Data Visualization Introduction

Prerequisite Skills:

- Lists
- DataFrames
- Pandas and Matplotlib

Objectives:

- Load provided data in Python.
- Learn how to create figures in Python based on real world data.
- Get inspiration for figures from https://matplotlib.org/stable/gallery/index.html

Sample Code and Figures:

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
df = pd.read_csv('your_dataset.csv')

# Visualization 1: Bar chart for top genres
top_genres = df['genre'].value_counts().nlargest(10)
plt.figure(figsize=(10, 6)) # optional - sets figure size
top_genres.plot(kind='bar', color='skyblue') # create bar plot
plt.title('Top 10 Book Genres on Amazon') # include title
plt.xlabel('Genre') # include x-axis label
plt.ylabel('Number of Books') # include y-axis label
plt.xticks(rotation=45) # rotate x-tick labels by 45 degrees for easier readability
plt.show() # show the plot!

# Visualization 2: Scatter plot for price vs. rating
plt.figure(figsize=(10, 6))
plt.scatter(df['price'], df['rating'], color='orange', alpha=0.5)
plt.title('Price vs. Rating of Top 100 Trending Books on Amazon')
plt.xlabel('Price ($)')
plt.ylabel('Rating')
plt.show()

# Visualization 3: Box plot for distribution of book prices
plt.figure(figsize=(10, 6))
plt.boxplot(df['price'], vert=False)
plt.title('Distribution of Book Prices on Amazon')
plt.xlabel('Price ($)')
plt.show()

# Visualization 4: Line chart for the number of books published each year
books_per_year = df['year_of_publication'].value_counts().sort_index()
plt.figure(figsize=(10, 6))
books_per_year.plot(kind='line', marker='o', color='green')
plt.title('Number of Books Published Each Year on Amazon')
plt.xlabel('Year of Publication')
plt.ylabel('Number of Books')
plt.show()
```


Price vs. Rating of Top 100 Trending Books on Amazon
Rating
Price ($)
5.0
4.8
4.6
4.4
4.2
10
20
30
40
50


Number of Books Published Each Year on Amazon
Number of Books
Year of Publication
25
20
15
10
5
0
1950
1960
1970
1980
1990
2000
2010
2020