

Mục tiêu

Sử dụng thư viện Seaborn để vẽ một số biểu đồ trên bộ dữ liệu Food Price in Turkey.

Bộ dữ liệu Food Price in Turkey mô tả giá cả các loại thức ăn tại mỗi thời điểm ở một số thành phố lớn Thổ Nhĩ Kỳ mà bạn đã được làm quen từ các buổi trước. Trong buổi thực hành này, chúng ta sẽ sử dụng thư viện Seaborn để vẽ một số loại biểu đồ nâng cao trên bộ dữ liệu này.

Khai báo thư viện cần dùng

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

Đọc dữ liệu

```
df = pd.read_csv("FoodPrice_in_Turkey.csv")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7381 entries, 0 to 7380
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Place       7381 non-null   object  
 1   ProductId   7381 non-null   int64   
 2   ProductName 7381 non-null   object  
 3   UmId        7381 non-null   int64   
 4   UmName      7381 non-null   object  
 5   Month       7381 non-null   int64   
 6   Year        7381 non-null   int64   
 7   Price       7381 non-null   float64  
dtypes: float64(1), int64(4), object(3)
memory usage: 461.4+ KB
```

```
df.describe()
```

	ProductId	UmId	Month	Year	Price
count	7381.000000	7381.000000	7381.000000	7381.000000	7381.000000
mean	254.509687	15.252811	6.479881	2017.458474	15.738709
std	141.008599	25.276665	3.530361	1.499256	22.418284
min	52.000000	5.000000	1.000000	2013.000000	0.253900
25%	114.000000	5.000000	3.000000	2017.000000	3.223600
50%	239.000000	5.000000	6.000000	2018.000000	5.909600
75%	388.000000	5.000000	10.000000	2019.000000	20.441900
max	502.000000	102.000000	12.000000	2019.000000	159.445800

```
df.head()
```

	Place	ProductId	ProductName	UmId	UmName	Month	Year	Price
0	National Average	52	Rice - Retail	5	KG	5	2013	4.4920
1	National Average	52	Rice - Retail	5	KG	6	2013	4.5786
2	National Average	52	Rice - Retail	5	KG	11	2013	4.7865
3	National Average	52	Rice - Retail	5	KG	12	2013	5.1337
4	National Average	52	Rice - Retail	5	KG	1	2014	5.5099

Loại bỏ dữ liệu bị khuyết

```
df = df.dropna()
```

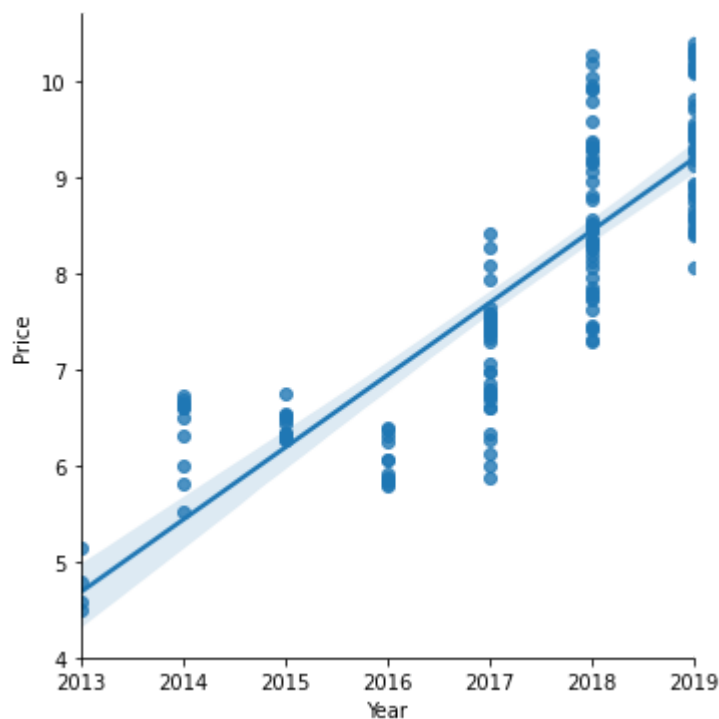
Vẽ biểu đồ xu hướng

Lọc dữ liệu sản phẩm gạo, vẽ biểu đồ xu hướng qua các năm

```
rice_df = df[df["ProductId"] == 52]
```

```
sns.lmplot(x="Year", y="Price", data = rice_df)
```

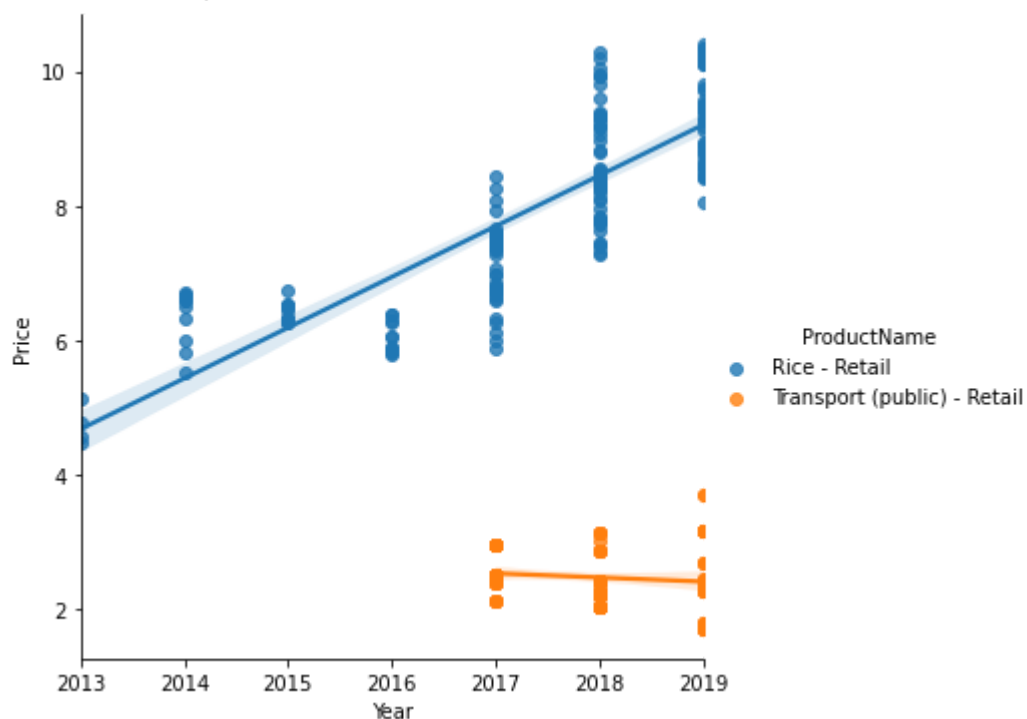
<seaborn.axisgrid.FacetGrid at 0x7fb0bde16f50>



Có thể nhận thấy rằng giá gạo tăng dần qua các năm

```
trans_df = df[(df["ProductName"] == "Transport (public) - Retail") | (df["ProductName"] == "Rice - Retail")]  
sns.lmplot(x="Year", y="Price", hue="ProductName", data = trans_df)
```

<seaborn.axisgrid.FacetGrid at 0x7fb0bde16310>



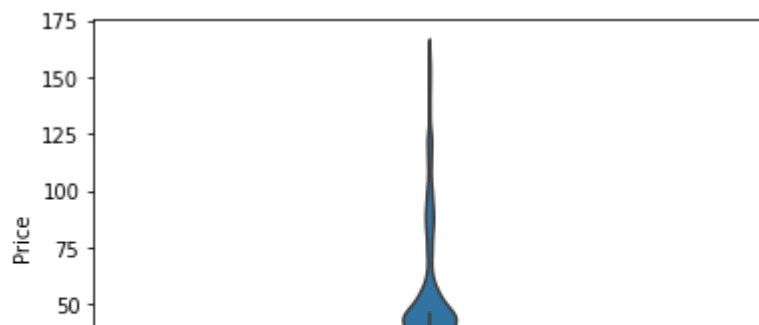
có thể thấy rằng dù chi phí vận chuyển có xu hướng giảm, tuy nhiên giá gạo vẫn tăng khá mạnh, chứng tỏ không có sự liên quan giữa 2 thuộc tính này

Vẽ biểu đồ phân bố

Biểu đồ phân bố cho giá sản phẩm

```
sns.violinplot(y = "Price", data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb0c770f850>
```

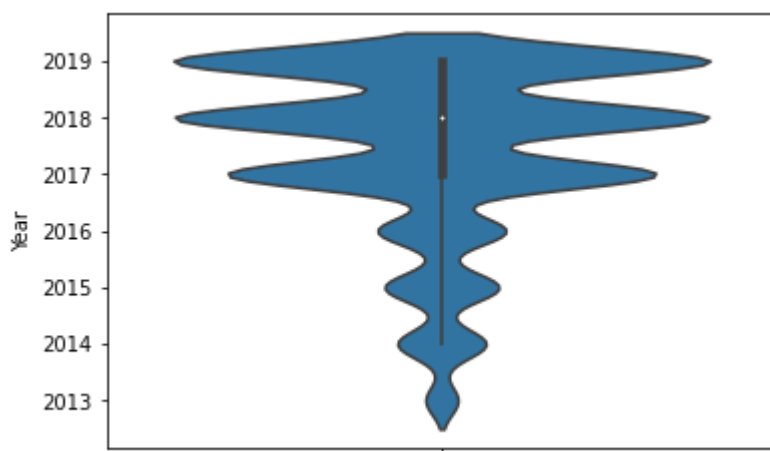


Có thể nhận thấy dữ liệu phân bố không được đều, nhiều sản phẩm giá thấp và ít sản phẩm giá cao.

Biểu đồ phân bố cho năm của sản phẩm

```
sns.violinplot(y = "Year", data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb0c76e4350>
```



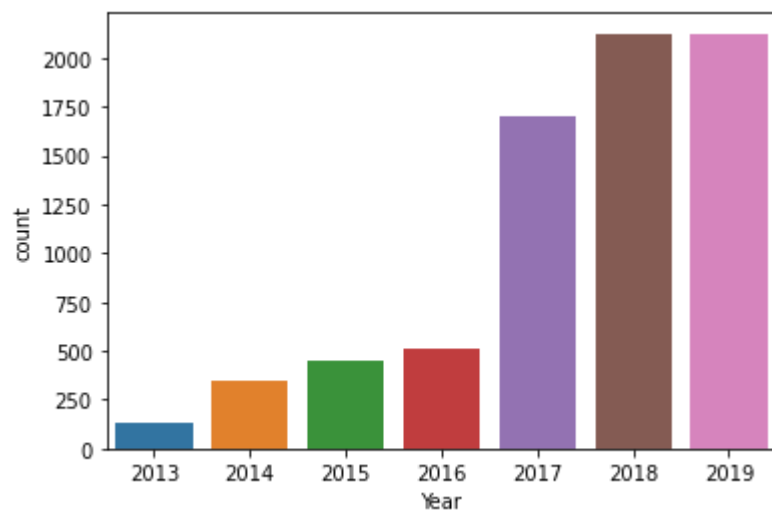
Có thể thấy rằng sản phẩm chủ yếu thuộc những năm gần đây, những năm càng xa thì số lượng càng ít

Vẽ biểu đồ tần số

vẽ biểu đồ tần số cho các sản phẩm theo năm

```
sns.countplot(x = "Year", data = df)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fb0c73ca850>

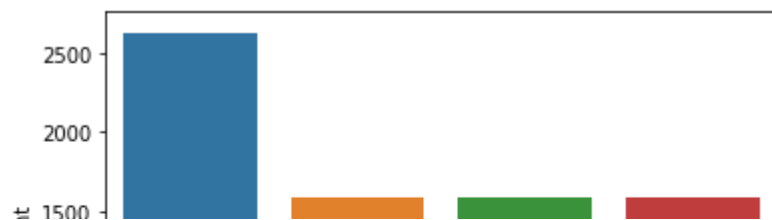


Tương tự như đã nhận xét ở phần trước, những năm càng xa thì càng ít sản phẩm

Thống kê sản phẩm theo địa điểm

```
sns.countplot(x = "Place", data = df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb0c71cb850>
```

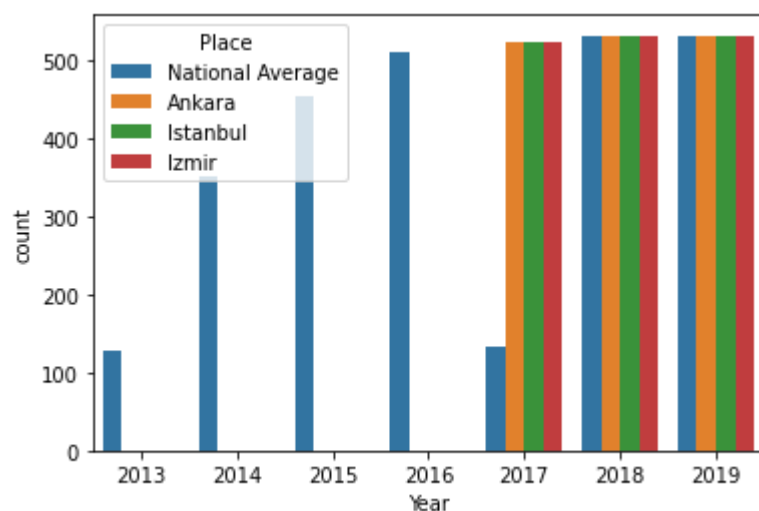


Hiển thị quan hệ giữa số lượng theo năm được nhóm theo địa điểm



```
sns.countplot(x = "Year", hue = "Place", data = df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb0c71a6410>
```



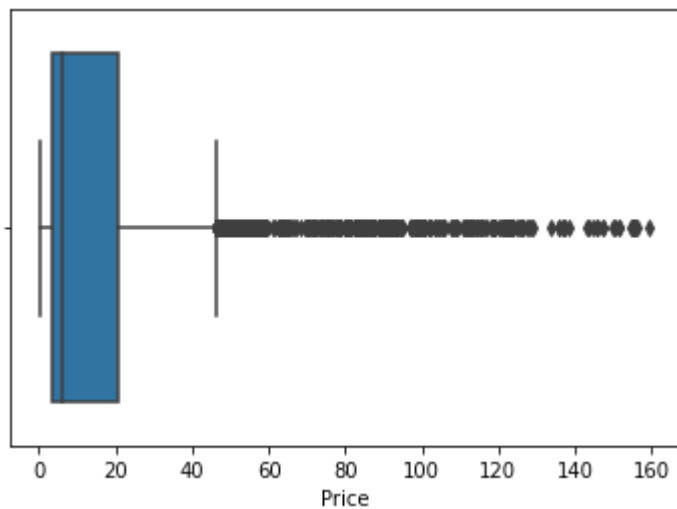
Có thể thấy rằng các năm 2013-2016 chỉ có dữ liệu cho sản phẩm ở National Average

Vẽ biểu đồ box plot

Vẽ biểu đồ box plot cho thuộc tính giá sản phẩm

```
sns.boxplot(x=df["Price"])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb0c2759710>
```



Tương tự như trên đã nhận xét, giá sản phẩm phân bố rất không đồng đều, tập trung ở giá thấp

Vẽ biểu đồ box plot giá sản phẩm phân bố theo năm

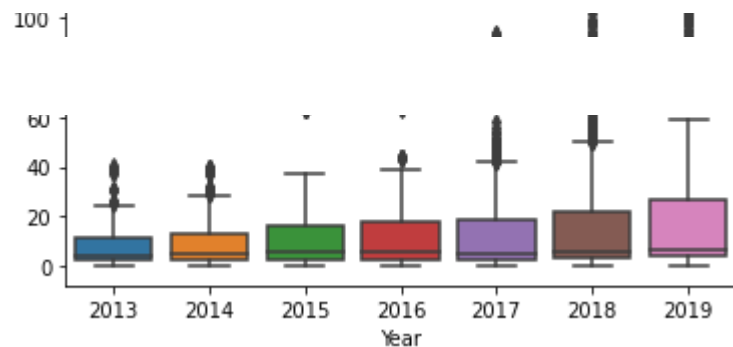
```
sns.boxplot(x = "Year", y = "Price", data=df)
```



```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb0be234310>
```

Tổng kết

Qua bài thực hành này, chúng ta đã ôn lại cách vẽ các biểu đồ đã được học với bộ dữ liệu Food Price in Turkey



[Các sản phẩm có tính phí của Colab](#) - [Huỷ hợp đồng tại đây](#)

