Actividad Evaluable: Mapas de calor y boxplots

Eduardo Rodríguez Gil - A01274913, Jose Manuel Neri Villeda - A01706450, Héctor Javier Calderón González - A01067542

In [89]:
```
import pandas as pd
import seaborn as sb
import numpy as np; np.random.seed(0)
import matplotlib.pyplot as plt
data = pd.read_csv('Data Bitcoin.csv')
from matplotlib import cm
plt.rcParams['figure.figsize'] = (16, 9)
plt.style.use('ggplot')
# Voy a revisar dimensiones
data.shape
```

Out[89]: (421, 7)

Son 421 registros con 7 columnas que no se me olvide que puedo intercalar texto y fórmulas.

In [90]:
```
data.head()
```

Out[90]:

	Date	Price	Open	High	Low	Volume	Change
0	Apr 25, 2021	49561.9	50088.2	50438.8	49226.5	66.26K	-1.05%
1	Apr 24, 2021	50088.9	51140.8	51183.0	48775.2	82.25K	-2.06%
2	Apr 23, 2021	51143.6	51707.1	52099.9	47659.4	214.46K	-1.13%
3	Apr 22, 2021	51729.5	53821.3	55408.4	50590.9	168.13K	-3.88%
4	Apr 21, 2021	53820.2	56479.5	56764.4	53657.6	100.26K	-4.71%

In [91]:
```
data.describe()
```

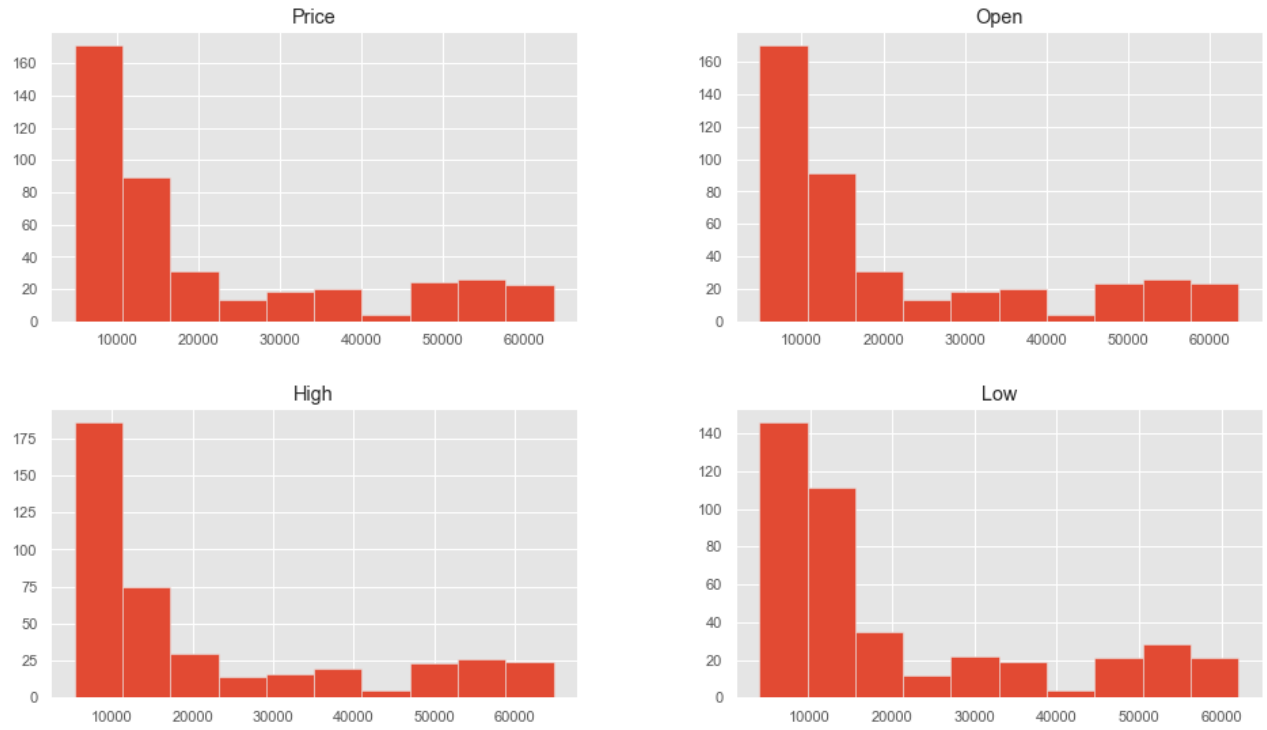
Out[91]:

	Price	Open	High	Low
count	421.000000	421.000000	421.000000	421.000000
mean	21471.073872	21372.344181	22028.754869	20687.659857
std	17492.702670	17448.718099	18024.928136	16785.882734
min	4826.000000	4815.200000	5369.300000	3869.500000
25%	9314.000000	9300.800000	9458.300000	9184.200000
50%	11557.200000	11533.500000	11766.900000	11315.900000
75%	32958.900000	32499.600000	34348.300000	30850.000000
max	63540.900000	63544.200000	64778.000000	62067.500000

Visualización general

Eliminar etiquetas de filas o columnas

In [92]:
```
data.drop ([0, 1]).hist()
plt.show ()
```



Filtros

In [93]:
```
mas_de_40 = data[data['Price'] > 4000]
mas_de_40
```

Out[93]:

	Date	Price	Open	High	Low	Volume	Change
0	Apr 25, 2021	49561.9	50088.2	50438.8	49226.5	66.26K	-1.05%
1	Apr 24, 2021	50088.9	51140.8	51183.0	48775.2	82.25K	-2.06%
2	Apr 23, 2021	51143.6	51707.1	52099.9	47659.4	214.46K	-1.13%
3	Apr 22, 2021	51729.5	53821.3	55408.4	50590.9	168.13K	-3.88%
4	Apr 21, 2021	53820.2	56479.5	56764.4	53657.6	100.26K	-4.71%
...	...	...	...	...	...	...	...
416	Mar 05, 2020	9060.3	8757.9	9147.3	8751.5	950.76K	3.45%
417	Mar 04, 2020	8757.9	8761.3	8840.3	8679.7	759.69K	-0.04%
418	Mar 03, 2020	8761.4	8906.1	8911.7	8669.3	1.01M	-1.61%
419	Mar 02, 2020	8904.8	8537.5	8961.8	8503.1	1.02M	4.27%
420	Mar 01, 2020	8540.0	8543.8	8737.2	8437.2	784.05K	-0.04%

421 rows × 7 columns

Doble filtro

In [94]:
```
doble_filtro = data[(data['Price'] > 4000) & (data['High'] > 35000)]
doble_filtro
```

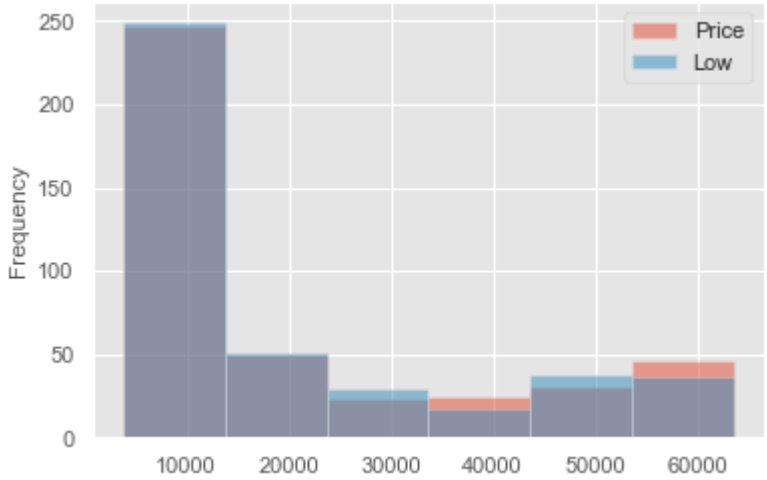
Out[94]:

	Date	Price	Open	High	Low	Volume	Change
0	Apr 25, 2021	49561.9	50088.2	50438.8	49226.5	66.26K	-1.05%
1	Apr 24, 2021	50088.9	51140.8	51183.0	48775.2	82.25K	-2.06%
2	Apr 23, 2021	51143.6	51707.1	52099.9	47659.4	214.46K	-1.13%
3	Apr 22, 2021	51729.5	53821.3	55408.4	50590.9	168.13K	-3.88%
4	Apr 21, 2021	53820.2	56479.5	56764.4	53657.6	100.26K	-4.71%
...	...	...	...	...	...	...	...
105	Jan 10, 2021	38192.2	40149.7	41362.4	35141.6	215.78K	-4.88%
106	Jan 09, 2021	40151.9	40607.2	41363.5	38775.1	128.42K	-1.10%
107	Jan 08, 2021	40599.3	39466.4	41921.7	36613.4	251.29K	2.89%
108	Jan 07, 2021	39460.2	36798.5	40340.9	36361.2	249.60K	7.25%
109	Jan 06, 2021	36793.2	33999.3	36934.8	33408.3	227.56K	8.24%

100 rows × 7 columns

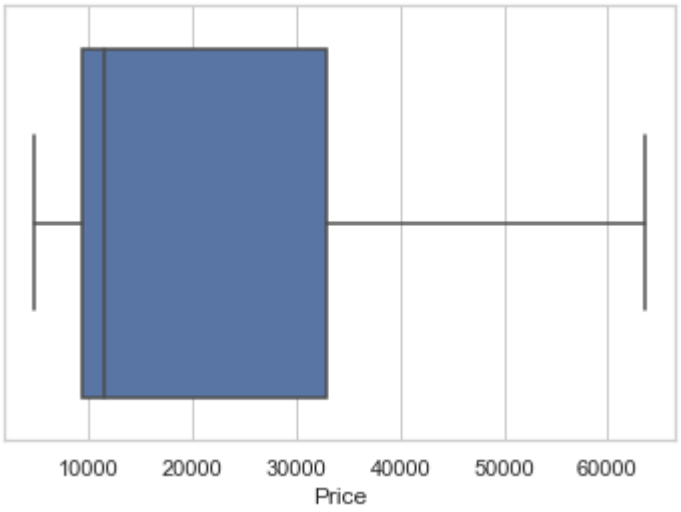
Visualización

In [95]:
```
%matplotlib inline
data[['Price', 'Low']].plot.hist(bins=6, alpha=0.5)
plt.show()
```

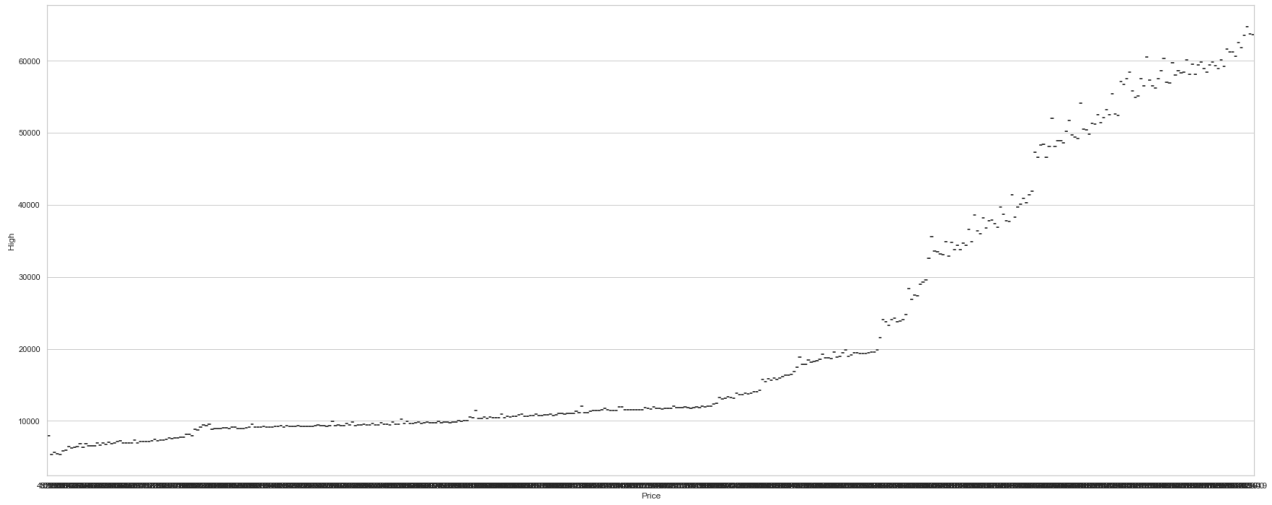


Boxplot para obtener un diagrama de cajas y bigotes

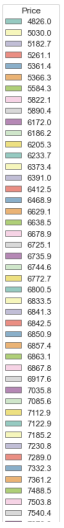
In [112…
```
sb.set_theme(style = "whitegrid")
Bitcoin = pd.read_csv('Data Bitcoin.csv')
ax = sb.boxplot(x = Bitcoin["Price"])
```

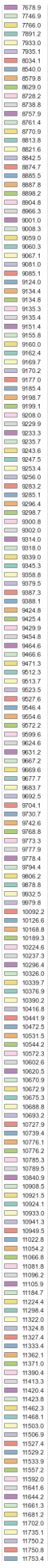


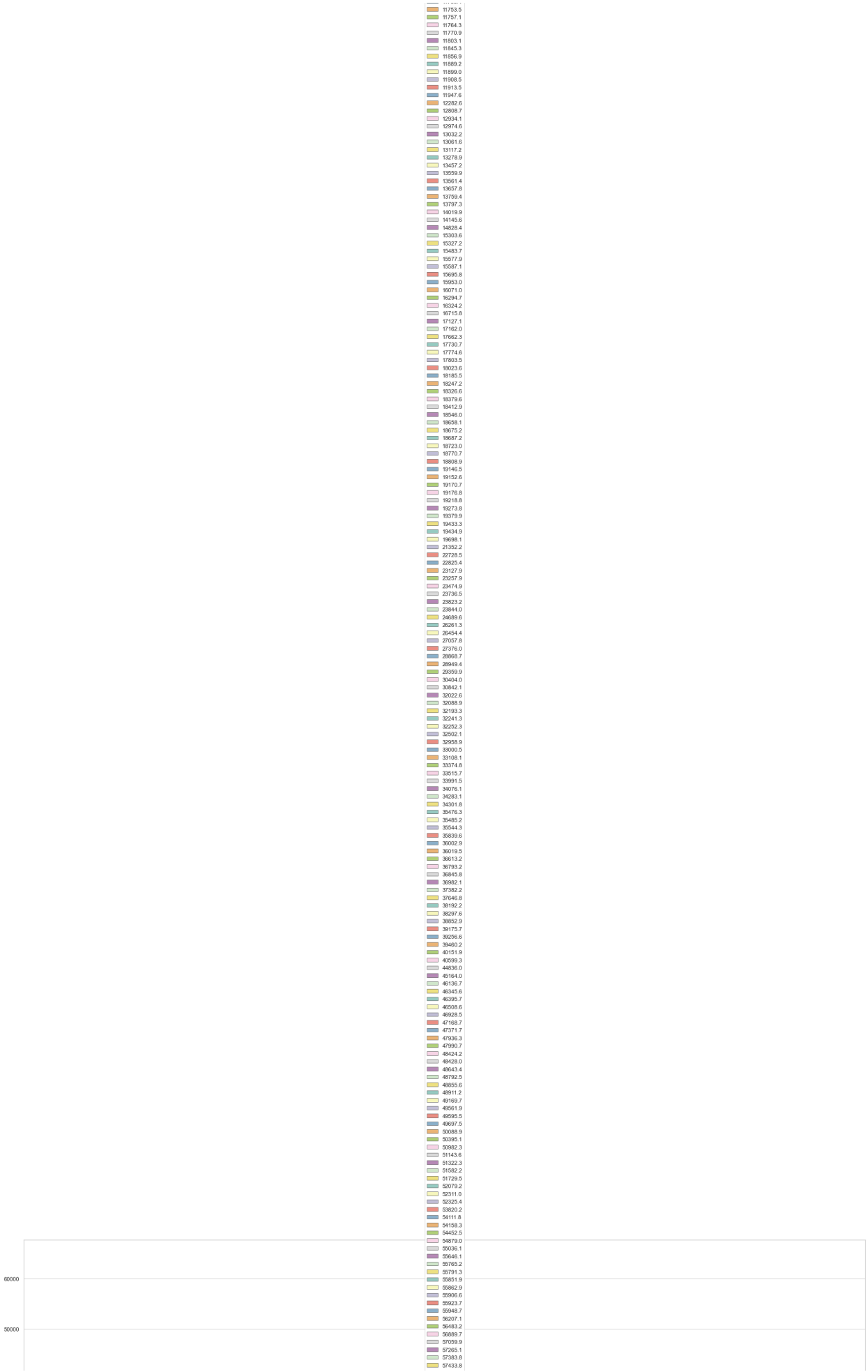
In [115…
```
plt.figure(figsize=(30,12))
ax = sb.boxplot(x = "Price", y = "High", data = Bitcoin)
```

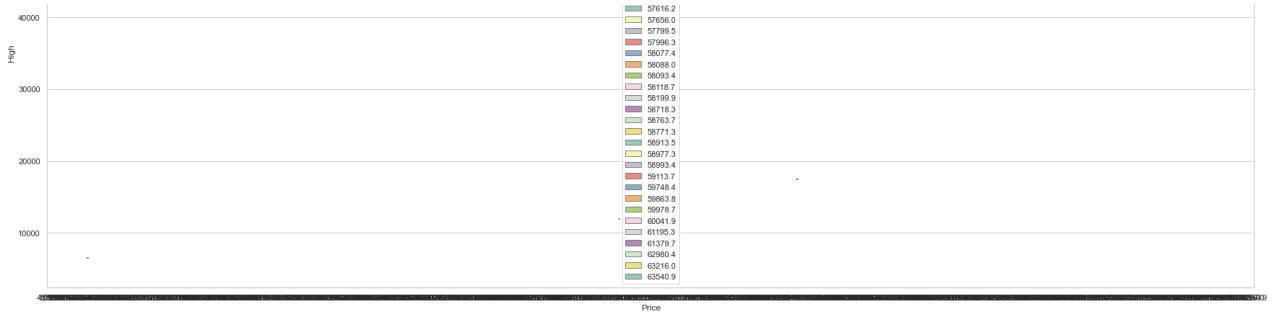


In [125…
```
plt.figure(figsize=(30,12))
ax = sb.boxplot(x = "Price", y = "High", hue = "Price",
                data = Bitcoin, palette = "Set3")
```




7678.9
7746.9
7766.0
7891.2
7933.0
7935.1
8034.1
8540.0
8579.8
8629.0
8728.2
8738.8
8757.9
8761.4
8770.9
8813.8
8821.6
8842.5
8874.7
8885.5
8887.8
8898.2
8904.8
8966.3
9001.0
9008.3
9059.0
9060.3
9067.1
9081.0
9085.1
9124.0
9134.4
9134.8
9135.3
9135.4
9151.4
9155.8
9160.0
9162.4
9169.7
9170.2
9177.0
9185.4
9198.7
9199.1
9208.0
9229.9
9233.3
9235.7
9243.6
9247.5
9253.4
9256.0
9283.2
9285.1
9296.4
9298.7
9300.8
9302.0
9314.0
9318.0
9339.0
9345.3
9358.8
9379.5
9387.3
9388.1
9424.8
9425.4
9429.9
9454.8
9464.6
9466.6
9471.3
9512.3
9513.7
9523.5
9527.6
9546.4
9554.6
9572.2
9599.6
9624.6
9631.2
9667.2
9669.6
9677.7
9683.7
9692.5
9704.1
9730.7
9742.6
9768.8
9773.3
9777.9
9778.4
9794.4
9806.2
9878.8
9932.5
9979.8
10092.2
10126.6
10168.8
10189.3
10224.6
10237.3
10296.4
10326.0
10339.7
10376.9
10390.2
10416.8
10441.9
10472.5
10531.5
10544.2
10572.3
10602.6
10620.5
10670.9
10672.9
10675.3
10688.8
10693.2
10727.9
10739.4
10776.1
10776.2
10785.3
10789.5
10840.9
10908.5
10921.5
10924.1
10933.0
10941.3
10949.5
11022.8
11054.2
11066.8
11081.8
11096.2
11105.9
11184.7
11224.4
11298.4
11322.0
11324.8
11327.4
11333.4
11362.1
11371.0
11390.4
11413.3
11420.4
11423.8
11462.3
11468.1
11503.0
11506.9
11527.4
11529.2
11533.9
11557.2
11592.0
11641.6
11644.2
11661.3
11681.2
11702.0
11735.1
11750.2
11750.8
11753.4

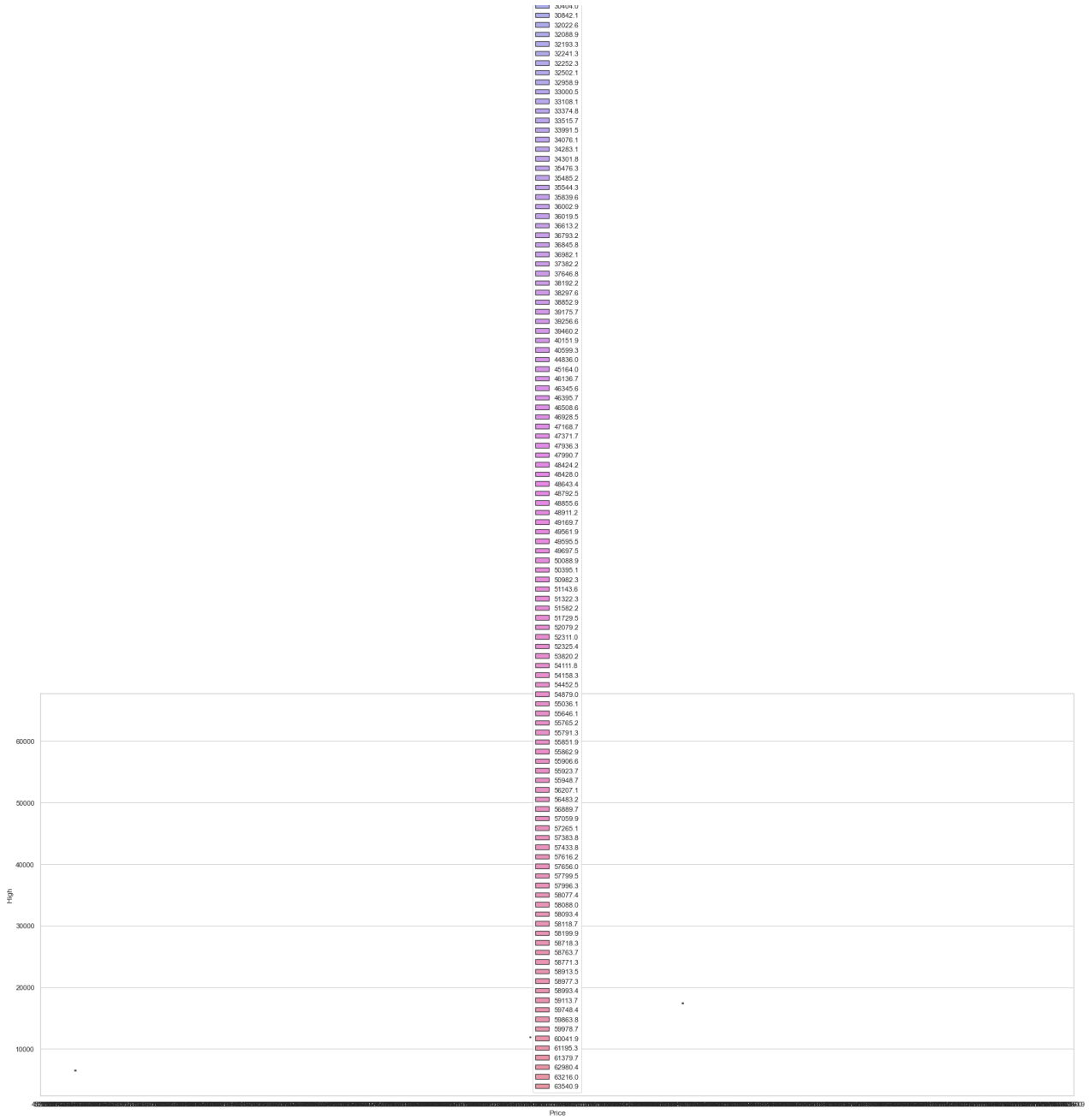

11753.5
11757.1
11764.3
11770.9
11803.1
11845.3
11856.9
11889.2
11899.0
11908.5
11913.5
11947.6
12282.6
12808.7
12934.1
12974.6
13032.2
13061.6
13117.2
13278.9
13457.2
13559.9
13561.4
13657.8
13759.4
13797.3
14019.9
14145.6
14828.4
15303.6
15327.2
15483.7
15577.9
15587.1
15695.8
15953.0
16071.0
16294.7
16324.2
16715.8
17127.1
17162.0
17662.3
17730.7
17774.6
17803.5
18023.6
18185.5
18247.2
18326.6
18379.6
18412.9
18546.0
18658.1
18675.2
18687.2
18723.0
18770.7
18808.9
19146.5
19152.6
19170.7
19176.8
19218.8
19273.8
19379.9
19433.3
19434.9
19698.1
21352.2
22728.5
22825.4
23127.9
23257.9
23474.9
23736.5
23823.2
23844.0
24689.6
26261.3
26454.4
27057.8
27376.0
28868.7
28949.4
29359.9
30404.0
30842.1
32022.6
32088.9
32193.3
32241.3
32252.3
32502.1
32958.9
33000.5
33108.1
33374.8
33515.7
33991.5
34076.1
34283.1
34301.8
35476.3
35485.2
35544.3
35839.6
36002.9
36019.5
36613.2
36793.2
36845.8
36982.1
37382.2
37646.8
38192.2
38297.6
38852.9
39175.7
39256.6
39460.2
40151.9
40599.3
44836.0
45164.0
46136.7
46345.6
46395.7
46508.6
46928.5
47168.7
47371.7
47936.3
47990.7
48424.2
48428.0
48643.4
48792.5
48855.6
48911.2
49169.7
49561.9
49595.5
49697.5
50088.9
50395.1
50982.3
51143.6
51322.3
51582.2
51729.5
52079.2
52311.0
52325.4
53820.2
54111.8
54158.3
54452.5
54879.0
55036.1
55646.1
55765.2
55791.3
55851.9
55862.9
55906.6
55923.7
55948.7
56207.1
56483.2
56889.7
57059.9
57265.1
57383.8
57433.8
60000
50000



In [126…

```
plt.figure(figsize=(30,12))
ax = sb.boxplot(x = "Price", y = "High", hue = "Price",
                data = Bitcoin, linewidth = 2.5)
```

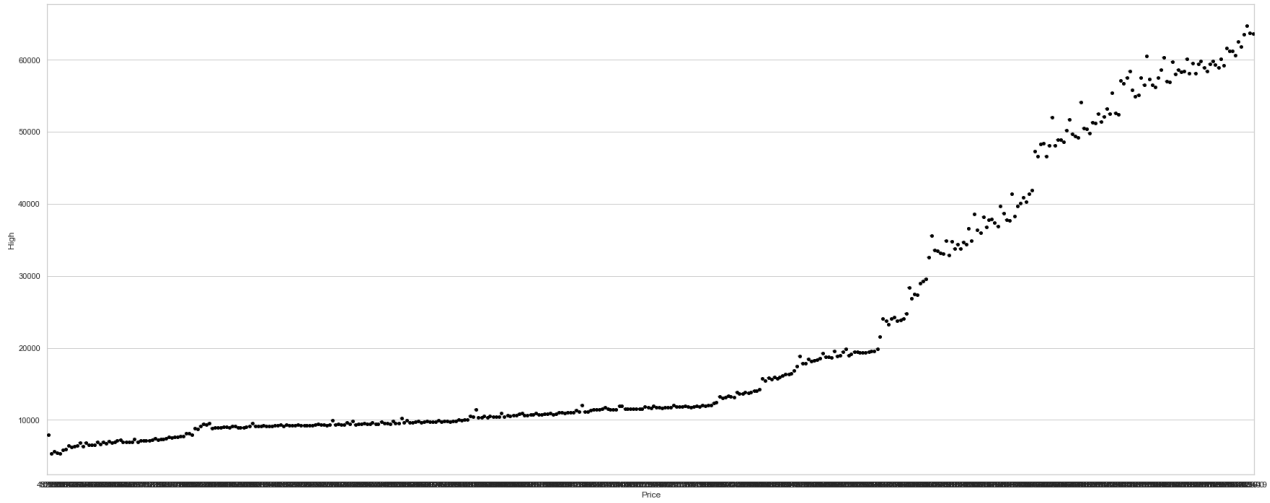

9677.7
9683.7
9692.5
9704.1
9730.7
9742.6
9768.8
9773.3
9777.9
9778.4
9794.4
9806.2
9878.8
9932.5
9979.8
10092.2
10126.6
10168.8
10189.3
10224.6
10237.3
10296.4
10326.0
10339.7
10376.9
10390.2
10416.8
10441.9
10472.5
10531.5
10544.2
10572.3
10602.6
10620.5
10670.9
10672.9
10675.3
10688.8
10693.2
10727.9
10739.4
10776.1
10776.2
10785.3
10789.5
10840.9
10908.5
10921.5
10924.1
10933.0
10941.3
10949.5
11022.8
11054.2
11066.8
11081.8
11096.2
11105.9
11184.7
11224.4
11298.4
11322.0
11324.8
11327.4
11333.4
11362.1
11371.0
11390.4
11413.3
11420.4
11423.8
11462.3
11468.1
11503.0
11506.9
11527.4
11529.2
11533.9
11557.2
11592.0
11641.6
11644.2
11661.3
11681.2
11702.0
11735.1
11750.2
11750.8
11753.4
11753.5
11757.1
11764.3
11770.9
11803.1
11845.3
11856.9
11889.2
11899.0
11908.5
11913.5
11947.6
12282.6
12808.7
12934.1
12974.6
13032.2
13061.6
13117.2
13278.9
13457.2
13559.9
13561.4
13657.8
13759.4
13797.3
14019.9
14145.6
14828.4
15303.6
15327.2
15483.7
15577.9
15587.1
15695.8
15953.0
16071.0
16294.7
16324.2
16715.8
17127.1
17162.0
17662.3
17730.7
17774.6
17803.5
18023.6
18185.5
18247.2
18326.6
18379.6
18412.9
18546.0
18658.1
18675.2
18687.2
18723.0
18770.7
18808.9
19146.5
19152.6
19170.7
19176.8
19218.8
19273.8
19379.9
19433.3
19434.9
19698.1
21352.2
22728.5
22825.4
23127.9
23257.9
23474.9
23736.5
23823.2
23844.0
24689.6
26261.3
26454.4
27057.8
27376.0
28868.7
28949.4
29359.9



Cajas

In [127…

```
plt.figure(figsize=(30,12))
ax = sb.boxplot(x = "Price", y = "High", data = Bitcoin)
ax = sb.stripplot(x = "Price", y = "High", data = Bitcoin, color = "0.0000005")
```



Correlación

In [135…
```
data.corr(method = 'pearson')
```

Out[135…

	Price	Open	High	Low
Price	1.000000	0.997997	0.999207	0.998767
Open	0.997997	1.000000	0.999032	0.998379
High	0.999207	0.999032	1.000000	0.998287
Low	0.998767	0.998379	0.998287	1.000000

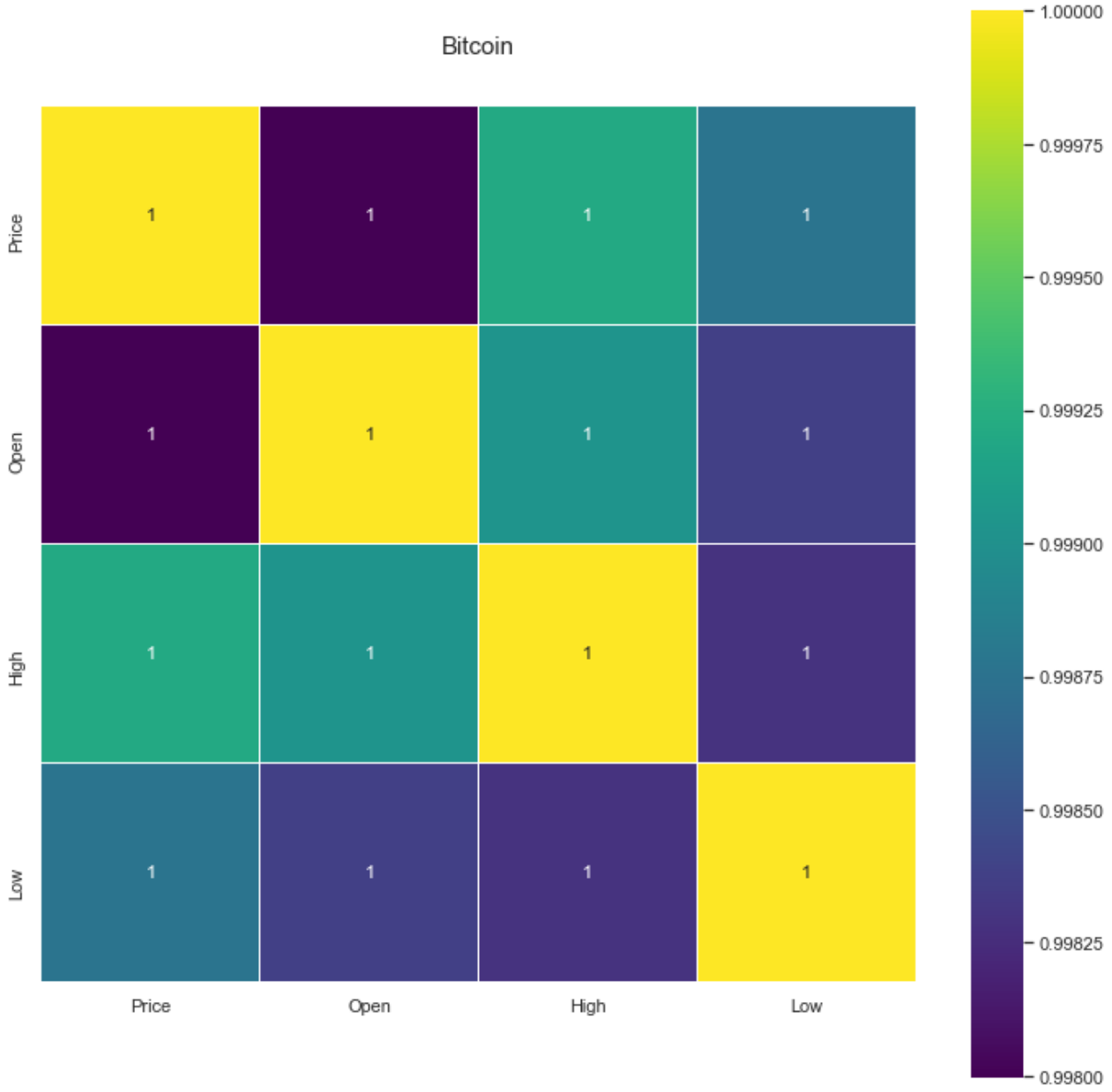
In [129…
```
data.corr(method = 'kendall')
```

Out[129…

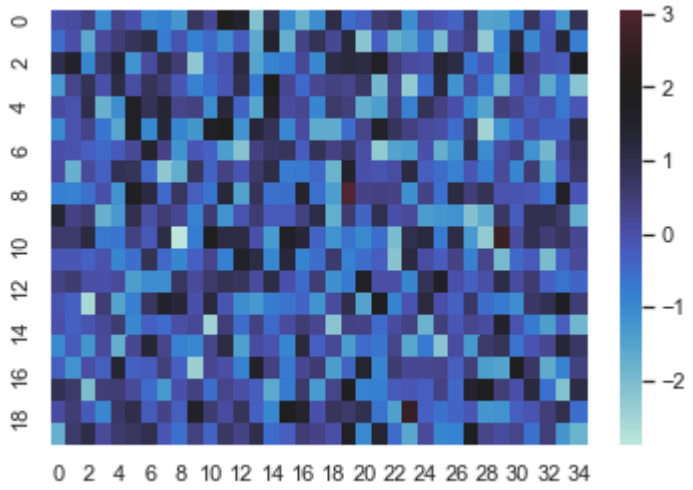
	Price	Open	High	Low
Price	1.000000	0.945413	0.966768	0.969574
Open	0.945413	1.000000	0.965298	0.962402
High	0.966768	0.965298	1.000000	0.957154
Low	0.969574	0.962402	0.957154	1.000000

Visualizar mapa de calor

In [146…
```
colormap = plt.cm.viridis
plt.figure(figsize = (12, 12))
plt.title('Bitcoin', y = 1.05, size = 15)
sb.heatmap(data.corr(), linewidths = 0.1, vmax = 1.0, square = True, cmap = colormap, l
```

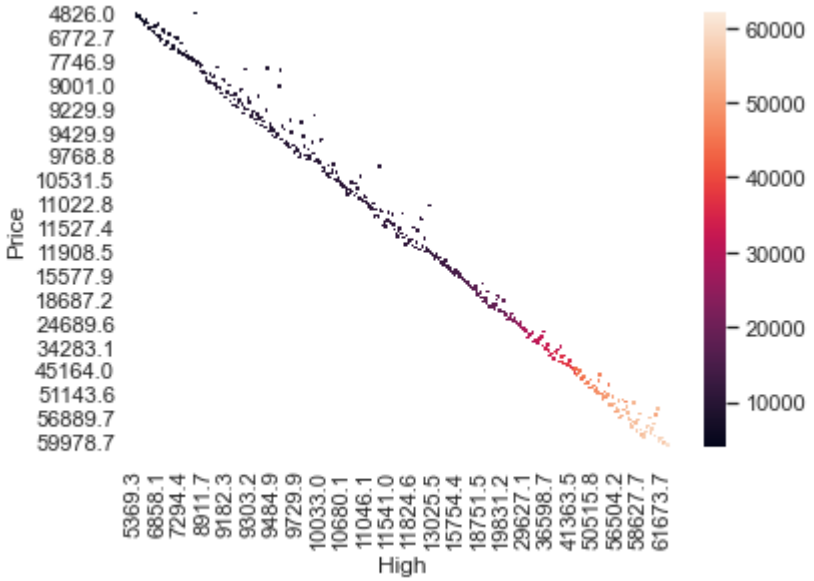


In [140…
```
normal_data = np.random.randn(20, 35)
ax = sb.heatmap(normal_data, center = 2)
```



In [141…
```
Bitcoin = pd.read_csv('Data Bitcoin.csv')
```

```
Bitcoin = Bitcoin.pivot("Price", "High", "Low")
ax = sb.heatmap(Bitcoin)
```



¿Hay alguna variable que no aporta información?

Todas las variables aportan información, pero con las que estamos trabajando que son numéricas son las variables de Price, Open, High y Low.

Si tuvieras que eliminar variables, ¿cuáles quitarías y por qué?

Principalmente quitaría la de Volume y Change porque no las utilizamos para nada y es información que no es tan relevante para nuestra base de datos. Pero de igual forma podríamos quitar la de Date, ya que tampoco es numérica y no la llegamos a utilizar, pero podría ser un poco más importante que las otras dos, ya que si queremos llegar a saber los cambios de precio de la Bitcoin en que fechas esta nos podría servir.

¿Existen variables que tengan datos extraños?

En si ninguna es tan extraña, podríamos decir que la de Change, ya que utiliza porcentaje y la de Date, ya que no son solo números, pero en si las 4 variables que utilizamos son numéricas y son similares.

Si comparas las variables, ¿todas están en rangos similares? ¿Crees que esto afecte?

Las que son más similares son las de Price, Open, High y Low sus rangos son muy idénticos y no llegaría a afectar en nada.

¿Puedes encontrar grupos qué se parezcan? ¿Qué grupos son estos?

Los grupos qué se parecen mucho son los de Price, Open, High y Low, ya que estos llegan a contener el precio de la Bitcoin y no llega a variar mucho.