2306aml133-kritika-assignment10

August 27, 2023

[10]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

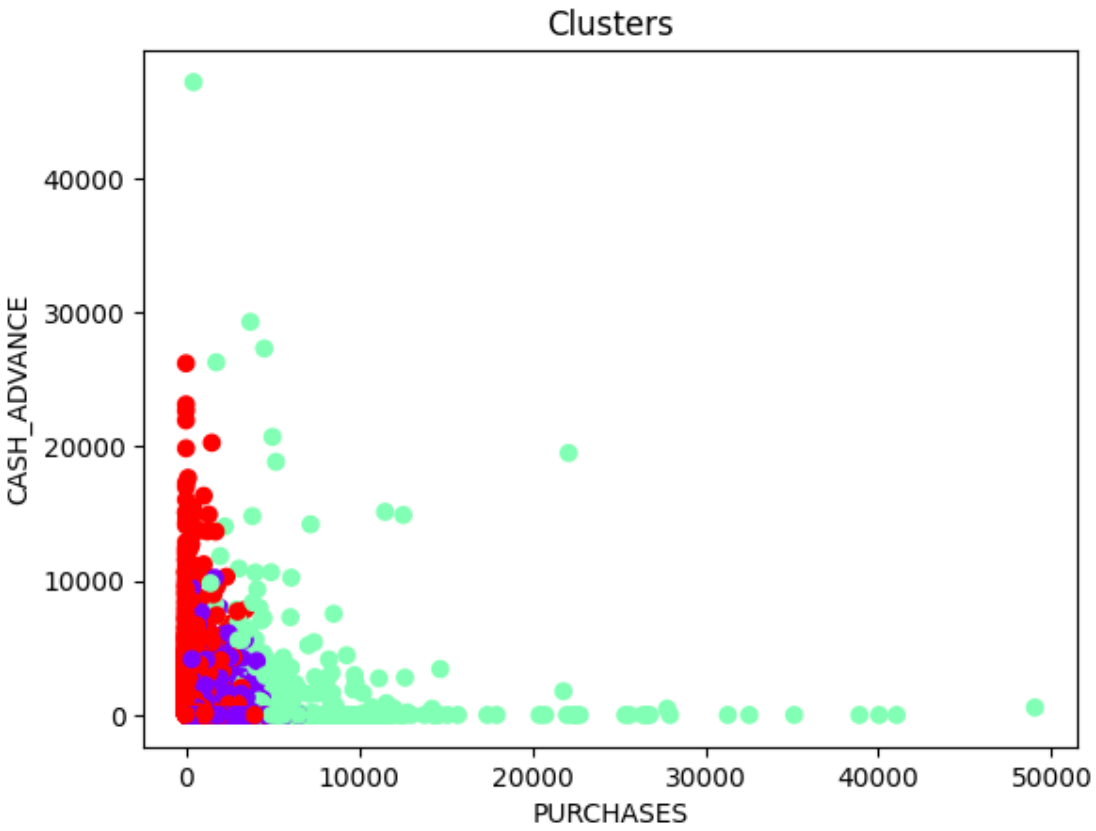
[22]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.impute import SimpleImputer

data_path = "/content/CC_GENERAL.csv"
data = pd.read_csv(data_path)

X = data.drop(['CUST_ID', 'BALANCE'], axis=1)
y = data['BALANCE']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)

imputer = SimpleImputer()
X_train_imputed = imputer.fit_transform(X_train)
X_test_imputed = imputer.transform(X_test)

reg_model = LinearRegression()
reg_model.fit(X_train_imputed, y_train)

y_pred = reg_model.predict(X_test_imputed)

mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
```

```
Mean Squared Error: 1527115.3749535694
```

[24]:
```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
```

```
data_path = "/content/CC_GENERAL.csv"
data = pd.read_csv(data_path)

X = data.drop(['CUST_ID', 'BALANCE'], axis=1)

imputer = SimpleImputer()
X_imputed = imputer.fit_transform(X)

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_imputed)

kmeans = KMeans(n_clusters=3, random_state=42)
data['cluster'] = kmeans.fit_predict(X_scaled)

plt.scatter(data['PURCHASES'], data['CASH_ADVANCE'], c=data['cluster'],␣
 ↪cmap='rainbow')
plt.xlabel('PURCHASES')
plt.ylabel('CASH_ADVANCE')
plt.title('Clusters')
plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
FutureWarning: The default value of `n_init` will change from 10 to 'auto' in
1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```


Clusters
CASH_ADVANCE
PURCHASES
0
10000
20000
30000
40000
50000