Name: Prerana.M.S Class: 3 E

SRN: PES2UG19CS296 Date: 19.08.2020

STATISTICS FOR DATA SCIENCE

ASSIGNMENT 6

1. Choose a Messy dataset as you wish.

Dataset be found in. ([www.kaggle.com](http://www.kaggle.com/), <https://data.world/datasets/open-data> )

a) Handle the missing values.

            a) Find the number of missing data points through suitable visualization technique.

            b) Figure out why data is missing

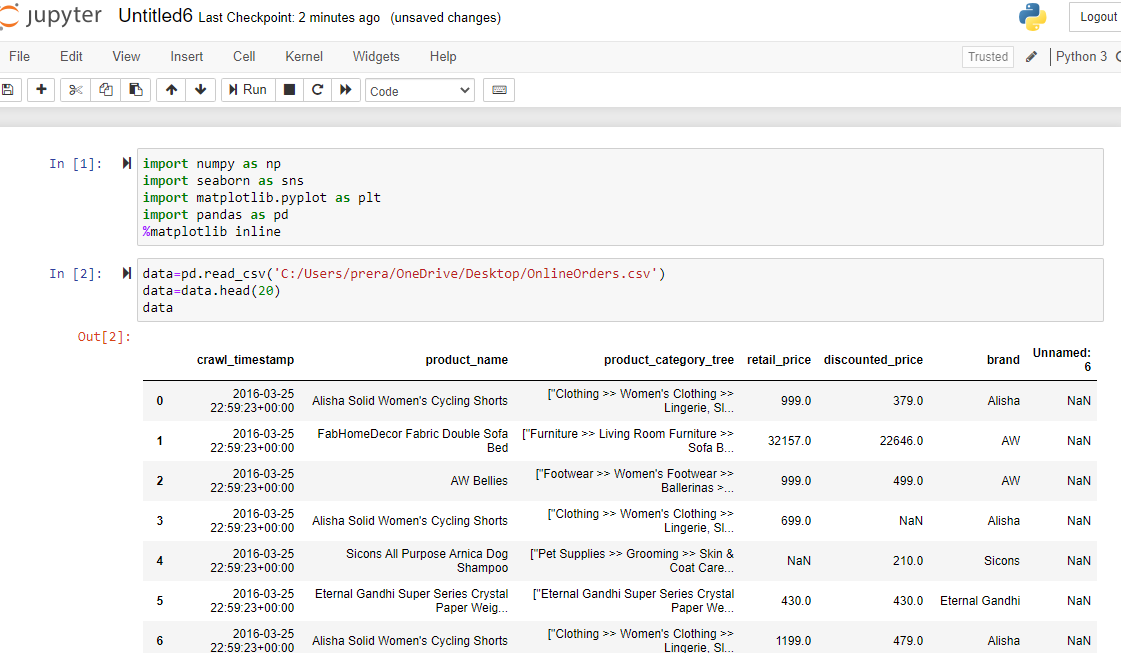
            c) Handle it by Dropping / Imputing / Replacing appropriately.

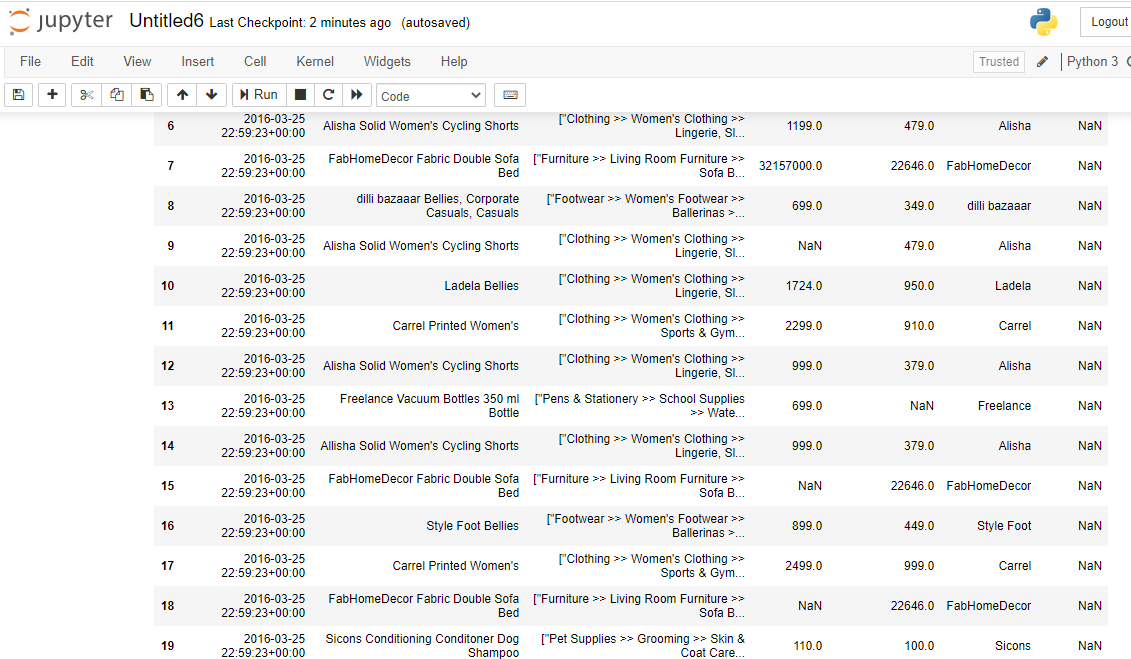
b) Filter the outliers

            a) Apply appropriate visualization technique and handle the outliers

Answer.

1. Handling missing data



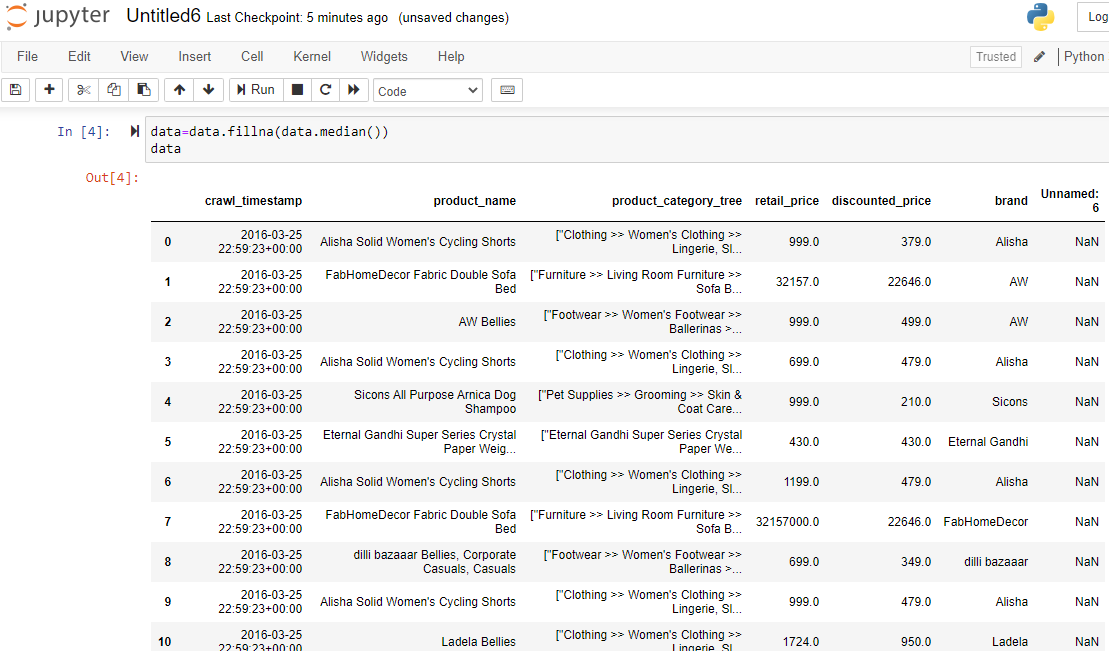




There are about 6 missing data from the dataset.

The most common reason for missing data is when it is not provided by the user and sometimes because of mishandling or negligence.

I will recover the loss by replacing the missing data with mean of the other data.



1. Handling Outliers

