Final_eda

June 24, 2020

1 EDA

Created by Gaurav Kumar IMPORT LIBRARY AND DATASET AS WELL AS INSTALL PANDAS PROFILING

```
[ ]: pip install pandas_profiling
```

```
[13]: import numpy as np
      import pandas as pd
      import matplotlib.pyplot as plt
```

```
[14]: df = pd.read_csv("C:/Users/admin.000/Desktop/Titan.csv")
```

```
[15]: df.head()
```

```
[15]:    PassengerId  Survived  Pclass  \
      0            1         0       3
      1            2         1       1
      2            3         1       3
      3            4         1       1
      4            5         0       3

                                                       Name     Sex   Age  SibSp  \
      0                            Braund, Mr. Owen Harris    male  22.0      1
      1  Cumings, Mrs. John Bradley (Florence Briggs Th…  female  38.0      1
      2                             Heikkinen, Miss. Laina  female  26.0      0
      3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
      4                           Allen, Mr. William Henry    male  35.0      0

         Parch            Ticket     Fare Cabin Embarked
      0      0         A/5 21171   7.2500   NaN        S
      1      0          PC 17599  71.2833   C85        C
      2      0  STON/O2. 3101282   7.9250   NaN        S
      3      0            113803  53.1000  C123        S
      4      0            373450   8.0500   NaN        S
```

```
[16]: from pandas_profiling import ProfileReport
      ProfileReport(df)
```

```
HBox(children=(FloatProgress(value=0.0, description='Summarize dataset', max=26.0, style=Progre
```

```
HBox(children=(FloatProgress(value=0.0, description='Generate report structure', max=1.0, style
```

```
HBox(children=(FloatProgress(value=0.0, description='Render HTML', max=1.0, style=ProgressStyle
```

```
<IPython.core.display.HTML object>
```

[16]:

1.1 Checking of null values

[17]: `missing_values_count = df.isnull().sum()`

[18]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    object
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64
 7   Parch        891 non-null    int64
 8   Ticket       891 non-null    object
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object
 11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 66.2+ KB
```

[8]: `df.dtypes.value_counts()`

[8]:
```
int64     5
object    5
```

```
float64      2
dtype: int64
```

[19]: `df.shape`

[19]: (891, 12)

[20]: `df.nunique(axis=0)`

[20]:
```
PassengerId    891
Survived         2
Pclass           3
Name           891
Sex              2
Age             88
SibSp            7
Parch            7
Ticket         681
Fare           248
Cabin          147
Embarked         3
dtype: int64
```

1.2 Drop unwanted coulumn from the dataset

[21]:
```python
cols = ['Name', 'Ticket', 'Cabin']
df1 = df.drop(cols, axis=1)
```

1.3 Data Imputations

[22]: `df1['Age'].fillna(df1['Age'].mean(), inplace=True)`

[ ]: `#df1['Age'].fillna(df1['Age'].median(), inplace=True)`

[23]:
```python
for column in df1[['Embarked']]:
    mode = df1[column].mode()
    df1[column] = df1[column].fillna(mode)
```

[24]: `df1.dropna()`

[24]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare \
0	1	0	3	male	22.000000	1	0	7.2500
1	2	1	1	female	38.000000	1	0	71.2833
2	3	1	3	female	26.000000	0	0	7.9250
3	4	1	1	female	35.000000	1	0	53.1000
4	5	0	3	male	35.000000	0	0	8.0500

```
 ..              …       …       …       …           …      …      …        …
886             887       0       2    male  27.000000       0      0   13.0000
887             888       1       1  female  19.000000       0      0   30.0000
888             889       0       3  female  29.699118       1      2   23.4500
889             890       1       1    male  26.000000       0      0   30.0000
890             891       0       3    male  32.000000       0      0    7.7500

      Embarked
0            S
1            C
2            S
3            S
4            S
..          …
886          S
887          S
888          S
889          C
890          Q

[889 rows x 9 columns]
```

1.4 Creation of Dummy variable

```python
[25]: dummies = []
      cols = ['Pclass', 'Sex', 'Embarked']
      for col in cols:
          dummies.append(pd.get_dummies(df[col]))
      titanic_dummies = pd.concat(dummies, axis=1)
      df1 = pd.concat((df1,titanic_dummies), axis=1)
```

```python
[26]: df1.head()
```

```
[26]:    PassengerId  Survived  Pclass     Sex   Age  SibSp  Parch     Fare  \
      0            1         0       3    male  22.0      1      0   7.2500
      1            2         1       1  female  38.0      1      0  71.2833
      2            3         1       3  female  26.0      0      0   7.9250
      3            4         1       1  female  35.0      1      0  53.1000
      4            5         0       3    male  35.0      0      0   8.0500

         Embarked  1  2  3  female  male  C  Q  S
      0         S  0  0  1       0     1  0  0  1
      1         C  1  0  0       1     0  1  0  0
      2         S  0  0  1       1     0  0  0  1
      3         S  1  0  0       1     0  0  0  1
      4         S  0  0  1       0     1  0  0  1
```

4

1.5 Feature Scaling

```
[27]: col_names = ['Age','Fare']
      features = df[col_names]
      from sklearn.preprocessing import StandardScaler
      scaler = StandardScaler().fit(features.values)
      features = scaler.transform(features.values)
```

```
[28]: scaled_features = pd.DataFrame(features, columns = col_names)
```

```
[29]: scaled_features.head()
```

```
[29]:         Age       Fare
      0 -0.530377 -0.502445
      1  0.571831  0.786845
      2 -0.254825 -0.488854
      3  0.365167  0.420730
      4  0.365167 -0.486337
```

```
[30]: from sklearn.preprocessing import MinMaxScaler
      scaler = MinMaxScaler(feature_range = (-3,3))
      scaler.fit_transform(df.Fare.values.reshape(-1,1))
```

```
[30]: array([[-2.91509365],
             [-2.16518559],
             [-2.90718858],
             [-2.37813422],
             [-2.90572468],
             [-2.90094299],
             [-2.39262685],
             [-2.75318604],
             [-2.86961547],
             [-2.64783424],
             [-2.80442263],
             [-2.68906711],
             [-2.90572468],
             [-2.63373159],
             [-2.90801774],
             [-2.81262048],
             [-2.65891072],
             [-2.84775414],
             [-2.78919804],
             [-2.91538644],
             [-2.69550828],
             [-2.84775414],
             [-2.90596827],
             [-2.58425169],
```

```
[-2.75318604],
[-2.63241408],
[-2.91538644],
[ 0.08005087],
[-2.90772495],
[-2.90753055],
[-2.67535561],
[-1.28406267],
[-2.90923804],
[-2.87703219],
[-2.03767968],
[-2.39101656],
[-2.91533725],
[-2.90572468],
[-2.78919804],
[-2.86834598],
[-2.88903619],
[-2.75406438],
[-2.90753055],
[-2.51305684],
[-2.90772495],
[-2.90572468],
[-2.81847609],
[-2.90923804],
[-2.74611012],
[-2.79154028],
[-2.53521095],
[-2.90865248],
[-2.10140745],
[-2.69550828],
[-2.27414795],
[-2.58425169],
[-2.87703219],
[-2.91533725],
[-2.67501364],
[-2.45074378],
[-2.91533725],
[-2.0631024 ],
[-2.02240591],
[-2.67325696],
[-2.67535561],
[-2.82145308],
[-2.87703219],
[-2.90445635],
[-2.90718858],
[-2.89855156],
[-2.87703219],
```

```
[-2.45074378],
[-2.13922533],
[-2.83072368],
[-2.33836525],
[-2.91040917],
[-2.90753055],
[-2.90572468],
[-2.66037462],
[-2.85390253],
[-2.89459902],
[-2.88874341],
[-2.90879887],
[-2.44840154],
[-2.87703219],
[-2.81437716],
[-2.59742681],
[-2.90572468],
[ 0.08005087],
[-2.90572468],
[-2.90572468],
[-2.90801774],
[-2.28356611],
[-2.75904165],
[-2.91509365],
[-2.90572468],
[-2.59415704],
[-2.25799701],
[-2.73064194],
[-2.69550828],
[-2.90753055],
[-2.90753055],
[-2.09486908],
[-2.89864876],
[-2.90718858],
[-2.90753055],
[-2.91040917],
[-2.90894526],
[-2.90753055],
[-2.71717404],
[-2.39101656],
[-2.83072368],
[-2.90572468],
[-2.88493726],
[-2.83067567],
[-2.90718858],
[-2.90923804],
[-2.75406438],
```

```
[-0.10122944],
[-2.63373159],
[-2.13922533],
[-2.90572468],
[-2.64783424],
[-2.84775414],
[-2.09486908],
[-2.86834598],
[-2.90923804],
[-2.91636198],
[-2.73815703],
[-2.91831424],
[-2.90753055],
[-2.9174359 ],
[-2.83018731],
[-2.69550828],
[-2.84775414],
[-2.82379533],
[-2.69219049],
[-2.37813422],
[-2.8920612 ],
[-2.07247137],
[-2.82145308],
[-2.90923804],
[-2.81437716],
[-2.92094926],
[-2.86532097],
[-2.56961266],
[-2.90870167],
[-2.59742681],
[-2.69550828],
[-2.84775414],
[-2.85331697],
[-2.22003274],
[-2.90572468],
[-2.83018731],
[-2.9143617 ],
[-2.28117468],
[-2.90943362],
[-2.90572468],
[-2.89855156],
[-2.18548465],
[-2.81144936],
[-2.81554828],
[-2.90894526],
[-2.89855156],
[-2.53521095],
```

```
[-2.75962721],
[-2.3558829 ],
[-2.67325696],
[-2.69638662],
[-2.33836525],
[-2.60767413],
[-2.65891072],
[-2.86961547],
[-2.90718858],
[-2.64051473],
[-2.90801774],
[-2.70175387],
[-2.66374159],
[-2.84775414],
[-3.        ],
[-2.18548465],
[-2.82374614],
[-2.63241408],
[-2.54326242],
[-2.74206038],
[-2.414439  ],
[-2.81847609],
[-2.68906711],
[-2.81847609],
[-2.90753055],
[-2.84775414],
[-2.84775414],
[-2.90801774],
[-2.69550828],
[-2.67535561],
[-1.28406267],
[-2.90923804],
[-2.90157656],
[-2.90923804],
[-2.84775414],
[-2.88874341],
[-2.18548465],
[-2.92392626],
[-2.91538644],
[-2.90572468],
[-2.87747136],
[-2.81437716],
[-2.77997545],
[-2.90923804],
[-2.63695218],
[-2.9174359 ],
[-2.75406438],
```

```
[-2.91509365],
[-2.84775414],
[-2.90923804],
[-1.67341155],
[-2.90718858],
[-2.68379706],
[-2.10653111],
[-2.87703219],
[-2.90572468],
[-2.84775414],
[-2.90572468],
[-2.90753055],
[-1.9459902 ],
[-2.89050009],
[-2.87703219],
[-2.91509365],
[-2.84775414],
[-2.70175387],
[-2.02240591],
[-2.90894526],
[-2.84189853],
[-2.63241408],
[-2.87703219],
[-2.91158029],
[-2.69550828],
[-2.69258047],
[-2.87703219],
[-2.85624477],
[-2.83072368],
[-2.81847609],
[-2.87703219],
[-2.91655756],
[-2.91538644],
[-1.9459902 ],
[-2.90894526],
[-2.83018731],
[-2.3845262 ],
[-2.69550828],
[-2.91509365],
[-2.87747136],
[-2.68906711],
[-2.81144936],
[-2.76328696],
[-2.82145308],
[-2.07247137],
[-1.98697947],
[ 3.        ],
```

```
   [-2.69550828],
   [-2.90923804],
   [-2.63241408],
   [-2.06720132],
   [-3.        ],
   [-2.90923804],
   [-2.87703219],
   [-2.53521095],
   [-2.90894526],
   [-1.20276689],
   [-1.41156858],
   [-2.63695218],
   [-3.        ],
   [-2.77163121],
   [-2.65217676],
   [-2.90923804],
   [-2.08701319],
   [-2.90923804],
   [-3.        ],
   [-2.65891072],
   [-2.76284779],
   [-2.90923804],
   [-2.90801774],
   [-2.88874341],
   [-2.90572468],
   [-2.69550828],
   [-2.89855156],
   [-2.88874341],
   [-2.90753055],
   [-2.84775414],
   [-2.90923804],
   [-2.0765703 ],
   [-1.93335145],
   [-2.84921804],
   [-2.8963557 ],
   [-2.90753055],
   [-2.67535561],
   [-2.91533725],
   [-1.2251646 ],
   [-2.64280779],
   [-0.10122944],
   [-2.90923804],
   [-2.72771413],
   [-3.        ],
   [-2.85536643],
   [-2.90572468],
   [-1.2251646 ],
```

```
[-1.70142127],
[-1.72464814],
[-2.71893072],
[-2.33328961],
[-2.02611485],
[ 0.07273136],
[-2.69550828],
[-2.90753055],
[-2.69258047],
[-2.90801774],
[-2.69550828],
[-2.83604292],
[-1.0692098 ],
[-1.4248409 ],
[-2.91509365],
[-2.90753055],
[-2.85536643],
[-2.66037462],
[-2.18548465],
[-1.41156858],
[-2.92695126],
[-2.84775414],
[-2.75962721],
[-2.32099283],
[-2.72771413],
[-2.66623023],
[-1.20276689],
[-2.78919804],
[-1.43479544],
[-2.90753055],
[-2.22003274],
[-1.4248409 ],
[-2.90572468],
[-2.58425169],
[-2.69550828],
[ 0.08005087],
[-2.84775414],
[-2.84775414],
[-2.84775414],
[-2.84775414],
[-2.84775414],
[-2.81144936],
[-2.8137916 ],
[-2.89855156],
[-2.891964  ],
[-2.5901073 ],
[-2.91533725],
```

```
[-2.79154028],
[-2.91538644],
[-2.88874341],
[-2.3558829 ],
[-2.84775414],
[-2.90772495],
[-2.90772495],
[-2.67325696],
[-2.67535561],
[-2.83072368],
[-2.9174359 ],
[-2.81847609],
[-2.91509365],
[-2.11873069],
[-2.91533725],
[-2.90923804],
[-2.18841245],
[-2.35071005],
[-2.92392626],
[-2.90572468],
[-1.41156858],
[-2.75318604],
[-2.03767968],
[-2.91509365],
[-0.52307696],
[-2.95300873],
[-2.90894526],
[-0.33540466],
[-2.81564549],
[-2.90718858],
[-2.39101656],
[-2.90753055],
[-2.13922533],
[-2.45074378],
[-2.84775414],
[-2.90948164],
[-2.85946536],
[-1.59465359],
[-2.90870167],
[-2.90718858],
[-1.67341155],
[-2.80442263],
[-2.90870167],
[-2.90801774],
[-2.69550828],
[-2.87703219],
[-2.85185307],
```

```
[-2.90718858],
[-2.90572468],
[-2.88493726],
[-2.81437716],
[-2.89855156],
[-2.75406438],
[-2.90923804],
[-2.78041462],
[-2.90894526],
[-2.70175387],
[-2.90753055],
[-2.91968094],
[-1.9459902 ],
[-3.        ],
[-2.90718858],
[-2.90572468],
[-2.61938535],
[-2.84775414],
[-2.84775414],
[-2.71717404],
[-2.90753055],
[-2.90943362],
[-2.90777414],
[-2.83135843],
[-2.76328696],
[-2.91509365],
[-2.69550828],
[-2.69550828],
[-2.90923804],
[-2.90572468],
[-2.68906711],
[-2.81144936],
[-2.69550828],
[-2.91655756],
[-2.3453428 ],
[-1.59465359],
[-2.59742681],
[-2.78041462],
[ 0.08005087],
[-2.87703219],
[-2.69258047],
[-2.88874341],
[-2.90894526],
[-2.84775414],
[-2.90499273],
[-2.04133944],
[-2.77163121],
```

```
[-2.68906711],
[-2.77446181],
[-2.64280779],
[-2.67501364],
[-2.76616558],
[-2.67501364],
[-1.95648111],
[-2.90572468],
[-2.90753055],
[-2.68906711],
[-2.39262685],
[-2.87703219],
[-2.90923804],
[-2.68906711],
[-2.90572468],
[-2.54911803],
[-2.84775414],
[-2.90572468],
[-2.9174359 ],
[-3.        ],
[-2.68906711],
[-2.90953083],
[-2.77446181],
[-2.91509365],
[-2.89855156],
[-2.67501364],
[-2.83848237],
[-2.88479087],
[-2.39101656],
[-2.75406438],
[-2.91748509],
[-2.91192226],
[-2.85609838],
[-2.45074378],
[-3.        ],
[-2.90572468],
[-2.88771868],
[-1.93335145],
[-2.70175387],
[-1.9459902 ],
[-2.65217676],
[-2.90572468],
[-2.8137916 ],
[-2.76616558],
[-2.91509365],
[-2.64280779],
[-2.42024542],
```

```
[-2.90572468],
[-2.83067567],
[-2.08340145],
[-2.82316058],
[-1.2251646 ],
[-2.90870167],
[-2.89855156],
[-2.90923804],
[-2.91065276],
[-2.88771868],
[-1.98697947],
[-1.72464814],
[-2.69550828],
[-2.68906711],
[-2.73620477],
[-2.33836525],
[-2.90923804],
[-2.90572468],
[-2.6921413 ],
[-2.30435353],
[-2.91221504],
[-2.60157492],
[-2.87703219],
[-2.71717404],
[-2.69550828],
[-2.90753055],
[-1.90500093],
[-2.90753055],
[-2.91538644],
[-2.32099283],
[-2.91533725],
[-2.90923804],
[-2.87703219],
[-0.40269499],
[-2.90718858],
[-2.86532097],
[-2.69550828],
[-2.91533725],
[-2.91533725],
[-2.73815703],
[-2.89855156],
[-2.69258047],
[-2.68906711],
[-1.75363341],
[-2.83018731],
[-2.42029461],
[-2.16850338],
```

```
[-2.63373159],
[-2.63373159],
[-2.69550828],
[-1.75363341],
[-2.69550828],
[-2.69550828],
[-2.83765321],
[-2.75962721],
[-2.56961266],
[-1.70142127],
[-2.69550828],
[-2.90831052],
[-2.91538644],
[-2.90894526],
[-2.68906711],
[-2.53623569],
[-0.33540466],
[-2.06720132],
[-2.79622477],
[-2.90923804],
[-2.90753055],
[-2.84189853],
[-2.90572468],
[-2.90572468],
[-2.71717404],
[-2.90753055],
[-2.75318604],
[-2.91533725],
[-2.90801774],
[-2.87703219],
[-2.39711576],
[-2.69097018],
[-2.90923804],
[-2.90572468],
[-2.83018731],
[-2.84775414],
[-2.3453428 ],
[-2.83067567],
[-2.90718858],
[-2.6486634 ],
[-1.70142127],
[-2.69550828],
[-2.5300873 ],
[-2.897966  ],
[-2.06720132],
[-2.8243317 ],
[-2.07247137],
```

```
[-2.90572468],
[-2.90572468],
[-2.91655756],
[-2.08340145],
[-2.91509365],
[-2.90923804],
[-2.69550828],
[-2.71717404],
[-2.61352974],
[-3.        ],
[-2.91538644],
[-2.33328961],
[-2.68379706],
[-2.90753055],
[-2.50344427],
[-2.90572468],
[-2.68906711],
[-2.81789053],
[-2.90753055],
[-2.64280779],
[-2.51305684],
[-1.20276689],
[-2.63373159],
[-2.9174359 ],
[-2.81847609],
[-2.90923804],
[-2.90572468],
[-2.2387707 ],
[-2.83135843],
[-2.81144936],
[-2.54326242],
[-2.87703219],
[-2.83072368],
[-2.3845262 ],
[-2.81564549],
[-2.90801774],
[-2.81144936],
[-2.621484  ],
[-2.85536643],
[-2.08701319],
[-2.90753055],
[-2.90943362],
[-2.6486634 ],
[-2.91738671],
[-2.64280779],
[-3.        ],
[-2.67325696],
```

```
[-2.84775414],
[-2.90718858],
[-2.69258047],
[-2.53521095],
[-2.81144936],
[-2.90801774],
[-2.18841245],
[-2.67325696],
[-2.33836525],
[-2.77446181],
[-2.10140745],
[-2.90753055],
[-2.58425169],
[-2.91158029],
[-2.91158029],
[-2.90753055],
[-2.73064194],
[-2.90123577],
[-2.90831052],
[-2.92094926],
[-2.13922533],
[-2.90753055],
[-2.81847609],
[-2.84775414],
[-1.67341155],
[-1.43479544],
[-2.91538644],
[-2.70033916],
[-2.91221504],
[-2.90718858],
[-2.13922533],
[-2.84775414],
[-2.90894526],
[-2.90572468],
[-2.39101656],
[-2.54326242],
[-2.39101656],
[-2.87703219],
[-2.84775414],
[-3.        ],
[-2.90894526],
[-2.90572468],
[-2.88474169],
[-2.45074378],
[ 3.        ],
[-2.90469995],
[-2.10140745],
```

```
[-2.891964  ],
[-2.45074378],
[-2.54326242],
[-2.51305684],
[-2.53521095],
[-2.88088752],
[-2.90870167],
[-0.52498003],
[-2.33246046],
[-2.84287407],
[-2.33836525],
[-2.91538644],
[-2.68906711],
[-2.84189853],
[-2.90572468],
[-2.90943362],
[-1.70142127],
[-2.91040917],
[-0.33540466],
[-2.6921413 ],
[-2.83072368],
[-2.90933525],
[-2.90801774],
[-2.69550828],
[-2.84189853],
[-2.6921413 ],
[-1.2251646 ],
[-2.82145308],
[-2.42024542],
[-2.68906711],
[-2.39101656],
[-2.88893899],
[-2.84775414],
[-2.91040917],
[-0.33540466],
[-2.87703219],
[-2.81847609],
[-2.90894526],
[-2.61352974],
[-2.91738671],
[-2.84775414],
[-2.84775414],
[-2.37813422],
[-2.89855156],
[-2.75406438],
[-2.90938443],
[-2.69550828],
```

```
[-2.90718858],
[-0.52498003],
[-2.77997545],
[-3.        ],
[-2.84775414],
[-2.84775414],
[-2.81144936],
[-2.59742681],
[ 3.        ],
[-2.90753055],
[-2.90753055],
[-2.6486634 ],
[-2.0765703 ],
[ 0.07273136],
[-2.81144936],
[-2.90718858],
[-2.16850338],
[-2.76284779],
[-2.84775414],
[-2.37813422],
[-2.90923804],
[-2.73064194],
[-2.85390253],
[-2.88874341],
[-2.90753055],
[-2.2387707 ],
[-2.83018731],
[-2.90870167],
[-2.86532097],
[-2.90572468],
[-1.98697947],
[-2.83018731],
[-2.91655756],
[-2.91533725],
[-1.59465359],
[-2.90894526],
[-2.08701319],
[-2.53623569],
[-2.90923804],
[-2.71717404],
[-2.90206492],
[-2.88874341],
[-2.90801774],
[-2.87703219],
[-2.91538644],
[-2.73064194],
[-2.90923804],
```

```
[-2.90923804],
[-2.85390253],
[-2.90938443],
[-0.52498003],
[-2.91533725],
[-2.33246046],
[-2.6486634 ],
[-2.72537189],
[-2.9174359 ],
[-2.91509365],
[-2.91221504],
[-2.65891072],
[-2.75904165],
[-2.07247137],
[-2.90923804],
[-2.69550828],
[-2.18548465],
[-2.64051473],
[-2.90753055],
[-2.84775414],
[-2.69633743],
[-2.89830796],
[-2.91533725],
[-2.71717404],
[-2.84775414],
[-2.69258047],
[-1.59465359],
[-2.90025905],
[-2.91831424],
[-2.90894526],
[-3.        ],
[-2.90894526],
[-2.84775414],
[-2.37813422],
[-2.90762775],
[-2.71717404],
[-2.87703219],
[-2.63373159],
[-2.90572468],
[-3.        ],
[-2.90718858],
[-2.56663567],
[-2.92446263],
[-2.67325696],
[-1.90500093],
[-2.89855156],
[-3.        ],
```

```
[-2.85390253],
[-2.53521095],
[-2.91860702],
[-2.33836525],
[-2.56663567],
[-2.90923804],
[-2.0631024 ],
[-2.83072368],
[-2.78041462],
[-2.91533725],
[-2.90801774],
[-2.90279687],
[-2.02611485],
[-2.89855156],
[-2.90572468],
[-2.33836525],
[-2.65217676],
[-2.90718858],
[-2.87703219],
[-2.63695218],
[-2.92460902],
[-2.89855156],
[-2.91158029],
[-2.18548465],
[-2.90753055],
[-2.61352974],
[-1.95648111],
[-2.63373159],
[-2.90894526],
[-2.82145308],
[-2.53857793],
[-2.69550828],
[-2.89050009],
[-1.0692098 ],
[-2.68906711],
[-2.77446181],
[-2.91533725],
[-2.83477459],
[-2.86532097],
[-2.69633743],
[-2.18548465],
[-2.84775414],
[-2.84775414],
[-2.8377024 ],
[-2.40863257],
[-2.88874341],
[-2.86961547],
```

```
         [-2.90753055],
         [-2.3845262 ],
         [-2.9414439 ],
         [-2.89459902],
         [-2.71893072],
         [-2.91538644],
         [-2.88469367],
         [-2.90753055],
         [-2.90753055],
         [-2.02611485],
         [-2.69550828],
         [-2.90753055],
         [-2.87683661],
         [-2.87703219],
         [-2.9174359 ],
         [-2.65891072],
         [-2.84775414],
         [-2.6486634 ],
         [-2.72537189],
         [-2.6486634 ],
         [-2.90923804]])
```

1.6   Data visualization after preprocessing

[31]:
```
pip install sweetviz
```

Requirement already satisfied: sweetviz in
c:\users\admin.000\anaconda3\lib\site-packages (1.0a4)
Requirement already satisfied: matplotlib>=3.1.3 in
c:\users\admin.000\anaconda3\lib\site-packages (from sweetviz) (3.2.1)
Requirement already satisfied: tqdm>=4.43.0 in
c:\users\admin.000\anaconda3\lib\site-packages (from sweetviz) (4.46.1)
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3 in
c:\users\admin.000\anaconda3\lib\site-packages (from sweetviz) (1.0.4)
Requirement already satisfied: importlib-resources>=1.2.0 in
c:\users\admin.000\anaconda3\lib\site-packages (from sweetviz) (1.5.0)
Requirement already satisfied: numpy>=1.16.0 in
c:\users\admin.000\anaconda3\lib\site-packages (from sweetviz) (1.18.1)
Requirement already satisfied: jinja2>=2.11.1 in
c:\users\admin.000\anaconda3\lib\site-packages (from sweetviz) (2.11.1)
Requirement already satisfied: scipy>=1.3.2 in
c:\users\admin.000\anaconda3\lib\site-packages (from sweetviz) (1.4.1)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in
c:\users\admin.000\anaconda3\lib\site-packages (from
matplotlib>=3.1.3->sweetviz) (2.4.6)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\admin.000\anaconda3\lib\site-packages (from

```
matplotlib>=3.1.3->sweetviz) (1.1.0)
Requirement already satisfied: python-dateutil>=2.1 in
c:\users\admin.000\anaconda3\lib\site-packages (from
matplotlib>=3.1.3->sweetviz) (2.8.1)
Requirement already satisfied: cycler>=0.10 in
c:\users\admin.000\anaconda3\lib\site-packages (from
matplotlib>=3.1.3->sweetviz) (0.10.0)
Requirement already satisfied: pytz>=2017.2 in
c:\users\admin.000\anaconda3\lib\site-packages (from
pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3->sweetviz) (2019.3)
Requirement already satisfied: zipp>=0.4; python_version < "3.8" in
c:\users\admin.000\anaconda3\lib\site-packages (from importlib-
resources>=1.2.0->sweetviz) (2.2.0)
Requirement already satisfied: importlib-metadata; python_version < "3.8" in
c:\users\admin.000\anaconda3\lib\site-packages (from importlib-
resources>=1.2.0->sweetviz) (1.5.0)
Requirement already satisfied: MarkupSafe>=0.23 in
c:\users\admin.000\anaconda3\lib\site-packages (from jinja2>=2.11.1->sweetviz)
(1.1.1)
Requirement already satisfied: setuptools in
c:\users\admin.000\anaconda3\lib\site-packages (from
kiwisolver>=1.0.1->matplotlib>=3.1.3->sweetviz) (45.2.0.post20200210)
Requirement already satisfied: six>=1.5 in
c:\users\admin.000\anaconda3\lib\site-packages (from python-
dateutil>=2.1->matplotlib>=3.1.3->sweetviz) (1.14.0)
Note: you may need to restart the kernel to use updated packages.
```

```python
[32]: import sweetviz as sv
my_report = sv.analyze(df)
my_report.show_html()
```

```
:FEATURES DONE:                    |                    | [  0%]    00:08  ->
(00:00 left)
:PAIRWISE DONE:                    |          | [100%]    00:00  ->
(00:00 left)

Creating Associations graph… DONE!
```

1.7 Correlation plot

```python
[ ]: import seaborn as sns
sns.pairplot(df)
```

```
[ ]: <seaborn.axisgrid.PairGrid at 0x1525cf90>
```

```python
[ ]:
```