In [1]:
```
import pandas as pd
```

In [2]:
```
mba1=pd.read_excel('BW_MBA_data.xlsx')
```

In [3]:
```
mba1.head()
```

Out[3]: —

In [4]:
```
mba=pd.ExcelFile('BW_MBA_data.xlsx')
mba_data=mba.parse('MBA Data')
mba_data.head()
```

Out[4]:

	Fulltime Business Week Ranking	School Name	State	Type	Enrollment	Avg GMAT	Resident Tuition, Fees	Pct International	Pct Female	P A
0	1	University of Chicago	Illinois	Private	1144	713.0	97165.0	35.0	35.0	
1	2	Harvard University	Massachusetts	Private	1801	720.0	101660.0	33.0	38.0	
2	3	Northwestern University	Illinois	Private	1200	711.0	93918.0	34.0	36.0	
3	4	University of Pennsylvania	Pennsylvania	Private	1651	714.0	104410.0	44.0	36.0	
4	5	University of Michigan	Michigan	Public	898	706.0	80879.0	27.0	34.0	

In [5]:
```
ug_data=mba.parse('Undergraduate Data')
ug_data.head()
```

Out[5]:

	2009 Rank	2008 Rank	School Name	Location	Type	Program Length	Annual Cost	Fulltime enrollment	Student Rank	Recruiter Rank	N S
0	1	2.0	Virginia (McIntire)	Charlottesville	Public	2	9490	655	1	52	
1	2	3.0	Notre Dame (Mendoza)	South Bend, Ind.	Private	3	36847	1669	2	12	
2	3	1.0	Pennsylvania (Wharton)	Philadelphia	Private	4	37526	2528	13	13	
3	4	6.0	Michigan (Ross)	Ann Arbor	Public	3	10848	1050	18	8	

	2009 Rank	2008 Rank	School Name	Location	Type	Program Length	Annual Cost	Fulltime enrollment	Student Rank	Recruiter Rank	N S
4	5	7.0	Brigham Young (Marriott)	Provo, Utah	Private	2	4110	1783	6	1	

In [6]:
```
NZstats=pd.read_csv('Nzdat.csv')
NZstats.head()
```

Out[6]:

	Year	Industry_aggregation_NZSIOC	Industry_code_NZSIOC	Industry_name_NZSIOC	Units	Variable
0	2019	Level 1	99999	All industries	Dollars (millions)	
1	2019	Level 1	99999	All industries	Dollars (millions)	
2	2019	Level 1	99999	All industries	Dollars (millions)	
3	2019	Level 1	99999	All industries	Dollars (millions)	
4	2019	Level 1	99999	All industries	Dollars (millions)	

In [7]:
```
employees=pd.read_excel('employee_data.xlsx')
employees.head()
```

Out[7]:

	Employee	Gender	Age	Prior Experience	Beta Experience	Education	Annual Salary
0	1	1	39	5	12	4	57700
1	2	0	44	12	8	6	76400
2	3	0	24	0	2	4	44000
3	4	1	25	2	1	4	41600
4	5	0	56	5	25	8	163900

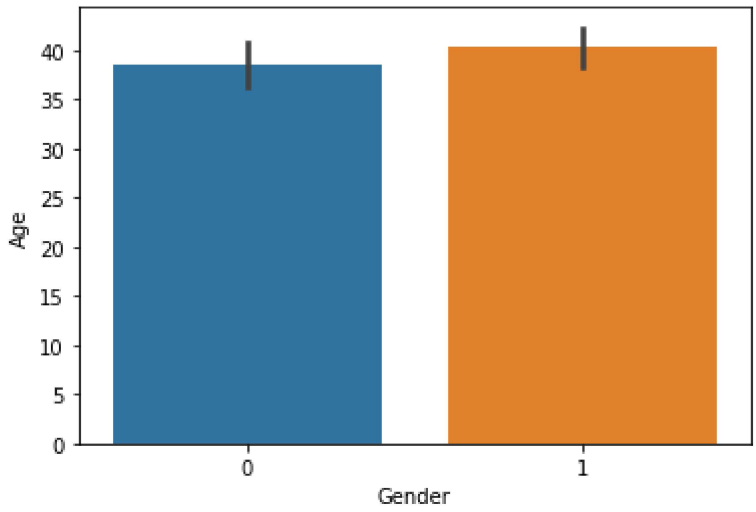
In [8]:
```
groups=['Gender', 'Education']
by_sex_edu = employees.groupby(groups,as_index=False).mean()
round(by_sex_edu[['Gender','Education','Age']],2)
```

Out[8]:

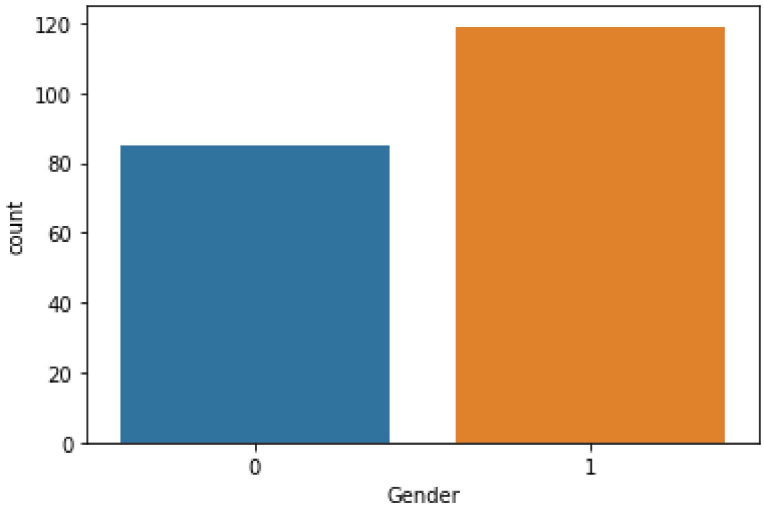
	Gender	Education	Age
0	0	0	33.50
1	0	2	37.22
2	0	4	37.22
3	0	6	40.75
4	0	8	45.50
5	1	0	30.38
6	1	2	39.70
7	1	4	42.72
8	1	6	38.26
9	1	8	35.40

In [9]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [10]:
```
sns.barplot(x='Gender', y='Age', data=employees)
plt.show()
```

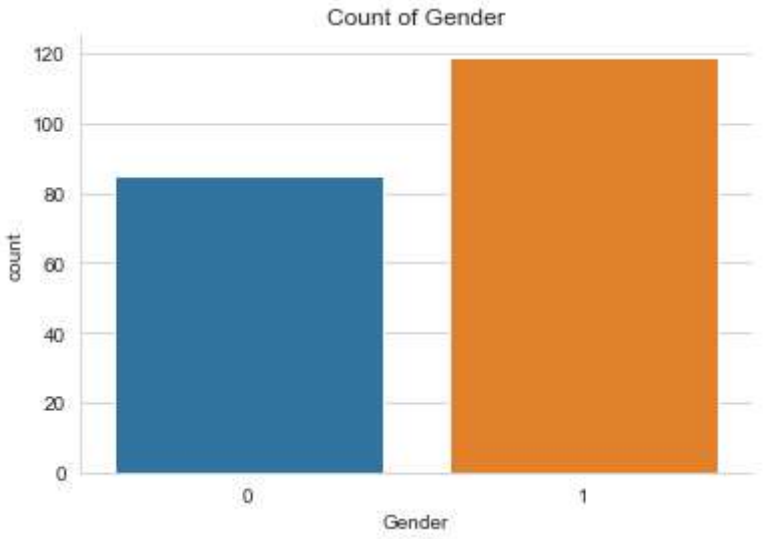


In [11]:
```
sns.countplot(x='Gender',data=employees)
plt.show()
```



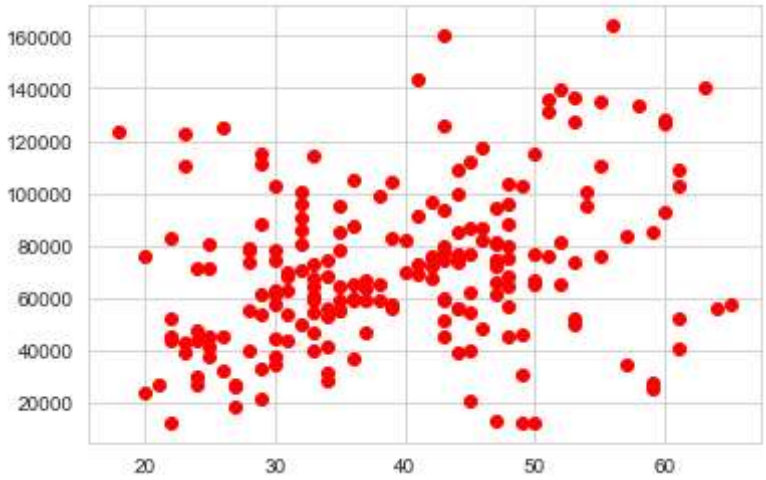
In [12]:

```
sns.set_style('whitegrid')
sns.countplot(x='Gender',data=employees)
plt.title('Count of Gender')
sns.despine()
plt.show()
```

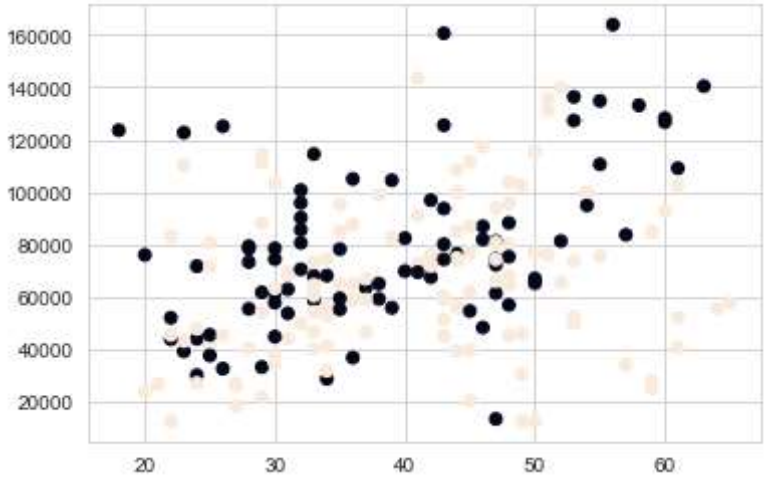


In [13]:

```
plt.scatter(employees['Age'], employees['Annual Salary'],color='red')
plt.show()
```



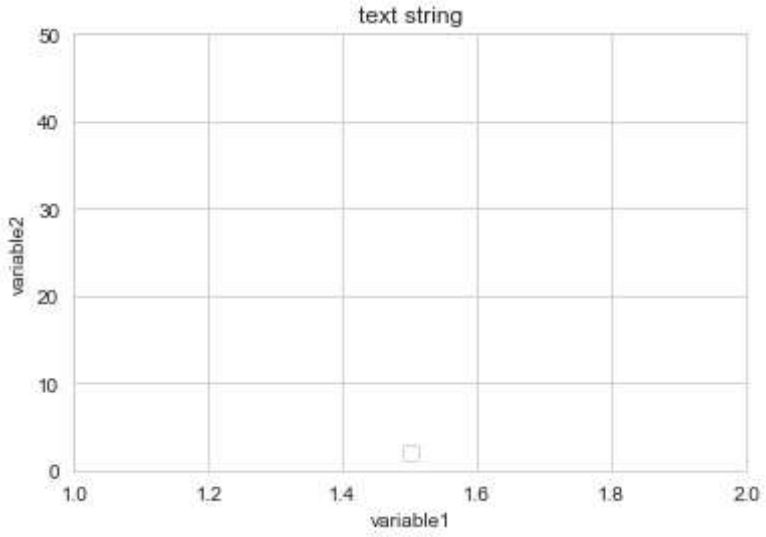
In [14]:
```
plt.scatter(employees['Age'], employees['Annual Salary'],c=employees['Gender'])
plt.show()
```



In [15]:
```
plt.xlabel('variable1')
plt.ylabel('variable2')
plt.xlim([1,2])
plt.ylim([0,50])
plt.title('text string')
plt.legend(loc='lower center')
```

```
No handles with labels found to put in legend.
```

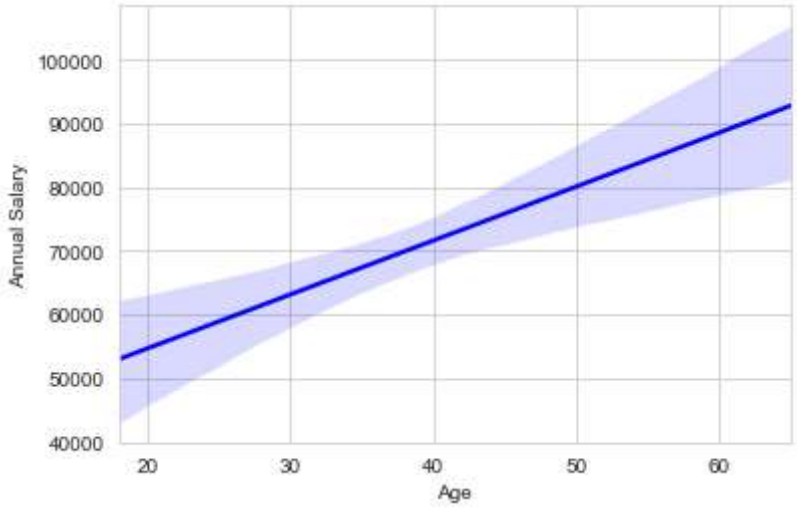
Out[15]:
```
<matplotlib.legend.Legend at 0x25d9c77a460>
```



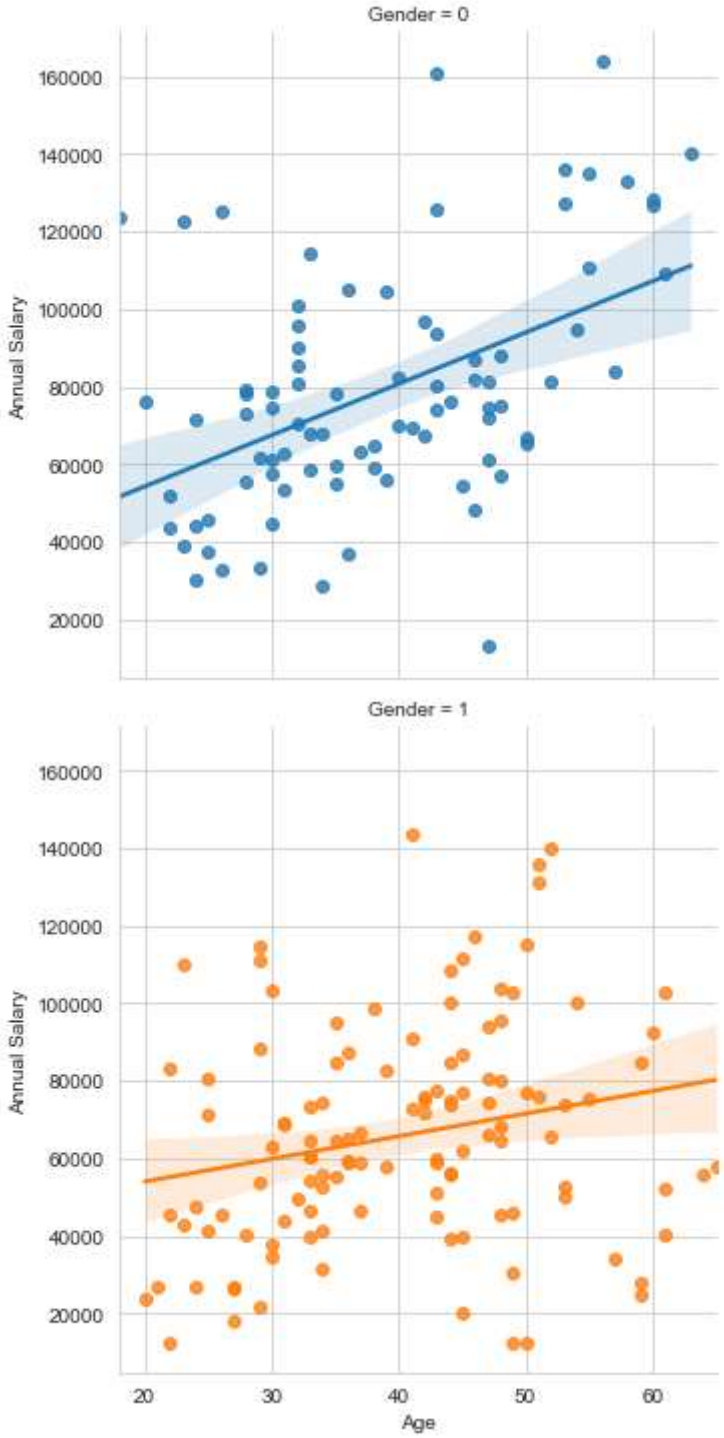
In [17]:
```
plt.savefig('name.png')
```

```
<Figure size 432x288 with 0 Axes>
```

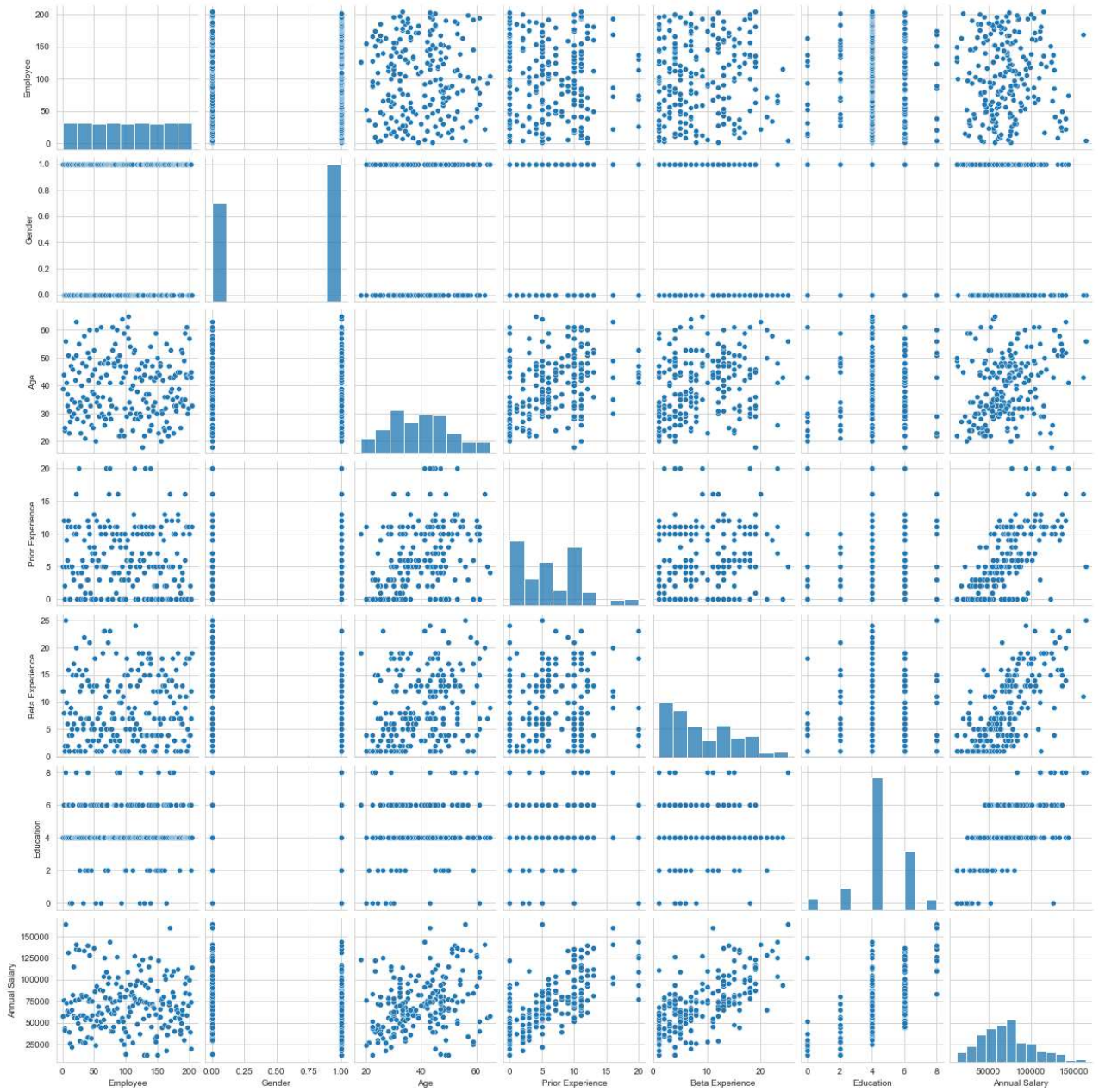
In [16]:
```
sns.regplot(x='Age',y='Annual Salary', data=employees, scatter=None, color='blue')
plt.show()
```



In [17]:
```
sns.lmplot(x='Age',y='Annual Salary', data=employees, hue='Gender', row='Gender')
plt.show()
```



In [18]:
```
sns.pairplot(employees)
plt.show()
```


Employee
Gender
Age
Prior Experience
Beta Experience
Education
Annual Salary

In [ ]: