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

```
dataset=pd.read_excel("/content/ANZ synthesised transaction dataset (4).xlsx")
```

```
dataset.head(10)
```

	date	gender	age	merchant_suburb	merchant_state	extraction	amount	
	2018-08-01	F	26	Ashmore	QLD	2018-08-01T01:01:15.000+0000	16.25	a
	2018-08-01	F	26	Sydney	NSW	2018-08-01T01:13:45.000+0000	14.19	132
	2018-08-01	M	38	Sydney	NSW	2018-08-01T01:26:15.000+0000	6.42	feb
	2018-08-01	F	40	Buderim	QLD	2018-08-01T01:38:45.000+0000	40.90	269
	2018-08-01	F	26	Mermaid Beach	QLD	2018-08-01T01:51:15.000+0000	3.25	32
	2018-08-01	M	20	NaN	NaN	2018-08-01T02:00:00.000+0000	163.00	10
	2018-08-01	F	43	Kalkallo	VIC	2018-08-01T02:23:04.000+0000	61.06	b79
	2018-08-01	F	43	Melbourne	VIC	2018-08-01T04:11:25.000+0000	15.61	e1
	2018-08-01	F	27	Yokine	WA	2018-08-01T04:40:00.000+0000	19.25	79
	2018-08-01	M	40	NaN	NaN	2018-08-01T06:00:00.000+0000	21.00	798

```
#We don't need all the columns' data so we will make a new dataset having all the needed c
```

```
#We don't need all the columns' data so we will make a new dataset having all the needed t

finaldata=dataset[['age','first_name','balance','amount','movement']]
finaldata.head(10)
```

	age	first_name	balance	amount	movement
0	26	Diana	35.39	16.25	debit
1	26	Diana	21.20	14.19	debit
2	38	Michael	5.71	6.42	debit
3	40	Rhonda	2117.22	40.90	debit
4	26	Diana	17.95	3.25	debit
5	20	Robert	1705.43	163.00	debit
6	43	Kristin	1248.36	61.06	debit
7	43	Kristin	1232.75	15.61	debit
8	27	Tonya	213.16	19.25	debit
9	40	Michael	466.58	21.00	debit

```
annual_salary=finaldata.amount[(finaldata.movement == "credit")].sum()

total_spending =finaldata.amount[(finaldata.movement == "debit")].sum()



print(annual_salary)
print(total_spending)
```

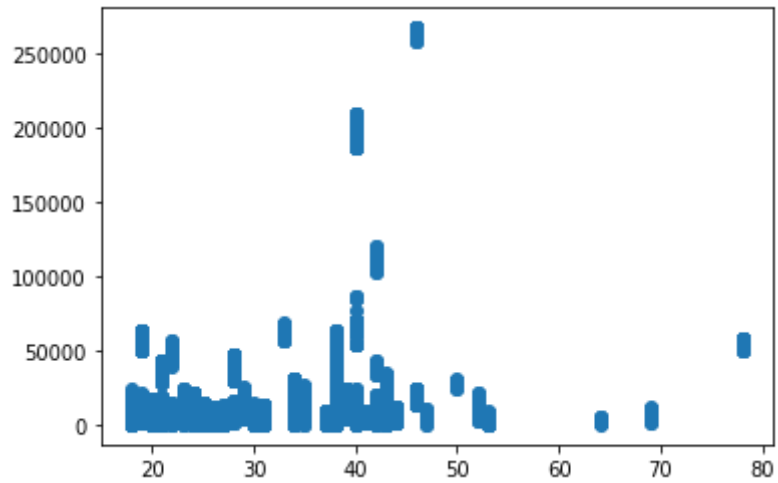
```
     1676576.85
     586707.35
```

```
label = ['Total Salary',
        'Total Spending']
x=[annual_salary,total_spending]
pie = [annual_salary, total_spending]
plt.pie(x,labels=label,autopct="%.1f%%")
```

```
([<matplotlib.patches.Wedge at 0x7f74caa153d0>,
  <matplotlib.patches.Wedge at 0x7f74caa15a90>],
 [Text(-0.7549434955041601, 0.8000376982342522, 'Total Salary'),
  Text(0.7549434205992215, -0.8000377689171599, 'Total Spending')],
 [Text(-0.41178736118408726, 0.43638419903686476, '74.1%'),
  Text(0.411787320326848, -0.4363842375911781, '25.9%')])
```

```
y=finaldata.balance
x=finaldata.age
plt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x7f74caa71510>
```

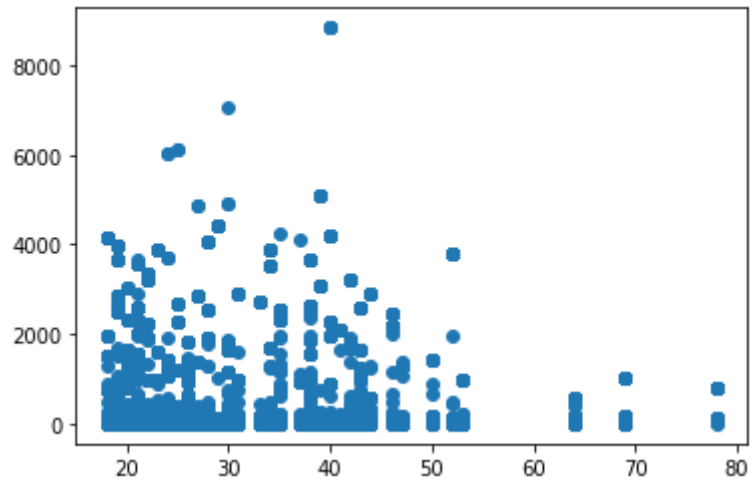


1) Age group of 40-50 has comparatively higher balance.

2) Most of the people are in the age group 20-30 with a balance of approx 50K

```
y=finaldata.amount
x=finaldata.age
plt.scatter(x,y)
```

```
<matplotlib.collections.PathCollection at 0x7f74cab4e290>
```



1) Old age group people make less transactions(in amount) as compared to younger age group

MODEL

```
x=finaldata[['age','amount']]
y=finaldata['balance']
```

```
import sklearn
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

```
from sklearn.linear_model import LinearRegression
```

```
model=LinearRegression()
model.fit(x_train,y_train)
pred=model.predict(x_test)
pred
```

```
array([ 7535.14447216, 12106.89326746,  6860.20531899, ...,
        6913.02616187,  8270.09041555, 19429.71212131])
```

✓ 0s completed at 6:40 PM