**Data Bootcamp Project 1 – Climate Study**

**Team Members**

* Will Davis
* Mohamed Mursal
* Thaimu Sesay
* Jessica Snowden

**Project Ideas**

* Question: How much has the average high temperature for the 5 largest cities in the USA changed over the past 20 years?
* Question: How much has the average annual precipitation for the 5 largest cities in the USA changed over the past 20 years?
* Question: Is there a relationship between average high temperature and average annual precipitation?

**Data Exploration**

* What are the largest 5 cities?
  + Use Census data to determine the 5 largest cities
    - Find .csv or API with population by city
    - Sort descending by population
    - Drop all but top 5
* What is the annual high temperature for each city?
  + Using OpenWeatherAPI, pull daily high temperature by city
  + Compress data into measure of central tendency
    - Average by month?
    - Average by year?
    - Average of the highest values? (set cut-off for highest values/year by using box/whisker analysis)?
* What is the annual precipitation by location?
  + Using OpenWeatherAPI, pull daily precipitation by city
  + Compress data into measure of central tendency
    - Sum by year?
* Notes / Considerations
  + 20 years of data = 7,305 daily values
  + We can compress/reduce this data to 20 rows by choosing a way to represent each year with a single value of temperature and precipitation

**Proposed Analysis and Output Graphics**

* Map (1) showing 5 largest cities
* Box-Whisker plot (1) to show how we decided how to reduce daily temperatures into a single representative value for each year
* Line graph (1) showing average temperature by year for all 5 cities (on same plot)
* Line graph (1) showing annual precipitation by year for all 5 cities (on same plot)
* Scatter plots (5) showing relationship between average temperature (x-axis) and annual precipitation (y-axis) for each city
  + Plot regression lines on each scatter plot

**Project Plan**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Date | Day | Will | Mohamed | Thaimu | Jessica |
| Oct 9 | Day 1: Define | Brainstorm ideas | Brainstorm ideas | Brainstorm ideas | Brainstorm ideas |
| Oct 10 | Day 2: Develop | Draft & submit proposal via Slack  Temperature Data   * Explore API (Open Weather) * Identify data fields for temperature * Understand parameters | Population Data   * Explore API * Identify data fields for population * Understand parameters | Precipitation Data   * Explore API (Open Weather) * Identify data fields for precipitation * Understand parameters | Temperature Data   * Explore API (Open Weather) * Identify data fields for temperature * Understand parameters |
| Oct 16 | Day 3: Develop | Develop dummy dataframe assuming values of precipitation and max temperature  Code approach to compress data   * Sum of precipitation data for annual value * Average of high temperature data for annual value   Output into dataframe | Identify API  Code data call to get population by city data  Identify 5 largest cities, capture these data elements:   * city\_name * city\_pop * city\_lat * city\_long   Output into .csv file for importing into a dataframe | Develop dataframe with timestamp values needed for API call; timestamps will be for 2023  Code data call to get precipitation data for NYC for 2023  Output into dataframe | Develop dataframe with timestamp values needed for API call; timestamps will be for 2023  Code data call to get max temperature data for NYC for 2023  Output into dataframe |
| Oct 17 | Day 4: Refine |  |  |  |  |
| Oct 21 | Day 5: Finalize |  |  |  |  |
| Oct 23 | Day 6: Deliver |  |  |  |  |

**Deliverables**

* Final code for project fully commented and annotated
* PowerPoint presentation that outlines approach, shows sample Dataframes, displays any output graphics, draws 2-3 observations/conclusions from the data analysis