**Project#2**

(Total Points: 15)

Due Date: 04/14/2020 (End of the day)

**Introduction:**

1. Weather data from 2010 to 2019 is available at /data/weather/<year>
2. Each year data is in the respective directory (i*.e., 2010 directory have data for year 2010*)
3. Additional data information can be found in the files:
   1. ish-history.txt
   2. readme.txt
   3. country-list.txt
4. Use spark to analyze the above-mentioned data.
5. Example commands used in the class is also attached: Temp\_Test.html
6. Ignore the missing values for the dataset (readme.txt files attached).
7. Drop columns which we dropped in the class (refer temp\_test.html file).
8. If you are unable to load all the data at once, load data for one year at a time.

**Tasks:**

1. Find the hottest and coldest day along the station code and date for each year.
2. Find the hottest and coldest day across all years (2010 - 2019) along with station code and date.
3. Maximum and minimum precipitation with station code and date for year 2015.
4. Count percentage missing values for mean station pressure (STP) for year 2019 and stations.
5. Station code with maximum wind gust and date for year 2019

Submission:

1. Your result.txt file with all the above information
2. Your spark code/command