```python
In [2]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
```

```python
In [4]: df=pd.read_csv(r"C:\Users\System21\Desktop\Social_Network_Ads.csv")
```

```python
In [6]: df
```

Out[6]:

	Age	EstimatedSalary	Purchased
0	19	19000	0
1	35	20000	0
2	26	43000	0
3	27	57000	0
4	19	76000	0
...	...	...	...
395	46	41000	1
396	51	23000	1
397	50	20000	1
398	36	33000	0
399	49	36000	1

400 rows × 3 columns

```python
In [8]: df.shape
```

Out[8]: (400, 3)

```python
In [10]: df.head()
```

Out[10]:

	Age	EstimatedSalary	Purchased
0	19	19000	0
1	35	20000	0
2	26	43000	0
3	27	57000	0
4	19	76000	0

In [13]: `df.Purchased.value_counts()`

Out[13]:
```
Purchased
0    257
1    143
Name: count, dtype: int64
```

In [17]: `df.dtypes`

Out[17]:
```
Age              int64
EstimatedSalary  int64
Purchased        int64
dtype: object
```

In [19]: `df.isnull().sum()`

Out[19]:
```
Age              0
EstimatedSalary  0
Purchased        0
dtype: int64
```
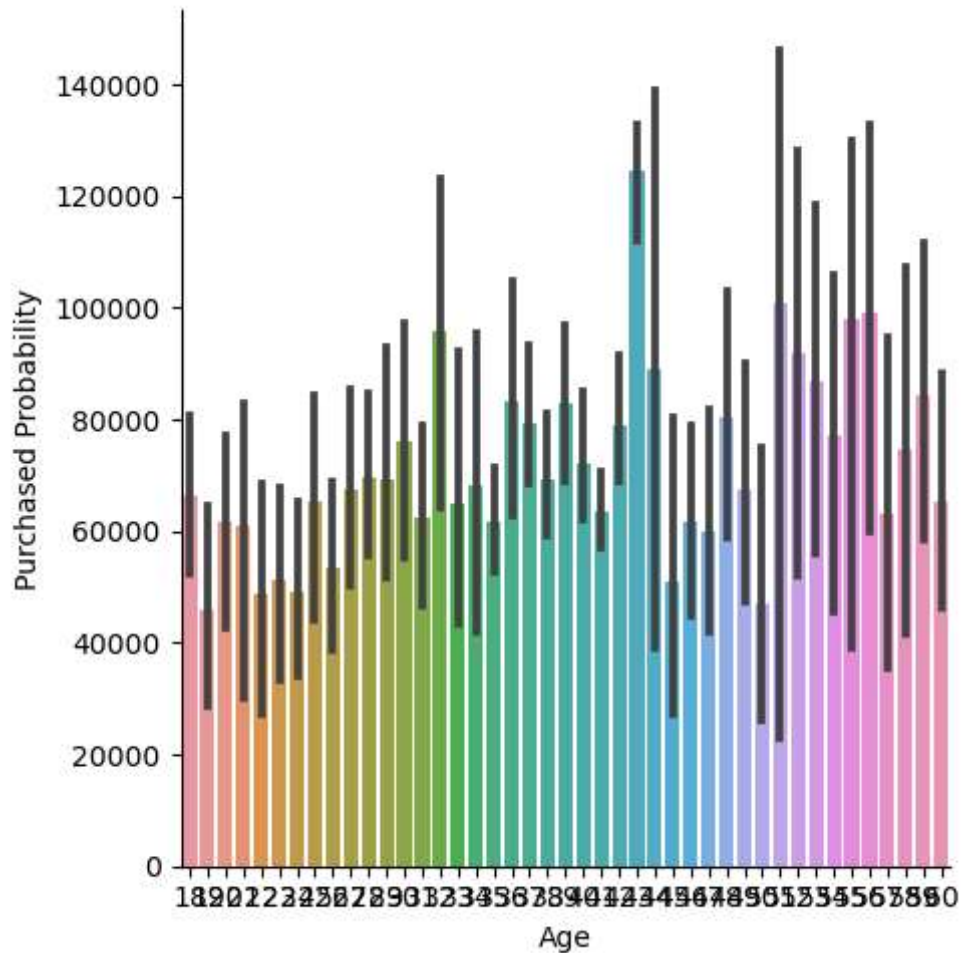
In [21]: `df.describe()`

Out[21]:

	Age	EstimatedSalary	Purchased
count	400.000000	400.000000	400.000000
mean	37.655000	69742.500000	0.357500
std	10.482877	34096.960282	0.479864
min	18.000000	15000.000000	0.000000
25%	29.750000	43000.000000	0.000000
50%	37.000000	70000.000000	0.000000
75%	46.000000	88000.000000	1.000000
max	60.000000	150000.000000	1.000000

In [30]:
```python
g = sns.catplot(x = "Age",y = "EstimatedSalary",data = df,kind = "bar",height =
g.set_ylabels("Purchased Probability")
plt.show
```

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarni
ng: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)

Out[30]: <function matplotlib.pyplot.show(close=None, block=None)>

In [32]:
```python
M2 = pd.crosstab(df.Age, df.EstimatedSalary, normalize='index')
print(M2)
M2.plot.bar(figsize=(6,4),stacked=True)
plt.legend(title='Age vs EstimatedSalary', loc='upper right')
plt.show()
```

EstimatedSalary	15000	16000	17000	18000	19000	20000
Age						
18	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.0000	0.000000	0.000000	0.142857	0.000000
20	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.2500	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.0000	0.000000	0.200000	0.000000	0.000000
23	0.000000	0.0000	0.000000	0.000000	0.000000	0.166667
24	0.000000	0.0000	0.000000	0.000000	0.111111	0.000000
25	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000
26	0.125000	0.0625	0.062500	0.000000	0.000000	0.000000
27	0.000000	0.0000	0.076923	0.000000	0.000000	0.076923
28	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000
29	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000
30	0.090909	0.0000	0.090909	0.000000	0.000000	0.000000
31	0.090909	0.0000	0.000000	0.090909	0.000000	0.000000
32	0.000000	0.0000	0.000000	0.222222	0.000000	0.000000
33	0.000000	0.0000	0.000000	0.000000	0.000000	0.000000

In [35]:

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[35], line 2
      1 from sklearn.model_selection import train_test_split
----> 2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size =
0.25, random_state = 0)

NameError: name 'X' is not defined
```