Chase Bank Branch Clustering Analysis

This project performs KMeans clustering on Chase bank branches based on their location and recent average deposits.

In [1]:
```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns

# Load data
df = pd.read_csv('branch_consolidation_results.csv')
```

Features and Scaling

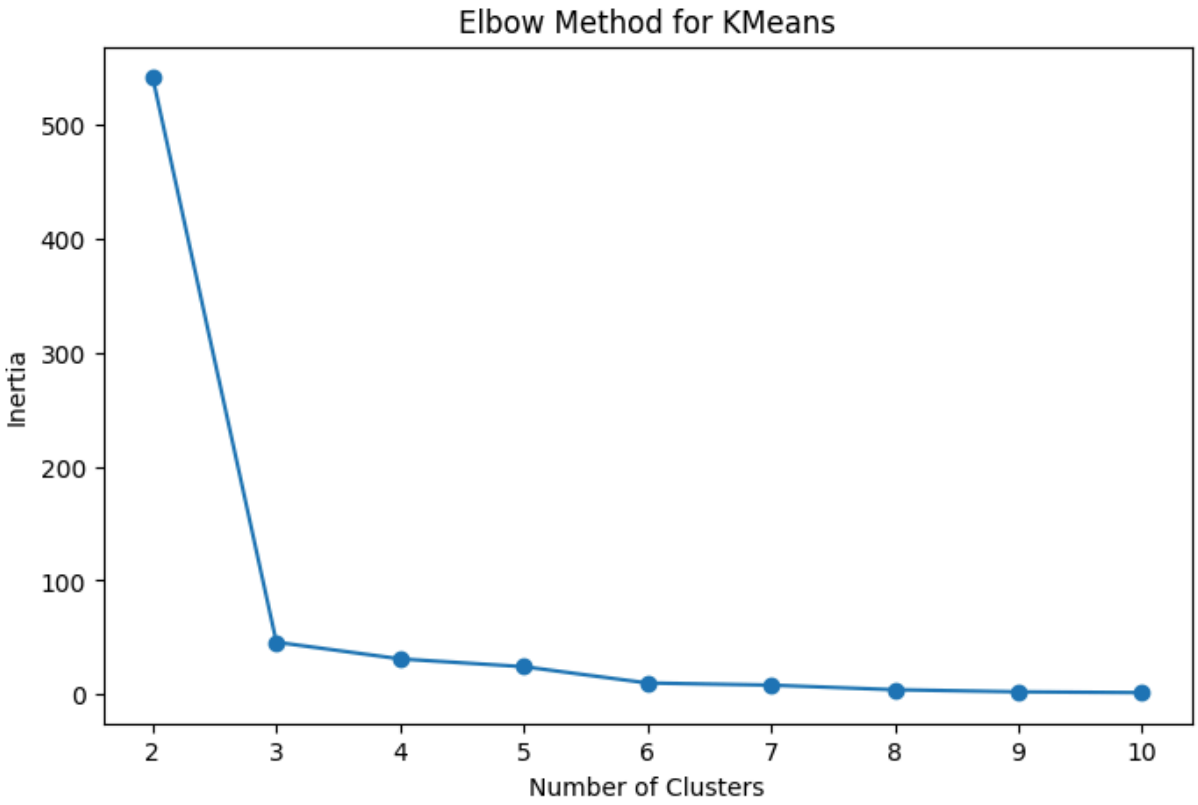
We'll use Latitude, Longitude, and recent_avg_deposits for clustering, scaling them for KMeans.

In [2]:
```
features = ['Latitude', 'Longitude', 'recent_avg_deposits']

# Scale features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df[features])
```

Elbow Method to Determine Number of Clusters

This helps us visually decide the number of clusters to use.

In [3]:
```
inertia = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
    inertia.append(kmeans.inertia_)

plt.figure(figsize=(8,5))
plt.plot(range(2, 11), inertia, marker='o')
plt.xlabel('Number of Clusters')
plt.ylabel('Inertia')
plt.title('Elbow Method for KMeans')
plt.show()
```



Fit KMeans with Chosen Number of Clusters

Here we select 2 clusters based on the elbow plot.

In [4]:
```
k = 2
kmeans = KMeans(n_clusters=k, random_state=1)
df['Cluster'] = kmeans.fit_predict(X_scaled)
```

Cluster Overview

We look at cluster sizes and statistics.

In [6]:
```
# Cluster sizes
print('Cluster Sizes:')
print(df['Cluster'].value_counts(), '\n')

# Cluster description statistics
print('Cluster Statistics:')
print(df.groupby('Cluster')[['recent_avg_deposits', 'Branches_Served', 'Deposits_Se

# Count by County per cluster
print('Branches by County per Cluster:')
print(df.groupby(['Cluster', 'State'])['County'].value_counts(), '\n')

# Average deposits per cluster
print('Average deposits per cluster:')
print(df.groupby('Cluster')['recent_avg_deposits'].mean(), '\n')
```

```
# Total deposits served per cluster
print('Total deposits served per cluster:')
print(df.groupby('Cluster')['Deposits_Served'].sum())
```

```
Cluster Sizes:
Cluster
0    395
1    105
Name: count, dtype: int64

Cluster Statistics:
        recent_avg_deposits                                                \
                      count          mean           std  min           25%
Cluster
0                     395.0  3.196510e+06  5.477004e+07  0.0  89558.000000
1                     105.0  7.878886e+05  6.522592e+06  0.0  64510.333333

                                                Branches_Served            \
                   50%            75%           max           count      mean
Cluster
0        162252.666667  262506.500000  1.085720e+09           395.0  4.281013
1        133620.333333  216507.666667  6.697700e+07           105.0  2.980952

         ...            Deposits_Served                              \
         ...  75%   max           count          mean           std
Cluster  ...
0        ...  5.0  14.0           395.0  4.403073e+06  5.503559e+07
1        ...  3.0   9.0           105.0  6.070312e+06  1.888936e+07

                                                                  \
                  min            25%            50%           75%
Cluster
0        29663.000000  343284.666667  603758.333333  1.123800e+06
1        64553.333333  205013.000000  310355.666667  4.160607e+05


                  max
Cluster
0        1.085720e+09
1        6.746669e+07

[2 rows x 24 columns]

Branches by County per Cluster:
Cluster  State  County
0        CT     Fairfield      25
                New Haven       2
         NJ     Bergen          9
                Essex           9
                Morris          9
                Passaic         3
                Union           3
         NY     New York       74
                Queens         41
                Kings          38
                Westchester    36
                Bronx          32
                Suffolk        32
                Nassau         26
                Monroe         23
```

```
                Onondaga       11
                Richmond        8
                Rockland        3
                Orange          2
                Putnam          2
                Wayne           2
                Madison         1
                Oneida          1
                Ontario         1
                Oswego          1
         OH     Delaware        1
1        TX     Harris         45
                Dallas         22
                El Paso         8
                Tarrant         8
                Travis          5
                Bexar           4
                Montgomery      4
                Cameron         2
                Hidalgo         2
                Collin          1
                Comal           1
                Denton          1
                Fort Bend       1
                Galveston       1
Name: count, dtype: int64

Average deposits per cluster:
Cluster
0    3.196510e+06
1    7.878886e+05
Name: recent_avg_deposits, dtype: float64

Total deposits served per cluster:
Cluster
0    1.739214e+09
1    6.373828e+08
Name: Deposits_Served, dtype: float64
```