In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv('gold_monthly.csv')
df
```

Out[2]:

	Date	Price
0	1950-01	34.730
1	1950-02	34.730
2	1950-03	34.730
3	1950-04	34.730
4	1950-05	34.730
...	...	...
842	2020-03	1593.764
843	2020-04	1680.030
844	2020-05	1715.697
845	2020-06	1734.032
846	2020-07	1840.807

847 rows × 2 columns

In [3]:
```
df.shape
```

Out[3]:
```
(847, 2)
```

In [4]:
```
df.head(10)
```

Out[4]:

	Date	Price
0	1950-01	34.73
1	1950-02	34.73
2	1950-03	34.73
3	1950-04	34.73
4	1950-05	34.73
5	1950-06	34.73
6	1950-07	34.73
7	1950-08	34.73
8	1950-09	34.73
9	1950-10	34.73

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 847 entries, 0 to 846
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Date    847 non-null    object 
 1   Price   847 non-null    float64
dtypes: float64(1), object(1)
memory usage: 13.4+ KB
```

In [6]:
```
#Checking for null values

pd.isnull(df).sum()
```

Out[6]:
```
Date     0
Price    0
dtype: int64
```

In [7]:
```
#New DataFrames with monthly dates as index

date_range = pd.date_range(start = '1/1/1950' , end = '8/1/2020', freq = 'M')
df['month'] = date_range
df.drop('Date', axis = 1, inplace = True)
df = df.set_index('month')
df.head()
```

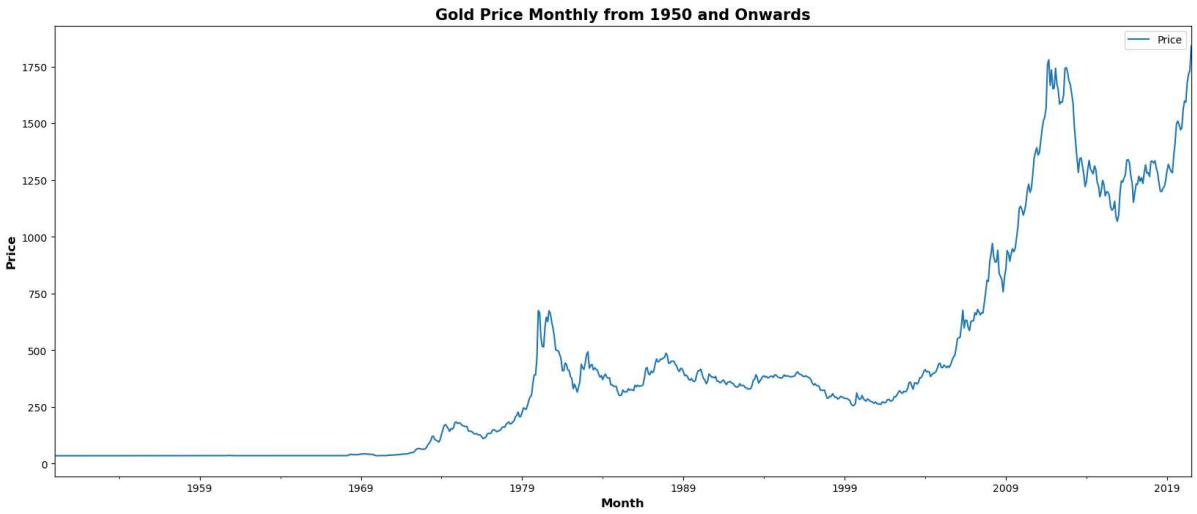
Out[7]:

month	Price
1950-01-31	34.73
1950-02-28	34.73
1950-03-31	34.73
1950-04-30	34.73
1950-05-31	34.73

In [8]:
```
#Plot gold prices over time
```

In [16]:
```
plt.figure(figsize=(20,8))
df.plot(figsize = (20,8))
plt.title("Gold Price Monthly from 1950 and Onwards", fontweight='bold', fontsize=1
plt.xlabel("Month" , fontweight='bold', fontsize=12)
plt.ylabel("Price" , fontweight='bold', fontsize=12)
plt.show()
```

```
<Figure size 2000x800 with 0 Axes>
```



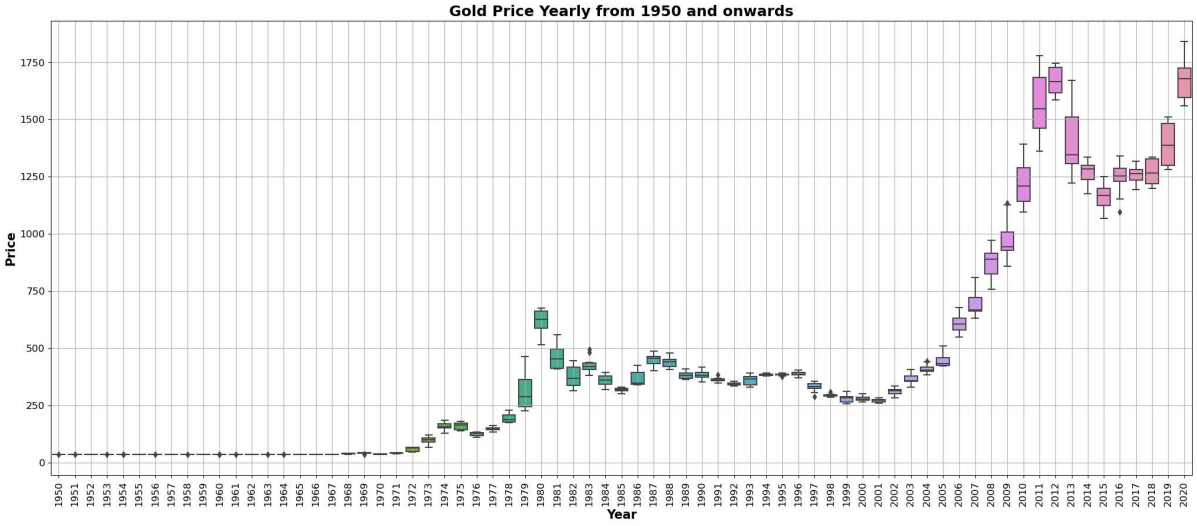
In [29]:
```
#Create a boxplot of the gold prices by year

plt.figure(figsize = (25, 10))

df['year'] = df.index.year

sns.boxplot(x= df.year, y=df.values[:,0] )

plt.title('Gold Price Yearly from 1950 and onwards', fontweight='bold', fontsize=18
plt.xlabel('Year', fontweight='bold', fontsize=15)
plt.ylabel('Price', fontweight='bold', fontsize=15)
plt.xticks(rotation=90, fontsize=12)
plt.yticks(fontsize=12)
plt.grid()
plt.show()
```

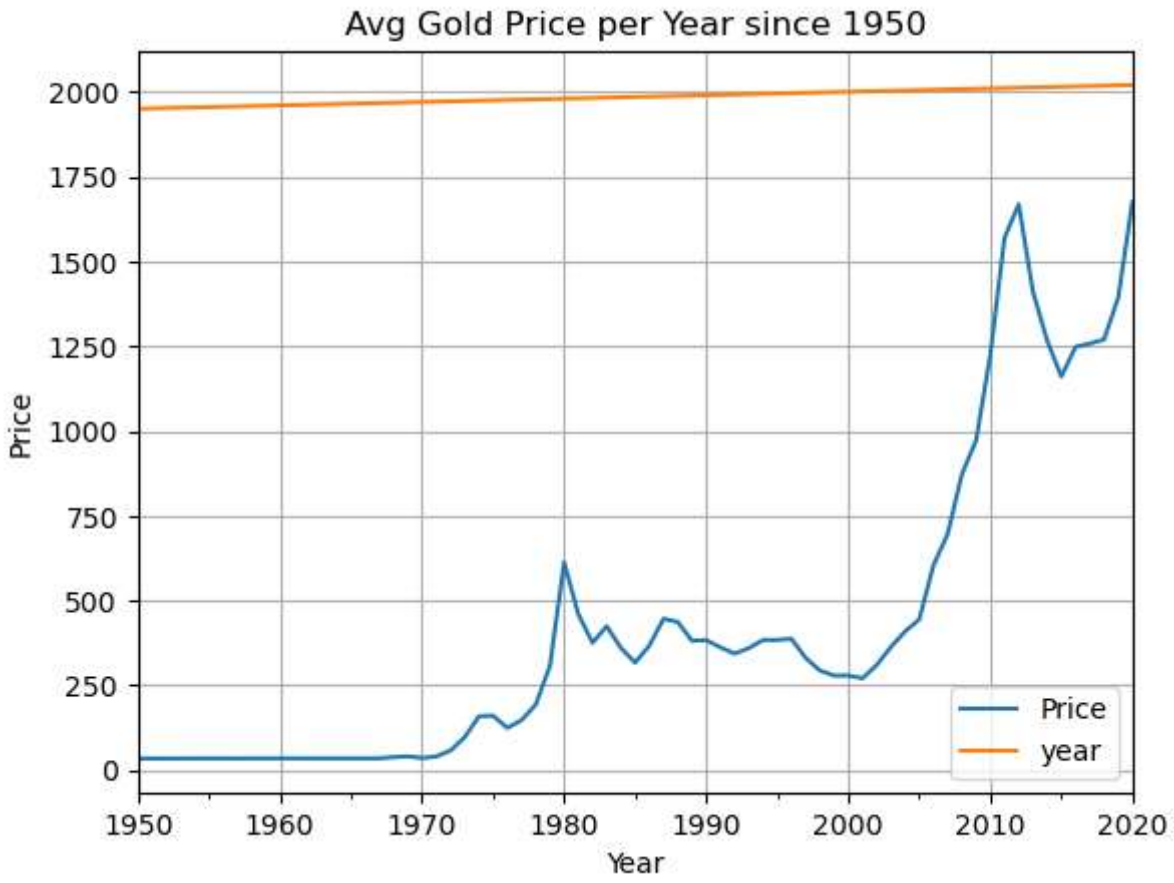


In [30]:
```
#Summary statistics of the gold prices

print("Summary statistics of the gold prices:\n", df.describe())
```

```
Summary statistics of the gold prices:
              Price         year
count   847.000000   847.000000
mean    416.556906  1984.793388
std     453.665313    20.388625
min      34.490000  1950.000000
25%      35.190000  1967.000000
50%     319.622000  1985.000000
75%     447.029000  2002.000000
max    1840.807000  2020.000000
```

In [31]:
```
df_yearly_sum = df.resample('A').mean()
df_yearly_sum.plot()
plt.title('Avg Gold Price per Year since 1950')
plt.xlabel('Year')
plt.ylabel('Price')
plt.grid()
```



In [35]:
```
df_quarterly_sum = df.resample('Q').mean()
df_quarterly_sum.plot()
plt.title('Avg Gold Price per Quarter since 1950')
plt.xlabel('Quarter')
plt.ylabel('Price')
plt.grid()
```

In [37]:
```
#Summary Statistics of Gold Prices:

print("Summary Statistics of Gold Prices:\n\n", df['Price'].describe())
```

```
Summary Statistics of Gold Prices:

 count     847.000000
mean      416.556906
std       453.665313
min        34.490000
25%        35.190000
50%       319.622000
75%       447.029000
max      1840.807000
Name: Price, dtype: float64
```

In [ ]: