feature-transformation

October 16, 2023

0.1 Feature Transformation

```
[1]: #importing necessary libraries
     import pandas as pd
```

```
[2]: df = pd.read_csv('supershops.csv')
```

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

```
[3]:    Marketing Spend  Administration  Transport     Area     Profit
     0        114523.61       136897.80  471784.10    Dhaka  192261.83
     1        162597.70       151377.59  443898.53      Ctg  191792.06
     2        153441.51       101145.55  407934.54  Rangpur  191050.39
     3        144372.41       118671.85  383199.62    Dhaka  182901.99
     4        142107.34        91391.77  366168.42  Rangpur  166187.94
```

0.1.1 Standardization

```
[4]: #dataframe for standardization
     df1 = df.copy()
```

```
[5]: from sklearn.preprocessing import StandardScaler
     scaler = StandardScaler()
```

```
[6]: #fit - calculating the mean and variance
     scaler.fit(df1[['Profit']])
```

```
[6]: StandardScaler(copy=True, with_mean=True, with_std=True)
```

```
[7]: df1.head()
```

```
[7]:    Marketing Spend  Administration  Transport     Area     Profit
     0        114523.61       136897.80  471784.10    Dhaka  192261.83
     1        162597.70       151377.59  443898.53      Ctg  191792.06
     2        153441.51       101145.55  407934.54  Rangpur  191050.39
     3        144372.41       118671.85  383199.62    Dhaka  182901.99
     4        142107.34        91391.77  366168.42  Rangpur  166187.94
```

```
[8]: #transforming using respective mean and variance
     x = df1['Profit'] = scaler.transform(df1[['Profit']])
```

```
[9]: df1.head()
```

```
[9]:    Marketing Spend  Administration  Transport     Area    Profit
     0        114523.61       136897.80  471784.10    Dhaka  2.011203
     1        162597.70       151377.59  443898.53      Ctg  1.999430
     2        153441.51       101145.55  407934.54  Rangpur  1.980842
     3        144372.41       118671.85  383199.62    Dhaka  1.776627
     4        142107.34        91391.77  366168.42  Rangpur  1.357740
```

```
[10]: #fit_transform - in a single shot
      df1['Marketing Spend'] = scaler.fit_transform(df1[['Marketing Spend']])
      df1['Administration'] = scaler.fit_transform(df1[['Administration']])
      df1['Transport'] = scaler.fit_transform(df1[['Transport']])
```

```
[11]: df1.head()
```

```
[11]:    Marketing Spend  Administration  Transport     Area    Profit
      0         0.897913        0.560753   2.165287    Dhaka  2.011203
      1         1.955860        1.082807   1.929843      Ctg  1.999430
      2         1.754364       -0.728257   1.626191  Rangpur  1.980842
      3         1.554784       -0.096365   1.417348    Dhaka  1.776627
      4         1.504937       -1.079919   1.273550  Rangpur  1.357740
```

```
[12]: #mean ~ 0
      x.mean()
```

```
[12]: -5.151434834260726e-16
```

```
[13]: #standard deviation ~ 1
      x.std()
```

```
[13]: 1.0
```

```
[14]: #median value
      df1.Profit.median()
```

```
[14]: -0.10111127105338139
```

```
[15]: #variance
      x.var()
```

```
[15]: 1.0
```

[16]:

```
#standard deviation of Profile column
x.std()
```

[16]:

```
1.0
```

0.1.2 Normalization

[17]:

```
#dataframe for normalization
df2 = df.copy()
```

[18]:

```
from sklearn.preprocessing import MinMaxScaler
m = MinMaxScaler(feature_range=(0, 1))

df2['Profit'] = m.fit_transform(df2[['Profit']])
df2['Marketing Spend'] = m.fit_transform(df2[['Marketing Spend']])
df2['Administration'] = m.fit_transform(df2[['Administration']])
df2['Transport'] = m.fit_transform(df2[['Transport']])
```

[19]:

```
df2.head()
```

[19]:

```
   Marketing Spend  Administration  Transport     Area    Profit
0         0.692617        0.651744   1.000000    Dhaka  1.000000
1         0.983359        0.761972   0.940893      Ctg  0.997355
2         0.927985        0.379579   0.864664  Rangpur  0.993178
3         0.873136        0.512998   0.812235    Dhaka  0.947292
4         0.859438        0.305328   0.776136  Rangpur  0.853171
```

0.1.3 Maximum Absolute Scaler

[20]:

```
#dataframe for Max Absolute Scaler
df3 = df.copy()
```

[21]:

```
from sklearn.preprocessing import MaxAbsScaler
mas = MaxAbsScaler()

df3['Marketing Spend'] = mas.fit_transform(df3[['Marketing Spend']])
df3['Administration'] = mas.fit_transform(df3[['Administration']])
df3['Transport'] = mas.fit_transform(df3[['Transport']])
df3['Profit'] = mas.fit_transform(df3[['Profit']])
```

[22]:

```
df3.head()
```

[22]:

```
   Marketing Spend  Administration  Transport     Area    Profit
0         0.692617        0.749527   1.000000    Dhaka  1.000000
1         0.983359        0.828805   0.940893      Ctg  0.997557
2         0.927985        0.553781   0.864664  Rangpur  0.993699
3         0.873136        0.649738   0.812235    Dhaka  0.951317
```

```
4          0.859438        0.500378   0.776136  Rangpur  0.864383
```

0.1.4 Robust Scaler

[23]:

```
#dataframe for Robust Scaler
df4 = df.copy()
```

[24]:

```
from sklearn.preprocessing import RobustScaler
RoSc = RobustScaler()

df4['Marketing Spend'] = RoSc.fit_transform(df4[['Marketing Spend']])
df4['Administration'] = RoSc.fit_transform(df4[['Administration']])
df4['Transport'] = RoSc.fit_transform(df4[['Transport']])
df4['Profit'] = RoSc.fit_transform(df4[['Profit']])
```

[25]:

```
df4.head()
```

[25]:

```
   Marketing Spend  Administration  Transport     Area    Profit
0         0.672530        0.345355   1.552016    Dhaka  1.698340
1         1.452113        0.697565   1.383714      Ctg  1.688874
2         1.303634       -0.524290   1.166654  Rangpur  1.673929
3         1.156567       -0.097977   1.017368    Dhaka  1.509736
4         1.119836       -0.761543   0.914576  Rangpur  1.172943
```