Data Import

In [37]:
```
import pandas as pd
```

In [39]:
```
# Load the dataset
df = pd.read_csv("bank_data.csv")
```

Data Exploration

In [41]:
```
# View All Rows
print (df.head())
```

```
  Customer ID  Age  Gender Location (City/State) Account Type  \
0    GTB00001   53  Female                 Lagos     Business
1    GTB00002   29  Female                 Enugu      Current
2    GTB00003   47    Male                 Abuja      Current
3    GTB00004   25    Male                 Abuja      Savings
4    GTB00005   29  Female                Ibadan      Current

   Average Monthly Balance  Transaction Frequency (per month) Loan Status  \
0                841737.10                                 11        Paid
1                 80786.61                                 16      Active
2                270180.37                                 24        Paid
3                954316.72                                  7      Active
4                745757.60                                 17      Active

  Preferred Channels  Service Feedback Score
0                ATM                       5
1                ATM                       4
2         Mobile App                       2
3         Mobile App                       4
4                ATM                       2
```

In [42]:
```
# View the first 10 rows of the dataset
display(df.head(10))
```

	Customer ID	Age	Gender	Location (City/State)	Account Type	Average Monthly Balance	Transaction Frequency (per month)	Loan Status	Preferred Channels	Service Feedback Score
0	GTB00001	53	Female	Lagos	Business	841737.10	11	Paid	ATM	5
1	GTB00002	29	Female	Enugu	Current	80786.61	16	Active	ATM	4
2	GTB00003	47	Male	Abuja	Current	270180.37	24	Paid	Mobile App	2
3	GTB00004	25	Male	Abuja	Savings	954316.72	7	Active	Mobile App	4
4	GTB00005	29	Female	Ibadan	Current	745757.60	17	Active	ATM	2
5	GTB00006	28	Male	Ibadan	Savings	11163.94	30	No Loans	Online Banking	3
6	GTB00007	36	Female	Enugu	Current	117460.00	27	No Loans	Mobile App	5
7	GTB00008	45	Female	Kano	Business	494237.64	22	No Loans	Online Banking	4
8	GTB00009	65	Male	Ibadan	Business	11791.96	7	No Loans	Branch	2
9	GTB00010	65	Female	Enugu	Current	334473.83	2	Paid	Mobile App	4

Check Unique Values in Categorical Columns

In [47]:
```
# Unique values in Age
for col in ["Age"]:
    print(f"{col} unique values: {df[col].unique()}")
```

```
Age unique values: [53 29 47 25 28 36 45 65 63 34 51 60 62 57 56 23 64 52 39 61 48 19 35 22
 46 27 30 42 32 41 20 40 33 49 44 43 54 26 37 59 58 55 38 31 21 50 24 18]
```

In [49]:
```
# Unique values in Gender
for col in ["Gender"]:
    print(f"{col} unique values: {df[col].unique()}")
```

```
Gender unique values: ['Female' 'Male']
```

In [51]:
```
# Unique values in Location
for col in ["Location (City/State)"]:
    print(f"{col} unique values: {df[col].unique()}")
```

```
Location (City/State) unique values: ['Lagos' 'Enugu' 'Abuja' 'Ibadan' 'Kano' 'Port Harcourt']
```

In [53]:
```
# Unique values in Account Type
for col in ["Account Type"]:
    print(f"{col} unique values: {df[col].unique()}")
```

```
Account Type unique values: ['Business' 'Current' 'Savings']
```

In [55]:
```
# Unique values in Loan Status
for col in ["Loan Status"]:
    print(f"{col} unique values: {df[col].unique()}")
```

```
Loan Status unique values: ['Paid' 'Active' 'No Loans']
```

In [57]:
```
# Unique values in Preferred Channels
for col in ["Preferred Channels"]:
    print(f"{col} unique values: {df[col].unique()}")
```

```
Preferred Channels unique values: ['ATM' 'Mobile App' 'Online Banking' 'Branch']
```

In [59]:
```
# Unique values in Service Feedback Score
for col in ["Service Feedback Score"]:
    print(f"{col} unique values: {df[col].unique()}")
```

```
Service Feedback Score unique values: [5 4 2 3 1]
```

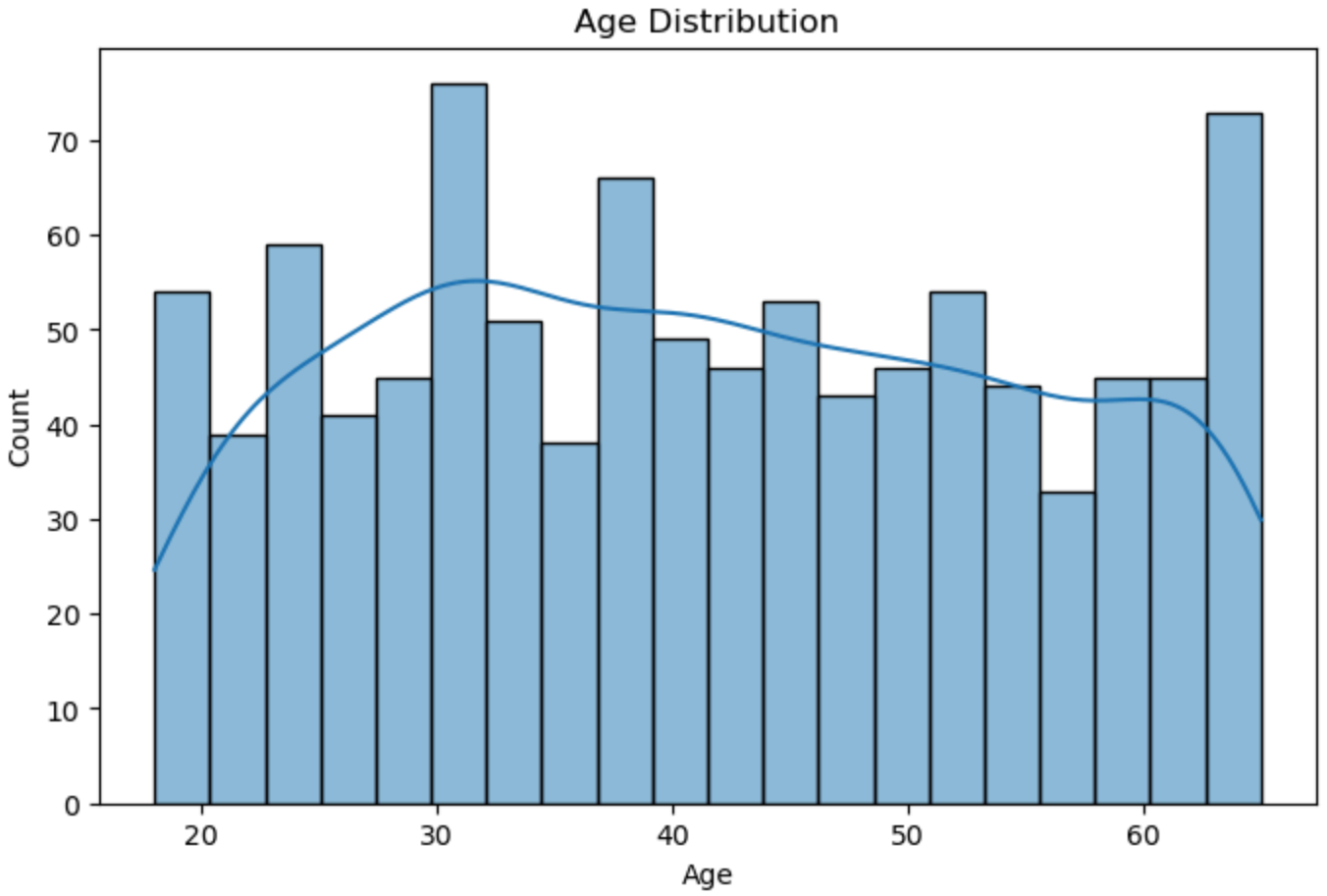
In [61]:
```
# Check for Missing Values in all columns
df.isnull().sum()
```

Out[61]:
```
Customer ID                          0
Age                                  0
Gender                               0
Location (City/State)                0
Account Type                         0
Average Monthly Balance              0
Transaction Frequency (per month)    0
Loan Status                          0
Preferred Channels                   0
Service Feedback Score               0
dtype: int64
```

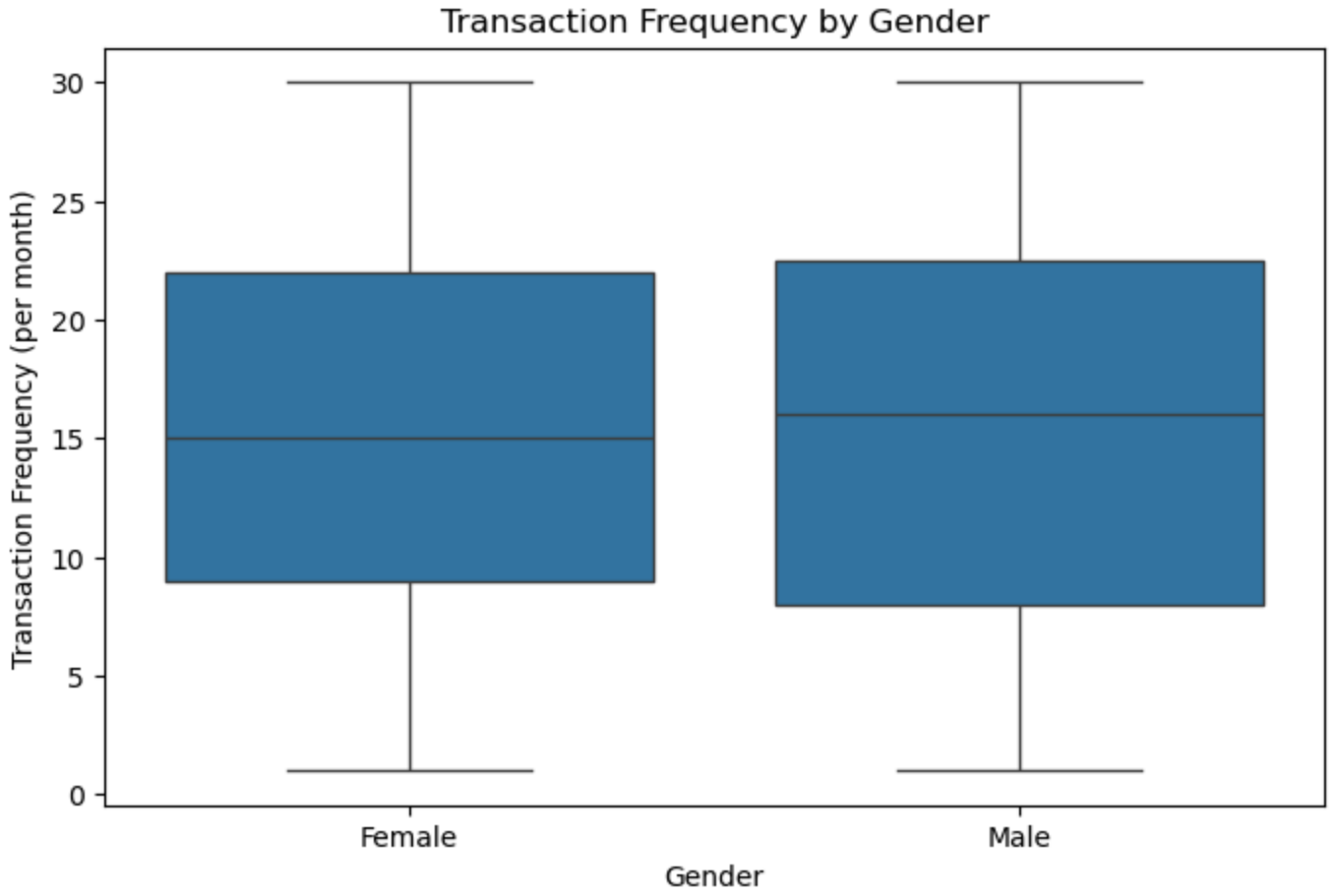
Visualise the Data

In [64]:
```
# Age Distribution
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(8,5))
sns.histplot(df["Age"], bins=20, kde=True)
plt.title("Age Distribution")
plt.show()
```

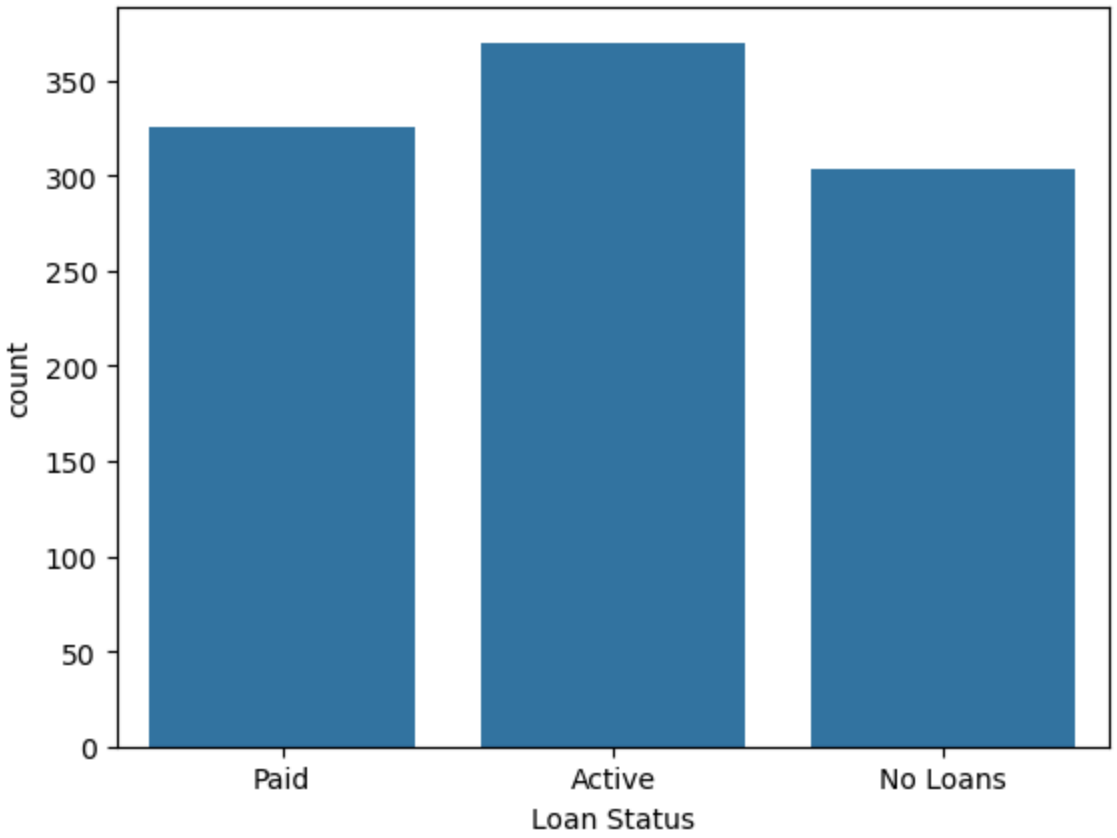


In [66]:
```
# Transaction Frequency by Gender
plt.figure(figsize=(8,5))
sns.boxplot(x=df["Gender"], y=df["Transaction Frequency (per month)"])
plt.title("Transaction Frequency by Gender")
plt.show()
```



In [67]:
```
# Loan Status Count
sns.countplot(x=df["Loan Status"])
```

Out[67]:
```
<Axes: xlabel='Loan Status', ylabel='count'>
```



Data Cleaning & Manipulation

Map States to Regions

In [72]:
```
# Extract the state from the "Location (City/State)
df["State"] = df["Location (City/State)"]
```

In [74]:
```
# Create a dictionary for state-to-region mapping
state_to_region = {
    "Lagos": "South-West",
    "Ibadan": "South-West",
    "Enugu": "South-East",
    "Kano": "North-West",
```

```
    "Abuja": "North-Central",
    "Port Harcourt": "South-South",
}
```

In [76]:
```
# Map states to regions
df["Region"] = df["State"].map(state_to_region)
```

Categorize Age Ranges

In [79]:
```
# Define age groups
bins = [0, 18, 25, 35, 45, 55, 65, 100]
labels = ["Under 18", "18-25", "26-35", "36-45", "46-55", "56-65", "65+"]
df["Age Range"] = pd.cut(df["Age"], bins=bins, labels=labels, right=False)
```

Segmentation based on account balance

In [82]:
```
# Segment customers into low, medium, and high-value groups by average monthly balance using percentiles
df["Value Segment"] = pd.qcut(df["Average Monthly Balance"], q=3, labels=["Low", "Medium", "High"])
```

In [84]:
```
print (df.head())
```

```
  Customer ID  Age  Gender Location (City/State) Account Type  \
0    GTB00001   53  Female                Lagos     Business
1    GTB00002   29  Female                Enugu      Current
2    GTB00003   47    Male                Abuja      Current
3    GTB00004   25    Male                Abuja      Savings
4    GTB00005   29  Female               Ibadan      Current

   Average Monthly Balance  Transaction Frequency (per month) Loan Status  \
0                841737.10                                  11        Paid
1                 80786.61                                  16      Active
2                270180.37                                  24        Paid
3                954316.72                                   7      Active
4                745757.60                                  17      Active

  Preferred Channels  Service Feedback Score   State         Region Age Range  \
0                ATM                       5   Lagos     South-West     46-55
1                ATM                       4   Enugu     South-East     26-35
2         Mobile App                       2   Abuja  North-Central     46-55
3         Mobile App                       4   Abuja  North-Central     26-35
4                ATM                       2  Ibadan     South-West     26-35

  Value Segment
0          High
1           Low
2           Low
3          High
4          High
```

In [86]:
```
display(df.head(10))
```

	Customer ID	Age	Gender	Location (City/State)	Account Type	Average Monthly Balance	Transaction Frequency (per month)	Loan Status	Preferred Channels	Service Feedback Score	State	Region
0	GTB00001	53	Female	Lagos	Business	841737.10	11	Paid	ATM	5	Lagos	South-West
1	GTB00002	29	Female	Enugu	Current	80786.61	16	Active	ATM	4	Enugu	South-East
2	GTB00003	47	Male	Abuja	Current	270180.37	24	Paid	Mobile App	2	Abuja	North-Central
3	GTB00004	25	Male	Abuja	Savings	954316.72	7	Active	Mobile App	4	Abuja	North-Central
4	GTB00005	29	Female	Ibadan	Current	745757.60	17	Active	ATM	2	Ibadan	South-West
5	GTB00006	28	Male	Ibadan	Savings	11163.94	30	No Loans	Online Banking	3	Ibadan	South-West
6	GTB00007	36	Female	Enugu	Current	117460.00	27	No Loans	Mobile App	5	Enugu	South-East
7	GTB00008	45	Female	Kano	Business	494237.64	22	No Loans	Online Banking	4	Kano	North-West
8	GTB00009	65	Male	Ibadan	Business	11791.96	7	No Loans	Branch	2	Ibadan	South-West
9	GTB00010	65	Female	Enugu	Current	334473.83	2	Paid	Mobile App	4	Enugu	South-East

In [63]:
```
# Save the analysis data to a single CSV file
df.to_csv("bank_customer_segmentation_analysis.csv", index=False)
```

Export To SQL Database For Querying & Analysis

In [99]:
```
import sqlite3
```

In [122…
```
conn = sqlite3.connect("bank_data.db")
df.to_sql("customers", conn, if_exists="replace", index=False)
```

Out[122… 1000

Analysis

Convert Segments into Features

In [105…
```
# Use one-hot encoding for categorical data
from sklearn.preprocessing import OneHotEncoder
import pandas as pd
```

In [107…
```
# Load Data
df = pd.read_sql("SELECT * FROM customers", conn)
```

In [113…
```
# Select Features for Recommendation Model
features = ["Preferred Channels", "Age Range", "Loan Status", "Value Segment"]
```

In [ ]:
```
# One-Hot Encoding to Convert Categorical Data
encoder = OneHotEncoder(sparse=False, drop="first")
encoded_features = encoder.fit_transform(df[features])
```

In [ ]:
```
# Convert Encoded Features into a DataFrame
feature_names = encoder.get_feature_names_out(features)
df_encoded = pd.DataFrame(encoded_features, columns=feature_names)
```

In [ ]:
```
# Add Customer IDs Back
df_encoded["Customer ID"] = df["Customer ID"]
```

Train a K-Means Clustering Model for Customer Segmentation

In [118…
```
from sklearn.cluster import KMeans
```

In [ ]:
```
# Define Number of Segments (Clusters)
kmeans = KMeans(n_clusters=5, random_state=42)
df_encoded["Cluster"] = kmeans.fit_predict(df_encoded.drop(columns=["Customer ID"]))
```

In [ ]:
```
# Merge Cluster Labels Back to Original Data
df["Customers"] = df_encoded["Cluster"]
```

Close SQL Database

In [126…
```
conn = sqlite3.connect("bank_data.db")
df.to_sql("customers", conn, if_exists="replace", index=False)
conn.close()
```