In [80]:
```
pip install pandas numpy faker
```

```
Requirement already satisfied: pandas in c:\users\aadesh\anaconda3\lib\site-packages (2.1.4)
Requirement already satisfied: numpy in c:\users\aadesh\anaconda3\lib\site-packages (1.26.4)
Requirement already satisfied: faker in c:\users\aadesh\anaconda3\lib\site-packages (37.0.0)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\aadesh\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\aadesh\anaconda3\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\aadesh\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in c:\users\aadesh\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

In [87]:
```
import pandas as pd
from faker import Faker
import numpy as np

def generate_synthetic_data(num_records):
    fake = Faker()
    data = {
        'Customer ID': [fake.uuid4() for _ in range(num_records)],
        'Name': [fake.name() for _ in range(num_records)],
        'Gender': [fake.random_element(elements=('Male', 'Female')) for _ in range(num_records)],
        'Age': [np.random.randint(18, 80) for _ in range(num_records)],
        'Product Purchased': [fake.random_element(elements=('Laptop', 'Phone', 'Tablet', 'Headphones')) for _ in range(num_records)],
        'Purchase Date': [fake.date_this_decade() for _ in range(num_records)],
        'Purchase Amount': [round(np.random.uniform(100, 2000), 2) for _ in range(num_records)]
    }

    df = pd.DataFrame(data)
    return df  # Make sure you have this return statement

# Generate the synthetic data
num_records = 100  # Example number of records
df_synthetic = generate_synthetic_data(num_records)

# Show a sample of the synthetic data
if df_synthetic is not None:  # Check if df_synthetic is not None
    print(df_synthetic.head())
else:
    print("Error: DataFrame was not generated.")
```

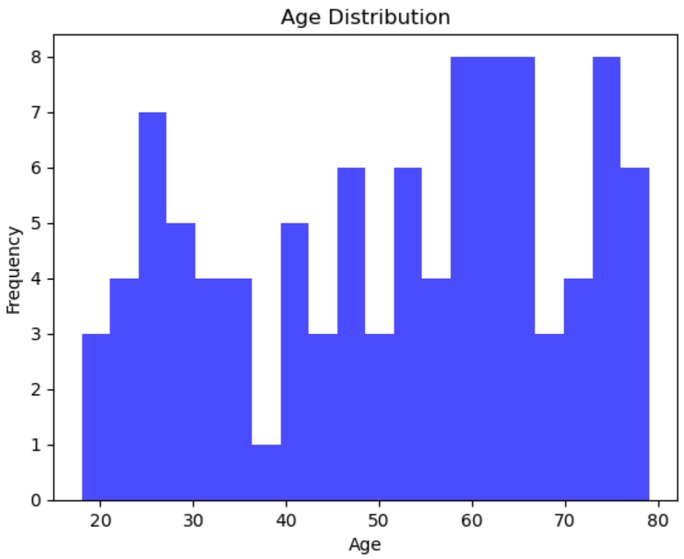
```
                            Customer ID            Name  Gender  Age  \
0  1c85ecd9-9220-40c7-8bed-3b791ba7f863      John Davis    Male   68
1  2dc3ec2d-e664-44c0-9ffa-58a51a638e12     Donald Tate    Male   41
2  a6956e90-584e-4180-93b1-504b50fc05c0  Tiffany Foster  Female   58
3  1d7229ff-b022-4d33-b46d-4c36dee3f2e7     Chloe Perez    Male   70
4  59a98bee-4514-4b04-b94b-4556e88bcb1d       Roy Kline    Male   35

  Product Purchased Purchase Date  Purchase Amount
0            Laptop    2021-07-31          1190.63
1             Phone    2025-02-26          1870.00
2        Headphones    2023-02-14           846.68
3            Tablet    2024-04-14           941.85
4            Tablet    2022-07-13           676.15
```

In [104…
```
# Save the data to a CSV file
df_synthetic.to_csv('synthetic_customer_data.csv', index=False)
```

In [88]:
```
import matplotlib.pyplot as plt

# Plot Age distribution
plt.hist(df_synthetic['Age'], bins=20, color='blue', alpha=0.7)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```


Age Distribution
Frequency
Age
0
1
2
3
4
5
6
7
8
20
30
40
50
60
70
80

In [ ]: