## **Project One**

|  |  |
| --- | --- |
| Project Title: | Micro-Brew Gentrification |

## **Team Members**

|  |
| --- |
| **Names:** |
| **Shawn Boehm** |
| **Mica Massie** |
| **Neida Flores** |
| **Michelle Hannah** |
|  |

## **Project Week Timeline**

|  |  |
| --- | --- |
| Date | Subject |
| 7/18/20 | Project Intro, Collaboration with Git, & Project Work |
| 7/20/20 | Pulling and Merging with Git & Project Work |
| 7/22/20 | Hypothesis Testing and Statistical Tests & Project Work |
| 7/25/20 | Project Work |
| 7/27/20 | Project Work |
| 7/29/20 | Project Presentations |

## **Team Effort**

Before anything, remember that Projects are a group effort: Working closely with your teammates is a requirement. This both helps teach real-world collaborative workflows, and enables you to tackle more difficult problems than you'd be able to working alone.

In other words, working in groups allows you to work smart and dream big. Take advantage of it!

**Note**:

* 100% participation and attendance is required in order to receive credit for the Group Project work.
* Participation in all Projects are course requirements.

## **Project Description**

Before you start writing any code, your group should outline the scope and purpose of your project. This helps provide direction and prevent [scope creep](https://en.wikipedia.org/wiki/Scope_creep).

Write this as a brief summary of your interests and intent, including:

* The kind of data you'd like to work with: map, economic data, census
* The kinds of questions you'll be asking of that data:
  + Is there any correlation between the number of microbreweries in a city and that cities’ gentrification?
  + What is our basis for determining gentrification? Income, property value?
* Possible source for such data: Kaggle, US News, Git, Zillow, Open API
* Breweries: Number of Microbreweries in each of the target areas. Also, data on brewpubs may also prove useful
  + Density of breweries in a given city needs to be over years
* Gentrification: In the given cities we need to identify property value increases over years. Median Income Data in these same areas may also prove useful.
* Graphics: Map showing gentrification in each city with breweries plotted over.
  + Number of Breweries and property value over time(years)

Strategy: Each team member finds one part of the data and create the appropriate csv and data frames in Jupyter and add to their respective branch of the project repository on Git.

Shawn: Brewery data, years

Mica: Average Property Values over years in each of the six cities

Nieda: Median Income Data over years, in each of the six cities

Michelle: Average Age of residents over years, in each of the six cities

Remember the **Analytics Paradigm**:

1. Decompose the Ask
2. Identify Data Sources
3. Define Strategy and Metrics
4. Build Data Retrieval Plan
5. Retrieve the Data
6. Assemble and Clean
7. Analyze for Trends
8. Acknowledge Limitations
9. Make the Call or Tell the Story

|  |
| --- |
| Your Project Description |
| Our project’s goal is to uncover, any correlation between Gentrification and the concentration of Micro-Breweries.  **Response**: Most gentrified cities according to US News: DC, San Diego, New York, Albuquerque, Atlanta. Also included will be Nashville. We can narrow our search and plotting to these major metropolitan areas.  Gentrification parameters: Income, property value, age of residents  **Response**: Maps showing locations of breweries in areas of gentrification  Density over time plotted with econ data over time  **Response**: |
|  |

## **Finding Data**

Once your group has written an outline, it's time to start hunting for data. You are free to use data from any source, but we recommend the following curated sources of high-quality data:

* [data.world](https://data.world/)
* [Kaggle](https://www.kaggle.com/)
* [Data.gov](https://www.data.gov/)
* [Public API Listing](https://github.com/toddmotto/public-apis)

Chances are you'll have to update your Project Outline as you explore the available data. This is fine—adjustments like this are part of the process! Just make sure everyone in the group is up-to-speed on the goals of the project as you make changes.

Make sure that your data is not too large for local analysis. Big Data datasets are difficult to manage locally, so consider a subset of that data or a different dataset altogether.

**Data Cleanup & Analysis**

With data in hand, it's time to tackle development and analysis. This is where the fun starts!

Inevitably, the analysis process can be broken into two broad phases: Exploration & Cleanup and Analysis proper.

As you've learned, you'll need to explore, clean, and reformat your data before you can begin to answer your research questions. We recommend keeping track of these exploration and cleanup steps in a dedicated Jupyter Notebook, both for organization's sake and to make it easier to present your work later.

Similarly, after you've massaged your data and are ready to start crunching numbers, you should keep track of your work in a Jupyter Notebook dedicated specifically to analysis.

During both phases, don't forget to include plots! Don't make the mistake of waiting to build figures until you're preparing your presentation. Creating them along the way can reveal insights and interesting trends in the data that you might not notice otherwise.

We recommend focusing your analysis on techniques such as aggregation, correlation, comparison, summary statistics, sentiment analysis, and time series analysis.

Finally, be sure that your projects meet the technical requirements.

## **Technical Requirements**

The technical requirements for Project 1 are as follows.

* Use Pandas to clean and format your data set(s)
* Create a Jupyter Notebook describing the data exploration and cleanup process
* Create a Jupyter Notebook illustrating the final data analysis
* Use Matplotlib to create a total of 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)
* Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation
* Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions
* Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots.
* Create a comprehensive README

## **Data Sources**

|  |  |  |
| --- | --- | --- |
| # | Description | URL or Resource Link |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |

## **Finding Data APIs to be Used**

|  |  |  |
| --- | --- | --- |
| # | API Description | API URL |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |

## **Repo Link**

|  |
| --- |
| Repo URL |
|  |

## **Presentation**

After you've analyzed your data to your satisfaction, you'll put together a presentation to show off your work, explain your process, and discuss your conclusions.

This presentation will be delivered as a slideshow, and should give your classmates and instructional staff an overview of your work. PowerPoint, Keynote, and Google Slides are all acceptable for building slides.

As long as your slides meet the presentation requirements, you are free to structure the presentation however you wish, but students are often successful with the format laid out in the [presentation guidelines](https://github.com/coding-boot-camp/DataViz-Lesson-Plans/blob/master/01-Lesson-Plans/08-Project-1/1/ProjectGuidelines/PresentationGuidelines.md).

# **Presentation Requirements**

The presentation requirements for Project 1 are as follows.

Your presentation must:

* Be at least 12-15 min. long
* Describe the core message or hypothesis for your project.
* Describe the questions you and your group found interesting, and what motivated you to answer them
* Summarize where and how you found the data you used to answer these questions
* Describe the data exploration and cleanup process (accompanied by your Jupyter Notebook)
* Describe the analysis process (accompanied by your Jupyter Notebook)
* Summarize your conclusions. This should include a numerical summary (i.e., what data did your analysis yield), as well as visualizations of that summary (plots of the final analysis data)
* Discuss the implications of your findings. This is where you get to have an open-ended discussion about what your findings "mean".
* Tell a good story! Storytelling through data analysis is no different than in literature. Find your narrative and use your analysis and visualization skills to highlight conflict and resolution in your data.