Hackerspace – The coding community of MSIT

Code Craft 3.0 Data Science

Grand Total = 50
Qualifying marks=40%
Presentation: 4 marks

Instructions:

- **File should be submitted in ".ipynb" format**
- **Report must be submitted in PDF**

Section A
(4 X 2 = 8)

1. Create a subset for your data frame and drop the column 'child_sex_ratio' and display the data.
2. Display all the column headings together.
3. Sort the data in an alphabetical order w.r.t. the column 'state_name'.
4. Calculate 'sex ratio' for each row and add it as a new column.
   [sex ratio= (Number of female/Number of male) X1000]

Section B
(4 X 5 = 20)

1. Provide a Bar Plot for 'state_name' vs 'literates_total', such that the states names do not overlap and label both x and y axes. Provide a suitable title for the plot.
2. Calculate the 'effective literacy rate'. Utilize the data to perform a scatter plot between total graduate vs effective literacy rate. Provide all the necessary labels & title.
   [Effective literacy rate= (total literates/total population) X100]
3. Provide a KDE plot for 'literates_total' taking 'state_name' as hue. Provide a necessary title.
4. How can you visualize the relationship between total population and effective literacy rate, while also representing the number of total graduates with bubble sizes in a scatter plot using Matplotlib?

Section C:
(2 X 7 = 14)

1. Create a dashboard for all the aforesaid plots.
2. Calculate the population ratio of male vs female in each state. Utilize the data to prepare a heatmap using top 10 states (according to population).

Report:
(4)

Generate an overall report of the EDA that you have performed, mentioning what you can infer from the different visualizations.