

In [1]:

```
import pandas as pd
df=pd.read_csv('salary.csv')
df
```

Out[1]:

	Name	Team	Number	Position	Age	DOJ	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	06-Feb	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	06-Jun	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	06-May	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	06-May	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	06-Oct	231	NaN	5000000.0
...
452	Trey Lyles	Utah Jazz	41	PF	20	06-Oct	234	Kentucky	2239800.0
453	Shelvin Mack	Utah Jazz	8	PG	26	06-Mar	203	Butler	2433333.0
454	Raul Neto	Utah Jazz	25	PG	24	06-Jan	179	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21	C	26	07-Mar	256	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24	C	26	7-0	231	Kansas	947276.0

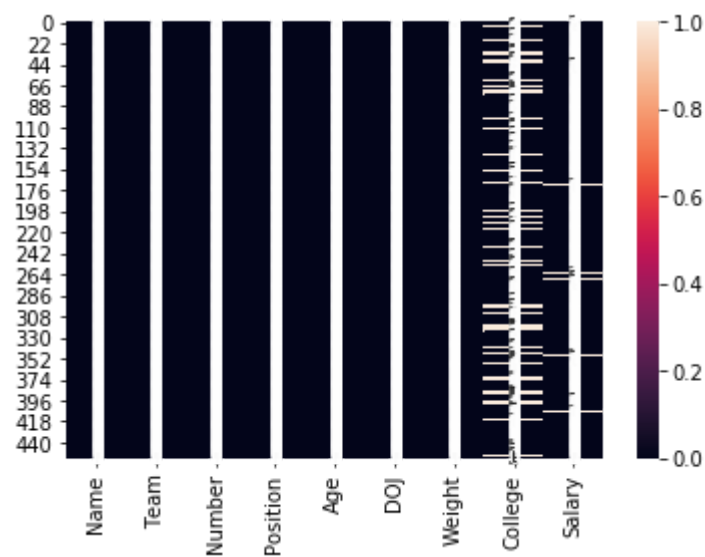
457 rows × 9 columns

In [3]:

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.heatmap(df.isnull(),annot=True)
```

Out[3]:

<matplotlib.axes._subplots.AxesSubplot at 0x23c2739db88>



In [4]:

```
df.head()
```

Out[4]:

	Name	Team	Number	Position	Age	DOJ	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0	PG	25	06-Feb	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99	SF	25	06-Jun	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30	SG	27	06-May	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28	SG	22	06-May	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8	PF	29	06-Oct	231	NaN	5000000.0

In [5]:

```
df.describe()
```

Out[5]:

	Number	Age	Weight	Salary
count	457.000000	457.000000	457.000000	4.460000e+02
mean	17.678337	26.938731	221.522976	4.842684e+06
std	15.966090	4.404016	26.368343	5.229238e+06
min	0.000000	19.000000	161.000000	3.088800e+04
25%	5.000000	24.000000	200.000000	1.044792e+06
50%	13.000000	26.000000	220.000000	2.839073e+06
75%	25.000000	30.000000	240.000000	6.500000e+06
max	99.000000	40.000000	307.000000	2.500000e+07

In [6]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 457 entries, 0 to 456
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Name        457 non-null    object
1   Team        457 non-null    object
2   Number      457 non-null    int64
3   Position    457 non-null    object
4   Age         457 non-null    int64
5   DOJ         457 non-null    object
6   Weight      457 non-null    int64
7   College     373 non-null    object
8   Salary      446 non-null    float64
dtypes: float64(1), int64(3), object(5)
memory usage: 32.3+ KB
```

In [7]:

```
df.isnull().sum()
```

Out[7]:

```
Name      0
Team      0
Number    0
Position  0
Age       0
DOJ       0
Weight    0
College   84
Salary    11
dtype: int64
```

In [8]:

```
df['Salary'].fillna(df['Salary'].mean(),inplace=True)
```

In [11]:

```
df['College'].fillna(df['College'].mode(),inplace=True)
```

In [12]:

```
df.isnull().sum()
```

Out[12]:

```
Name          0
Team          0
Number        0
Position      0
Age           0
DOJ           0
Weight        0
College      84
Salary        0
dtype: int64
```

In [13]:

```
df.shape
```

Out[13]:

```
(457, 9)
```

In [14]:

```
df.drop(columns=['College'],inplace=True)
```

In [15]:

```
df.isnull().sum()
```

Out[15]:

```
Name          0
Team          0
Number        0
Position      0
Age           0
DOJ           0
Weight        0
Salary        0
dtype: int64
```

In [24]:

```
df.skew()
```

Out[24]:

```
Number    1.668386  
Age        0.626349  
Weight     0.113788  
Salary     1.595511  
dtype: float64
```

In [26]:

```
x=df['Age']  
y=df['Salary']
```

In [33]:

```
sns.countplot(df['Age'],hue=df['Team'])  
pd.crosstab(df['Team'],df['Age']).apply(lambda r:round((r/r.sum())*100,1),axis=1)
```

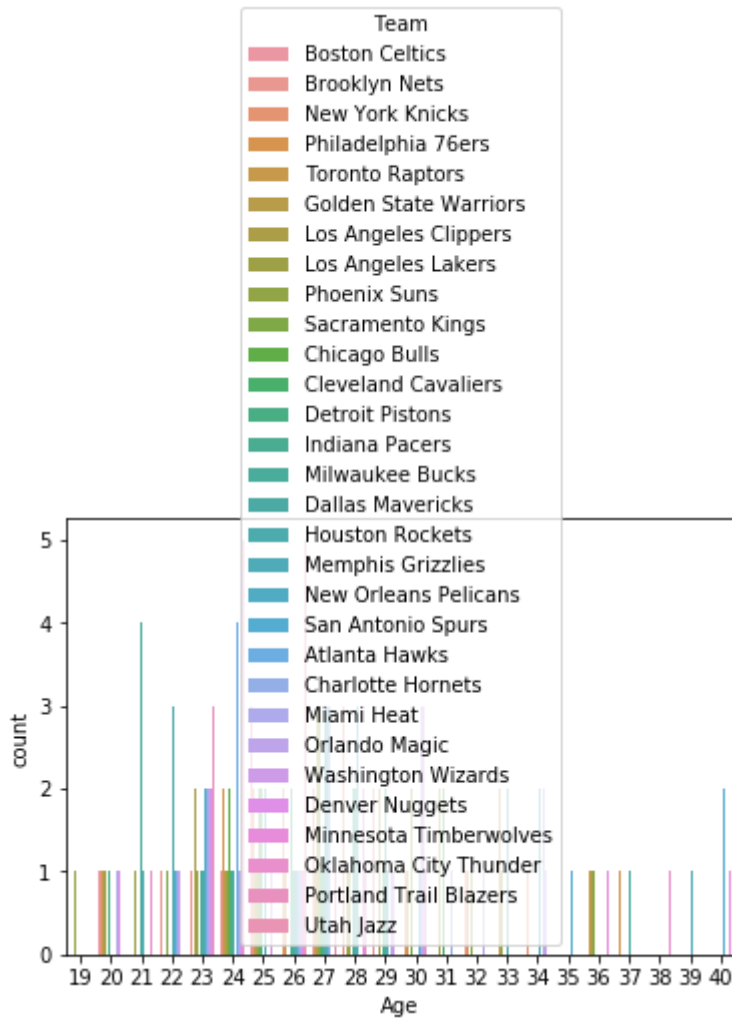
Out[33]:

	Age	19	20	21	22	23	24	25	26	27	28	...	31	32	33
Team															
Atlanta Hawks	0.0	0.0	0.0	6.7	0.0	26.7	0.0	6.7	13.3	0.0	...	20.0	6.7	0.0	
Boston Celtics	0.0	6.7	6.7	20.0	0.0	6.7	20.0	6.7	20.0	0.0	...	0.0	0.0	0.0	
Brooklyn Nets	0.0	0.0	13.3	6.7	6.7	6.7	6.7	20.0	13.3	20.0	...	0.0	6.7	0.0	
Charlotte Hornets	0.0	0.0	6.7	6.7	13.3	13.3	0.0	6.7	20.0	6.7	...	6.7	0.0	0.0	
Chicago Bulls	0.0	0.0	6.7	0.0	6.7	13.3	13.3	6.7	20.0	0.0	...	13.3	0.0	0.0	
Cleveland Cavaliers	0.0	0.0	0.0	0.0	0.0	6.7	26.7	0.0	6.7	0.0	...	13.3	0.0	13.3	
Dallas Mavericks	0.0	0.0	0.0	6.7	0.0	6.7	0.0	0.0	6.7	13.3	...	26.7	6.7	13.3	
Denver Nuggets	0.0	6.7	20.0	0.0	13.3	6.7	6.7	6.7	6.7	13.3	...	0.0	0.0	0.0	
Detroit Pistons	0.0	6.7	0.0	6.7	26.7	0.0	13.3	13.3	0.0	6.7	...	6.7	0.0	6.7	
Golden State Warriors	0.0	6.7	0.0	0.0	6.7	6.7	6.7	20.0	0.0	13.3	...	6.7	6.7	13.3	
Houston Rockets	0.0	0.0	0.0	20.0	6.7	6.7	6.7	6.7	20.0	0.0	...	0.0	0.0	0.0	
Indiana Pacers	0.0	6.7	0.0	6.7	6.7	6.7	13.3	6.7	6.7	13.3	...	0.0	0.0	0.0	
Los Angeles Clippers	0.0	0.0	0.0	0.0	13.3	0.0	6.7	0.0	20.0	6.7	...	13.3	0.0	0.0	
Los Angeles Lakers	0.0	6.7	6.7	0.0	13.3	13.3	6.7	0.0	6.7	0.0	...	13.3	0.0	6.7	
Memphis Grizzlies	0.0	0.0	5.6	5.6	11.1	11.1	11.1	0.0	0.0	16.7	...	5.6	0.0	0.0	
Miami Heat	0.0	6.7	0.0	6.7	6.7	6.7	0.0	6.7	0.0	0.0	...	6.7	6.7	6.7	
Milwaukee Bucks	6.2	0.0	25.0	0.0	6.2	12.5	12.5	6.2	12.5	6.2	...	0.0	6.2	0.0	
Minnesota Timberwolves	0.0	14.3	14.3	0.0	7.1	0.0	21.4	7.1	0.0	7.1	...	0.0	0.0	0.0	
New Orleans Pelicans	0.0	0.0	0.0	0.0	10.5	0.0	21.1	10.5	15.8	15.8	...	5.3	0.0	0.0	
New York Knicks	0.0	6.2	0.0	0.0	18.8	6.2	12.5	12.5	0.0	12.5	...	0.0	12.5	6.2	
Oklahoma City Thunder	0.0	0.0	6.7	6.7	0.0	33.3	0.0	6.7	13.3	6.7	...	0.0	6.7	0.0	
Orlando Magic	0.0	7.1	7.1	7.1	14.3	14.3	7.1	21.4	0.0	0.0	...	0.0	7.1	0.0	
Philadelphia 76ers	0.0	13.3	0.0	33.3	0.0	13.3	20.0	0.0	6.7	0.0	...	0.0	6.7	0.0	
Phoenix Suns	6.7	0.0	6.7	13.3	6.7	6.7	13.3	6.7	6.7	6.7	...	6.7	6.7	6.7	

	Age	19	20	21	22	23	24	25	26	27	28	...	31	32	33
Team															
Portland Trail Blazers	0.0	13.3	0.0	0.0	20.0	20.0	13.3	6.7	6.7	6.7	...	0.0	0.0	0.0	
Sacramento Kings	0.0	0.0	0.0	6.7	6.7	13.3	20.0	0.0	20.0	6.7	...	0.0	0.0	0.0	
San Antonio Spurs	0.0	0.0	0.0	6.7	0.0	6.7	0.0	6.7	13.3	6.7	...	0.0	0.0	6.7	
Toronto Raptors	0.0	6.7	0.0	0.0	20.0	20.0	6.7	6.7	6.7	0.0	...	0.0	0.0	0.0	
Utah Jazz	0.0	13.3	0.0	0.0	20.0	20.0	0.0	33.3	0.0	13.3	...	0.0	0.0	0.0	
Washington Wizards	0.0	6.7	0.0	6.7	6.7	6.7	6.7	6.7	6.7	0.0	...	0.0	6.7	13.3	

30 rows × 22 columns





In [34]:

```
print(df['Salary'].skew())
```

1.5955114443826113

In [35]:

```
print(df['Age'].skew())
```

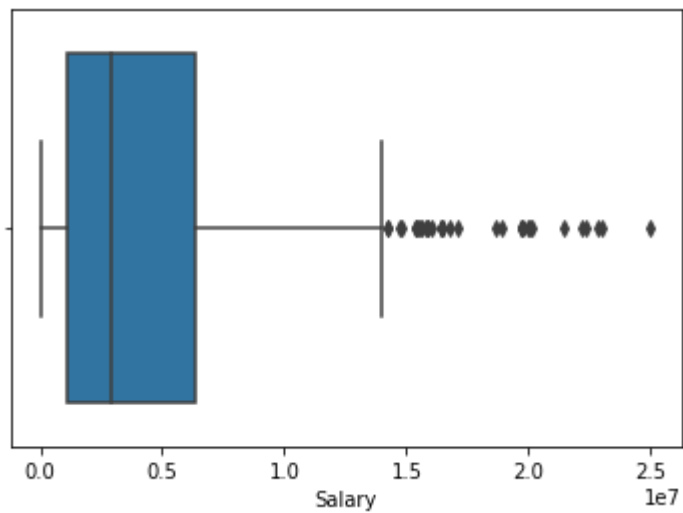
0.6263487611614392

In [36]:

```
sns.boxplot(df['Salary'])
```

Out[36]:

<matplotlib.axes._subplots.AxesSubplot at 0x23c2bd32b08>



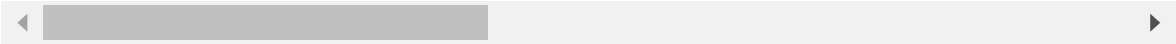
In [38]:

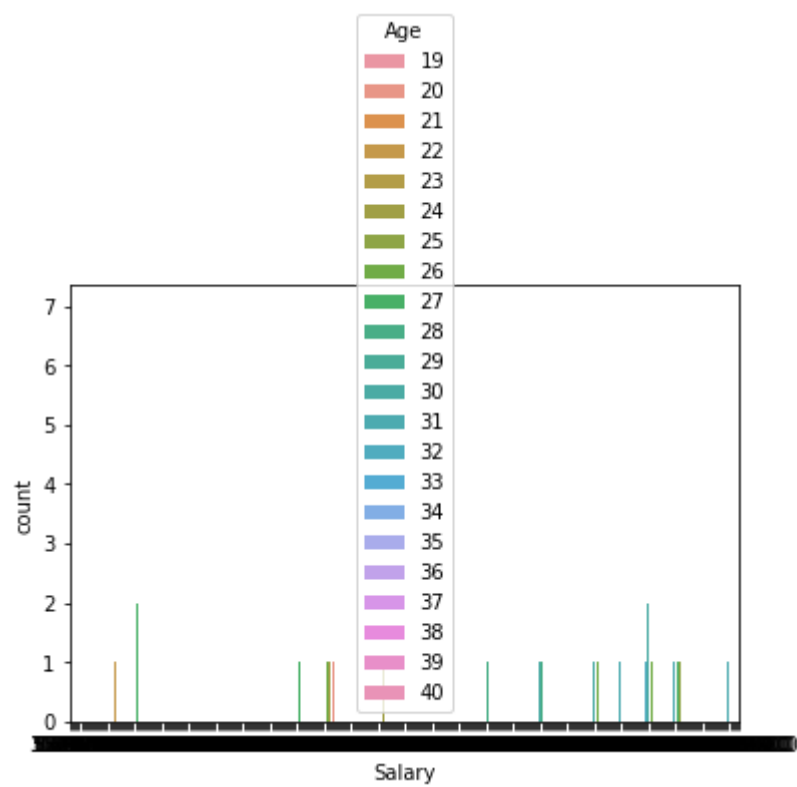
```
sns.countplot(df['Salary'],hue=df['Age'])  
pd.crosstab(df['Age'],df['Salary']).apply(lambda r:round((r/r.sum())*100,1),axis=0)
```

Out[38]:

Salary	30888.0	55722.0	83397.0	111196.0	111444.0	134215.0	167406.0	169883.0	189455.0
Age									
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	100.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0
24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25	0.0	50.0	0.0	100.0	50.0	0.0	0.0	0.0	0.0
26	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
27	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
28	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29	0.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0
30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
34	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
36	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
37	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
39	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

22 rows × 310 columns



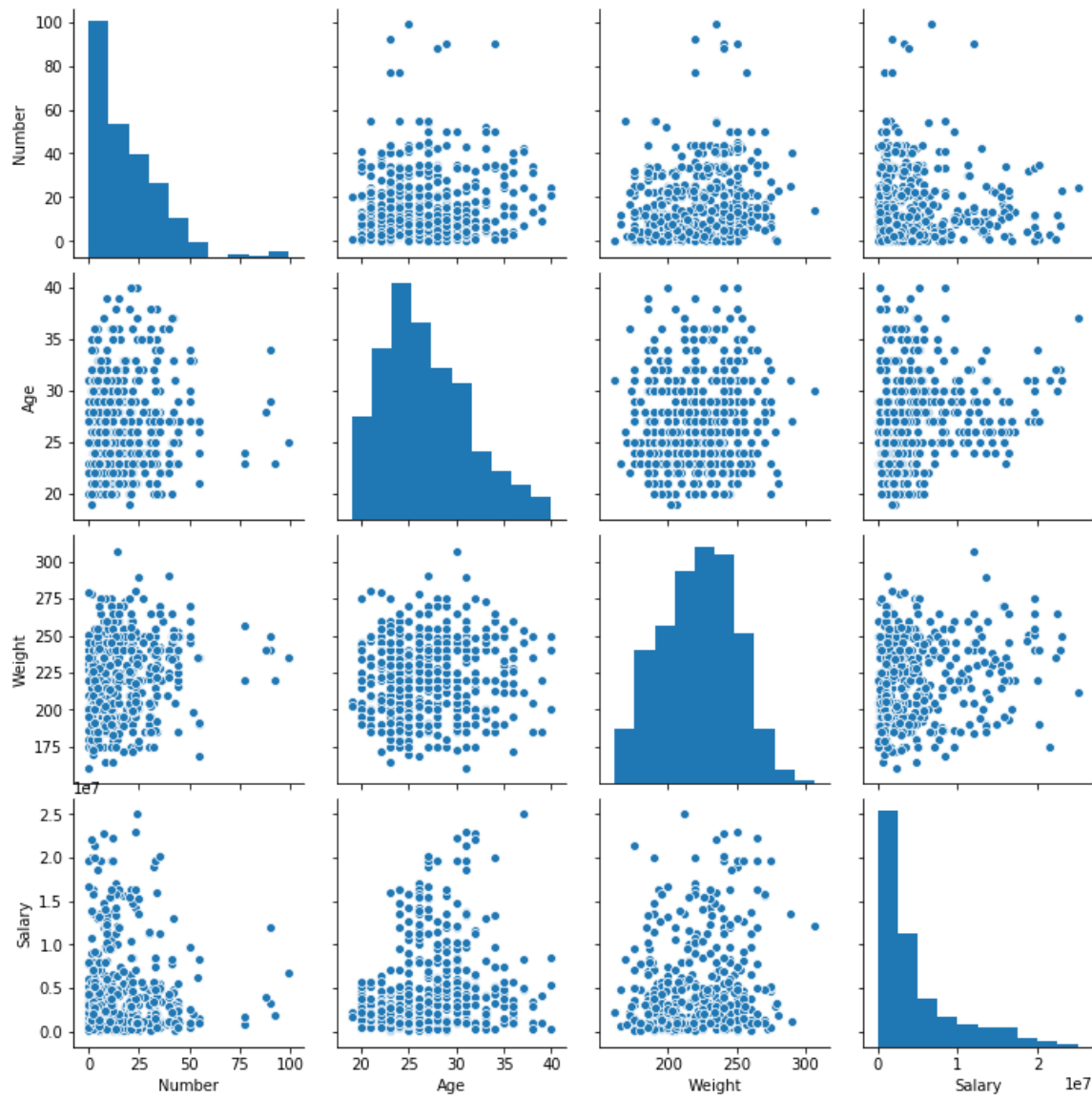


In [39]:

```
sns.pairplot(df)
```

Out[39]:

<seaborn.axisgrid.PairGrid at 0x23c2be3bf48>



In [40]:

```
dfcorr=df.corr()  
dfcorr
```

Out[40]:

	Number	Age	Weight	Salary
Number	1.000000	0.028724	0.206921	-0.111436
Age	0.028724	1.000000	0.087183	0.210627
Weight	0.206921	0.087183	1.000000	0.135552
Salary	-0.111436	0.210627	0.135552	1.000000

In [41]:

```
sns.heatmap(df.corr())
```

Out[41]:

<matplotlib.axes._subplots.AxesSubplot at 0x23c2be7b088>



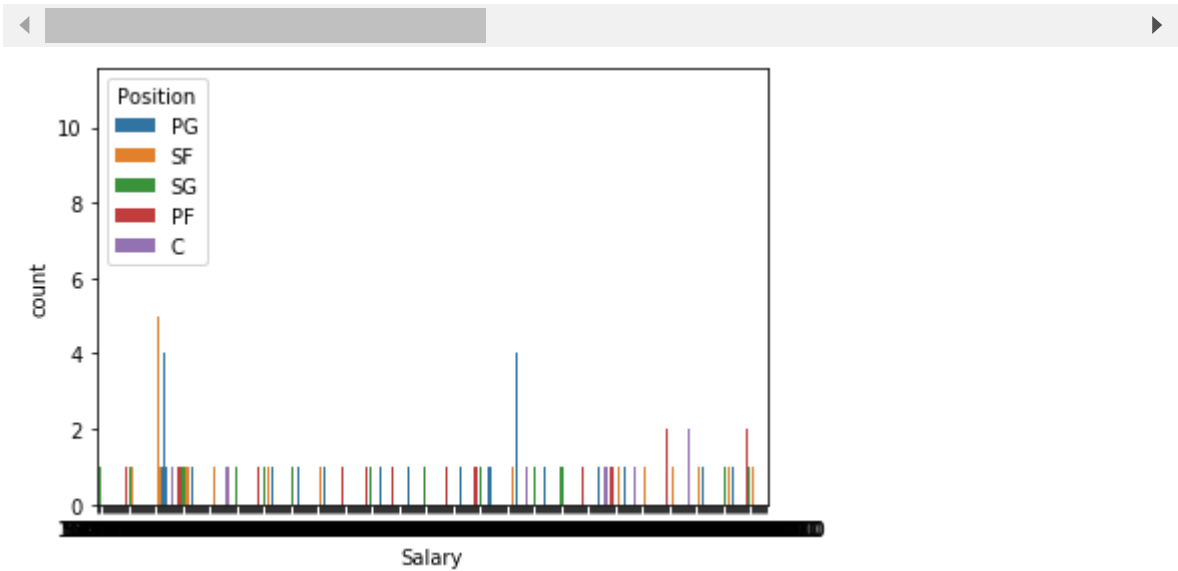
In [43]:

```
sns.countplot(df['Salary'],hue=df['Position'])
pd.crosstab(df['Position'],df['Salary']).apply(lambda r:round((r/r.sum())*100,1),axis=1)
```

Out[43]:

Salary	30888.0	55722.0	83397.0	111196.0	111444.0	134215.0	167406.0	169883.0	189451.0
Position									
C	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0
PF	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
PG	0.0	1.1	0.0	0.0	1.1	0.0	0.0	0.0	0.0
SF	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SG	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0

5 rows × 310 columns



In [44]:

```
df.sample(5)
```

Out[44]:

	Name	Team	Number	Position	Age	DOJ	Weight	Salary
431	Ed Davis	Portland Trail Blazers	17	C	27	06-Oct	240	6980802.0
40	Kristaps Porzingis	New York Knicks	6	PF	20	07-Mar	240	4131720.0
200	George Hill	Indiana Pacers	3	PG	30	06-Mar	188	8000000.0
334	Courtney Lee	Charlotte Hornets	1	SG	30	06-May	200	5675000.0
312	Al Horford	Atlanta Hawks	15	C	30	06-Oct	245	12000000.0

In [92]:

```
x=df.drop('Salary',1)
df['Age']=df['Age'].astype('category')
y=df['Age']
```

In [93]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.20,random_state=58)
```

In [94]:

```
x_train
```

Out[94]:

	Name	Team	Number	Position	Age	DOJ	Weight
176	Timofey Mozgov	Cleveland Cavaliers	20	C	29	07-Jan	275
75	Delon Wright	Toronto Raptors	55	PG	24	06-May	190
254	Donatas Motiejunas	Houston Rockets	20	PF	25	7-0	222
192	Stanley Johnson	Detroit Pistons	3	SF	20	06-Jul	245
36	Langston Galloway	New York Knicks	2	SG	24	06-Feb	200
...
398	Nemanja Bjelica	Minnesota Timberwolves	88	PF	28	06-Oct	240
25	Willie Reed	Brooklyn Nets	33	PF	26	06-Oct	220
279	Norris Cole	New Orleans Pelicans	30	PG	27	06-Feb	175
32	Thanasis Antetokounmpo	New York Knicks	43	SF	23	06-Jul	205
291	Orlando Johnson	New Orleans Pelicans	0	SG	27	06-May	220

365 rows × 7 columns

In [95]:

y_train

Out[95]:

176 29
75 24
254 25
192 20
36 24

..
398 28
25 26
279 27
32 23
291 27

Name: Age, Length: 365, dtype: category
Categories (22, int64): [19, 20, 21, 22, ..., 37, 38, 39, 40]

In [96]:

x_test

Out[96]:

	Name	Team	Number	Position	Age	DOJ	Weight
319	Mike Scott	Atlanta Hawks	32	PF	27	06-Aug	237
341	Goran Dragic	Miami Heat	7	PG	30	06-Mar	190
363	Victor Oladipo	Orlando Magic	5	SG	24	06-Apr	210
356	Aaron Gordon	Orlando Magic	0	PF	20	06-Sep	220
284	James Ennis	New Orleans Pelicans	4	SF	25	06-Jul	210
...
213	Michael Carter-Williams	Milwaukee Bucks	5	PG	24	06-Jun	190
77	Harrison Barnes	Golden State Warriors	40	SF	24	06-Aug	225
69	Patrick Patterson	Toronto Raptors	54	PF	27	06-Sep	235
348	Amar'e Stoudemire	Miami Heat	5	PF	33	06-Oct	245
198	Monta Ellis	Indiana Pacers	11	SG	30	06-Mar	185

92 rows × 7 columns

In [97]:

```
y_test
```

Out[97]:

```
319    27
341    30
363    24
356    20
284    25
```

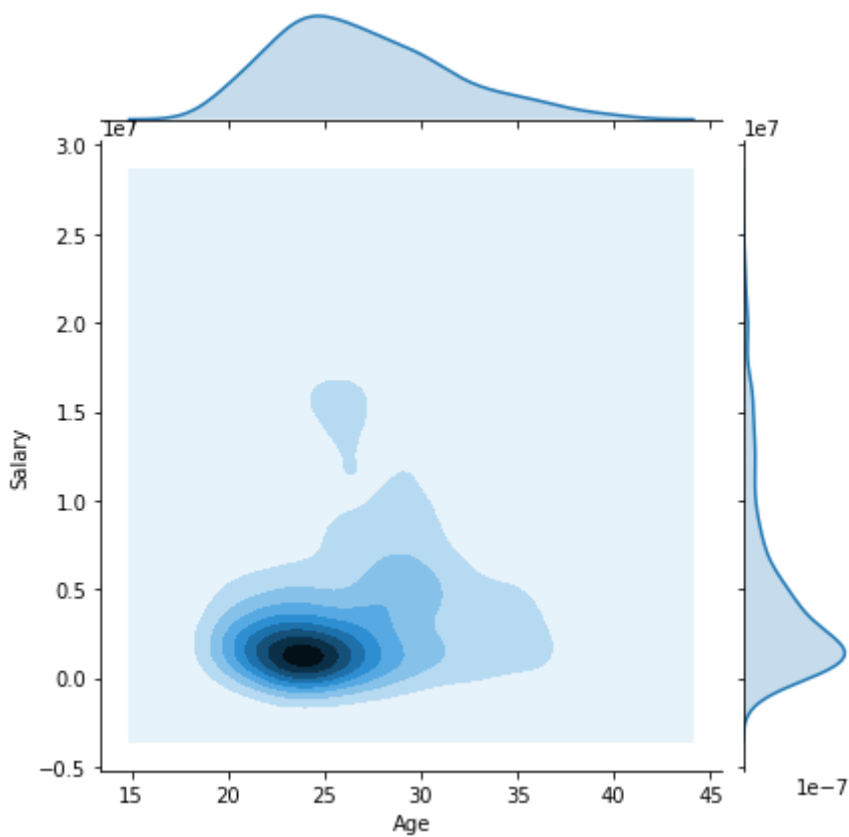
```
..
213    24
77     24
69     27
348    33
198    30
```

Name: Age, Length: 92, dtype: category

Categories (22, int64): [19, 20, 21, 22, ..., 37, 38, 39, 40]

In [107]:

```
sns.jointplot(x='Age',y='Salary',kind="kde",data=df)
plt.show()
```



In []: