**Term Project**: Milestone #5

**Date**: August 2, 2020

**Name**: Maddie Bauer

Throughout the course of this semester, I have truly learned a lot when it comes to data wrangling. Not only am I now more comfortable with the terms associated with data wrangling, but I feel as if I have been successful in wrangling data from different file sources.

For milestone 2, I imported my CSV dataset from Kaggle. This dataset had information regarding average monthly temperatures for major world-wide cities. I cleaned this dataframe by dropping missing values, dropping unnecessary columns, looked for duplications and outliers, and created a function to convert all temperature data to Fahrenheit.

For milestone 3, I web scraped a Wikipedia page that provided me with a list of United States cities ranked by their population. I cleaned this dataframe by renaming city names due to hyperlinks on Wikipedia, removing missing values, creating new columns due to merged headers on Wikipedia and checked for duplicates. This was by far the hardest milestone for me to complete as I had absolutely no experience with web scraping or HTML.

For milestone 4, I connected to the OpenWeatherMap.com API to gather current weather data for different United States cities. I was able to create a dataframe by looping through the JSON data from my API requests.

For milestone 5, I connected to a database and placed my dataframes within the database. I was then able to make several queries to use and merge the data to create different visualizations. I created the following visualizations:

* A bar chart which compares the current temperature with the average temperature for different cities.
* A line chart which displays the average monthly temperatures for Milwaukee, WI along with the average yearly temperature displayed.
* A line chart which displays the current temperature, feels like temperature and the average August temperature for different cities.
* A scatter plot of the average yearly temperatures along with the mean average yearly temperature displayed.
* A pie chart to show how prevalent each continent is within one of my dataframes.

Throughout this process I have learned the following things:

* How to clean a dataframe
* How to check for missing values
* How to remove outliers by looking at histograms and box plots
* How to group data based on certain columns or values
* How to web scrape and use BeautifulSoup
* How to conduct an API request and parse through JSON to create a dataframe
* How to load a dataframe into a database
* How to query data from a database
* How to plot more characteristics using matplotlib, such as adjusting x-axis rotation, color of lines, style of lines, legend placement, etc.
* How to ask others for help (I’ve never been good at this one!)

Overall, I feel more confident in my Python skills as well. I know there is always more to learn and this course has definitely helped me realize that there is always information out there to help you when you’re stuck.