Advanced Regression Techniques in Econometrics
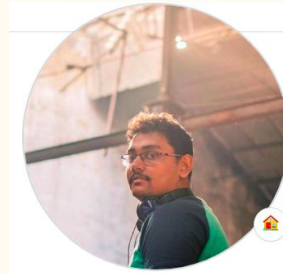
Implementation on a Real-world Dataset

Code Implementation

You can refer to my github account for more detailed and documented codes on the topic.

Check Out my Github : github

github.com/Nilotpal1998

www.linkedin.com/in/nilotpal-sarkar-b27085192/
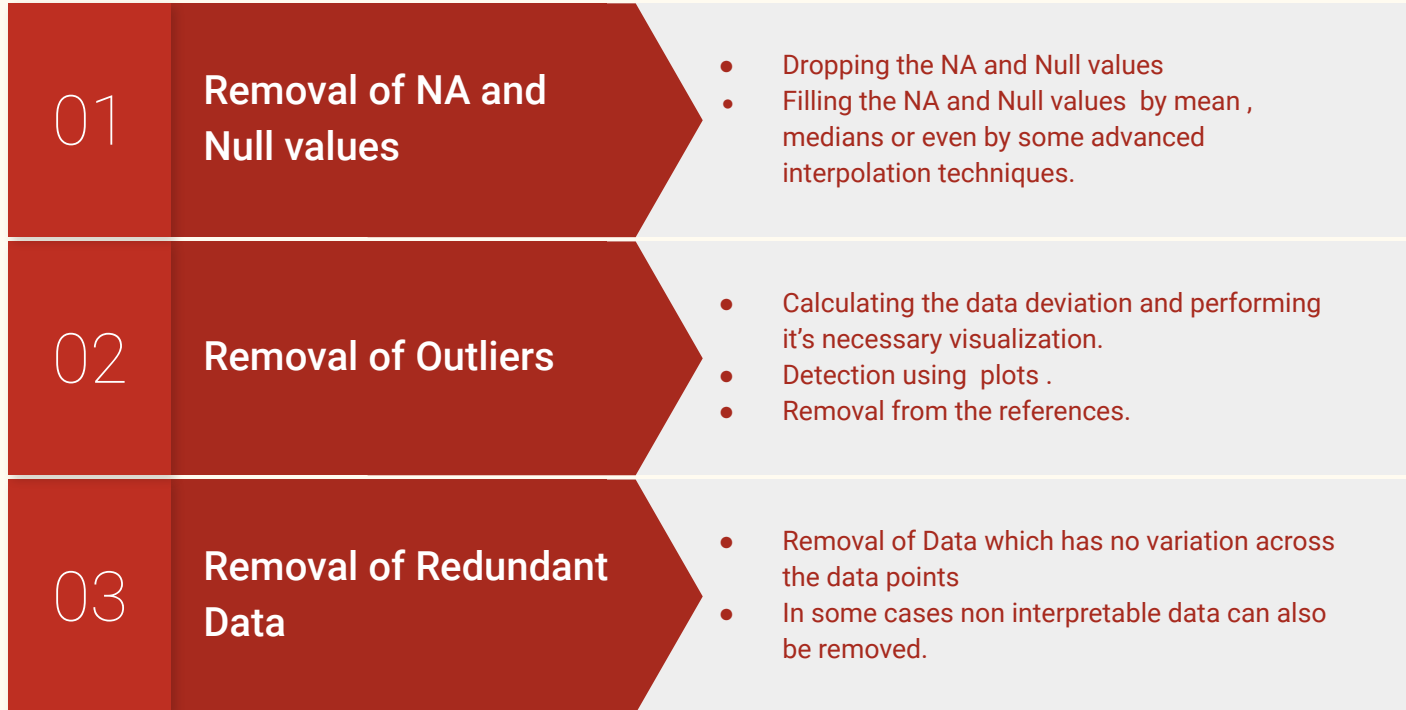
versatilenilot@gmail.com

Contents

This mainly contains some of advanced regression techniques that are taught in our econometrics course and I have tried to implement some of these into my present work on a simple real world data. These will mainly take care of the following topics:

1. Data Cleaning Details
2. Some simple EDA
3. Multicollinearity : Detection and Removal
4. Categorical Variables Treatment
5. Heteroskedasticity : Detection and Removal
6. Endogeneity : Detection and Removal

Data Cleaning

01 **Removal of NA and Null values**
- Dropping the NA and Null values
- Filling the NA and Null values by mean , medians or even by some advanced interpolation techniques.

02 **Removal of Outliers**
- Calculating the data deviation and performing it's necessary visualization.
- Detection using plots .
- Removal from the references.

03 **Removal of Redundant Data**
- Removal of Data which has no variation across the data points
- In some cases non interpretable data can also be removed.

Exploratory Data Analysis

These may contain necessary plots and inferences from the provided data . As in my case I have extensively used barplots , correlation plots , histograms , scatterplots , box plots , pie charts , doughnut plots , density plots and many more which are easily available in python seaborn and matplotlib packages.
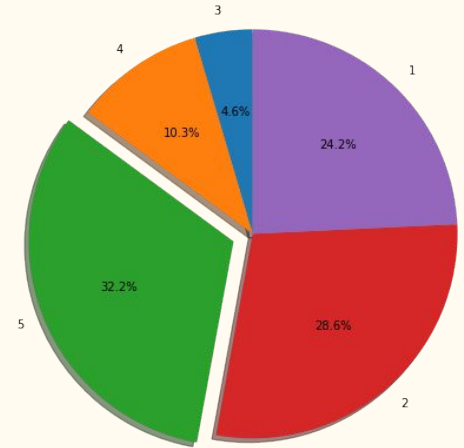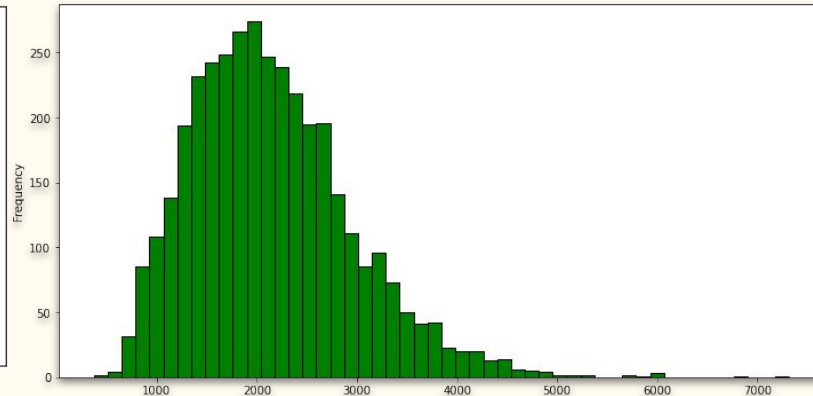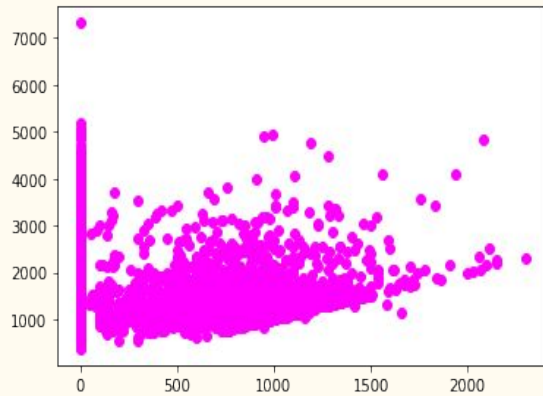
MultiCollinearity

Multicollinearity (or collinearity) occurs when one independent variable in a regression model is linearly correlated with another independent variable.
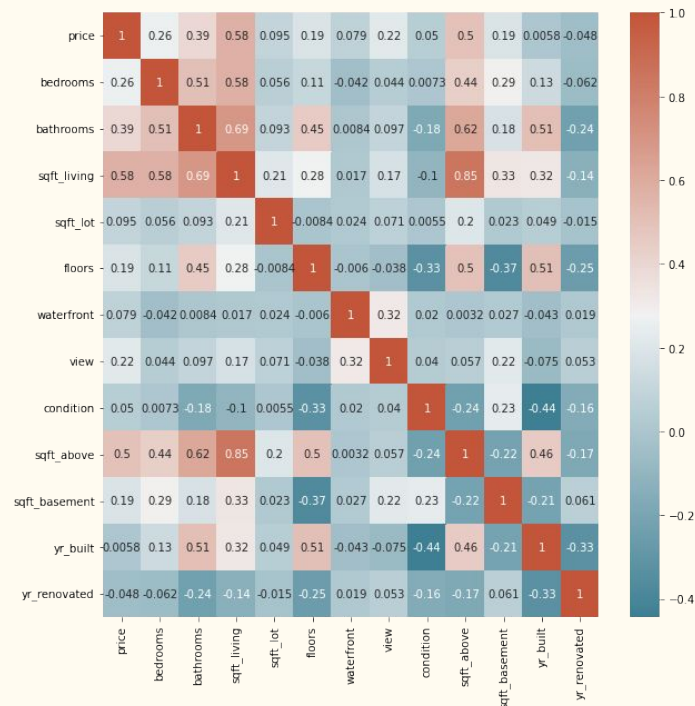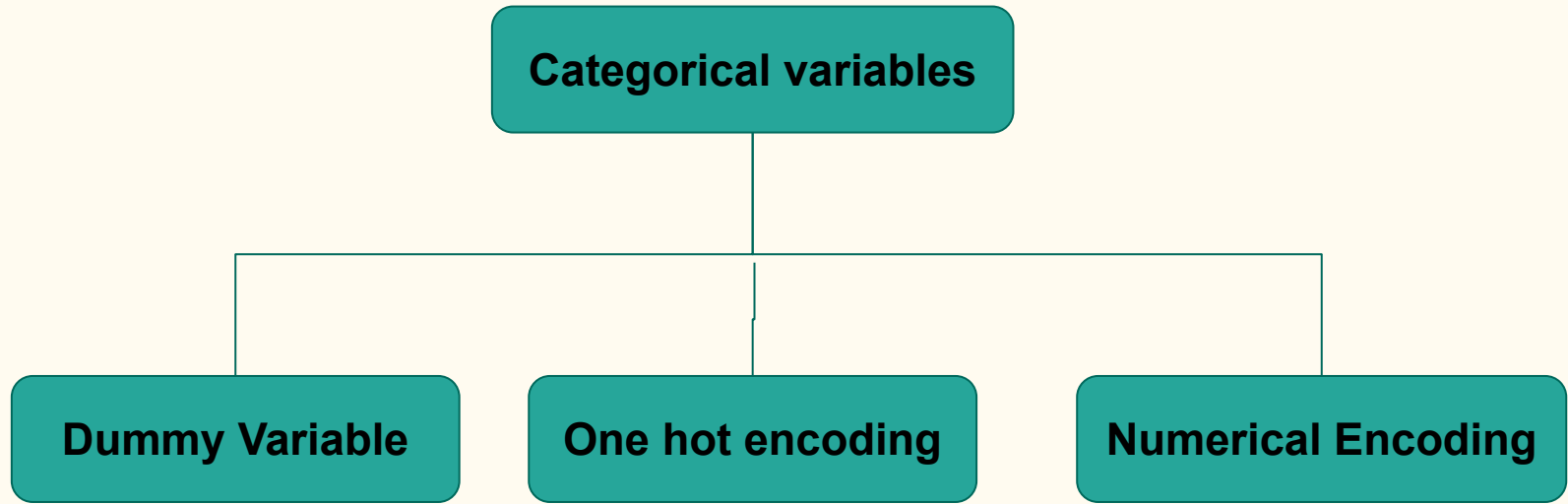
Detection :

Variance Inflation factor

$$VIF_i = \frac{1}{1 - R_i^2}$$

Solution :

Principal Component Analysis

Categorical Variables Treatment
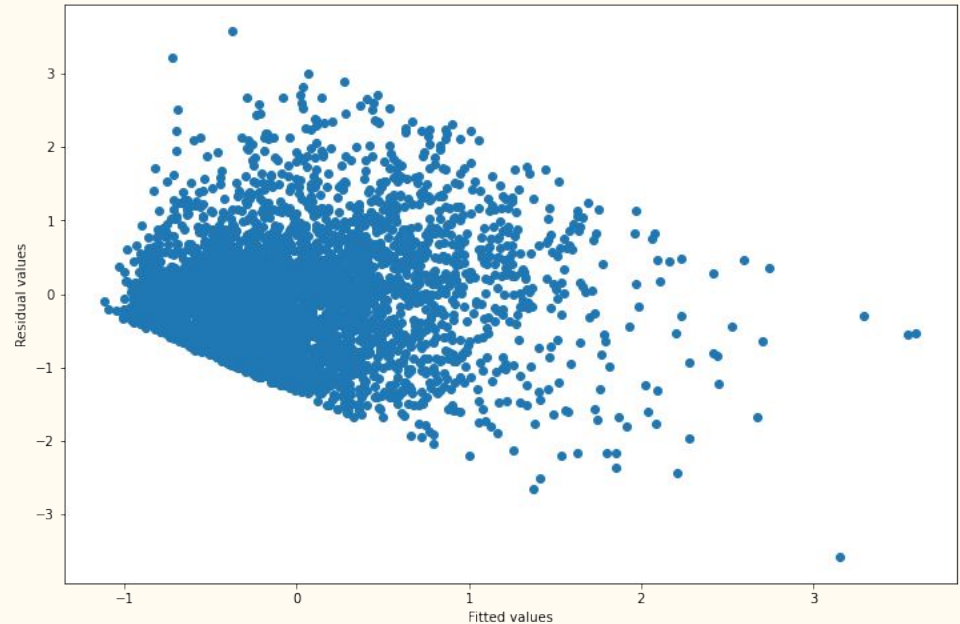
Heteroskedasticity

In simple terms, **heteroscedasticity** technically, refers to data with unequal variability (scatter) across a set of second, predictor variables.

Detection :

1.   Breusch-Pagan Test
2.   White Test

Solution :

Log transformation of the dependent variable.

Endogeneity

In econometrics endogeneity refers to the situation where the explanatory variables become correlated with the error term.

Detection :

Hausman Wu Test

Solution :

IV Regression Technique

Coding Compilation