Đỗ Minh Quân - 207CT10276

```
import pandas as pd
data = {
    'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eva', 'Frank', 'Grace', 'Hannah', 'Ivan', 'Jack', 'Kelly', 'Liam', 'Mona', 'Nina', 'Oscar'],
    'Age': [25, 30, 35, 28, 22, 45, 34, 31, 27, 29, 33, 40, 26, 32, 36],
    'Salary': [50000, 60000, 70000, 55000, 52000, 80000, 72000, 68000, 61000, 59000, 63000, 77000, 53000, 66000, 75000]
}

df = pd.DataFrame(data)
df
```

	Name	Age	Salary
0	Alice	25	50000
1	Bob	30	60000
2	Charlie	35	70000
3	David	28	55000
4	Eva	22	52000
5	Frank	45	80000
6	Grace	34	72000
7	Hannah	31	68000
8	Ivan	27	61000
9	Jack	29	59000
10	Kelly	33	63000
11	Liam	40	77000
12	Mona	26	53000
13	Nina	32	66000
14	Oscar	36	75000

Câu 2

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
import statsmodels.api as sm
df = pd.read_csv("datakiemtra.csv")
print(df)
```

```
   Unnamed: 0     Name  Age  Salary
0           1      Bob   30   60000
1           2  Charlie   35   70000
2           5    Frank   45   80000
3           6    Grace   34   72000
4           7   Hannah   31   68000
5           9     Jack   29   59000
6          10    Kelly   33   63000
7          11     Liam   40   77000
8          13     Nina   32   66000
9          14    Oscar   36   75000
```

Câu 3:lọc age lớn hơn 28

```
import pandas as pd

file_path = 'datakiemtra.csv'
data = pd.read_csv(file_path)
filtered_data = data[data['Age'] > 28]

filtered_data
```

	Unnamed: 0	Name	Age	Salary
0	1	Bob	30	60000
1	2	Charlie	35	70000
2	5	Frank	45	80000
3	6	Grace	34	72000
4	7	Hannah	31	68000
5	9	Jack	29	59000
6	10	Kelly	33	63000
7	11	Liam	40	77000
8	13	Nina	32	66000
9	14	Oscar	36	75000

câu 4

```
import pandas as pd
```

```
import pandas as pd

file_path = 'datakiemtra.csv'
data = pd.read_csv(file_path)
salary_mean = data['Salary'].mean()
salary_mean
```

```
69000.0
```

Câu 5

```
import pandas as pd

file_path = 'datakiemtra.csv'
data = pd.read_csv(file_path)
grouped_df = df.groupby('Age')['Salary'].sum().reset_index()
print(grouped_df)
```

```
   Age  Salary
0   29   59000
1   30   60000
2   31   68000
3   32   66000
4   33   63000
5   34   72000
6   35   70000
7   36   75000
8   40   77000
9   45   80000
```

Câu 6

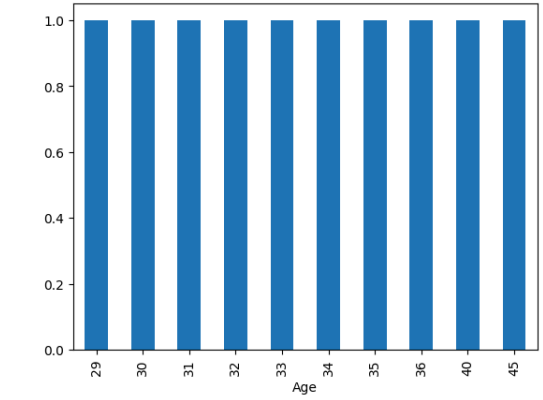
```
import pandas as pd

file_path = 'datakiemtra.csv'
data = pd.read_csv(file_path)

sorted_df = df.sort_values(by='Salary', ascending=False)
print(sorted_df)
```

```
   Unnamed: 0     Name  Age  Salary
2           5    Frank   45   80000
7          11     Liam   40   77000
9          14    Oscar   36   75000
3           6    Grace   34   72000
1           2  Charlie   35   70000
4           7   Hannah   31   68000
8          13     Nina   32   66000
6          10    Kelly   33   63000
0           1      Bob   30   60000
5           9     Jack   29   59000
```

Câu 7

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)

age = df.groupby('Age')['Age'].agg('count')
age.plot.bar()
```

```
<Axes: xlabel='Age'>
```

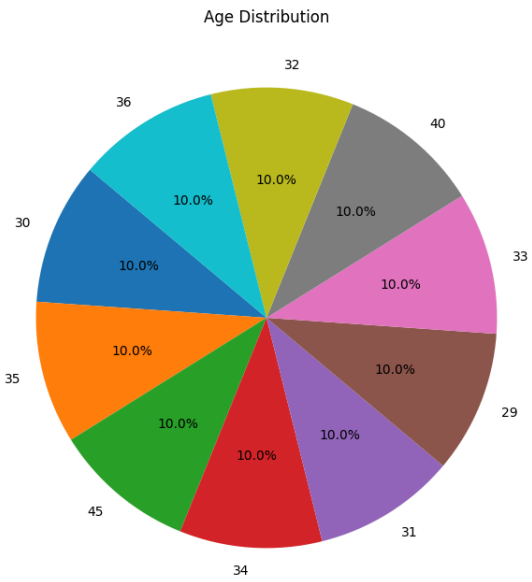


Câu 8

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
plt.figure(figsize=(10, 6))
plt.plot(df.index, df['Salary'], marker='o', linestyle='-', color='b')
plt.title('Salary')
plt.xlabel('Index')
plt.ylabel('Salary')
plt.grid(True)
plt.show()
```

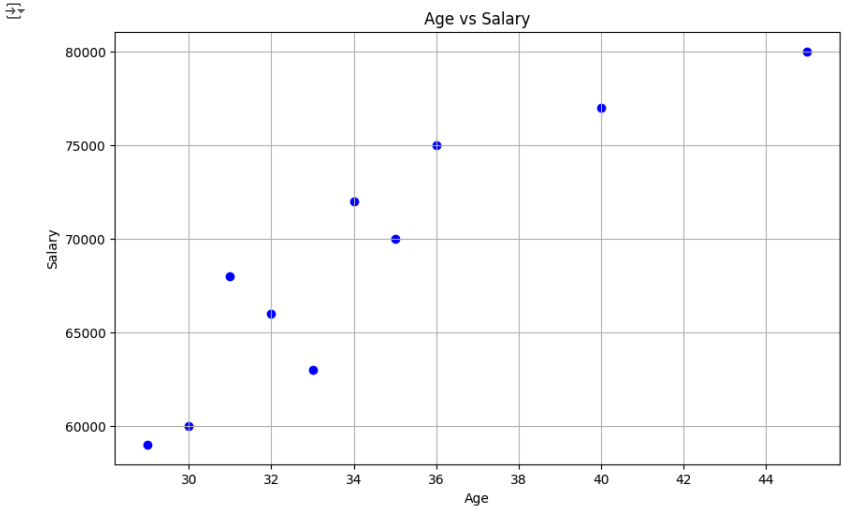
câu 9

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
age_counts = df['Age'].value_counts()

plt.figure(figsize=(8, 8))
plt.pie(age_counts, labels=age_counts.index, autopct='%1.1f%%', startangle=140)
plt.title('Age Distribution')
plt.show()
```



Câu 10

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)

plt.figure(figsize=(10, 6))
plt.scatter(df['Age'], df['Salary'], color='b')
plt.title('Age vs Salary')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.grid(True)
plt.show()
```



Câu 11

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
nan_exists = df.isnull().values.any()

if nan_exists:
    print("Có giá trị NaN trong DataFrame.")
else:
    print("Không có giá trị NaN trong DataFrame.")
```

```
Không có giá trị NaN trong DataFrame.
```

Câu 12

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
mean_age = df['Age'].mean()

df.loc[df['Age'] > 30, 'Age'] = mean_age

print(df)
```

```
   Unnamed: 0     Name   Age  Salary
0           1      Bob  30.0   60000
1           2  Charlie  34.5   70000
2           5    Frank  34.5   80000
3           6    Grace  34.5   72000
4           7   Hannah  34.5   68000
5           9     Jack  29.0   59000
6          10    Kelly  34.5   63000
7          11     Liam  34.5   77000
8          13     Nina  34.5   66000
9          14    Oscar  34.5   75000
```

Câu 13

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
scaler = MinMaxScaler()
df['Age'] = scaler.fit_transform(df[['Age']])

print(df)
```

```
   Unnamed: 0     Name     Age  Salary
0           1      Bob  0.0625   60000
1           2  Charlie  0.3750   70000
2           5    Frank  1.0000   80000
3           6    Grace  0.3125   72000
4           7   Hannah  0.1250   68000
5           9     Jack  0.0000   59000
6          10    Kelly  0.2500   63000
7          11     Liam  0.6875   77000
8          13     Nina  0.1875   66000
9          14    Oscar  0.4375   75000
```

Câu 14

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
def categorize_age(age):
    if age < 30:
        return 'Young'
    elif 30 <= age <= 60:
        return 'Middle-aged'
    else:
        return 'Old'
df['Age_group'] = df['Age'].apply(categorize_age)

print(df)
```

```
   Unnamed: 0     Name  Age  Salary    Age_group
0           1      Bob   30   60000  Middle-aged
1           2  Charlie   35   70000  Middle-aged
2           5    Frank   45   80000  Middle-aged
3           6    Grace   34   72000  Middle-aged
4           7   Hannah   31   68000  Middle-aged
5           9     Jack   29   59000        Young
```

```
6          10    Kelly   33   63000  Middle-aged
7          11     Liam   40   77000  Middle-aged
8          13     Nina   32   66000  Middle-aged
9          14    Oscar   36   75000  Middle-aged
```

Bắt đầu lập trình hoặc tạo mã bằng trí tuệ nhân tạo (AI).

Câu 15

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
df['Salary_pct_change'] = df['Salary'].pct_change()

print(df)
```

```
   Unnamed: 0     Name  Age  Salary  Salary_pct_change
0           1      Bob   30   60000                NaN
1           2  Charlie   35   70000           0.166667
2           5    Frank   45   80000           0.142857
3           6    Grace   34   72000          -0.100000
4           7   Hannah   31   68000          -0.055556
5           9     Jack   29   59000          -0.132353
6          10    Kelly   33   63000           0.067797
7          11     Liam   40   77000           0.222222
8          13     Nina   32   66000          -0.142857
9          14    Oscar   36   75000           0.136364
```

Câu 16

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

file_path = 'datakiemtra.csv'
df = pd.read_csv('datakiemtra.csv')
data = pd.read_csv(file_path)
df = df.drop_duplicates(subset=['Name'], keep='first')


print(df)
```

```
   Unnamed: 0     Name  Age  Salary
0           1      Bob   30   60000
```