Andrea Khoury

Draft your approach to the project using the analyst’s workflow, including what questions you need to answer at each stage

-Analyst workflow-

* **Understanding Core Question-** Can it be proven that those public schools not in the best areas of Buffalo, have low graduation rates. The different parts of the city and suburbs of Buffalo have high and low crime rates, along with multiple schools in each district. Do the high crime areas have a lower graduation rate? If so, money needs to be put into the Buffalo Public school systems to create and environment for kids to want to learn. Will the investment in the schools create a higher socioeconomic standard and will it increase the graduation rate in the schools most effected?
* **Data Ingest-** The first place I looked when beginning this project was at the Buffalo Open Data website. I figured out my topic through this because you may have an idea of what you want but the data may not be there. On there I found Buffalo Crime data, and Buffalo Public school data. The Public-school data was not enough for me to make a conclusion or story about, so I looked elsewhere. Then data.nysed.gov gave a lot of information on not just Buffalo data but all of New York State. **Question?** What sources are credible? Will this data be useful/support my idea for this project?
* **Data Preparation**- From the different databases I then started to fish out what information I am going to need for this project. After saving the sources, applying them to a Jupyter notebook helps see what data is in the data sets and how it is going to support my idea. **Questions?** What am I trying to get out of this data?
* **Basic Data Analysis-** The Buffalo Crime data had information such as Zip codes, type of crime, where and when they took place and so much more. Buffalo Public schools from Buffalo open data website gave me information such as how many schools are in all zip codes in Buffalo, what kind of school they are like, elementary, middle or high school. The .gov database gave me information such as, school district, averages of graduation rates, counties, grades, dropouts and so much more than can lead to more interesting things in my project overall. **Questions?** What data would be beneficial to me? How do they connect with on another to make conclusions?
* **Preparation of Results-** The results that I collected form the data have all been put into a Jupyter notebook and I extracted each part of the whole data set that I felt was important to my project. **Questions?** Is this enough data? Do I need more data to fit my conclusion? Do I fully understand the data and what it’s telling me?
* **Reproducibility-** Each data database has similarities where the three sets can connect in a way to one another. I have attempted to merge two of the data sets bases on the same zip codes because they were both in Buffalo. Now with this third set I am finding schools in this area and seeing their graduation rate compared to the amount of crime done in the same area. **Questions?** How are the different data sources relatable?
* **Presentation-** The presentation on my data will be on a Jupyter Notebook and it will display my findings from each data source that is fitting to my project. The written report might be included on the notebook or on a separate word document them I will make a PowerPoint presentation for the class of my research findings. **Questions?** Does the final presentation of the data and report cover everything in the instructions/rubric for the project? Is it interesting and appealing to the eye?