In [1]:
```
from completejourney_py import get_data
import matplotlib.pyplot as plt
cj_data = get_data()

import pandas as pd
transactions = cj_data['transactions']
products = cj_data['products']
demographics = cj_data['demographics']
df1 = transactions.merge(demographics, how='left', on='household_id')
pizza_filter = products['product_type'].str.contains('pizza', case=False, na=False)
df2 = (products[pizza_filter].merge(df1 , how='inner', on='product_id')
       .groupby(['marital_status'], as_index=False)
       .agg({'product_type': 'count', 'sales_value': 'sum'})
      )
df2 = df2.rename(columns={'product_type': 'pizza_product_purchased'})
df2
```

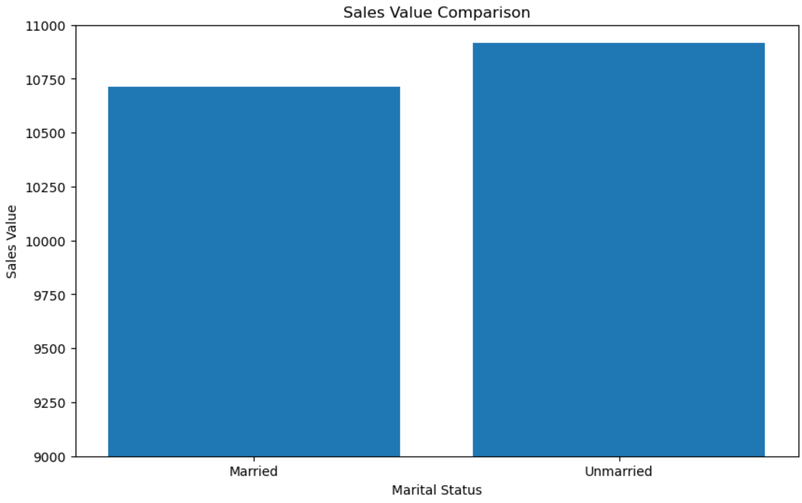
Out[1]:

| | marital_status | pizza_product_purchased | sales_value |
|---|---|---|---|
| **0** | Married | 2788 | 10711.37 |
| **1** | Unmarried | 3037 | 10917.73 |

In [8]:
```
fig = plt.figure(figsize = (10, 6))
plt.bar(df2['marital_status'], df2['sales_value'])

plt.title('Sales Value Comparison')
plt.xlabel('Marital Status')
plt.ylabel('Sales Value')
plt.ylim(9000, 11000)

plt.show()
```

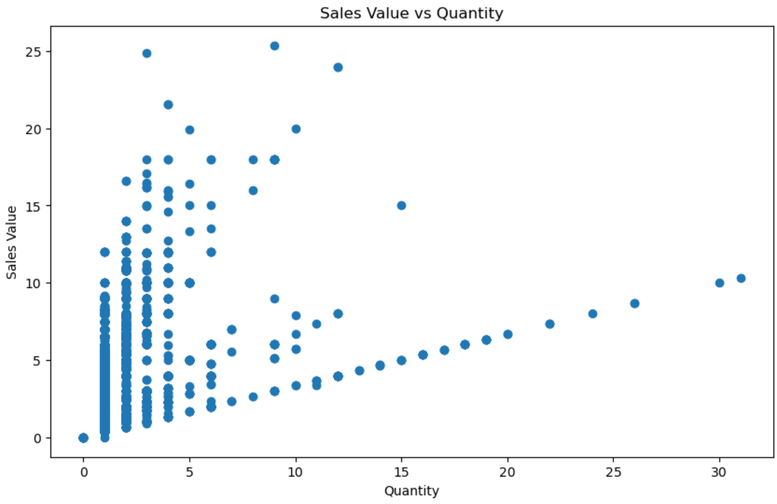



In [10]:
```
df3 = (products[pizza_filter]
       .merge(transactions , how='inner', on='product_id')
       .merge(demographics, how='inner', on='household_id')
      )

fig = plt.figure(figsize = (10, 6))

plt.scatter('quantity', 'sales_value', data=df3)

plt.title('Sales Value vs Quantity')
plt.xlabel('Quantity')
plt.ylabel('Sales Value')
plt.show()
```

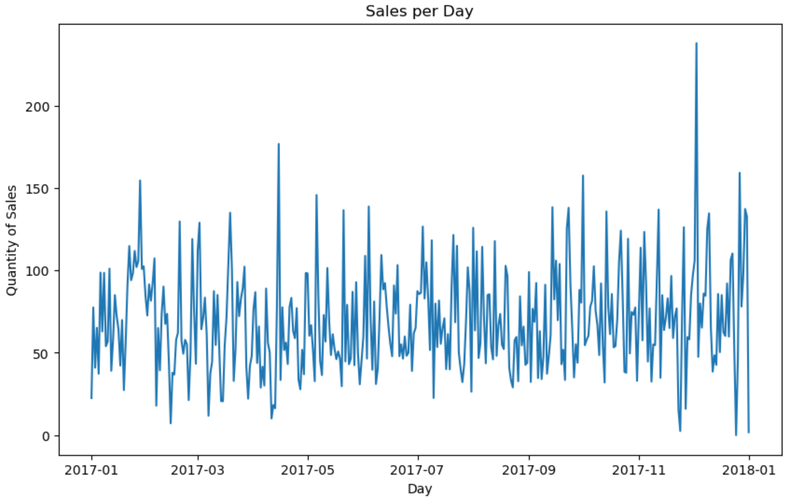



In [13]:
```
df4 = df3.set_index('transaction_timestamp')['sales_value']
sale_resample = df4.resample('D').sum()
```

In [17]:
```
fig = plt.figure(figsize = (10, 6))

plt.plot(sale_resample)
plt.title('Sales per Day')
plt.xlabel('Day')
plt.ylabel('Quantity of Sales')
```

Out[17]:
```
Text(0, 0.5, 'Quantity of Sales')
```




In [ ]: