

E-commerce Furniture Data Analysis Project

Name: Isha Koolwal

Date: 15/08/2025

Tools: ML, Python, SQL, Excel

1. Introduction

This project analyzes an E-commerce furniture dataset containing details about various products, their prices, shipping information, and the number of units sold. The goal is to explore the dataset, visualize trends, and identify factors influencing sales.

2. Dataset Overview

The dataset includes columns such as:

- productTitle: Name/description of the product
- price: Current selling price of the item
- sold: Number of units sold
- tagText: Shipping or promotional tags
- originalPrice (optional): Original price before discounts

3. Data Cleaning & Preprocessing

Steps performed:

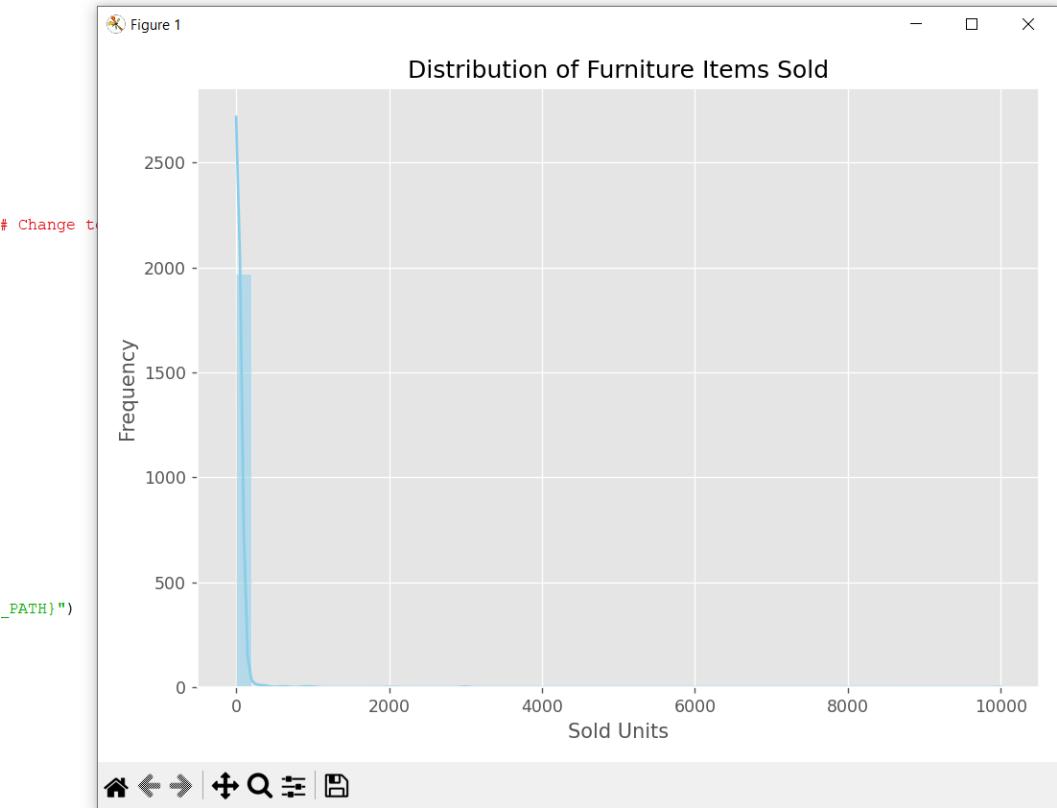
1. Removed rows with missing critical values.
2. Converted 'price' to numeric by removing currency symbols.
3. Converted 'sold' to numeric.
4. Encoded 'tagText' values and grouped rare categories into 'Others'.
5. Added text-based features: title length and word count.

4. Exploratory Data Analysis

Below are some of the visualizations created during EDA:

Distribution of Furniture Items Sold

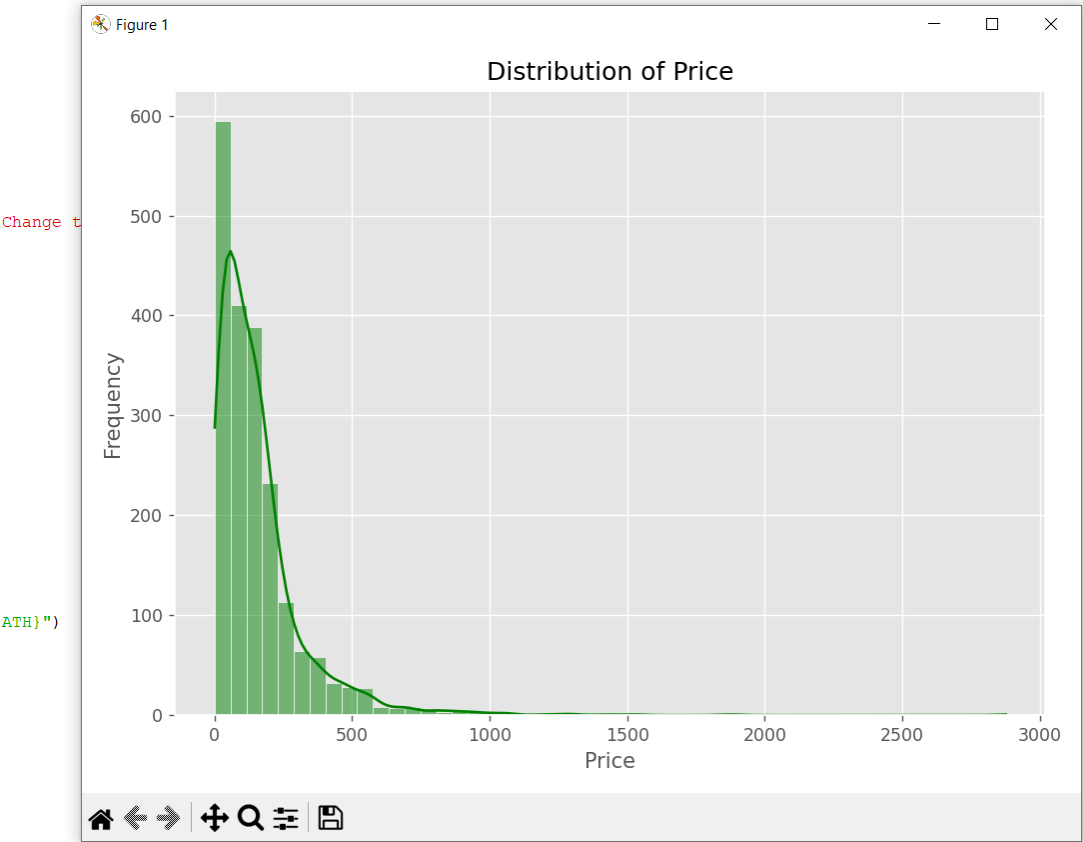
3/ecommerce.py (3.13.5)



'productTitle']])

Distribution of Price

commerce.py (3.13.5)



Price Vs Sold

```
. 'sold', 'tagText']
```

```
title ...  
.. .... Fr  
.. .... Fr  
.. .... Fr  
.. .... Fr  
.. .... Fr
```

```
.png  
s.png
```

```
'Python313/  
iridis', or
```

```
cated and w
```

```
f  
old.png  
.png
```

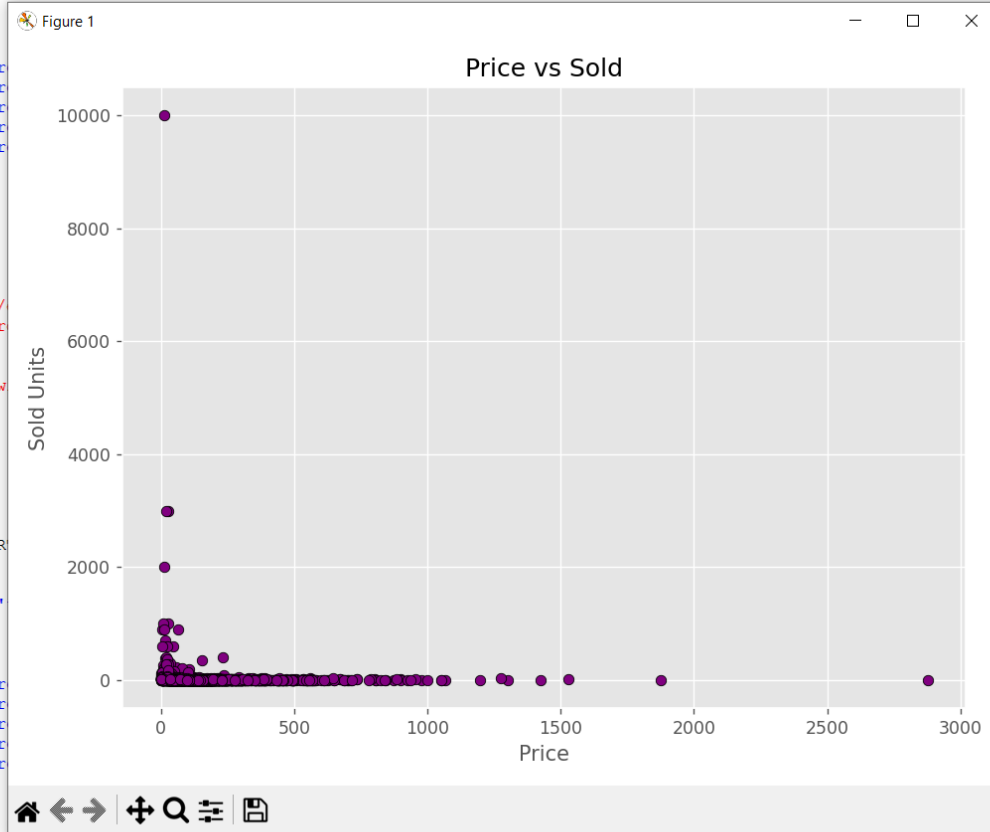
```
figures
```

```
==== RESTART
```

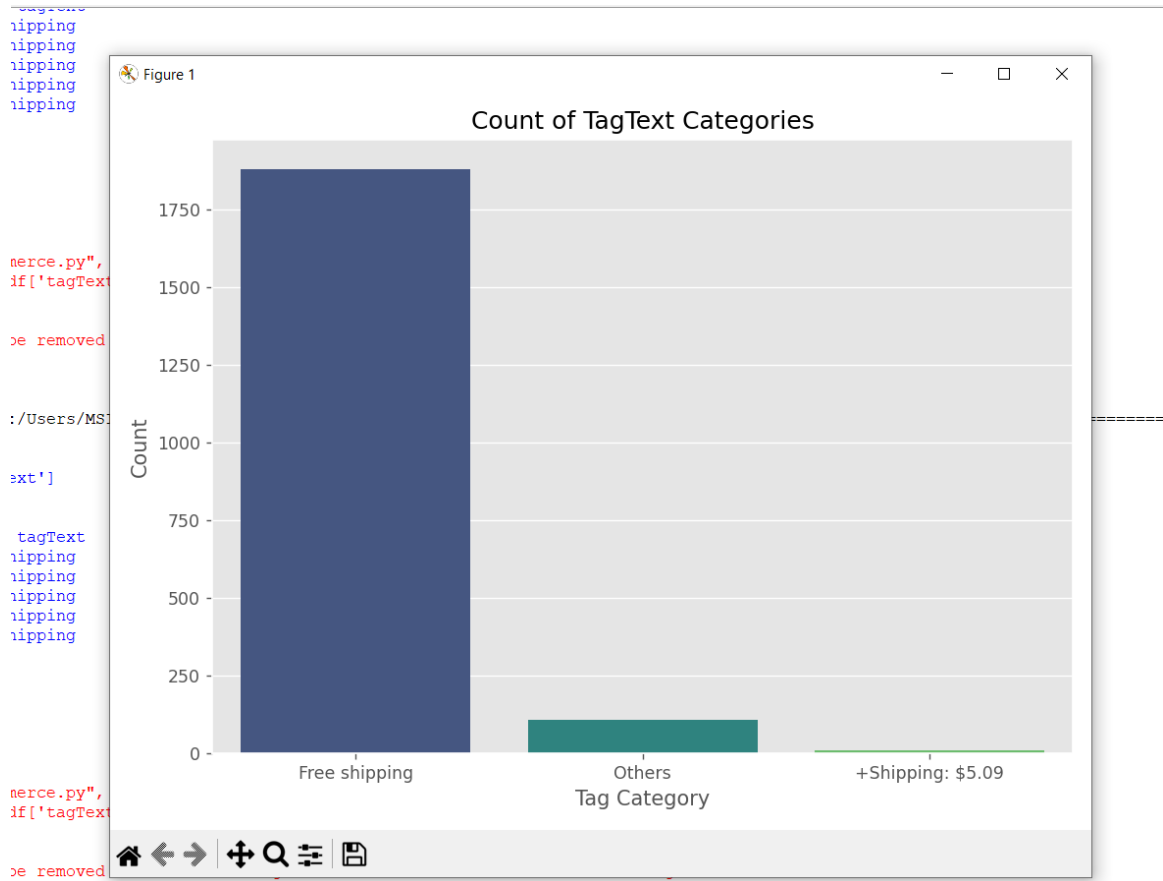
```
. 'sold', '
```

```
title ...  
.. .... Fr  
.. .... Fr  
.. .... Fr  
.. .... Fr  
.. .... Fr
```

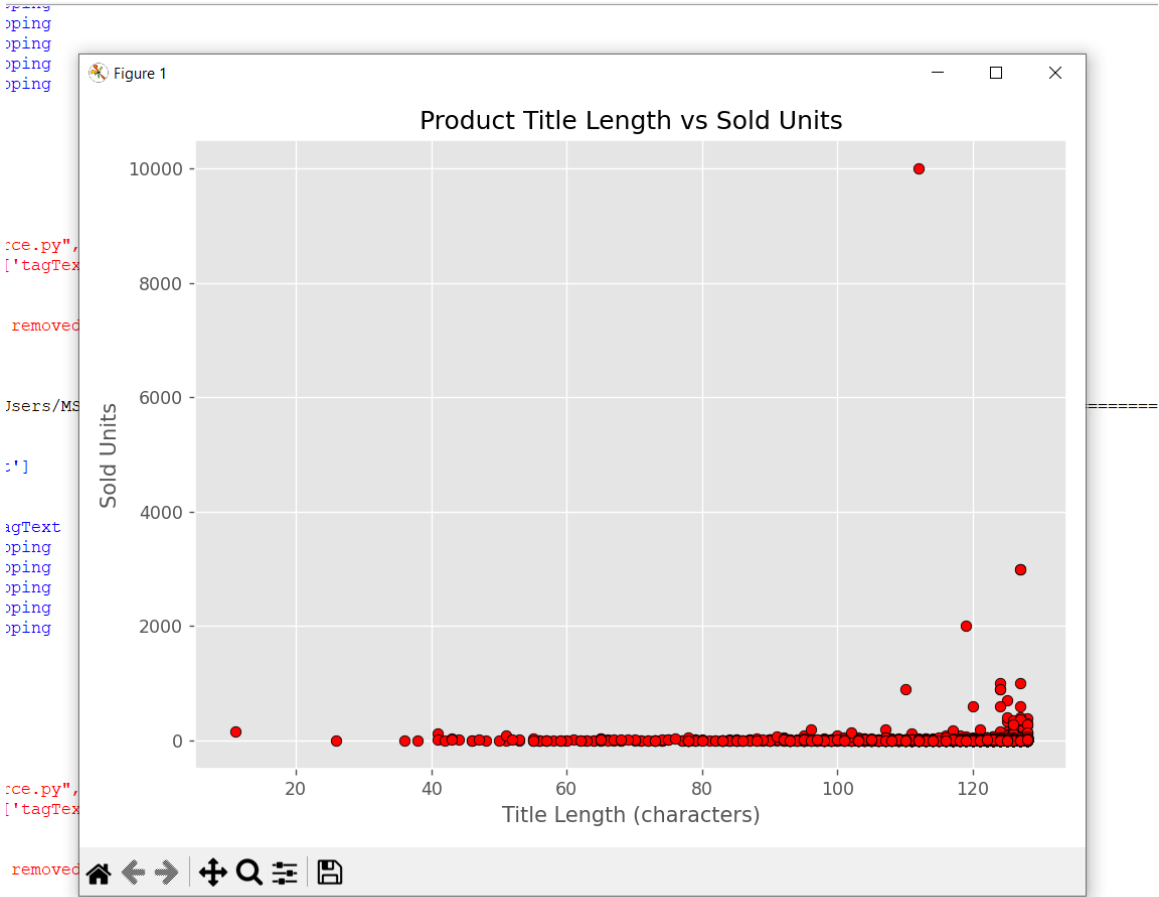
```
.png  
s.png
```



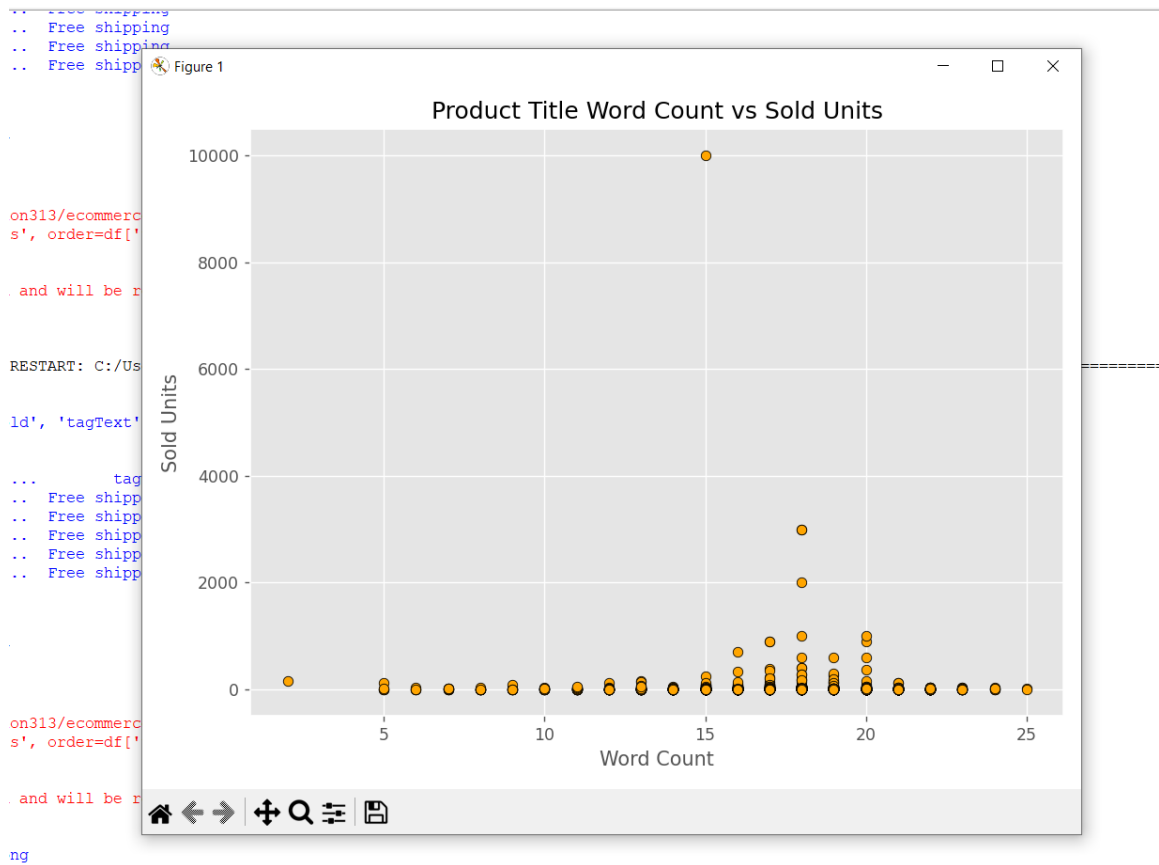
Count of Tag Text Categories



Product Title Length Vs Sold Units



Product Title Word Count Vs Sold Units



5. Conclusion

The analysis revealed that 'Free shipping' products dominate the listings. Price has a varying impact on sales, and some low-priced items still have low sales, suggesting other factors like product quality and demand play a role. Text features like title length and word count can provide additional signals for predictive modeling.