OK. You’ll now see a more condensed chart focusing on the past few years.



22. 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? 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.
23. Congratulations! You’ve done some data cleaning and preparation, analysis, and forecasting. It’s important to note that in Tableau, you cannot put multiple charts on the same sheet. Each sheet has its own chart/visual and we combine them into a dashboard using the “dashboard button at the bottom of the main screen.



24. In the dashboard view, you can drag and drop your sheets and format/organize them to your liking.



Trying it on your own

1. 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.

2. Change the color of the line chart to something other than blue. Paste a picture of the new line chart in your report.
3. Create a forecast, setting the model to custom and the season to additive. Paste a picture of the line chart in your report.
4. Change your line chart to filter for a **Relative date** showing items **since 2019**.
5. 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.

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

Creating a dashboard

Your client wants to assess the viability of the job market across New York state, particularly considering the Erie County region as compared to statewide statistics. The client wants to understand what the data for 2016-2022 ***looked like*** for Erie County and what the 2022-2024 looks like, including any forecasted pitfalls in the employment statistics. The client wants more than just a simple "number of employed" and would like you to provide some insights into unemployment rates across various counties and where Erie county has ranked over this time period.

- 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). It can include any visuals you would like (it does not have to include the two sample line charts above), but ***must include*** at least one line chart with a forecast. You can use any of the sheets in the provided Excel file to help you build your dashboard. As well, you can find out more about this data from the NYS Labor Statistics website: https://dol.ny.gov/labor-force-and-unemployment-data-technical-notes and https://dol.ny.gov/new-york-state-geography
    - Hint: Make your dashboard clean and focused. A good dashboard focuses on telling a specific story about the data that may be useful for business decisions. Use tables, bar charts, line charts, etc. to tell your story!
- You must leverage the design principles we talked about in class (e.g., Gestalt principles). If you're reading this on the first day of class, we'll get to these, that's why the homework has a due date near the end of the course. 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.
- Paste a screenshot of your dashboard in your final report. You can also use the built in "Export to PDF" option in Tableau
- Write 1 page (double-spaced) describing your dashboard (tell your story!).

Appendix A: Complete deliverable requirements

NOTE: The below are the same requirements found throughout the assignment. There are no additional requirements. You should use this as a checklist for the assignment deliverables.

- Scroll down the data preview table, and 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.
- What is the resulting output? Why might this be bad for your data analysis?
- We'll learn how to make these charts more aesthetically pleasing in a bit, but for now, hover over the different points on the chart. On average, which month had the highest employment numbers across the dataset?
- 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? 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.
- 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.
- Write 2-4 sentences in your report detailing any insights you can draw from these two line charts.
- 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). It can include any visuals you would like (it

does not have to include the two sample line charts above), but ***must include*** at least one line chart with a forecast. You can use any of the sheets in the provided Excel file to help you build your dashboard. As well, you can find out more about this data from the NYS Labor Statistics website: https://dol.ny.gov/labor-force-and-unemployment-data-technical-notes and https://dol.ny.gov/new-york-state-geography

    - Hint: Make your dashboard clean and focused. A good dashboard focuses on telling a specific story about the data that may be useful for business decisions. Use tables, bar charts, line charts, etc. to tell your story!

- You must leverage the design principles we talked about in class (e.g., Gestalt principles). If you're reading this on the first day of class, we'll get to these, that's why the homework has a due date near the end of the course. 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.
- Paste a screenshot of your dashboard in your final report. You can also use the built in "Export to PDF" option in Tableau
- Write 1 page (double-spaced) describing your dashboard (tell your story!).