Matthew McCutchen

ISMN 5650

Final Project

1. The two datasets I chose were the “Water Consumption in the City” and “Drinking Water Quality Distribution Monitoring Data” datasets from the NYC open datasets. I thought that these two sets would go well together when creating the data and is also something I think would be interesting to look at. I think their relation could be seen just up front, but if more things are found in the water, then consumption might be down, and vice versa. Water is a huge part of any city or living place and seeing the quality and how that might impact total consumption is important, as that seems to be a bit more of a problem in a lot of bigger cities like NYC. Knowing what might be in your water is helpful, and this can be insightful into other environmental aspects/issues that could have impact on it. Seeing what is found in the water over time could help point in the direction of those very issues.
2. Comments on notebook.
3. I think this project has helped a lot with my personal knowledge of python and pandas. Being able to take data and have to clean it and merge it without it being on a kind of guideline feels close to something I might have to do outside of school and getting that experience this early I think can be really helpful. More to the coding, the biggest thing to me was being able to convert various data columns into others and using that as a basis to merge. Unlike in DMBS, where you just make a bunch of keys and call it a day, using the same data but being able to convert it without a bunch of extra clutter was interesting. I also took the time to dive into Seaborn with my graphs and see how many different types of graphs there are with a bunch of different ways to customize how the data is visualized.