addis

May 11, 2022

[19]:
```
#impoerterer pakker..
import numpy as np
import pandas as pd
import plotly.express as px

#her har jeg funnet en basic og enkel pakke som heter plotly. denne er␣
 ↪hovedpakken til resultatene
```

[22]:
```
#leser datasaett til dataframe
df = pd.read_csv('avocado-updated-2020.csv')
df.info()
#her viser innholdet til priser
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33045 entries, 0 to 33044
Data columns (total 13 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   date           33045 non-null  object
 1   average_price  33045 non-null  float64
 2   total_volume   33045 non-null  float64
 3   4046           33045 non-null  float64
 4   4225           33045 non-null  float64
 5   4770           33045 non-null  float64
 6   total_bags     33045 non-null  float64
 7   small_bags     33045 non-null  float64
 8   large_bags     33045 non-null  float64
 9   xlarge_bags    33045 non-null  float64
 10  type           33045 non-null  object
 11  year           33045 non-null  int64
 12  geography      33045 non-null  object
dtypes: float64(9), int64(1), object(3)
memory usage: 3.3+ MB
```

[43]:
```
#plotter en tabell som data er oppsatt.
avocado.head()
```

[43]:

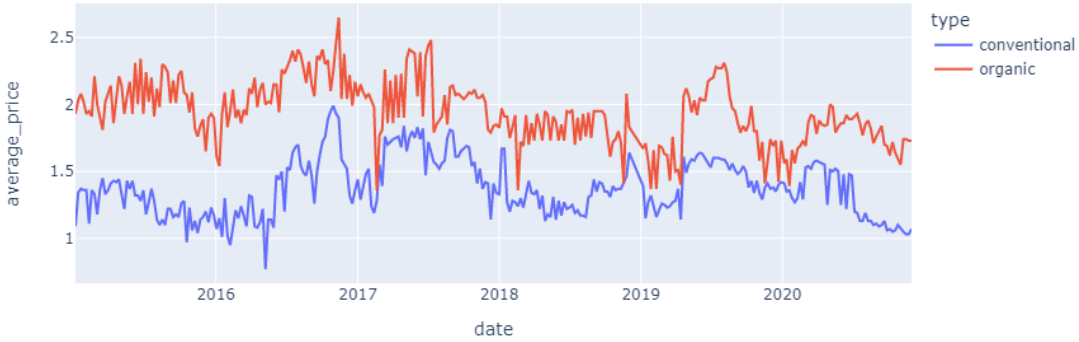
	date	average_price	total_volume	4046	4225	4770	total_bags	small_bags	large_bags	xlarge_bags	type	year	geography
0	2015-01-04	1.22000	40873.28000	2819.50000	28287.42000	49.90000	9716.46000	9186.93000	529.53000	0.00000	conventional	2015	Albany
1	2015-01-04	1.79000	1373.95000	57.42000	153.88000	0.00000	1162.65000	1162.65000	0.00000	0.00000	organic	2015	Albany
2	2015-01-04	1.00000	435021.49000	364302.39000	23821.16000	82.15000	46815.79000	16707.15000	30108.64000	0.00000	conventional	2015	Atlanta
3	2015-01-04	1.76000	3846.69000	1500.15000	938.35000	0.00000	1408.19000	1071.35000	336.84000	0.00000	organic	2015	Atlanta
4	2015-01-04	1.08000	788025.06000	53987.31000	552906.04000	39995.03000	141136.68000	137146.07000	3990.61000	0.00000	conventional	2015	Baltimore/Washington

[24]:
```
#viser katogoriserte variabler.
print(df['type'].value_counts(dropna=False))
print(df['geography'].value_counts(dropna=False))
```

```
conventional    16524
organic         16521
Name: type, dtype: int64
Philadelphia            612
Northeast               612
Syracuse                612
Seattle                 612
San Diego               612
Baltimore/Washington    612
Midsouth                612
Chicago                 612
Grand Rapids            612
New York                612
Houston                 612
Denver                  612
Los Angeles             612
Sacramento              612
Harrisburg/Scranton     612
Dallas/Ft. Worth        612
Buffalo/Rochester       612
Southeast               612
Pittsburgh              612
```

```
Atlanta                 612
St. Louis               612
Boston                  612
Jacksonville            612
Albany                  612
Cincinnati/Dayton       612
Spokane                 612
Orlando                 612
Raleigh/Greensboro      612
Richmond/Norfolk        612
New Orleans/Mobile      612
Las Vegas               612
San Francisco           612
Northern New England    612
Phoenix/Tucson          612
South Carolina          612
California              612
Plains                  612
Boise                   612
Indianapolis            612
Detroit                 612
Louisville              612
Tampa                   612
Great Lakes             612
Roanoke                 612
Nashville               612
Total U.S.              612
South Central           612
West                    612
Hartford/Springfield    612
Charlotte               612
Miami/Ft. Lauderdale    612
Portland                612
Columbus                612
West Tex/New Mexico     609
Name: geography, dtype: int64
```

[56]:
```
#her vises første polttet. her har jeg valgt New Yourk. men kan endes veldig␣
↪nekelt ved å Endre på df['geography'] == 'HER'
msk = df['geography'] == 'New York'
px.line(df[msk], x='date', y='average_price', color='type')
```



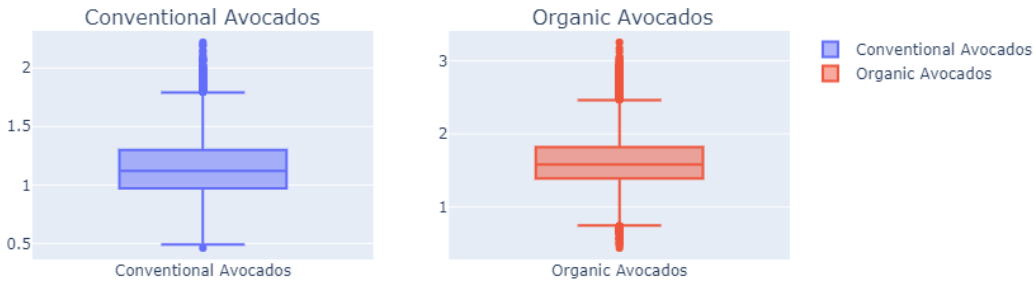
[31]:
```
#Her lager jeg en Boxplot av Conventional and Organic Avocados
# der kan man se dyreste og billigste priser på Conventional and Organic␣
 ↪Avocados.
conventional_avocado = avocado[avocado['type']=='conventional']
organic_avocado = avocado[avocado['type']=='organic']

fig = make_subplots(rows=1, cols=2, subplot_titles=('Conventional Avocados','␣
 ↪Organic Avocados'))

trace1 = go.Box(y=conventional_avocado['average_price'],name ='Conventional␣
 ↪Avocados')

trace2 = go.Box(y=organic_avocado['average_price'], name ='Organic Avocados')

fig.append_trace(trace1, row = 1, col=1)
fig.append_trace(trace2, row = 1, col=2)
fig.show()
```



[ ]:

[ ]:

[ ]:

[ ]: