

## HW3

Greyson Sequino, Zach Fuchs, Videep Venkatesha

### Question 1

-Results:

Lowest Day : MONDAY	30.881996324161367
Highest Day : FRIDAY	31.389048478112457

-Process:

The day was extracted from each Epoch value. Each day was a key and all the corresponding AQI scores for that day were values (Mapper). The Average aqi score for each day was calculated. The Highest and Lowest average values were found from all the averages(Reducer).

### Question 2

-Results:

Lowest : OCTOBER	28.63588478835831
Highest : MAY	32.89839018129146

-Process:

The month of each record was extracted from the Epoch time. We first mapped each month to the AQI score and then in the reducer we calculated the mean aqi score for each month. After calculating the mean aqi score we used a cleanup method to sort the mean aqi scores and output the largest and smallest of the means.

### Question 3

-Results:

G0800430	12.647058823529411
G0801130	14.521367521367521
G0800970	14.8989898989899
G5300130	15.649122807017545
G0801070	15.972067039106145
G1300550	16.166666666666668
G1201070	16.407103825136613
G1300730	17.09795918367347
G1200230	17.20357142857143
G5300570	17.62676056338028

-Process:

This question can be answered with one job. The mappers map all AQIs in 2020 to their respective county. The reducer takes all the AQI values per county and calculates an average. It then adds this average to a list. Once all keys have been added to the list, the list is sorted and the lowest 10 county AQI pairs are returned.

#### Question 4

-Results:

G0600310	64.37978142076503
G0600190	61.5
G0600290	56.98907103825137
G0600390	56.505464480874316
G0600710	56.275956284153004
G0600510	54.93593314763231
G0601070	54.53825136612022
G0600630	54.09274193548387
G0600070	53.38251366120219
G0601010	51.97540983606557

-Process:

This question can be answered with one job. The mappers map all AQIs in 2020 to their respective county. The reducer takes all the AQI values per county and calculates an average. It then adds this average to a list. Once all keys have been added to the list, the list is sorted and the highest 10 county AQI pairs are returned.

#### Question 5

-Results:

G0600010	112.71428571428571
G0600050	-35.142857142857146
G0600070	142.71428571428572
G0600090	-81.85714285714286
G0600110	96.28571428571429
G0600130	119.42857142857142
G0600150	46.5
G0600170	-58.85714285714286
G0600190	104.14285714285714
G0600210	-37.52380952380952

-Process:

1. Unique weeks were extracted through the Epoch values ("County+Week+year"). Each week has 7 Aqi scores for each day in the week.
2. Mapper1 mapped Unique Week and corresponding AQI scores and passed to Reducer1.
3. Reducer1 calculated the Average Aqi score for each week.
4. Unique Week mapped to corresponding to average aqi was passed to Mapper2.
5. Mapper 2 mapped <county, unique week + averageAqi>.
6. Reducer2 sorted unique week chronologically using a TreeMap.
7. The difference between average aqi each week and the previous week was calculated.
8. That value was used as a key in reducer 2 and was mapped with county.
9. Result - <County, change in Aqi Score>

## Question 6

-Results:

```
0 186.42857200,G0600070,46,2017-2018
1 140.28571400,G0600070,38,2019-2020
2 133.71428600,G0601010,37,2019-2020
3 130.71428600,G0600070,37,2019-2020
4 113.14285700,G0601010,26,2007-2008
5 105.28571500,G0600610,38,2019-2020
6 103.14285700,G0601010,38,2019-2020
7 101.42857100,G0600570,38,2019-2020
8 100.85714300,G0600070,35,2019-2020
9 100.28571400,G0601010,40,2019-2020
10 99.16666700,G0600570,37,2019-2020
11 96.71428500,G0601010,46,2017-2018
12 95.57142900,G0601010,35,2019-2010
13 93.71428600,G0600070,40,2019-2020
14 93.28571400,G0600070,32,2018-2019
15 91.85714300,G0600570,35,2019-2020
16 91.28571400,G0601010,28,2008-2009
17 89.71428600,G0600610,37,2019-2020
18 89.00000000,G0600610,35,2019-2020
19 85.57142900,G0600570,46,2017-2018
20 83.57142900,G0601010,34,2019-2020
21 80.42857200,G0601010,33,2001-2002
22 80.28571400,G0600610,34,2019-2020
23 79.71428500,G0600610,33,2013-2014
24 79.14285800,G0800690,41,2019-2020
```

-Question:

Do wildfires cause a large year to year AQI change for the same week in each year

-Process:

We started by looking up large wildfires across counties in California and Colorado then finding counties nearby to where the fire burned at. Two fires that we focused our search on were the Cameron peak (Colorado) and Dixie (California) fires which affected many counties in both states. From this point we asked if it was possible to detect wildfires within these counties by taking the difference from a week's mean aqi score one year and the same week's mean aqi score the next year. To start processing the aqi data provided to find the answer to this question we filter out any data not from Nevada, Placer, El Dorado, Yuba, Butte, Sierra, Sutter, Larimer, Weld, and Boulder counties. If the data does originate from one of those counties we calculate the average aqi for each unique year, week, and county. After getting the mean aqi for each year, week, and county we map the results to each week and county as the key and the year and aqi as value to calculate the year to year difference in mean aqi of a county's week. With the results we sort to get the largest difference in mean aqis and compare the resulting dates with the dates of the fires in that county.