practical-3

March 10, 2024

[ ]:
```
import pandas as pd
import numpy as np
```

[ ]:
```
df=pd.read_csv("/employee_data.csv")
```

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

[ ]:

	Unnamed: 0	id	groups	age	healthy_eating	active_lifestyle	salary
0	0	0	A	36	5	5	2297
1	1	1	A	55	3	5	1134
2	2	2	A	61	8	1	4969
3	3	3	O	29	3	6	902
4	4	4	O	34	6	2	3574

[ ]:
```
def categorize_age(age):
    if 20 <= age < 30:
        return '20-30'
    elif 30 <= age < 40:
        return '30-40'
    elif 40 <= age < 50:
        return '40-50'
    elif 50 <= age < 60:
        return '50-60'
    else:
        return 'Unknown'
```

[ ]:
```
df['AgeGroup'] = df['age'].apply(categorize_age)
```

[ ]:
```
statistics = df.groupby('AgeGroup')['salary'].describe()
```

[ ]:
```
print("Summary Statistics for Salary Grouped by Age Group:\n", statistics)
```

```
Summary Statistics for Salary Grouped by Age Group:
            count         mean          std    min     25%     50%     75%  \
AgeGroup
20-30       192.0  2327.848958  1070.521966  553.0  1369.0  2297.0  3110.0
30-40       221.0  2278.687783  1136.339093  553.0  1369.0  2065.0  3110.0
40-50       213.0  2235.464789  1091.709666  553.0  1252.0  2174.0  2878.0
```

```
50-60     223.0  2174.309417   999.019866  553.0  1366.0  2174.0  2878.0
Unknown   151.0  2092.046358  1103.201174  553.0  1191.5  1948.0  2878.0

              max
AgeGroup
20-30      5435.0
30-40      5435.0
40-50      5550.0
50-60      5435.0
Unknown    5204.0
```

[ ]:
```
df.groupby('AgeGroup')['salary'].median()
```

[ ]:
```
AgeGroup
20-30      2297.0
30-40      2065.0
40-50      2174.0
50-60      2174.0
Unknown    1948.0
Name: salary, dtype: float64
```

[ ]:
```
df.groupby('AgeGroup')['salary'].apply(lambda x: x.mode().iloc[0])
```

[ ]:
```
AgeGroup
20-30      2646
30-40       902
40-50      1020
50-60      1833
Unknown    1252
Name: salary, dtype: int64
```

[ ]:
```
max_age=df.groupby('AgeGroup')['salary'].max()
max_age
```

[ ]:
```
AgeGroup
20-30      5435
30-40      5435
40-50      5550
50-60      5435
Unknown    5204
Name: salary, dtype: int64
```

[ ]:
```
min_age=df.groupby('AgeGroup')['salary'].min()
min_age
```

[ ]:
```
AgeGroup
20-30      553
```

```
30-40      553
40-50      553
50-60      553
Unknown    553
Name: salary, dtype: int64
```

[ ]:
```
range=(max_age)-(min_age)
```

[ ]:
```
range
```

[ ]:
```
AgeGroup
20-30      4882
30-40      4882
40-50      4997
50-60      4882
Unknown    4651
Name: salary, dtype: int64
```

[ ]:
```
mid_range=range/2
mid_range
```

[ ]:
```
AgeGroup
20-30      2441.0
30-40      2441.0
40-50      2498.5
50-60      2441.0
Unknown    2325.5
Name: salary, dtype: float64
```

[ ]:
```
df.groupby('AgeGroup')['salary'].var()
```

[ ]:
```
AgeGroup
20-30      1.146017e+06
30-40      1.291267e+06
40-50      1.191830e+06
50-60      9.980407e+05
Unknown    1.217053e+06
Name: salary, dtype: float64
```

[ ]:
```
df.groupby('AgeGroup')['salary'].std()
```

[ ]:
```
AgeGroup
20-30      1070.521966
30-40      1136.339093
40-50      1091.709666
50-60       999.019866
Unknown    1103.201174
```

```
Name: salary, dtype: float64
```

[ ]:
```
df1=pd.read_csv("/Iris.csv")
```

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

[ ]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

[ ]:
```
def categorize_Species(Species):
    if (Species=='Iris-setosa'):
        return 'Iris-setosa'
    elif (Species=='Iris-versicolor'):
        return 'Iris-versicolor'


    else:
        return 'Iris-virginica'
        df1['SpeciesAge'] = df1['Species'].apply(categorize_Species)
```

[ ]:
```
df1['SpeciesGroup'] = df1['Species'].apply(categorize_Species)
```

[ ]:
```
statistics1 = df1.groupby('SpeciesGroup')['SepalLengthCm'].describe()
```

[ ]:
```
print("Summary Statistics for SepalLengthCm Grouped by species Group:\n",␣
 ↪statistics1)
```

```
Summary Statistics for SepalLengthCm Grouped by species Group:
                 count   mean       std  min    25%  50%  75%  max
SpeciesGroup
Iris-setosa       50.0  5.006  0.352490  4.3  4.800  5.0  5.2  5.8
Iris-versicolor   50.0  5.936  0.516171  4.9  5.600  5.9  6.3  7.0
Iris-virginica    50.0  6.588  0.635880  4.9  6.225  6.5  6.9  7.9
```

[ ]:
```
df1.groupby('SpeciesGroup')['SepalLengthCm'].median()
```

[ ]:
```
SpeciesGroup
Iris-setosa        5.0
Iris-versicolor    5.9
Iris-virginica     6.5
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
max_species=df1.groupby('SpeciesGroup')['SepalLengthCm'].max()
max_species
```

[ ]:
```
SpeciesGroup
Iris-setosa        5.8
Iris-versicolor    7.0
Iris-virginica     7.9
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
min_species=df1.groupby('SpeciesGroup')['SepalLengthCm'].min()
min_species
```

[ ]:
```
SpeciesGroup
Iris-setosa        4.3
Iris-versicolor    4.9
Iris-virginica     4.9
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
range1=(max_species)-(min_species)
```

[ ]:
```
range1
```

[ ]:
```
SpeciesGroup
Iris-setosa        1.5
Iris-versicolor    2.1
Iris-virginica     3.0
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
mid_range1=range1/2
mid_range1
```

[ ]:
```
SpeciesGroup
Iris-setosa        0.75
Iris-versicolor    1.05
Iris-virginica     1.50
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
df1.groupby('SpeciesGroup')['SepalLengthCm'].apply(lambda x: x.mode().iloc[0])
```

[ ]:
```
SpeciesGroup
Iris-setosa        5.0
Iris-versicolor    5.5
Iris-virginica     6.3
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
df1.groupby('SpeciesGroup')['SepalLengthCm'].var()
```

[ ]:
```
SpeciesGroup
Iris-setosa        0.124249
Iris-versicolor    0.266433
Iris-virginica     0.404343
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
df1.groupby('SpeciesGroup')['SepalLengthCm'].std()
```

[ ]:
```
SpeciesGroup
Iris-setosa        0.352490
Iris-versicolor    0.516171
Iris-virginica     0.635880
Name: SepalLengthCm, dtype: float64
```

[ ]:
```
df1.groupby('SpeciesGroup')['SepalLengthCm'].mean()
```

[ ]:
```
SpeciesGroup
Iris-setosa        5.006
Iris-versicolor    5.936
Iris-virginica     6.588
Name: SepalLengthCm, dtype: float64
```