|  |  |  |
| --- | --- | --- |
|  | Anthony Olivieri  Syracuse iSchool  Applied Data Science  SUID: 370774094  aolivier@syr.edu | |
|  | |  |

Portfolio

# PROJECTS

## Data Visualization

The main goal of this project was to convey meaningful insights through visual communication. Data as raw numbers or text requires the audience to stop, focus, concentrate, and possibly calculate values to derive meaning. To express that value quicker and in a more memorable fashion the data can be arranged in symbols and colors. Humans can evaluate colors and symbols –sizes, positions, patterns—in a preattentive manner, meaning communication is instant and more pleasing to the audience.

**Technologies**: R, ggplot, Illustrator

**Techniques**: data mining, descriptive analytics, grouping and aggregation, plotting, data transformation, illustration, color theory

**Ethical consideration**: The charts on this poster can only convey insights into the data as it is: raw arrest records from Pittsburgh City Police. There is no way to display the effects of generational poverty or policing patterns.

**Project document**: [PDF Poster](https://github.com/oxenfree/Syracuse-iSchool-Data-Science-Portfolio/blob/master/Projects/Data-Vizualization/Pittsburgh-Crime-Stats-Poster.pdf)

## National Real Estate Analysis

This project aims to solve a financial investment problem. Given an investment opportunity in real estate, what three zip codes would yield the best return on that investment. This is obviously a prediction problem and because the data is median home values per zip code reported each month, it’s also a time-series problem. Time-series analysis offers interesting statistical challenges such transforming data to achieve stationarity and testing stationarity using the Dicky-Fuller test as well as testing for auto-correlation. The main model used for time-series, ARIMA, was not used in this project in favor of a new library by Facebook called Prophet.

**Technologies**: python, jupyter notebook, Facebook Prophet, pyplot, seaborn, basemap

**Techniques**: data mining, plotting, data transformation, predictive analytics, machine learning, time-series, test statistics

**Ethical consideration**: Real estate investment using market indicators or demographic data may worsen economic conditions for residents of that area.

**Project Document**: [Jupyter Notebook](https://github.com/oxenfree/Syracuse-iSchool-Data-Science-Portfolio/blob/master/Projects/National-Real-Estate-Analysis/main.ipynb)

1. Describe a broad overview of the major practice areas in data science.

Learning the application domain:

Knowing how data can be represented:

Seeing the big picture of a complex system:

Data transformation and analysis:

Attention to quality:

Ethical reasoning:

1. Collect and organize data.
2. Identify patterns in data via visualization, statistical analysis, and data mining.
3. Develop alternative strategies based on the data.
4. Develop a plan of action to implement the business decisions derived from the analyses.
5. Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization.
6. Synthesize the ethical dimensions of data science practice (e.g., privacy).