In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

In [92]:
```
df=pd.read_csv("C:/Users/NITISH/Downloads/Online_Shoping.csv")
df
```

Out[92]:

	City	Gender	Online_Shoping	Amazon_User	Flipkart_User	Myntra_User	Meesho_U
0	Lucknow	Male	Yes	Amazon	Flipkart	Myntra	N
1	Lucknow	Male	Yes	Amazon	Flipkart	NaN	N
2	Kolkata	Female	Yes	Amazon	Flipkart	NaN	N
3	Indore	Male	Yes	Amazon	Flipkart	Myntra	N
4	Kolkata	Female	Yes	Amazon	Flipkart	Myntra	Mee
...	...	...	...	...	...	...	
195	Indore	Male	Yes	Amazon	Flipkart	Myntra	N
196	Mumbai	Male	Yes	NaN	NaN	NaN	N
197	Delhi	Male	Yes	Amazon	Flipkart	NaN	N
198	Kolkata	Female	Yes	Amazon	Flipkart	NaN	N
199	Delhi	Male	Yes	NaN	Flipkart	NaN	Mee

200 rows × 13 columns

In [29]:
```
print("Original DataFrame shape:", df.shape)
```

```
Original DataFrame shape: (200, 13)
```

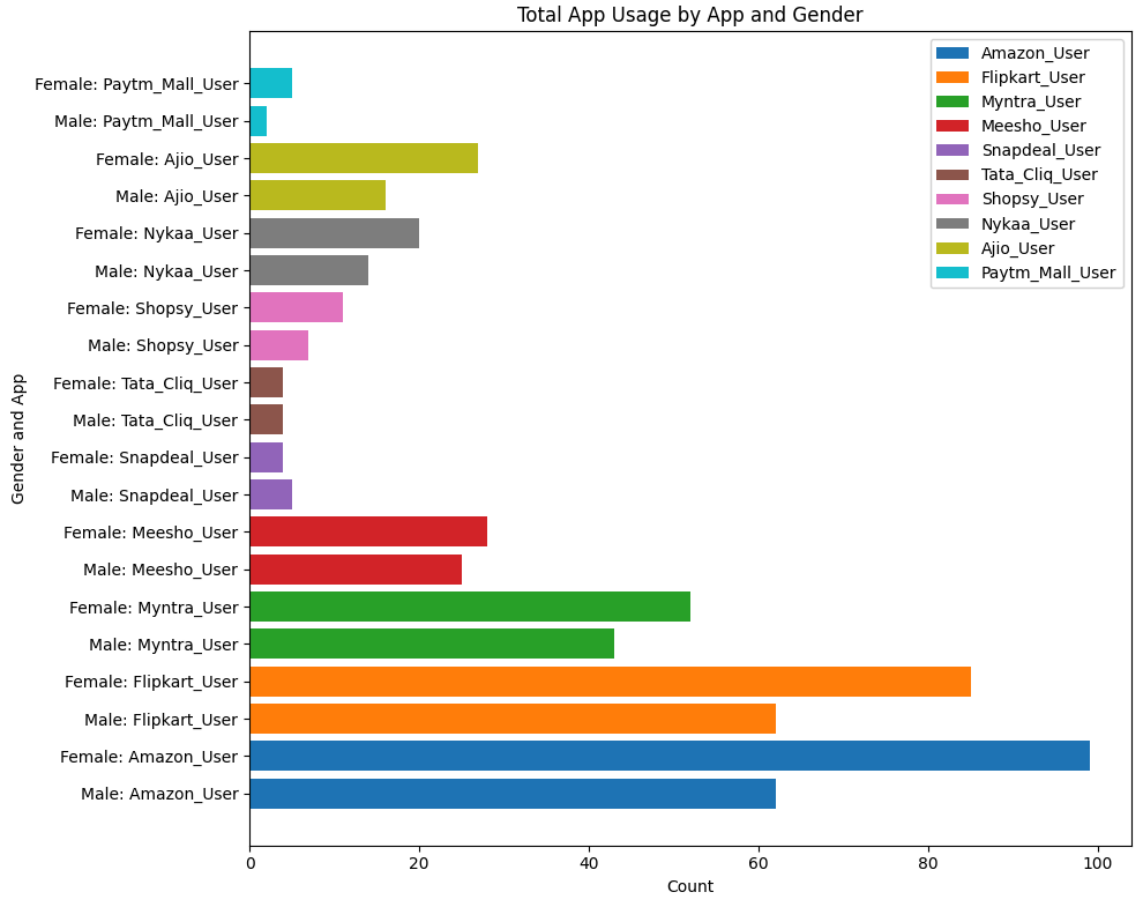
In [30]:
```
app_columns = df.columns[3:]
app_columns
```

Out[30]:
```
Index(['Amazon_User', 'Flipkart_User', 'Myntra_User', 'Meesho_User',
       'Snapdeal_User', 'Tata Cliq_User', 'Shopsy (by Flipkart)_User',
       'Nykaa_User', 'Ajio_User', 'Paytm Mall_User'],
      dtype='object')
```

In [6]:
```
app_counts = df.groupby(['Gender']).agg({'Amazon_User':'count', 'Flipkart_U
print(app_counts)
```

```
   Gender  Amazon_User  Flipkart_User  Myntra_User  Meesho_User  \
0  Female           62             62           43           25
1    Male           99             85           52           28

   Snapdeal_User  Tata Cliq_User  Shopsy (by Flipkart)_User  Nykaa_User  \
0              5               4                          7          14
1              4               4                         11          20

   Ajio_User  Paytm Mall_User
0         16                2
1         27                5
```

In [11]:
```
import matplotlib.pyplot as plt
apps = ['Amazon_User', 'Flipkart_User', 'Myntra_User', 'Meesho_User', 'Snap
genders = df['Gender'].unique()

# Set the figure size
plt.figure(figsize=(10, 8))

# Create a horizontal bar plot
for i, app in enumerate(apps):
    plt.barh([f"{Gender}: {app}" for Gender in genders], app_counts.iloc[:,

# Add labels and title
plt.xlabel('Count')
plt.ylabel('Gender and App')
plt.title('Total App Usage by App and Gender')

# Add legend
plt.legend()

# Show the plot
plt.tight_layout()
plt.show()
```

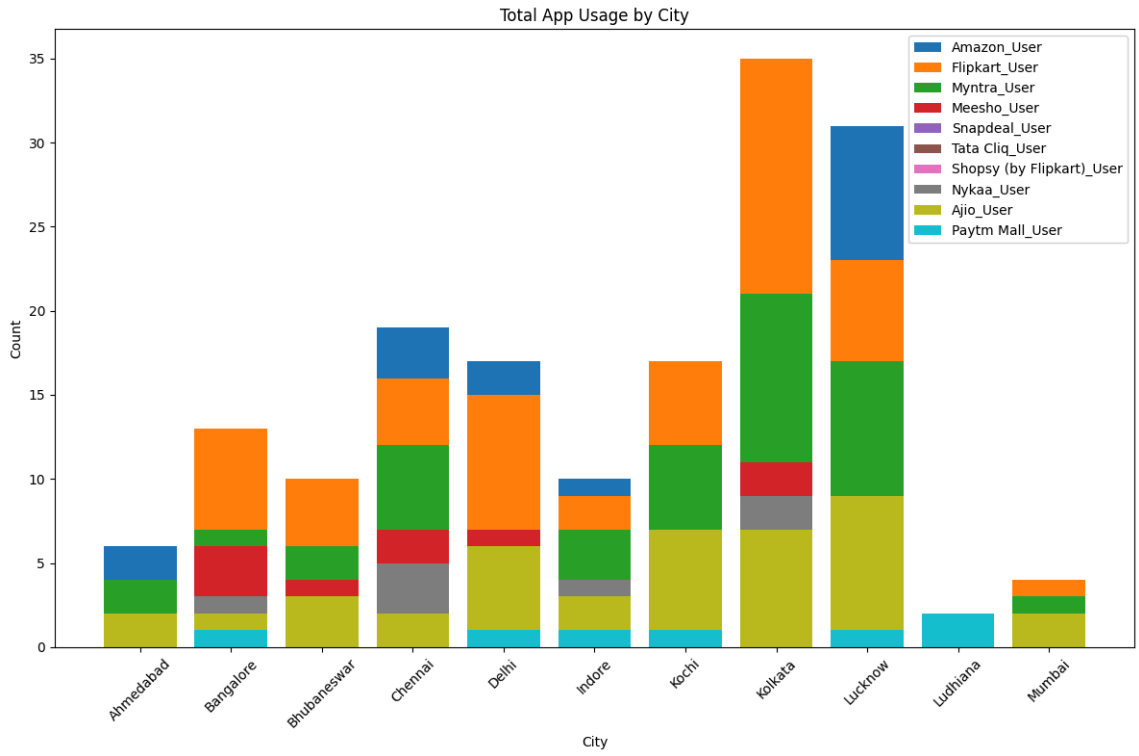


In [12]:
```
app_counts1 = df.groupby(['City']).agg({'Amazon_User':'count', 'Flipkart_Us
app_counts1
```

Out[12]:

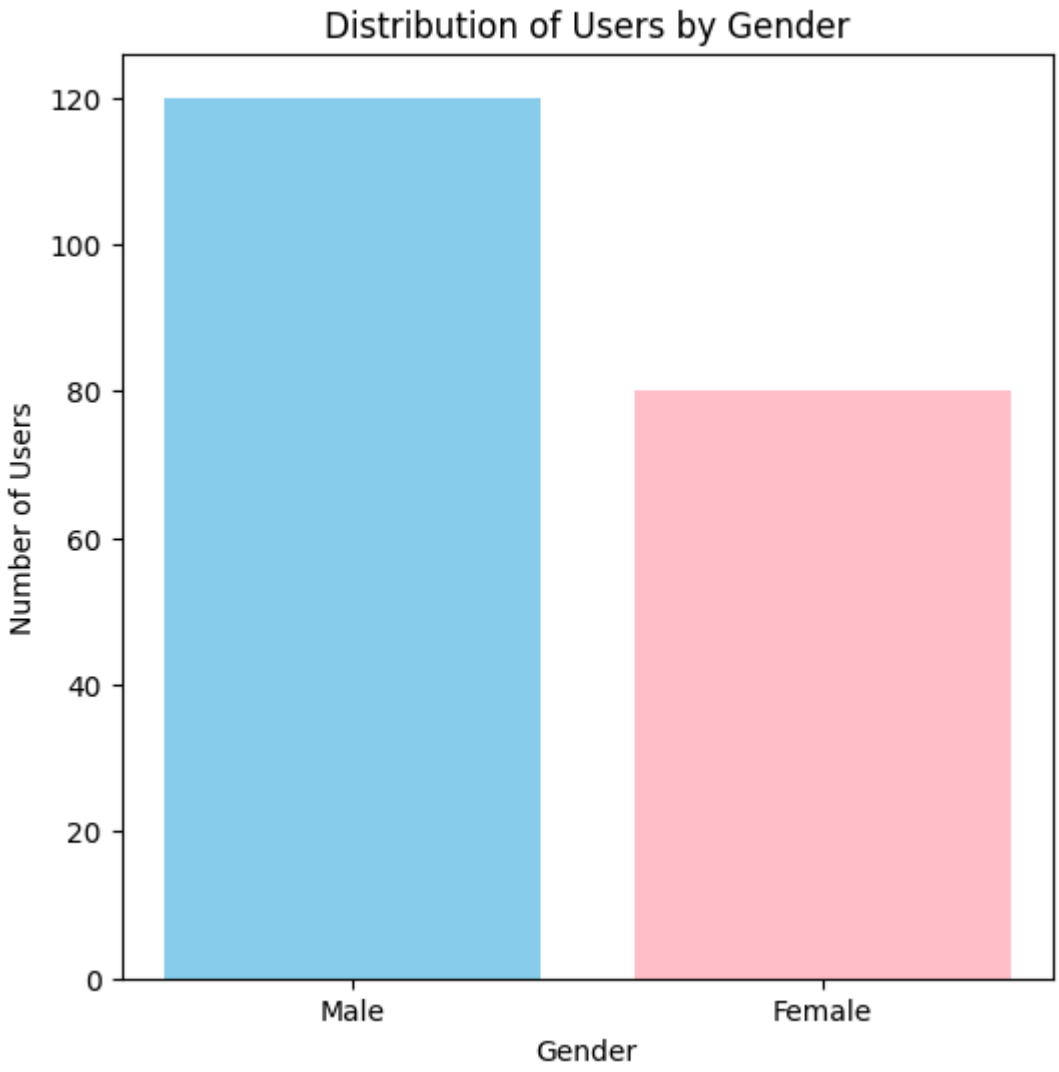
	City	Amazon_User	Flipkart_User	Myntra_User	Meesho_User	Snapdeal_User	C
0	Ahmedabad	6	4	4	0	0	
1	Bangalore	12	13	7	6	0	
2	Bhubaneswar	10	10	6	4	0	
3	Chennai	19	16	12	7	0	
4	Delhi	17	15	6	7	2	
5	Indore	10	9	7	2	0	
6	Kochi	17	17	12	6	2	
7	Kolkata	34	35	21	11	2	
8	Lucknow	31	23	17	9	2	
9	Ludhiana	2	1	0	0	0	
10	Mumbai	3	4	3	1	1	

In [13]:
```
city_counts = df.groupby('City').count().drop(columns=['Gender', 'Online_Sh

# Plot the data
plt.figure(figsize=(12, 8))
for app in city_counts.columns[1:]:
    plt.bar(city_counts['City'], city_counts[app], label=app)

plt.xlabel('City')
plt.ylabel('Count')
plt.title('Total App Usage by City')
plt.xticks(rotation=45)
plt.legend()
plt.tight_layout()
plt.show()
```



In [ ]:
```

```

In [15]:
```
# 1. Correlation between Gender and Online Shopping

# Count the number of users by gender
gender_counts = df['Gender'].value_counts()

# Plot the distribution of users by gender
plt.figure(figsize=(6, 6))
plt.bar(gender_counts.index, gender_counts.values, color=['skyblue', 'pink'
plt.xlabel('Gender')
plt.ylabel('Number of Users')
plt.title('Distribution of Users by Gender')
plt.show()
```



In [16]:
```
'City' in df.columns
```

Out[16]: True

In [17]:
```
df=pd.read_csv("C:/Users/NITISH/Downloads/Online_Shoping.csv")
df
```

Out[17]:

	City	Gender	Online_Shoping	Amazon_User	Flipkart_User	Myntra_User	Meesho_U
0	Lucknow	Male	Yes	Amazon	Flipkart	Myntra	N
1	Lucknow	Male	Yes	Amazon	Flipkart	NaN	N
2	Kolkata	Female	Yes	Amazon	Flipkart	NaN	N
3	Indore	Male	Yes	Amazon	Flipkart	Myntra	N
4	Kolkata	Female	Yes	Amazon	Flipkart	Myntra	Mee
...	...	...	...	...	...	...	
195	Indore	Male	Yes	Amazon	Flipkart	Myntra	N
196	Mumbai	Male	Yes	NaN	NaN	NaN	N
197	Delhi	Male	Yes	Amazon	Flipkart	NaN	N
198	Kolkata	Female	Yes	Amazon	Flipkart	NaN	N
199	Delhi	Male	Yes	NaN	Flipkart	NaN	Mee

200 rows × 13 columns

In [18]:
```
# Extract app column names from the dataframe
app_columns = ['Amazon_User', 'Flipkart_User', 'Myntra_User', 'Meesho_User'

# Concatenate the city column with the app columns
df_apps = df[['City'] + app_columns]

# Melt the dataframe to convert wide format to long format
df_melted = pd.melt(df_apps, id_vars=['City'], value_vars=app_columns, var_

# Remove rows where User_Count is 0
df_melted = df_melted[df_melted['User_Count'] != 0]

# Display the preprocessed dataframe
print(df_melted.head())
```

```
      City          App User_Count
0  Lucknow  Amazon_User     Amazon
1  Lucknow  Amazon_User     Amazon
2  Kolkata  Amazon_User     Amazon
3   Indore  Amazon_User     Amazon
4  Kolkata  Amazon_User     Amazon
```

In [19]:
```
# Assuming the 'User_Count' column contains the app names instead of counts
# we need to rename the column to 'App' and create a new column 'User_Count

# Rename the 'User_Count' column to 'App'
df_melted.rename(columns={'User_Count': 'App'}, inplace=True)

# Display the preprocessed dataframe
print(df_melted.head())
```

```
      City          App     App
0  Lucknow  Amazon_User  Amazon
1  Lucknow  Amazon_User  Amazon
2  Kolkata  Amazon_User  Amazon
3   Indore  Amazon_User  Amazon
4  Kolkata  Amazon_User  Amazon
```

In [20]:
```
# Rename the columns to clarify their meaning
df.rename(columns={'App': 'App_Name', 'App.1': 'User_Count'}, inplace=True)

# Display the preprocessed dataframe
print(df.head())
```

```
      City  Gender Online_Shoping Amazon_User Flipkart_User Myntra_User  \
0  Lucknow    Male            Yes      Amazon      Flipkart      Myntra
1  Lucknow    Male            Yes      Amazon      Flipkart         NaN
2  Kolkata  Female            Yes      Amazon      Flipkart         NaN
3   Indore    Male            Yes      Amazon      Flipkart      Myntra
4  Kolkata  Female            Yes      Amazon      Flipkart      Myntra

  Meesho_User Snapdeal_User Tata Cliq_User Shopsy (by Flipkart)_User  \
0         NaN           NaN            NaN                        NaN
1         NaN           NaN            NaN                        NaN
2         NaN           NaN            NaN                        NaN
3         NaN           NaN      Tata Cliq                        NaN
4      Meesho           NaN            NaN                        NaN

  Nykaa_User Ajio_User Paytm Mall_User
0        NaN      Ajio             NaN
1        NaN       NaN             NaN
2        NaN       NaN             NaN
3        NaN      Ajio             NaN
4      Nykaa       NaN             NaN
```

In [25]:
```
df_filtered = df.dropna(subset=app_columns, how='any')
df_filtered
```

Out[25]:

	City	Gender	Online_Shoping	Amazon_User	Flipkart_User	Myntra_User	Meesho_Use
177	Kochi	Female	Yes	Amazon	Flipkart	Myntra	Meesho

In [26]:
```
print("Filtered DataFrame shape:", df_filtered.shape)
```

```
Filtered DataFrame shape: (1, 13)
```

In [33]:
```
app_counts_by_city = df.groupby("City").sum()
app_counts_by_city_numeric = app_counts_by_city.drop(columns=['Gender', 'On
```

In [35]:
```
import seaborn as sns
```

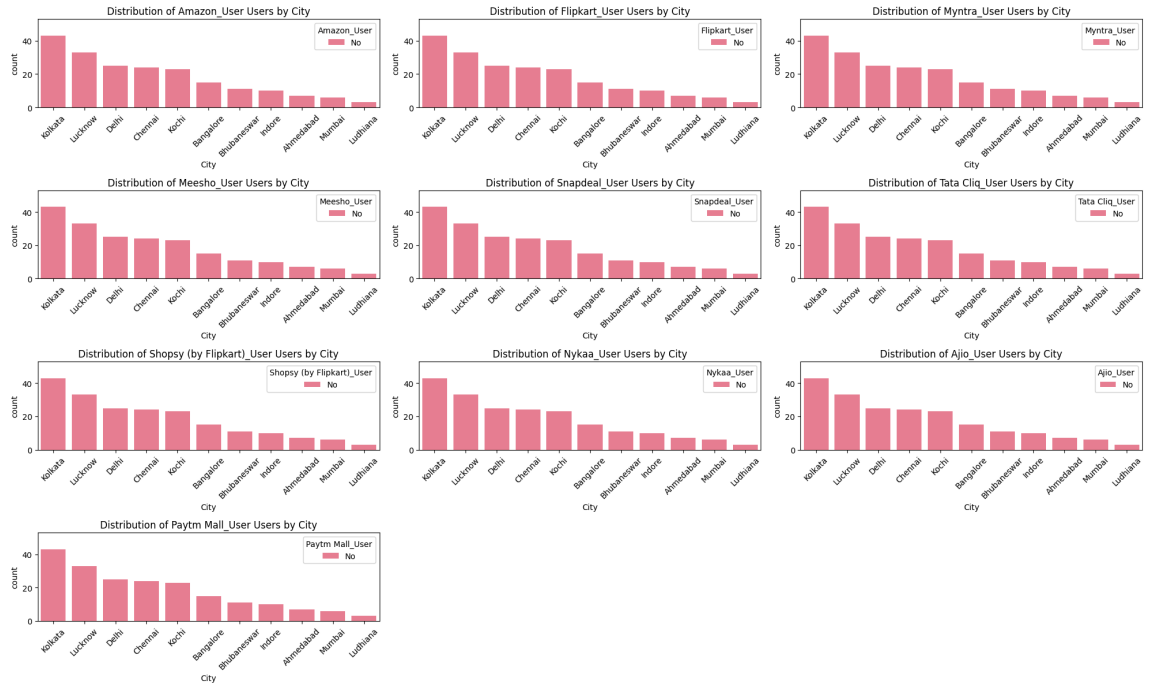
In [36]:
```
categorical_columns = ['Gender', 'Online_Shoping', 'Amazon_User', 'Flipkart
                       'Myntra_User', 'Meesho_User', 'Snapdeal_User', 'Tata
                       'Shopsy (by Flipkart)_User', 'Nykaa_User', 'Ajio_Use
                       'Paytm_Mall_User']
```

In [45]:
```
print(df.columns)
```

```
Index(['City', 'Gender', 'Online_Shoping', 'Amazon_User', 'Flipkart_User',
       'Myntra_User', 'Meesho_User', 'Snapdeal_User', 'Tata Cliq_User',
       'Shopsy (by Flipkart)_User', 'Nykaa_User', 'Ajio_User',
       'Paytm Mall_User'],
      dtype='object')
```

In [46]:
```
for app in df.columns[3:]:
    df[app] = df[app].apply(lambda x: 'Yes' if x == 'Yes' else 'No')
```

In [53]:
```
city_palette = sns.color_palette("husl", len(df['City'].unique()))

# Set the figure size
plt.figure(figsize=(20, 12))  # Increase the figure size

# Iterate over each app column and create a countplot
for i, app in enumerate(df.columns[3:], 1):
    plt.subplot(4, 3, i)
    sns.countplot(data=df, x='City', hue=app, order=df['City'].value_counts
    plt.title(f'Distribution of {app} Users by City')
    plt.xlabel('City')
    plt.xticks(rotation=45)  # Rotate x-axis labels for better readability
    plt.ylim(0, df['City'].value_counts().max() + 10)  # Adjust y-axis limi

# Adjust layout
plt.tight_layout()

plt.show()
```



In [55]:
```
contingency_table = pd.crosstab(df['Gender'], df['Amazon_User'])
```

In [57]:
```
from scipy.stats import chi2_contingency
```

In [58]:
```
# Perform chi-square test
chi2_stat, p_val, dof, expected = chi2_contingency(contingency_table)
print("\nChi-Square Test:")
print("Chi-Square Statistic:", chi2_stat)
print("P-value:", p_val)
print("Degrees of Freedom:", dof)
print("Expected Frequencies:\n", expected)
```

```
Chi-Square Test:
Chi-Square Statistic: 0.0
P-value: 1.0
Degrees of Freedom: 0
Expected Frequencies:
 [[ 80.]
 [120.]]
```

From the test we came to know that there is no relationship between gender and Amazon users

In [93]:
```
df
```

Out[93]:

	City	Gender	Online_Shoping	Amazon_User	Flipkart_User	Myntra_User	Meesho_U
0	Lucknow	Male	Yes	Amazon	Flipkart	Myntra	N
1	Lucknow	Male	Yes	Amazon	Flipkart	NaN	N
2	Kolkata	Female	Yes	Amazon	Flipkart	NaN	N
3	Indore	Male	Yes	Amazon	Flipkart	Myntra	N
4	Kolkata	Female	Yes	Amazon	Flipkart	Myntra	Mee
...	...	...	...	...	...	...	
195	Indore	Male	Yes	Amazon	Flipkart	Myntra	N
196	Mumbai	Male	Yes	NaN	NaN	NaN	N
197	Delhi	Male	Yes	Amazon	Flipkart	NaN	N
198	Kolkata	Female	Yes	Amazon	Flipkart	NaN	N
199	Delhi	Male	Yes	NaN	Flipkart	NaN	Mee

200 rows × 13 columns

In [ ]:
```

```