Data Analysis Steps

Raw Dataset Parameters: Refer to “Identifying Relevant Data: Using the Multiple Regression Model” section of report for list of datasets and data sources

For each dataset:

1. Reorganize Data: reorder and filter tracts by neighborhoods within city boundaries (using Text-to-Column, Sort-and-Filter)
2. Pivot Table: insert reorganized data into Pivot Table – filtered by neighborhood, values averaged to account for neighborhoods with multiple data points, expressed in percentage (%) or dollar value ($)
3. Filter Data: filter neighborhoods with available data points for all datasets (neighborhoods with incomplete datasets were excluded from data analysis)

Data Analysis:

1. Multiple linear regression: 15 datasets using Data Analysis Toolkit, with “Percent Non-White” as y-value for regression model.
2. Correlation function: used “CORREL” Excel function to compare statistically significant variables from multiple regression model with “Percent Non-White”
3. Cluster Analysis: run cluster analysis with Solver using statistically significant variables, categorize neighborhoods by three clusters
4. Geospatial Analysis (refer to Google Colaboratory Notebook for data analysis steps):
   1. Import relevant libraries
   2. Import cluster dataset of “Neighborhood Name” and “Cluster Number”
   3. Import geospatial data of Baltimore City map from Baltimore Neighborhood Indicators Alliance portal
   4. Merge geospatial polygon shape data onto cluster dataset
   5. Create static choropleth map using geoJSON format
5. Data Trends Before and After 2017:
   * 1. For all variables, average data points for 2015 -2018
     2. Pivot Table: insert data into Pivot Table filtered by neighborhood and values
     3. Create Pivot Chart