Wolt Data Science Internship – Task 2 Analysis

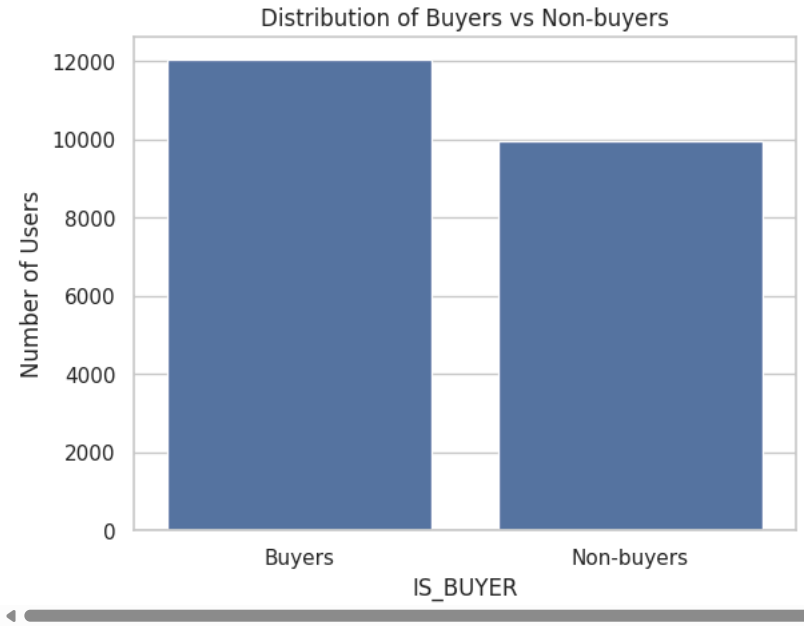
Exploratory Data Analysis and Insights from user data.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style='whitegrid')

# Load dataset
df = pd.read_csv('dataset_for_datascience_assignment.csv')
df['IS_BUYER'] = df['PURCHASE_COUNT'] > 0
df.head()
```

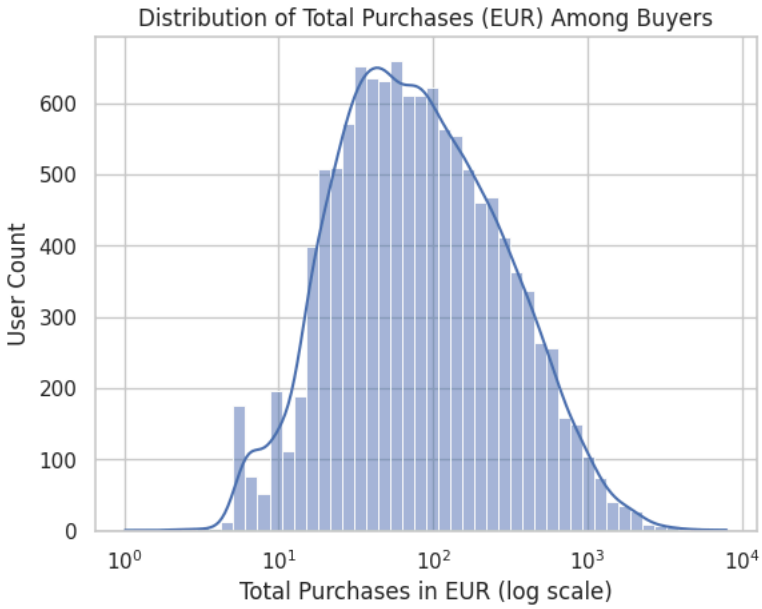
Buyer vs Non-Buyer Distribution

```
buyer_counts = df['IS_BUYER'].value_counts().rename({True: 'Buyers', False: 'Non-buyers'})
sns.barplot(x=buyer_counts.index, y=buyer_counts.values)
plt.title('Distribution of Buyers vs Non-buyers')
plt.ylabel('Number of Users')
plt.show()
```



Total Purchase Value Distribution Among Buyers (Log Scale)

```
buyers_df = df[df['IS_BUYER']]
sns.histplot(buyers_df['TOTAL_PURCHASES_EUR'], bins=50, kde=True, log_scale=True)
plt.title('Distribution of Total Purchases (EUR) Among Buyers')
plt.xlabel('Total Purchases in EUR (log scale)')
plt.ylabel('User Count')
plt.show()
```


Distribution of Total Purchases (EUR) Among Buyers
User Count
0
100
200
300
400
500
600
10^0
10^1
10^2
10^3
10^4
Total Purchases in EUR (log scale)