

CS677 HW3

Create a Jupyter notebook CS677_HW3_lastName and add the python code and output cells for the following. Upload the notebook to the Assignments section once completed. **Only one attempt is available to upload.**

Use Seaborn for plotting functions.

All code should be done without using any loops.

For Part1, do not use any aggregation, grouping, merge/join operations.

For Part2, use the aggregation, merging and other operations covered during the class.

Initialize the USA Daily Average temperatures dataset as shown below:

```
df = pd.read_csv('http://people.bu.edu/kalathur/datasets/usa_daily_avg_temps.csv.gz', header=0)
```

PART 1

Q1. Show the distribution plot for the 'avgtemp' attribute. What anomaly do you notice from this plot?

Q2. Display the rows in the dataframe which have this anomaly. What do you infer about the data collection process?

Q3. Do an in-place replacement in the dataframe for the above value with the NaN value.

Q4. Do an in-place fill of the missing value (NaN value) in the dataframe with the forward fill approach. Show the distribution plot for the 'avgtemp' attribute to make sure the observed anomaly is gone.

Q5. For the city Boston, which New Year day is the coldest? warmest? Show the dataframe with the New Year days for Boston. Show an appropriate plot for the 'avgtemp' attribute all the available New Years days for Boston.

Q6. Create a subset of the dataframe for the cities Boston and Orlando with all the data for the first day of each month for all the available years (use the isin method for the cities).

Q7. Using the data from Q6, show the strip plot for the avgtemp attribute of the two cities using the year for the hue semantic.

Q8. Using the data from Q6, show the swarm plot for the avgtemp attribute of the two cities using the year for the hue semantic. What do you infer from the plot?

Q9. Using the data from Q6, show the monthly breakdown of the strips plots for the avgtemp attribute of the two cities using the year for the hue semantic. Use col_wrap=4 for the plot. Write at least two conclusions from the plot.

Q10. Using the data from Q6, show the monthly box plots for the avgtemp attribute of the two cities using the city for the hue semantic. Write at least two conclusions from the plot.

PART2

Q11. What are the maximum yearly temperatures recorded? Show the resulting dataframe and an appropriate plot.

Q12. Show the dataframe with the state, city, month, day, year, avgtemp reflecting the maximum yearly temperatures computed with the dataframe from Q11.

Q13. Using the appropriate pandas method, show the relationship of the frequencies between the state and city values for result from Q12.

Q14. What are the maximum temperatures recorded for each state? Show the resulting dataframe and an appropriate plot.

Q15. Show the dataframe with the state, city, month, day, year, avgtemp reflecting the maximum temperatures computed with the dataframe from Q14.