**Individual Project 2**

**DS160-02**

**Introduction to Data Science**

**Spring 2022**

**Working with Pandas (100 points)**

**Goal:** The goal of this project is to use some of the features/functions provided by the Python Pandas library

**Instructions:** Create a new notebook titled **IP2\_XXX**, where **XXX** are your initials. Also create a GitHub repository titled **IP2\_XXX** to which you can push your code. Then complete the following:

1. Attached to this assignment in Moodle is a file called weather\_2012.csv. Download this file and put it in the same directory as your notebook. Read the file into a dataset named “df”.
2. Write the code to display the first five rows of the dataset, the last five rows of the dataset and a sample of five random rows from the dataset.
3. Write the code to display a list of all of the columns in the dataset.
4. The column names are unwieldy for our purposes. Rename them according to this list {date\_time, temp\_c, dew\_point, relative\_humidity, wind\_speed, visibility, pressure, and weather}. K**eep a copy of the original columns so we can change them back if needed at the end of our analysis.**
5. Most of the data in the dataset are numerical, write the code to get some standard statistics for each column (e.g. count, mean, etc.)
6. Write the code to add a column named temp\_f to the dataset that contains the temperature from temp\_c represented in Fahrenheit.
7. Write the code to rearrange the columns in the dataset so temp\_f appears immediately before temp\_c.
8. Write the code to display the data types of all the columns in the dataset.
9. You’ll notice that the date\_time column is not actually considered a date by Pandas. Write the code to convert this column to a date/time column. Verify the conversion was successful.
10. Write the code to get a count of the different weather designations in the weather column and display top 10 weather categories.
11. Which month of the year, the temp\_f more than 90, humidity less than 50 , dew point more than 17, and the weather is clear? Extract that data and save it to a new data frame.
12. **Extract** the month information from the date\_time column and **add** a month column at the data frame. (example: dt.month)
13. Save your original dataframe to ‘weather\_2012\_new.csv’) – ensure your file does not contain the index column of the dataset.
14. Add, commit and push your code to GitHub.

**Project Submission:** Upload a link to your GitHub repository for the project in the area provided in Moodle by the deadline specified.