Project: Exploratory Data Analysis on Car Price

1. Read the dataset "CarPrice.csv" and find the information of data, the shape of the data, and which attributes have the missing values.

2. Remove all missing values with memory updating using single-line code. How many rows have been eliminated, and why?

3. Read the dataset "CarPrice.csv" again, and remove the missing values with respect to the "wheelbase" attribute with memory updating. How to prove the number of data/tuples correctly removed?

4. Read the dataset "CarPrice.csv" again, and remove the missing values with respect to the "wheelbase" and "drivewheel" with memory updating. Is this problem the same as question 2? Prove it.

5. Read the dataset "CarPrice.csv" again, regarding the "drivewheel" and "wheelbase" attributes, which one is categorical, and which is numerical?

6. What are the primary central tendency methods are used for filling the categorical and numerical missing values?

7. Fill the missing values of "drivewheel" and "wheelbase" using the appropriate central tendency method with memory updating.

8. Read the dataset "CarPrice.csv" again, and fill the missing values of the "drivewheel" group by "Carname" with memory updating. Show the proof that your fill is unique according to category.

9. Read the dataset "CarPrice.csv" again, and fill the missing values of the "wheelbase" attribute group by "CarName" and "Symboling". Show the proof that your fill is unique according to category.

10. Extract from the data, which car is more popular. How have you decided it? Each line of code should be presented.

Descriptive Analytics:

11. What are the summary statistics for the 'price' attribute, including mean, median, standard deviation, and range?

12. What is the distribution of 'citympg' and 'highwaympg' in the dataset?

13. How many cars of each 'fueltype' are there in the dataset?

Data Visualization:

14. Create a histogram for the 'horsepower' attribute to visualize the distribution of engine power.

15. Plot a scatter plot between 'price' and 'enginesize' to explore the relationship between engine size and car price.

16. Generate a box plot for 'price' by 'carbody' to see how different car body types affect prices.

Segmentation and Grouping:

17. What is the average 'price' of cars segmented by 'fueltype'?

18. Group cars by 'cylindernumber' and calculate the mean 'horsepower' within each group.

19. Segment cars by 'enginetype' and 'aspiration' to understand the distribution of engine types in turbocharged and non-turbocharged cars.

Customer Segmentation:

20. Segment customers based on their preferred car attributes, such as 'cylindernumber,' 'carbody,' and 'fueltype,' and analyze the characteristics of each segment.

Cost-Benefit Analysis:

21. How does 'enginesize' impact 'price,' and is it a cost-effective feature for customers?

22. Investigate the relationship between 'curbweight' and 'citympg' to understand the trade-offs between car weight and fuel efficiency.