Bookstore Dataset Analysis

By Ajith c

Introduction

This dataset contains information about a bookstore's inventory and sales. Each entry includes details like the book's title, price, and availability, as well as data on sales and customer preferences. This dataset can help us understand popular books, pricing trends, and customer buying patterns. It's useful for analyzing bookstore performance, planning stock, or recommending books to customers.

Overview of Dataset

Rows=1000

Columns=7

Key features

- Title
- Book Category
- Star rating
- Price
- Stock
- Quantity
- Price category

Feature Engineering

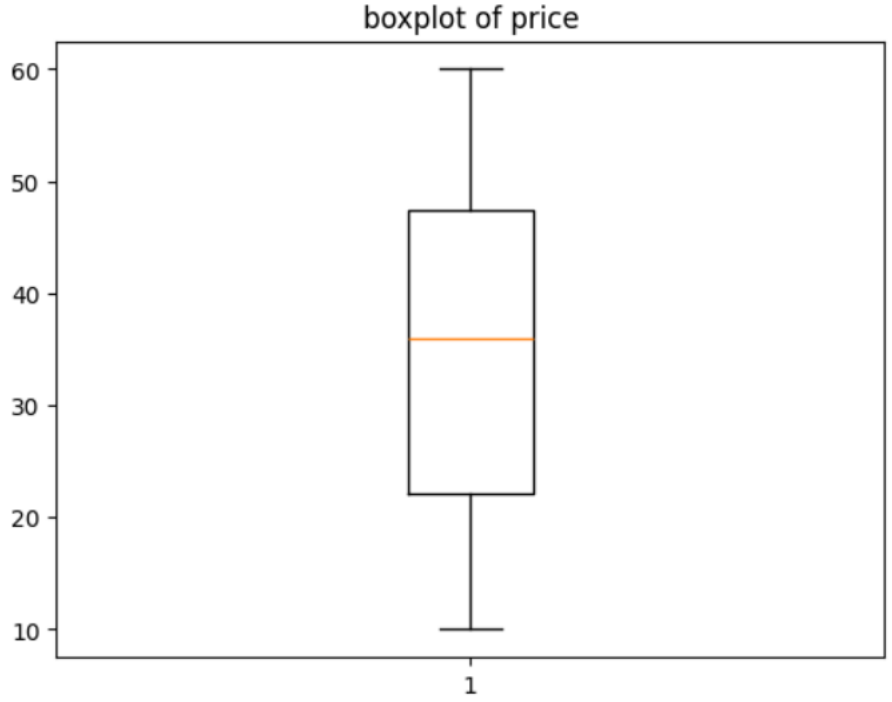
- Added a new column named ‘price category’.

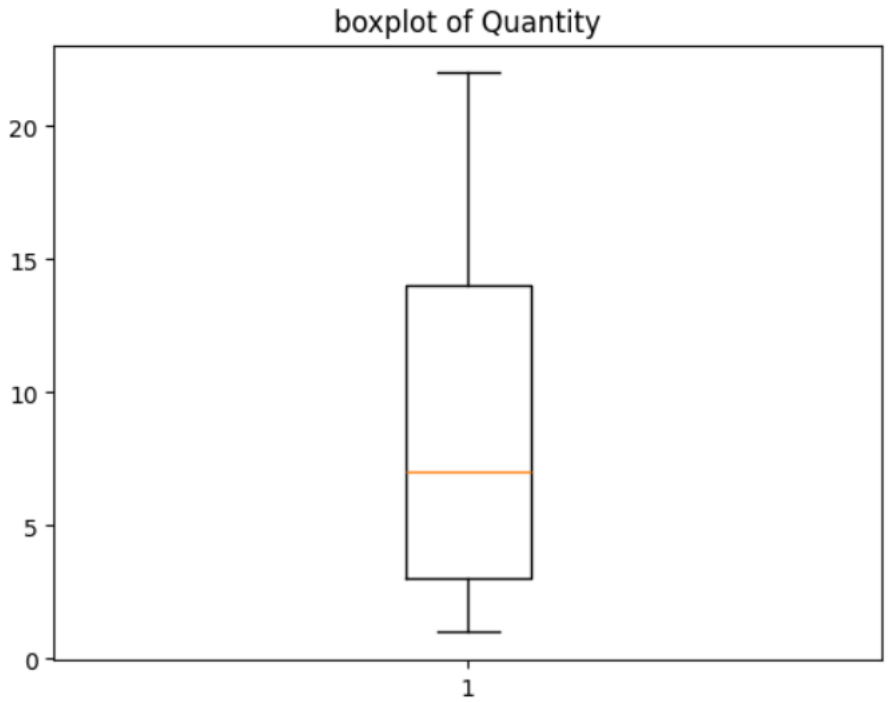
```
def price Category(Price):
if Price<20:
return 'Budget’
elif Price>=20 and Price<40:
return 'Medium’
else:    return "Premium’
```

OUTLIER DETECTION

This visualization helps you quickly understand the spread and skewness of your price data, as well as identify any potential outliers. And there is no outlier found.

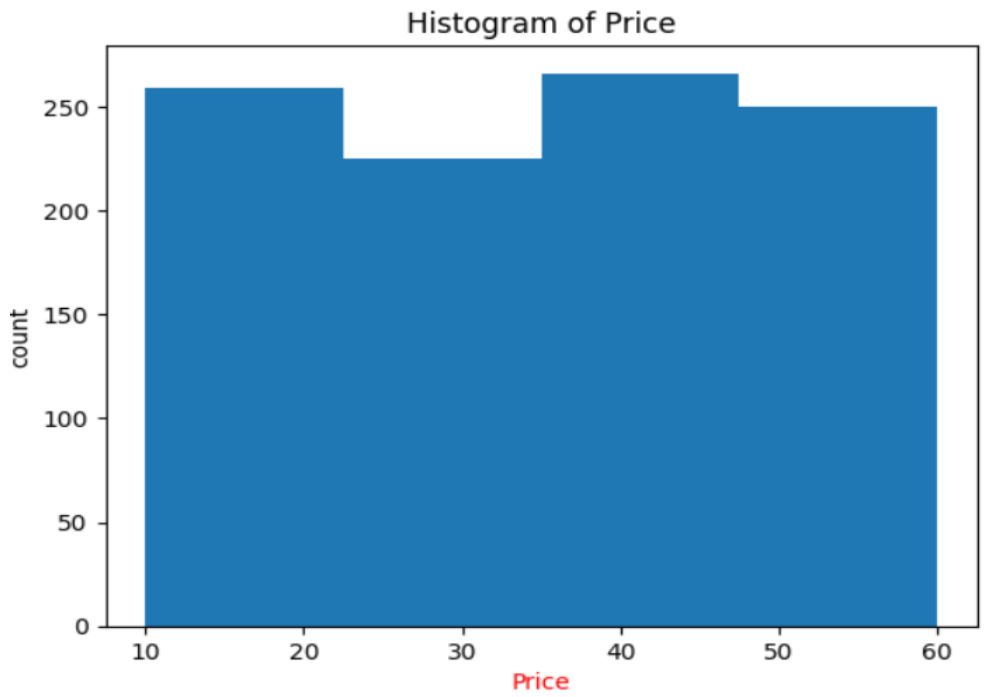
```
‘import matplotlib.pyplot as plt
 plt.boxplot(df['Price’])
plt.title ('boxplot of price’)
 plt.show ‘
```

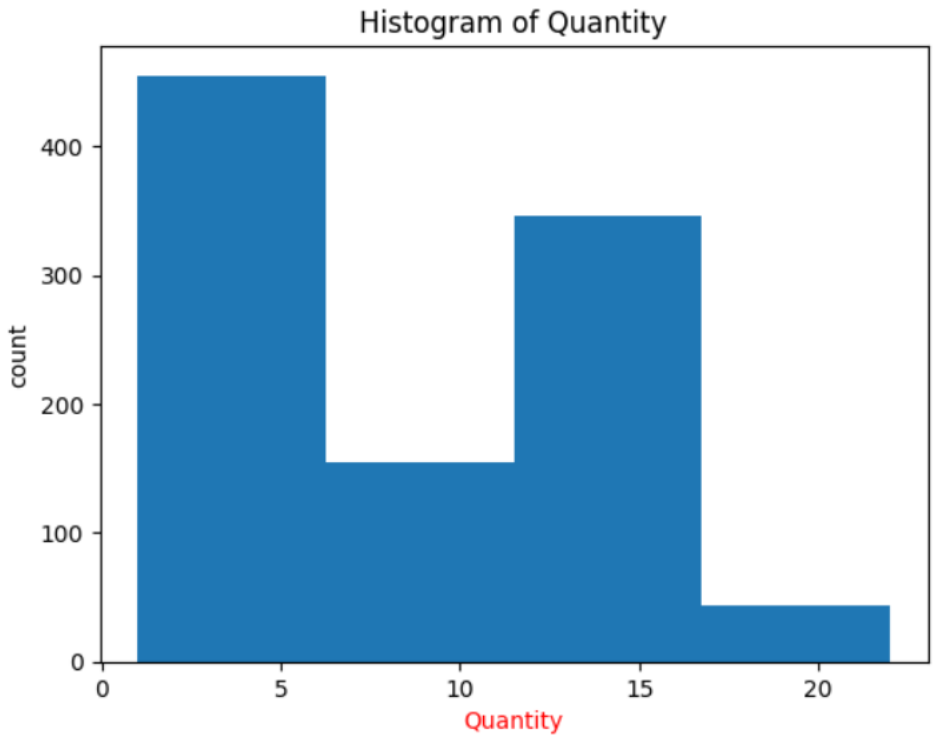




HISTPLOT

This graph represent each column in a Data Frame, checks if the column has numbers, and if it does, it makes a histogram with 4 bins. It also sets the title and x-axis label for each histogram. On the basis of the chart when the stock of books in the store is price between 35-45 .



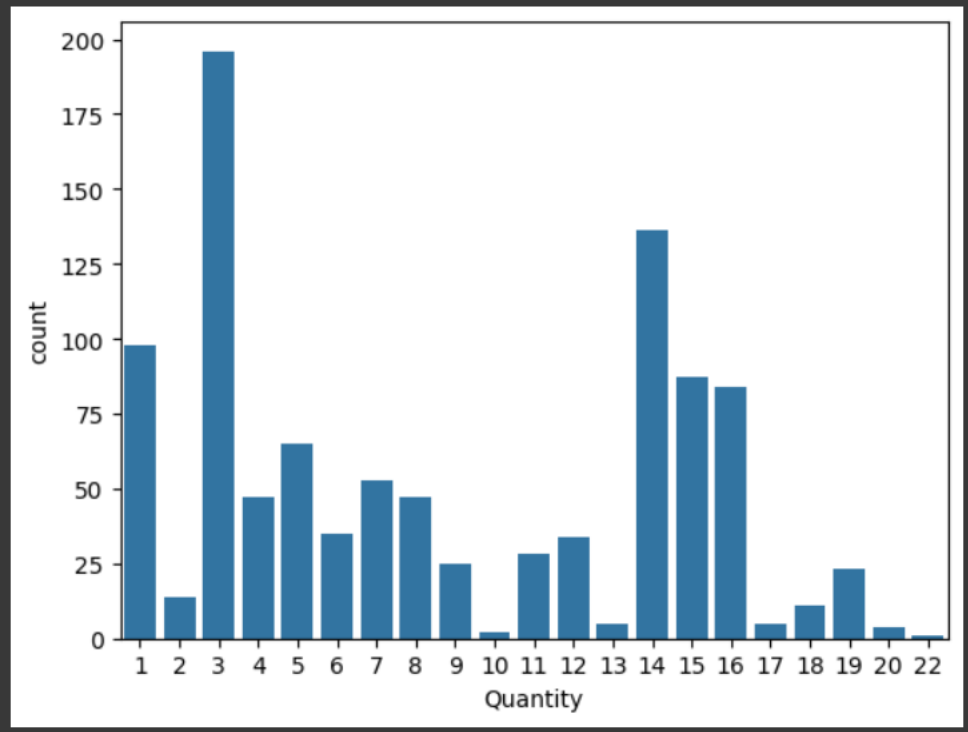


COUNT PLOT

The overall effect of this code is to generate a bar plot that visually represents the frequency distribution of the quantity column in each data frame.

Each bar shows how often each quantity value occurs, making it easier to understand the distribution of your data at a glance.

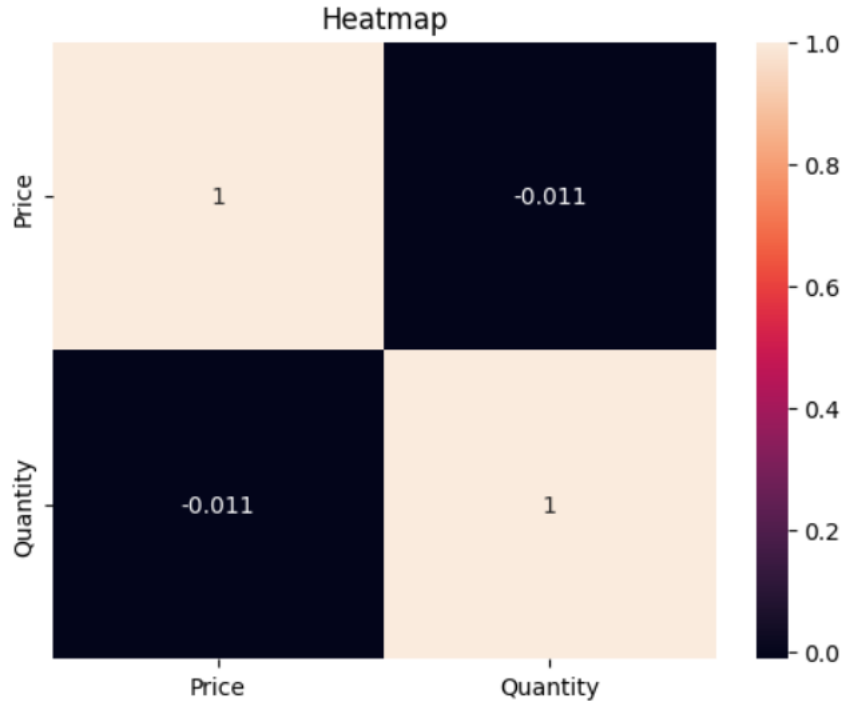
```
#univeriate analysis
import seaborn as sns
sns.countplot(x=df['Quantity'])
plt.show()
```



HEAT MAP

The resulting output is a heatmap that visually represents the correlation matrix of the Data Frame numeric columns. Each cell in the heatmap shows the correlation coefficient between a pair of variables, with colors representing the strength and direction of the correlations.

```
import seaborn as sns
 var1=df.corr(numeric_only=True)
 plt.title('Heatmap’)
 sns.heatmap(var1,annot=True)
 plt.show()
```



SCATTER PLOT

The resulting output is a scatter plot where each dot represents an individual data point. The x-axis shows the price values, and the y-axis shows the quantity values. This plot helps you visually identify any relationships or patterns between the price and quantity data.

```
plt.scatter(x=df['Price'],y=df['Quantity’])
 plt.title('scatter plots of Price and Quantity ‘)
 plt.xlabel('Price') plt.ylabel('Quantity')
 plt.show()
```

LINE PLOT

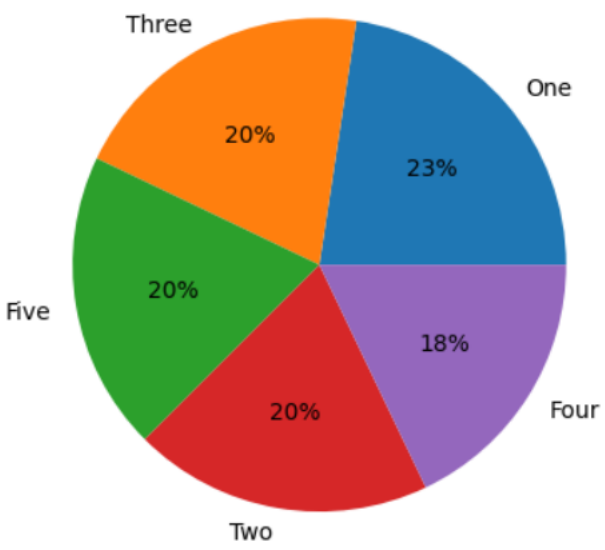
The resulting output is a line plot that shows how the average price changes with different quantities. This visualization helps you understand the relationship between price and quantity.

```
#Line plot
#time series analysis
var1=df[['Quantity','Price']].groupby ('Quantity').mean()
plt.plot(var1.index,var1)
plt.title(' Quantity vs average Price  ')
plt.xlabel('Quantity')
plt.ylabel('Price')
plt.show()
```

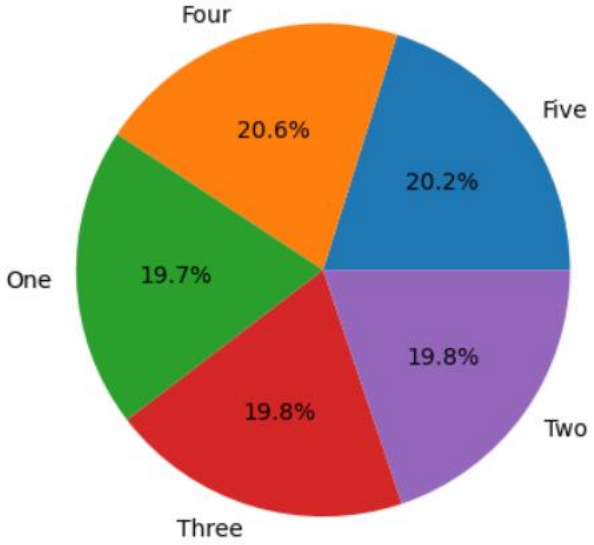
PIE PLOT

Pie plot is a circular graph that shows the relative proportions of different categories in dataset. Each slice of the pie represents a category, and its size corresponds to the category's proportion of the total. which allows you to visualize the distribution of categorical data in a clear and intuitive way.

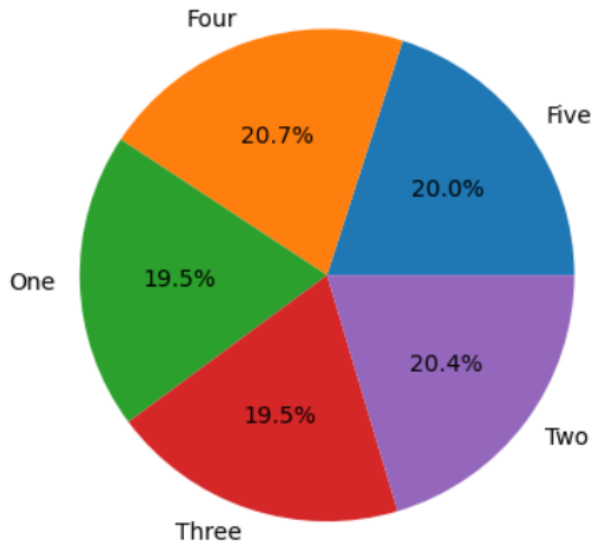
pie plot of Star_rating categorical count



pie plot of quantity and star rating



pie plot of price and star rating



FINDINGS

Analyze the relationship between book categories and their average star ratings to determine which genres are most favored by customers.

Investigate the pricing strategy by comparing book prices across different categories and their respective price categories to optimize pricing.

Evaluate stock levels in relation to book popularity (star ratings) and sales quantity to improve inventory management.

Assess the impact of star ratings on sales performance by comparing the quantity sold of books with varying star ratings.

CONCLUSION

The bookstore's inventory, pricing, and popularity metrics (like star ratings and quantities sold) can reveal valuable insights into customer preferences and sales performance. Analyzing these data points helps identify which book categories are most popular, how pricing strategies affect sales, and how stock levels correspond to demand. This information can guide better inventory management, marketing strategies, and pricing decisions to enhance overall store performance.