Summary

For the term project I choose 3 historical data sets for the US zip 43212 (Columbus, Ohio). The first data set (second milestone) I downloaded from <https://www.worldweatheronline.com/> (in csv format). The data contains historical weather data from 07/01/2008 to 10/25/2020 in hourly basis. I selected only data for noon for each day. Next, historical weather data from 09/30/2011 to 09/29/2014 were extracted. The dataset was split on three parts

1. 09/30/2011-09/29/2012
2. 09/30/2012-09/29/2013
3. 09/30/2013-09/29/2014

The second dataset (the third milestone) I downloaded from <https://openweathermap.org/> again in csv format, because nowadays all historical weather data are present only in the csv or json format I created the html format from the csv file. The html file contains historical weather data from 09/29/2014 to 09/29/2017. The dataset was split on three parts

1. 09/30/2014-09/29/2015
2. 09/30/2015-09/29/2016
3. 09/30/2016-09/29/2017

The third data set (the fourth milestone) I downloaded from [https://www.visualcrossing.com](https://www.visualcrossing.com/) API

1. API key:NCG9V4G5X7Y1HN3S5U2PAWNNR The data contains historical weather data from 09/29/2017 to 09/29/2020. The dataset was split on three parts
2. 09/30/2019-09/29/2020
3. 09/30/2018-09/29/2019
4. 09/30/2017-09/29/2018

The whole idea of this project is to compare temperature for 9 consecutive periods. I also wanted to check if temperature correlates with wind speed. Each period starts from September 30 of a corresponding year and ends on September 29 of next year. During this project I learn many important tools and concepts such as pandas’ data frame indexing, how to find index corresponding to a particular value, how to use regular expressions to select particular rows. I also learn how to build grouped bar plots with error bars. I learn how to groupby a data frame by a column categorical values, how to aggregate data.