```
In [1]: import numpy as bp
        import os
        import pandas as pd
        import matplotlib.pyplot as plt
        %matplotlib inline
        from matplotlib import style
        import seaborn as sns
```

```
In [174]: #import csv file
          df = pd.read_csv('car purchasing.csv')
```

Data cleaning process

```
In [175]: #check the data type
          df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Customer name       500 non-null    object
 1   Customer e-mail     500 non-null    object
 2   Country             500 non-null    object
 3   Gender              500 non-null    object
 4   Age group           500 non-null    object
 5   Age                 500 non-null    int64
 6   Annual Salary       500 non-null    int64
 7   Credit card debit   500 non-null    float64
 8   Net worth           500 non-null    int64
 9   Car Purchase Amount 500 non-null    int64
dtypes: float64(1), int64(4), object(5)
memory usage: 39.2+ KB
```

```
In [176]: #check the null values
          pd.isnull(df).sum()
```

```
Out[176]: Customer name          0
          Customer e-mail        0
          Country                0
          Gender                 0
          Age group              0
          Age                    0
          Annual Salary          0
          Credit card debit      0
          Net worth              0
          Car Purchase Amount    0
          dtype: int64
```

```
In [14]: df.shape

Out[14]: (500, 9)
```

```
In [15]: #change from float to int
         df['Credit card debit'] = df['Credit card debit'].astype('int')
```

```
In [177…: df.columns

Out[177…: Index(['Customer name', 'Customer e-mail', 'Country', 'Gender', 'Age group',
                 'Age', 'Annual Salary', 'Credit card debit', 'Net worth',
                 'Car Purchase Amount'],
                dtype='object')
```

```
In [178…: #Drop duplicate values
          df.drop_duplicates(subset = 'Customer e-mail', keep = "first").head(2)
```

Out[178…:

	Customer name	Customer e-mail	Country	Gender	Age group	Age	Annual Salary	Credit card debit	Net worth	Car Purchase Amount
0	Martina Avila	cubilia.Curae.Phasellus@quisaccumsanconvallis.edu	Bulgaria	F	Adult	42	62812	11609.380910	238961	35322
1	Harlan Barnes	eu.dolor@diam.co.uk	Belize	F	Adult	41	66647	9572.957136	530974	45116

Exploratory Data Analysis

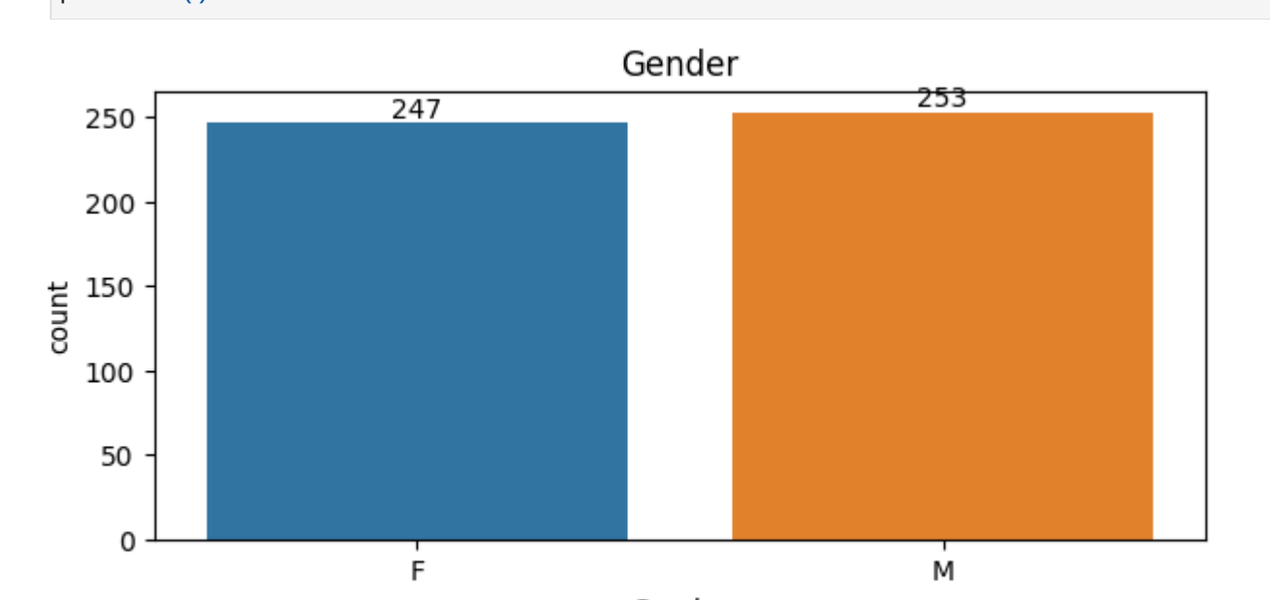
```
In [26]: df.columns

Out[26]: Index(['Customer name', 'Customer e-mail', 'Country', 'Gender', 'Age',
                'Annual Salary', 'Credit card debit', 'Net worth',
                'Car Purchase Amount'],
               dtype='object')
```

```
In [29]: import warnings
         warnings.filterwarnings('ignore')
```

```
In [41]: plt.figure(figsize = (7,3))
         ax = sns.countplot(x = 'Gender', data = df)

         for bars in ax. containers:
             ax.bar_label(bars)
         plt.title('Gender')
         plt.show()
```



The count of the female is around 247 and male's count is 253.

```
In [49]: print(df['Car Purchase Amount'].max())
         df['Customer name'] [df['Car Purchase Amount'].idxmax()]

80000
Out[49]: 'Cameron, Kimberley P.'
```

```
In [50]: print(df['Car Purchase Amount'].max())
         df['Age'] [df['Car Purchase Amount'].idxmax()]

80000
Out[50]: 55
```

```
In [51]: print(df['Car Purchase Amount'].max())
         df['Net worth'] [df['Car Purchase Amount'].idxmax()]

80000
Out[51]: 1000000
```

```
In [54]: print(df['Car Purchase Amount'].max())
         df['Annual Salary'] [df['Car Purchase Amount'].idxmax()]

80000
Out[54]: 83334
```

The highest amount for the car was paid by Cameron, Kimberley P. a Male around the age of 55 from Namibia. His net worth is around 1000000 and his annual salary is 83334.

```
In [56]: df[(df.Age <= 19)]
```

Out[56]:

Customer name	Customer e-mail	Country	Gender	Age	Annual Salary	Credit card debit	Net worth	Car Purchase Amount

```
In [62]: df[(df.Age <= 50)]
```

Out[62]:

	Customer name	Customer e-mail	Country	Gender	Age	Annual Salary	Credit card debit	Net worth	Car Purchase Amount
0	Martina Avila	cubilia.Curae.Phasellus@quisaccumsanconvallis.edu	Bulgaria	F	42	62812	11609	238961	35322
1	Harlan Barnes	eu.dolor@diam.co.uk	Belize	F	41	66647	9572	530974	45116
2	Naomi Rodriquez	vulputate.mauris.sagittis@ametconsectetueradip...	Algeria	M	43	53799	11160	638467	42926
6	Griffin Rivera	vehicula@at.co.uk	Syria	M	47	39815	5958	326373	28926
7	Orli Casey	nunc.est.mollis@Suspendissetristiqueneque.co.uk	Czech Republic	M	50	51752	10985	629312	47435
...	...	...	...	...	...	...	...	...	...
491	Merrill	dolor.sit@turpisIn.com	Egypt	M	50	78518	10072	294506	52786
493	Winter	egestas.urna.justo@maurissagittis.edu	Wallis and Futuna	F	43	77665	13308	349589	47761
495	Walter	ligula@Cumsociis.ca	Nepal	F	41	71942	6995	541670	48902
496	Vanna	Cum.sociis.natoque@Sedmolestie.edu	Zimbabwe	M	38	56039	12301	360419	31492
499	Marla	Camaron.marla@hotmail.com	marlal	M	47	61371	9391	462946	45108

344 rows × 9 columns

Total 344 customers purchased the cars under the age of 50 and 75 customers above the age of 50. The maximum age is 70 and minimum is 20.

```
In [68]: df[(df.Age >= 50)]
```

Out[68]:

	Customer name	Customer e-mail	Country	Gender	Age	Annual Salary	Credit card debit	Net worth	Car Purchase Amount
3	Jade Cunningham	malesuada@dignissim.com	Cook Islands	M	58	79370	14426	548599	67423
4	Cedric Leach	felis.ullamcorper.viverra@egetmollislectus.net	Brazil	M	57	59729	5358	560304	55916
5	Carla Hester	mi@Aliquamerat.edu	Liberia	M	57	68500	14179	428485	56612
7	Orli Casey	nunc.est.mollis@Suspendissetristiqueneque.co.uk	Czech Republic	M	50	51752	10985	629312	47435
10	Jerome Rowe	ipsum.cursus@dui.org	Sint Maarten	M	50	73349	8270	612739	59046
...	...	...	...	...	...	...	...	...	...
491	Merrill	dolor.sit@turpisIn.com	Egypt	M	50	78518	10072	294506	52786
492	Nolan	Donec.at@neccursus.co.uk	Latvia	F	55	72425	9831	523681	60118
494	Rigel	egestas.blandit.Nam@semvitaealiquam.com	Sao Tome and Principe	F	52	77346	6736	665099	64189
497	Pearl	penatibus.et@massanonante.com	Philippines	M	54	68889	10611	764531	64148
498	Nell	Quisque.varius@arcuVivamussit.net	Botswana	M	59	49812	14013	337827	45443

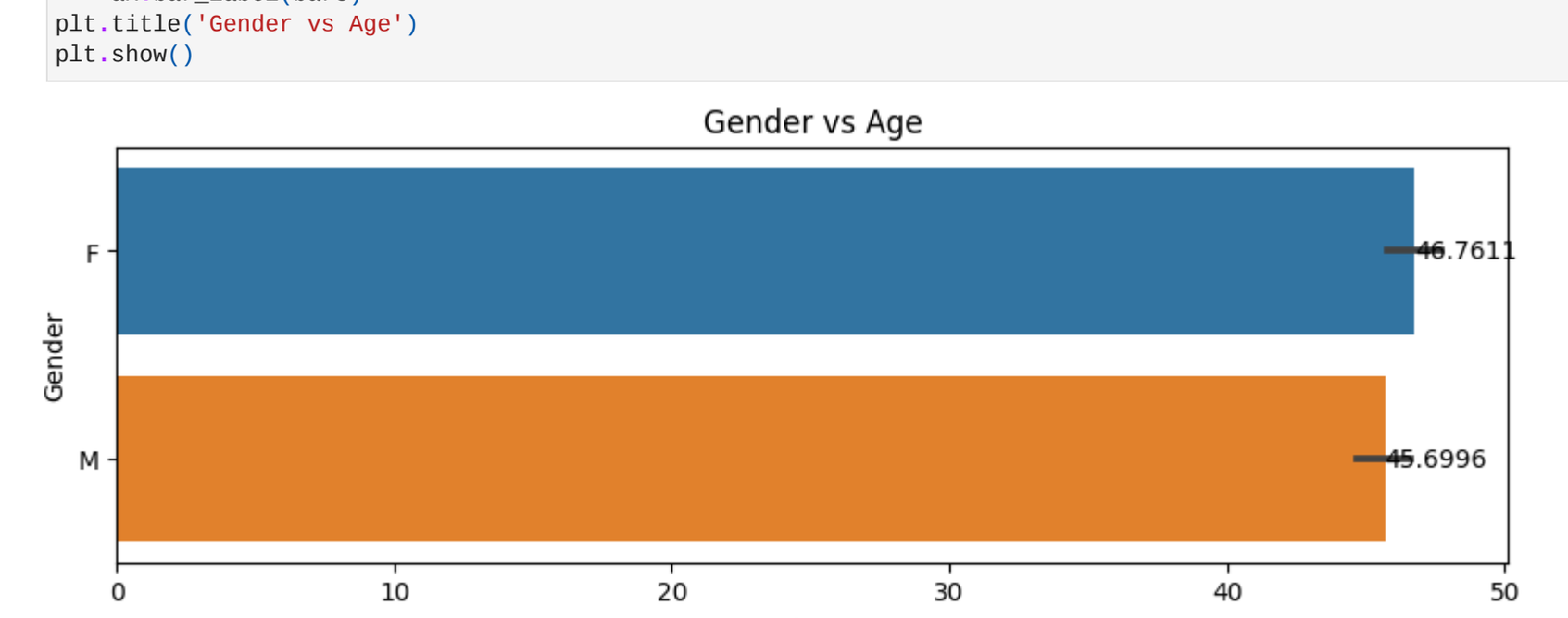
175 rows × 9 columns

```
In [85]: print(df['Car Purchase Amount'].min())
         df['Customer name'] [df['Car Purchase Amount'].idxmin()]

9000
Out[85]: 'Tatyana'
```

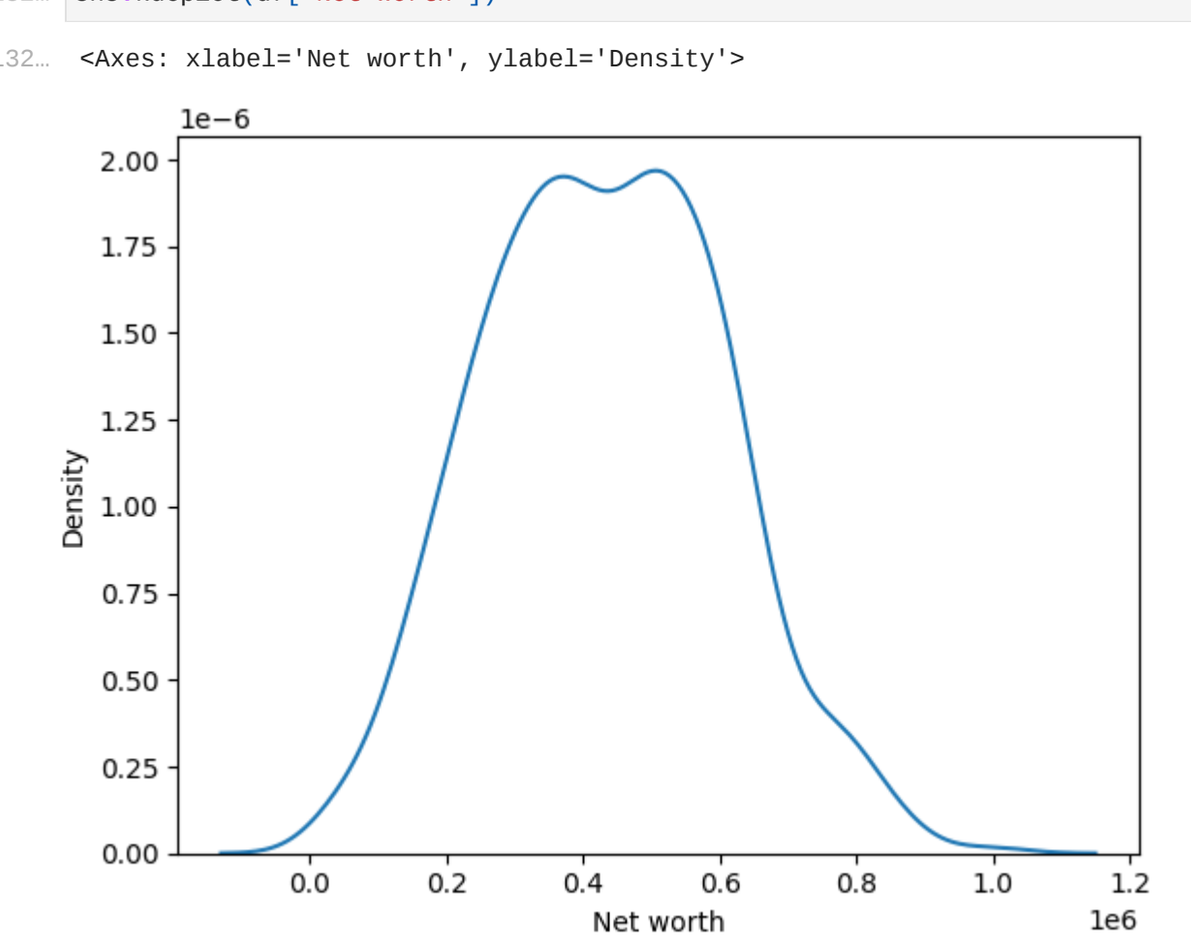
Overall the cheapest price of car is only 9000 purchased by Tatyana whose age is 25 and his annual salary is 45093.

```
In [137…: plt.figure(figsize = (10,3))
          ax = sns.barplot(x = 'Age', y = 'Gender', data = df)
          for bars in ax.containers:
              ax.bar_label(bars)
          plt.title('Gender vs Age')
          plt.show()
```




```
In [132…: sns.kdeplot(df['Net worth'])

Out[132…: <Axes: xlabel='Net worth', ylabel='Density'>
```



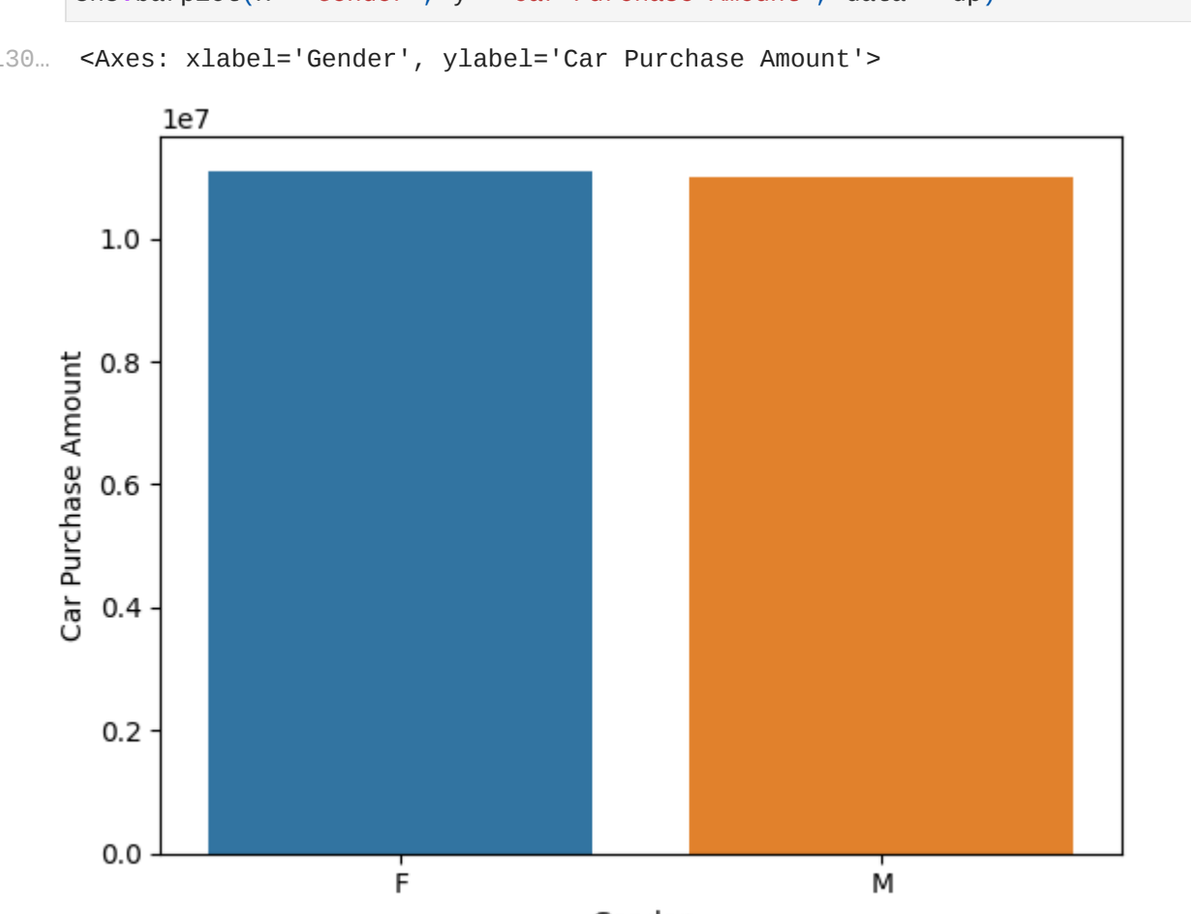

```
In [128…: df.groupby(['Gender'], as_index=False)['Car Purchase Amount'].sum().sort_values(by='Car Purchase Amount', ascending=False)
```

Out[128…:

	Gender	Car Purchase Amount
0	F	11098602
1	M	11006542

```
In [130…: dp=df.groupby(['Gender'], as_index=False)['Car Purchase Amount'].sum().sort_values(by='Car Purchase Amount', ascending=False)
          sns.barplot(x= 'Gender', y= 'Car Purchase Amount', data = dp)

Out[130…: <Axes: xlabel='Gender', ylabel='Car Purchase Amount'>
```



Most of the buyers are females.

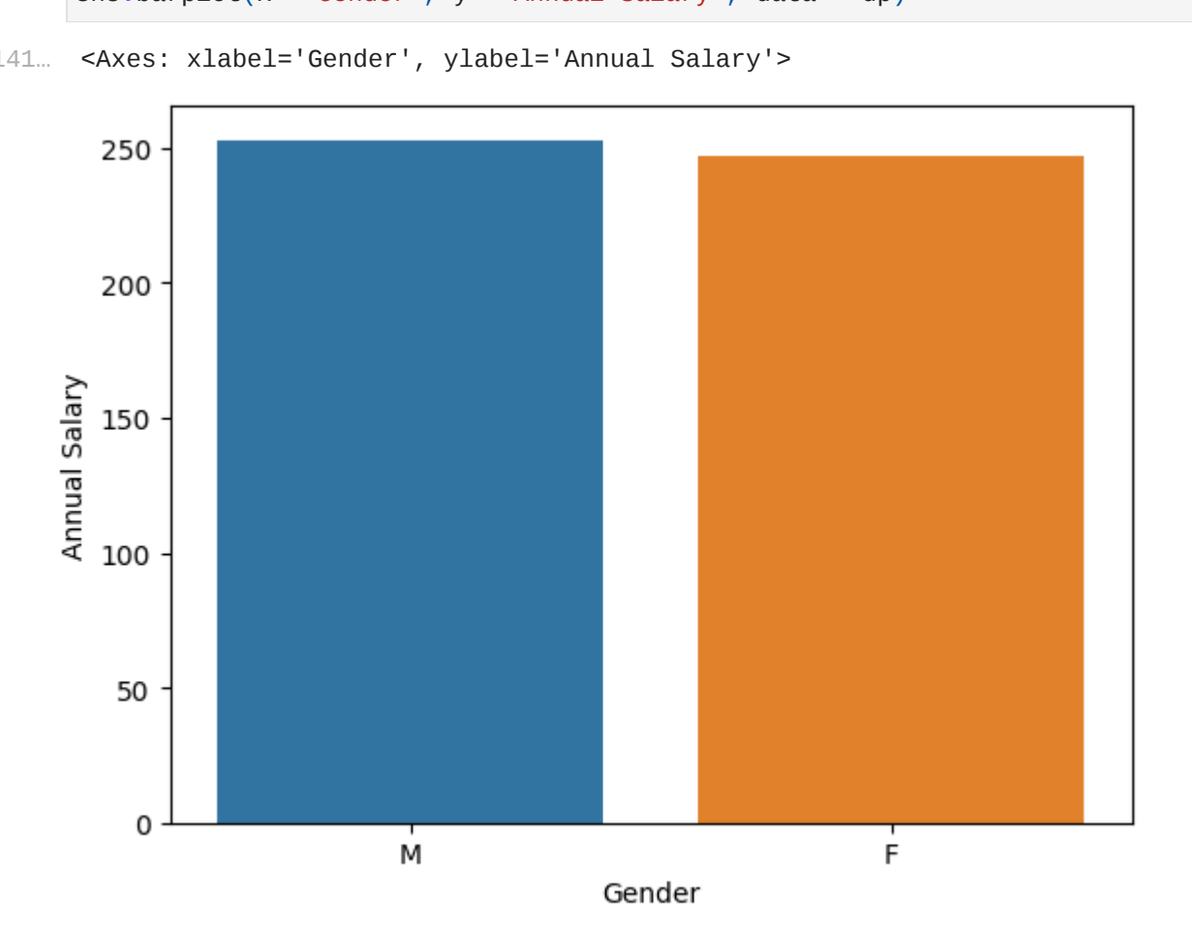
```
In [139…: df[['Age', 'Car Purchase Amount', 'Net worth', 'Customer name']].describe()
```

Out[139…:

	Age	Car Purchase Amount	Net worth
count	500.000000	500.000000	500.000000
mean	46.224000	44210.288000	431475.700000
std	7.990339	10773.182576	173536.758223
min	20.000000	9000.000000	20000.000000
25%	41.000000	37630.500000	299824.000000
50%	46.000000	43998.500000	426750.000000
75%	52.000000	51255.500000	557324.500000
max	70.000000	80000.000000	1000000.000000

```
In [141…: dp=df.groupby(['Gender'], as_index=False)['Annual Salary'].count().sort_values(by='Annual Salary', ascending=False)
          sns.barplot(x= 'Gender', y= 'Annual Salary', data = dp)

Out[141…: <Axes: xlabel='Gender', ylabel='Annual Salary'>
```

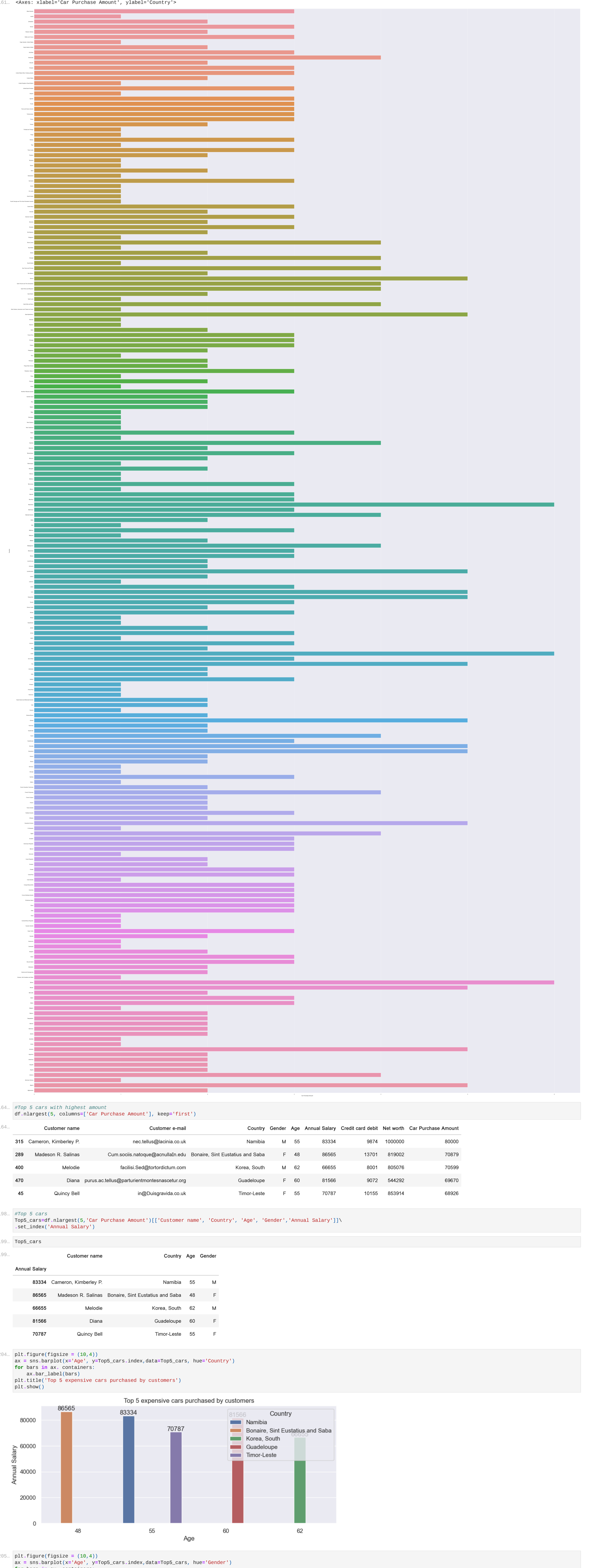


Count of the female's salary is higher than the male's salary

```
In [161…: dp=df.groupby(['Country'], as_index=False)['Car Purchase Amount'].count().sort_values(by='Country', ascending=False)
          sns.barplot(x= 'Car Purchase Amount', y= 'Country', data = dp)

          sns.set(rc={'figure.figsize':(50,200)})
          sns.barplot(data = dp, x = 'Car Purchase Amount', y= 'Country')

Out[161…: <Axes: xlabel='Car Purchase Amount', ylabel='Country'>
```



```
In [164…: #Top 5 cars with highest amount
          df.nlargest(5, columns=['Car Purchase Amount'], keep='first')
```

Out[164…:

	Customer name	Customer e-mail	Country	Gender	Age	Annual Salary	Credit card debit	Net worth	Car Purchase Amount
315	Cameron, Kimberley P.	nec.tellus@lacinia.co.uk	Namibia	M	55	83334	9874	1000000	80000
289	Madeson R. Salinas	Cum.sociis.natoque@acnullaIn.edu	Bonaire, Sint Eustatius and Saba	F	48	86565	13701	819002	70879
400	Melodie	facilisi.Sed@tortordictum.com	Korea, South	M	62	66655	8001	805076	70599
470	Diana	purus.ac.tellus@parturientmontesnascetur.org	Guadeloupe	F	60	81566	9072	544292	69670
45	Quincy Bell	in@Duisgravida.co.uk	Timor-Leste	F	55	70787	10155	853914	68926

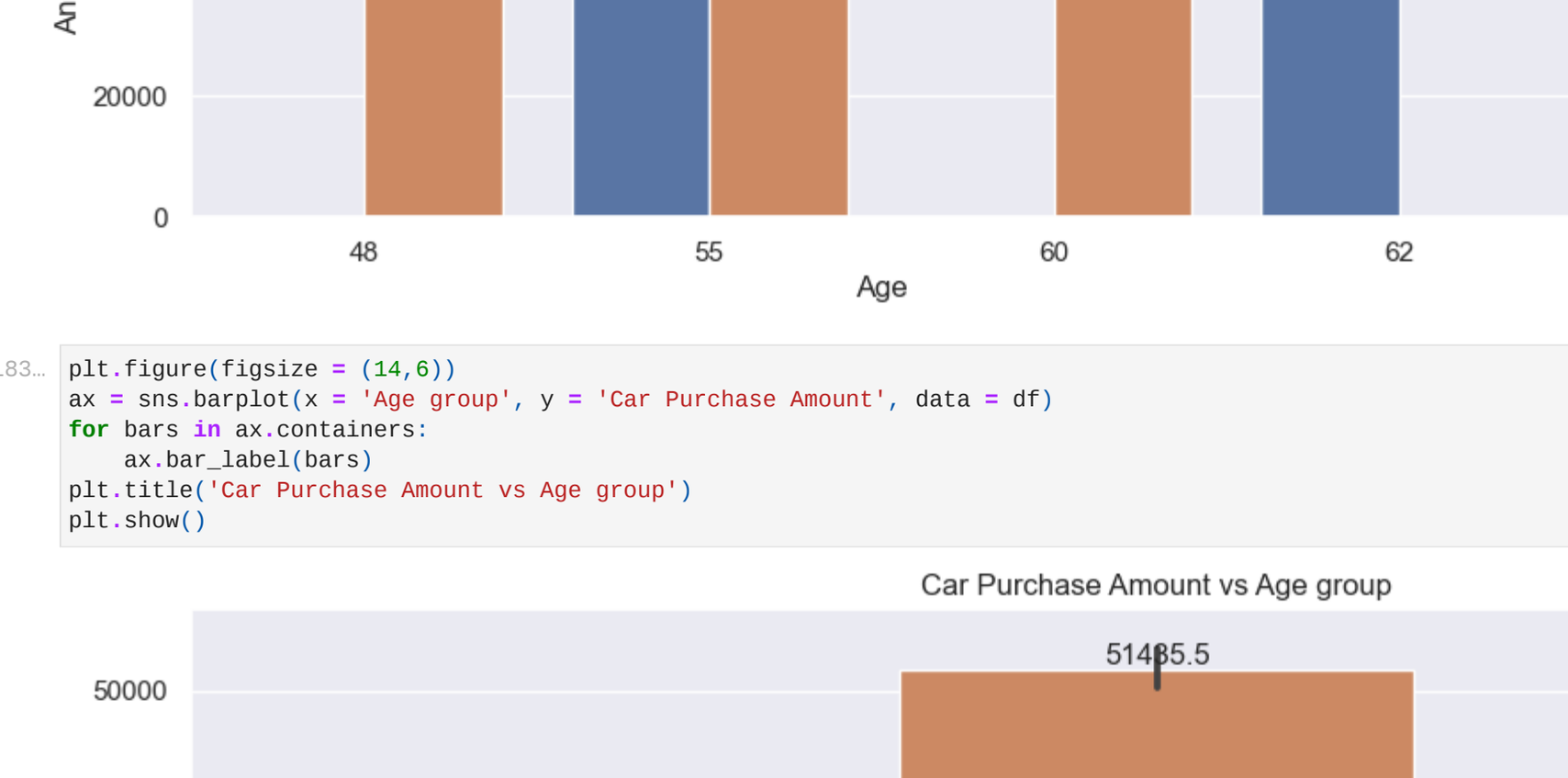
```
In [198…: #Top 5 cars
          Top5_cars=df.nlargest(5,'Car Purchase Amount')[['Customer name', 'Country', 'Age', 'Gender','Annual Salary']]\
          .set_index('Annual Salary')
```

```
In [199…: Top5_cars
```

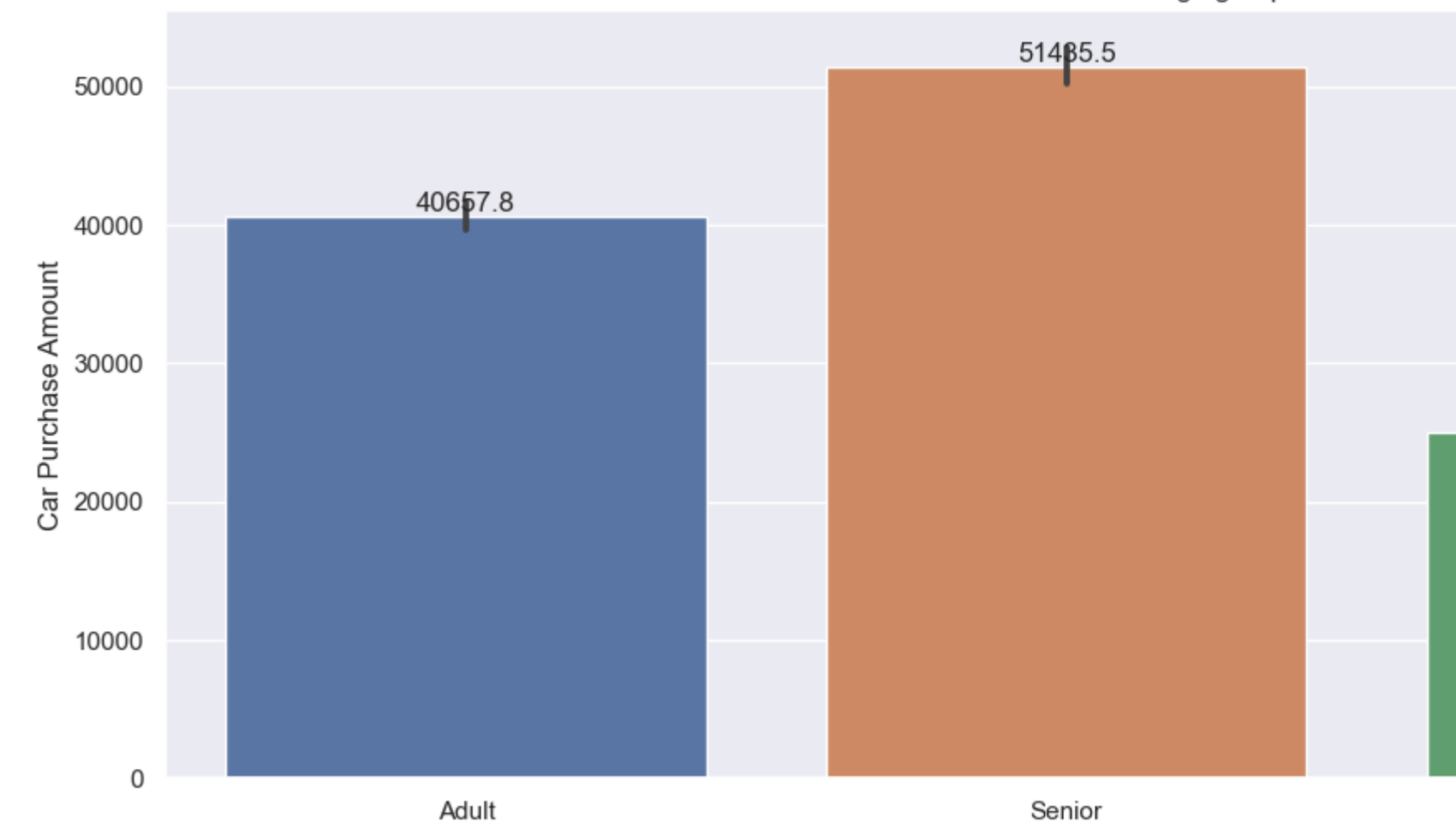
Out[199…:

Annual Salary	Customer name	Country	Age	Gender
83334	Cameron, Kimberley P.	Namibia	55	M
86565	Madeson R. Salinas	Bonaire, Sint Eustatius and Saba	48	F
66655	Melodie	Korea, South	62	M
81566	Diana	Guadeloupe	60	F
70787	Quincy Bell	Timor-Leste	55	F

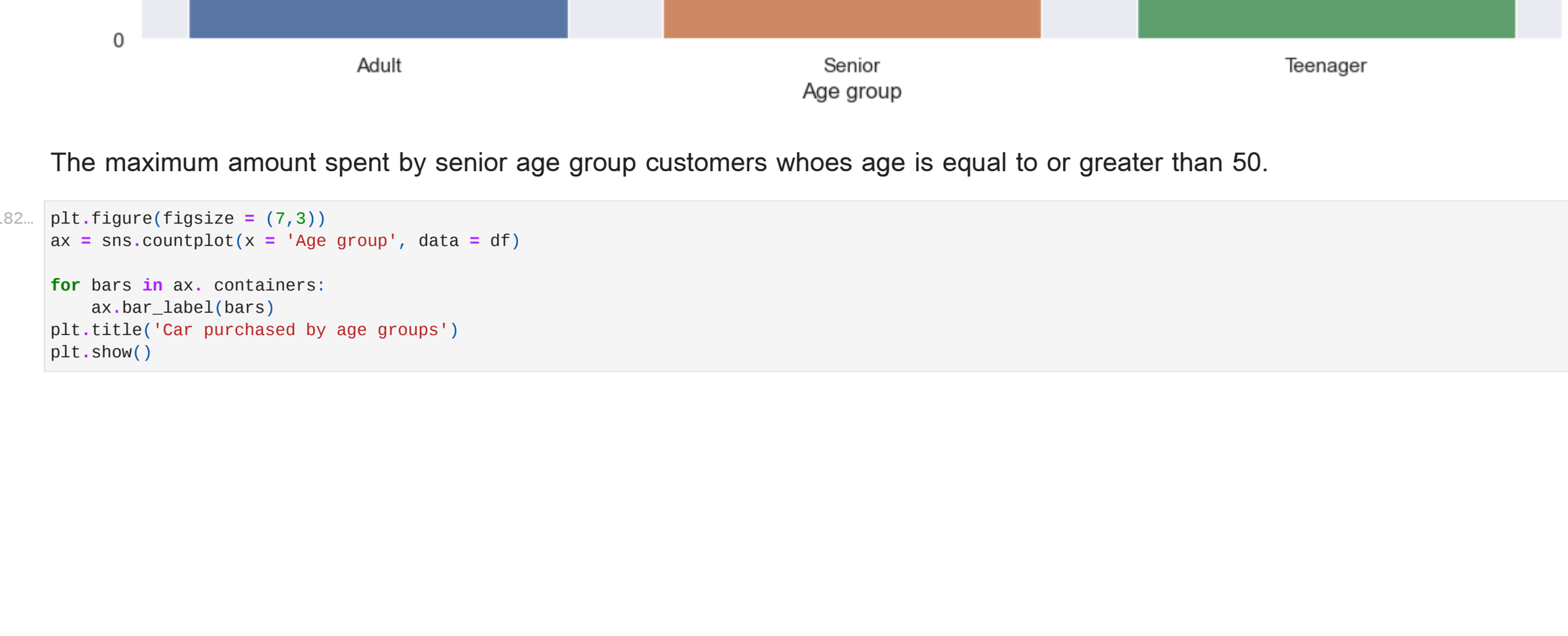
```
In [204…: plt.figure(figsize = (10,4))
          ax = sns.barplot(x='Age', y=Top5_cars.index,data=Top5_cars, hue='Country')
          for bars in ax. containers:
              ax.bar_label(bars)
          plt.title('Top 5 expensive cars purchased by customers')
          plt.show()
```




```
In [205…: plt.figure(figsize = (10,4))
          ax = sns.barplot(x='Age', y=Top5_cars.index,data=Top5_cars, hue='Gender')
          for bars in ax. containers:
              ax.bar_label(bars)
          plt.title('Top 5 expensive cars purchased by customers's details ')
          plt.show()
```



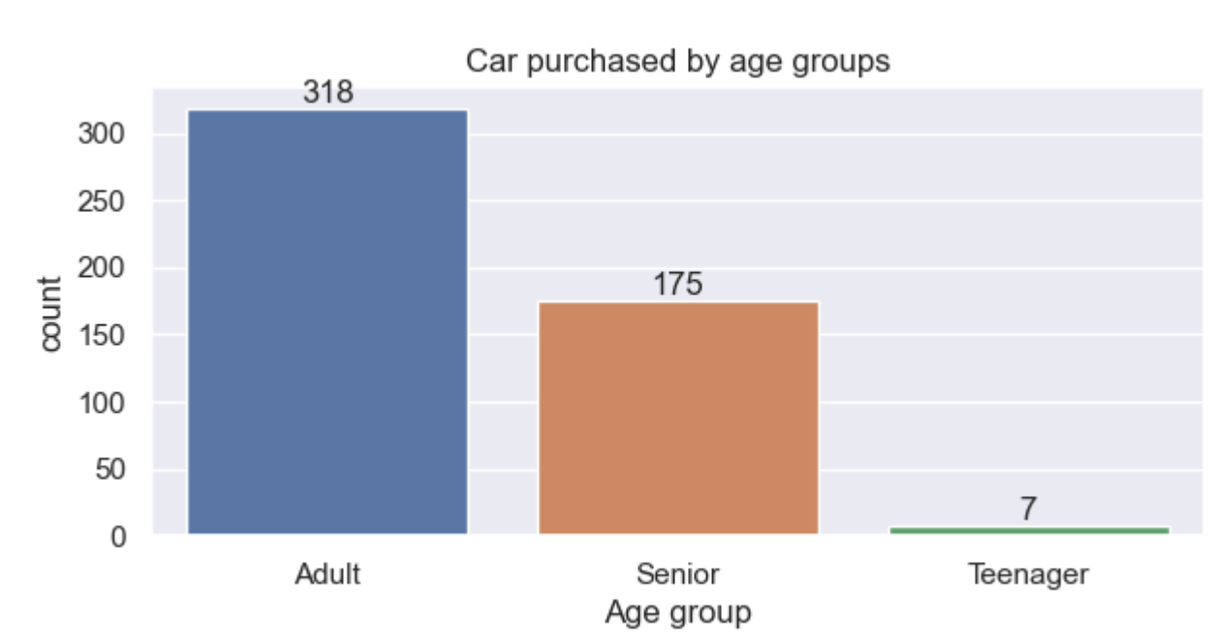

```
In [183…: plt.figure(figsize = (14,6))
          ax = sns.barplot(x = 'Age group', y = 'Car Purchase Amount', data = df)
          for bars in ax.containers:
              ax.bar_label(bars)
          plt.title('Car Purchase Amount vs Age group')
          plt.show()
```



The maximum amount spent by senior age group customers whoes age is equal to or greater than 50.

```
In [182…: plt.figure(figsize = (7,3))
          ax = sns.countplot(x = 'Age group', data = df)

          for bars in ax. containers:
              ax.bar_label(bars)
          plt.title('Car purchased by age groups')
          plt.show()
```



The cars purchased by adults (30 to 49 age) are more than senior (>=50 age) or teenager (<=30 age) .

Insight

The count of the female is around 247 and the male count is 253.

The highest amount for the car was paid by Cameron, Kimberley P. a male around the age of 55 from Namibia. His net worth is around 1000000 and his annual salary is 83334.

A total 344 customers purchased cars under the age of 50 and 75 customers above the age of 50. The maximum age is 70 and the minimum is 20.

Overall the cheapest price of the car is only 9000 purchased by Tatyana whose age is 25 and his annual salary is 45093.

Most of the buyers are females. Females spent around 11098602 and the count of the female's salary is higher than the male's salary.

The maximum amount spent by senior age group customers there's age is equal to or greater than 50.

The cars purchased by adults (30 to 49 age) are more than seniors (>=50 age) or teenagers (<=30 age).

Graphically presented the top 5 expensive cars details.

```
In [ ]:
```