CS677 HW2

Create a Jupyter notebook CS677_HW2_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.**

Initialize the US Zipcode dataset as shown below:

usZipCodeData = pd.read_csv('http://people.bu.edu/kalathur/datasets/uszips.csv',
    converters={'zip': lambda x: str(x)})

Q1. Show the count, mean, standard deviation, min, max, and quartiles for the population attribute

Q2. How many city zip codes have fewer than 1000 people?

Q3. Re-assign usZipCodeData variable to have only the rows where population is 1000 or above. Change the index values of this new dataframe so that they are contiguous.

Q4. Compare the count, mean, standard deviation, min, max, and quartiles for the population attribute w.r.t. Q1.

Q5. Assign the data containing all cities for Massachusetts to the variable ma_StateData

Q6. For Massachusets data, compare the mean, standard deviation, min, max, and quartiles for the population attribute w.r.t. Q4.

Q7. Show the histogram for the population for the Massachusetts cities. Provide brief interpretation of the figure.

Q8. Show the top 20 zip codes for Massachusetts by the decreasing order of density attribute.

Q9. Show the top 20 zip codes for Massachusetts by the decreasing order of population attribute.

Q10. What zip codes are common between Q8 and Q9.

Q11. For Massachusetts, show a scatter plot of latitute versus longitude using color as the log of the population and size as 1/25 of the density.

Q12. For the top 75 populous zip codes in the usZipCodeData, show a pie chart with the distribution of the states and the frequencies of the zip codes in those states.
(Hint: Use value_counts. Do not use any aggregate functions not yet covered).

Q13. Using the last four digits of your BU ID as the seed, pick 10 random rows from the Massachusetts data. Show the resulting data frame. Show the horizontal bar chart of the populations with the city as the label.