Introduction:

A new family is currently living in the Bushwick Neighborhood of Brooklyn, New York. However, upon the expansion of their family through the birth of two children, they are seeking to return to Upstate, New York where they originally hail. The couple would like to live in a neighborhood that is similar in amenities to their current neighborhood (Bushwick), and is currently investigating Troy, Schenectady, and Albany, New York as potentially suitable communities.

Audience:

This analysis would apply to any individual also interested in relocating from the New York City to the tri-city area in Upstate, NY, and provides a point of reference for comparing the cities of Upstate, NY to the neighborhoods of New York City.

Data Sources:

- **City Geolocation Data:** Geolocation data for New York Cities' 5 boroughs, Albany, Schenectady, and Troy, NY used to identify, map, and search for venue information using the Foursquare Application Programming Interface (API)
- **New York City Neighborhood Data:** A list of all NYC neighborhoods and Boroughs acquired from the following linked location: https://cocl.us/new_york_dataset
- **Shops/Venues:** Foursquare API Data for all identified cities and neighborhoods.
    - Venue Name and Category by Neighborhood in New York City, and by City for all Upstate, NY Venues.
        - All venues were searched within 2000 meters of Upstate, NY cities (due to size of the neighborhood), and within 1Km of Bushwick due to the size of Bushwick
- **Housing Price/Cost Information:** Mean Single Family Housing price/cost information from Zillow data, obtained from the following linked location: https://www.zillow.com/research/data/
    - Monthly data from 1996 through 2020. Data cleaned and transformed from horizontal to vertical structure for analysis.
    - Housing data filtered based on Bushwick, Albany, Schenectady, Troy zip-codes, using zip-codes.com to locate all zip-codes associated with the neighborhood/city being analyzed.
    - Federal regional cost of living adjustment data – obtained from open source data on the Federal Housing and Finance Agency (FHFA) available at: https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx

Methodology and Results:

To compare the neighborhood of Bushwick in Brooklyn, NY with Albany, Schenectady, and Troy, NY, the first step that was taken was to clean and transform the venue data from Foursquare based on Category Type (e.g. Bar, American Food, etc.…). The method used to compare these neighborhoods was an unclassified machine learning algorithm, K-Means clustering. The K-Means

algorithm was executed against the mean count of Venue Category for each Neighborhood in Brooklyn, NY and Albany, Schenectady, and Troy, NY. After an initial K-Means cluster run determined that Bushwick was most similar to all three Upstate, NY cities, the following additional clustering variables were also reviewed to determine if any particular city would distinguish itself more closely to Bushwick, NY. All combinations of the below # of Clusters and # of Top Venues were Analyzed.

Number of Clusters	**# of Top Venues**
3	5, 10, 15, 20, 25
6	5, 10, 15, 20, 25
9	5, 10, 15, 20, 25
12	5, 10, 15, 20, 25
15	5, 10, 15, 20, 25

In all instances of the above table, Bushwick, and all three Upstate, NY cities were repeatedly clustered together, resulting in a lack of the ability to distinguish Upstate, NY city as more comparable to Bushwick than another. As a result, a further analysis was conducted to determine if one city may be more desirable than the others based on home values.

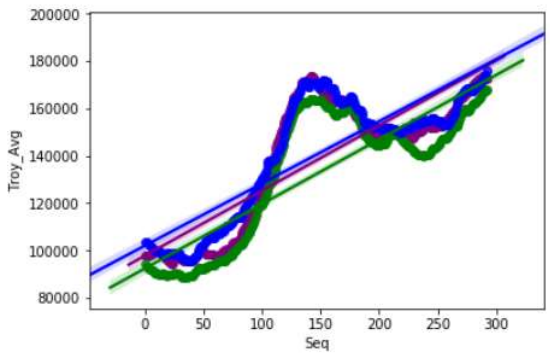
To execute the home value analysis, Median Home Value prices were parsed by month from the Zillow Homes Data source. The resulting dataframe, provided all median home values by month since 1996 by Zip-Code. The dataframe was subset to include only the zip codes associated with our neighborhoods of interest, Bushwick, Albany, Schenectady, and Troy. This data was then compared by taking the average home value for all zip codes associated with each neighborhood and normalizing the data to the cost of living adjustment of Bushwick vs. Metropolitan Albany, NY. Based on the Federal Housing and Finance Agency (FHFA), this meant all Upstate, NY city home values had to be transformed by 315%.

Average Home Prices				
Date	Bushwick	Troy	Schenectady	Albany
2020-04-30	$846,830	$553,754	$528,666	$543,415

As shown, the results of this analysis do not show a *clearly* preferable option between the three Upstate, NY cities based solely on their adjusted average home price. So, the question arose, is there a significant difference in the value of a home in one of the three cities over time? To examine whether or not this was true, a linear regression of average home values for each city was executed, and produced results with R^2 values of ~.68 for all 3 cities, and upon visual inspection, the slope of each line is not significantly different from one another.

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
0.6870479600179096
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
         normalize=False)
0.6819788974041504

]: <matplotlib.axes._subplots.AxesSubplot at 0x7fdb380d7780>
```



Due to the inability to distinguish cities based upon their apparent home values over time, the next step was to compare the three cities of Upstate, NY against one another to determine whether or not the home prices were significantly different from each other. As the underlying dataset is not provided, and the distribution of that dataset is unknown, the best method for determining significance between each set of home values was deemed to be a Mann-Whitney U test. This statistical analysis was produced by creating an array of home values for all zip codes in Albany, Schenectady, and Troy, NY, and then running the MannWhitneyU test to solve whether or not any of the home prices were significantly different from one another. The test was only run to compare Albany and Schenectady, and Troy and Schenectady, as these city pairs have the largest difference between them compared to Albany and Troy, NY alone. The null hypotheses being tested were:

- Null Hypothesis 1: The home prices of Albany, NY are the same as Schenectady, NY
- Null Hypothesis 2: The home prices of Troy, NY are the same as Schenectady, NY

Results were as follows:

```
Statistics=3.00000000, p=0.40848066
Same distribution (fail to reject H0)
Given failure to reject the Null Hypothesis, it is assumed the values of homes in Albany vs. Schenectady Homes are the same.
Statistics=2.00000000, p=0.34926768
Same distribution (fail to reject H0)
Given failure to reject the Null Hypothesis, it is assumed the values of homes in Troy vs. Schenectady Homes are the same.
```

As shown in the screen print above, in both cases of the Whitney-Mann U test, it could not be concluded that the Null Hypothesis could be rejected in either the Albany nor the Troy cases, meaning that the home prices are statistically the same in all three cities.

Discussion Section:

Data analysis of Bushwick, Albany, Schenectady, and Troy interestingly resulted in the fact that there is no real difference between the three Upstate, NY cities when it comes to home value and similarity of businesses/venues compared to the Bushwick neighborhood of Brooklyn, NY. As such, all three cities are equally similar in venue make-up to one another, and similar in home value and home price over time. Thus, no recommendations can be made based on this analysis, and a

recommendation for further analyses based on demographic and crime rate data is recommended for the future.

Conclusion:

Results of the clustering and housing price analyses comparing the Bushwick neighborhood of Bushwick, NY to Albany, Schenectady, and Troy, NY clearly shows that all three cities have similar economic make-ups to Bushwick in Brooklyn. Statistical analyses attempted to distinguish the three cities based on their home values further revealed no statistically significant difference between all three Upstate, NY cities, meaning that distinguishing factors must be further explored through future analyses.