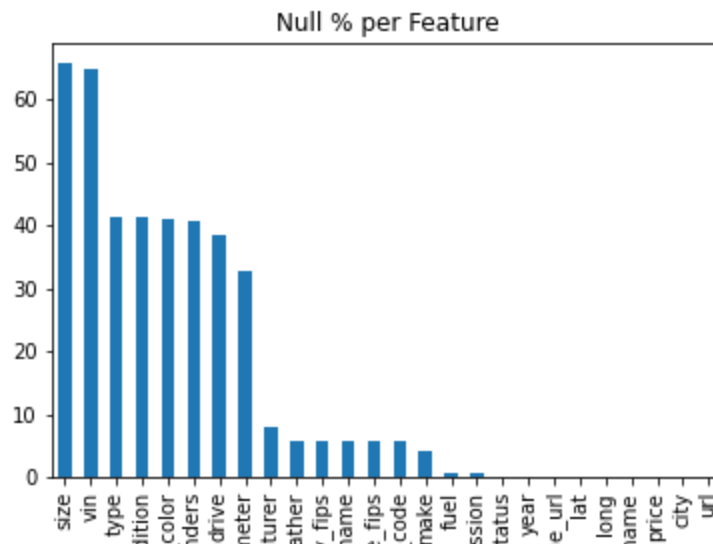**1. Ask the following questions and look for the answers using code and plots:**
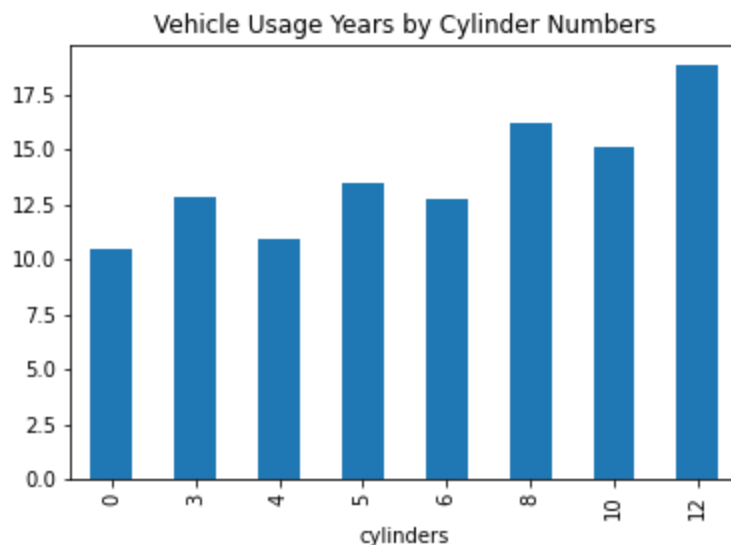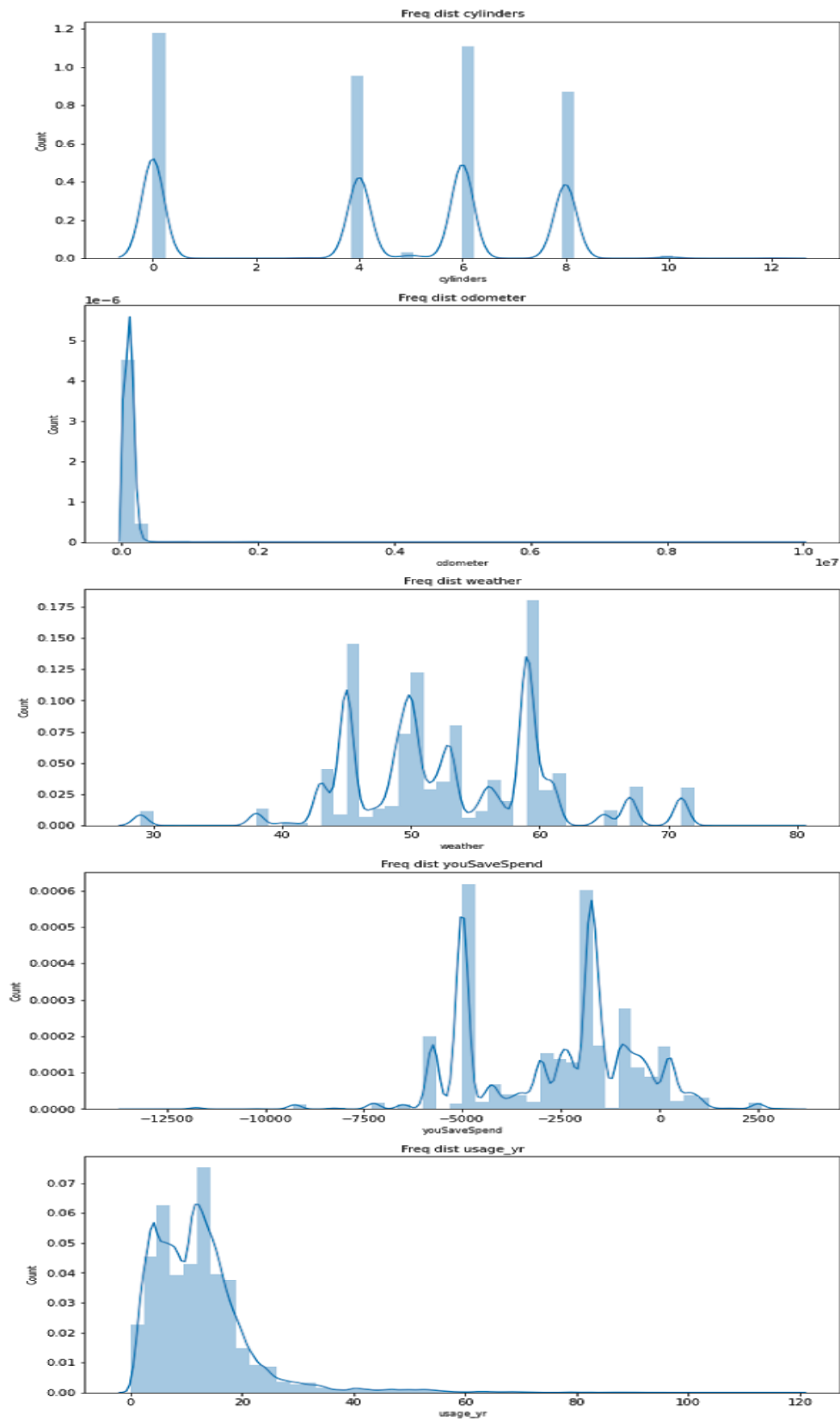**1. Can you count something interesting?**

"Which features do this dataset have the most number of null values?" This is my first question I had because a successful data cleansing process depends not only on how much null values are distributed over features, but also on how the null values per feature are pre-processed.

2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?
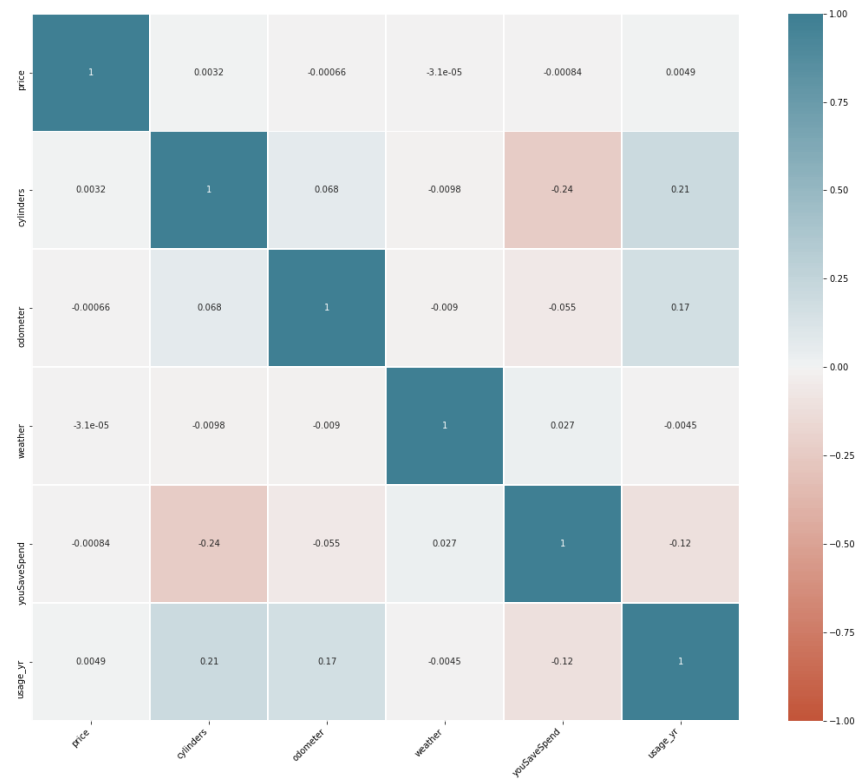
The more cylinders your car has, the older your car is.

3. Can you make a bar plot or a histogram?

Freq dist cylinders

Freq dist odometer

Freq dist weather

Freq dist youSaveSpend

Freq dist usage_yr

4. Can you compare two related quantities?

With negative correlation, the more numbers of 'cylinders', the longer 'usage_yr' and the less 'youSaveSpend'. This can be interpreted that once you purchased a car with more than average numbers of cylinders, you would use the car longer, surely along with less fuel economy and efficiency ('youSaveSpend).

5. Can you make a scatterplot?
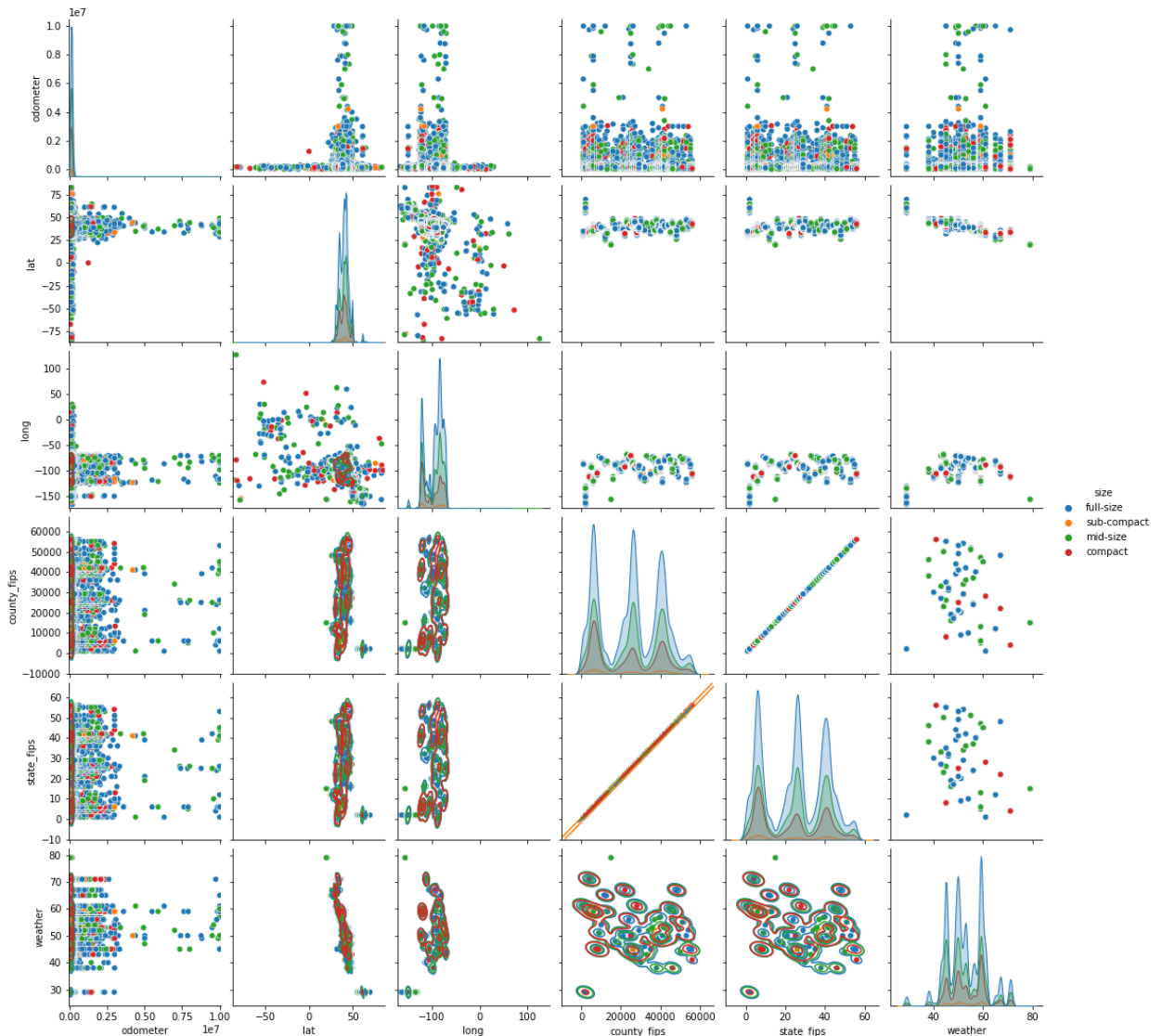
6. Can you make a time-series plot?
Sames as the above number 2.

2. Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you'd like to investigate further? What other questions do the insights lead you to ask?

In order to meet the purpose of this project, "to predict used cars' resale values', it seems that there are almost no correlations with such other other features as 'odometer' or 'usage_yr'. But people usually purchase used cars on bases of 'odometer' or 'usage_yr'. I had an insight that a kind of machine learning algorithm can identify the correlation between 'price' and other features. And I thought fuel economy can also play a key role in deciding used cars' resale

values, so I tried to merge the original dataset with fuel economy information data, 'youSaveSpend' on a basis of 'make' and 'manufacturer'.

3. Now that you've asked questions, hopefully you've found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?

In addition to the above, there are many null values over 'manufacturer' and 'type' and etc. By adopting 'VClass' data information from other dataset, 'type' was replaced by 'VClass' for identifying more specific car types.

And regarding null values over 'manufacturers' and other features, I filled in the null values with the most frequent values of each feature by utilizing for loops. By doing so, the dataset's null value size could be minimized, and it could be more guaranteeing a reliable dataset.