

Filling Missing Values Automatically Based On Outliers

In [1]:

1 ls

Volume in drive C has no label.
Volume Serial Number is A007-E621

Directory of C:\Users\tswar\Documents\Data_Science_programming_interview\Pipeline\Automatic Null Filling

```
09/19/2022  10:47 AM    <DIR>          .
09/19/2022  10:47 AM    <DIR>          ..
09/19/2022  10:47 AM    <DIR>          .ipynb_checkpoints
09/19/2022  10:47 AM                4,998 Automatic Filling Missing Values Based o
n Outliers.ipynb
09/17/2022   09:53 PM                616,390 Automatic Filling Missing Values Based o
n Outliers.pdf
09/17/2022   09:01 PM                947,430 Avenger_Assemble.csv
09/17/2022   08:59 PM                29,822 claimants.csv
09/17/2022   09:49 PM                46,645 Filled_file.csv
                5 File(s)          1,645,285 bytes
                3 Dir(s)  84,174,589,952 bytes free
```

In [2]:

```
1 import pandas as pd
2 import numpy as np
```

In [3]:

```
1 def checkoutlier(dataseries):
2     outlier=[]
3     for i in dataseries:
4         z = (i - np.mean(dataseries))/np.std(dataseries)
5         if np.abs(z) > 2:
6             outlier.append(i)
7         else:
8             pass
9     return outlier
```

```
In [4]: 1 def starter(dataframe):
2         from sklearn.impute import SimpleImputer
3         meen = SimpleImputer(strategy="mean")
4         meed = SimpleImputer(strategy="median")
5         modee = SimpleImputer(strategy='most_frequent')
6         for i in dataframe.columns:
7             if type(dataframe[i][0]) == str:
8                 dataframe[i] = modee.fit_transform(dataframe[[i]])
9             else:
10                if len(checkoutlier(dataframe[i])) != 0:
11                    dataframe[i] = meed.fit_transform(dataframe[[i]])
12                else:
13                    dataframe[i] = meen.fit_transform(dataframe[[i]])
14         return dataframe
```

```
In [5]: 1 def mainer():
2         inputer = input("Please Path or File_Name")
3         df = pd.read_csv(inputer)
4         new_df = starter(df)
5         new_df.to_csv("Filled_file.csv", index=False)
```

```
In [6]: 1 mainer()
```

Please Path or File_NameAvenger_Assemble.csv

```
In [7]: 1 ls
```

Volume in drive C has no label.
Volume Serial Number is A007-E621

Directory of C:\Users\tswar\Documents\Data_Science_programming_interview\Pipeline\Automatic Null Filling

```
09/19/2022  10:47 AM    <DIR>          .
09/19/2022  10:47 AM    <DIR>          ..
09/19/2022  10:47 AM    <DIR>          .ipynb_checkpoints
09/19/2022  10:47 AM                4,998 Automatic Filling Missing Values Based o
n Outliers.ipynb
09/17/2022  09:53 PM                616,390 Automatic Filling Missing Values Based o
n Outliers.pdf
09/17/2022  09:01 PM                947,430 Avenger_Assemble.csv
09/17/2022  08:59 PM                29,822 claimants.csv
09/19/2022  10:49 AM                985,634 Filled_file.csv
           5 File(s)          2,584,274 bytes
           3 Dir(s)  84,173,803,520 bytes free
```

```
In [8]: 1 df_o = pd.read_csv("Avenger_Assemble.csv")
        2 df_o.isnull().sum()
```

```
Out[8]: Unnamed: 0      0
        bike_price    98
        km_driven     96
        make_year    102
        bike_name     102
        bike_model    102
        bike_address   98
        state         0
        dtype: int64
```

```
In [9]: 1 df = pd.read_csv("Filled_file.csv")
        2 df.isnull().sum()
```

```
Out[9]: Unnamed: 0      0
        bike_price     0
        km_driven      0
        make_year      0
        bike_name      0
        bike_model     0
        bike_address    0
        state          0
        dtype: int64
```

In [10]:

1df

Out[10]:

	Unnamed: 0	bike_price	km_driven	make_year	bike_name	bike_model	bike_address	
0	0.0	₹ 45,000	34,361 km	2019.0	Bajaj	Pulsar	Alapati Doctor Colony, Visakhapatnam, Andhra P...	Andhr
1	1.0	₹ 2,00,000	9,000 km	2019.0	KTM	390 Duke ABS	MVP Colony, Visakhapatnam, Andhra Pradesh	Andhr
2	2.0	₹ 1,35,000	17,000 km	2019.0	KTM	Others	NAD, Visakhapatnam, Andhra Pradesh	Andhr
3	3.0	₹ 1,10,000	28,000 km	1992.0	Royal Enfield	Bullet	Vajiapalem, Visakhapatnam, Andhra Pradesh	Andhr
4	4.0	₹ 50,000	51,400 km	2015.0	Bajaj	Avenger	Dasanna Pet, Vizianagaram, Andhra Pradesh	Andhr
...	
9826	9826.0	₹ 87,500	1,700 km	2019.0	TVS	Apache RTR	Azadpur, Delhi, Delhi	
9827	9827.0	₹ 40,000	55,000 km	2015.0	Suzuki	Gixxer	Abul Fazal Enclave Part 1, Delhi, Delhi	
9828	9828.0	₹ 80,000	25,000 km	2017.0	Yamaha	YZF R	Ranhola, Delhi, Delhi	
9829	9829.0	₹ 1,80,000	15,700 km	2019.0	Royal Enfield	Classic	Mayur Vihar Phase 3, Delhi, Delhi	
9830	9830.0	₹ 1,25,000	35,296 km	2017.0	Royal Enfield	Classic	Tilak Nagar, Delhi, Delhi	

9831 rows × 8 columns



In []:

1

In []:

1

