Malik Downey

Lealena Befikadu

Jean Didier Ernest

Ruiqi Lai

TianYi Li

Project Proposal

A study performed at University of Minnesota law found that most cities who experience redevelopment do not push low income residents out, Washington D.C is not most cities ([Study](https://www.law.umn.edu/sites/law.umn.edu/files/metro-files/american_neighborhood_change_in_the_21st_century_-_full_report_-_4-1-2019.pdf)). “For all the talk of gentrification happening in cities all over the country, what we found is that it really isn’t,” stated by Myron Orfield, director of the Minnesota law program. This study was conducted to explore and investigate growing social and economic disparities in American cities. Orfield said that “Washington is one of the few places in the country where real displacement is actually occurring. It’s quite rare.” Popular areas such as Capitol Hill, Navy Yard, the Southwest Waterfront and parts of downtown have been affected along with many others. Specifically Ward 6 has been affected the most.

Our goal with this project is to explore the trend of redevelopment over time and the outcomes of redevelopment within DC. We will do this by examining specific wards, income and population density of neighborhoods undergoing redevelopment. This is important because we want to learn if redevelopment in DC is resulting in negative outcomes. The purpose of redevelopment in DC is to “revitalize” neighborhoods to increase tourism and improve the local economy, not to displace natives (<https://www.dcfpi.org/all/a-broken-foundation-affordable-housing-crisis-threatens-dcs-lowest-income-residents-2/>). Our results will show the overall trends in the redevelopment in DC. If redevelopment in DC is resulting in the displacement of natives, it is important to be aware of this, so that future plans for redevelopment plans can include solutions to this such as possibly providing low-income housing for the people who were already living in DC.

We found four datasets that would be helpful to accomplish our goal. All the data come from Open Data DC (<https://opendata.dc.gov/>). Existing Land Use is the first dataset we select from Open Data. It represents those developed areas between 2004 to 2019. (<https://opendata.dc.gov/datasets/existing-land-use?geometry=-78.003%2C38.707%2C-76.026%2C39.081>). Future Land Use dataset is the Comprehensive Plan Amendment Act of 2010 (<https://opendata.dc.gov/datasets/future-land-use?geometry=-78.003%2C38.712%2C-76.025%2C39.086&page=2>). These two datasets will help us compare and define which area is planned to redevelop, but we are not trying to combine these two datasets. They do not contain a primary key that we can merge. We found another dataset called the DC Comprehensive Plan. It identified the downtown development in 1984. This dataset has only ten rolls, which represent ten main areas in Washington DC (<https://opendata.dc.gov/datasets/downtown-development-comprehensive-plan>). The last dataset we found was the ACS 2018 Poverty Status Variables Tract (<https://opendata.dc.gov/datasets/acs-2018-poverty-status-variables-tract?selectedAttribute=B17020_calc_numChildPovE>). This dataset shows the percent population and population income in Washington, DC. This dataset will be used to discover trends between population and income over the past year in DC. We will mainly use the Existing Land Use dataset and the ACS 2018 Poverty Status Variables Tract to accomplish our goal. We will use the other two datasets as reference.

To get our results, we will first extract data from online resources and open APIs, including Open Data DC. We will download the required data files from the listed data source website and save them as CSV files. If our required data are not presented in CSV files, we will use the python web scraping technique to load online data into our local computer. We can also use the data transformation ability of OpenRefine to transform data into our required data format.

Then, we will publish and save our required raw data into the Github repository. It allows our team to view and manipulate data collaboratively. The link of our Github repository is <https://github.com/Litianyi123/INST447-Project>, and we will update our data manipulation processes into this repository.

We may group and combine data from different data sets, and we will only reserve data that helps us analyze the trend. We can use Pandas DataFrame of Python to clean and manipulate the data, and we also can use OpenRefine to achieve our goals. Specifically, Pandas DataFrame can help us to subset, merge, reshape our data.

After collecting required data, we will visually present our data by using tableau and publish the story on tableau public repository. The relative information will also be presented on the Github repository. By observing the data visualization, we should be able to have insight of the trend, and we will be able to summarize the data. If we need further statistical analysis to observe the trend, we can use RStudio to achieve certain requirements.

An additional dataset about Affordable Housing will be used to test the result of University of Minnesot’s study. This data set contains data about specific location, different ward in Washington DC, number of affordable units. With the full address and the ward section, we will be able to figure out which ward is the most affected by gentrification. University of Minnesota found that Ward 6 faced tougher gentrification experience than the other seven wards.

The data set about Affordable Housing requires challenging tasks to be made. The first one is data cleaning. We will make sure that all the datasets will be correct, usable, and free of all errors. Deleting duplicate data, changing column names, trimming white space are the necessary steps to clean the data we have. The second challenging task is merging two or more datasets. This process can also be called syncing data; it is called *join* is SQL.

The steps we will take to process our data will be appropriate to draw significant insight. First, we will get data from multiple sources and combine them to make it into one dataset. We will prepare the data after combining them. The major part of preparing data is data cleaning. After collecting, and preparing the data, one of the programming languages like R, Python, will be used for calculation or analysis. At the end, the result will be visually presented via Tableau.