

```
[20]: #upgrading Libraries
!pip install --upgrade sagemaker
!pip install --upgrade s3fs

Requirement already satisfied: sagemaker in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (2.254.1)
Requirement already satisfied: attrs<26,>=24 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (25.4.0)
Requirement already satisfied: boto3<2.0,>=1.39.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (1.40.69)
Requirement already satisfied: cloudpickle>=2.2.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (3.1.2)
Requirement already satisfied: docker in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (7.1.0)
Requirement already satisfied: fastapi in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (0.121.0)
Requirement already satisfied: google-pasta in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (0.2.0)
Requirement already satisfied: graphene<4,>=3 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (3.4.3)
Requirement already satisfied: importlib-metadata<7.0,>=1.4.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (6.11.0)
Requirement already satisfied: jsonschema in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (4.25.1)
Requirement already satisfied: numpy<3.0,>=1.26.4 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (1.26.4)
Requirement already satisfied: omegaconf<3,>=2.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (2.3.0)
Requirement already satisfied: packaging<25,>=23.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (24.2)
Requirement already satisfied: pandas>=2.3.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (2.3.3)
Requirement already satisfied: pathos in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (0.3.4)
Requirement already satisfied: platformdirs in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (4.5.0)
Requirement already satisfied: protobuf<6.32,>=3.12 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from sagemaker) (6.31.1)

[21]: #importing Libraries
import pandas as pd #read csv file
import boto3 #to communicate with S3
from sagemaker import get_execution_role #share AWS permissions

[22]: #Loading csv file into dataframe
role = get_execution_role()
data_location = 's3://aws-assignment-025/employee_attitudes.csv'

#setting up dataframe
df_ratings = pd.read_csv(data_location)

[27]: #data preview of rows and columns
df_ratings.shape

[27]: (30, 5)

[25]: #checking first few rows
df_ratings.head(5)

[25]:
  rating_overall  rating_training  rating_raises  rating_benefits  rating_management
0              43              39              61              30              51
1              63              54              63              51              64
2              71              69              76              68              70
3              61              47              54              45              63
4              81              66              71              56              78

[24]: #checking column attributes
df_ratings.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ---
 0   rating_overall  30 non-null    int64
 1   rating_training 30 non-null    int64
 2   rating_raises   30 non-null    int64
 3   rating_benefits 30 non-null    int64
 4   rating_management 30 non-null    int64
dtypes: int64(5)
memory usage: 1.3 KB

[28]: #generating statistics of dataset

df_ratings.describe()

[28]:
  rating_overall  rating_training  rating_raises  rating_benefits  rating_management
count    30.000000     30.000000     30.000000     30.000000     30.000000
mean     64.633333     56.366667     64.633333     53.133333     66.600000
std      12.172562     11.737013     10.397226     12.235430     13.314757
min      40.000000     34.000000     43.000000     30.000000     37.000000
25%      58.750000     47.000000     58.250000     45.000000     58.500000
50%      65.500000     56.500000     63.500000     51.500000     65.000000
75%      71.750000     66.750000     71.000000     62.500000     77.000000
max      85.000000     75.000000     88.000000     83.000000     90.000000

[29]: #generating correlation matrix

df_ratings.corr()

[29]:
  rating_overall  rating_training  rating_raises  rating_benefits  rating_management
rating_overall    1.000000      0.623678      0.590139      0.426117      0.825418
rating_training    0.623678      1.000000      0.640314      0.493331      0.596736
rating_raises      0.590139      0.640314      1.000000      0.445478      0.669197
rating_benefits    0.426117      0.493331      0.445478      1.000000      0.558288
rating_management  0.825418      0.596736      0.669197      0.558288      1.000000

[ ]: 
```