Assignment 3

January 30, 2025

[48]:
```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(1.2.0)
Requirement already satisfied: cycler>=0.10 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(1.4.4)
Requirement already satisfied: numpy>=1.21 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(1.26.4)
Requirement already satisfied: packaging>=20.0 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(23.2)
Requirement already satisfied: pillow>=8 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from matplotlib)
(2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[49]:
```
pip install seaborn
```

```
Requirement already satisfied: seaborn in
```

```
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from seaborn)
(1.26.4)
Requirement already satisfied: pandas>=1.2 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from seaborn)
(2.2.2)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from seaborn)
(3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.2.0)
Requirement already satisfied: cycler>=0.10 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (23.2)
Requirement already satisfied: pillow>=8 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
pandas>=1.2->seaborn) (2024.1)
Requirement already satisfied: tzdata>=2022.7 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from
pandas>=1.2->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in
/Users/delta/Downloads/anaconda3/lib/python3.12/site-packages (from python-
dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

[50]:
```
import matplotlib.pyplot as plt
import numpy as np
```

```
import pandas as pd
import seaborn as sns
```

[51]:
```
df=pd.read_csv("Online Sales Data.csv")
```

[52]:
```
df.tail(-60)
```

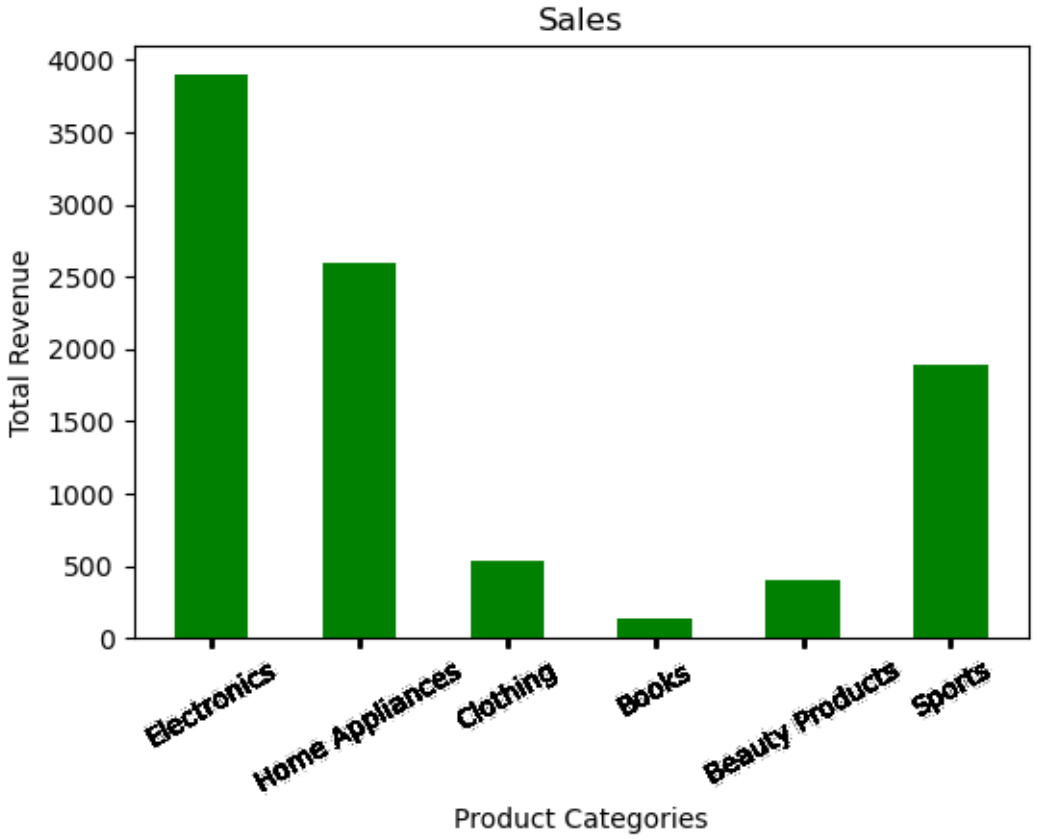
[52]:
```
     Transaction ID        Date Product Category  \
60            10061  2024-03-01      Electronics
61            10062  2024-03-02  Home Appliances
62            10063  2024-03-03         Clothing
63            10064  2024-03-04            Books
64            10065  2024-03-05  Beauty Products
..              …         …              …
235           10236  2024-08-23  Home Appliances
236           10237  2024-08-24         Clothing
237           10238  2024-08-25            Books
238           10239  2024-08-26  Beauty Products
239           10240  2024-08-27           Sports

                                      Product Name  Units Sold  Unit Price  \
60                                 Nintendo Switch           3      299.99
61                              Philips Airfryer XXL           2      199.99
62                          Hanes ComfortSoft T-Shirt          10        9.99
63              Where the Crawdads Sing by Delia Owens           4       18.99
64                           Lancome La Vie Est Belle           1      102.00
..                                                 …         …          …
235  Nespresso Vertuo Next Coffee and Espresso Maker           1      159.99
236                          Nike Air Force 1 Sneakers           3       90.00
237            The Handmaid's Tale by Margaret Atwood           3       10.99
238              Sunday Riley Luna Sleeping Night Oil           1       55.00
239                         Yeti Rambler 20 oz Tumbler           2       29.99

     Total Revenue         Region Payment Method
60          899.97  North America    Credit Card
61          399.98         Europe         PayPal
62           99.90           Asia     Debit Card
63           75.96  North America    Credit Card
64          102.00         Europe         PayPal
..             …              …              …
235         159.99         Europe         PayPal
236         270.00           Asia     Debit Card
237          32.97  North America    Credit Card
238          55.00         Europe         PayPal
239          59.98           Asia    Credit Card

[180 rows x 9 columns]
```

[53]:
```
df.describe()
```

[53]:

	Transaction ID	Units Sold	Unit Price	Total Revenue
count	240.00000	240.000000	240.000000	240.000000
mean	10120.50000	2.158333	236.395583	335.699375
std	69.42622	1.322454	429.446695	485.804469
min	10001.00000	1.000000	6.500000	6.500000
25%	10060.75000	1.000000	29.500000	62.965000
50%	10120.50000	2.000000	89.990000	179.970000
75%	10180.25000	3.000000	249.990000	399.225000
max	10240.00000	10.000000	3899.990000	3899.990000

[54]:
```
plt.figure(figsize=(6,4))
plt.bar(df['Product Category'],df['Total Revenue'],color="green",width=0.5)
plt.title("Sales")
plt.xlabel("Product Categories")
plt.ylabel("Total Revenue")
plt.xticks(df['Product Category'], rotation=30)
plt.show()
```

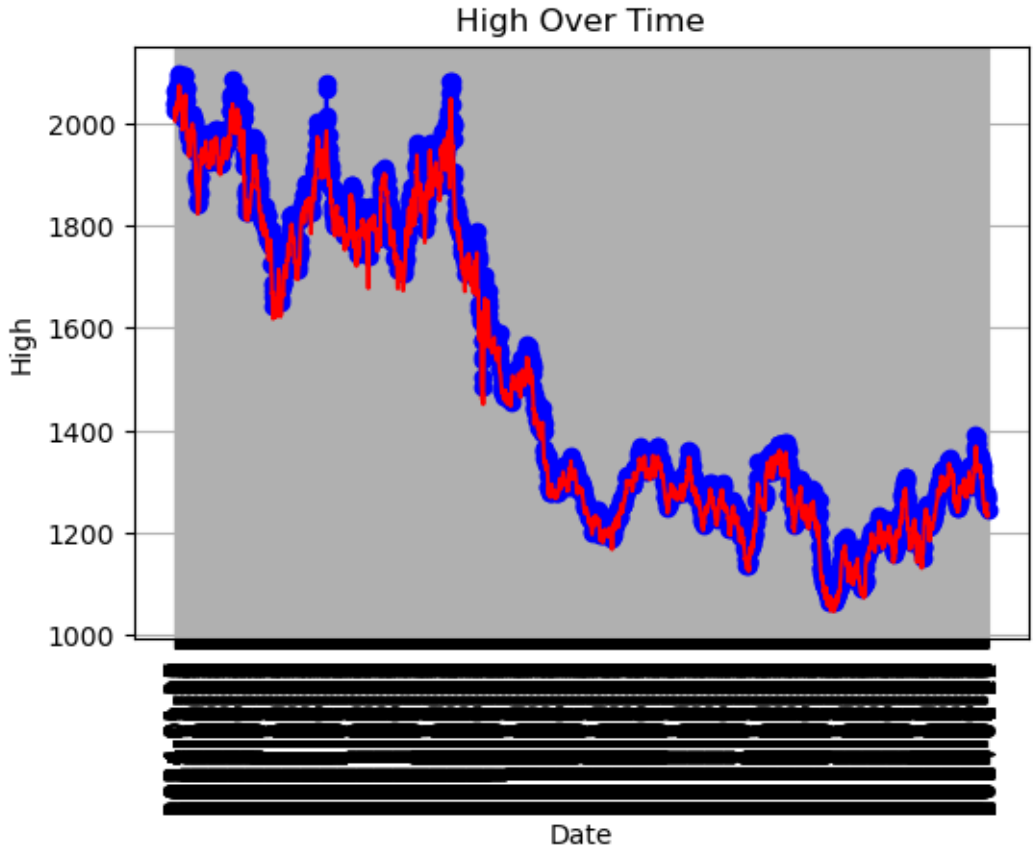


[55]:
```
df=pd.read_csv("goldstock v1.csv")
df.head(2)
```

[55]:

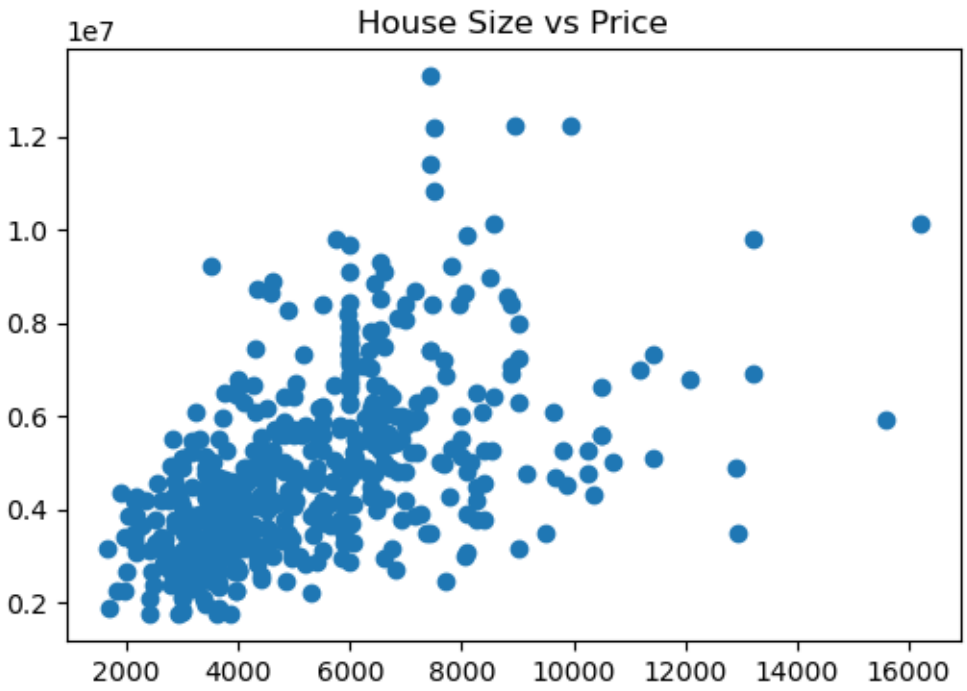
	Unnamed: 0	Date	Close	Volume	Open	High	Low
0	0	2024-01-19	2029.3	166078.0	2027.4	2041.9	2022.2
1	1	2024-01-18	2021.6	167013.0	2009.1	2025.6	2007.7

[56]:
```
plt.figure(figsize=(6, 4))
plt.plot(df['Date'], df['High'], marker='o', color='b')
plt.plot(df['Date'], df['Low'],color='r')
plt.grid(True)
plt.title('High Over Time')
plt.xlabel('Date')
plt.ylabel('High')
plt.xticks(rotation=90)
plt.show()
```



[61]:
```
df=pd.read_csv("Housing.csv")
```

[63]:
```
plt.figure(figsize=(6,4))
plt.scatter(df['area'],df['price'])
plt.title("House Size vs Price")
plt.show()
```



[66]:
```
df=pd.read_csv("heart_disease.csv")
df.head(2)
```

[66]:

	Age	Gender	Blood Pressure	Cholesterol Level	Exercise Habits	Smoking	\
0	56.0	Male	153.0	155.0	High	Yes	
1	69.0	Female	146.0	286.0	High	No	

	Family Heart Disease	Diabetes	BMI	High Blood Pressure	…	\
0	Yes	No	24.991591	Yes	…	
1	Yes	Yes	25.221799	No	…	

	High LDL Cholesterol	Alcohol Consumption	Stress Level	Sleep Hours	\
0	No	High	Medium	7.633228	
1	No	Medium	High	8.744034	

	Sugar Consumption	Triglyceride Level	Fasting Blood Sugar	CRP Level	\
0	Medium	342.0	NaN	12.969246	
1	Medium	133.0	157.0	9.355389	

```
   Homocysteine Level  Heart Disease Status
0           12.387250                    No
1           19.298875                    No

[2 rows x 21 columns]
```

[82]:
```
fig,ax=plt.subplots(figsize=(6,4))
sns.heatmap(df.corr(numeric_only=True),annot=True,cmap="Greens")
plt.show()
```