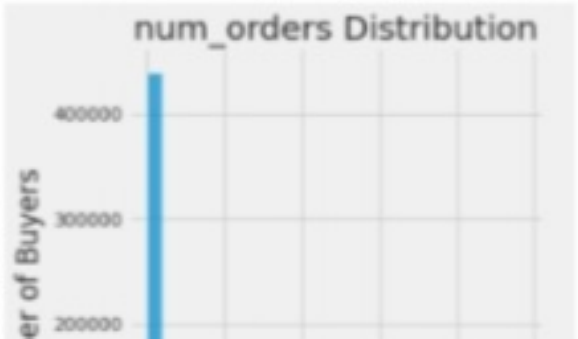
Data Visualization

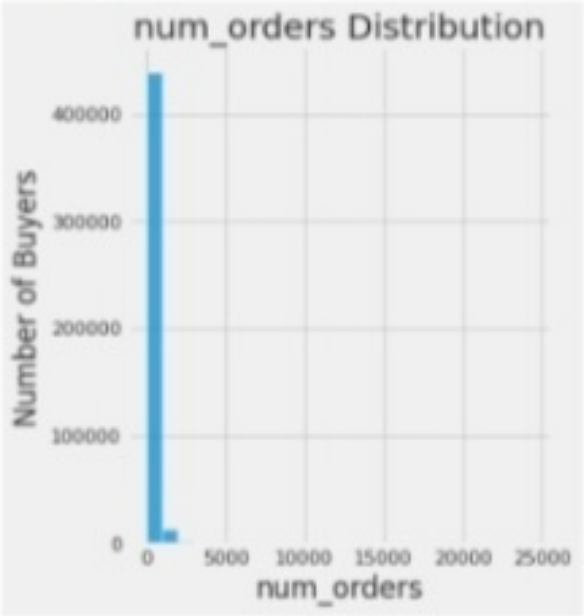
In [118]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [119]:
```
plt.style.use('fivethirtyeight')
plt.figure(figsize=(12,7))
sns.displot(trainfinal.num_orders, bins = 25)
plt.xlabel("num_orders")
plt.ylabel("Number of Buyers")
plt.title("num_orders Distribution")
```

Out[119]: Text(0.5, 1.0, 'num_orders Distribution')

<Figure size 864x504 with 0 Axes>





Drop the column "id" and find the correlation between the columns.

In [120]:
```
trainfinal2 = trainfinal.drop(['id'],axis=1)
correlation = trainfinal2.corr(method='pearson')
columns = correlation.nlargest(8, 'num_orders').index
columns
```