diamondBook

April 2, 2021

[1]:

```python
#Import necessary packages
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

[2]:

```python
#Reading in diamond data
diamonds = pd.read_csv("diamonds.csv")
uniqueCuts = diamonds["cut"].unique()
print(uniqueCuts)
```

```
['Ideal' 'Premium' 'Good' 'Very Good' 'Fair']
```

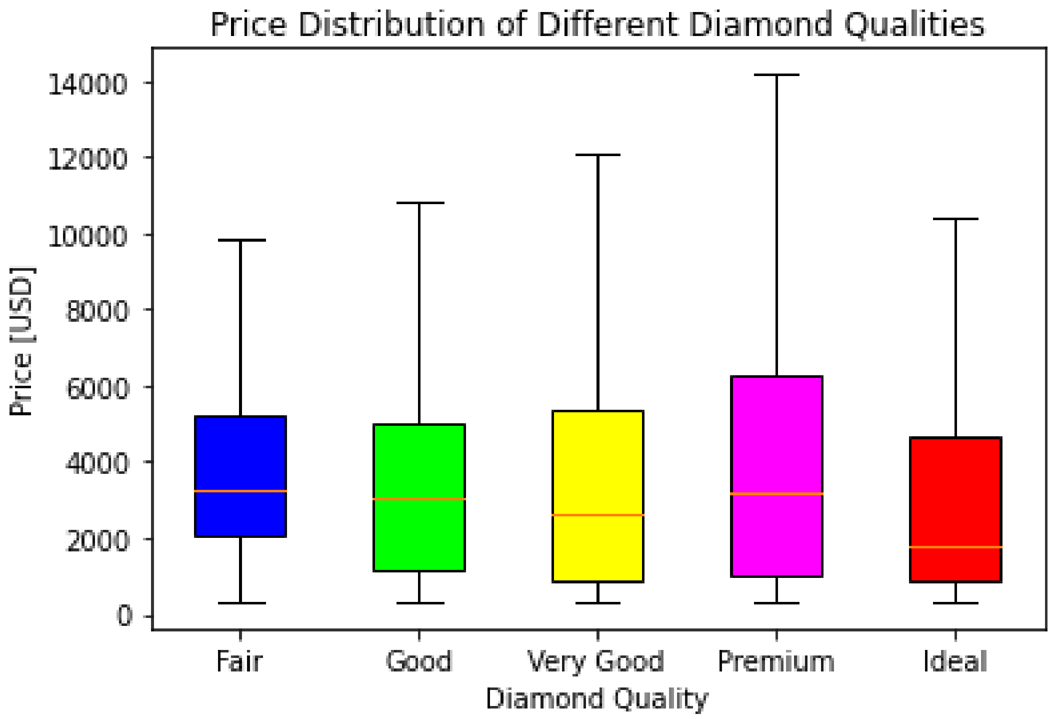
[3]:

```python
#Cleaning data to have diamond info per cut
fair = diamonds[diamonds["cut"] == "Fair"].to_numpy()
fairP = fair[:,7]
good = diamonds[diamonds["cut"] == "Good"].to_numpy()
goodP = good[:,7]
veryg = diamonds[diamonds["cut"] == "Very Good"].to_numpy()
verygP = veryg[:,7]
prem = diamonds[diamonds["cut"] == "Premium"].to_numpy()
premP = prem[:,7]
ideal = diamonds[diamonds["cut"] == "Ideal"].to_numpy()
idealP = ideal[:,7]
```

[4]:

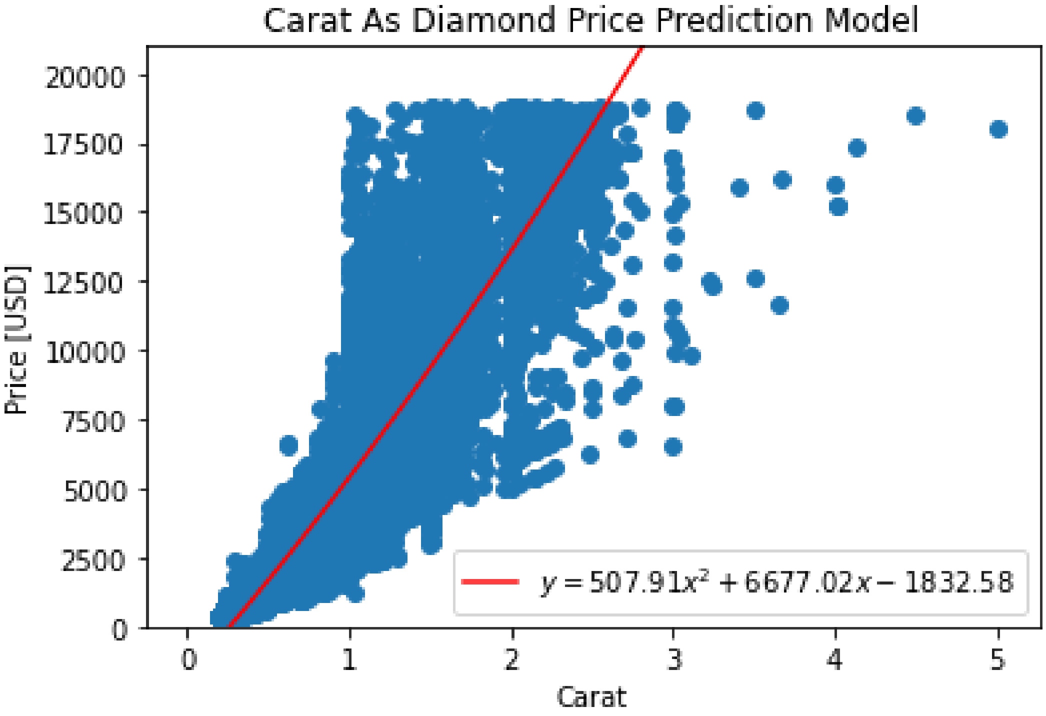
```python
#Boxplot Generation
boxData = [fairP,goodP,verygP,premP,idealP]

fig, axis = plt.subplots()

BoxPlot = axis.boxplot(boxData, patch_artist = True, showfliers = False)
fig.patch.set_alpha(1.0)
plt.xticks([1,2,3,4,5],["Fair","Good","Very Good","Premium","Ideal"])
plt.ylabel("Price [USD]")
plt.xlabel("Diamond Quality")
plt.title("Price Distribution of Different Diamond Qualities")

colors = ['#0000FF', '#00FF00',
          '#FFFF00', '#FF00FF','#FF0000']
```

```
for patch, color in zip(BoxPlot['boxes'], colors):
    patch.set_facecolor(color)

plt.savefig('boxPlot.png',dpi = 300)
plt.show()
```



[12]:
```
#Linear Regression of carat as a price predictor
carat = diamonds["carat"].to_numpy()
price = diamonds["price"].to_numpy()
print(carat)
print(price)

fig, axis = plt.subplots()
fig.patch.set_alpha(1.0)

caratFit = np.polyfit(carat,price,2)
print(caratFit)

xvals = np.linspace(0,3,301)
axis.plot(xvals, caratFit[0]*xvals**2 + caratFit[1]*xvals + caratFit[2], color␣
 ↪= 'red')
plt.scatter(carat,price)
plt.xlabel("Carat")
plt.ylabel("Price [USD]")
plt.ylim([0,21000])
plt.title("Carat As Diamond Price Prediction Model")
plt.legend(["$y = 507.91x^2 + 6677.02x - 1832.58$"])
plt.savefig('caratModel.png',dpi = 300)
plt.show()
```

```
[0.23 0.21 0.23 … 0.7  0.86 0.75]
[ 326  326  327 … 2757 2757 2757]
[  507.91326733  6677.02734194 -1832.57737161]
```


Carat As Diamond Price Prediction Model
Price [USD]
$y = 507.91x^2 + 6677.02x - 1832.58$
Carat