PREDICTIVE ANALYSIS BY AKSHAYA

In [ ]:

```
# Importing Libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
```

In [3]:

```
# Importing data
data = pd.read_csv('D:ANZ synthesised transaction dataset.csv')
```

In [11]:

```
data.head()
```

Out[11]:

	status	card_present_flag	bpay_biller_code	account	currency	long_lat	txn_description	merchant_id	merchant_code	first
0	authorized	1.0	NaN	ACC-1598451071	AUD	153.41 -27.95	POS	81c48296-73be-44a7-befa-d053f48ce7cd	NaN	
1	authorized	0.0	NaN	ACC-1598451071	AUD	153.41 -27.95	SALES-POS	830a451c-316e-4a6a-bf25-e37caedca49e	NaN	
2	authorized	1.0	NaN	ACC-1222300524	AUD	151.23 -33.94	POS	835c231d-8cdf-4e96-859d-e9d571760cf0	NaN	
3	authorized	1.0	NaN	ACC-1037050564	AUD	153.10 -27.66	SALES-POS	48514682-c78a-4a88-b0da-2d6302e64673	NaN	
4	authorized	1.0	NaN	ACC-1598451071	AUD	153.41 -27.95	SALES-POS	b4e02c10-0852-4273-b8fd-7b3395e32eb0	NaN	

5 rows × 23 columns

In [12]:

```
data= data[['age','amount','balance']]
data.head()
```

Out[12]:

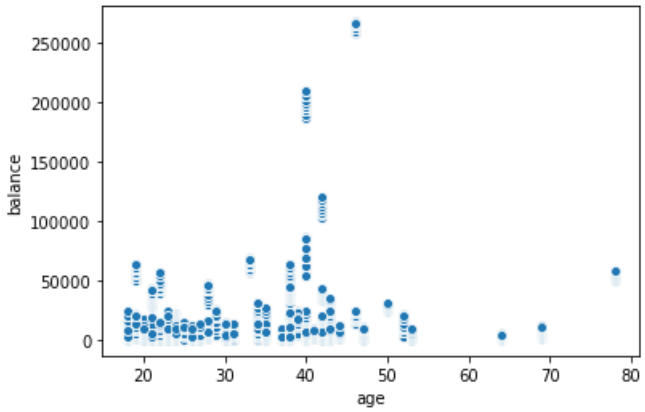
	age	amount	balance
0	26	16.25	35.39
1	26	14.19	21.20
2	38	6.42	5.71
3	40	40.90	2117.22
4	26	3.25	17.95

In [14]:

```
sns.scatterplot(x= data['age'], y= data['balance'], data= data)
```

Out[14]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x14509a0a3d0>
```

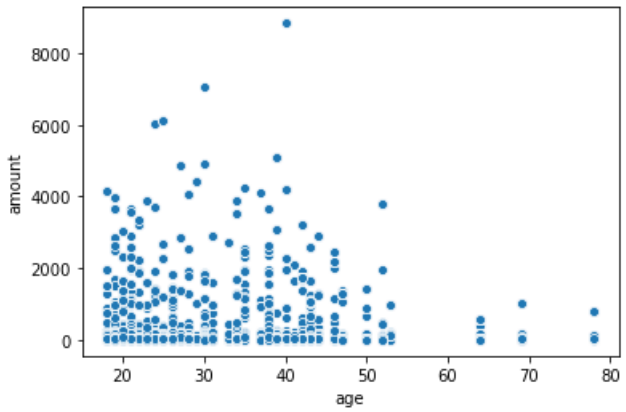


In [15]:

```
sns.scatterplot(x= data['age'],y=data['amount'],data=data)
```

Out[15]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x14509b49df0>
```



In [4]:

```
# --- Balance vs Age ---
y = data.iloc[:,10:11].values # Balance
X_age = data.iloc[:, 13:14].values # Age
```

In [5]:

```
# Spliting the dataset
from sklearn.model_selection import train_test_split
X_age_train, X_age_test, y_train, y_test = train_test_split(X_age, y, test_size =0.2, random_state
=0)
```

In [6]:

```
# Fitting Decision Tree Regression to the dataset
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor()
regressor.fit(X_age_train,y_train)
```

Out[6]:

```
DecisionTreeRegressor()
```

In [7]:

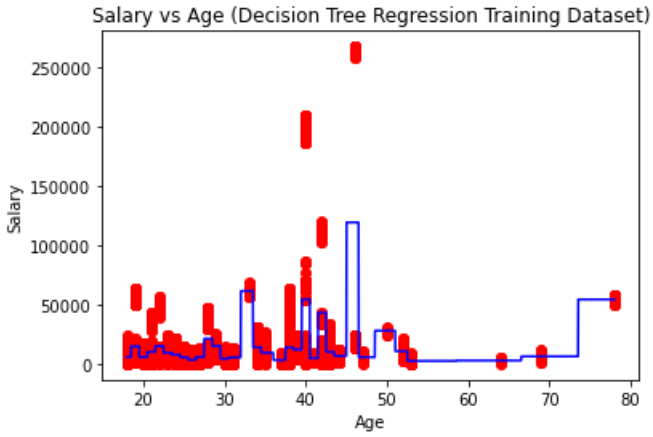
```
# predicting a new result
y_pred = regressor.predict(X_age_test)
```

In [8]:

```
# Calculating MSE
from sklearn.metrics import mean_squared_error
ms = mean_squared_error(y_test, y_pred)
```

In [9]:

```
# Visualizing Decision Tree Regression in High resolution
X_grid = np.arange(min(X_age_train), max(X_age_train), 0.01)
X_grid = X_grid.reshape(len(X_grid), 1)
plt.scatter(X_age_train, y_train, color = 'red')
plt.plot(X_grid, regressor.predict(X_grid), color = 'blue')
plt.title('Salary vs Age (Decision Tree Regression Training Dataset)')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```



In [10]:

```
# Visualizing Training Dataset
X_grid1 = np.arange(min(X_age_test), max(X_age_test), 0.001)
X_grid1 = X_grid1.reshape(len(X_grid1), 1)
plt.scatter(X_age_test, y_test, color = 'red')
plt.plot(X_grid1, regressor.predict(X_grid1), color = 'blue')
plt.title('Salary vs Age (Decision Tree Regression Testing Dataset)')
plt.xlabel('Age')
plt.ylabel('Salary')
plt.show()
```