In this file we Practice how to deal with inappropriate Data:

We learnt handling with:

1- Numerical (Continuous ND & Discrete ND) Columns

2- Categorical Columns

3- Ordinal Columns

4- Date Columns

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 1 | Import libraries | import pandas as pd  import numpy as np |
| 2 | Read dataset | data = pd.read\_csv('datasetExample.csv') |
| 3 | information about the Dataset like:   * the number of columns * column labels * column data types * memory usage | data.info() |
| 4 | remove all duplicate records | data.drop\_duplicates(inplace=True) |
| 5 | Reset the Index of a Dataframe | data.set\_index(np.arange(0,len(data)) , inplace=True) |
| 6 | remove all duplicate columns | data.drop(['Age\_Group.1'] , axis = 1, inplace=True) |
| 7 | check a column has Negative Numbers? | data[data.CustomerID < 0] |
| 8 | check a column type | data.CustomerID.dtype |
| 9 | Replace negative value with NaN | data.Bill.loc[data.Bill < 0] = np.nan |
| 10 | Print Rating < 0 or Rating > 5 | data[ (data['Rating(1-5)'] < 1) | (data['Rating(1-5)'] > 5) ] |
| 11 | Replace Rating < 0 or Rating > 5 with nan | data['Rating(1-5)'].loc[ (data['Rating(1-5)'] < 1) | (data['Rating(1-5)'] > 5) ] = np.nan  data |
| 12 | Print unique values of an column | data['Age\_Group'].unique() |
| 13 | Replace values in a column | data.Hotel.replace(['Ibys'],'Ibis',inplace=True) |
| 14 |  |  |