NPV – Mini-Project

A Company is having its offices across all locations of India. These locations are categorized with respect to zones. The company has decided to give medical expenditure benefits to their employees. The company has collected the medical expenditure of all the employees for the year 2019. The company also collected some other health and family related data of all employees like number of dependents, age, bmi, smoking and alcoholic habits.

Answer the following questions to provide insight into the collected dat.

Dataset details **

id: Employee id - unique

age: age of primary beneficiary

sex: Gender data - female, male

bmi: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height,

objective index of body weight (kg / m ^ 2) using the ratio of height to weight, ideally 18.5 to 24.9

dependent: Number of children other family members who are not earning and dependent on employee.

alcohol: Alcohol consumption habit. Possible values - daily, weekend, party, no, rarely

smoker: Smoking habit.

zone: The office location zones. Possible values - east, west, south, north, northeast, southeast, southwest, northwest.

expenditure: Individual medical expenditure by employees during year 2019

Read the dataset and do primary study of data. Perform following steps:

 a. Reading some rows of dataset

 b. Learn data types of all features

 c. See the statistical properties of the data.

1. How many employees above 50 age smoke, what is their average expenditure on health management?

 2. How many female employees did not consume alcohol daily. What is their proportion with respect to total female employees?

 3. Display zone wise average expenditure of employees on health

4. Employees of which zone did maximum expenditure on health

5. What is the minimum expenditure done by any employee in organization

6. What is average expenditure of employees who have 2 or lesser dependents

7. How many employees are below age of 20 years. What is their average BMI

8. Create a dataframe "non_smoker" for all non-smoker male employees

9. Create a dataframe "non_alcoholic" for all male employees who do not consume alcohol

10. Create a dataframe called "non_smoking_alcoholic" with the men employees who do not smoke and consume alcohol. The dataframe should contain columns id, smoke, alcohol

(Hint:Use join and dataframes created in Q8 and Q9)

11.What is the count for each zone with respect to number of dependent

12. Classify the employees into 3 categories.

Category 1 : employees below 18.5 bmi - underweight

Category 2 : employees with bmi in the range of 18.5 to 24.9 - fit

Category 3 : employee with bmi over 24.9 - overweight

What is the average expenditure of employees in each category.

Display the bar chart for the same

13. Draw line plot of expenditure against age and BMI

14. Draw a bar chart to show zone wise average expenditure

15. Draw a bar chart to show smoking and alcoholic habits against average expenditure

16. Draw pie chart for alcohol consumption habits

17. Draw scatterplot to see the relation between age and expenditure, keep smoking habit as hue

18. Draw distribution plot of expenditure

19. Display the pairplot

20. Display heatmap for the data