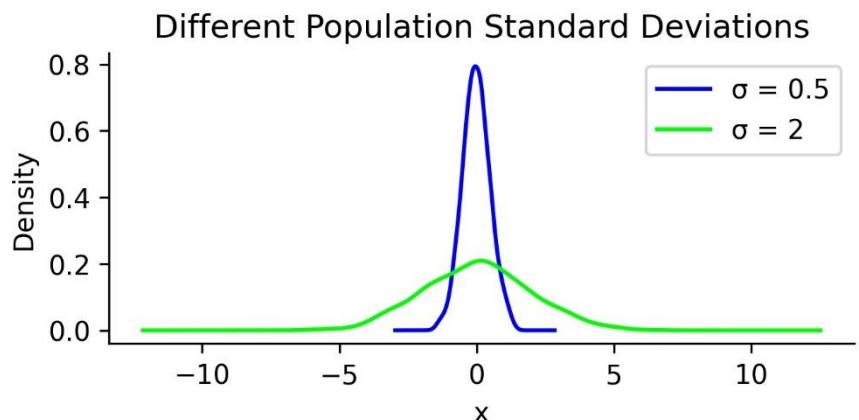


Chapter 1: Introduction to Data Analysis



[stefmolin / Hands-On-Data-Analysis-with-Pandas-2nd-edition](#)

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Security](#) [Insights](#)

[master](#) [1 branch](#) [2 tags](#)

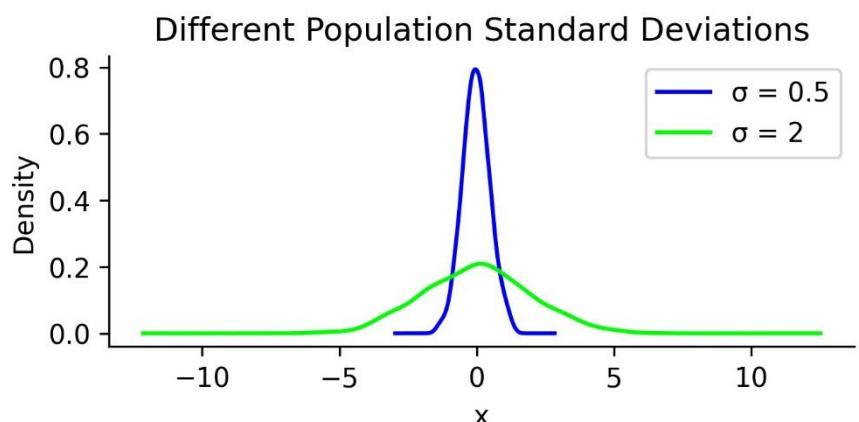
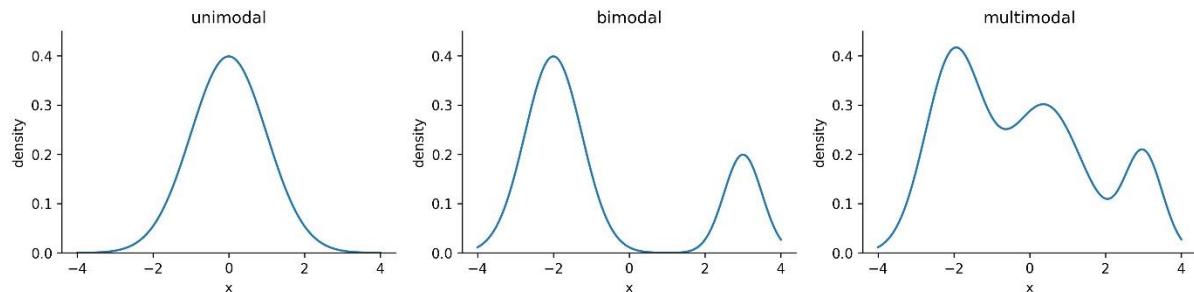
[Go to file](#) [Code](#)

About

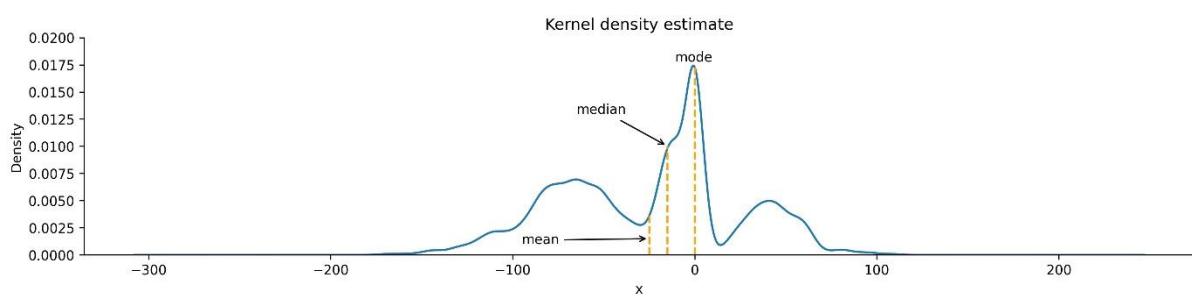
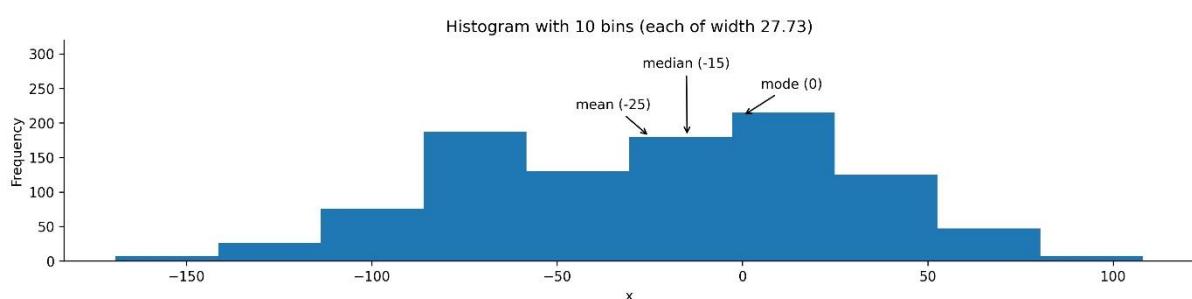
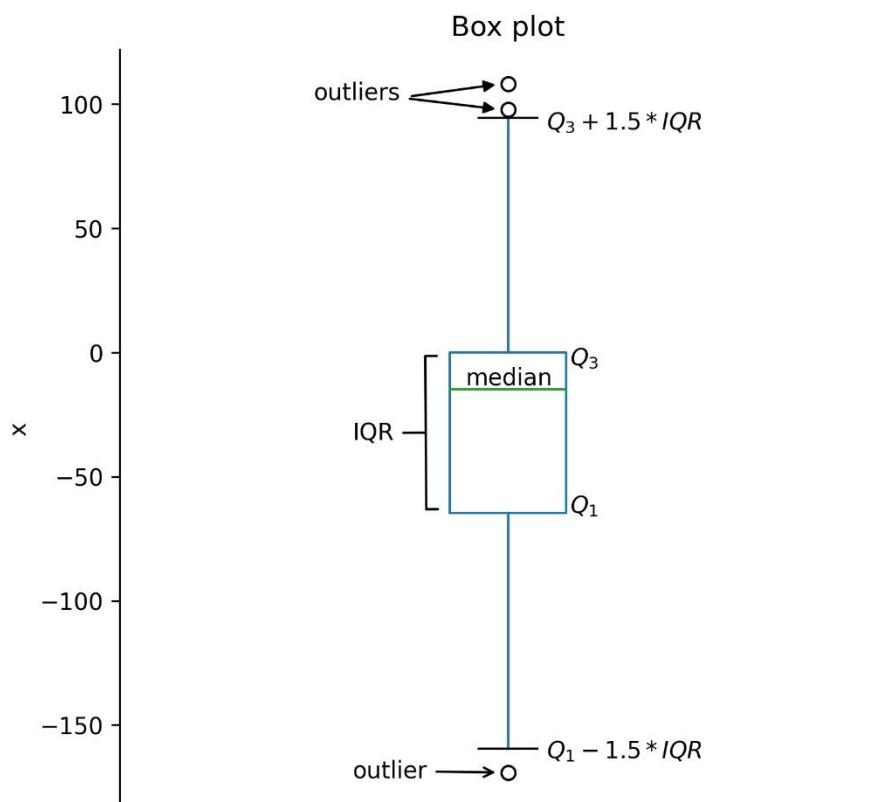
Materials for following along with
Hands-On Data Analysis with Pandas,
2nd edition

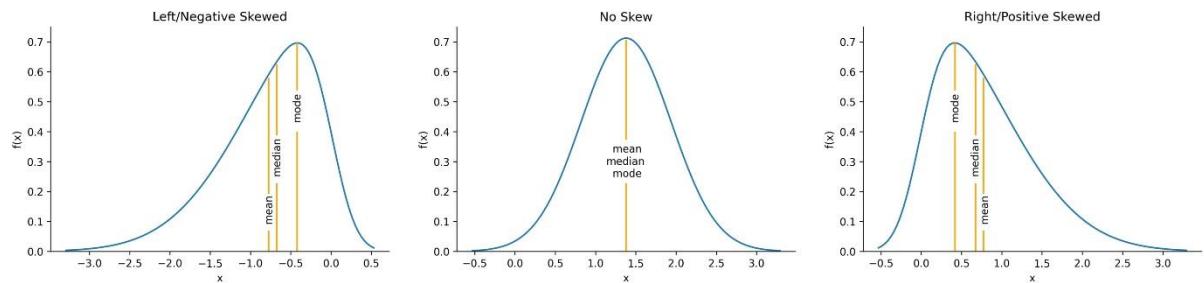
[www.packtpub.com/product/hands-o...](#)

[data-analysis](#) [data-analysis-python](#)
[data-analysis-pandas](#) [data-science](#)

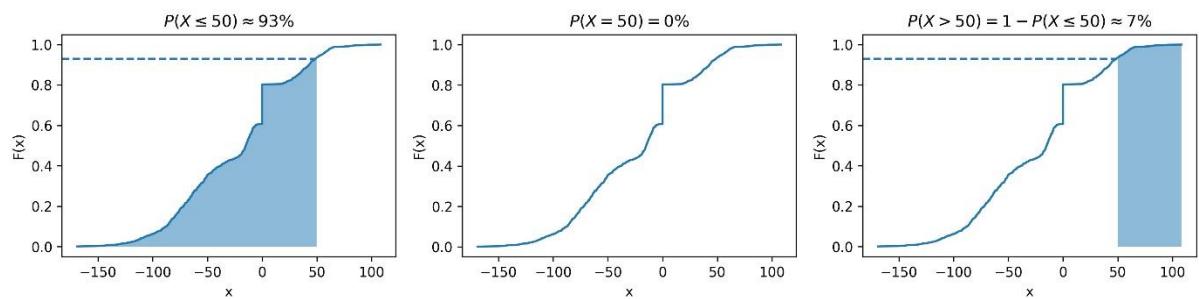


Quartile	Statistic	Percentile
1. Q_0	minimum	0 th
2. Q_1	N/A	25 th
3. Q_2	median	50 th
4. Q_3	N/A	75 th
5. Q_4	maximum	100 th

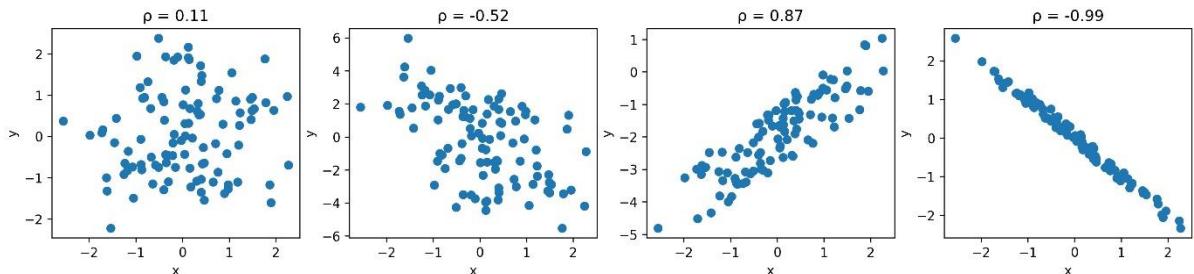
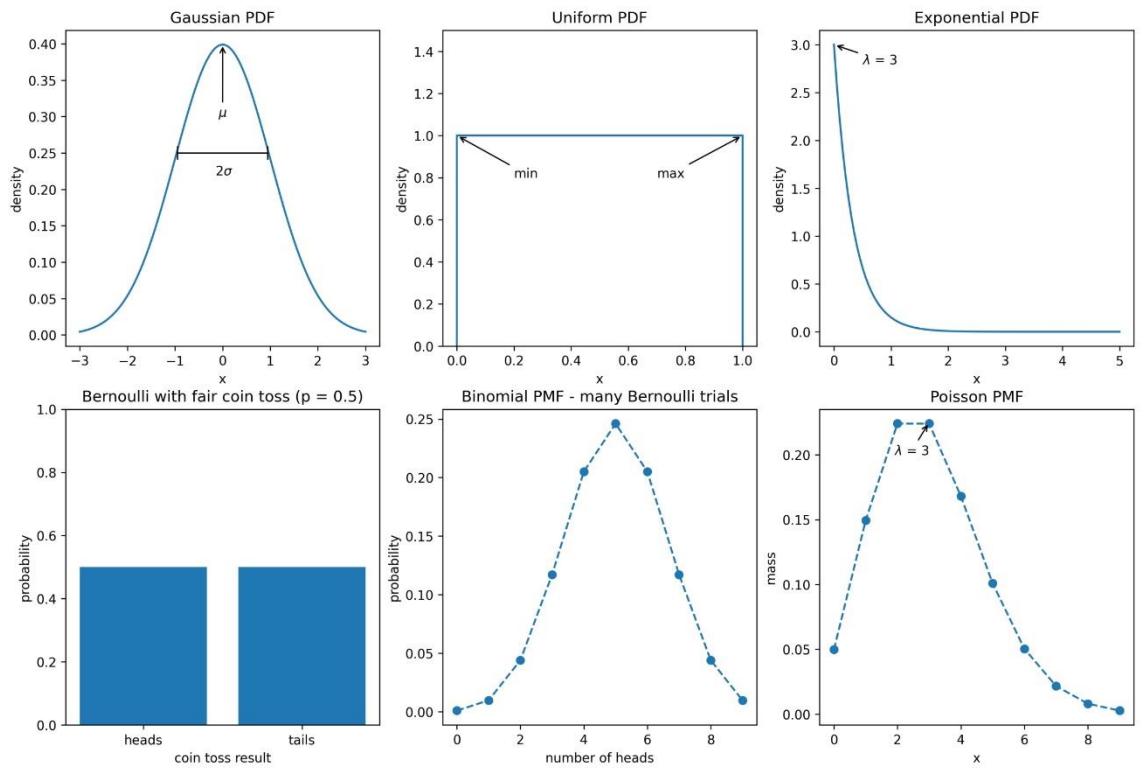


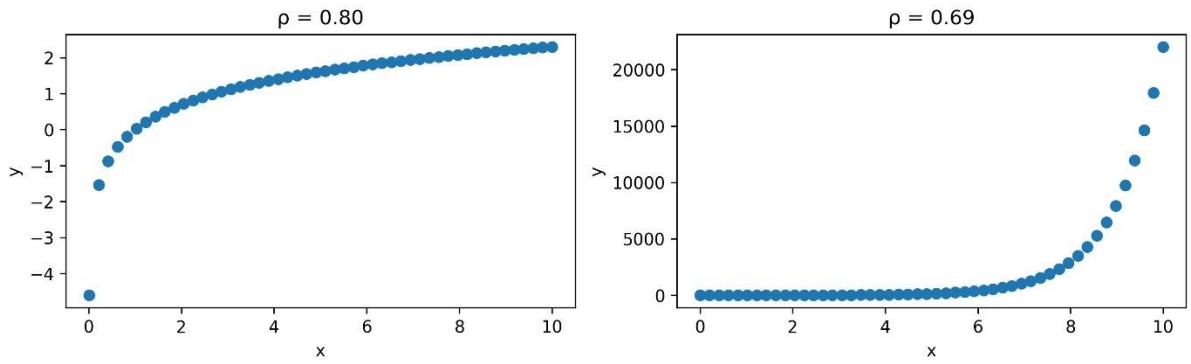


Understanding the CDF

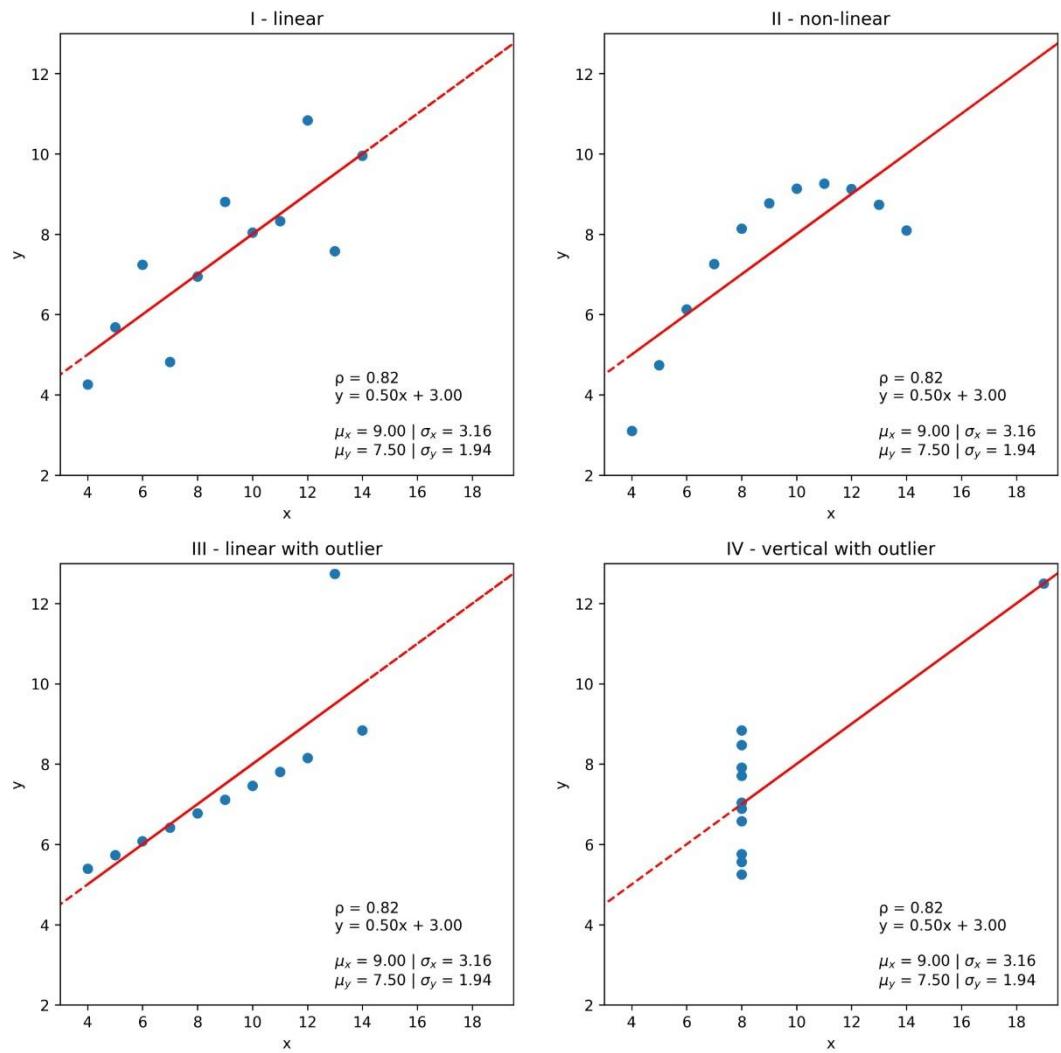


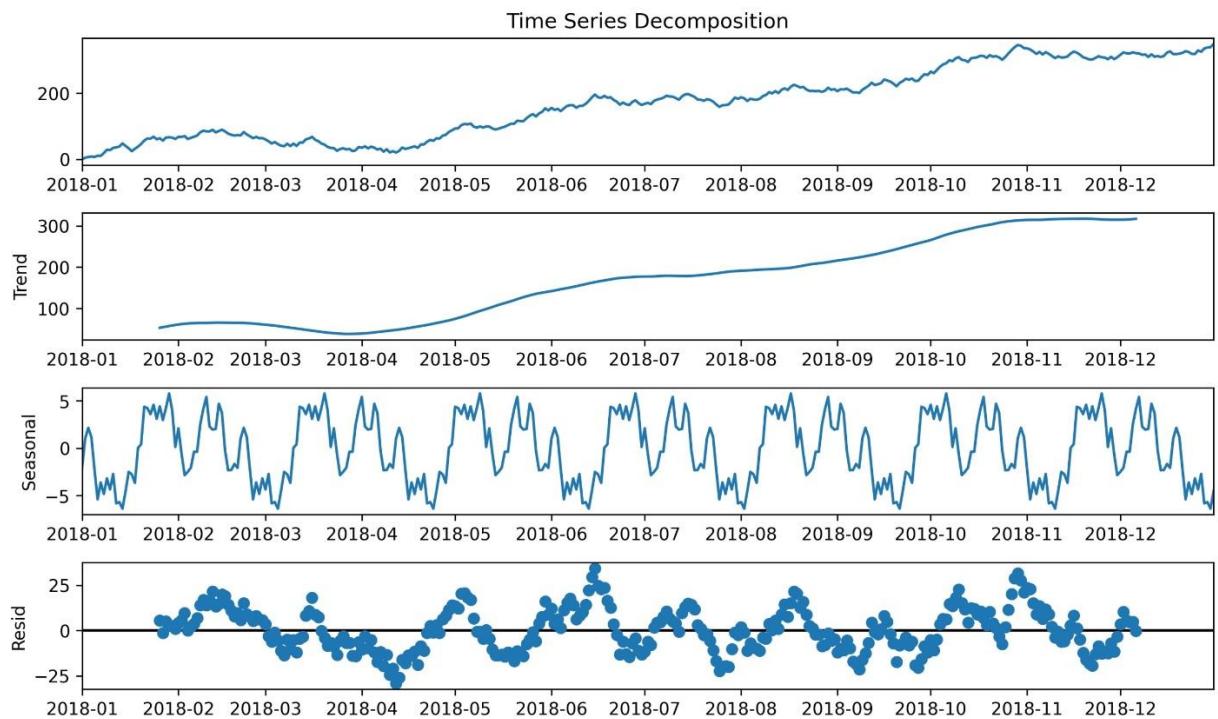
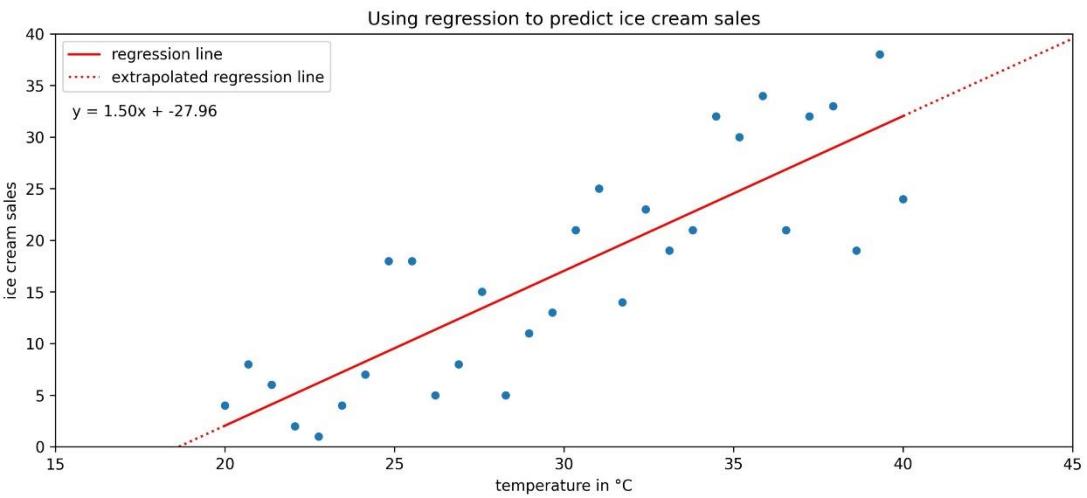
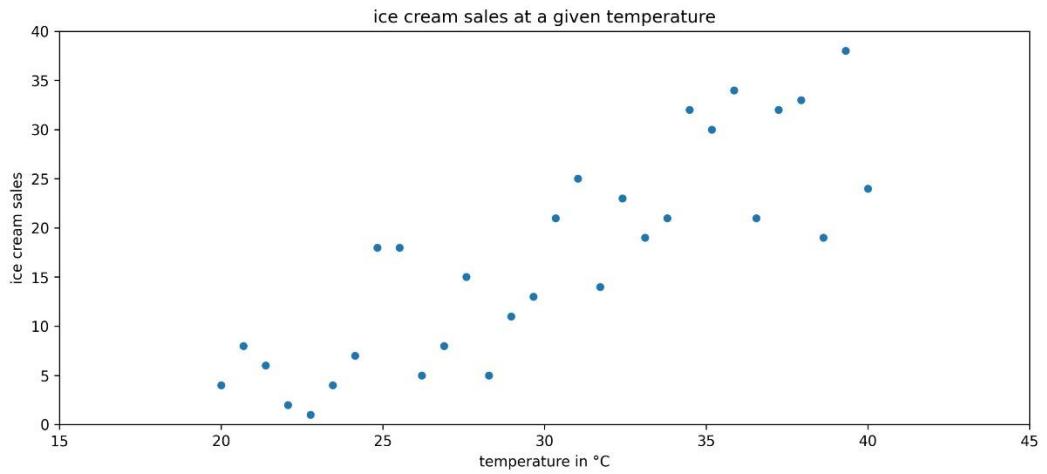
Some commonly used distributions

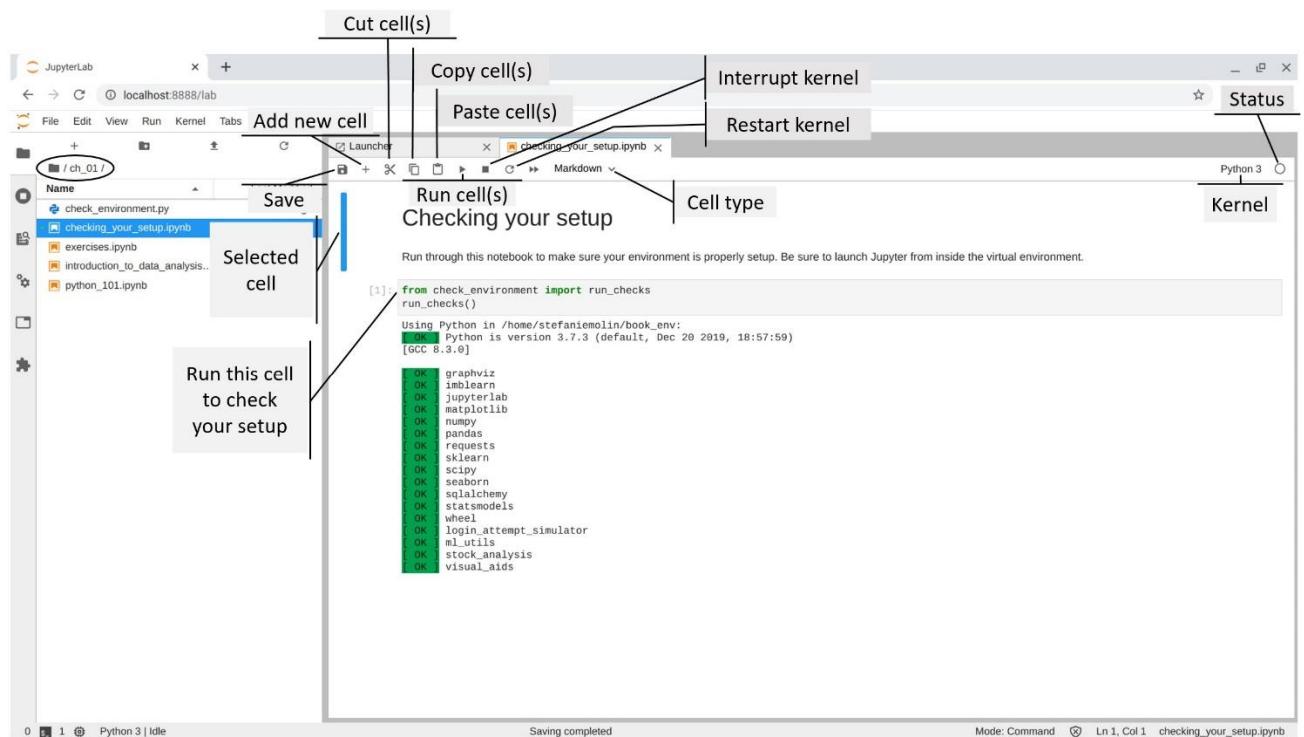
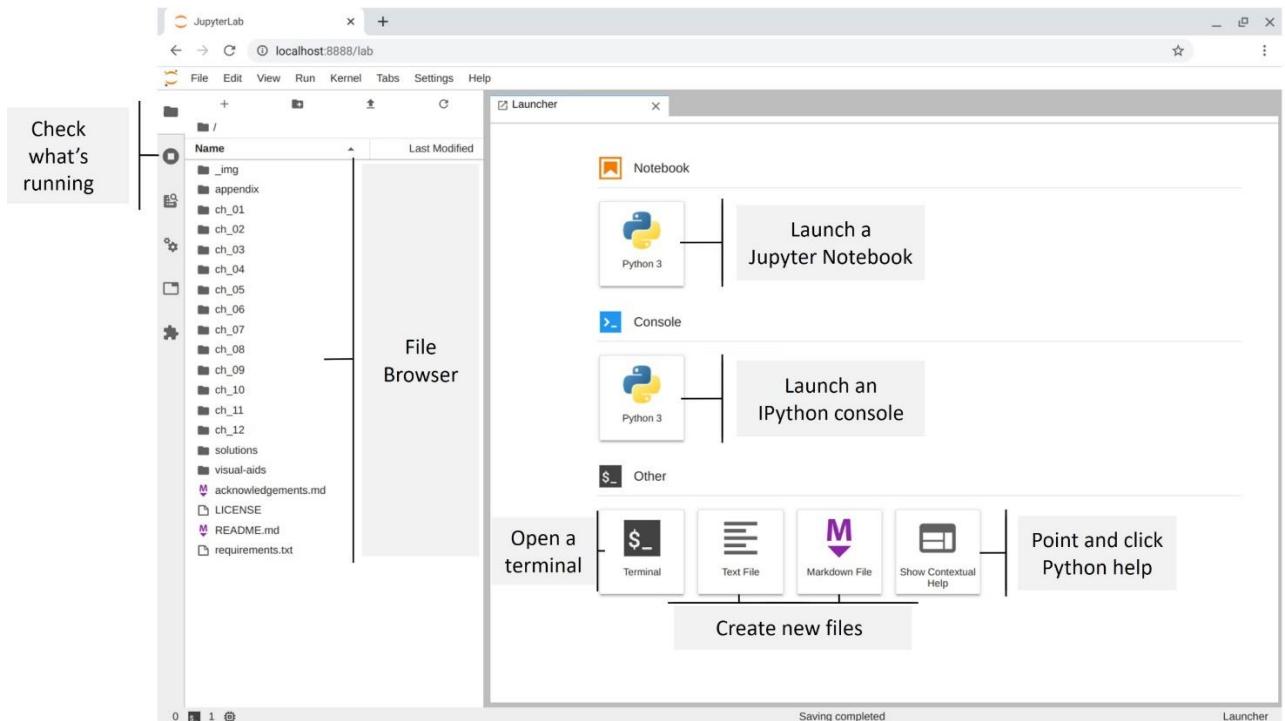


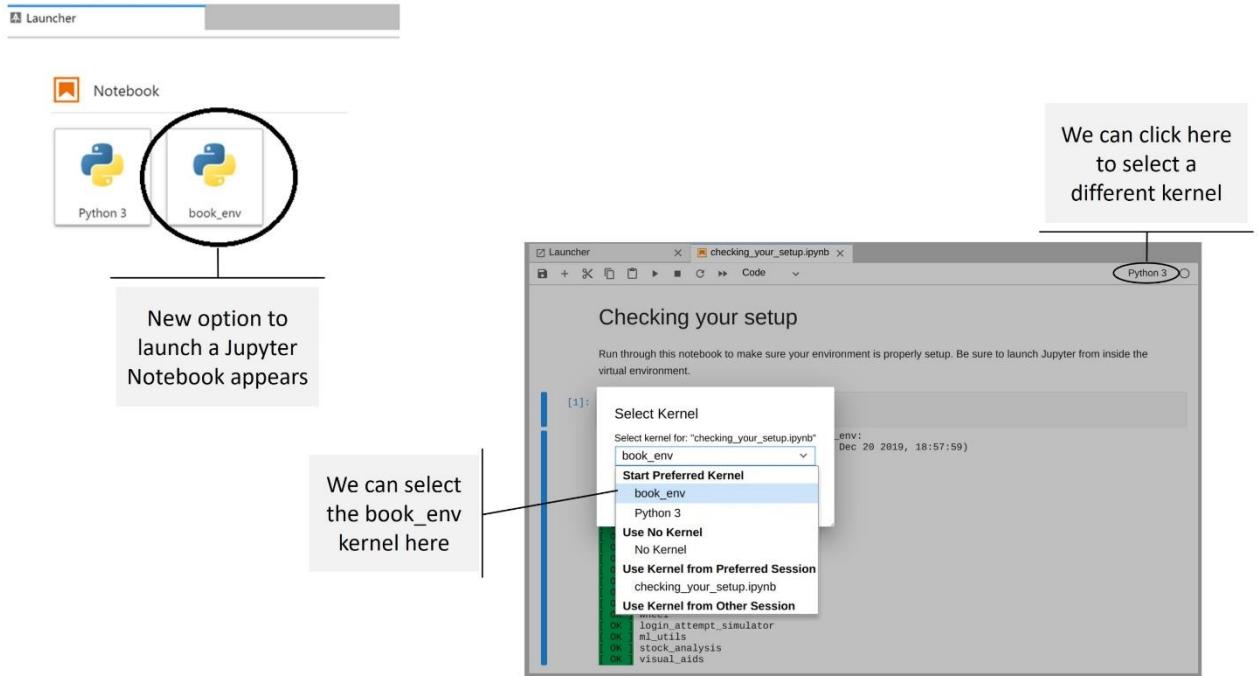


Anscombe's Quartet









Chapter 2: Working with Pandas DataFrames

Attribute	Returns
name	The name of the <code>Series</code> object
dtype	The data type of the <code>Series</code> object
shape	Dimensions of the <code>Series</code> object in a tuple of the form <code>(number of rows,)</code>
index	The <code>Index</code> object that is part of the <code>Series</code> object
values	The data in the <code>Series</code> object

Attribute	Returns
name	The name of the <code>Index</code> object
dtype	The data type of the <code>Index</code> object
shape	Dimensions of the <code>Index</code> object
values	The data in the <code>Index</code> object
is_unique	Check if the <code>Index</code> object has all unique values

	time	place	magType	mag	alert	tsunami
0	2018-10-13 11:10:23.560	262km NW of Ozernovskiy, Russia	mww	6.7	green	1
1	2018-10-13 04:34:15.580	25km E of Bitung, Indonesia	mww	5.2	green	0
2	2018-10-13 00:13:46.220	42km WNW of Sola, Vanuatu	mww	5.7	green	0
3	2018-10-12 21:09:49.240	13km E of Nueva Concepcion, Guatemala	mww	5.7	green	0
4	2018-10-12 02:52:03.620	128km SE of Kimbe, Papua New Guinea	mww	5.6	green	1

Attribute	Returns
dtypes	The data types of each column
shape	Dimensions of the <code>DataFrame</code> object in a tuple of the form <code>(number of rows, number of columns)</code>
index	The <code>Index</code> object along the rows of the <code>DataFrame</code> object
columns	The name of the columns (as an <code>Index</code> object)
values	The data in the <code>DataFrame</code> object
empty	Check if the <code>DataFrame</code> object is empty

	time	place	magType	mag	alert	tsunami			
0	2018-10-13 11:10:23.560	2018-10-13 11:10:23.560	262km NW of Ozernovskiy, Russia	262km NW of Oze...	mwwmww	13.4	green	green	2
1	2018-10-13 04:34:15.580	2018-10-13 04:34:15.580	25km E of Bitung, Indonesia	25km E of Bitung, I...	mwwmww	10.4	green	green	0
2	2018-10-13 00:13:46.220	2018-10-13 00:13:46.220	42km WNW of Sola, Vanuatu	42km WNW of Sola, Van...	mwwmww	11.4	green	green	0
3	2018-10-12 21:09:49.240	2018-10-12 21:09:49.240	13km E of Nueva Concepcion, Guatemala	13km E of...	mwwmww	11.4	green	green	0
4	2018-10-12 02:52:03.620	2018-10-12 02:52:03.620	128km SE of Kimbe, Papua New Guinea	128km SE of...	mwwmww	11.2	green	green	2

random text truth

date

2019-04-17	0.548814	hot	False
2019-04-18	0.715189	warm	True
2019-04-19	0.602763	cool	True
2019-04-20	0.544883	cold	False
2019-04-21	0.423655	None	True

mag place

0	5.2	California
1	1.2	Alaska
2	0.2	California

n n_squared n_cubed

0	0	0	0
1	1	1	1
2	2	4	8
3	3	9	27
4	4	16	64

Parameter	Purpose
<code>sep</code>	Specifies the delimiter
<code>header</code>	Row number where the column names are located; the default option has <code>pandas infer</code> whether they are present
<code>names</code>	List of column names to use as the header
<code>index_col</code>	Column to use as the index
<code>usecols</code>	Specifies which columns to read in
<code>dtype</code>	Specifies data types for the columns
<code>converters</code>	Specifies functions for converting data in certain columns
<code>skiprows</code>	Rows to skip
<code>nrows</code>	Number of rows to read at a time (combine with <code>skiprows</code> to read a file bit by bit)
<code>parse_dates</code>	Automatically parse columns containing dates into datetime objects
<code>chunksize</code>	For reading the file in chunks
<code>compression</code>	For reading in compressed files without extracting beforehand
<code>encoding</code>	Specifies the file encoding

	<code>alert</code>	<code>type</code>	<code>title</code>	<code>place</code>	<code>magType</code>	<code>mag</code>	<code>time</code>									
0	None	earthquake	M 5.0 - 165km NNW of Flying Fish Cove, Christm...	165km NNW of Flying Fish Cove, Christmas Island	mww	5.0	1539459504090									
1	green	earthquake	M 6.7 - 262km NW of Ozernovskiy, Russia	262km NW of Ozernovskiy, Russia	mww	6.7	1539429023560									
2	green	earthquake	M 5.6 - 128km SE of Kimbe, Papua New Guinea	128km SE of Kimbe, Papua New Guinea	mww	5.6	1539312723620									
3	green	earthquake	M 6.5 - 148km S of Severo-Kuril'sk, Russia	148km S of Severo-Kuril'sk, Russia	mww	6.5	1539213362130									
4	green	earthquake	M 6.2 - 94km SW of Kokopo, Papua New Guinea	94km SW of Kokopo, Papua New Guinea	mww	6.2	1539208835130									
	<code>alert</code>	<code>...</code>	<code>dmin</code>	<code>felt</code>	<code>...</code>	<code>mag</code>	<code>magType</code>	<code>...</code>	<code>place</code>	<code>...</code>	<code>time</code>	<code>title</code>	<code>tsunami</code>	<code>...</code>	<code>updated</code>	<code>url</code>
0	NaN	...	0.008693	NaN	...	1.35	ml	...	9km NE of Aguanga, CA	...	1539475168010	M 1.4 - 9km NE of Aguanga, CA	0	...	1539475395144	https...
1	NaN	...	0.020030	NaN	...	1.29	ml	...	9km NE of Aguanga, CA	...	1539475129610	M 1.3 - 9km NE of Aguanga, CA	0	...	1539475253925	https...
2	NaN	...	0.021370	28.0	...	3.42	ml	...	8km NE of Aguanga, CA	...	1539475062610	M 3.4 - 8km NE of Aguanga, CA	0	...	1539536756176	https...
3	NaN	...	0.026180	NaN	...	0.44	ml	...	9km NE of Aguanga, CA	...	1539474978070	M 0.4 - 9km NE of Aguanga, CA	0	...	1539475196167	https...
4	NaN	...	0.077990	NaN	...	2.16	md	...	10km NW of Avenal, CA	...	1539474716050	M 2.2 - 10km NW of Avenal, CA	0	...	1539477547926	https...

	alert	...	dmin	felt	...	mag	magType	...	place	...	time	title	tsunami	...	updated	url
9330	NaN	...	0.01865	NaN	...	1.10	ml	...	9km NE of Aguanga, CA	...	1537229545350	M 1.1 - 9km NE of Aguanga, CA	0	...	1537230211640	https...
9331	NaN	...	0.01698	NaN	...	0.66	ml	...	9km NE of Aguanga, CA	...	1537228864470	M 0.7 - 9km NE of Aguanga, CA	0	...	1537305830770	https...
	cdi	dmin	felt	gap	mag	...	sig	time	tsunami	tz	updated					
count	329.000000	6139.000000	329.000000	6164.000000	9331.000000	...	9332.000000	9.332000e+03	9332.000000	9331.000000	9.332000e+03					
mean	2.754711	0.544925	12.310030	121.506588	1.497345	...	56.899914	1.538284e+12	0.006537	-451.990140	1.538537e+12					
std	1.010637	2.214305	48.954944	72.962363	1.203347	...	91.872163	6.080306e+08	0.080589	231.752571	6.564135e+08					
min	0.000000	0.000648	0.000000	12.000000	-1.260000	...	0.000000	1.537229e+12	0.000000	-720.000000	1.537230e+12					
25%	2.000000	0.020425	1.000000	66.142500	0.720000	...	8.000000	1.537793e+12	0.000000	-540.000000	1.537996e+12					
50%	2.700000	0.059050	2.000000	105.000000	1.300000	...	26.000000	1.538245e+12	0.000000	-480.000000	1.538621e+12					
75%	3.300000	0.177250	5.000000	159.000000	1.900000	...	56.000000	1.538766e+12	0.000000	-480.000000	1.539110e+12					
max	8.400000	53.737000	580.000000	355.910000	7.500000	...	2015.000000	1.539475e+12	1.000000	720.000000	1.539537e+12					
	alert	code	detail	ids	magType	net	place	sources	status	title	type	types	...			
count	59	9332	9332	9332	9331	9332	9332	9332	9332	9332	9332	9332	9332	9332	9332	9332
unique	2	9332	9332	9332	10	14	5433	52	2	7807	5	42	42	42	42	9332
top	green	70628507	https://ear...	,pr201827...	ml	ak	10km NE of Aguanga, CA	,ak	reviewed	M 0.4 - 10km NE of Aguanga, CA	earthquake	,geoserve, origin, phase-data,	,geoserve, origin, phase-data,	,geoserve, origin, phase-data,	,geoserve, origin, phase-data,	https://ear...
freq	58	1	1	1	6803	3166	306	2981	7797	55	9081	5301	5301	5301	5301	1
	Method			Description						Data types						
	<code>count()</code>	The number of non-null observations						Any								
	<code>nunique()</code>	The number of unique values						Any								
	<code>sum()</code>	The total of the values						Numerical or Boolean								
	<code>mean()</code>	The average of the values						Numerical or Boolean								
	<code>median()</code>	The median of the values						Numerical								
	<code>min()</code>	The minimum of the values						Numerical								
	<code>idxmin()</code>	The index where the minimum values occurs						Numerical								
	<code>max()</code>	The maximum of the values						Numerical								
	<code>idxmax()</code>	The index where the maximum value occurs						Numerical								
	<code>abs()</code>	The absolute values of the data						Numerical								
	<code>std()</code>	The standard deviation						Numerical								
	<code>var()</code>	The variance						Numerical								
	<code>cov()</code>	The covariance between two <code>Series</code> , or a covariance matrix for all column combinations in a <code>DataFrame</code>						Numerical								
	<code>corr()</code>	The correlation between two <code>Series</code> , or a correlation matrix for all column combinations in a <code>DataFrame</code>						Numerical								
	<code>quantile()</code>	Calculates a specific quantile						Numerical								
	<code>cumsum()</code>	The cumulative sum						Numerical or Boolean								
	<code>cummin()</code>	The cumulative minimum						Numerical								
	<code>cummax()</code>	The cumulative maximum						Numerical								

Method	Description
<code>argmax() / argmin()</code>	Find the location of the maximum/minimum value in the index
<code>equals()</code>	Compare the index to another <code>Index</code> object for equality
<code>isin()</code>	Check if the index values are in a list of values and return an array of Booleans
<code>max() / min()</code>	Find the maximum/minimum value in the index
<code>nunique()</code>	Get the number of unique values in the index
<code>to_series()</code>	Create a <code>Series</code> object from the index
<code>unique()</code>	Find the unique values of the index
<code>value_counts()</code>	Create a frequency table for the unique values in the index

	mag	title
0	1.35	M 1.4 - 9km NE of Aguanga, CA
1	1.29	M 1.3 - 9km NE of Aguanga, CA
2	3.42	M 3.4 - 8km NE of Aguanga, CA
3	0.44	M 0.4 - 9km NE of Aguanga, CA
4	2.16	M 2.2 - 10km NW of Avenal, CA
...
9327	0.62	M 0.6 - 9km ENE of Mammoth Lakes, CA
9328	1.00	M 1.0 - 3km W of Julian, CA
9329	2.40	M 2.4 - 35km NNE of Hatillo, Puerto Rico
9330	1.10	M 1.1 - 9km NE of Aguanga, CA
9331	0.66	M 0.7 - 9km NE of Aguanga, CA

		title	time	mag	magType
0		M 1.4 - 9km NE of Aguanga, CA	1539475168010	1.35	ml
1		M 1.3 - 9km NE of Aguanga, CA	1539475129610	1.29	ml
2		M 3.4 - 8km NE of Aguanga, CA	1539475062610	3.42	ml
3		M 0.4 - 9km NE of Aguanga, CA	1539474978070	0.44	ml
4		M 2.2 - 10km NW of Avenal, CA	1539474716050	2.16	md
...	
9327		M 0.6 - 9km ENE of Mammoth Lakes, CA	1537230228060	0.62	md
9328		M 1.0 - 3km W of Julian, CA	1537230135130	1.00	ml
9329		M 2.4 - 35km NNE of Hatillo, Puerto Rico	1537229908180	2.40	md
9330		M 1.1 - 9km NE of Aguanga, CA	1537229545350	1.10	ml
9331		M 0.7 - 9km NE of Aguanga, CA	1537228864470	0.66	ml

	alert	...	dmin	felt	...	mag	magType	...	place	...	time	title	tsunami	...	updated	url
100	NaN	...	NaN	NaN	...	1.20	ml	...	25km NW of Ester, Alaska	...	1539435449480	M 1.2 - 25km NW of Ester, Alaska	0	...	1539443551010	https...
101	NaN	...	0.01355	NaN	...	0.59	md	...	8km ESE of Mammoth Lakes, CA	...	1539435391320	M 0.6 - 8km ESE of Mammoth Lakes, CA	0	...	1539439802162	https...
102	NaN	...	0.02987	NaN	...	1.33	ml	...	8km ENE of Aguanga, CA	...	1539435293090	M 1.3 - 8km ENE of Aguanga, CA	0	...	1539435940470	https...

		title	time
100		M 1.2 - 25km NW of Ester, Alaska	1539435449480
101		M 0.6 - 8km ESE of Mammoth Lakes, CA	1539435391320
102		M 1.3 - 8km ENE of Aguanga, CA	1539435293090

		title	mag
10		M 0.5 - 10km NE of Aguanga, CA	0.50
11		M 2.8 - 53km SE of Punta Cana, Dominican Republic	2.77
12		M 0.5 - 9km NE of Aguanga, CA	0.50
13		M 4.5 - 120km SSW of Banda Aceh, Indonesia	4.50
14		M 2.1 - 14km NW of Parkfield, CA	2.13
15		M 2.0 - 156km WNW of Haines Junction, Canada	2.00

		title	mag
10		M 0.5 - 10km NE of Aguanga, CA	0.50
11		M 2.8 - 53km SE of Punta Cana, Dominican Republic	2.77
12		M 0.5 - 9km NE of Aguanga, CA	0.50
13		M 4.5 - 120km SSW of Banda Aceh, Indonesia	4.50
14		M 2.1 - 14km NW of Parkfield, CA	2.13

	gap	ids	mag	magType
10	57.0	,ci37389162,	0.50	ml
11	186.0	,pr2018286010,	2.77	md
12	76.0	,ci37389146,	0.50	ml
13	157.0	,us1000hbti,	4.50	mb
14	71.0	,nc73096921,	2.13	md

	alert	...	dmin	felt	...	mag	magType	...	place	...	time	title	tsunami	...	updated	url
837	green	...	1.763	3.0	...	7.0	mww	...	117km E of Kimbe, Papua New Guinea	...	1539204500290	M 7.0 - 117km E of Kimbe, Papua New Guinea	1	...	1539378744253	https...
5263	red	...	1.589	18.0	...	7.5	mww	...	78km N of Palu, Indonesia	...	1538128963480	M 7.5 - 78km N of Palu, Indonesia	1	...	1539123134531	https...

	alert	mag	magType	title	tsunami	type
837	green	7.0	mww	M 7.0 - 117km E of Kimbe, Papua New Guinea	1	earthquake
5263	red	7.5	mww	M 7.5 - 78km N of Palu, Indonesia	1	earthquake

	alert	mag	magType	title	tsunami	type
5263	red	7.5	mww	M 7.5 - 78km N of Palu, Indonesia	1	earthquake

	alert	mag	magType	title	tsunami	type
36	NaN	5.0	mww	M 5.0 - 165km NNW of Flying Fish Cove, Christm...	1	earthquake
118	green	6.7	mww	M 6.7 - 262km NW of Ozernovskiy, Russia	1	earthquake
501	green	5.6	mww	M 5.6 - 128km SE of Kimbe, Papua New Guinea	1	earthquake
799	green	6.5	mww	M 6.5 - 148km S of Severo-Kuril'sk, Russia	1	earthquake
816	green	6.2	mww	M 6.2 - 94km SW of Kokopo, Papua New Guinea	1	earthquake
...
8561	NaN	5.4	mb	M 5.4 - 228km S of Taron, Papua New Guinea	1	earthquake
8624	NaN	5.1	mb	M 5.1 - 278km SE of Pondaguitan, Philippines	1	earthquake
9133	green	5.1	ml	M 5.1 - 64km SSW of Kaktovik, Alaska	1	earthquake
9175	NaN	5.2	mb	M 5.2 - 126km N of Dili, East Timor	1	earthquake
9304	NaN	5.1	mb	M 5.1 - 34km NW of Finschhafen, Papua New Guinea	1	earthquake

	alert	mag	magType		title	tsunami	type
1015	green	5.0	ml	M 5.0 - 61km SSW of Chignik Lake, Alaska		1	earthquake
1273	green	4.0	ml	M 4.0 - 71km SW of Kaktovik, Alaska		1	earthquake
1795	green	4.0	ml	M 4.0 - 60km WNW of Valdez, Alaska		1	earthquake
2752	green	4.0	ml	M 4.0 - 67km SSW of Kaktovik, Alaska		1	earthquake
3260	green	3.9	ml	M 3.9 - 44km N of North Nenana, Alaska		0	earthquake
4101	green	4.2	ml	M 4.2 - 131km NNW of Arctic Village, Alaska		0	earthquake
6897	green	3.8	ml	M 3.8 - 80km SSW of Kaktovik, Alaska		0	earthquake
8524	green	3.8	ml	M 3.8 - 69km SSW of Kaktovik, Alaska		0	earthquake
9133	green	5.1	ml	M 5.1 - 64km SSW of Kaktovik, Alaska		1	earthquake

	alert	mag	magType		title	tsunami	type
1465	green	3.83	mw	M 3.8 - 109km WNW of Trinidad, CA		0	earthquake
2414	green	3.83	mw	M 3.8 - 5km SW of Tres Pinos, CA		1	earthquake

	alert	mag	magType		title	tsunami	type
118	green	6.7	mww	M 6.7 - 262km NW of Ozernovskiy, Russia		1	earthquake
799	green	6.5	mww	M 6.5 - 148km S of Severo-Kuril'sk, Russia		1	earthquake
837	green	7.0	mww	M 7.0 - 117km E of Kimbe, Papua New Guinea		1	earthquake
4363	green	6.7	mww	M 6.7 - 263km NNE of Ndri Island, Fiji		1	earthquake
5263	red	7.5	mww	M 7.5 - 78km N of Palu, Indonesia		1	earthquake

	alert	mag	magType		title	tsunami	type
995	NaN	3.35	mw	M 3.4 - 9km WNW of Cobb, CA		0	earthquake
1465	green	3.83	mw	M 3.8 - 109km WNW of Trinidad, CA		0	earthquake
2414	green	3.83	mw	M 3.8 - 5km SW of Tres Pinos, CA		1	earthquake
4988	green	4.41	mw	M 4.4 - 1km SE of Delta, B.C., MX		1	earthquake
6307	green	5.80	mwb	M 5.8 - 297km NNE of Ndri Island, Fiji		0	earthquake
8257	green	5.70	mwb	M 5.7 - 175km SSE of Lambasa, Fiji		0	earthquake

	alert	mag	magType		title	tsunami	type
2409	NaN	-1.26	ml	M -1.3 - 41km ENE of Adak, Alaska		0	earthquake
5263	red	7.50	mww	M 7.5 - 78km N of Palu, Indonesia		1	earthquake

	alert	mag	magType	place	time	title	tsunami	source			
0	NaN	1.35	ml	9km NE of Aguanga, CA	1539475168010	M 1.4 - 9km NE of Aguanga, CA	0	USGS API			
1	NaN	1.29	ml	9km NE of Aguanga, CA	1539475129610	M 1.3 - 9km NE of Aguanga, CA	0	USGS API			
2	NaN	3.42	ml	8km NE of Aguanga, CA	1539475062610	M 3.4 - 8km NE of Aguanga, CA	0	USGS API			
3	NaN	0.44	ml	9km NE of Aguanga, CA	1539474978070	M 0.4 - 9km NE of Aguanga, CA	0	USGS API			
4	NaN	2.16	md	10km NW of Avenal, CA	1539474716050	M 2.2 - 10km NW of Avenal, CA	0	USGS API			
	alert	mag	magType	place	time	title	tsunami	source	mag_negative		
0	NaN	1.35	ml	9km NE of Aguanga, CA	1539475168010	M 1.4 - 9km NE of Aguanga, CA	0	USGS API	False		
1	NaN	1.29	ml	9km NE of Aguanga, CA	1539475129610	M 1.3 - 9km NE of Aguanga, CA	0	USGS API	False		
2	NaN	3.42	ml	8km NE of Aguanga, CA	1539475062610	M 3.4 - 8km NE of Aguanga, CA	0	USGS API	False		
3	NaN	0.44	ml	9km NE of Aguanga, CA	1539474978070	M 0.4 - 9km NE of Aguanga, CA	0	USGS API	False		
4	NaN	2.16	md	10km NW of Avenal, CA	1539474716050	M 2.2 - 10km NW of Avenal, CA	0	USGS API	False		
	alert	mag	magType	place	time	title	tsunami	source	parsed_place	in_ca	in_alaska
7207	NaN	4.80	mwr	73km SSW of Masachapa, Nicaragua	1537749595210	M 4.8 - 73km SSW of Masachapa, Nicaragua	0	USGS API	False	Nicaragua	False
4755	NaN	1.09	ml	28km NNW of Packwood, Washington	1538227540460	M 1.1 - 28km NNW of Packwood, Washington	0	USGS API	False	Washington	False
4595	NaN	1.80	ml	77km SSW of Kaktovik, Alaska	1538259609862	M 1.8 - 77km SSW of Kaktovik, Alaska	0	USGS API	False	Alaska	False
3566	NaN	1.50	ml	102km NW of Arctic Village, Alaska	1538464751822	M 1.5 - 102km NW of Arctic Village, Alaska	0	USGS API	False	Alaska	False
2182	NaN	0.90	ml	26km ENE of Pine Valley, CA	1538801713880	M 0.9 - 26km ENE of Pine Valley, CA	0	USGS API	False	California	True
	alert	mag	magType	place	time	title	tsunami	source	mag_negative	neither	
7207	NaN	4.80	mwr	73km SSW of Masachapa, Nicaragua	1537749595210	M 4.8 - 73km SSW of Masachapa, Nicaragua	0	USGS API	False	Nicaragua	False
4755	NaN	1.09	ml	28km NNW of Packwood, Washington	1538227540460	M 1.1 - 28km NNW of Packwood, Washington	0	USGS API	False	Washington	False
4595	NaN	1.80	ml	77km SSW of Kaktovik, Alaska	1538259609862	M 1.8 - 77km SSW of Kaktovik, Alaska	0	USGS API	False	Alaska	True
3566	NaN	1.50	ml	102km NW of Arctic Village, Alaska	1538464751822	M 1.5 - 102km NW of Arctic Village, Alaska	0	USGS API	False	Alaska	True
2182	NaN	0.90	ml	26km ENE of Pine Valley, CA	1538801713880	M 0.9 - 26km ENE of Pine Valley, CA	0	USGS API	False	California	False

	alert	mag	magType	place	time	title	tsunami	source	mag_negative	parsed_place	felt	ids	tz
0	NaN	1.35	ml	9km NE of Aguanga, CA	1539475168010	M 1.4 - 9km NE of Aguanga, CA	0	USGS API	False	California	Nan	,ci37389218,	-480.0
1	NaN	1.29	ml	9km NE of Aguanga, CA	1539475129610	M 1.3 - 9km NE of Aguanga, CA	0	USGS API	False	California	Nan	,ci37389202,	-480.0
	alert	mag	magType	place	time	title	tsunami	source	mag_negative	parsed_place	felt	ids	tz
0	NaN	1.35	ml	9km NE of Aguanga, CA	1.539475e+12	M 1.4 - 9km NE of Aguanga, CA	0.0	USGS API	False	California	Nan	Nan	Nan
1	NaN	1.29	ml	9km NE of Aguanga, CA	1.539475e+12	M 1.3 - 9km NE of Aguanga, CA	0.0	USGS API	False	California	Nan	Nan	Nan
1539475129610	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	,ci37389202,	-480.0
1539475168010	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	,ci37389218,	-480.0
	alert	mag	magType	place	time	title	tsunami	source	mag_negative	parsed_place			
36	NaN	5.00	mww	165km NNW of Flying Fish Cove, Christmas Island	1539459504090	M 5.0 - 165km NNW of Flying Fish Cove, Christm...	1	USGS API	False	Christmas Island			
118	green	6.70	mww	262km NW of Ozernovskiy, Russia	1539429023560	M 6.7 - 262km NW of Ozernovskiy, Russia	1	USGS API	False	Russia			
0	NaN	1.35	ml	9km NE of Aguanga, CA	1539475168010	M 1.4 - 9km NE of Aguanga, CA	0	USGS API	False	California			
1	NaN	1.29	ml	9km NE of Aguanga, CA	1539475129610	M 1.3 - 9km NE of Aguanga, CA	0	USGS API	False	California			
	alert	mag	magType	place	time	title	tsunami	source	mag_negative	parsed_place			
0	NaN	5.00	mww	165km NNW of Flying Fish Cove, Christmas Island	1539459504090	M 5.0 - 165km NNW of Flying Fish Cove, Christm...	1	USGS API	False	Christmas Island			
1	green	6.70	mww	262km NW of Ozernovskiy, Russia	1539429023560	M 6.7 - 262km NW of Ozernovskiy, Russia	1	USGS API	False	Russia			
2	NaN	1.35	ml	9km NE of Aguanga, CA	1539475168010	M 1.4 - 9km NE of Aguanga, CA	0	USGS API	False	California			
3	NaN	1.29	ml	9km NE of Aguanga, CA	1539475129610	M 1.3 - 9km NE of Aguanga, CA	0	USGS API	False	California			
	alert	mag	magType	place	time	title	tsunami	source	mag_negative	parsed_place			
39	NaN	-0.10	ml	6km NW of Lemmon Valley, Nevada	1539458844506	M -0.1 - 6km NW of Lemmon Valley, Nevada	0	Nevada					
49	NaN	-0.10	ml	6km NW of Lemmon Valley, Nevada	1539455017464	M -0.1 - 6km NW of Lemmon Valley, Nevada	0	Nevada					
135	NaN	-0.40	ml	10km SSE of Beatty, Nevada	1539422175717	M -0.4 - 10km SSE of Beatty, Nevada	0	Nevada					
161	NaN	-0.02	md	20km SSE of Ronan, Montana	1539412475360	M -0.0 - 20km SSE of Ronan, Montana	0	Montana					
198	NaN	-0.20	ml	60km N of Pahrump, Nevada	1539398340822	M -0.2 - 60km N of Pahrump, Nevada	0	Nevada					

	alert	mag	magType	place	time	title	tsunami	parsed_place
2	NaN	3.42	ml	8km NE of Aguanga, CA	1539475062610	M 3.4 - 8km NE of Aguanga, CA	0	California
3	NaN	0.44	ml	9km NE of Aguanga, CA	1539474978070	M 0.4 - 9km NE of Aguanga, CA	0	California

	alert	mag	time	title	tsunami
0	NaN	1.35	1539475168010	M 1.4 - 9km NE of Aguanga, CA	0
1	NaN	1.29	1539475129610	M 1.3 - 9km NE of Aguanga, CA	0
2	NaN	3.42	1539475062610	M 3.4 - 8km NE of Aguanga, CA	0
3	NaN	0.44	1539474978070	M 0.4 - 9km NE of Aguanga, CA	0
4	NaN	2.16	1539474716050	M 2.2 - 10km NW of Avenal, CA	0

Chapter 3: Data Wrangling with Pandas

File	Description	Source
<code>bitcoin.csv</code>	Daily opening, high, low, and closing price of bitcoin, along with volume traded and market capitalization for 2017 through 2018.	CoinMarketCap
<code>dirty_data.csv</code>	2018 weather data for New York City, manipulated to introduce data issues.	Modified version of the data from the NCEI API's GHCND dataset.
<code>long_data.csv</code>	Long format temperature data for New York City in October 2018 from the Boonton 1 station, containing daily temperature at time of observation, minimum temperature, and maximum temperature.	The NCEI API's GHCND dataset
<code>nyc_temperatures.csv</code>	Temperature data for New York City in October 2018 measured from LaGuardia airport, containing daily minimum, maximum, and average temperature.	The NCEI API's GHCND dataset
<code>sp500.csv</code>	Daily opening, high, low, and closing price of the S&P 500 stock index, along with volume traded and adjusted close for 2017 through 2018.	The <code>stock_analysis</code> package (see <i>Chapter 7, Financial Analysis – Bitcoin and the Stock Market</i>).
<code>wide_data.csv</code>	Wide format temperature data for New York City in October 2018 from the Boonton 1 station, containing daily temperature at time of observation, minimum temperature, and maximum temperature.	The NCEI API's GHCND dataset

variables

	date	TMAX	TMIN	TOBS
0	2018-10-01	21.1	8.9	13.9
1	2018-10-02	23.9	13.9	17.2
2	2018-10-03	25.0	15.6	16.1
3	2018-10-04	22.8	11.7	11.7
4	2018-10-05	23.3	11.7	18.9
5	2018-10-06	20.0	13.3	16.1

repeated values for `date` column

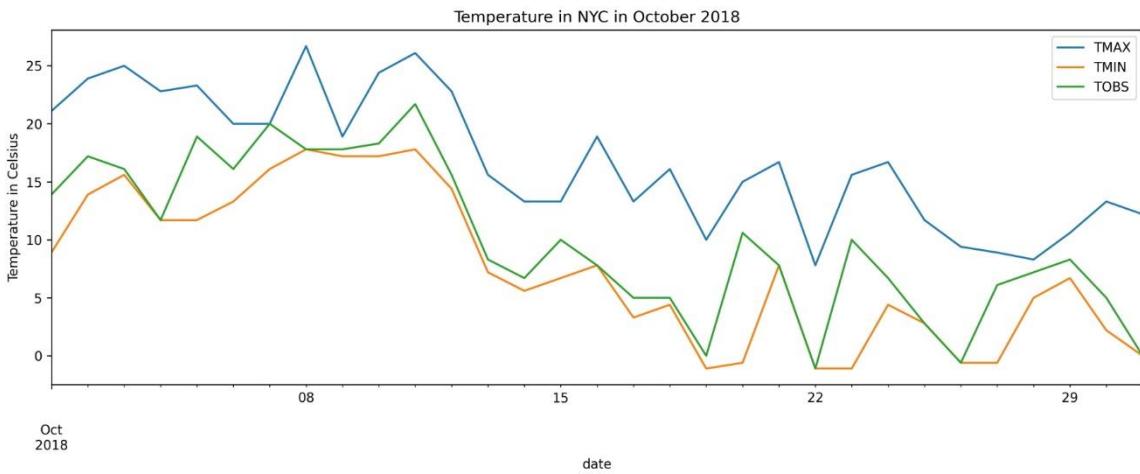
variable names variable values

	date	datatype	value
0	2018-10-01	TMAX	21.1
1	2018-10-01	TMIN	8.9
2	2018-10-01	TOBS	13.9
3	2018-10-02	TMAX	23.9
4	2018-10-02	TMIN	13.9
5	2018-10-02	TOBS	17.2

observations

	date	TMAX	TMIN	TOBS
0	2018-10-01	21.1	8.9	13.9
1	2018-10-02	23.9	13.9	17.2
2	2018-10-03	25.0	15.6	16.1
3	2018-10-04	22.8	11.7	11.7
4	2018-10-05	23.3	11.7	18.9
5	2018-10-06	20.0	13.3	16.1

	date	TMAX	TMIN	TOBS
count	31	31.000000	31.000000	31.000000
mean	2018-10-16 00:00:00	16.829032	7.561290	10.022581
min	2018-10-01 00:00:00	7.800000	-1.100000	-1.100000
25%	2018-10-08 12:00:00	12.750000	2.500000	5.550000
50%	2018-10-16 00:00:00	16.100000	6.700000	8.300000
75%	2018-10-23 12:00:00	21.950000	13.600000	16.100000
max	2018-10-31 00:00:00	26.700000	17.800000	21.700000
std	NaN	5.714962	6.513252	6.596550

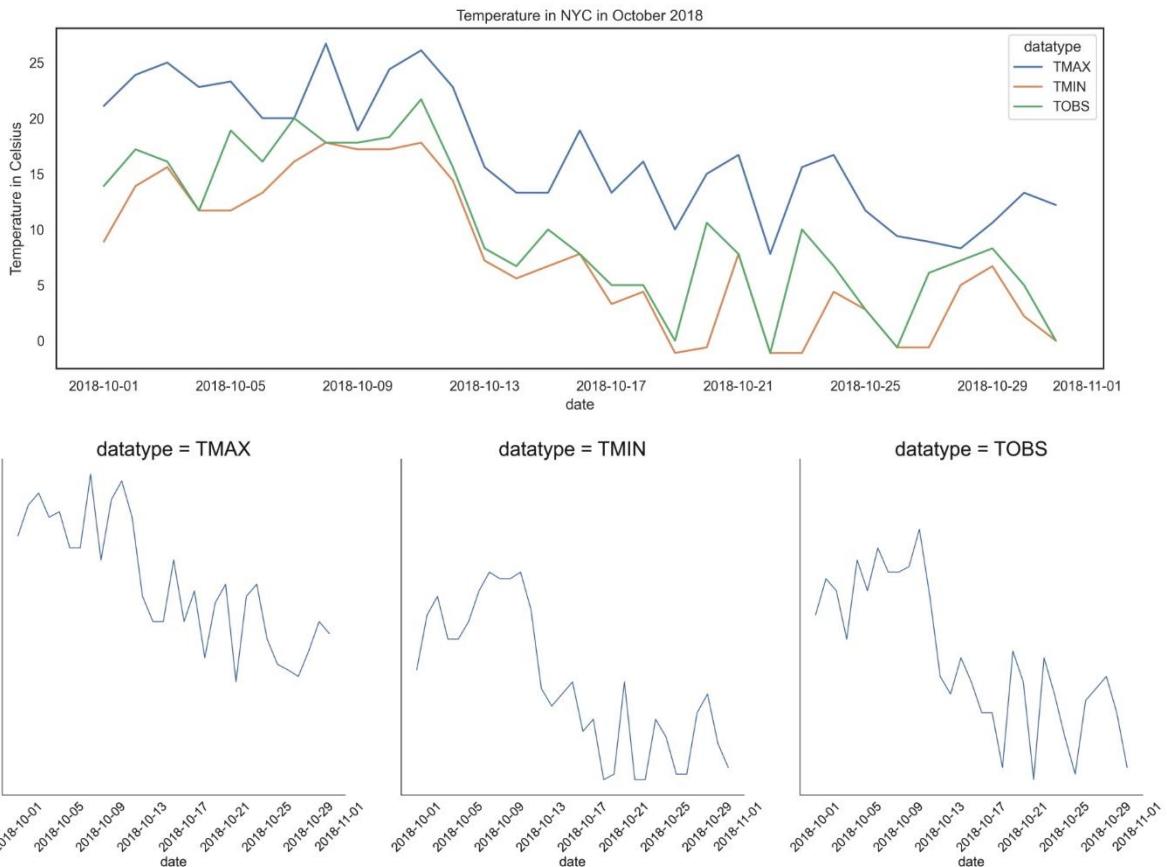


date datatype value

0	2018-10-01	TMAX	21.1
1	2018-10-01	TMIN	8.9
2	2018-10-01	TOBS	13.9
3	2018-10-02	TMAX	23.9
4	2018-10-02	TMIN	13.9
5	2018-10-02	TOBS	17.2

date datatype value

count	93	93	93.000000
unique	NaN	3	NaN
top	NaN	TOBS	NaN
freq	NaN	31	NaN
mean	2018-10-16 00:00:00	NaN	11.470968
min	2018-10-01 00:00:00	NaN	-1.100000
25%	2018-10-08 00:00:00	NaN	6.700000
50%	2018-10-16 00:00:00	NaN	11.700000
75%	2018-10-24 00:00:00	NaN	17.200000
max	2018-10-31 00:00:00	NaN	26.700000
std	NaN	NaN	7.362354



	date	datatype	station	attributes	value
0	2018-10-01T00:00:00	TMAX	GHCND:USW00094728	,,W,2400	24.4
1	2018-10-01T00:00:00	TMIN	GHCND:USW00094728	,,W,2400	17.2
2	2018-10-02T00:00:00	TMAX	GHCND:USW00094728	,,W,2400	25.0
3	2018-10-02T00:00:00	TMIN	GHCND:USW00094728	,,W,2400	18.3
4	2018-10-03T00:00:00	TMAX	GHCND:USW00094728	,,W,2400	23.3

	date	datatype	station	attributes	value
0	2018-10-01T00:00:00	TAVG	GHCND:USW00014732	H,,S,	21.2
1	2018-10-01T00:00:00	TMAX	GHCND:USW00014732	,,W,2400	25.6
2	2018-10-01T00:00:00	TMIN	GHCND:USW00014732	,,W,2400	18.3
3	2018-10-02T00:00:00	TAVG	GHCND:USW00014732	H,,S,	22.7
4	2018-10-02T00:00:00	TMAX	GHCND:USW00014732	,,W,2400	26.1

	datatype	station	attributes	value
	date			
2018-10-01 00:00:00-05:00	TAVG	GHCND:USW00014732	H,,S,	21.2
2018-10-01 00:00:00-05:00	TMAX	GHCND:USW00014732	,,W,2400	25.6
2018-10-01 00:00:00-05:00	TMIN	GHCND:USW00014732	,,W,2400	18.3
2018-10-02 00:00:00-05:00	TAVG	GHCND:USW00014732	H,,S,	22.7
2018-10-02 00:00:00-05:00	TMAX	GHCND:USW00014732	,,W,2400	26.1

	datatype	station	attributes	value
date				
2018-10-01 05:00:00+00:00	TAVG	GHCND:USW00014732	H,,S,	21.2
2018-10-01 05:00:00+00:00	TMAX	GHCND:USW00014732	,,W,2400	25.6
2018-10-01 05:00:00+00:00	TMIN	GHCND:USW00014732	,,W,2400	18.3
2018-10-02 05:00:00+00:00	TAVG	GHCND:USW00014732	H,,S,	22.7
2018-10-02 05:00:00+00:00	TMAX	GHCND:USW00014732	,,W,2400	26.1

	date	datatype	station	flags	temp_C	temp_F
0	2018-10-01	TAVG	GHCND:USW00014732	H,,S,	21.2	70.16
1	2018-10-01	TMAX	GHCND:USW00014732	,,W,2400	25.6	78.08
2	2018-10-01	TMIN	GHCND:USW00014732	,,W,2400	18.3	64.94
3	2018-10-02	TAVG	GHCND:USW00014732	H,,S,	22.7	72.86
4	2018-10-02	TMAX	GHCND:USW00014732	,,W,2400	26.1	78.98

	date	datatype	station	flags	temp_C	temp_C_whole	temp_F	temp_F_whole
0	2018-10-01	TAVG	GHCND:USW00014732	H,,S,	21.2	21	70.16	70
1	2018-10-01	TMAX	GHCND:USW00014732	,,W,2400	25.6	25	78.08	78
2	2018-10-01	TMIN	GHCND:USW00014732	,,W,2400	18.3	18	64.94	64
3	2018-10-02	TAVG	GHCND:USW00014732	H,,S,	22.7	22	72.86	72
4	2018-10-02	TMAX	GHCND:USW00014732	,,W,2400	26.1	26	78.98	78

	datatype	station
	count	93
	unique	1
	top	TAVG GHCND:USW00014732
	freq	31

	date	datatype	station	flags	temp_C	temp_C_whole	temp_F	temp_F_whole
19	2018-10-07	TMAX	GHCND:USW00014732	,,W,2400	27.8	27	82.04	82
28	2018-10-10	TMAX	GHCND:USW00014732	,,W,2400	27.8	27	82.04	82
31	2018-10-11	TMAX	GHCND:USW00014732	,,W,2400	26.7	26	80.06	80
10	2018-10-04	TMAX	GHCND:USW00014732	,,W,2400	26.1	26	78.98	78
4	2018-10-02	TMAX	GHCND:USW00014732	,,W,2400	26.1	26	78.98	78
1	2018-10-01	TMAX	GHCND:USW00014732	,,W,2400	25.6	25	78.08	78
25	2018-10-09	TMAX	GHCND:USW00014732	,,W,2400	25.6	25	78.08	78
7	2018-10-03	TMAX	GHCND:USW00014732	,,W,2400	25.0	25	77.00	77
13	2018-10-05	TMAX	GHCND:USW00014732	,,W,2400	22.8	22	73.04	73
22	2018-10-08	TMAX	GHCND:USW00014732	,,W,2400	22.8	22	73.04	73

	date	datatype	station	flags	temp_C	temp_C_whole	temp_F	temp_F_whole
19	2018-10-07	TMAX	GHCND:USW00014732	,,W,2400	27.8	27	82.04	82
28	2018-10-10	TMAX	GHCND:USW00014732	,,W,2400	27.8	27	82.04	82
31	2018-10-11	TMAX	GHCND:USW00014732	,,W,2400	26.7	26	80.06	80
4	2018-10-02	TMAX	GHCND:USW00014732	,,W,2400	26.1	26	78.98	78
10	2018-10-04	TMAX	GHCND:USW00014732	,,W,2400	26.1	26	78.98	78
1	2018-10-01	TMAX	GHCND:USW00014732	,,W,2400	25.6	25	78.08	78
25	2018-10-09	TMAX	GHCND:USW00014732	,,W,2400	25.6	25	78.08	78
7	2018-10-03	TMAX	GHCND:USW00014732	,,W,2400	25.0	25	77.00	77
13	2018-10-05	TMAX	GHCND:USW00014732	,,W,2400	22.8	22	73.04	73
22	2018-10-08	TMAX	GHCND:USW00014732	,,W,2400	22.8	22	73.04	73

	date	datatype	station	flags	temp_C	temp_C_whole	temp_F	temp_F_whole
27	2018-10-10	TAVG	GHCND:USW00014732	H,,S,	23.8	23	74.84	74
30	2018-10-11	TAVG	GHCND:USW00014732	H,,S,	23.4	23	74.12	74
18	2018-10-07	TAVG	GHCND:USW00014732	H,,S,	22.8	22	73.04	73
3	2018-10-02	TAVG	GHCND:USW00014732	H,,S,	22.7	22	72.86	72
6	2018-10-03	TAVG	GHCND:USW00014732	H,,S,	21.8	21	71.24	71
24	2018-10-09	TAVG	GHCND:USW00014732	H,,S,	21.8	21	71.24	71
9	2018-10-04	TAVG	GHCND:USW00014732	H,,S,	21.3	21	70.34	70
0	2018-10-01	TAVG	GHCND:USW00014732	H,,S,	21.2	21	70.16	70
21	2018-10-08	TAVG	GHCND:USW00014732	H,,S,	20.9	20	69.62	69
12	2018-10-05	TAVG	GHCND:USW00014732	H,,S,	20.3	20	68.54	68

datatype	date	flags	station	temp_C	temp_C_whole	temp_F	temp_F_whole	
0	TAVG	2018-10-01	H,,S,	GHCND:USW00014732	21.2	21	70.16	70
1	TMAX	2018-10-01	,,W,2400	GHCND:USW00014732	25.6	25	78.08	78
2	TMIN	2018-10-01	,,W,2400	GHCND:USW00014732	18.3	18	64.94	64
3	TAVG	2018-10-02	H,,S,	GHCND:USW00014732	22.7	22	72.86	72
4	TMAX	2018-10-02	,,W,2400	GHCND:USW00014732	26.1	26	78.98	78

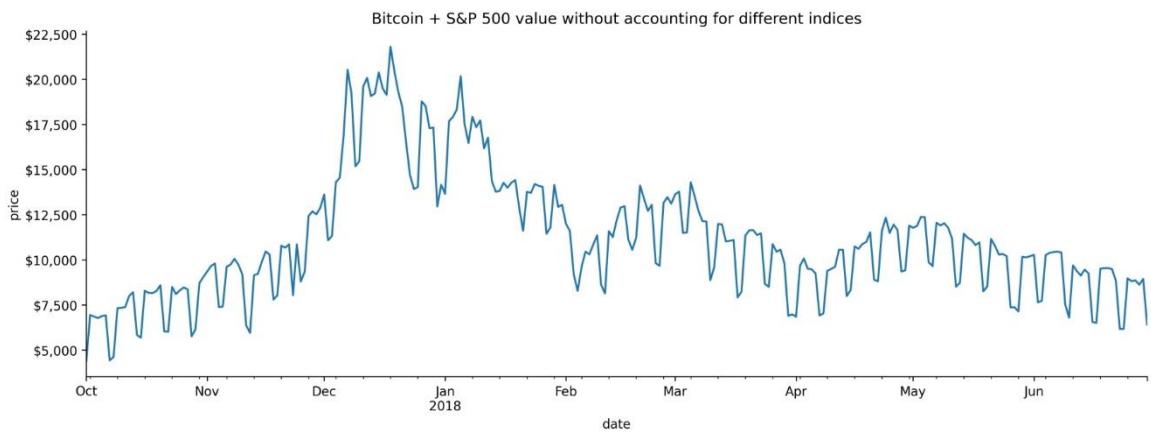
datatype	date	station	flags	temp_C	temp_C_whole	temp_F	temp_F_whole
2018-10-01	TAVG	GHCND:USW00014732	H,,S,	21.2	21	70.16	70
2018-10-01	TMAX	GHCND:USW00014732	,,W,2400	25.6	25	78.08	78
2018-10-01	TMIN	GHCND:USW00014732	,,W,2400	18.3	18	64.94	64
2018-10-02	TAVG	GHCND:USW00014732	H,,S,	22.7	22	72.86	72
2018-10-02	TMAX	GHCND:USW00014732	,,W,2400	26.1	26	78.98	78

	datatype	station	flags	temp_C	temp_C_whole	temp_F	temp_F_whole
	date						
2018-10-11	TAVG	GHCND:USW00014732	H,,S,	23.4	23	74.12	74
2018-10-11	TMAX	GHCND:USW00014732	,,W,2400	26.7	26	80.06	80
2018-10-11	TMIN	GHCND:USW00014732	,,W,2400	21.7	21	71.06	71
2018-10-12	TAVG	GHCND:USW00014732	H,,S,	18.3	18	64.94	64
2018-10-12	TMAX	GHCND:USW00014732	,,W,2400	22.2	22	71.96	71
2018-10-12	TMIN	GHCND:USW00014732	,,W,2400	12.2	12	53.96	53

	date	datatype	station	flags	temp_C	temp_C_whole	temp_F	temp_F_whole
	date	high	low	open	close	volume	day_of_week	
0	2018-10-11	TAVG	GHCND:USW00014732	H,,S,	23.4	23	74.12	74
1	2018-10-11	TMAX	GHCND:USW00014732	,,W,2400	26.7	26	80.06	80
2	2018-10-11	TMIN	GHCND:USW00014732	,,W,2400	21.7	21	71.06	71
3	2018-10-12	TAVG	GHCND:USW00014732	H,,S,	18.3	18	64.94	64
4	2018-10-12	TMAX	GHCND:USW00014732	,,W,2400	22.2	22	71.96	71
5	2018-10-12	TMIN	GHCND:USW00014732	,,W,2400	12.2	12	53.96	53

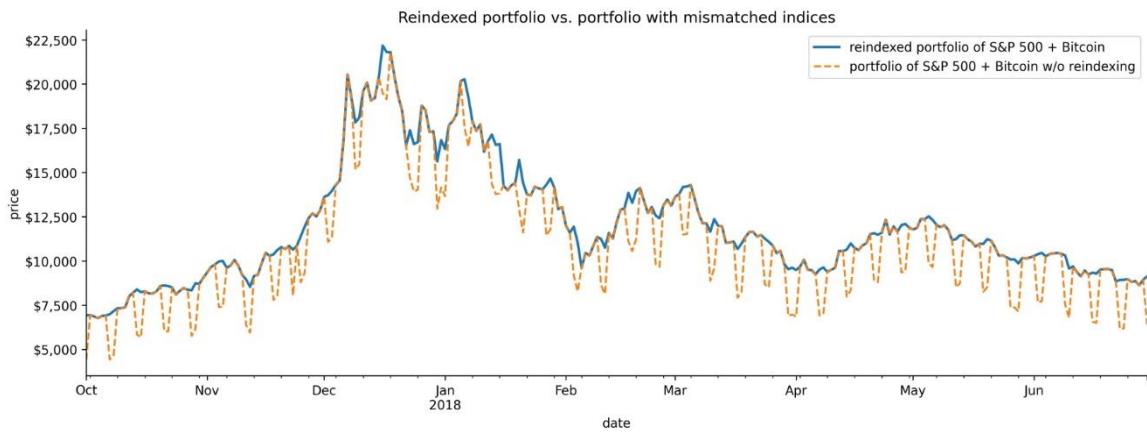
	date	high	low	open	close	volume	day_of_week
2017-01-03	2263.879883	2245.129883	2251.570068	2257.830078	3770530000	Tuesday	
2017-01-04	2272.820068	2261.600098	2261.600098	2270.750000	3764890000	Wednesday	
2017-01-05	2271.500000	2260.449951	2268.179932	2269.000000	3761820000	Thursday	
2017-01-06	2282.100098	2264.060059	2271.139893	2276.979980	3339890000	Friday	
2017-01-09	2275.489990	2268.899902	2273.590088	2268.899902	3217610000	Monday	
2017-01-10	2279.270020	2265.270020	2269.719971	2268.899902	3638790000	Tuesday	
2017-01-11	2275.320068	2260.830078	2268.600098	2275.320068	3620410000	Wednesday	
2017-01-12	2271.780029	2254.250000	2271.139893	2270.439941	3462130000	Thursday	
2017-01-13	2278.679932	2271.510010	2272.739990	2274.639893	3081270000	Friday	
2017-01-17	2272.080078	2262.810059	2269.139893	2267.889893	3584990000	Tuesday	

	high	low	open	close	volume	day_of_week
date						
2017-01-01	1003.080000	958.700000	963.660000	998.330000	147775008	Sunday
2017-01-02	1031.390000	996.700000	998.620000	1021.750000	222184992	Monday
2017-01-03	3307.959883	3266.729883	3273.170068	3301.670078	3955698000	Tuesday
2017-01-04	3432.240068	3306.000098	3306.000098	3425.480000	4109835984	Wednesday
2017-01-05	3462.600000	3170.869951	3424.909932	3282.380000	4272019008	Thursday
2017-01-06	3328.910098	3148.000059	3285.379893	3179.179980	3691766000	Friday
2017-01-07	908.590000	823.560000	903.490000	908.590000	279550016	Saturday
2017-01-08	942.720000	887.250000	908.170000	911.200000	158715008	Sunday
2017-01-09	3189.179990	3148.709902	3186.830088	3171.729902	3359486992	Monday
2017-01-10	3194.140020	3166.330020	3172.159971	3176.579902	3754598000	Tuesday



	high	low	open	close	volume	day_of_week
date						
2017-01-01	NaN	NaN	NaN	NaN	NaN	Sunday
2017-01-02	NaN	NaN	NaN	NaN	NaN	Monday
2017-01-03	2263.879883	2245.129883	2251.570068	2257.830078	3.770530e+09	Tuesday
2017-01-04	2272.820068	2261.600098	2261.600098	2270.750000	3.764890e+09	Wednesday
2017-01-05	2271.500000	2260.449951	2268.179932	2269.000000	3.761820e+09	Thursday
2017-01-06	2282.100098	2264.060059	2271.139893	2276.979980	3.339890e+09	Friday
2017-01-07	2282.100098	2264.060059	2271.139893	2276.979980	3.339890e+09	Saturday
2017-01-08	2282.100098	2264.060059	2271.139893	2276.979980	3.339890e+09	Sunday
2017-01-09	2275.489990	2268.899902	2273.590088	2268.899902	3.217610e+09	Monday
2017-01-10	2279.270020	2265.270020	2269.719971	2268.899902	3.638790e+09	Tuesday

	high	low	open	close	volume	day_of_week
date						
2017-01-01	NaN	NaN	NaN	NaN	0.000000e+00	Sunday
2017-01-02	NaN	NaN	NaN	NaN	0.000000e+00	Monday
2017-01-03	2263.879883	2245.129883	2251.570068	2257.830078	3.770530e+09	Tuesday
2017-01-04	2272.820068	2261.600098	2261.600098	2270.750000	3.764890e+09	Wednesday
2017-01-05	2271.500000	2260.449951	2268.179932	2269.000000	3.761820e+09	Thursday
2017-01-06	2282.100098	2264.060059	2271.139893	2276.979980	3.339890e+09	Friday
2017-01-07	2276.979980	2276.979980	2276.979980	2276.979980	0.000000e+00	Saturday
2017-01-08	2276.979980	2276.979980	2276.979980	2276.979980	0.000000e+00	Sunday
2017-01-09	2275.489990	2268.899902	2273.590088	2268.899902	3.217610e+09	Monday
2017-01-10	2279.270020	2265.270020	2269.719971	2268.899902	3.638790e+09	Tuesday



	ticker	date	high	low	open	close	volume
0	AAPL	2018-01-02	43.075001	42.314999	42.540001	43.064999	102223600
0	AMZN	2018-01-02	1190.000000	1170.510010	1172.000000	1189.010010	2694500
0	FB	2018-01-02	181.580002	177.550003	177.679993	181.419998	18151900
0	GOOG	2018-01-02	1066.939941	1045.229980	1048.339966	1065.000000	1237600
0	NFLX	2018-01-02	201.649994	195.419998	196.100006	201.070007	10966900

	datatype	date	temp_C	temp_F
0	TMAX	2018-10-01	21.1	69.98
1	TMIN	2018-10-01	8.9	48.02
2	TOBS	2018-10-01	13.9	57.02
3	TMAX	2018-10-02	23.9	75.02
4	TMIN	2018-10-02	13.9	57.02

date	2018-10-01	2018-10-01	2018-10-01	2018-10-02	2018-10-02	2018-10-02
datatype	TMAX	TMIN	TOBS	TMAX	TMIN	TOBS
temp_C	21.10	8.90	13.90	23.90	13.90	17.20
temp_F	69.98	48.02	57.02	75.02	57.02	62.96

datatype	TMAX	TMIN	TOBS
----------	------	------	------

date			
------	--	--	--

2018-10-01	21.1	8.9	13.9
2018-10-02	23.9	13.9	17.2
2018-10-03	25.0	15.6	16.1
2018-10-04	22.8	11.7	11.7
2018-10-05	23.3	11.7	18.9

datatype	TMAX	TMIN	TOBS
----------	------	------	------

count	31.000000	31.000000	31.000000
mean	16.829032	7.561290	10.022581
std	5.714962	6.513252	6.596550
min	7.800000	-1.100000	-1.100000
25%	12.750000	2.500000	5.550000
50%	16.100000	6.700000	8.300000
75%	21.950000	13.600000	16.100000
max	26.700000	17.800000	21.700000

temp_C	temp_F
--------	--------

datatype	TMAX	TMIN	TOBS	TMAX	TMIN	TOBS
----------	------	------	------	------	------	------

date						
2018-10-01	21.1	8.9	13.9	69.98	48.02	57.02
2018-10-02	23.9	13.9	17.2	75.02	57.02	62.96
2018-10-03	25.0	15.6	16.1	77.00	60.08	60.98
2018-10-04	22.8	11.7	11.7	73.04	53.06	53.06
2018-10-05	23.3	11.7	18.9	73.94	53.06	66.02

temp_C	temp_F
--------	--------

date	datatype
------	----------

2018-10-01	TMAX	21.1	69.98
	TMIN	8.9	48.02
	TOBS	13.9	57.02
2018-10-02	TMAX	23.9	75.02
	TMIN	13.9	57.02

	temp_C				temp_F		
datatype	TMAX	TMIN	TOBS	TMAX	TMIN	TOBS	
date							
2018-10-01	21.1	8.9	13.9	69.98	48.02	57.02	
2018-10-02	23.9	13.9	17.2	75.02	57.02	62.96	
2018-10-03	25.0	15.6	16.1	77.00	60.08	60.98	
2018-10-04	22.8	11.7	11.7	73.04	53.06	53.06	
2018-10-05	23.3	11.7	18.9	73.94	53.06	66.02	

	temp_C temp_F	
date	datatype	
2018-10-01	TAVG	10.0 50.00
	TMAX	21.1 69.98
	TMIN	8.9 48.02
	TOBS	13.9 57.02
2018-10-02	TMAX	23.9 75.02
	TMIN	13.9 57.02
	TOBS	17.2 62.96

	temp_C				temp_F			
datatype	TAVG	TMAX	TMIN	TOBS	TAVG	TMAX	TMIN	TOBS
date								
2018-10-01	10.0	21.1	8.9	13.9	50.0	69.98	48.02	57.02
2018-10-02	NaN	23.9	13.9	17.2	NaN	75.02	57.02	62.96
2018-10-03	NaN	25.0	15.6	16.1	NaN	77.00	60.08	60.98
2018-10-04	NaN	22.8	11.7	11.7	NaN	73.04	53.06	53.06
2018-10-05	NaN	23.3	11.7	18.9	NaN	73.94	53.06	66.02

	temp_C				temp_F			
datatype	TAVG	TMAX	TMIN	TOBS	TAVG	TMAX	TMIN	TOBS
date								
2018-10-01	10.0	21.1	8.9	13.9	50.0	69.98	48.02	57.02
2018-10-02	-40.0	23.9	13.9	17.2	-40.0	75.02	57.02	62.96
2018-10-03	-40.0	25.0	15.6	16.1	-40.0	77.00	60.08	60.98
2018-10-04	-40.0	22.8	11.7	11.7	-40.0	73.04	53.06	53.06
2018-10-05	-40.0	23.3	11.7	18.9	-40.0	73.94	53.06	66.02

	date	TMAX	TMIN	TOBS
0	2018-10-01	21.1	8.9	13.9
1	2018-10-02	23.9	13.9	17.2
2	2018-10-03	25.0	15.6	16.1
3	2018-10-04	22.8	11.7	11.7
4	2018-10-05	23.3	11.7	18.9

	date	measurement	temp_C
0	2018-10-01	TMAX	21.1
1	2018-10-02	TMAX	23.9
2	2018-10-03	TMAX	25.0
3	2018-10-04	TMAX	22.8
4	2018-10-05	TMAX	23.3

	values		
	date		
2018-10-01	TMAX	21.1	
	TMIN	8.9	
	TOBS	13.9	
2018-10-02	TMAX	23.9	
	TMIN	13.9	

	date	station	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
0	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
1	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
2	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
3	2018-01-02T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-8.3	-16.1	-12.2	NaN	False
4	2018-01-03T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-4.4	-13.9	-13.3	NaN	False
	PRCP	SNOW	SNWD	TMAX		TMIN		TOBS		WESF
count	765.000000	577.000000	577.0	765.000000	765.000000	398.000000	11.000000			
mean	5.360392	4.202773	NaN	2649.175294	-15.914379	8.632161	16.290909			
std	10.002138	25.086077	NaN	2744.156281	24.242849	9.815054	9.489832			
min	0.000000	0.000000	-inf	-11.700000	-40.000000	-16.100000	1.800000			
25%	0.000000	0.000000	NaN	13.300000	-40.000000	0.150000	8.600000			
50%	0.000000	0.000000	NaN	32.800000	-11.100000	8.300000	19.300000			
75%	5.800000	0.000000	NaN	5505.000000	6.700000	18.300000	24.900000			
max	61.700000	229.000000	inf	5505.000000	23.900000	26.100000	28.700000			

	date	station	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
0	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
1	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
2	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
3	2018-01-02T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-8.3	-16.1	-12.2	NaN	False
4	2018-01-03T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-4.4	-13.9	-13.3	NaN	False
5	2018-01-03T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-4.4	-13.9	-13.3	NaN	False
6	2018-01-03T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-4.4	-13.9	-13.3	NaN	False
7	2018-01-04T00:00:00		?	20.6	229.0	inf	5505.0	-40.0	NaN	19.3
8	2018-01-04T00:00:00		?	20.6	229.0	inf	5505.0	-40.0	NaN	19.3
9	2018-01-05T00:00:00		?	0.3	NaN	NaN	5505.0	-40.0	NaN	NaN

	count	mean	std	min	25%	50%	75%	max
np.inf Snow Depth	24.0	101.041667	74.498018	13.0	25.0	120.5	152.0	229.0
-np.inf Snow Depth	553.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0

	date	station	inclement_weather
count	765	765	408
unique	324	2	2
top	2018-07-05T00:00:00	GHCND:USC00280907	False
freq	8	398	384

	date	station	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
1	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
2	2018-01-01T00:00:00		?	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN
5	2018-01-03T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-4.4	-13.9	-13.3	NaN	False
6	2018-01-03T00:00:00	GHCND:USC00280907	0.0	0.0	-inf	-4.4	-13.9	-13.3	NaN	False
8	2018-01-04T00:00:00		?	20.6	229.0	inf	5505.0	-40.0	NaN	19.3

PRCP SNOW SNWD TMAX TMIN TOBS WESF inclement_weather

	date									
	2018-01-01	0.0	0.0	-inf	5505.0	-40.0	NaN	NaN	NaN	NaN
	2018-01-02	0.0	0.0	-inf	-8.3	-16.1	-12.2	NaN	NaN	False
	2018-01-03	0.0	0.0	-inf	-4.4	-13.9	-13.3	NaN	NaN	False
	2018-01-04	20.6	229.0	inf	5505.0	-40.0	NaN	19.3	NaN	True
	2018-01-05	14.2	127.0	inf	-4.4	-13.9	-13.9	NaN	NaN	True

PRCP SNOW SNWD TMAX TMIN TOBS WESF inclement_weather

	date									
	2018-01-01	0.0	0.0	-inf	5505.0	-40.0	NaN	0.0	NaN	NaN
	2018-01-02	0.0	0.0	-inf	-8.3	-16.1	-12.2	0.0	NaN	False
	2018-01-03	0.0	0.0	-inf	-4.4	-13.9	-13.3	0.0	NaN	False
	2018-01-04	20.6	229.0	inf	5505.0	-40.0	NaN	19.3	NaN	True
	2018-01-05	14.2	127.0	inf	-4.4	-13.9	-13.9	0.0	NaN	True

	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
date								
2018-01-01	0.0	0.0	-inf	NaN	NaN	NaN	0.0	NaN
2018-01-02	0.0	0.0	-inf	-8.3	-16.1	-12.2	0.0	False
2018-01-03	0.0	0.0	-inf	-4.4	-13.9	-13.3	0.0	False
2018-01-04	20.6	229.0	inf	-4.4	-13.9	NaN	19.3	True
2018-01-05	14.2	127.0	inf	-4.4	-13.9	-13.9	0.0	True
	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
date								
2018-01-01	0.0	0.0	-1.797693e+308	NaN	NaN	NaN	0.0	NaN
2018-01-02	0.0	0.0	-1.797693e+308	-8.3	-16.1	-12.2	0.0	False
2018-01-03	0.0	0.0	-1.797693e+308	-4.4	-13.9	-13.3	0.0	False
2018-01-04	20.6	229.0	1.797693e+308	NaN	NaN	NaN	19.3	True
2018-01-05	14.2	127.0	1.797693e+308	-4.4	-13.9	-13.9	0.0	True
	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
date								
2018-01-01	0.0	0.0	0.0	NaN	NaN	NaN	0.0	NaN
2018-01-02	0.0	0.0	0.0	-8.3	-16.1	-12.2	0.0	False
2018-01-03	0.0	0.0	0.0	-4.4	-13.9	-13.3	0.0	False
2018-01-04	20.6	229.0	229.0	NaN	NaN	NaN	19.3	True
2018-01-05	14.2	127.0	127.0	-4.4	-13.9	-13.9	0.0	True
	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
date								
2018-01-01	0.0	0.0	-inf	14.4	5.6	10.0	0.0	NaN
2018-01-02	0.0	0.0	-inf	-8.3	-16.1	-12.2	0.0	False
2018-01-03	0.0	0.0	-inf	-4.4	-13.9	-13.3	0.0	False
2018-01-04	20.6	229.0	inf	14.4	5.6	10.0	19.3	True
2018-01-05	14.2	127.0	inf	-4.4	-13.9	-13.9	0.0	True

	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
date								
2018-01-01	0.0	0.0	-inf	NaN	NaN	NaN	0.0	NaN
2018-01-02	0.0	0.0	-inf	-8.30	-16.1	-12.20	0.0	False
2018-01-03	0.0	0.0	-inf	-4.40	-13.9	-13.30	0.0	False
2018-01-04	20.6	229.0	inf	-6.35	-15.0	-12.75	19.3	True
2018-01-05	14.2	127.0	inf	-4.40	-13.9	-13.90	0.0	True
2018-01-06	0.0	0.0	-inf	-10.00	-15.6	-15.00	0.0	False
2018-01-07	0.0	0.0	-inf	-11.70	-17.2	-16.10	0.0	False
2018-01-08	0.0	0.0	-inf	-7.80	-16.7	-8.30	0.0	False
2018-01-10	0.0	0.0	-inf	5.00	-7.8	-7.80	0.0	False
2018-01-11	0.0	0.0	-inf	4.40	-7.8	1.10	0.0	False

	PRCP	SNOW	SNWD	TMAX	TMIN	TOBS	WESF	inclement_weather
date								
2018-01-01	0.0	0.0	-inf	NaN	NaN	NaN	0.0	NaN
2018-01-02	0.0	0.0	-inf	-8.3	-16.10	-12.20	0.0	False
2018-01-03	0.0	0.0	-inf	-4.4	-13.90	-13.30	0.0	False
2018-01-04	20.6	229.0	inf	-4.4	-13.90	-13.60	19.3	True
2018-01-05	14.2	127.0	inf	-4.4	-13.90	-13.90	0.0	True
2018-01-06	0.0	0.0	-inf	-10.0	-15.60	-15.00	0.0	False
2018-01-07	0.0	0.0	-inf	-11.7	-17.20	-16.10	0.0	False
2018-01-08	0.0	0.0	-inf	-7.8	-16.70	-8.30	0.0	False
2018-01-09	0.0	0.0	-inf	-1.4	-12.25	-8.05	0.0	NaN
2018-01-10	0.0	0.0	-inf	5.0	-7.80	-7.80	0.0	False

Chapter 4: Aggregating Pandas DataFrames

File	Description	Source
<code>dirty_data.csv</code>	Dirty weather data from the <i>Handling duplicate, missing, or invalid data</i> section in <i>Chapter 3, Data Wrangling with Pandas</i>	Adapted from the NCEI API's GHCND dataset
<code>fb_2018.csv</code>	Facebook stock's opening, high, low, and closing price daily, along with volume traded for 2018.	The <code>stock_analysis</code> package (see <i>Chapter 7, Financial Analysis – Bitcoin and the Stock Market</i>).
<code>fb_week_of_may_20_per_minute.csv</code>	Facebook stock's opening, high, low, and closing price per minute, along with volume traded for May 20, 2019 through May 24, 2019.	Nasdaq
<code>melted_stock_data.csv</code>	The contents of <code>fb_week_of_may_20_per_minute.csv</code> melted into a single column for the price and another for the timestamp.	Adapted from Nasdaq
<code>nyc_weather_2018.csv</code>	Long format weather data for New York City across various stations.	The NCEI API's GHCND dataset.
<code>stocks.db</code>	The <code>fb_prices</code> and <code>aapl_prices</code> tables contain the stock prices for Facebook and Apple, respectively, for May 20, 2019 through May 24, 2019. Facebook is at a minute granularity, whereas Apple has timestamps that include (fictitious) seconds.	Adapted from Nasdaq
<code>weather_by_station.csv</code>	Long format weather data for New York City across various stations, along with station information.	The NCEI API's GHCND dataset and the <code>stations</code> endpoint.
<code>weather_stations.csv</code>	Information on all the stations providing weather data for New York City.	The NCEI API's <code>stations</code> endpoint.
<code>weather.db</code>	The <code>weather</code> table contains New York City weather data, while the <code>stations</code> table contains information on the stations.	The NCEI API's GHCND dataset and the <code>stations</code> endpoint.

	date	datatype	station	attributes	value
0	2018-01-01T00:00:00	PRCP	GHCND:US1CTFR0039	,,N,	0.0
1	2018-01-01T00:00:00	PRCP	GHCND:US1NJBG0015	,,N,	0.0
2	2018-01-01T00:00:00	SNOW	GHCND:US1NJBG0015	,,N,	0.0
3	2018-01-01T00:00:00	PRCP	GHCND:US1NJBG0017	,,N,	0.0
4	2018-01-01T00:00:00	SNOW	GHCND:US1NJBG0017	,,N,	0.0

	date	datatype	station	attributes	value
114	2018-01-01T00:00:00	SNOW	GHCND:US1NYWC0019	,,N,	25.0
789	2018-01-04T00:00:00	SNOW	GHCND:US1NYNS0007	,,N,	41.0
794	2018-01-04T00:00:00	SNOW	GHCND:US1NYNS0018	,,N,	10.0
798	2018-01-04T00:00:00	SNOW	GHCND:US1NYNS0024	,,N,	89.0
800	2018-01-04T00:00:00	SNOW	GHCND:US1NYNS0030	,,N,	102.0

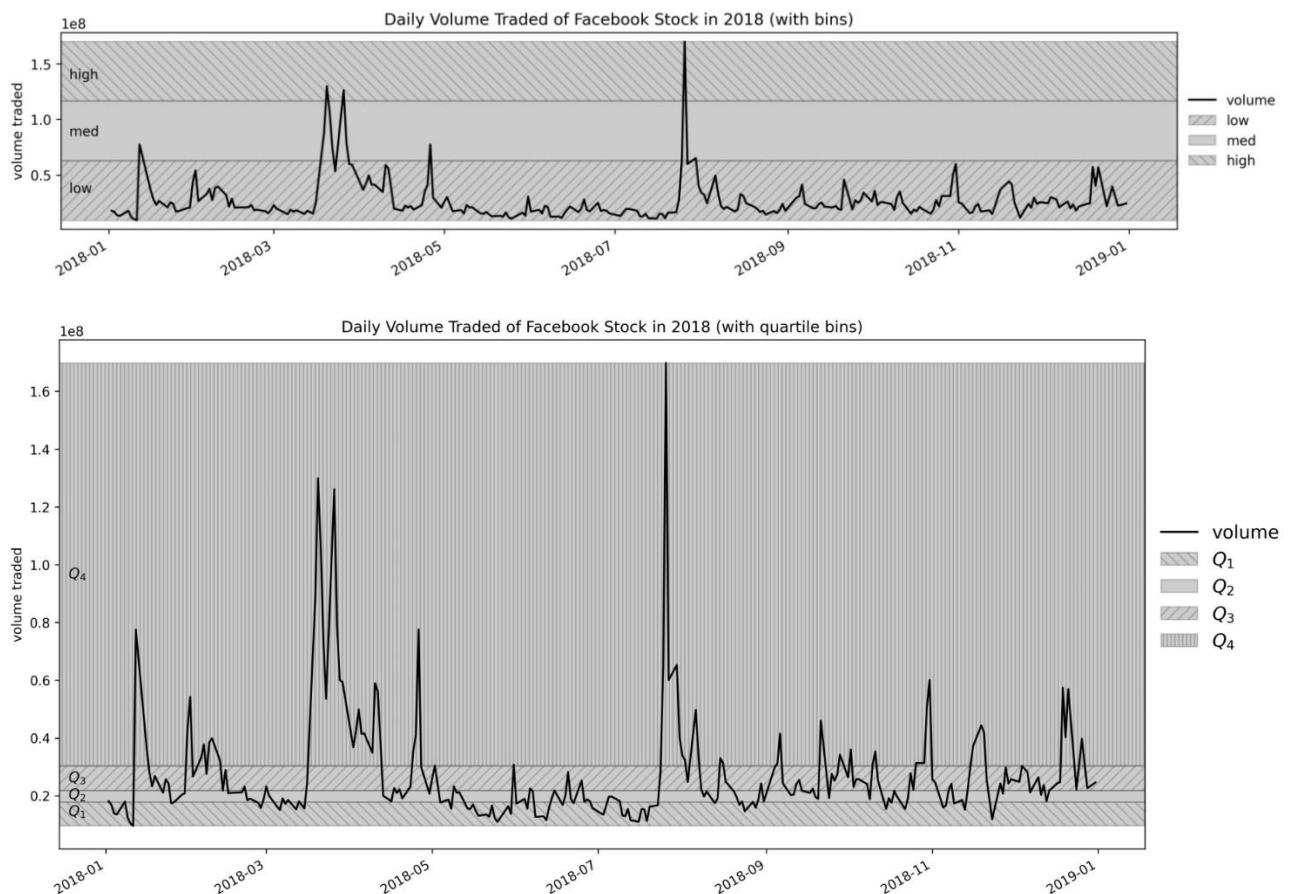
	id		name	latitude	longitude	elevation				
0	GHCND:US1CTFR0022	STAMFORD	2.6 SSW, CT US	41.064100	-73.577000	36.6				
1	GHCND:US1CTFR0039	STAMFORD	4.2 S, CT US	41.037788	-73.568176	6.4				
2	GHCND:US1NJBG0001	BERGENFIELD	0.3 SW, NJ US	40.921298	-74.001983	20.1				
3	GHCND:US1NJBG0002	SADDLE BROOK TWP	0.6 E, NJ US	40.902694	-74.083358	16.8				
4	GHCND:US1NJBG0003	TENAFLY	1.3 W, NJ US	40.914670	-73.977500	21.6				
date	datatype	station	attributes	value	id	name	latitude	longitude	elevation	
10739	2018-08-07T00:00:00	SNOW	GHCND:US1NJMN0069	,,N,	0.0	GHCND:US1NJMN0069	LONG BRANCH 1.7 SSW, NJ US	40.275368	-74.006027	9.4
45188	2018-12-21T00:00:00	TMAX	GHCND:USW00014732	,,W,2400	16.7	GHCND:USW00014732	LAGUARDIA AIRPORT, NY US	40.779440	-73.880350	3.4
59823	2018-01-15T00:00:00	WDF5	GHCND:USW00094741	,,W,	40.0	GHCND:USW00094741	TETERBORO AIRPORT, NJ US	40.850000	-74.061390	2.7
10852	2018-10-31T00:00:00	PRCP	GHCND:US1NJMN0069	T,,N,	0.0	GHCND:US1NJMN0069	LONG BRANCH 1.7 SSW, NJ US	40.275368	-74.006027	9.4
46755	2018-05-05T00:00:00	SNOW	GHCND:USW00014734	,,W,	0.0	GHCND:USW00014734	NEWARK LIBERTY INTERNATIONAL AIRPORT, NJ US	40.682500	-74.169400	2.1
date	datatype	station	attributes	value	name	latitude	longitude	elevation		
10739	2018-08-07T00:00:00	SNOW	GHCND:US1NJMN0069	,,N,	0.0	LONG BRANCH 1.7 SSW, NJ US	40.275368	-74.006027	9.4	
45188	2018-12-21T00:00:00	TMAX	GHCND:USW00014732	,,W,2400	16.7	LAGUARDIA AIRPORT, NY US	40.779440	-73.880350	3.4	
59823	2018-01-15T00:00:00	WDF5	GHCND:USW00094741	,,W,	40.0	TETERBORO AIRPORT, NJ US	40.850000	-74.061390	2.7	
10852	2018-10-31T00:00:00	PRCP	GHCND:US1NJMN0069	T,,N,	0.0	LONG BRANCH 1.7 SSW, NJ US	40.275368	-74.006027	9.4	
46755	2018-05-05T00:00:00	SNOW	GHCND:USW00014734	,,W,	0.0	NEWARK LIBERTY INTERNATIONAL AIRPORT, NJ US	40.682500	-74.169400	2.1	
date	datatype	station	attributes	value	id	name	latitude	longitude	elevation	
0	NaN	NaN	NaN	NaN	NaN	GHCND:US1CTFR0022	STAMFORD 2.6 SSW, CT US	41.064100	-73.577000	36.6
344	NaN	NaN	NaN	NaN	NaN	GHCND:US1NJBG0001	BERGENFIELD 0.3 SW, NJ US	40.921298	-74.001983	20.1
345	NaN	NaN	NaN	NaN	NaN	GHCND:US1NJBG0002	SADDLE BROOK TWP 0.6 E, NJ US	40.902694	-74.083358	16.8
718	NaN	NaN	NaN	NaN	NaN	GHCND:US1NJBG0005	WESTWOOD 0.8 ESE, NJ US	40.983041	-74.015858	15.8
719	NaN	NaN	NaN	NaN	NaN	GHCND:US1NJBG0006	RAMSEY 0.6 E, NJ US	41.058611	-74.134068	112.2

		date	datatype	station	attributes	value		id	name	latitude	longitude	elevation	_merge
23634		2018-04-12T00:00:00	PRCP	GHCND:US1NYNS0043	,,N,	0.0	GHCND:US1NYNS0043	PLAINVIEW 0.4 ENE, NY US	40.785919	-73.466873	56.7	both	
25742		2018-03-25T00:00:00	PRCP	GHCND:US1NYSF0061	,,N,	0.0	GHCND:US1NYSF0061	CENTERPORT 0.9 SW, NY US	40.891689	-73.383133	53.6	both	
60645		2018-04-16T00:00:00	TMIN	GHCND:USW00094741	,,W,	3.9		NaN	NaN	NaN	NaN	NaN	left_only
70764		2018-03-23T00:00:00	SNWD	GHCND:US1NJHD0002	,,N,	203.0		NaN	NaN	NaN	NaN	NaN	left_only
78790		Nan	Nan		Nan		GHCND:US1NYQN0033	HOWARD BEACH 0.4 NNW, NY US	40.662099	-73.841345	2.1	right_only	
78800		Nan	Nan		Nan		GHCND:US1NYWC0009	NEW ROCHELLE 1.3 S, NY US	40.904000	-73.777000	21.9	right_only	

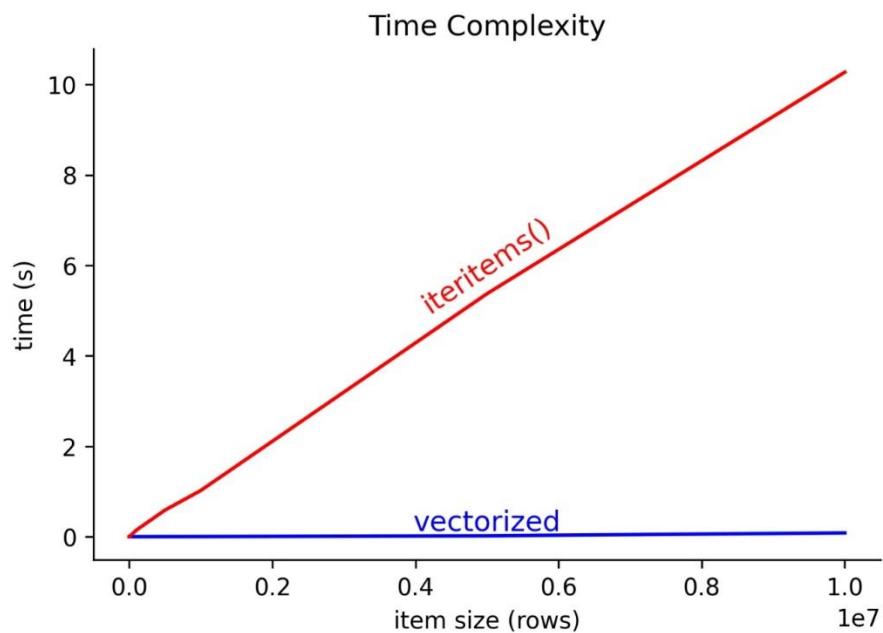
		station	PRCP	SNOW	TMAX	TMIN	TOBS	WESF	inclement_weather			
	date											
2018-01-01T00:00:00			?	0.0	0.0	5505.0	-40.0	Nan	Nan			
2018-01-02T00:00:00		GHCND:USC00280907	0.0	0.0	-8.3	-16.1	-12.2	Nan	False			
2018-01-03T00:00:00		GHCND:USC00280907	0.0	0.0	-4.4	-13.9	-13.3	Nan	False			
2018-01-04T00:00:00			?	20.6	229.0	5505.0	-40.0	Nan	19.3			
2018-01-05T00:00:00			?	0.3	Nan	5505.0	-40.0	Nan	Nan			
		PRCP_x	SNOW_x	TMAX	TMIN	TOBS	inclement_weather_x	PRCP_y	SNOW_y	WESF	inclement_weather_y	
	date											
2018-01-30T00:00:00		0.0	0.0	6.7	-1.7	-0.6		False	1.5	13.0	1.8	True
2018-03-08T00:00:00		48.8	Nan	1.1	-0.6	1.1		False	28.4	Nan	28.7	Nan
2018-03-13T00:00:00		4.1	51.0	5.6	-3.9	0.0		True	3.0	13.0	3.0	True
2018-03-21T00:00:00		0.0	0.0	2.8	-2.8	0.6		False	6.6	114.0	8.6	True
2018-04-02T00:00:00		9.1	127.0	12.8	-1.1	-1.1		True	14.0	152.0	15.2	True
		PRCP	SNOW	TMAX	TMIN	TOBS	inclement_weather	PRCP_?	SNOW_?	WESF	inclement_weather_?	
	date											
2018-01-30T00:00:00		0.0	0.0	6.7	-1.7	-0.6		False	1.5	13.0	1.8	True
2018-03-08T00:00:00		48.8	Nan	1.1	-0.6	1.1		False	28.4	Nan	28.7	Nan
2018-03-13T00:00:00		4.1	51.0	5.6	-3.9	0.0		True	3.0	13.0	3.0	True
2018-03-21T00:00:00		0.0	0.0	2.8	-2.8	0.6		False	6.6	114.0	8.6	True
2018-04-02T00:00:00		9.1	127.0	12.8	-1.1	-1.1		True	14.0	152.0	15.2	True

	open	high	low	close	volume	abs_z_score_volume
date						
2018-03-19	177.01	177.17	170.06	172.56	88140060	3.145078
2018-03-20	167.47	170.20	161.95	168.15	129851768	5.315169
2018-03-21	164.80	173.40	163.30	169.39	106598834	4.105413
2018-03-26	160.82	161.10	149.02	160.06	126116634	5.120845
2018-07-26	174.89	180.13	173.75	176.26	169803668	7.393705
	open	high	low	close	volume	volume_pct_change
date						
2018-01-12	178.06	181.48	177.40	179.37	77551299	7.087876
2018-03-19	177.01	177.17	170.06	172.56	88140060	2.611789
2018-07-26	174.89	180.13	173.75	176.26	169803668	1.628841
2018-09-21	166.64	167.25	162.81	162.93	45994800	1.428956
2018-03-26	160.82	161.10	149.02	160.06	126116634	1.352496
	open	high	low	close	volume	
date						
2018-01-11	188.40	188.40	187.38	187.77	9588587	
2018-01-12	178.06	181.48	177.40	179.37	77551299	
	open	high	low	close	volume	
date						
2018-07-26	174.89	180.13	173.75	176.26	169803668	
2018-03-20	167.47	170.20	161.95	168.15	129851768	
2018-03-26	160.82	161.10	149.02	160.06	126116634	
	open	high	low	close	volume	
date						
2018-07-25	215.715	218.62	214.27	217.50	64592585	
2018-07-26	174.890	180.13	173.75	176.26	169803668	

	open	high	low	close	volume
date					
2018-03-16	184.49	185.33	183.41	185.09	24403438
2018-03-19	177.01	177.17	170.06	172.56	88140060
2018-03-20	167.47	170.20	161.95	168.15	129851768



	count	mean	std	min	25%	50%	75%	max
datatype								
TMIN	31.0	-1.790682e-16	1.0	-1.339112	-0.751019	-0.474269	1.065152	1.843511
TMAX	31.0	1.951844e-16	1.0	-1.305582	-0.870013	-0.138258	1.011643	1.604016
PRCP	31.0	4.655774e-17	1.0	-0.394438	-0.394438	-0.394438	-0.240253	3.936167



	date	2018-10-01	2018-10-02	2018-10-03	2018-10-04	2018-10-05	2018-10-06	2018-10-07
	datatype							
	PRCP	0.0	17.5	0.0	1.0	0.0	0.0	0.0
	rolling_PRCP	0.0	17.5	17.5	18.5	1.0	1.0	0.0
	datatype	AWND	PRCP	SNOW	SNWD	TMAX	TMIN	
	date							
	2018-10-01	0.900000	0.000000	0.0	0.0	24.400000	17.200000	
	2018-10-02	0.900000	8.750000	0.0	0.0	24.700000	17.750000	
	2018-10-03	0.966667	5.833333	0.0	0.0	24.233333	17.566667	
	2018-10-04	0.800000	6.166667	0.0	0.0	24.233333	17.200000	
	2018-10-05	1.033333	0.333333	0.0	0.0	23.133333	16.300000	
	2018-10-06	0.833333	0.333333	0.0	0.0	22.033333	16.300000	
	2018-10-07	1.066667	0.000000	0.0	0.0	22.600000	17.400000	

	AWNND	AWNND_rolling	PRCP	PRCP_rolling	TMAX	TMAX_rolling	TMIN	TMIN_rolling
--	--------------	----------------------	-------------	---------------------	-------------	---------------------	-------------	---------------------

date								
-------------	--	--	--	--	--	--	--	--

2018-10-01	0.9	0.900000	0.0	0.0	24.4	24.4	17.2	17.2
2018-10-02	0.9	0.900000	17.5	17.5	25.0	25.0	18.3	17.2
2018-10-03	1.1	0.966667	0.0	17.5	23.3	25.0	17.2	17.2
2018-10-04	0.4	0.800000	1.0	18.5	24.4	25.0	16.1	16.1
2018-10-05	1.6	1.033333	0.0	1.0	21.7	24.4	15.6	15.6
2018-10-06	0.5	0.833333	0.0	1.0	20.0	24.4	17.2	15.6
2018-10-07	1.1	1.066667	0.0	0.0	26.1	26.1	19.4	15.6

date	2018-06-01	2018-06-02	2018-06-03	2018-06-04	2018-06-05	2018-06-06	2018-06-07	2018-06-08	2018-06-09	2018-06-10
-------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

datatype										
-----------------	--	--	--	--	--	--	--	--	--	--

PRCP	6.9	2.00	6.4	4.10	0.00	0.000000	0.000000	0.000	0.000000	0.30
TOTAL_PRCP	6.9	8.90	15.3	19.40	19.40	19.400000	19.400000	19.400	19.400000	19.70
AVG_PRCP	6.9	4.45	5.1	4.85	3.88	3.233333	2.771429	2.425	2.155556	1.97

	AWNND	AWNND_expanding	PRCP	PRCP_expanding	TMAX	TMAX_expanding	TMIN	TMIN_expanding
--	--------------	------------------------	-------------	-----------------------	-------------	-----------------------	-------------	-----------------------

date								
-------------	--	--	--	--	--	--	--	--

2018-10-01	0.9	0.900000	0.0	0.0	24.4	24.4	17.2	17.2
2018-10-02	0.9	0.900000	17.5	17.5	25.0	25.0	18.3	17.2
2018-10-03	1.1	0.966667	0.0	17.5	23.3	25.0	17.2	17.2
2018-10-04	0.4	0.825000	1.0	18.5	24.4	25.0	16.1	16.1
2018-10-05	1.6	0.980000	0.0	18.5	21.7	25.0	15.6	15.6
2018-10-06	0.5	0.900000	0.0	18.5	20.0	25.0	17.2	15.6
2018-10-07	1.1	0.928571	0.0	18.5	26.1	26.1	19.4	15.6

date	2018-09-29	2018-09-30	2018-10-01	2018-10-02	2018-10-03	2018-10-04	2018-10-05	2018-10-06	2018-10-07	2018-10-08
-------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

datatype										
-----------------	--	--	--	--	--	--	--	--	--	--

TMAX	22.200000	21.100000	24.400000	25.000000	23.300000	24.400000	21.700000	20.000000	26.100000	23.300000
EWMA	24.410887	24.197281	24.210360	24.261304	24.199285	24.212234	24.050154	23.788854	23.937960	23.896802
AVG	24.723333	24.573333	24.533333	24.460000	24.163333	23.866667	23.533333	23.070000	23.143333	23.196667

	open	high	low	close	volume
--	-------------	-------------	------------	--------------	---------------

date					
-------------	--	--	--	--	--

2018-01-02	177.68	181.580	177.5500	181.420	18151903.0
2018-01-03	179.78	183.180	179.4400	183.045	17519233.0
2018-01-04	181.88	184.780	181.3300	184.330	16886563.0
2018-01-05	183.39	185.495	182.7148	184.500	15383729.5
2018-01-08	184.90	186.210	184.0996	184.670	16886563.0

	open	high	low	close	volume
date					
2018-01-02	177.680000	181.580000	177.550000	181.420000	1.815190e+07
2018-01-03	180.480000	183.713333	180.070000	183.586667	1.730834e+07
2018-01-04	183.005714	185.140000	182.372629	184.011429	1.534980e+07
2018-01-05	184.384000	186.078667	183.736560	185.525333	1.440299e+07
2018-01-08	185.837419	187.534839	185.075110	186.947097	1.625679e+07

	TMAX	TMIN	AWND	PRCP
date				
2018-10-01	24.4	17.2	0.900000	0.0
2018-10-02	25.0	17.2	0.900000	17.5
2018-10-03	25.0	17.2	0.966667	17.5
2018-10-04	25.0	16.1	0.800000	18.5
2018-10-05	24.4	15.6	1.033333	1.0

	datatype	station	value	station_name
date				
2018-01-01	PRCP	GHCND:US1CTFR0039	0.0	STAMFORD 4.2 S, CT US
2018-01-01	PRCP	GHCND:US1NJBG0015	0.0	NORTH ARLINGTON 0.7 WNW, NJ US
2018-01-01	SNOW	GHCND:US1NJBG0015	0.0	NORTH ARLINGTON 0.7 WNW, NJ US
2018-01-01	PRCP	GHCND:US1NJBG0017	0.0	GLEN ROCK 0.7 SSE, NJ US
2018-01-01	SNOW	GHCND:US1NJBG0017	0.0	GLEN ROCK 0.7 SSE, NJ US

	open	high	low	close
mean	171.45	NaN	NaN	171.51
min	NaN	129.74	123.02	NaN
max	NaN	218.62	214.27	NaN

	open	high	low	close	volume		
trading_volume	low	171.36	173.46	169.31	171.43	24547207.71	
	med	175.82	179.42	172.11	175.14	79072559.12	
	high	167.73	170.48	161.57	168.16	141924023.33	
	min	max	mean				
trading_volume	low	124.06	214.67	171.43			
	med	152.22	217.50	175.14			
	high	160.06	176.26	168.16			
	open	high	low	close			
trading_volume	mean	min	max	min	max	mean	
trading_volume	low	171.36	129.74	216.20	123.02	212.60	171.43
	med	175.82	162.85	218.62	150.75	214.27	175.14
	high	167.73	161.10	180.13	149.02	173.75	168.16
	open_mean	high_min	high_max	low_min	low_max	close_mean	
trading_volume	low	171.36	129.74	216.20	123.02	212.60	171.43
	med	175.82	162.85	218.62	150.75	214.27	175.14
	high	167.73	161.10	180.13	149.02	173.75	168.16
	value						
	date	2018-03-31	2018-06-30	2018-09-30	2018-12-31		
station_name							
WANTAGH 1.1 NNE, NY US		279.90	216.80	472.50	277.20		
STATEN ISLAND 1.4 SE, NY US		379.40	295.30	438.80	409.90		
SYOSSET 2.0 SSW, NY US		323.50	263.30	355.50	459.90		
STAMFORD 4.2 S, CT US		338.00	272.10	424.70	390.00		
WAYNE TWP 0.8 SSW, NJ US		246.20	295.30	620.90	422.00		

prcp

date

2018-01-28	69.31
2018-01-29	69.31
2018-01-30	69.31
2018-01-31	69.31
2018-02-01	158.11
2018-02-02	158.11
2018-02-03	158.11

prcp total_prcp_in_month pct_monthly_prcp

date

2018-10-12	34.77	105.63	0.33
2018-01-13	21.66	69.31	0.31
2018-03-02	38.77	137.46	0.28
2018-04-16	39.34	140.57	0.28
2018-04-17	37.30	140.57	0.27

trading_volume low med high

close	171.43	175.14	168.16
high	173.46	179.42	170.48
low	169.31	172.11	161.57
open	171.36	175.82	167.73

volume 24547207.71 79072559.12 141924023.33

datatype	date	station	station_name	AWND	DAPR	MDPR	PGTM	PRCP	SNOW	SNWD	...
28740	2018-12-31	GHCND:USW00054787	FARMINGDALE REPUBLIC AIRPORT, NY US	5.00	NaN	NaN	2052.00	28.70	NaN	NaN	...
28741	2018-12-31	GHCND:USW00094728	NY CITY CENTRAL PARK, NY US	NaN	NaN	NaN	NaN	25.90	0.00	0.00	...
28742	2018-12-31	GHCND:USW00094741	TETERBORO AIRPORT, NJ US	1.70	NaN	NaN	1954.00	29.20	NaN	NaN	...
28743	2018-12-31	GHCND:USW00094745	WESTCHESTER CO AIRPORT, NY US	2.70	NaN	NaN	2212.00	24.40	NaN	NaN	...
28744	2018-12-31	GHCND:USW00094789	JFK INTERNATIONAL AIRPORT, NY US	4.10	NaN	NaN	NaN	31.20	0.00	0.00	...

month	1	2	3	4	5	6	7	8	9	10	11	12
trading_volume												
low	20	19	15	20	22	21	18	23	19	23	21	19
med	1	0	4	1	0	0	2	0	0	0	0	0
high	0	0	2	0	0	0	1	0	0	0	0	0
month	1	2	3	4	5	6	7	8	9	10	11	12
trading_volume												
low	185.24	180.27	177.07	163.29	182.93	195.27	201.92	177.49	164.38	154.19	141.64	137.16
med	179.37	NaN	164.76	174.16	NaN	NaN	194.28	NaN	NaN	NaN	NaN	NaN
high	NaN	NaN	164.11	NaN	NaN	NaN	176.26	NaN	NaN	NaN	NaN	NaN
month	1	2	3	4	5	6	7	8	9	10	11	12
total observations of snow												
station_name												
ALBERTSON 0.2 SSE, NY US	3.00	1.00	3.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	9.00
AMITYVILLE 0.1 WSW, NY US	1.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.00
AMITYVILLE 0.6 NNE, NY US	3.00	1.00	3.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.00
ARMONK 0.3 SE, NY US	6.00	4.00	6.00	3.00	0.00	0.00	0.00	0.00	0.00	1.00	3.00	23.00
BLOOMINGDALE 0.7 SSE, NJ US	2.00	1.00	3.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	8.00
...
WESTFIELD 0.6 NE, NJ US	3.00	0.00	4.00	1.00	0.00	NaN	0.00	0.00	0.00	1.00	NaN	9.00
WOODBRIDGE TWP 1.1 ESE, NJ US	4.00	1.00	3.00	2.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	11.00
WOODBRIDGE TWP 1.1 NNE, NJ US	2.00	1.00	3.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	7.00
WOODBRIDGE TWP 3.0 NNW, NJ US	NaN	0.00	0.00	NaN	NaN	0.00	NaN	NaN	0.00	0.00	NaN	0.00
total observations of snow	190.00	97.00	237.00	81.00	0.00	0.00	0.00	0.00	0.00	49.00	13.00	667.00
open	high	low	close	volume	trading_volume							
date												
2018-10-11	150.13	154.81	149.1600	153.35	35338901							low
2018-10-12	156.73	156.89	151.2998	153.74	25293492							low
2018-10-15	153.32	155.57	152.5500	153.52	15433521							low

	open	high	low	close	volume	trading_volume
date						
2018-01-02	177.68	181.58	177.5500	181.42	18151903	low
2018-01-03	181.88	184.78	181.3300	184.67	16886563	low
2018-01-04	184.90	186.21	184.0996	184.33	13880896	low
2018-01-05	185.59	186.90	184.9300	186.85	13574535	low

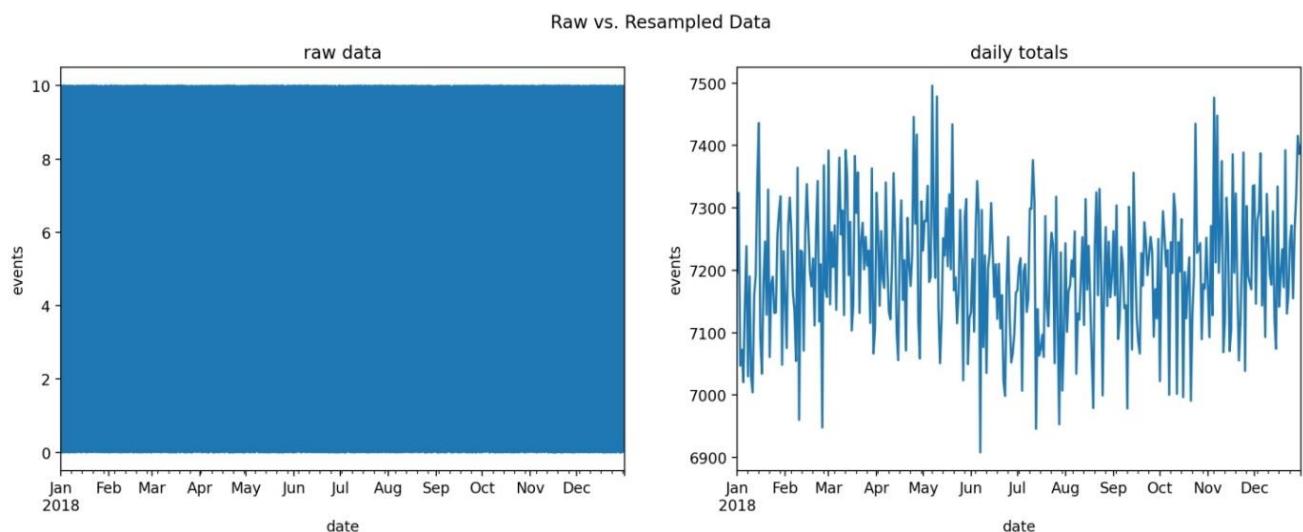
	open	high	low	close	volume	trading_volume
date						
2018-12-31	134.45	134.64	129.95	131.09	24625308	low

	open	high	low	close	volume
date					
2019-05-20 09:30:00	181.6200	181.6200	181.6200	181.6200	159049.0
2019-05-20 09:31:00	182.6100	182.6100	182.6100	182.6100	468017.0
2019-05-20 09:32:00	182.7458	182.7458	182.7458	182.7458	97258.0
2019-05-20 09:33:00	182.9500	182.9500	182.9500	182.9500	43961.0
2019-05-20 09:34:00	183.0600	183.0600	183.0600	183.0600	79562.0

	open	high	low	close	volume
date					
2019-05-20	181.62	184.1800	181.6200	182.72	10044838.0
2019-05-21	184.53	185.5800	183.9700	184.82	7198405.0
2019-05-22	184.81	186.5603	184.0120	185.32	8412433.0
2019-05-23	182.50	183.7300	179.7559	180.87	12479171.0
2019-05-24	182.33	183.5227	181.0400	181.06	7686030.0

	open	high	low	close	volume				
date									
2019-05-20 09:30:00	181.62	181.62	181.62	181.62	159049.0				
2019-05-21 09:30:00	184.53	184.53	184.53	184.53	58171.0				
2019-05-22 09:30:00	184.81	184.81	184.81	184.81	41585.0				
2019-05-23 09:30:00	182.50	182.50	182.50	182.50	121930.0				
2019-05-24 09:30:00	182.33	182.33	182.33	182.33	52681.0				
	open	high	low	close	volume				
date									
2019-05-20 15:59:00	182.915	182.915	182.915	182.915	134569.0				
2019-05-20 16:00:00	182.720	182.720	182.720	182.720	1113672.0				
2019-05-21 15:59:00	184.840	184.840	184.840	184.840	61606.0				
2019-05-21 16:00:00	184.820	184.820	184.820	184.820	801080.0				
2019-05-22 15:59:00	185.290	185.290	185.290	185.290	96099.0				
2019-05-22 16:00:00	185.320	185.320	185.320	185.320	1220993.0				
2019-05-23 15:59:00	180.720	180.720	180.720	180.720	109648.0				
2019-05-23 16:00:00	180.870	180.870	180.870	180.870	1329217.0				
2019-05-24 15:59:00	181.070	181.070	181.070	181.070	52994.0				
2019-05-24 16:00:00	181.060	181.060	181.060	181.060	764906.0				
	open	high	low	close	volume	trading_volume	prior_close	after_hours_change_in_price	abs_change
date									
2018-07-26	174.89	180.13	173.75	176.26	169803668	high	217.50	-42.61	42.61
2018-04-26	173.22	176.27	170.80	174.16	77556934	med	159.69	13.53	13.53
2018-01-12	178.06	181.48	177.40	179.37	77551299	med	187.77	-9.71	9.71
2018-10-31	155.00	156.40	148.96	151.79	60101251	low	146.22	8.78	8.78
2018-03-19	177.01	177.17	170.06	172.56	88140060	med	185.09	-8.08	8.08

	open	high	low	close	volume
date					
2018-01-02	NaN	NaN	NaN	NaN	NaN
2018-01-03	4.20	3.20	3.7800	3.25	-1265340.0
2018-01-04	3.02	1.43	2.7696	-0.34	-3005667.0
2018-01-05	0.69	0.69	0.8304	2.52	-306361.0
2018-01-08	1.61	2.00	1.4000	1.43	4420191.0



	open	high	low	close	volume
date					
2019-05-20	181.62	184.1800	181.6200	182.72	10044838.0
2019-05-21	184.53	185.5800	183.9700	184.82	7198405.0
2019-05-22	184.81	186.5603	184.0120	185.32	8412433.0
2019-05-23	182.50	183.7300	179.7559	180.87	12479171.0
2019-05-24	182.33	183.5227	181.0400	181.06	7686030.0

	open	high	low	close	volume
date					
2018-03-31	179.472295	181.794659	177.040428	179.551148	3.292640e+07
2018-06-30	180.373770	182.277689	178.595964	180.704687	2.405532e+07
2018-09-30	180.812130	182.890886	178.955229	181.028492	2.701982e+07
2018-12-31	145.272460	147.620121	142.718943	144.868730	2.697433e+07

	open	high	low	close	volume
date					
2018-03-31	-22.53	-20.1600	-23.410	-21.63	41282390
2018-06-30	39.51	38.3997	39.844	38.93	-20984389
2018-09-30	-25.04	-28.6600	-29.660	-32.90	20304060
2018-12-31	-28.58	-31.2400	-31.310	-31.35	-1782369

	price
	date
	2019-05-20 09:30:00 181.6200
	2019-05-20 09:31:00 182.6100
	2019-05-20 09:32:00 182.7458
	2019-05-20 09:33:00 182.9500
	2019-05-20 09:34:00 183.0600

	open	high	low	close
date				
2019-05-20	181.62	184.1800	181.6200	182.72
2019-05-21	184.53	185.5800	183.9700	184.82
2019-05-22	184.81	186.5603	184.0120	185.32
2019-05-23	182.50	183.7300	179.7559	180.87
2019-05-24	182.33	183.5227	181.0400	181.06

	open	high	low	close	volume	trading_volume
date						
2018-01-02 00:00:00	177.68	181.58	177.55	181.42	18151903.0	low
2018-01-02 06:00:00	NaN	NaN	NaN	NaN	NaN	NaN
2018-01-02 12:00:00	NaN	NaN	NaN	NaN	NaN	NaN
2018-01-02 18:00:00	NaN	NaN	NaN	NaN	NaN	NaN
2018-01-03 00:00:00	181.88	184.78	181.33	184.67	16886563.0	low

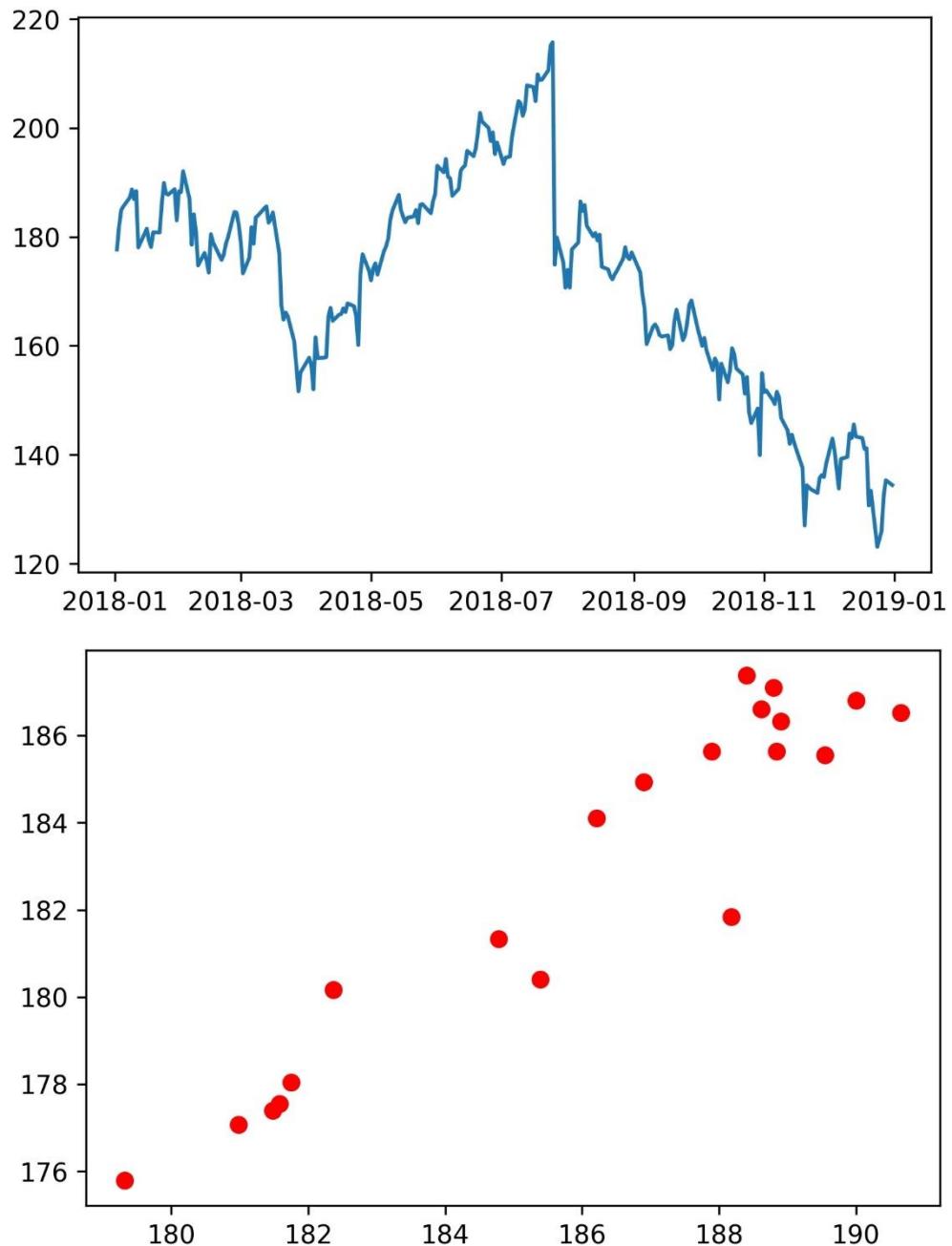
FB AAPL

	date		
2019-05-20 09:30:00	181.6200	183.5200	
2019-05-20 09:31:00	182.6100	NaN	
2019-05-20 09:32:00	182.7458	182.8710	
2019-05-20 09:33:00	182.9500	182.5000	
2019-05-20 09:34:00	183.0600	182.1067	

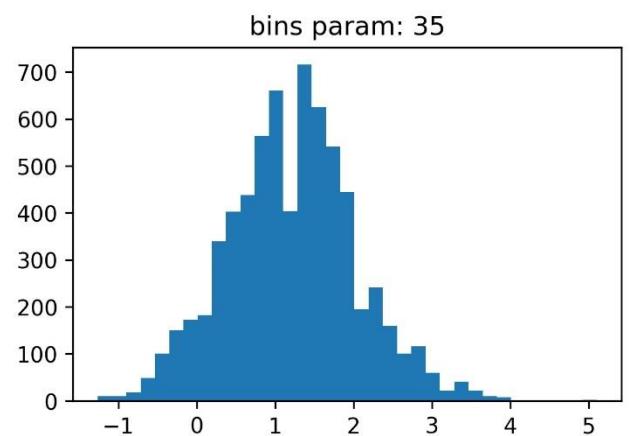
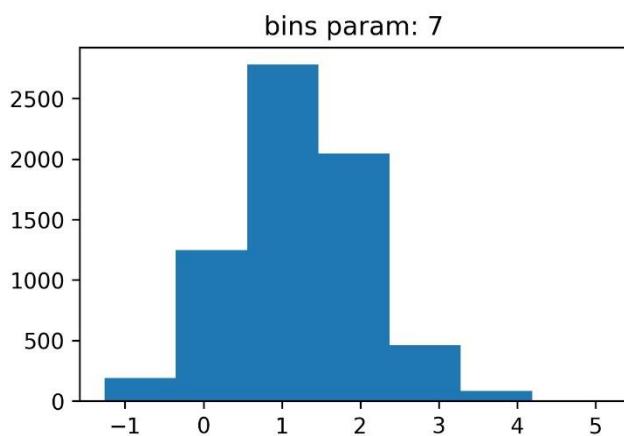
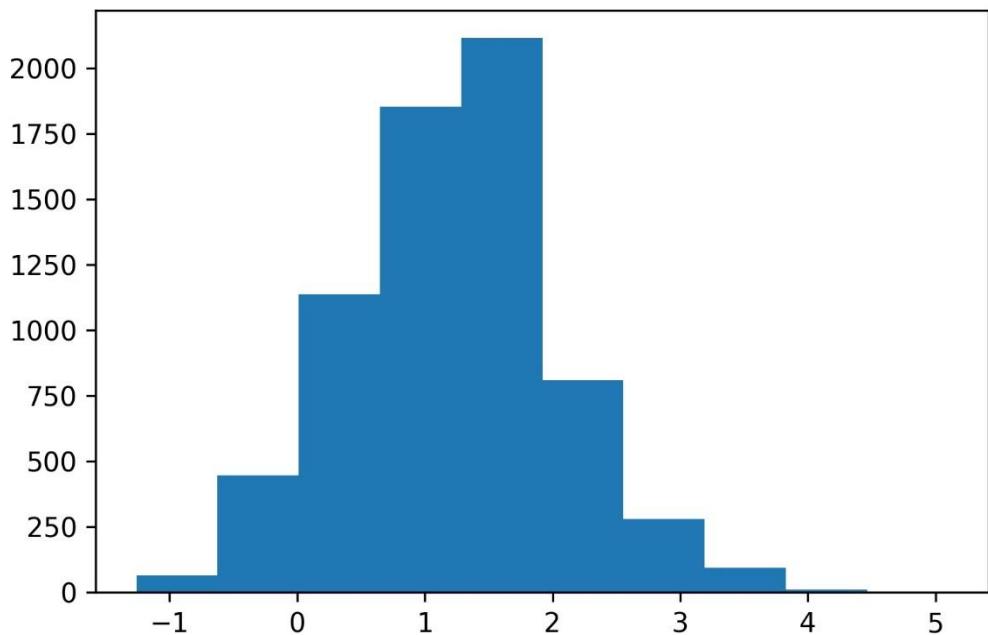
FB AAPL

	date		
2019-05-20 09:30:00	181.6200	183.520	
2019-05-20 09:31:00	182.6100	NaN	
2019-05-20 09:31:52	NaN	182.871	
2019-05-20 09:32:00	182.7458	NaN	
2019-05-20 09:32:36	NaN	182.500	

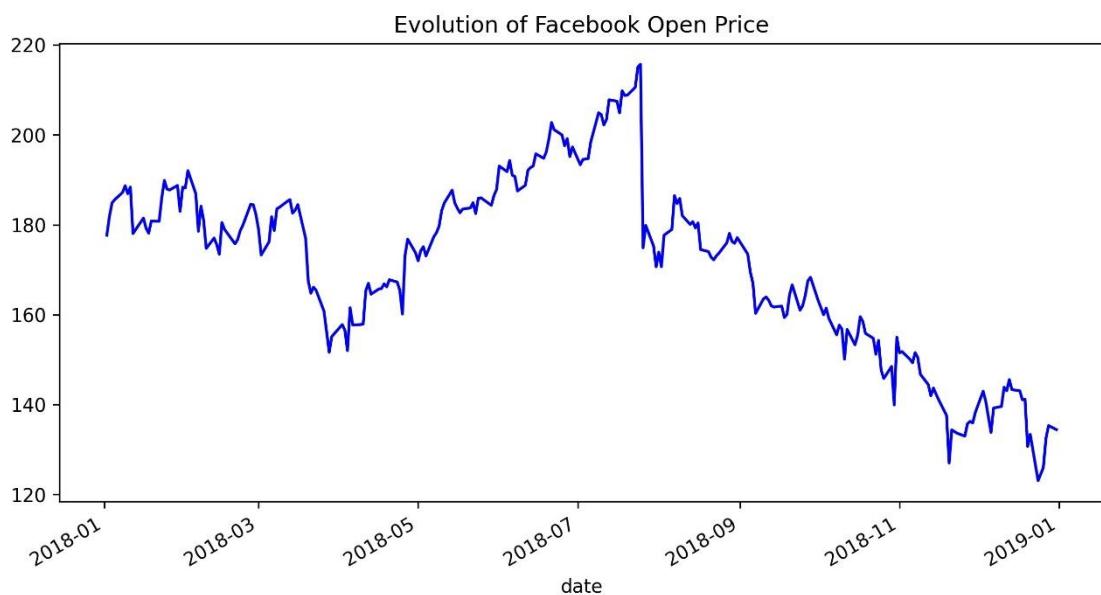
Chapter 5: Visualizing Data with Pandas and Matplotlib



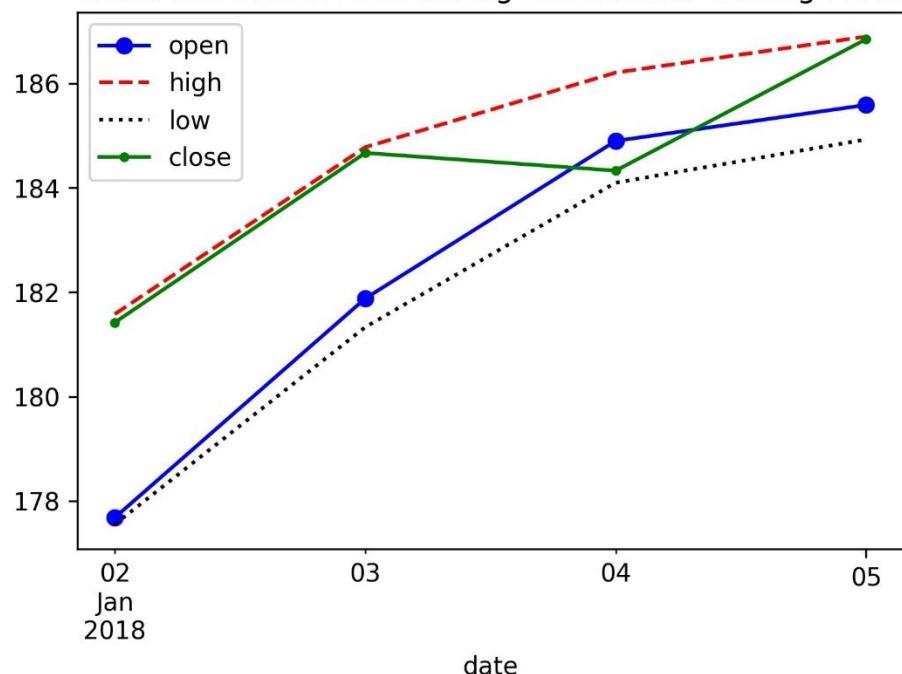
Marker	Linestyle	Color	Format String	Result
-	b	-b	blue solid line	
.	k	.k	black points	
--	r	--r	red dashed line	
o	-	g	o-g	green solid line with circles
:	m	:m	magenta dotted line	
x	-.	c	x-.c	cyan dot-dashed line with x's



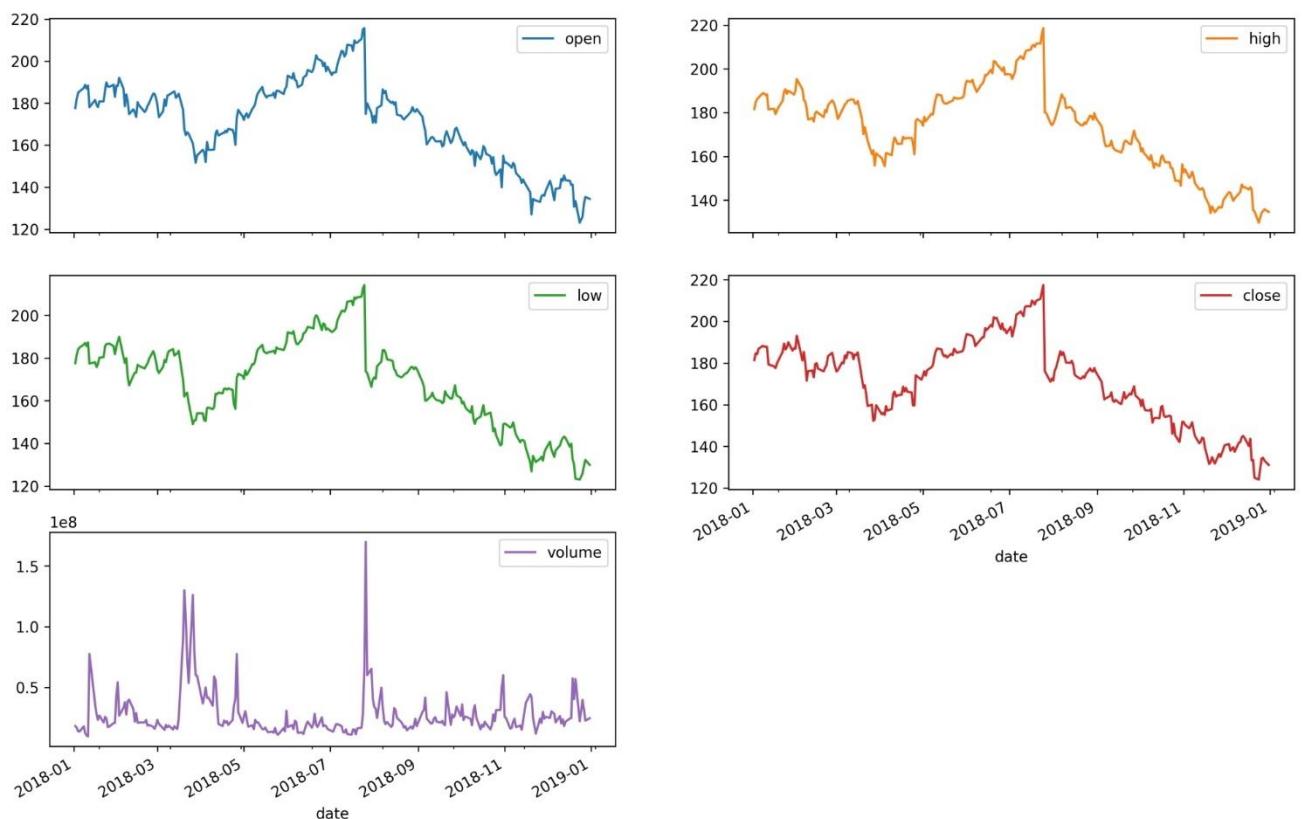
Parameter	Purpose	Data Type
<code>kind</code>	Determines the plot type	String
<code>x / y</code>	Column(s) to plot on the x-axis/y-axis	String or list
<code>ax</code>	Draws the plot on the <code>Axes</code> object provided	<code>Axes</code>
<code>subplots</code>	Determines whether to make subplots	Boolean
<code>layout</code>	Specifies how to arrange the subplots	Tuple of <code>(rows, columns)</code>
<code>figsize</code>	Size to make the <code>Figure</code> object	Tuple of <code>(width, height)</code>
<code>title</code>	The title of the plot or subplots	String for the plot title or a list of strings for subplot titles
<code>legend</code>	Determines whether to show the legend	Boolean
<code>label</code>	What to call an item in the legend	String if a single column is being plotted; otherwise, a list of strings
<code>style</code>	<code>matplotlib</code> style strings for each item being plotted	String if a single column is being plotted; otherwise, a list of strings
<code>color</code>	The color to plot the item in	String or red, green, blue tuple if a single column is being plotted; otherwise, a list
<code>colormap</code>	The colormap to use	String or <code>matplotlib</code> colormap object
<code>logx / logy / loglog</code>	Determines whether to use a logarithmic scale for the x-axis, y-axis, or both	Boolean
<code>xticks / yticks</code>	Determines where to draw the ticks on the x-axis/y-axis	List of values
<code>xlim / ylim</code>	The axis limits for the x-axis/y-axis	Tuple of the form <code>(min, max)</code>
<code>rot</code>	The angle to write the tick labels at	Integer
<code>sharex / sharey</code>	Determines whether to have subplots share the x-axis/y-axis	Boolean
<code>fontsize</code>	Controls the size of the tick labels	Integer
<code>grid</code>	Turns on/off the grid lines	Boolean

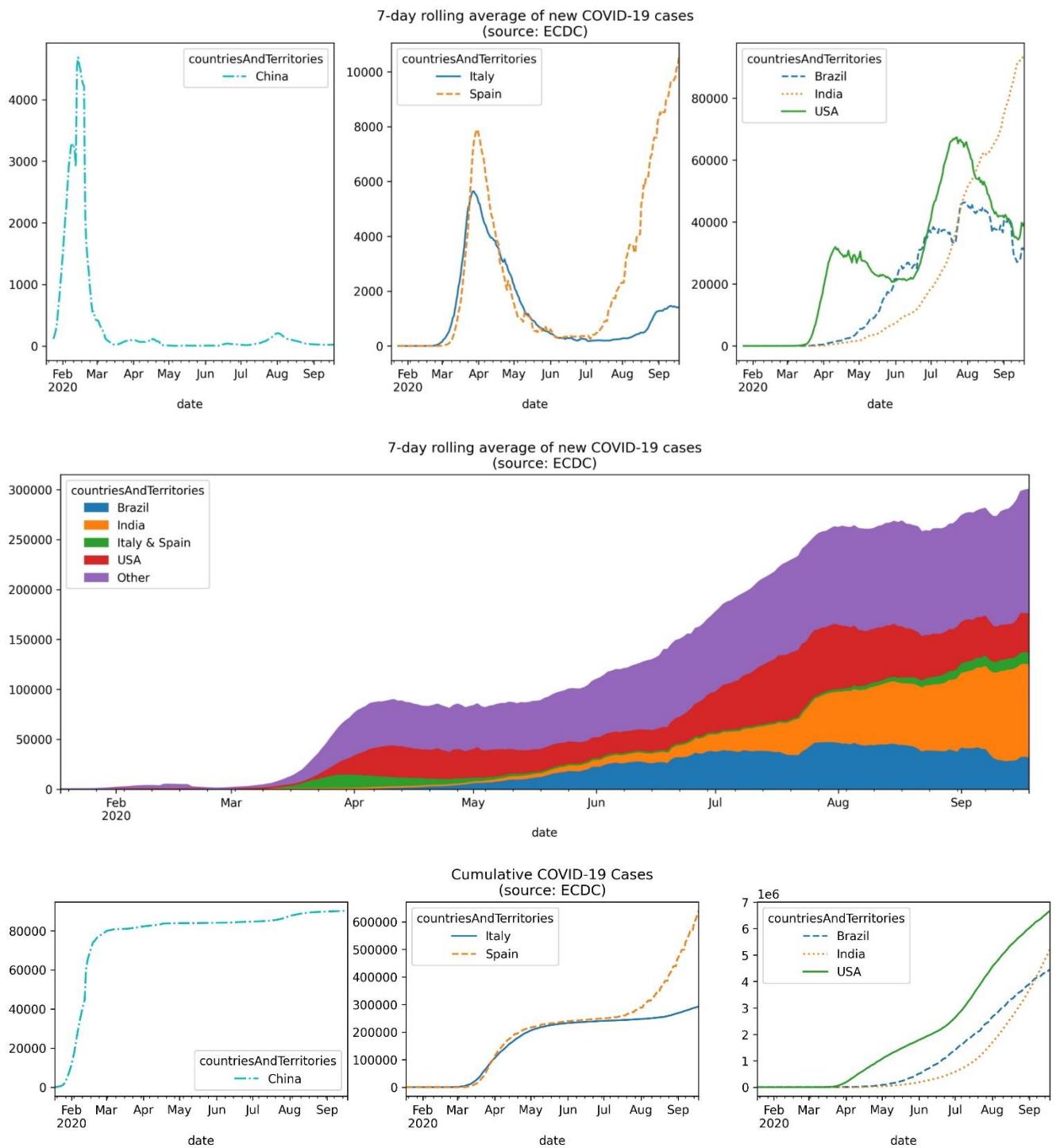


Facebook OHLC Prices during 1st Week of Trading 2018

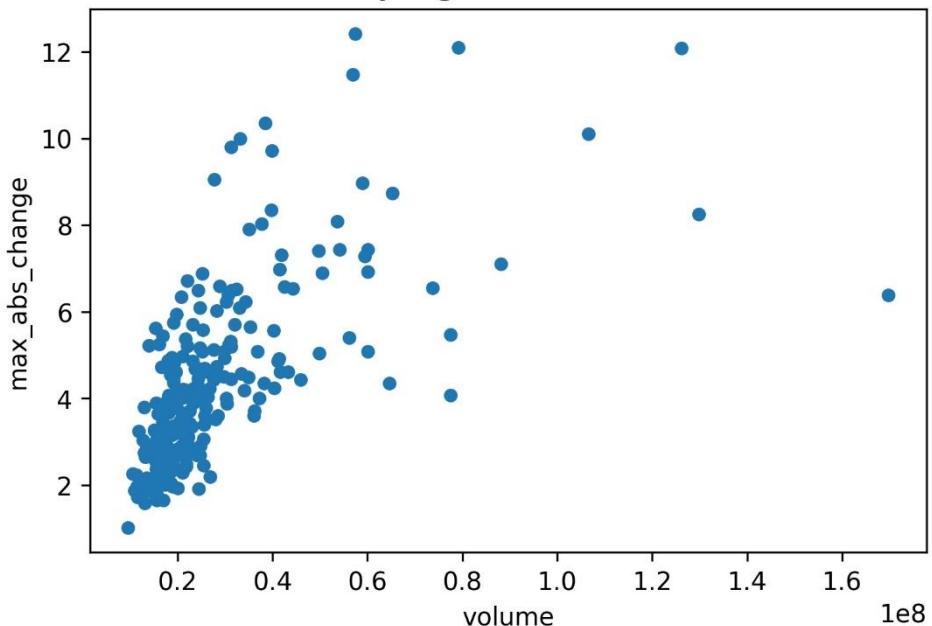


Facebook Stock 2018

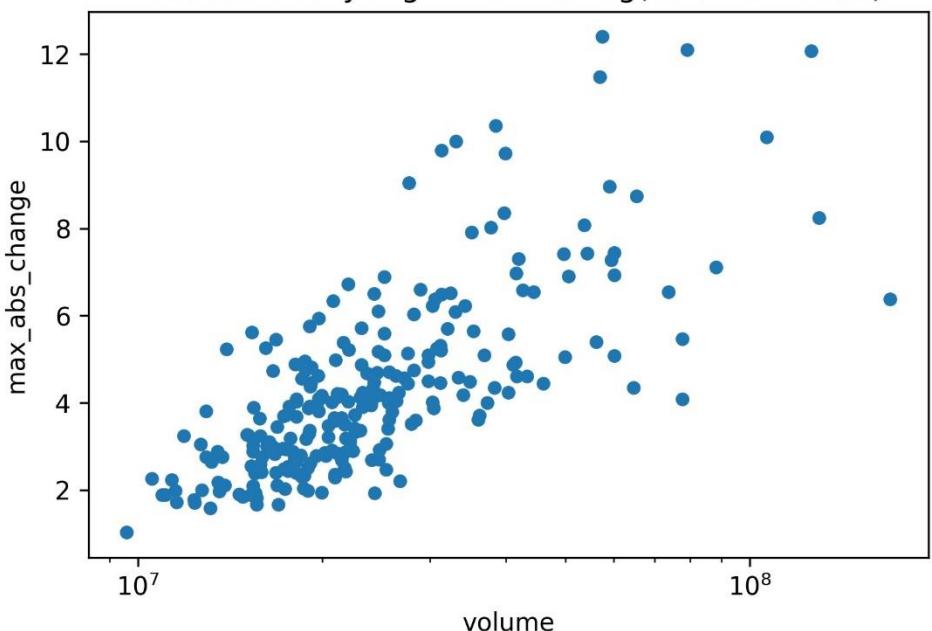




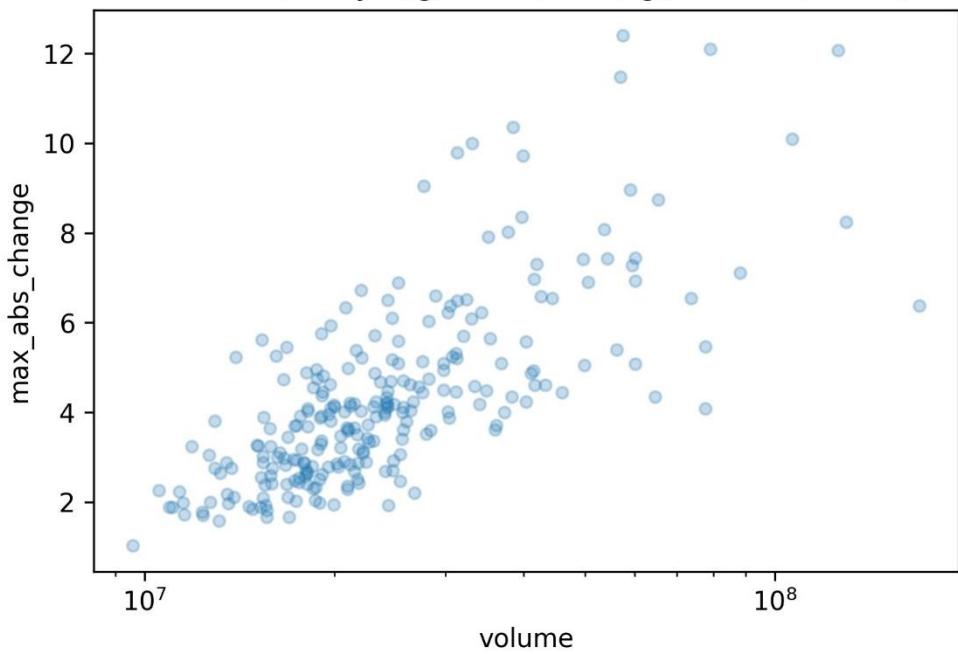
Facebook Daily High - Low vs. Volume Traded



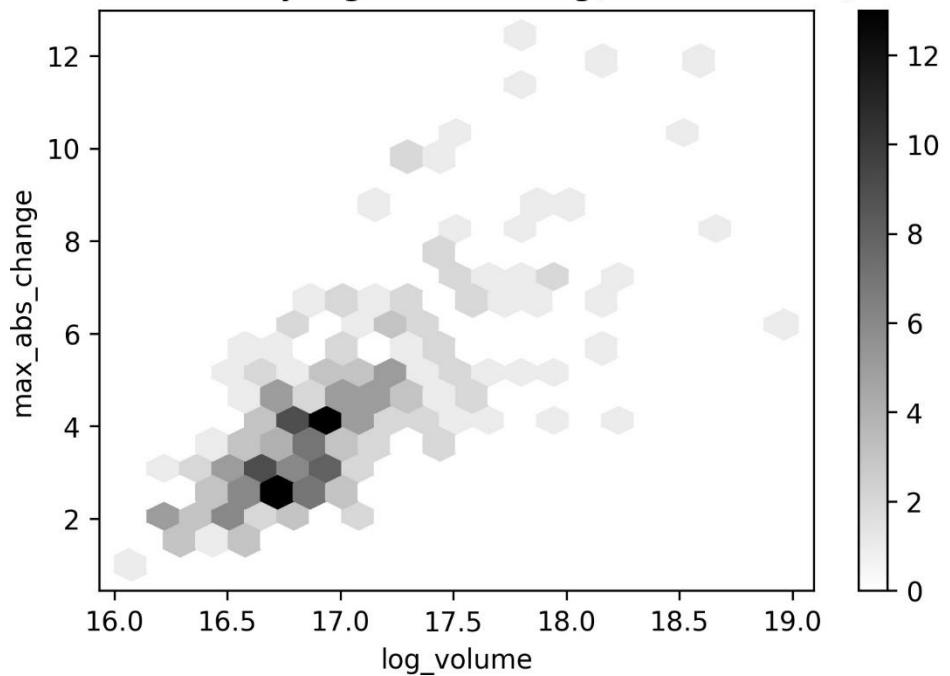
Facebook Daily High - Low vs. log(Volume Traded)

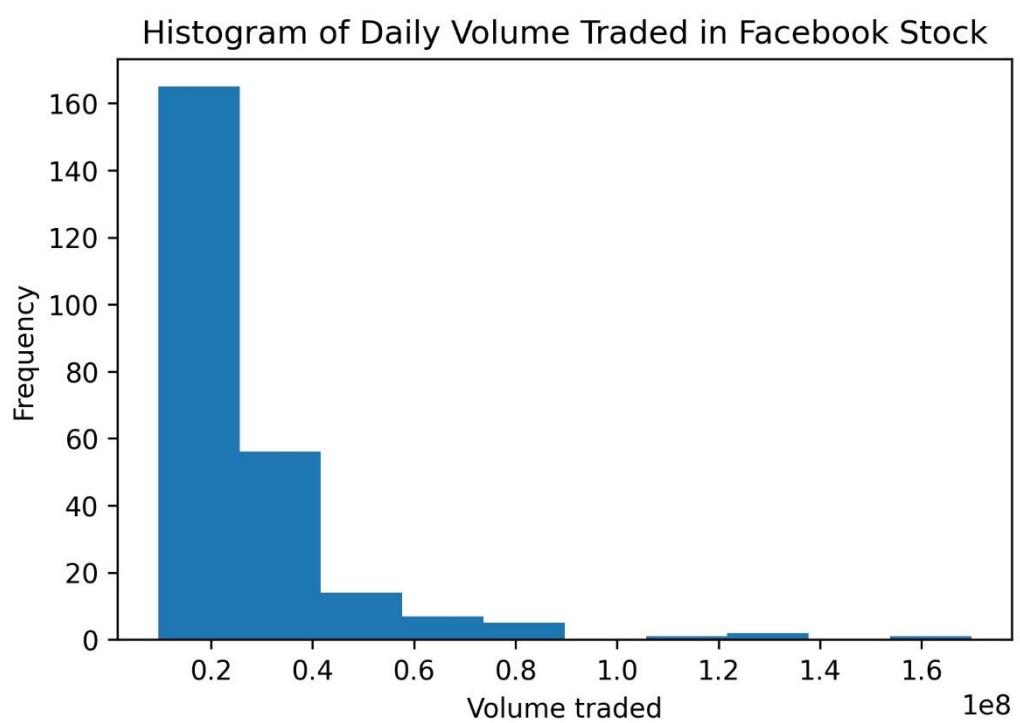
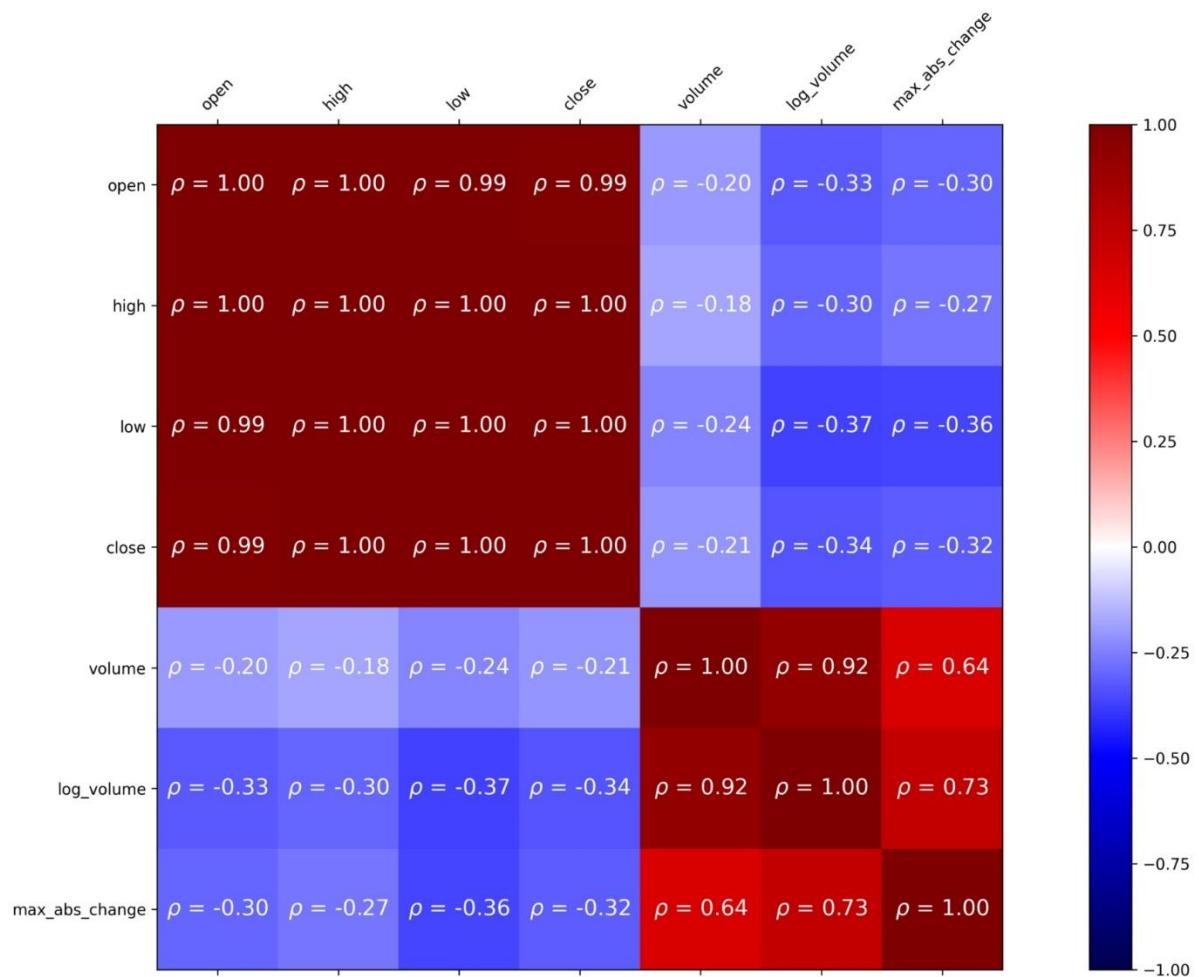


Facebook Daily High - Low vs. log(Volume Traded)

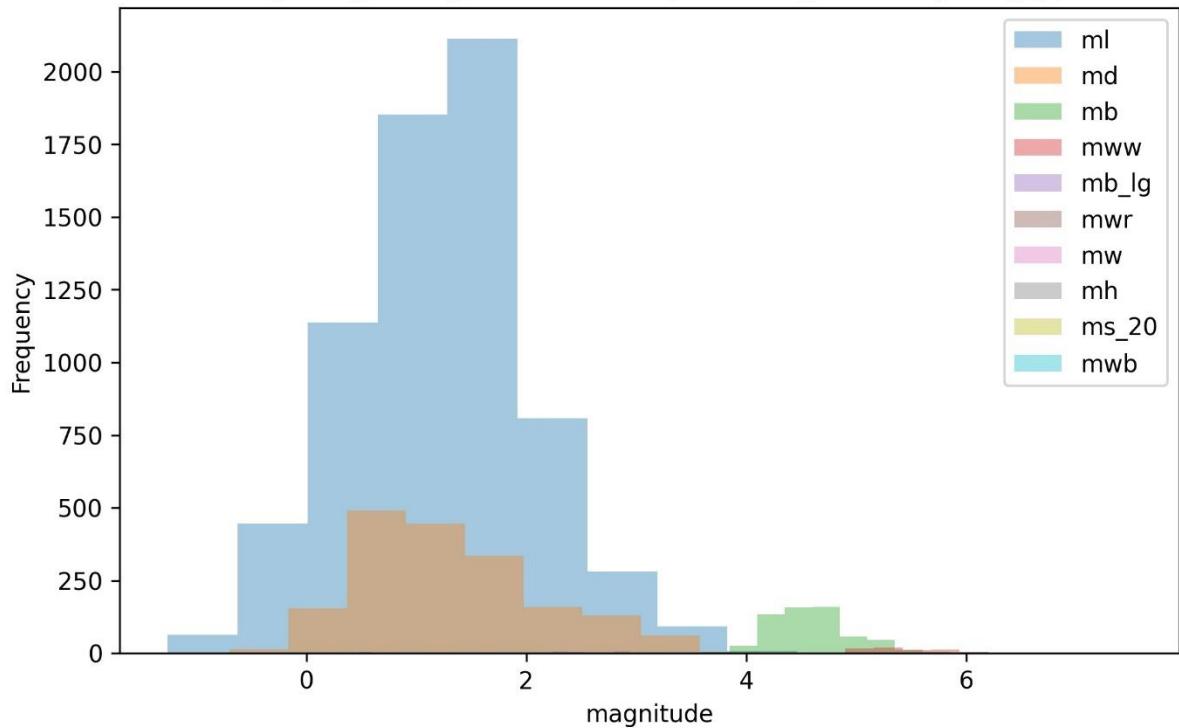


Facebook Daily High - Low vs. log(Volume Traded)

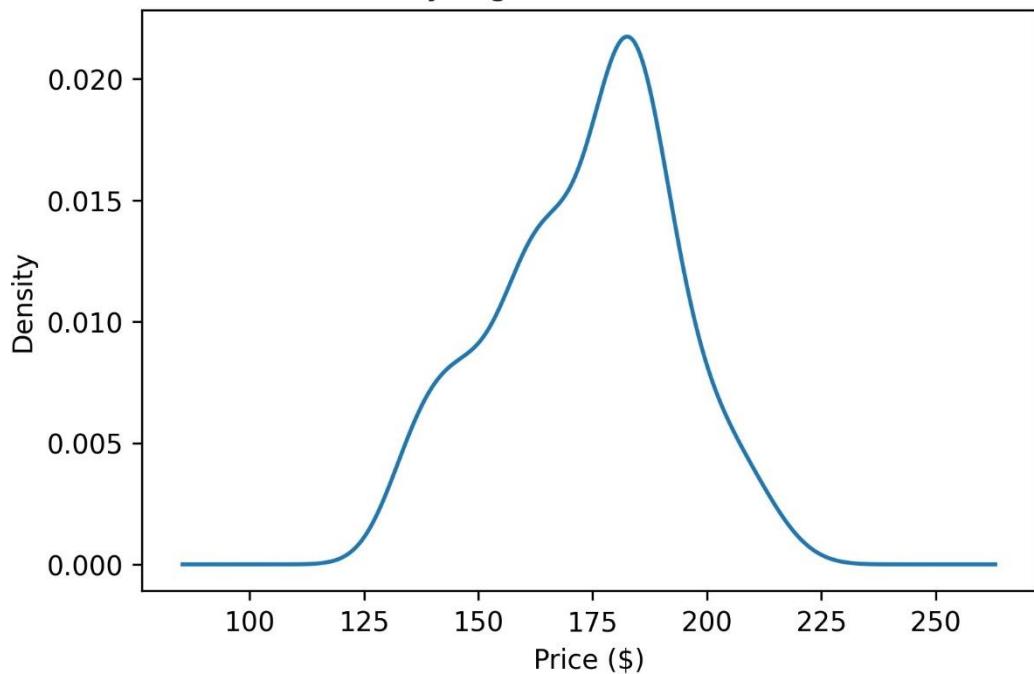




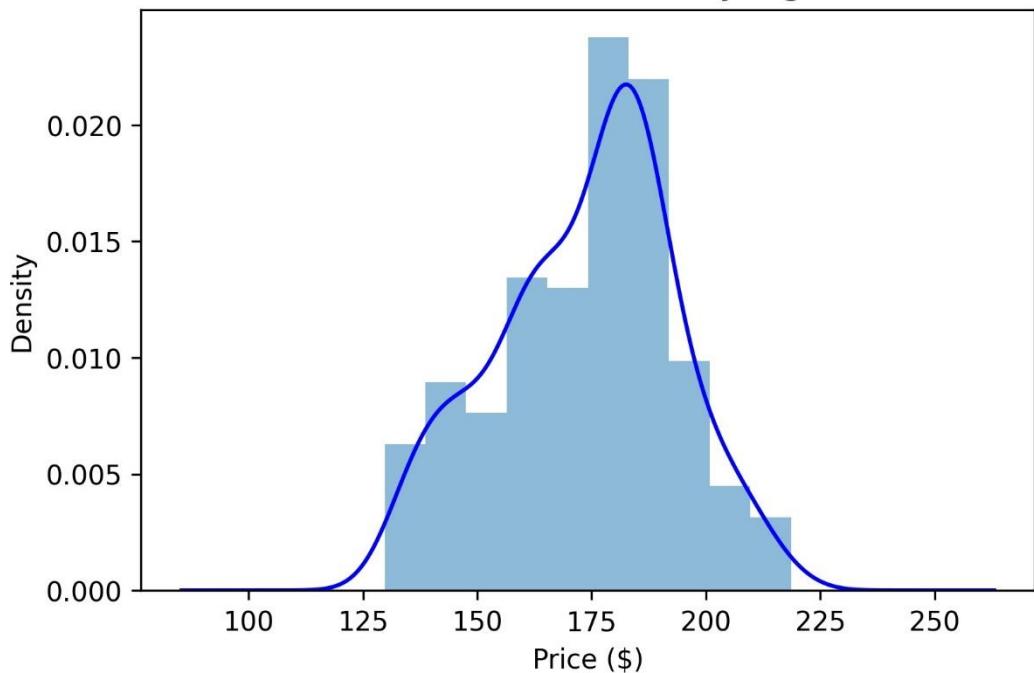
Comparing histograms of earthquake magnitude by magType



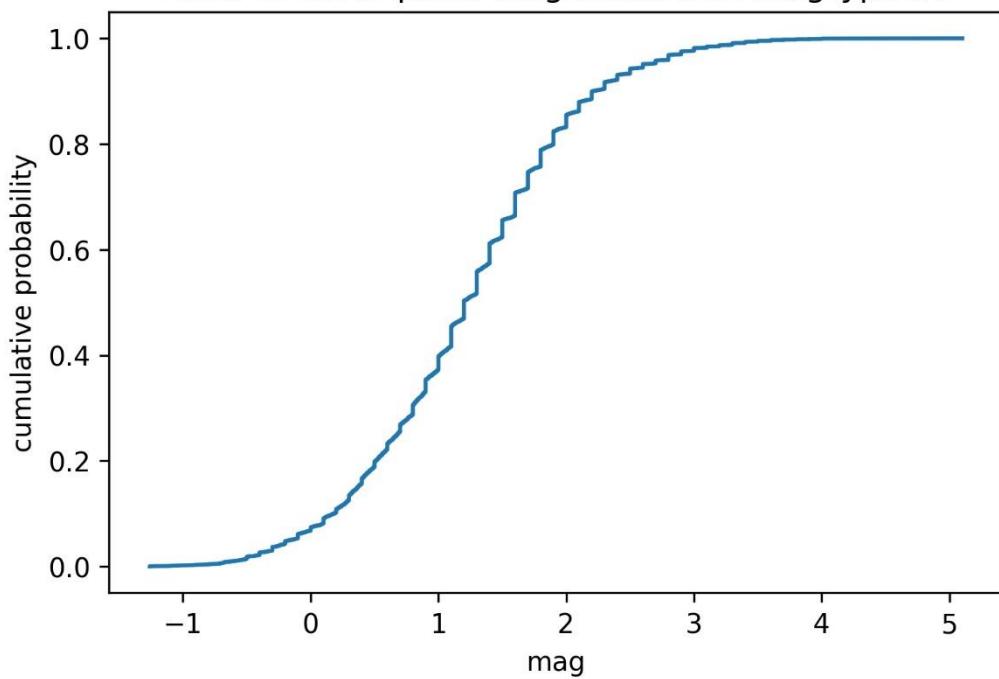
KDE of Daily High Price for Facebook Stock

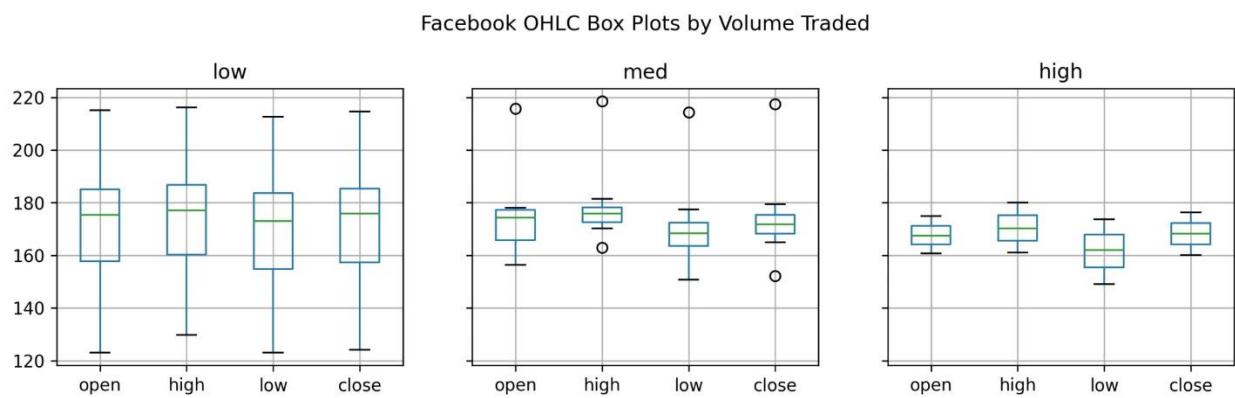
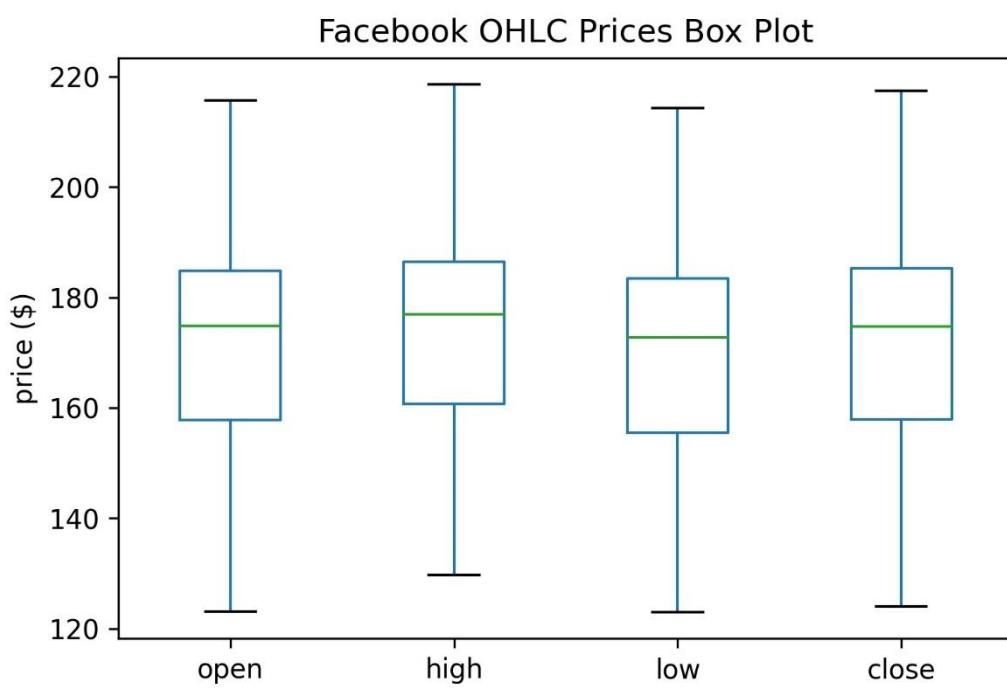
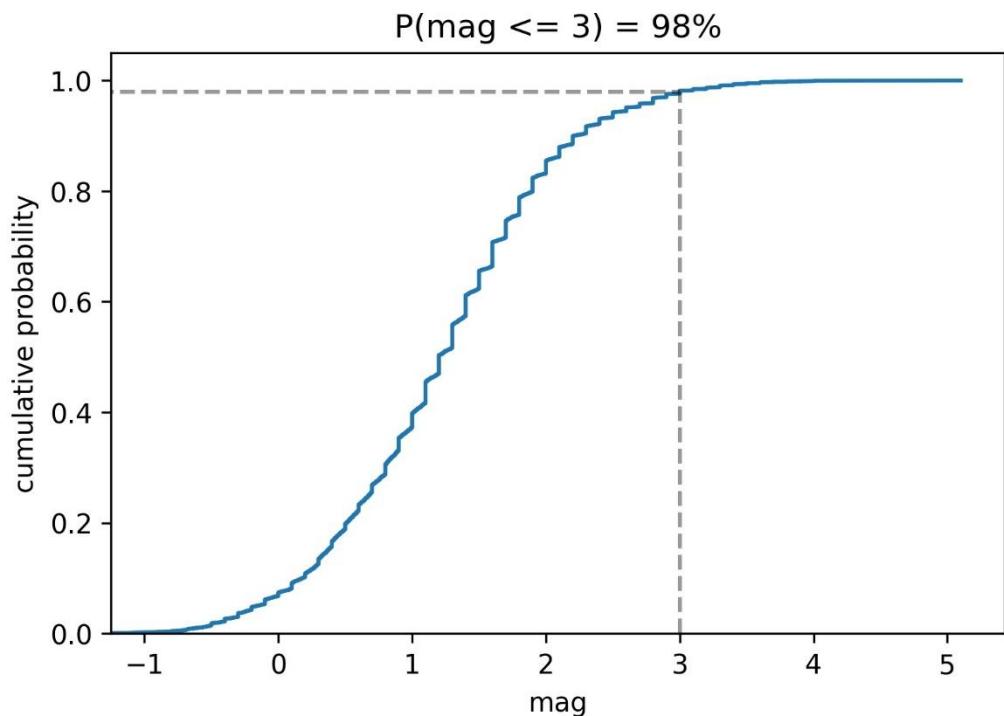


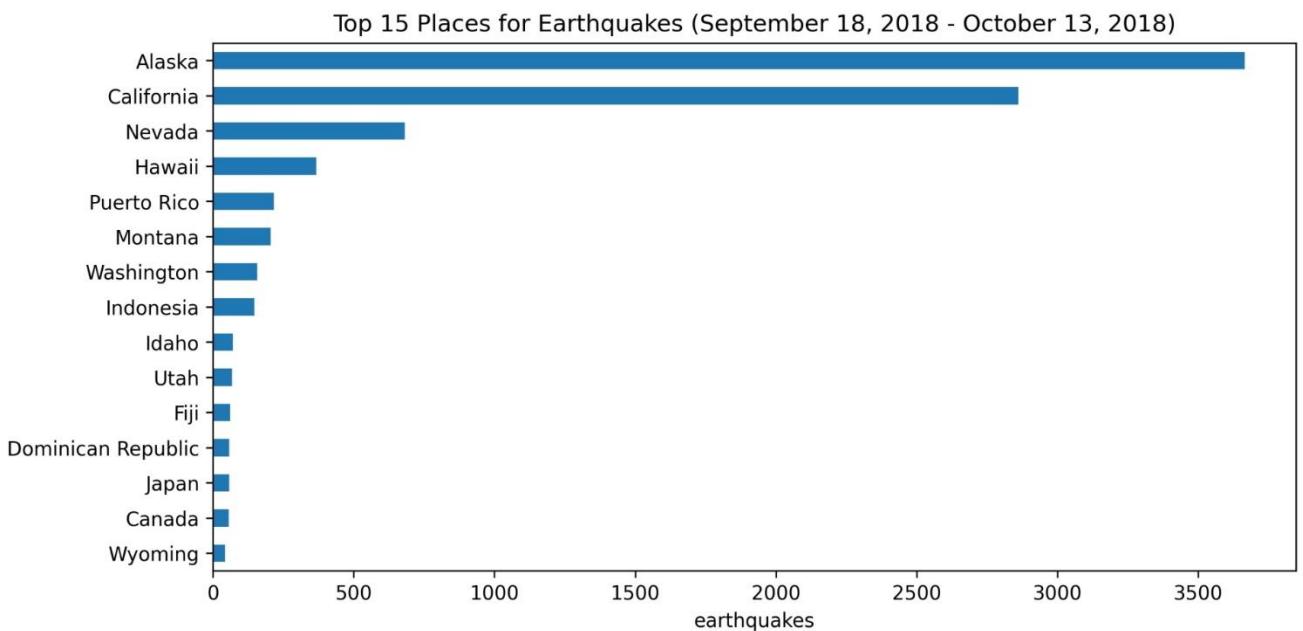
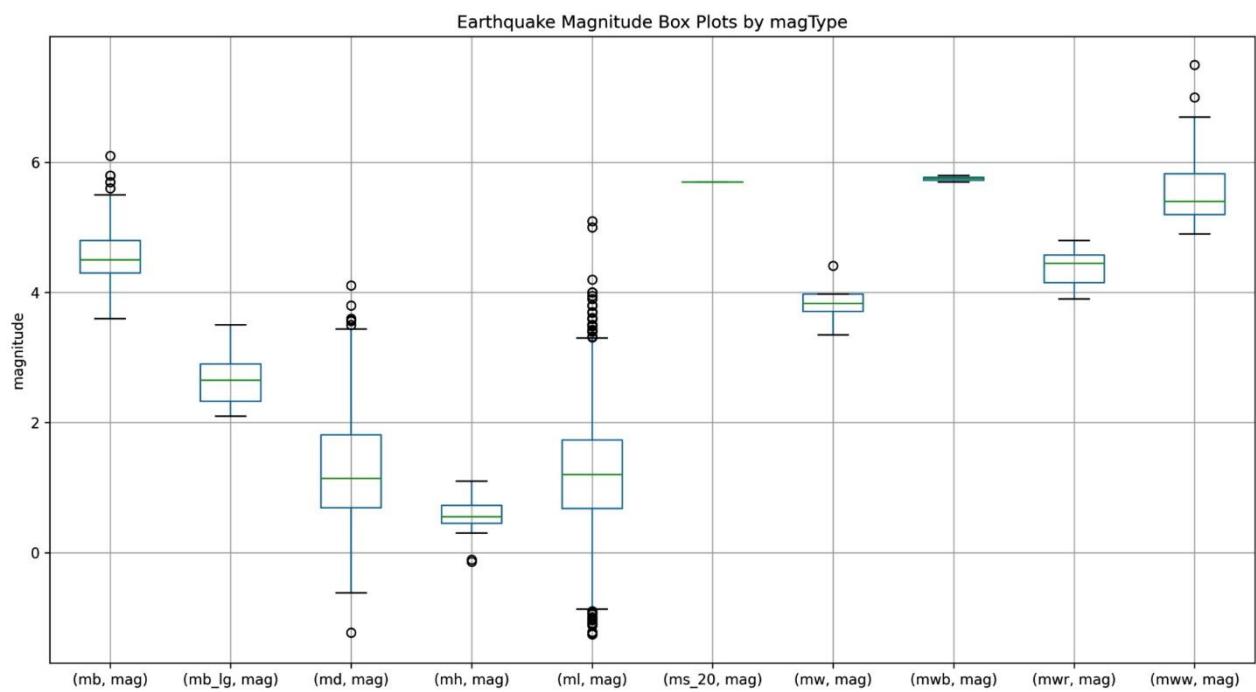
Distribution of Facebook Stock's Daily High Price in 2018

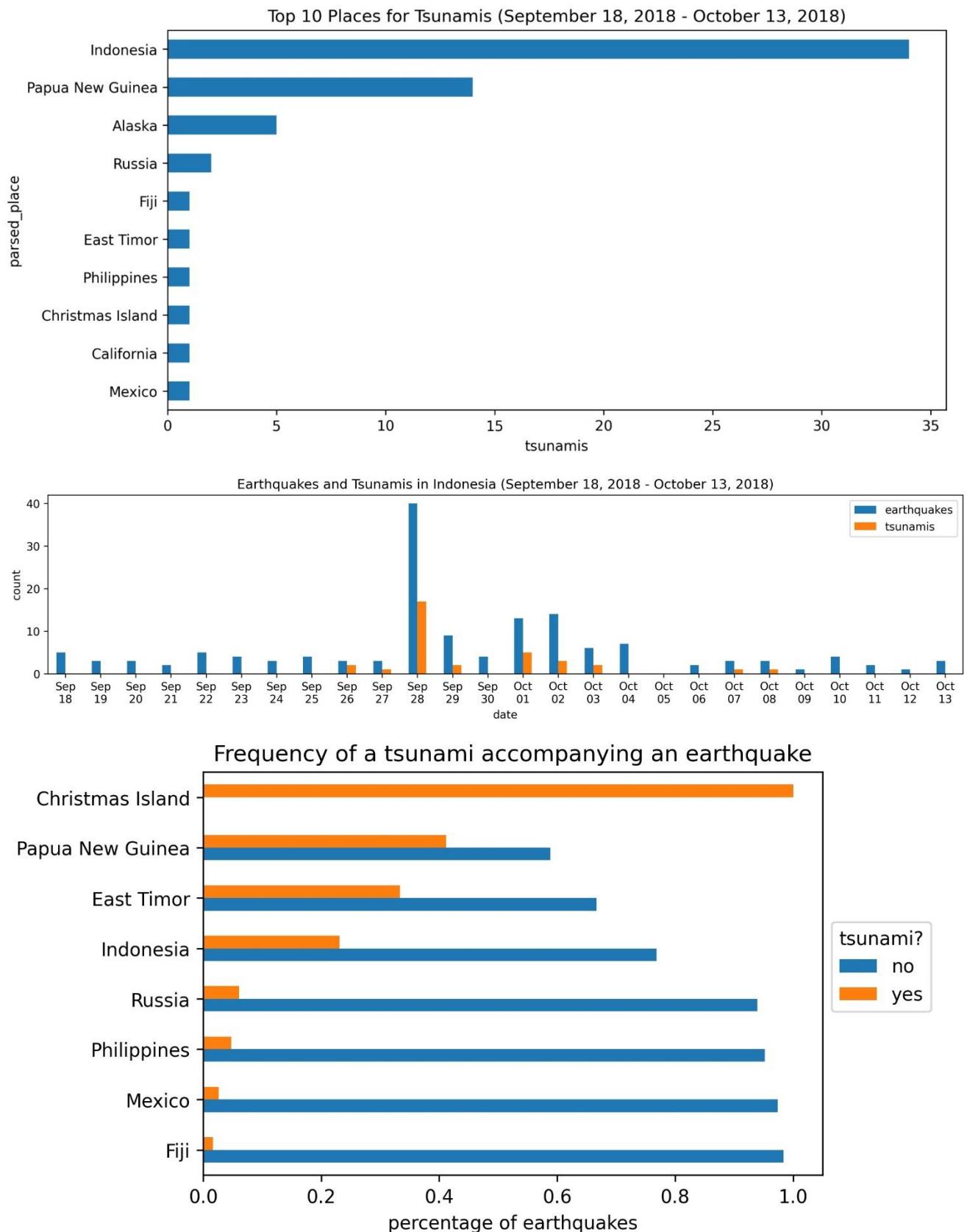


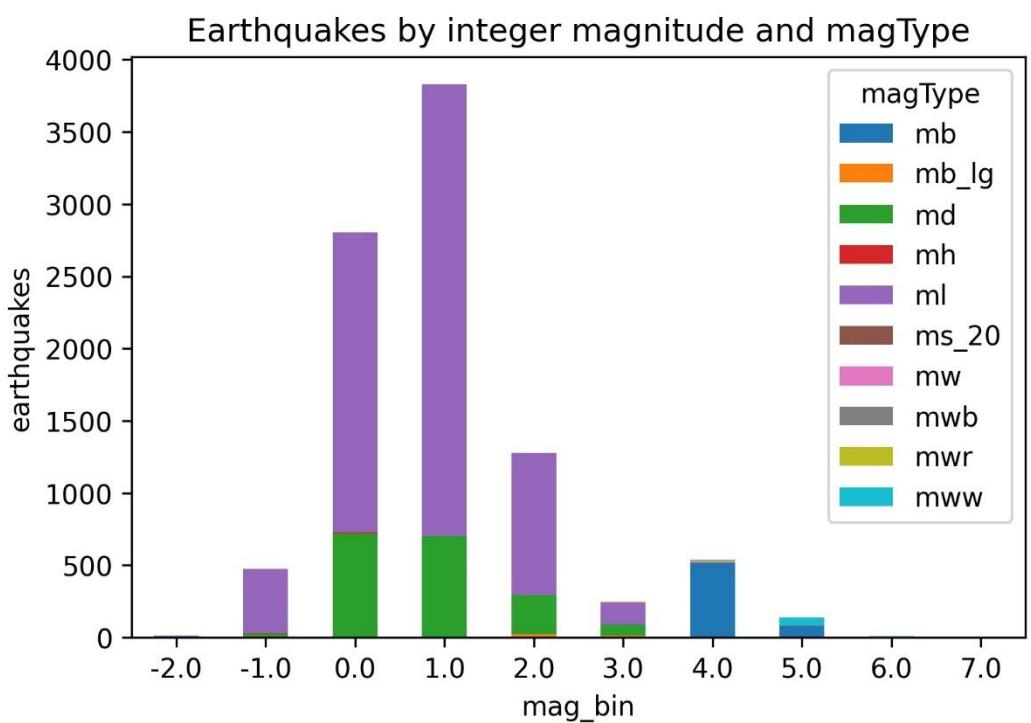
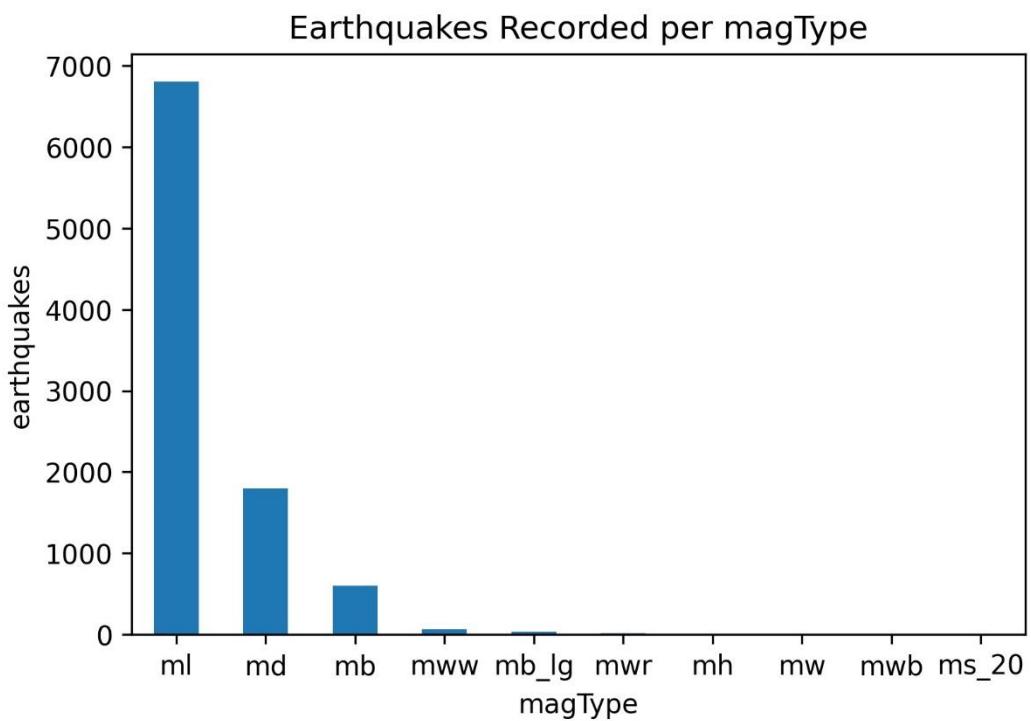
ECDF of earthquake magnitude with magType ml

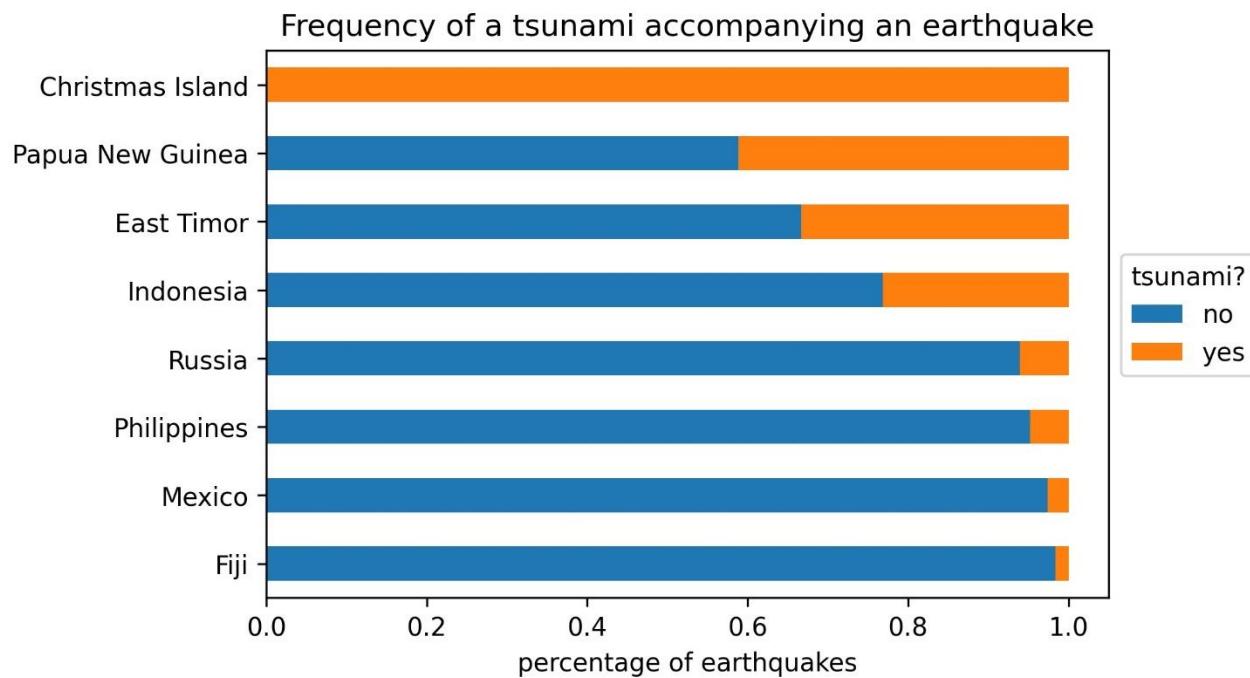
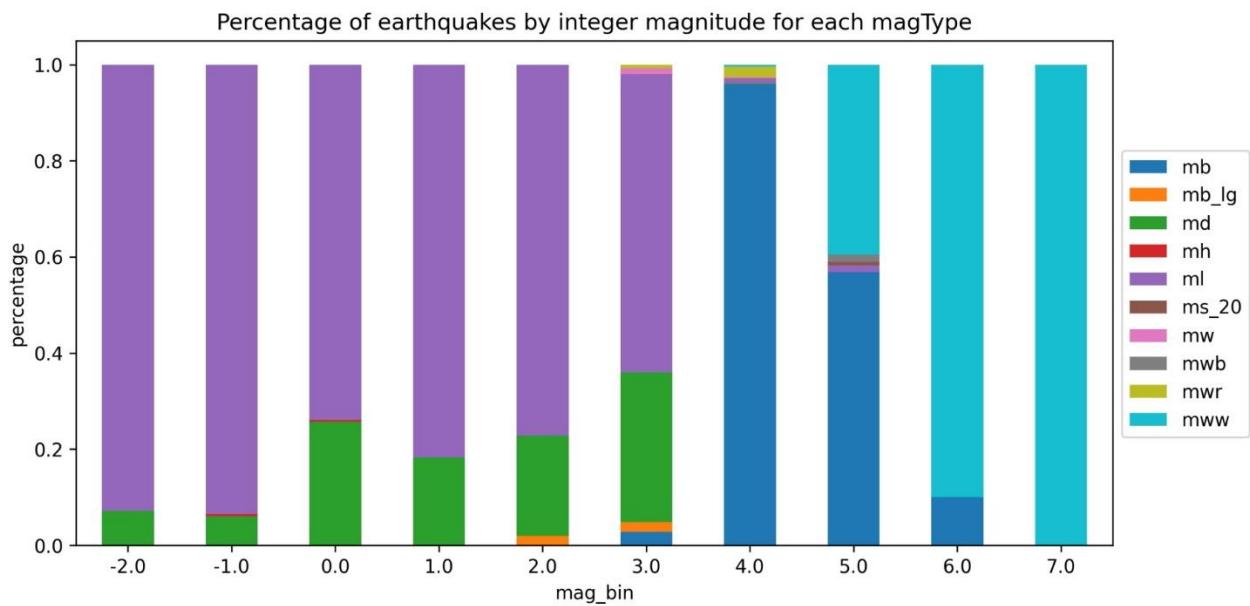


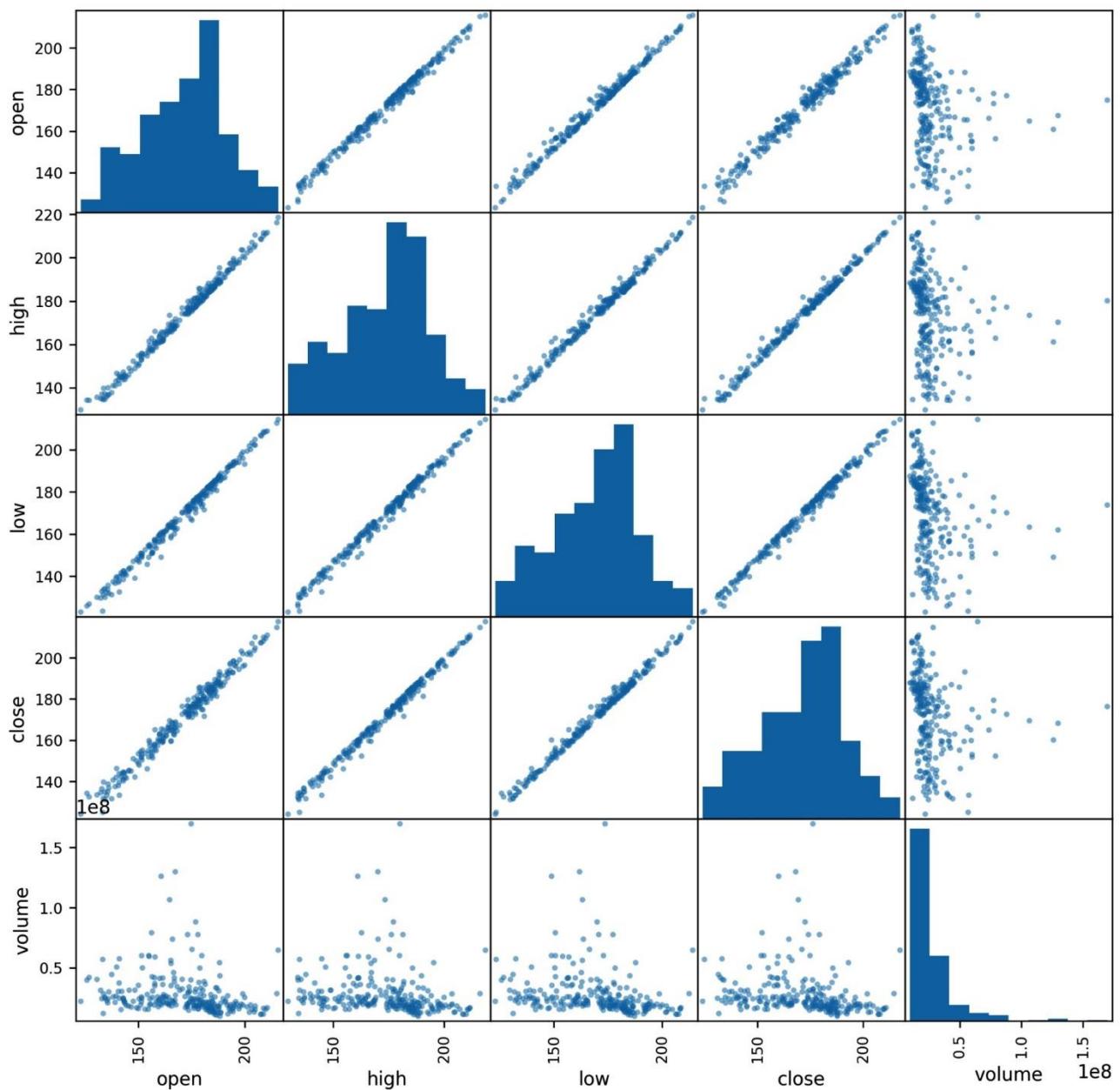


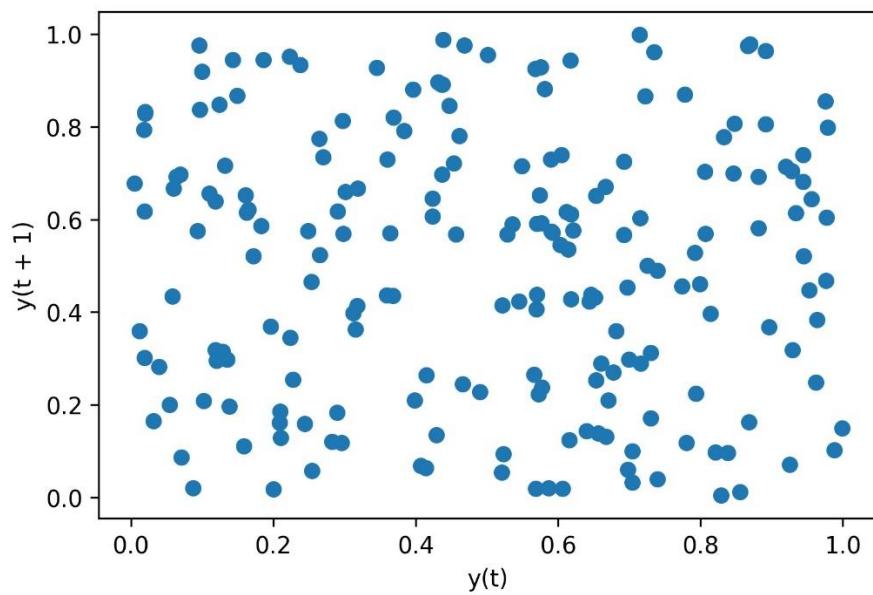
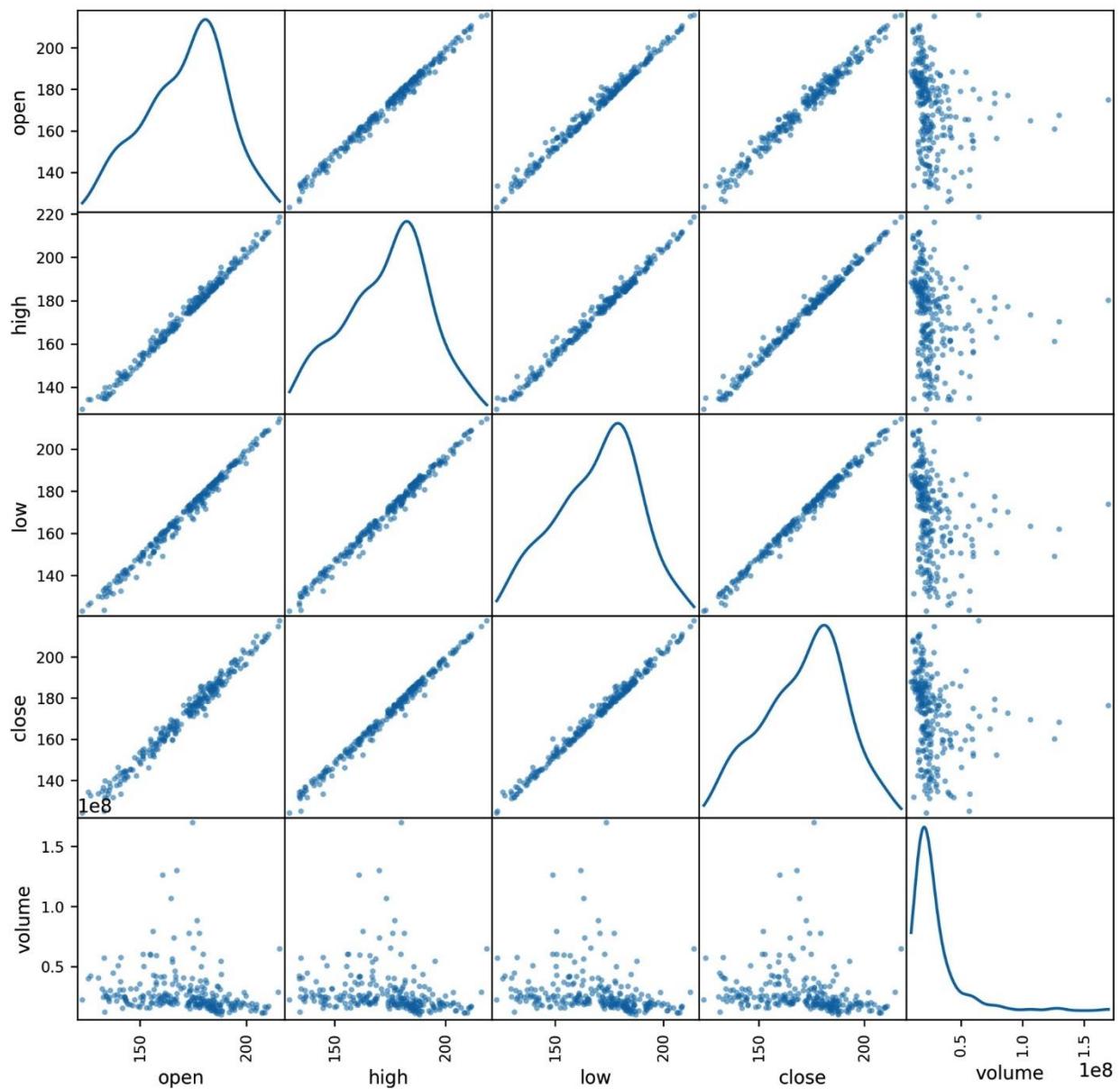


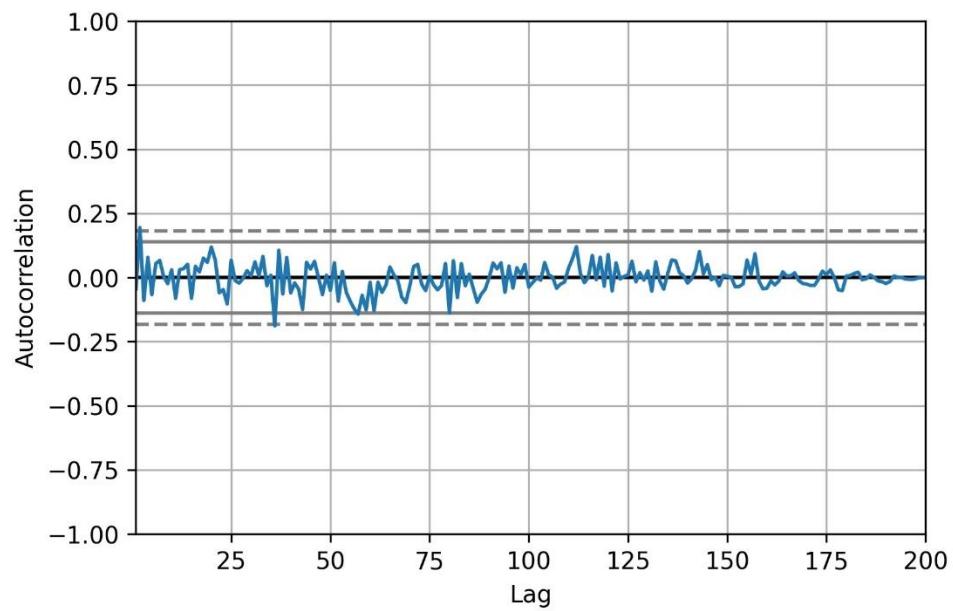
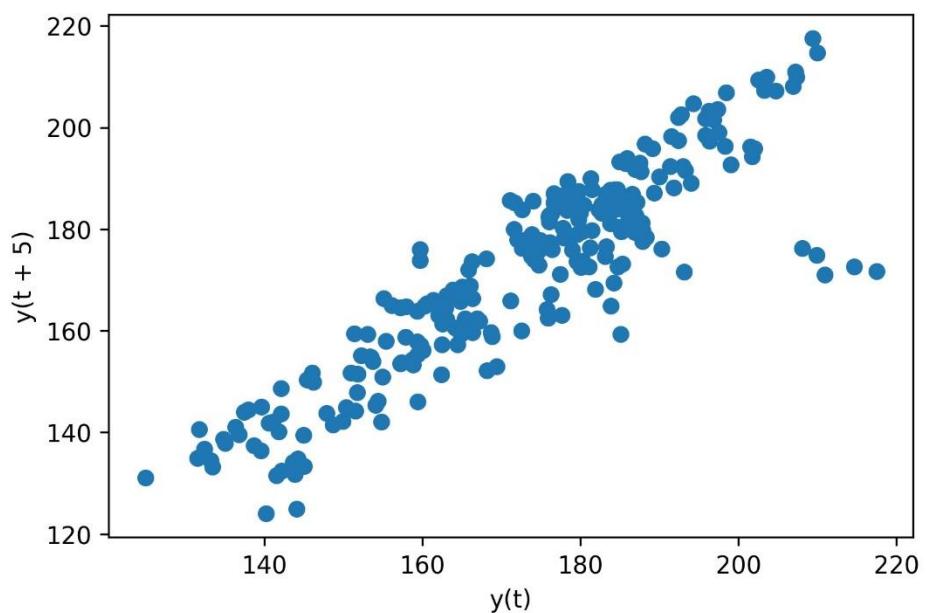
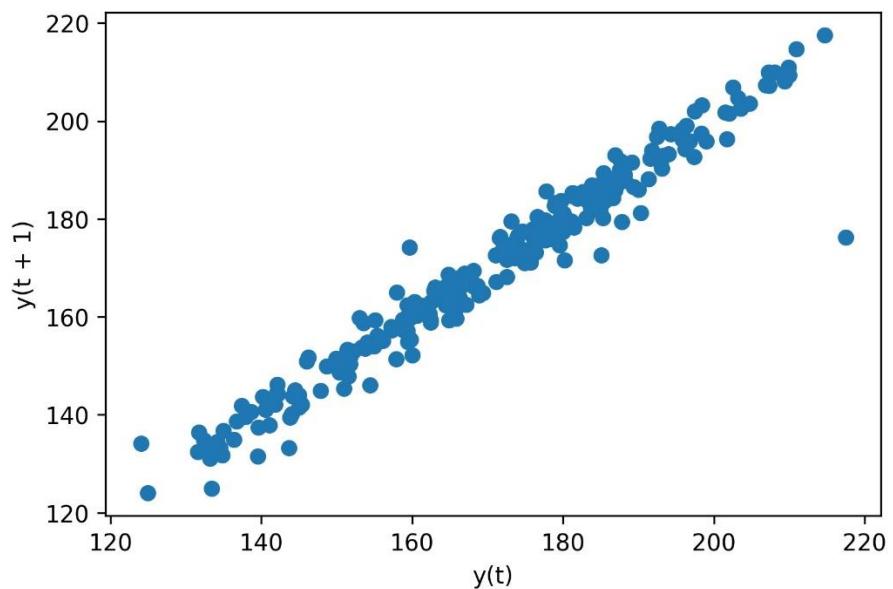


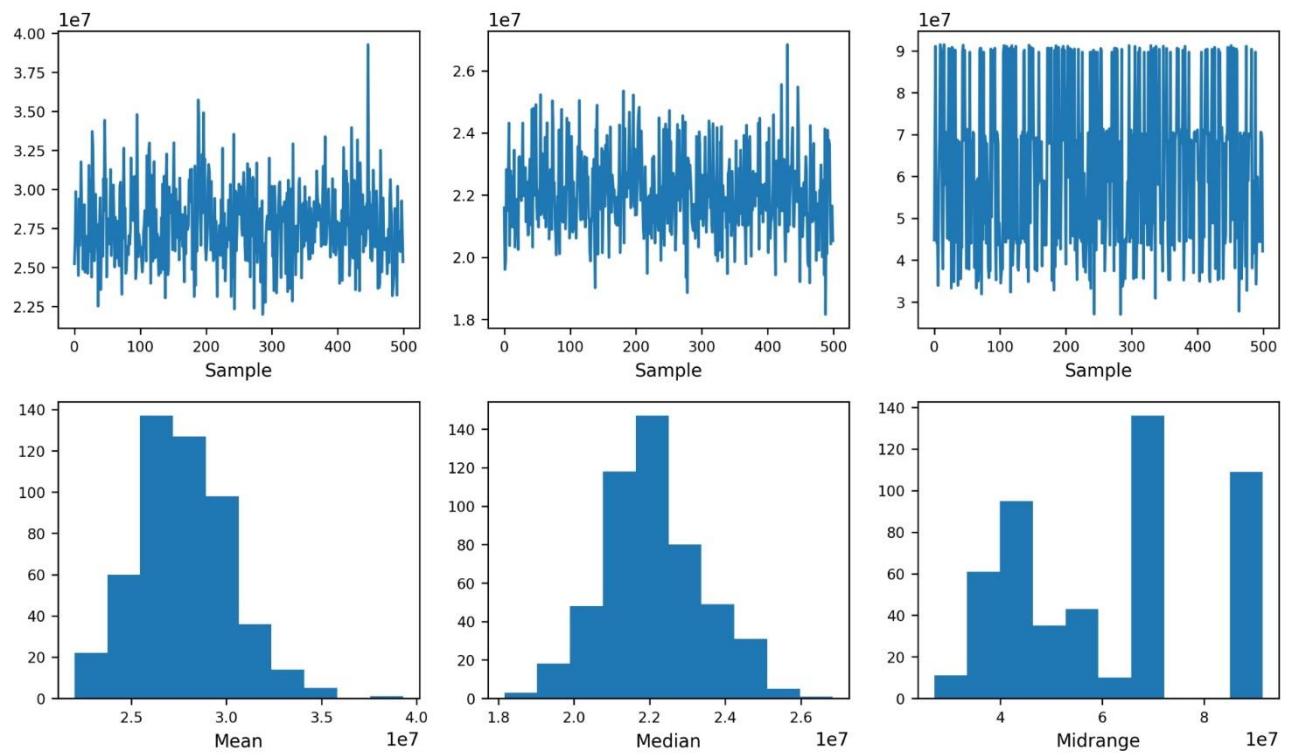
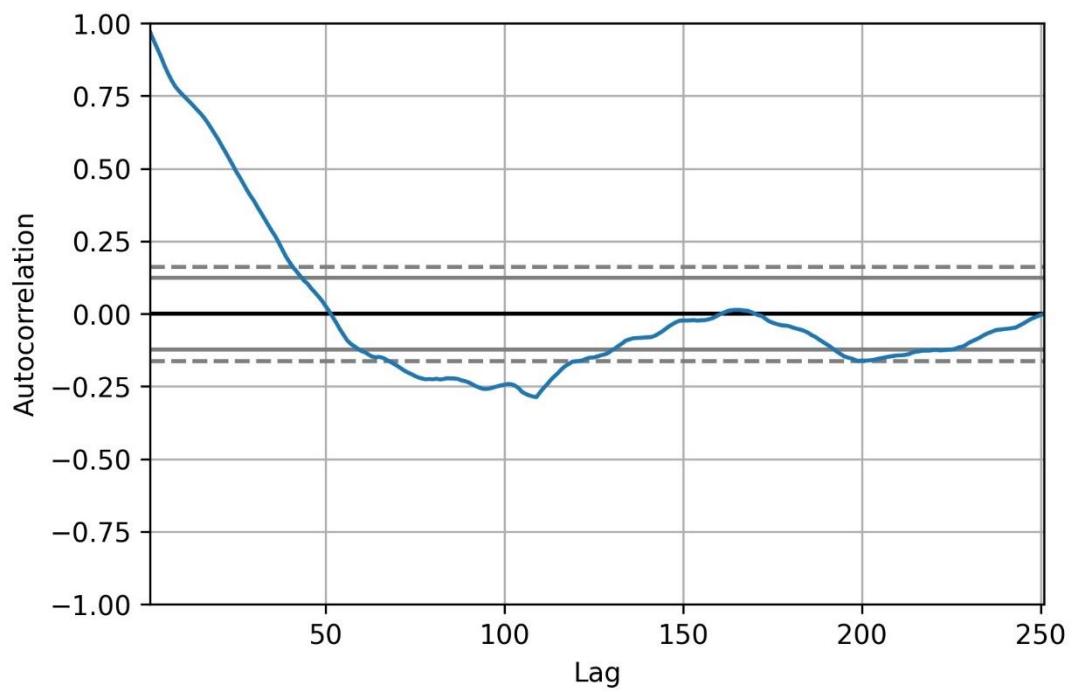




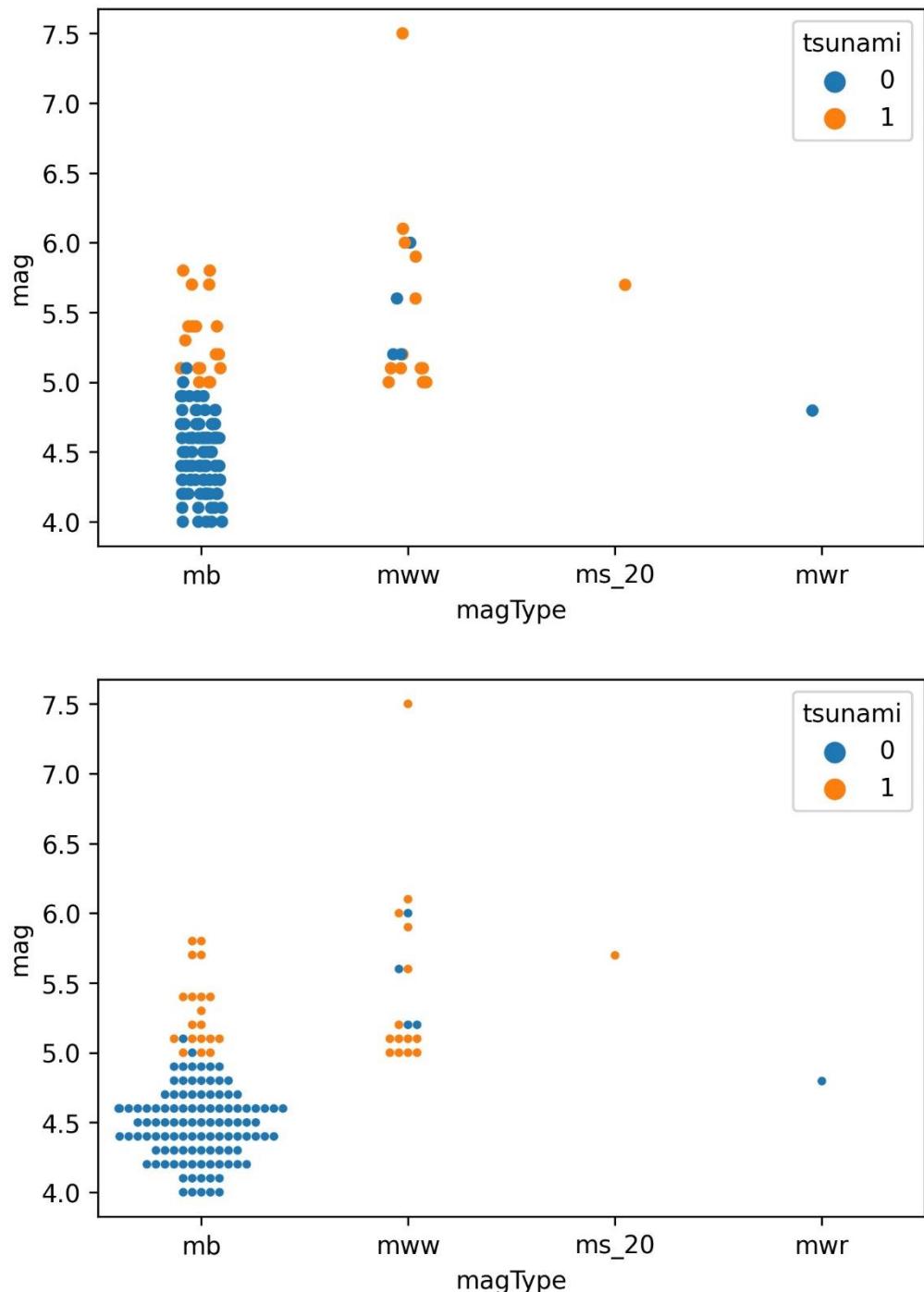




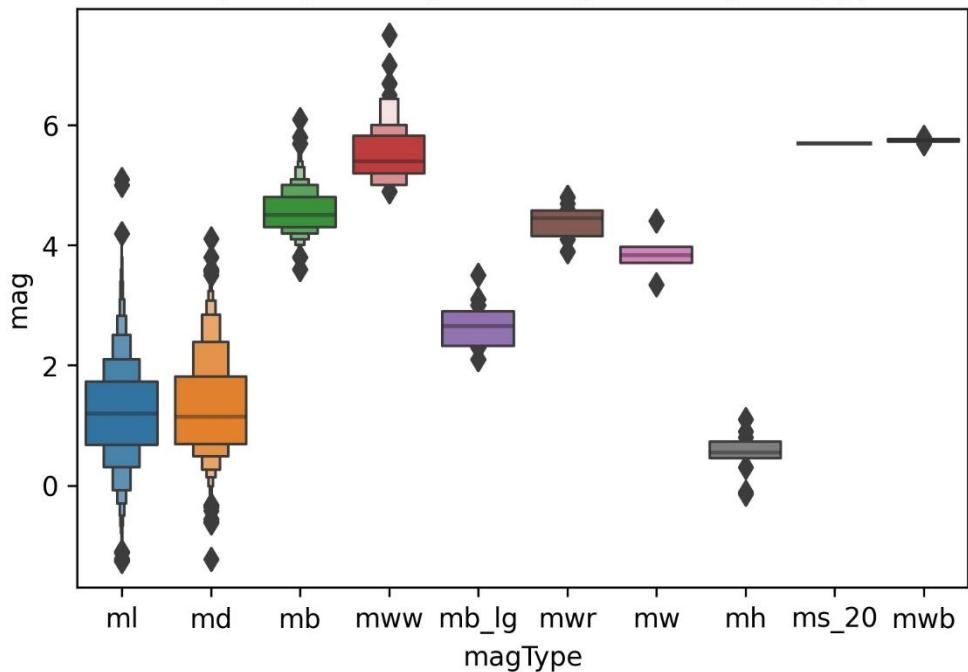




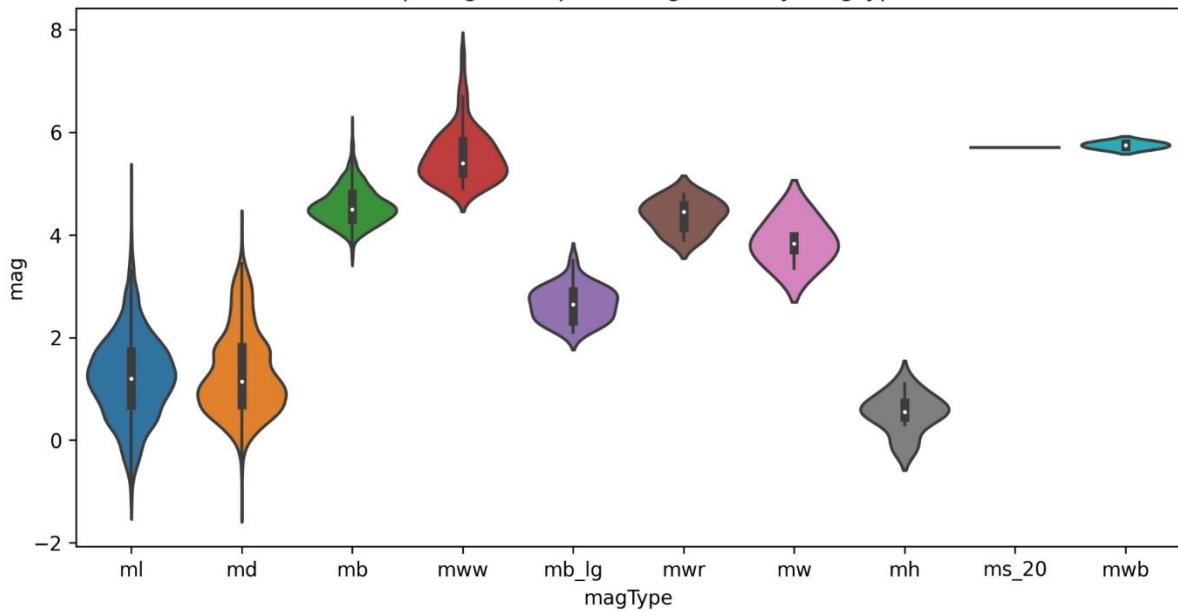
Chapter 6: Plotting with Seaborn and Customization Techniques

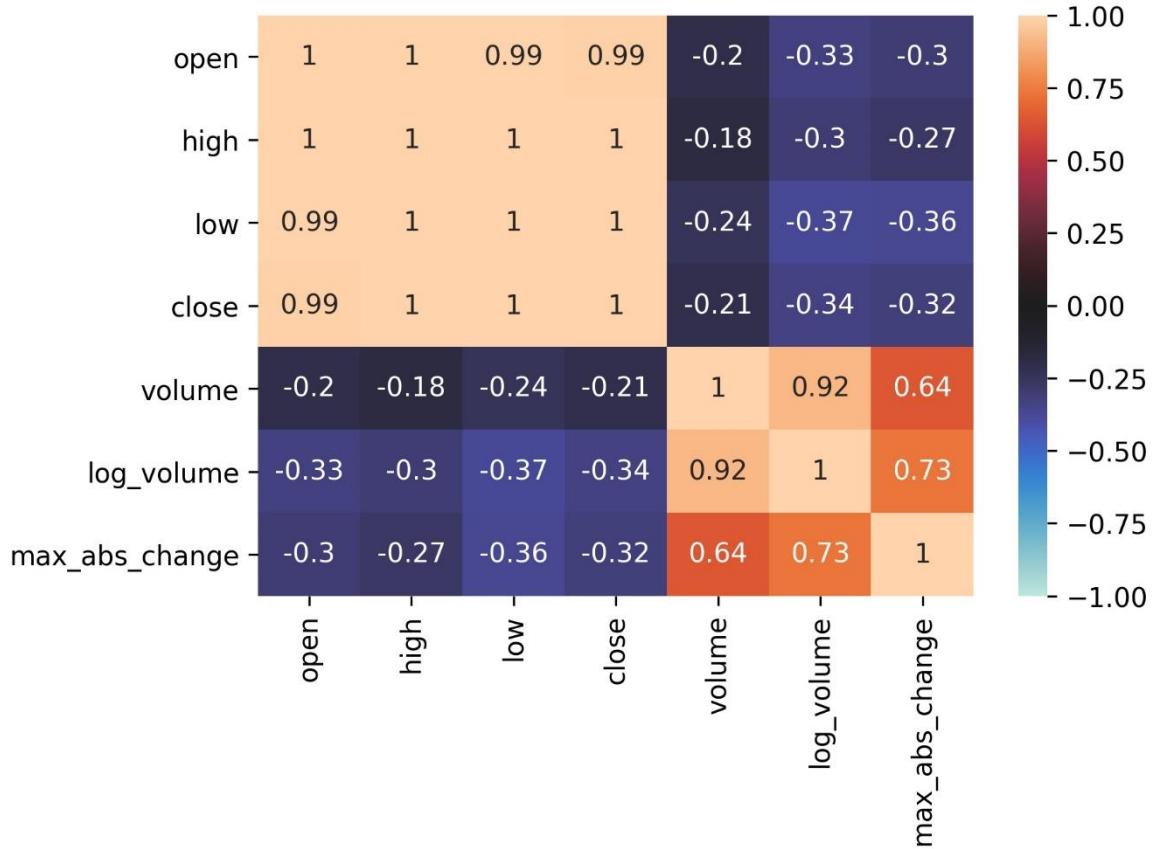


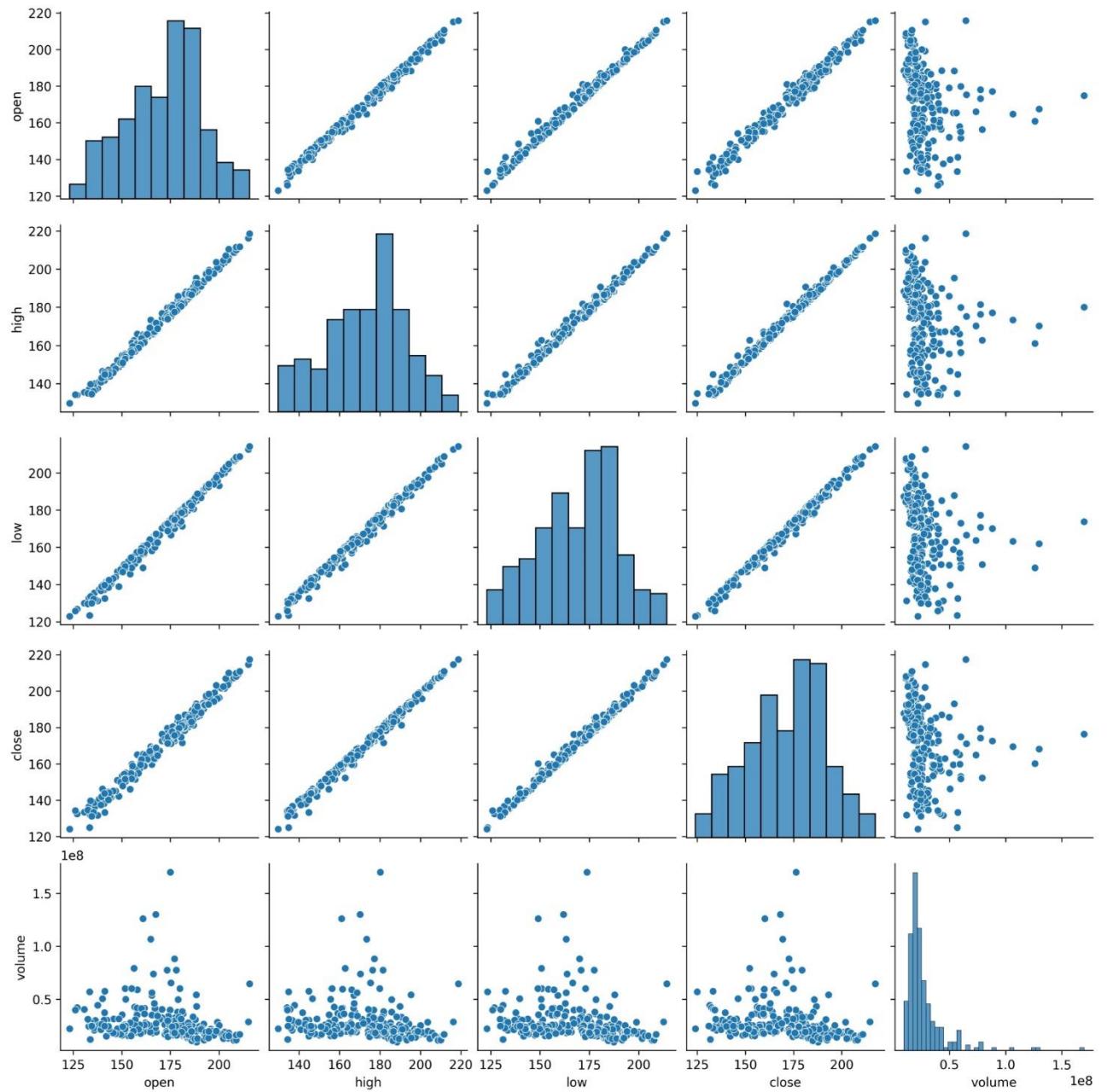
Comparing earthquake magnitude by magType

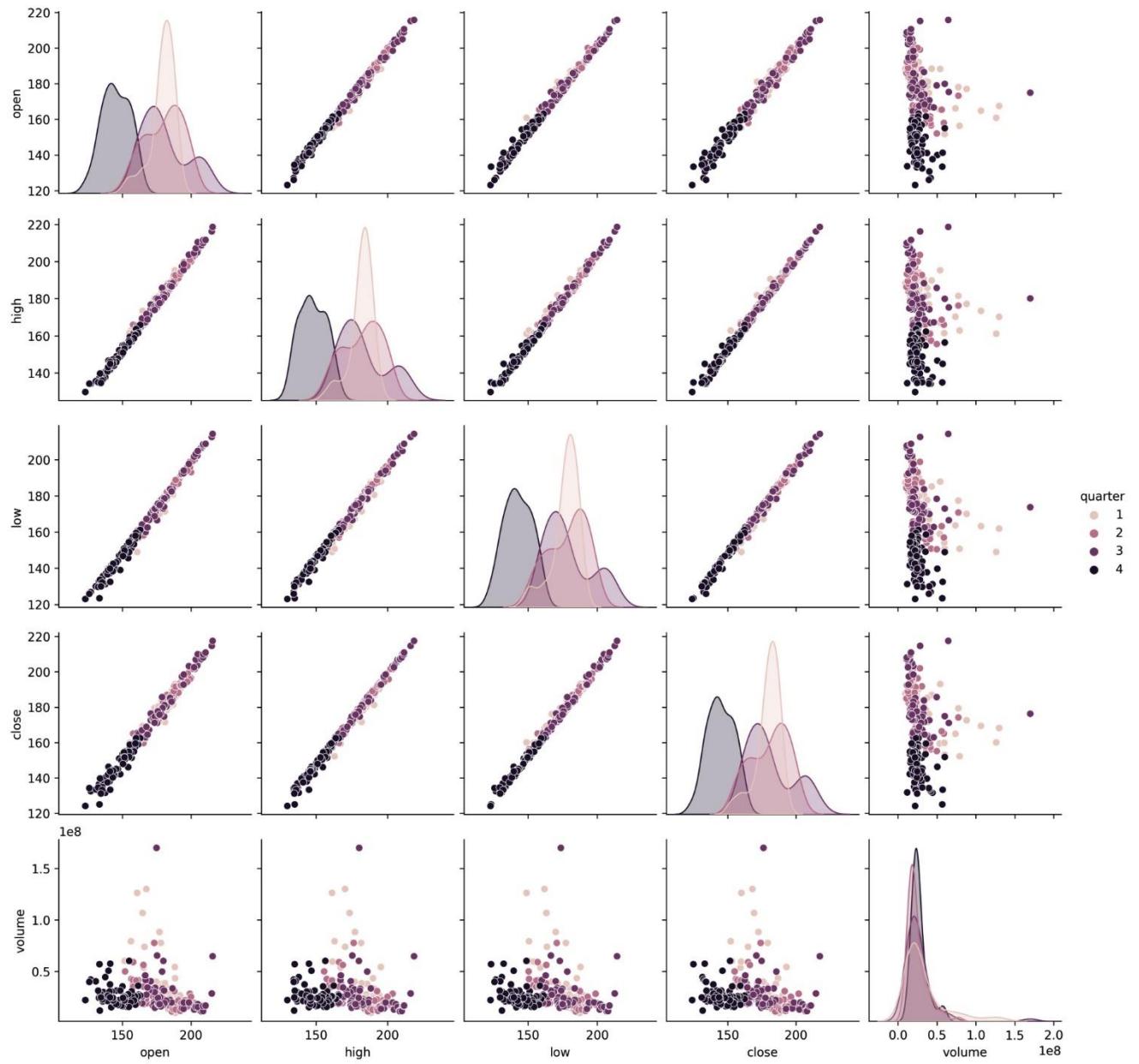


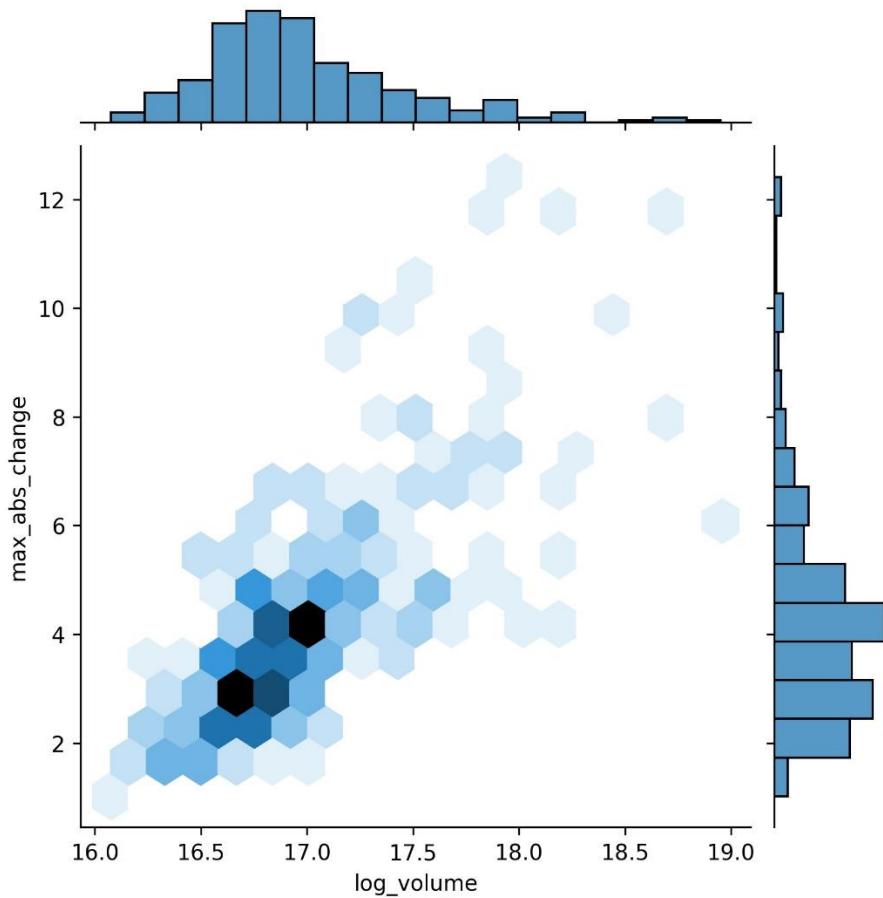
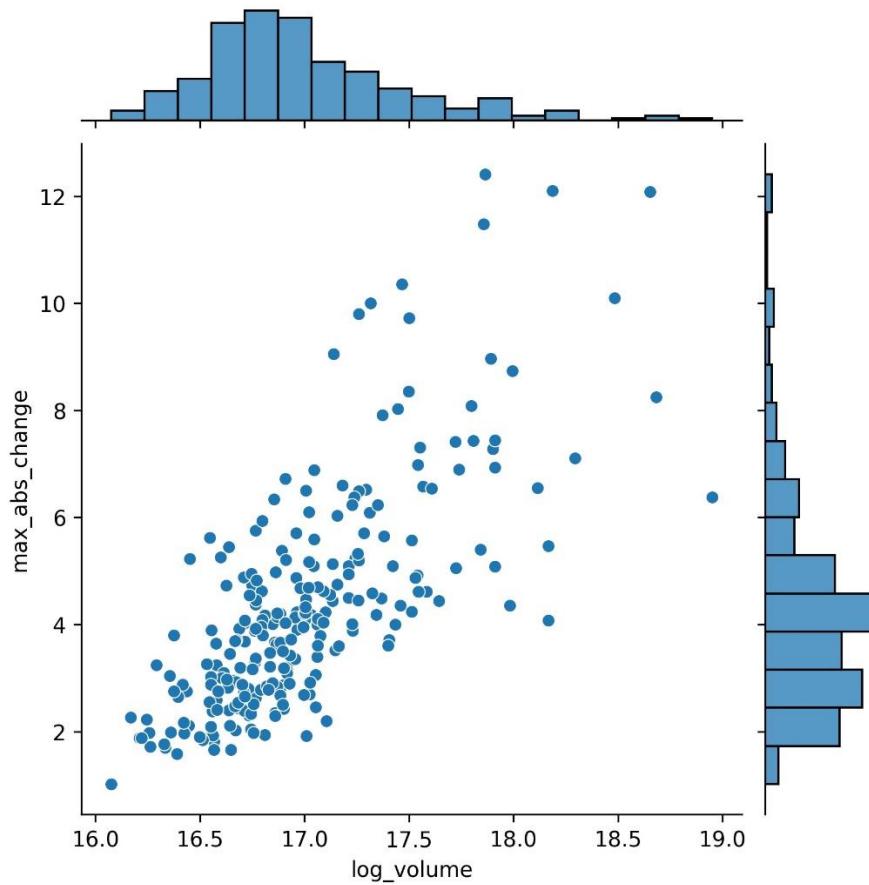
Comparing earthquake magnitude by magType

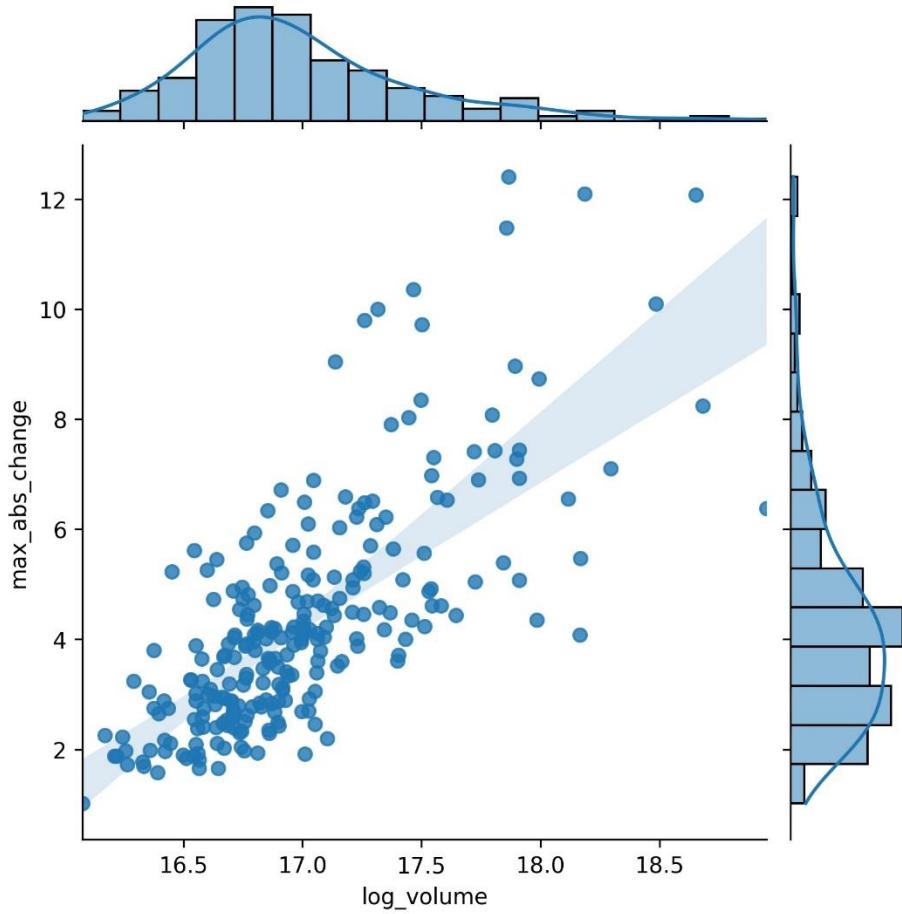
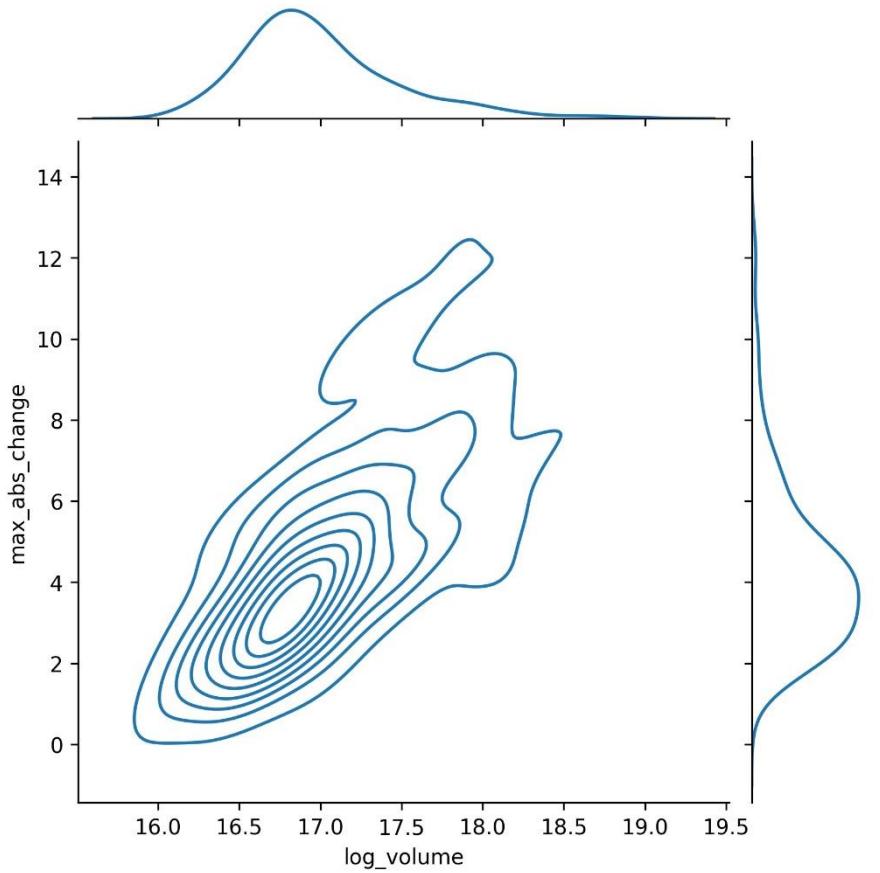


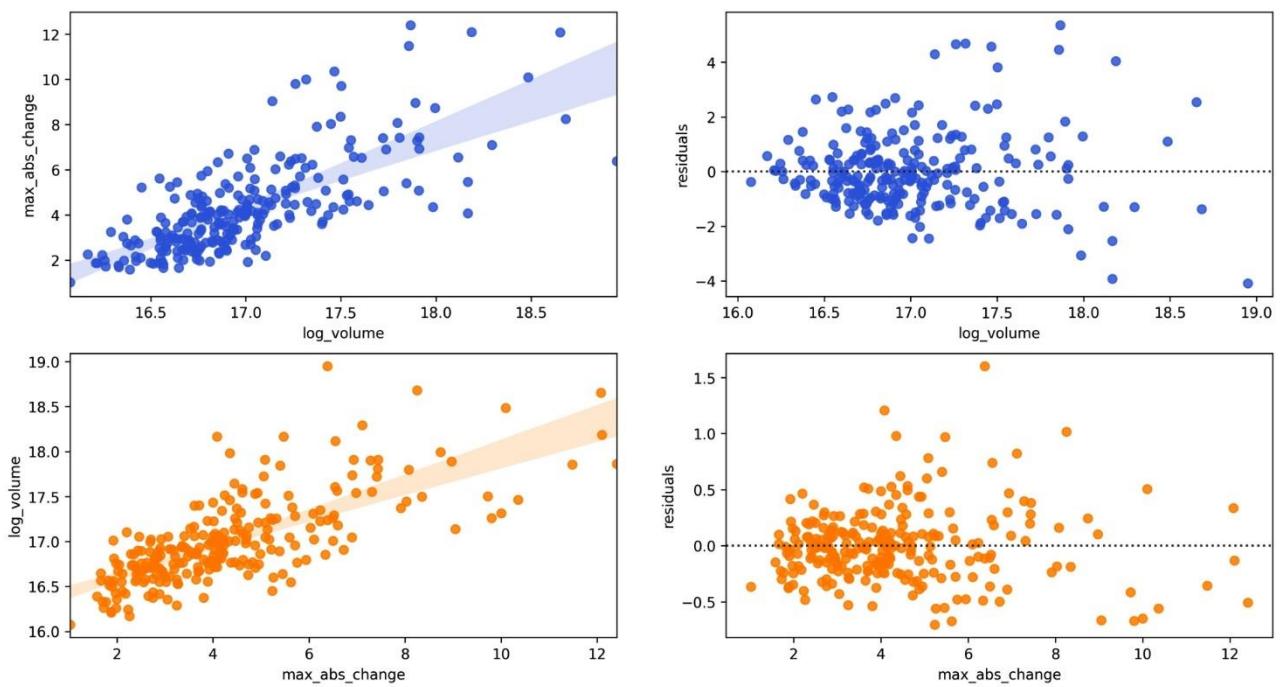
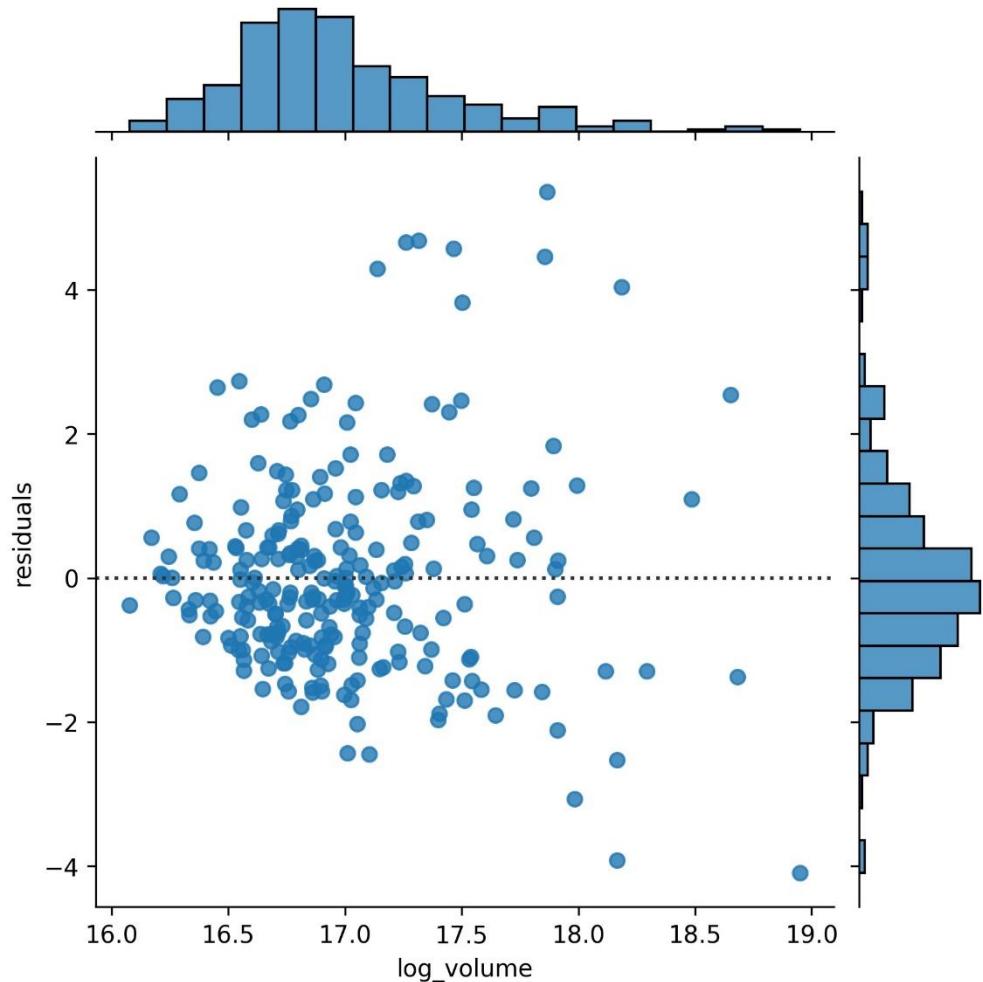


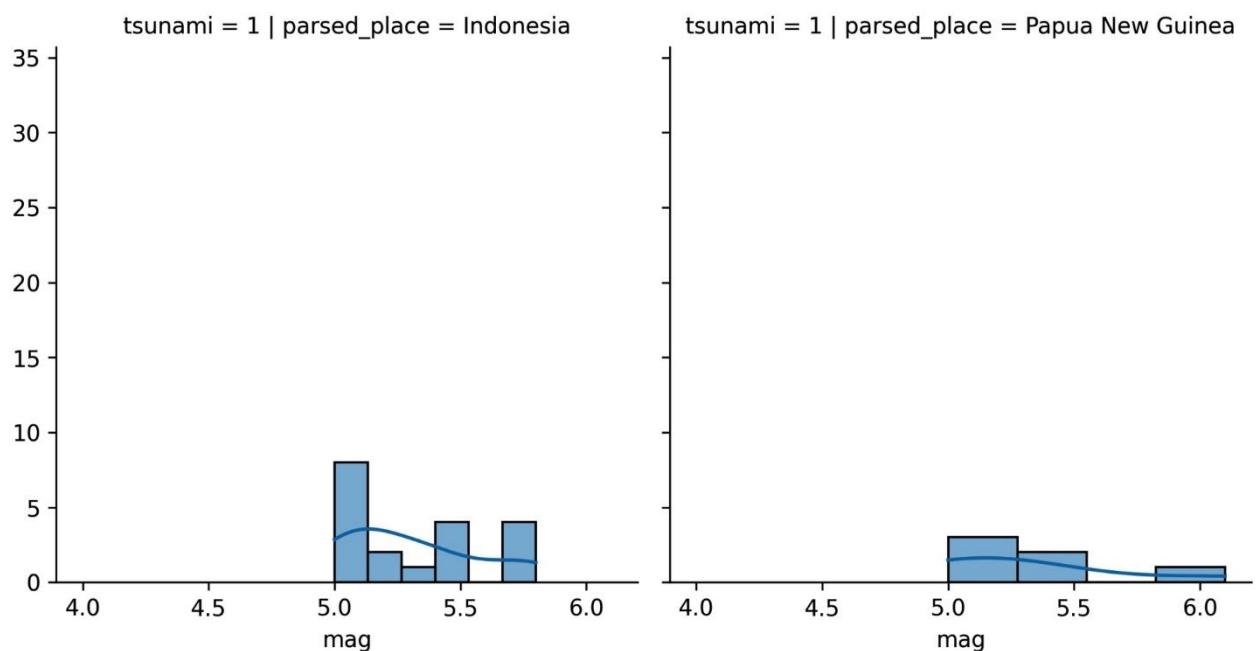
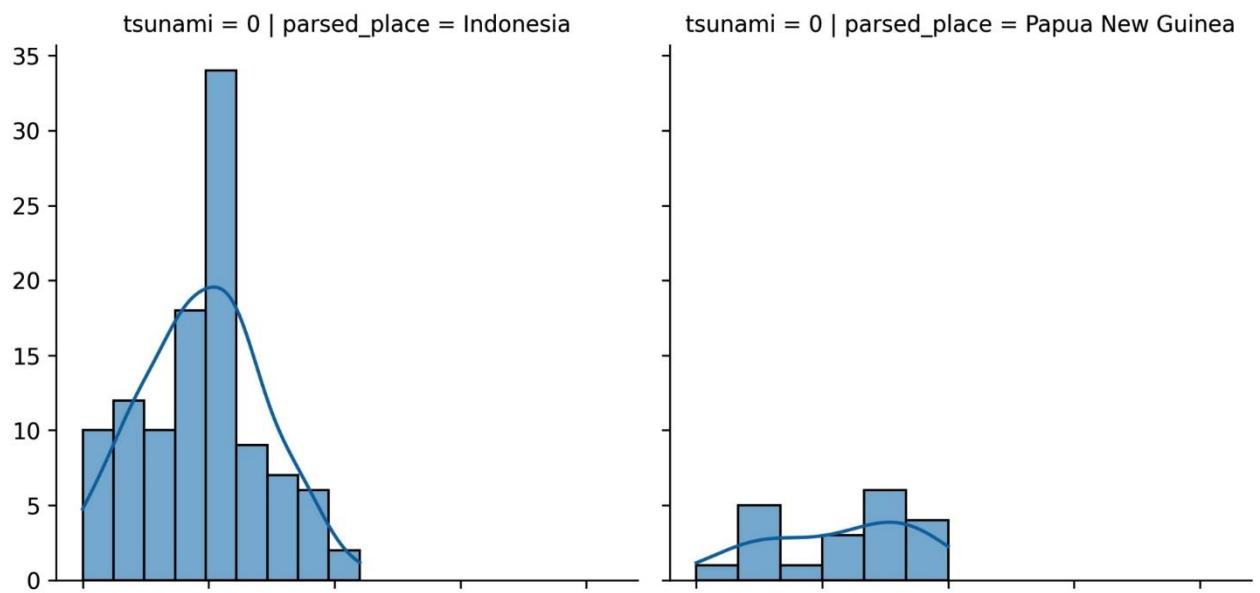
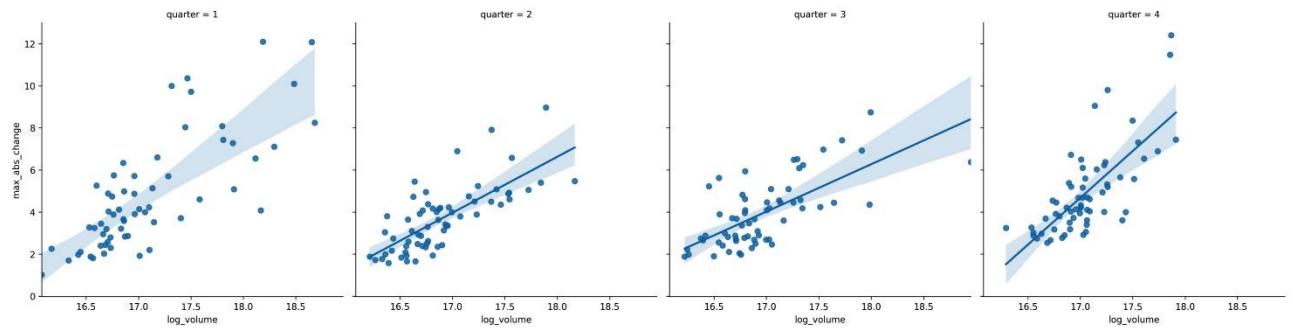


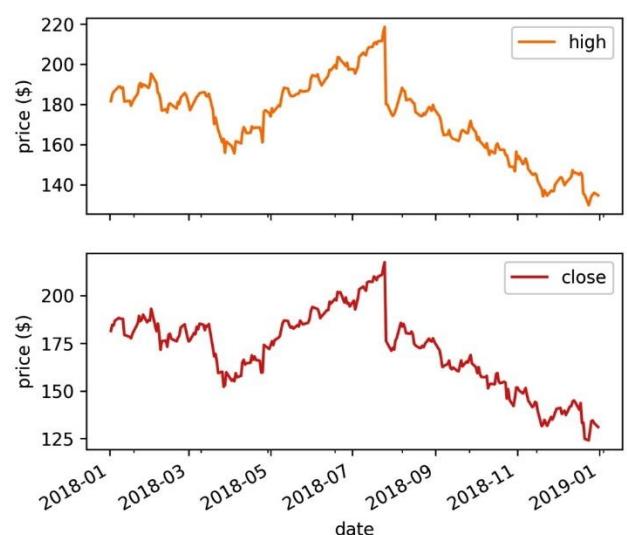
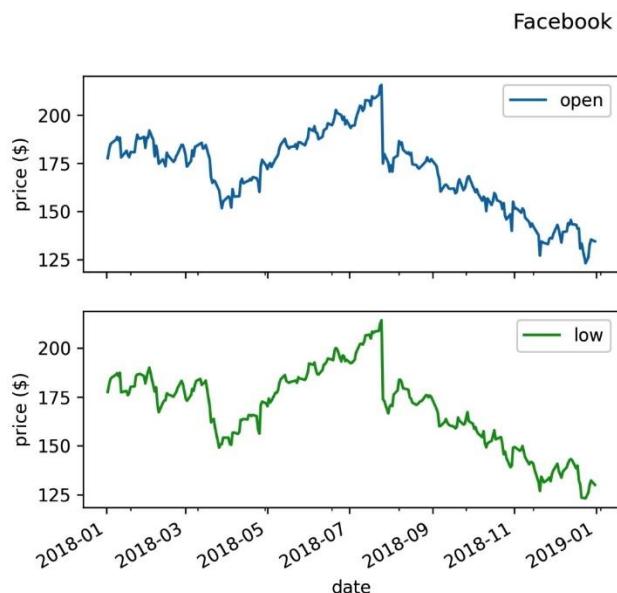
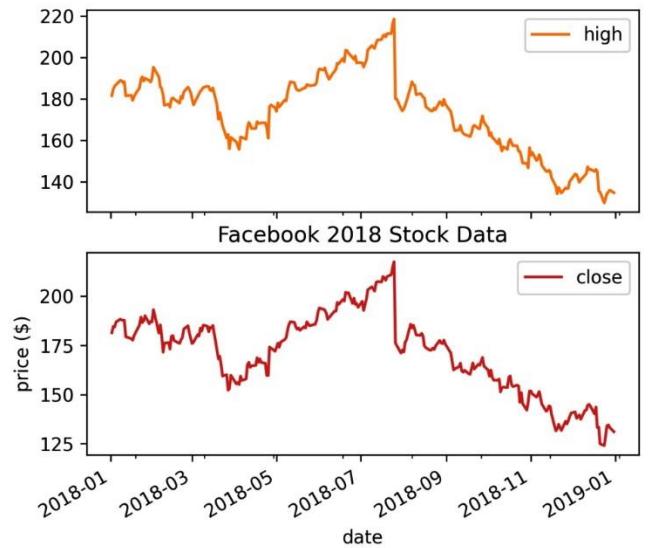
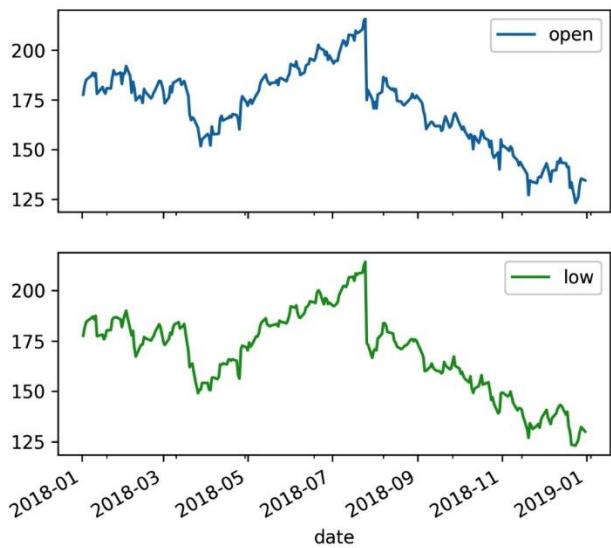
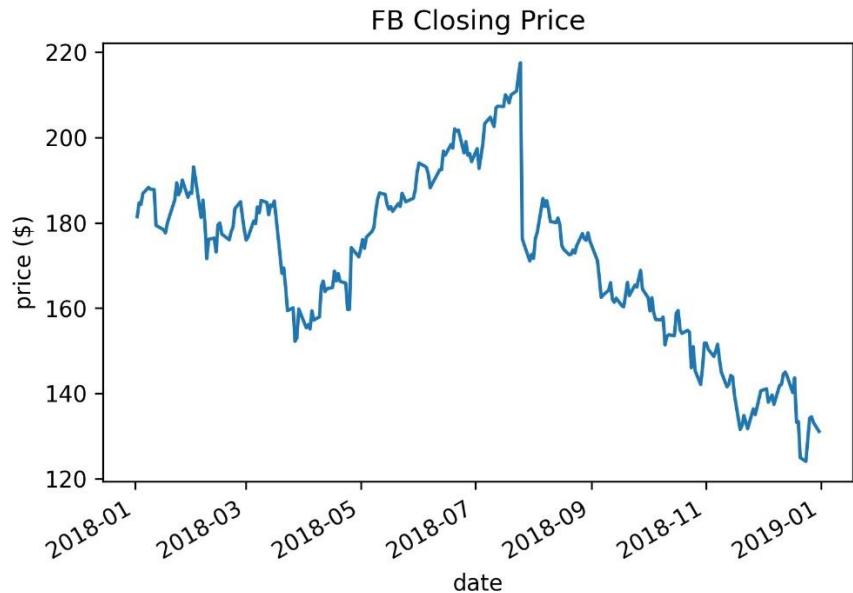




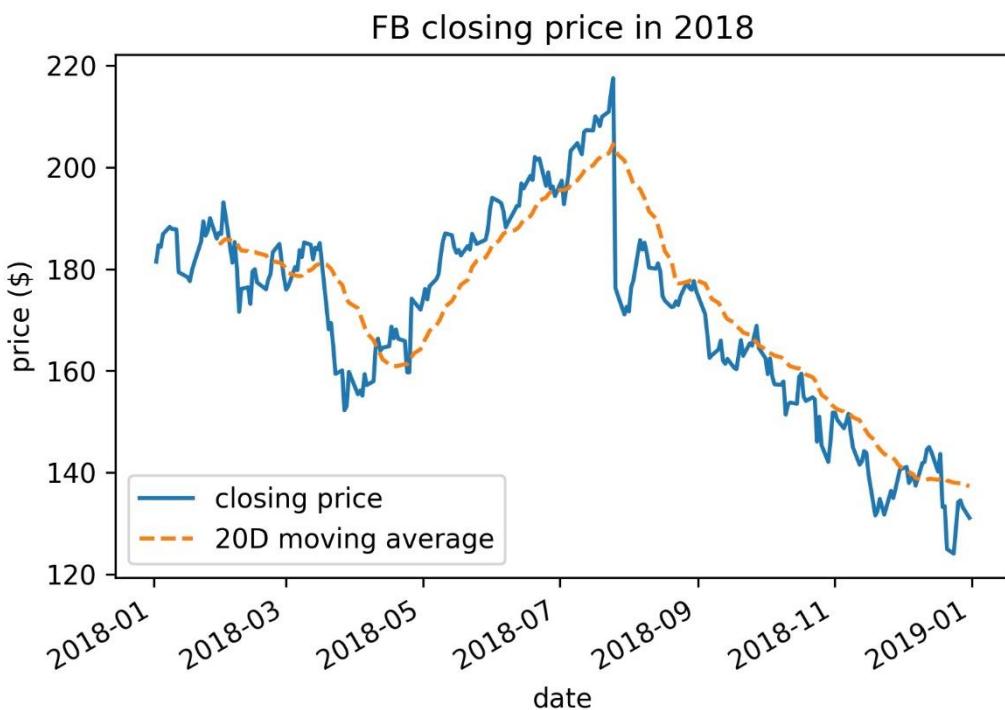




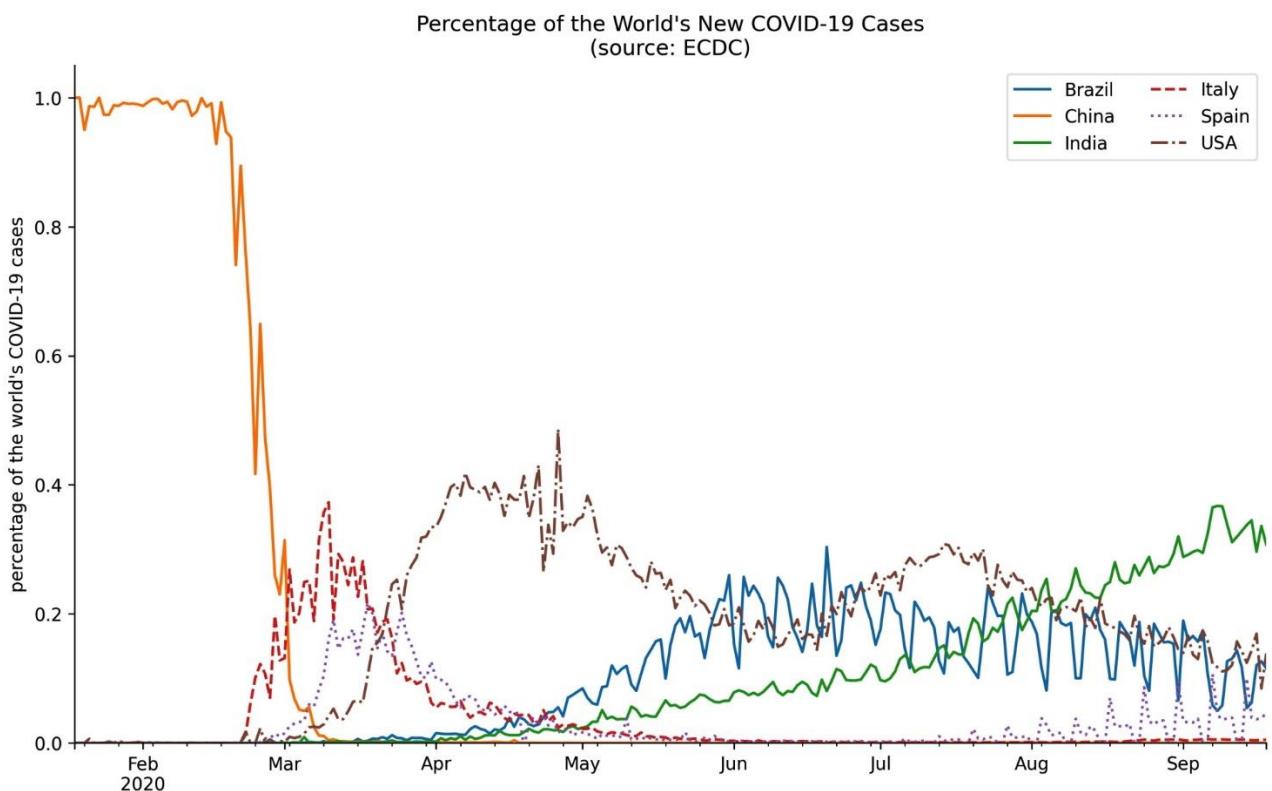
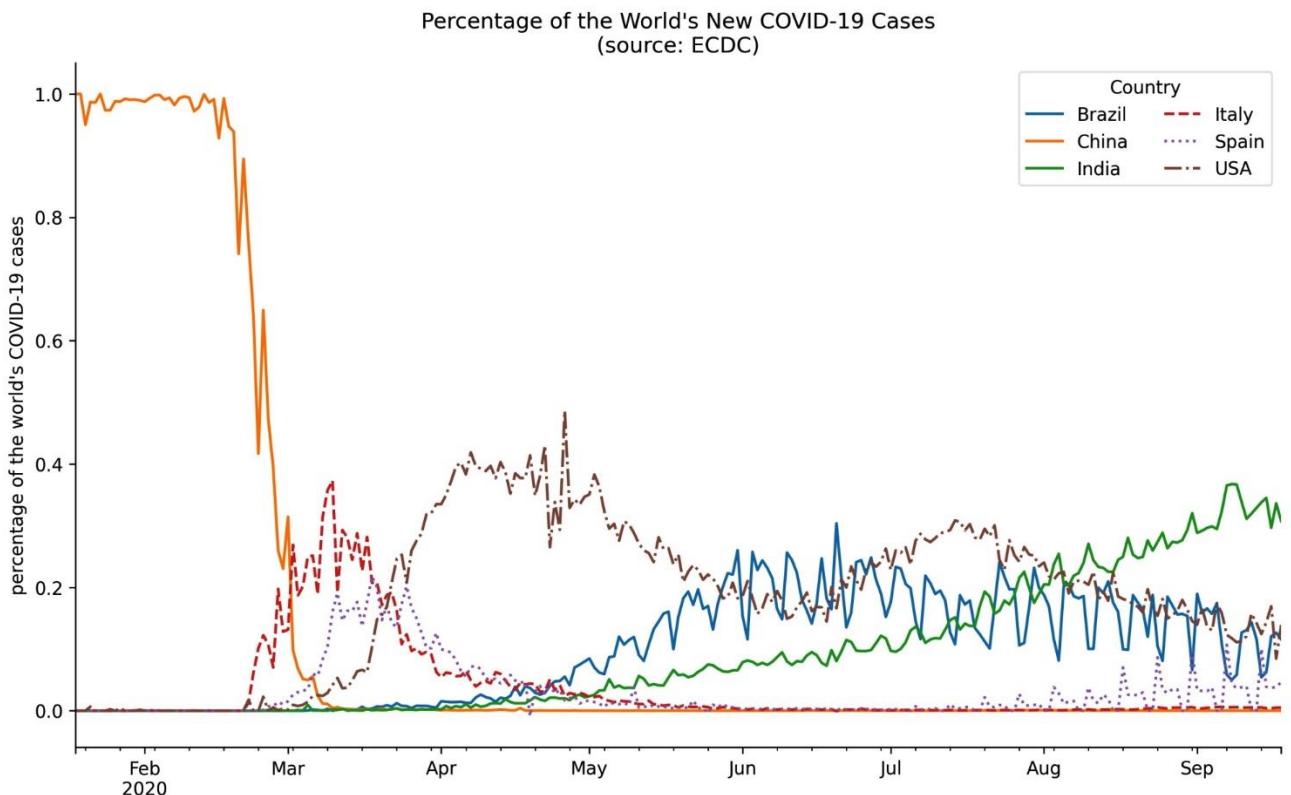


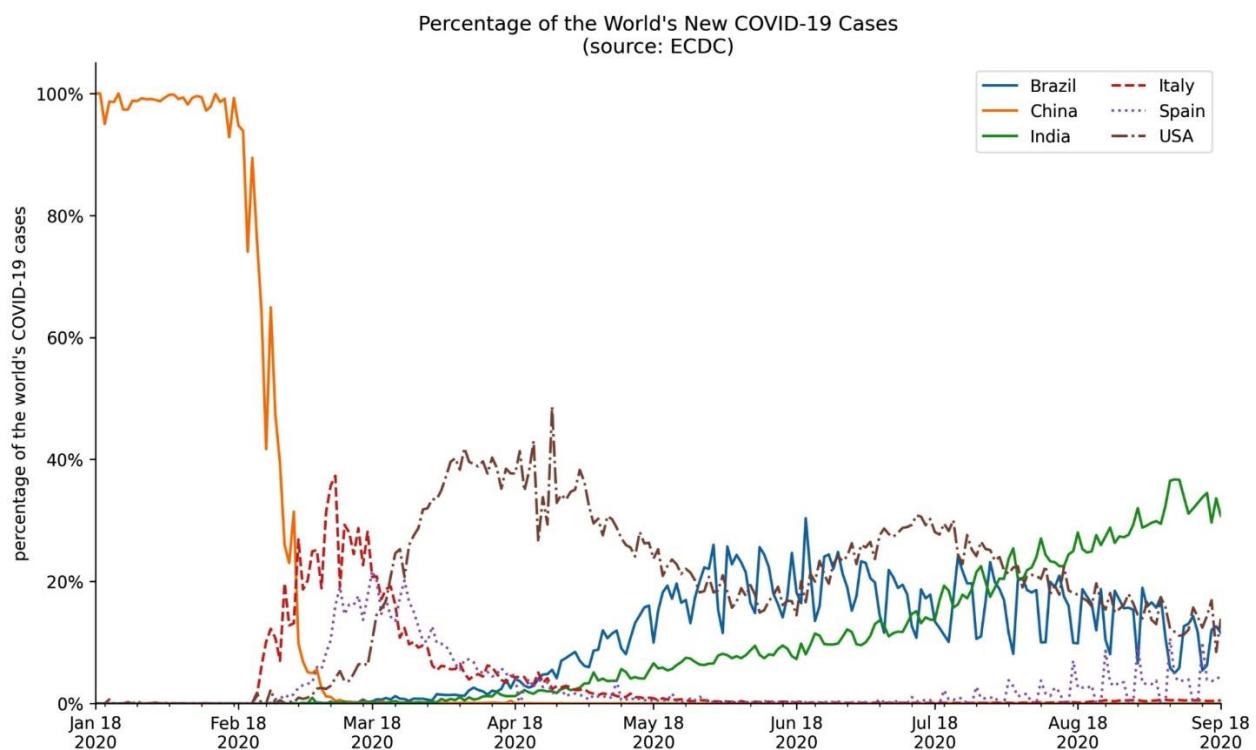
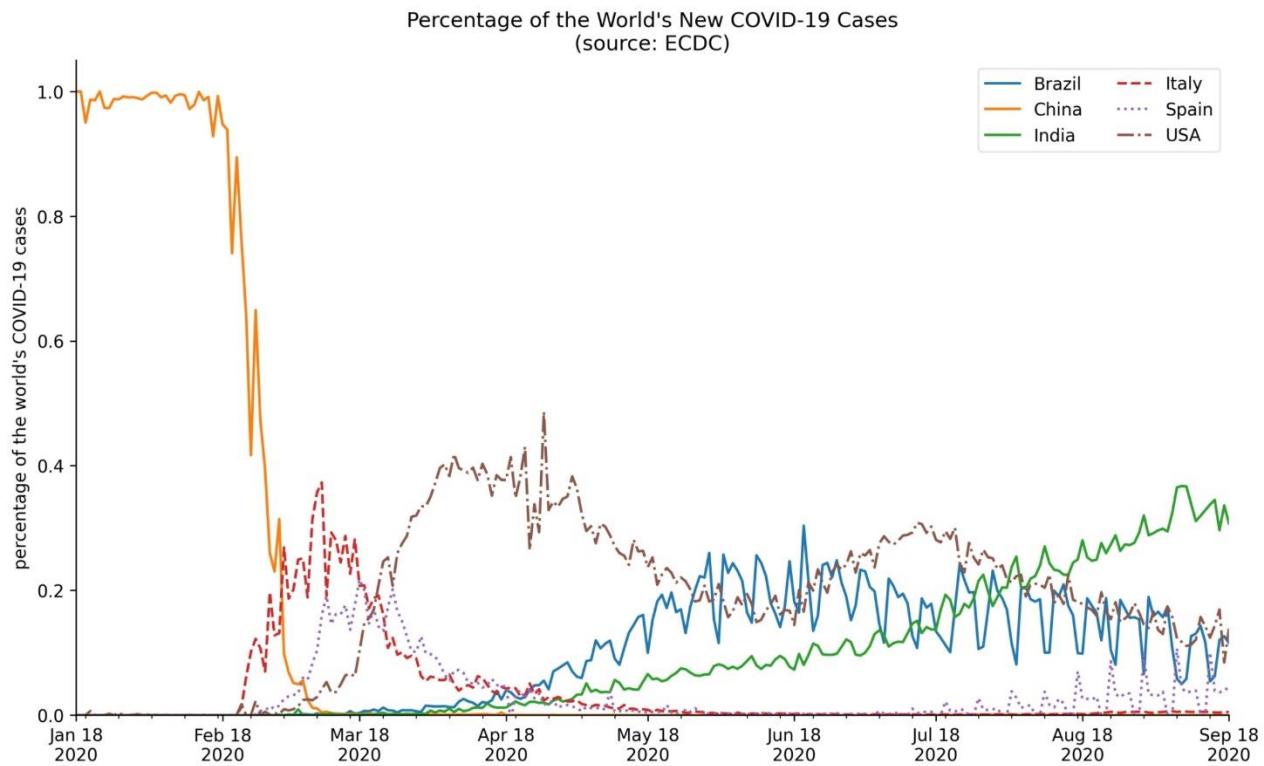


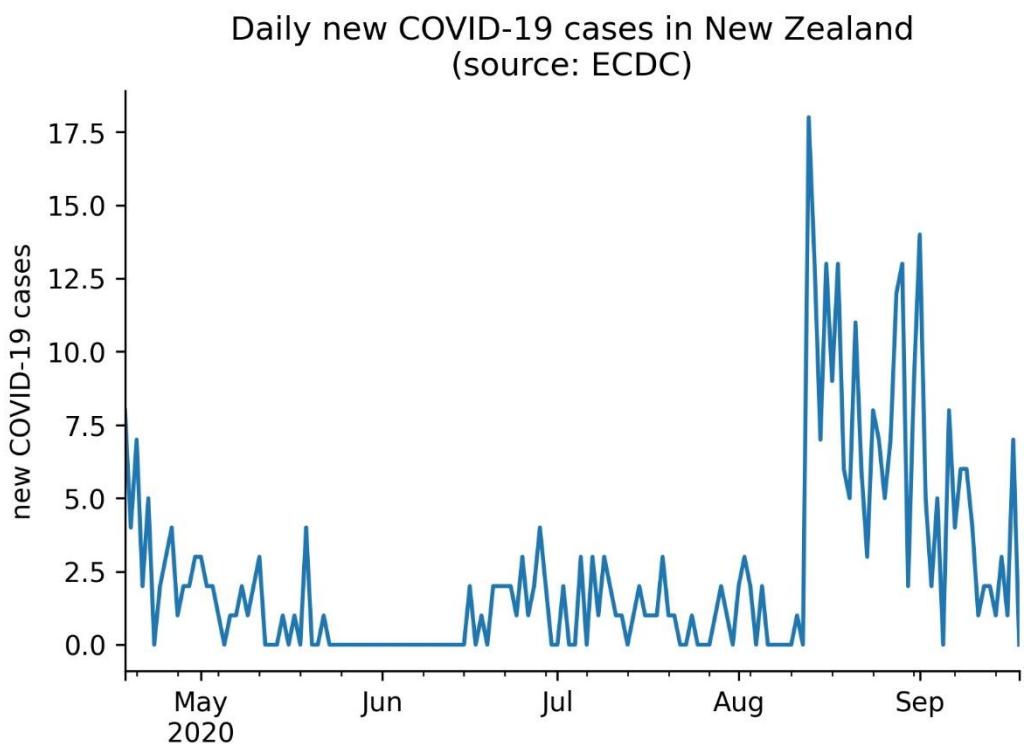
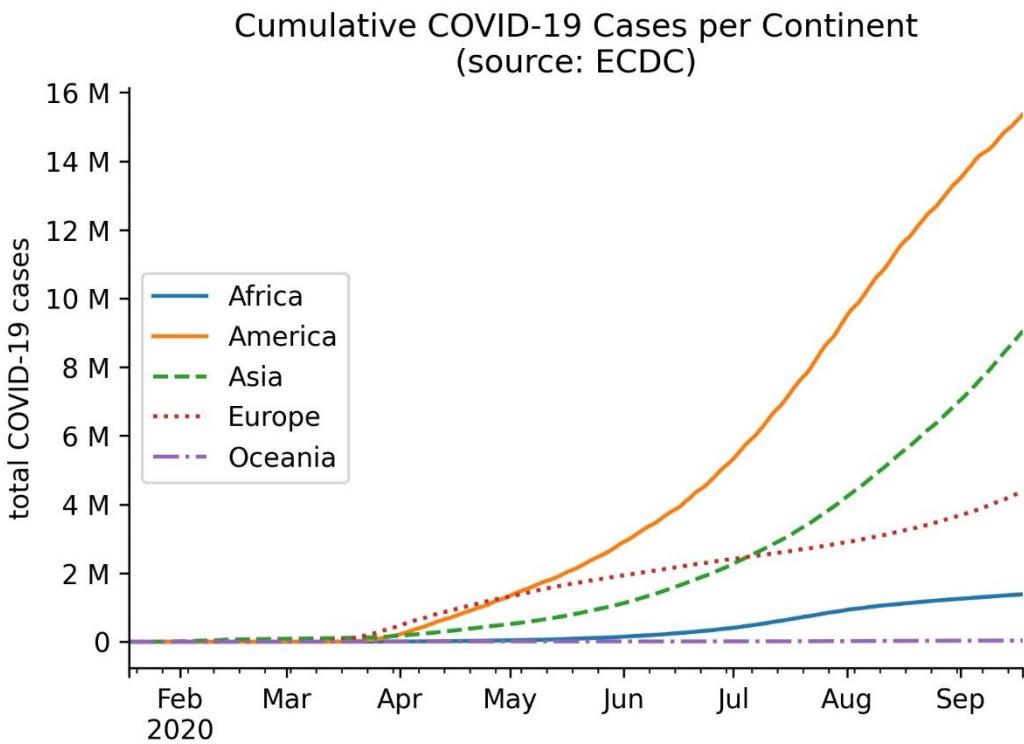
Parameter	Purpose
<code>loc</code>	Specify the location of the legend
<code>bbox_to_anchor</code>	Used in conjunction with <code>loc</code> to specify legend location
<code>ncol</code>	Set the number of columns the labels will be broken into, default is 1
<code>framealpha</code>	Control the transparency of the legend's background
<code>title</code>	Give the legend a title



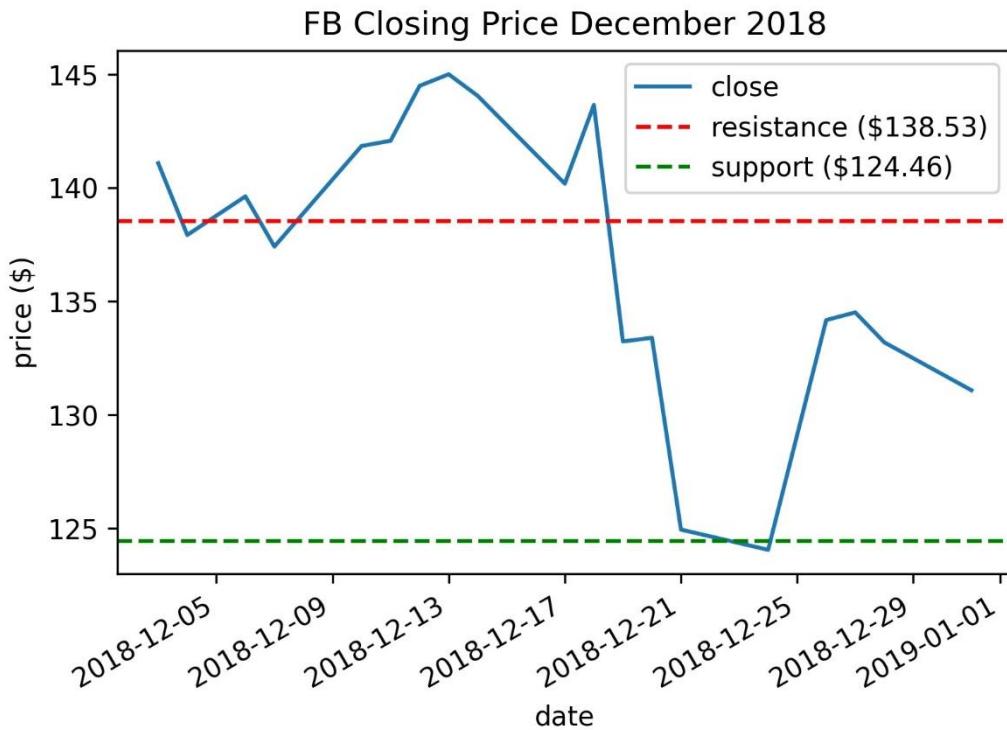
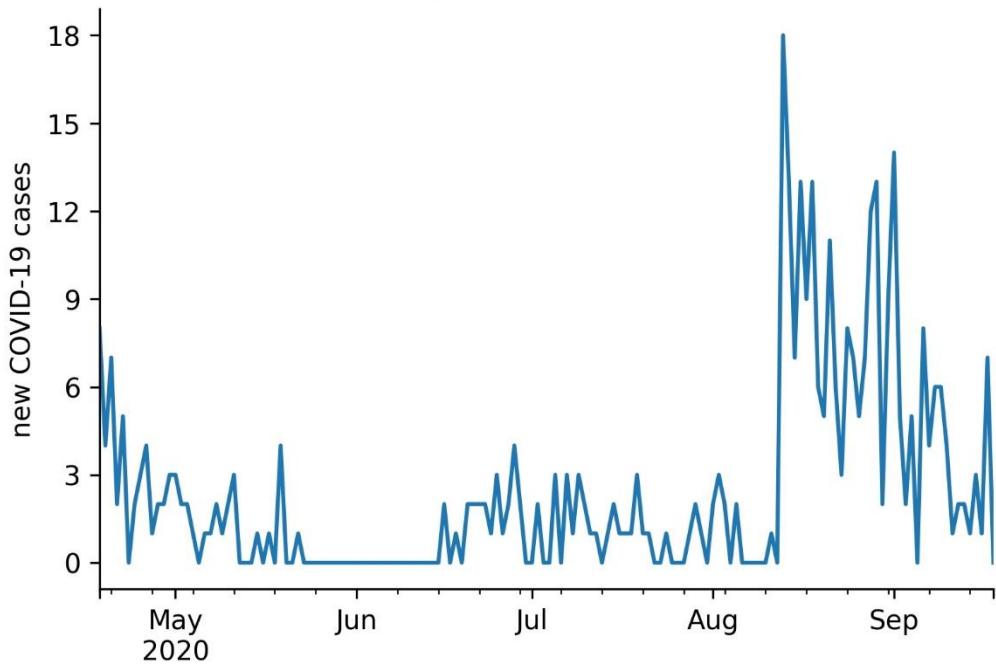
Location String	Location Code
'best'	0
'upper right'	1
'upper left'	2
'lower left'	3
'lower right'	4
'right'	5
'center left'	6
'center right'	7
'lower center'	8
'upper center'	9
'center'	10

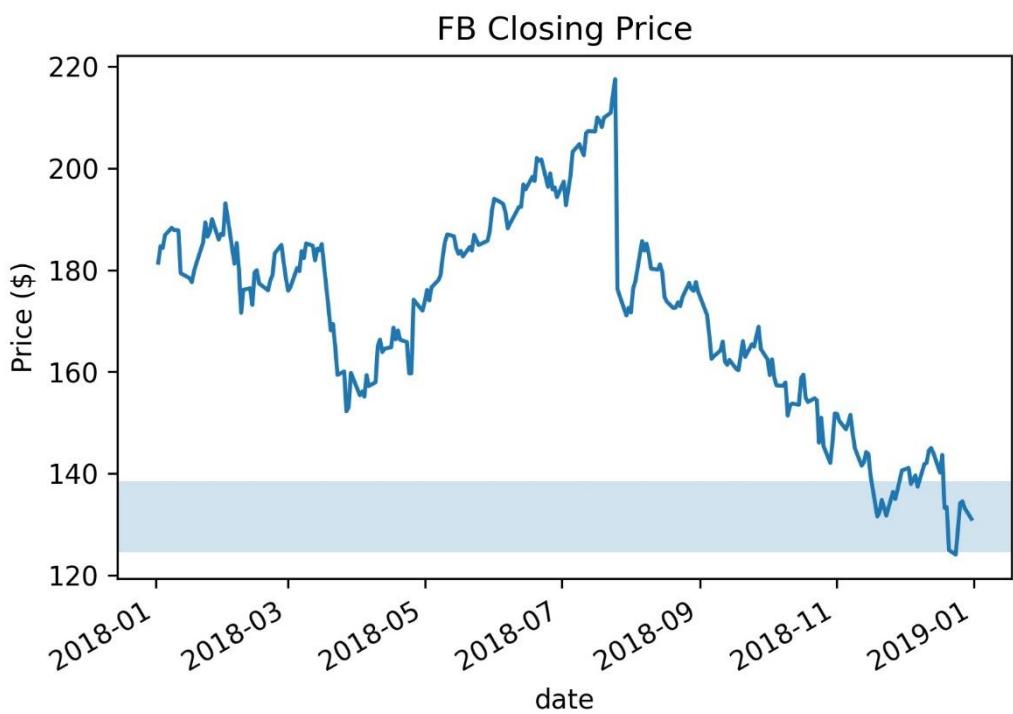
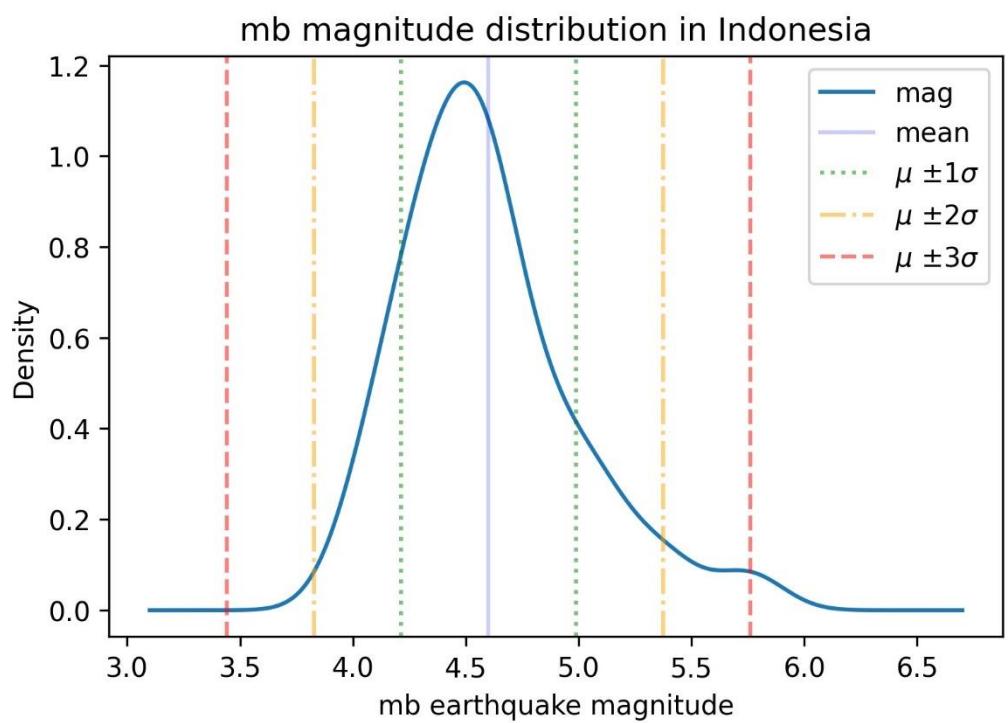




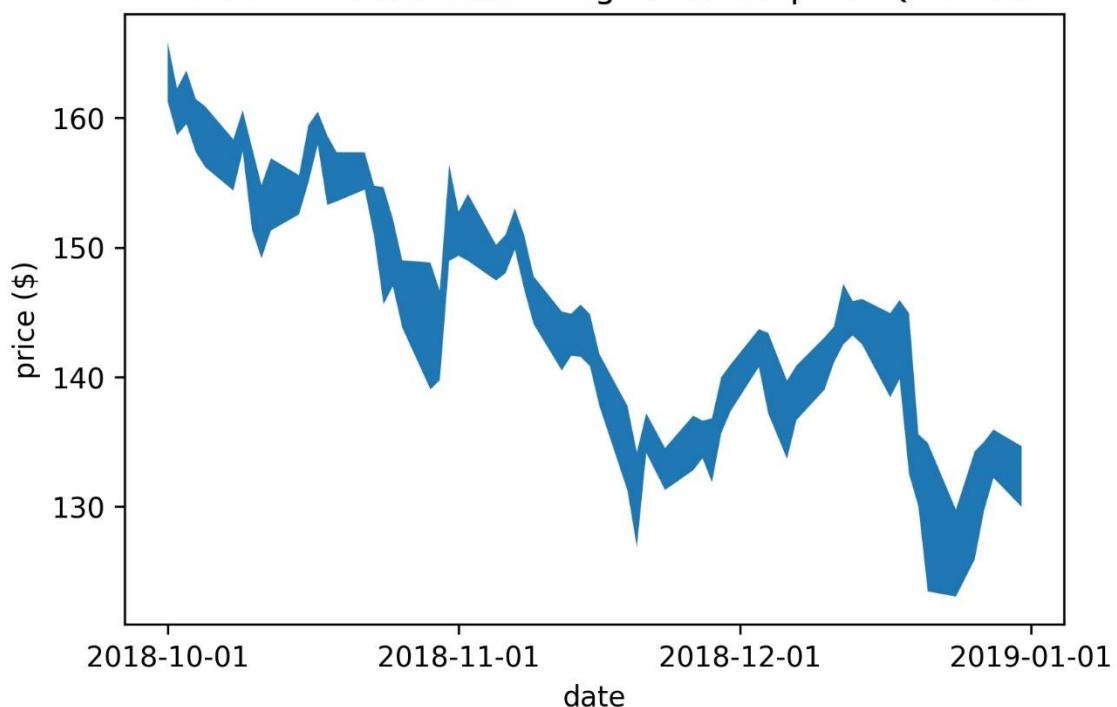


Daily new COVID-19 cases in New Zealand
(source: ECDC)

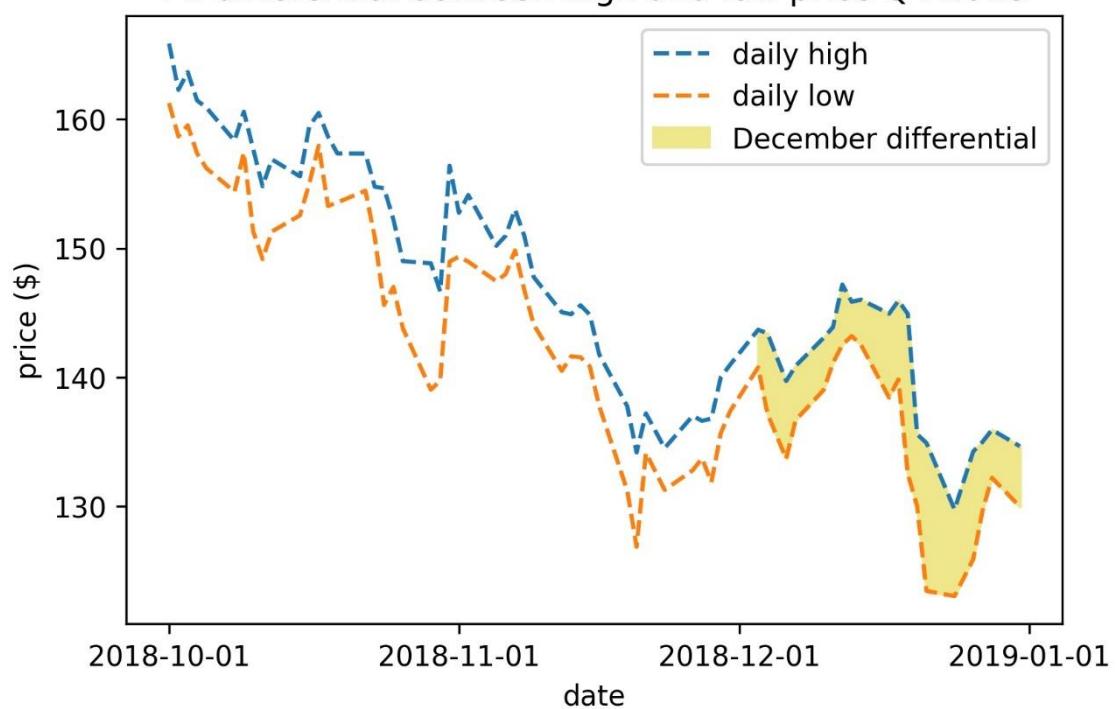


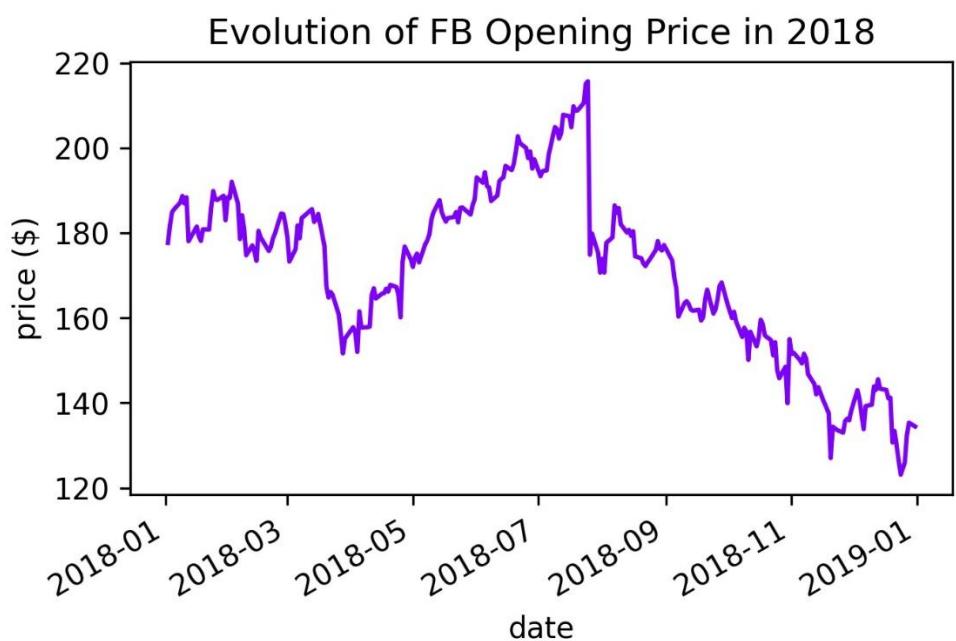
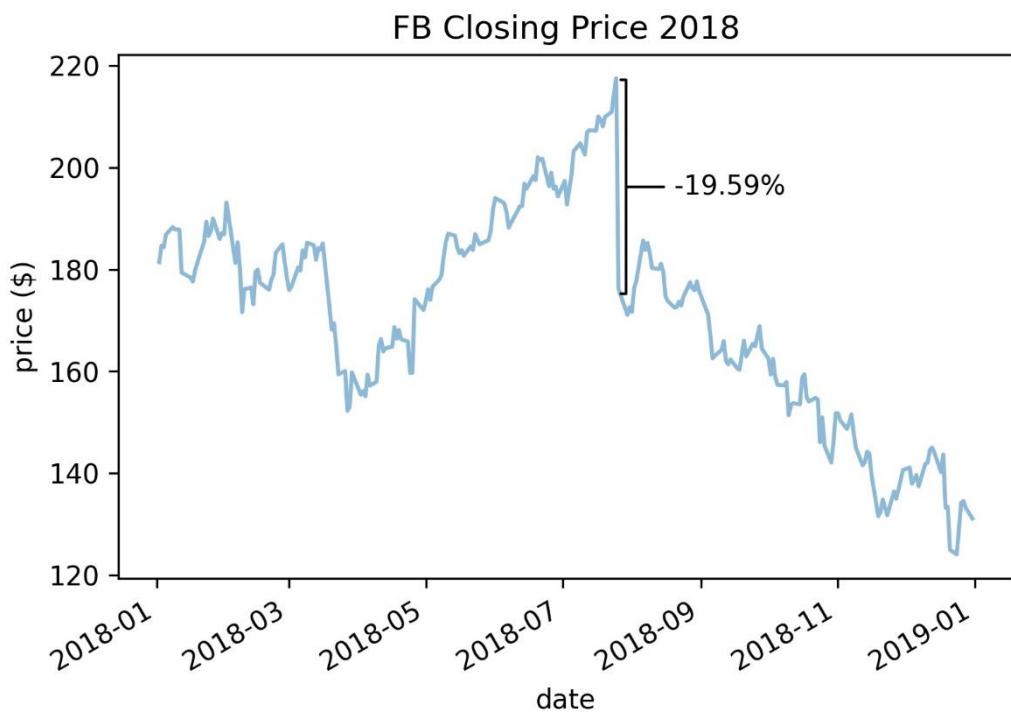
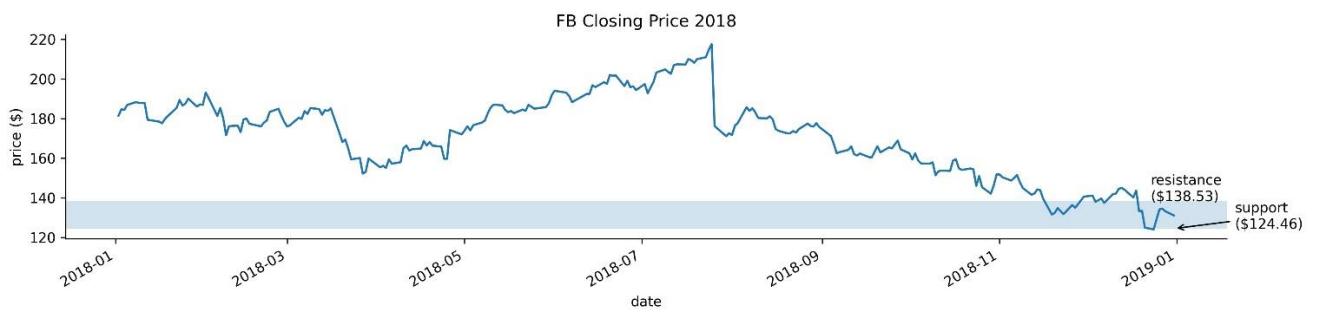


FB differential between high and low price Q4 2018

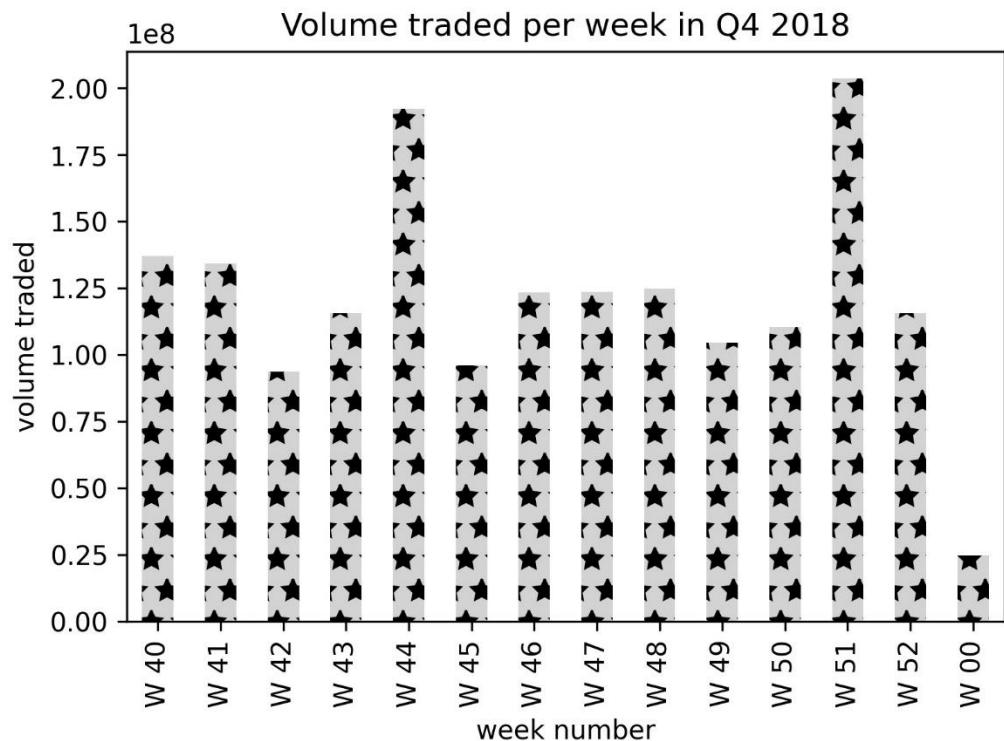
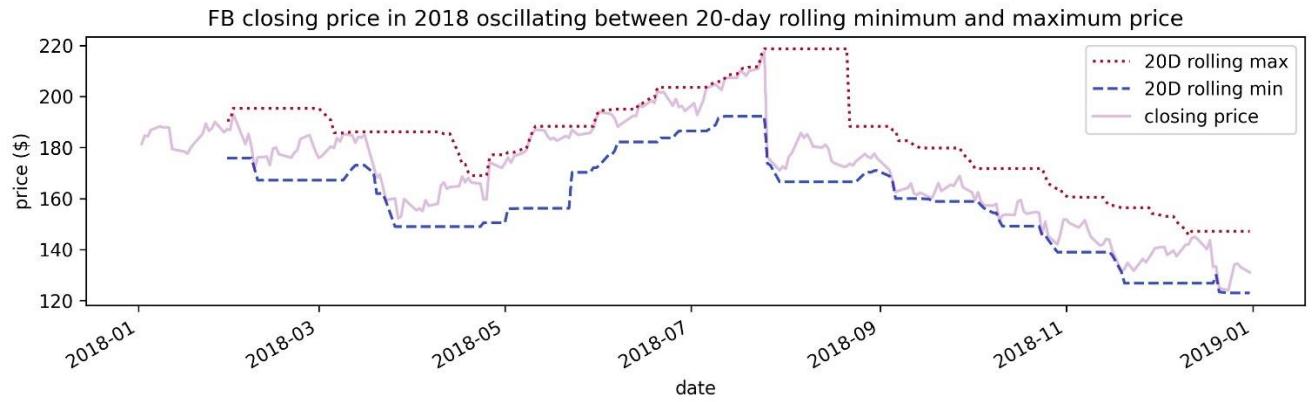


FB differential between high and low price Q4 2018

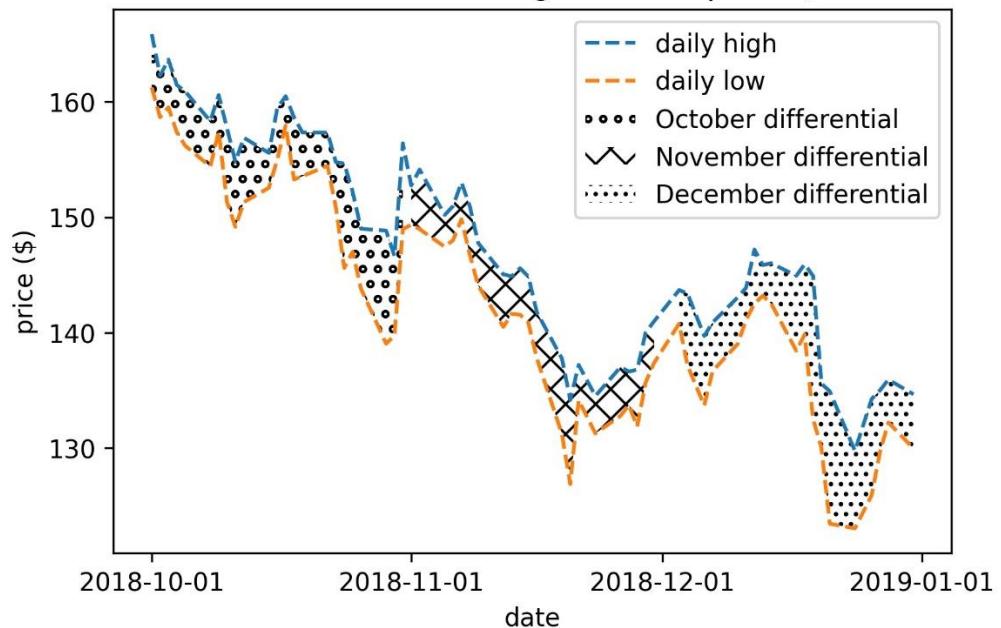




Class	Purpose
Qualitative	No ordering or relationship between colors; just used to distinguish between groups
Sequential	For information with ordering, such as temperature
Diverging	There is a middle value between two extremes that has meaning; for example, correlation coefficients are bounded in the range $[-1, 1]$, and 0 has meaning (no correlation)

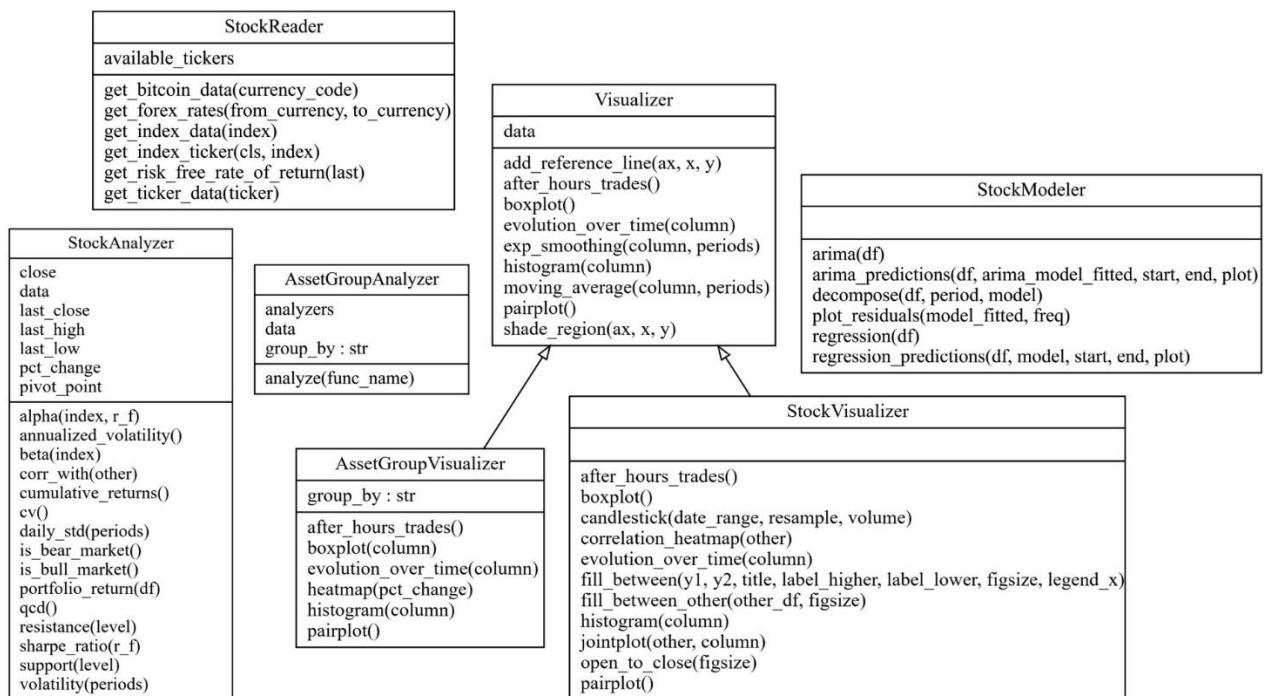


FB differential between high and low price Q4 2018

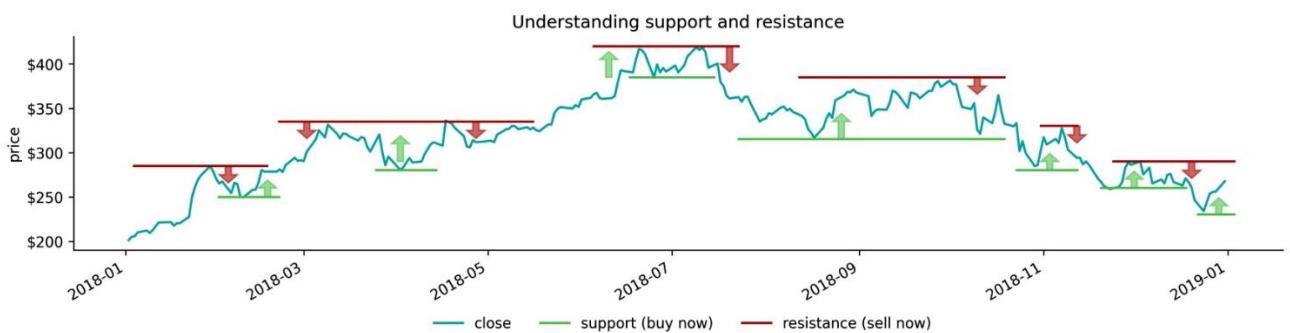
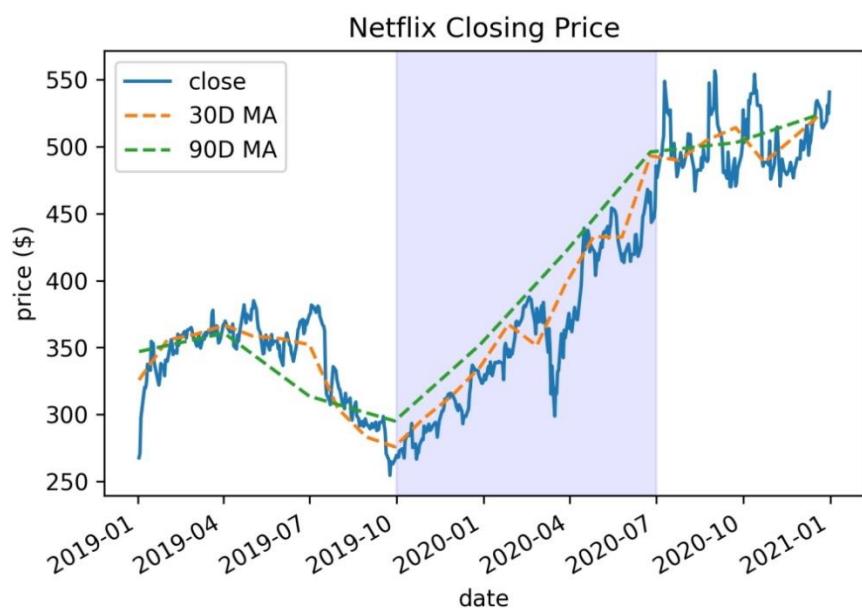
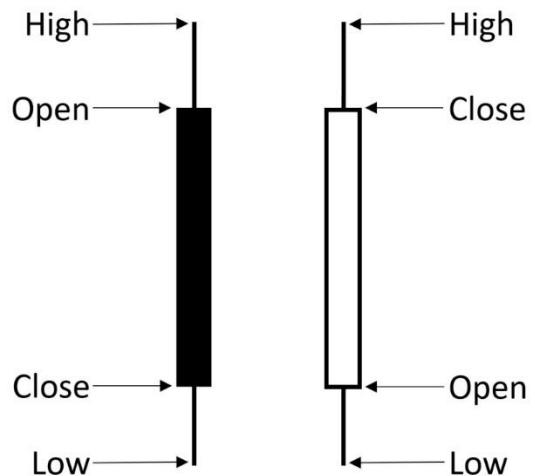


Chapter 7: Financial Analysis – Bitcoin and the Stock Market

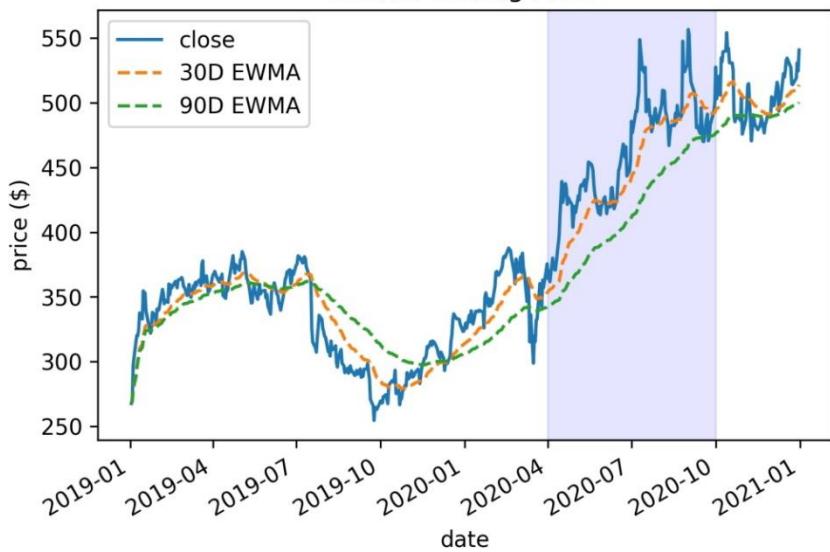
Purpose	Class(es)	Module
Collecting the data from various sources	StockReader	stock_reader.py
Visualizing the data	Visualizer , StockVisualizer , AssetGroupVisualizer	stock_visualizer.py
Calculating financial metrics	StockAnalyzer , AssetGroupAnalyzer	stock_analyzer.py
Modeling the data	StockModeler	stock_modeler.py



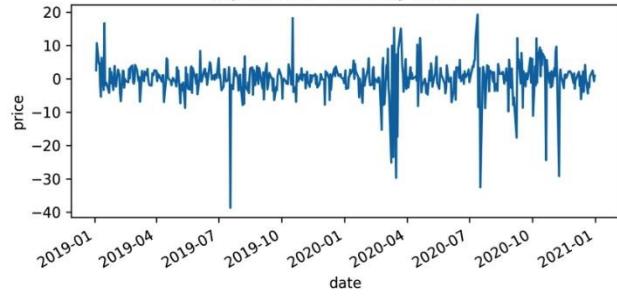
	Amazon	Apple	Bitcoin	Facebook	Google	Netflix	S&P 500
count	505.000000	505.000000	727.000000	505.000000	505.000000	505.000000	505.000000
mean	2235.904988	73.748386	9252.825408	208.146574	1335.188544	387.966593	3065.907599
std	594.306346	27.280933	4034.014685	39.665111	200.793911	78.931238	292.376435
min	1500.280029	35.547501	3399.471680	131.740005	1016.059998	254.589996	2237.399902
25%	1785.660034	50.782501	7218.593750	180.029999	1169.949951	329.089996	2870.719971
50%	1904.280029	66.730003	9137.993164	196.770004	1295.280029	364.369995	3005.469971
75%	2890.300049	91.632500	10570.513184	235.940002	1476.229980	469.959991	3276.020020
max	3531.449951	136.690002	29001.720703	303.910004	1827.989990	556.549988	3756.070068



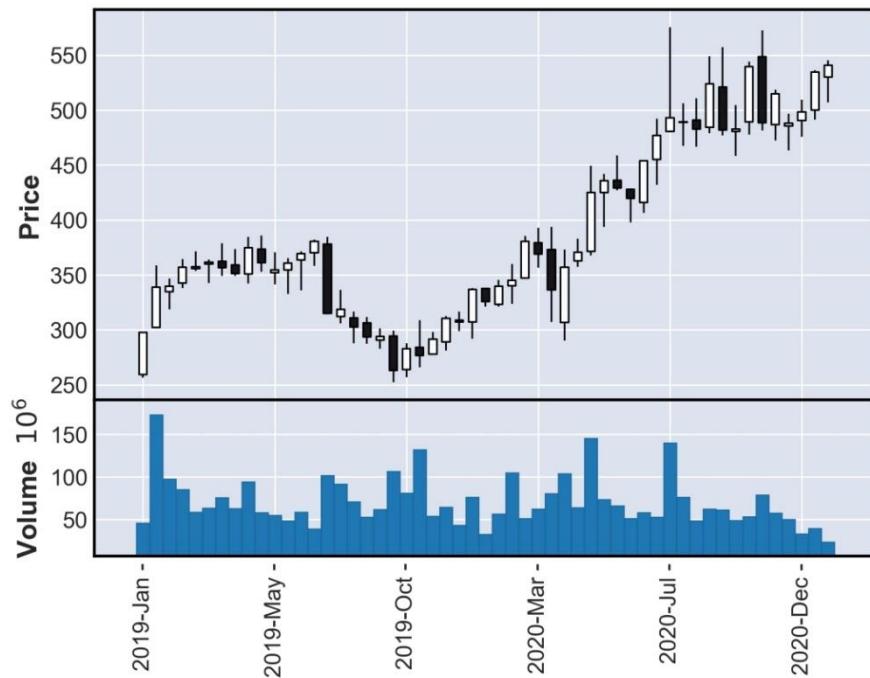
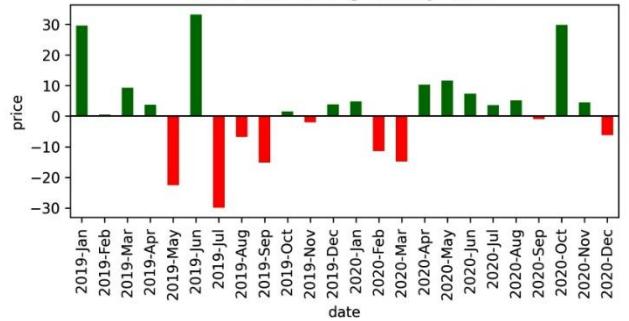
Netflix Closing Price

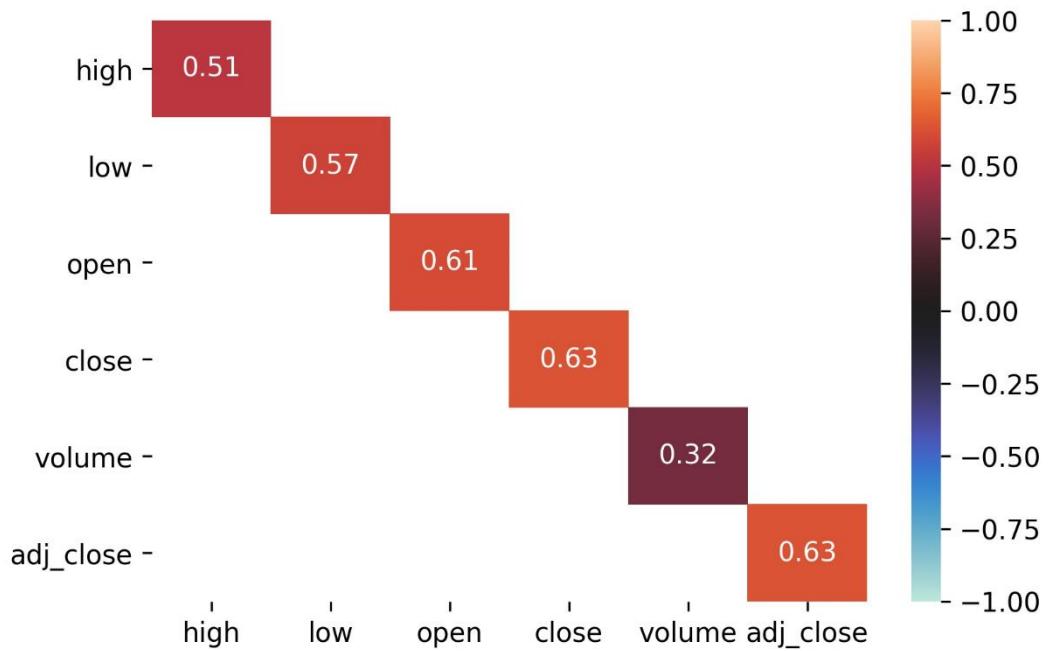
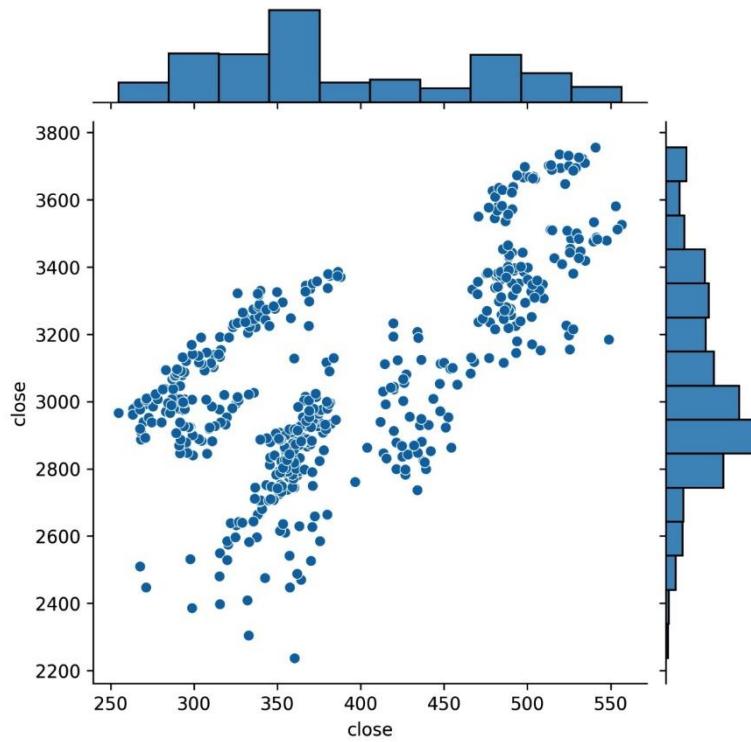


After-hours trading
(Open Price - Prior Day's Close)

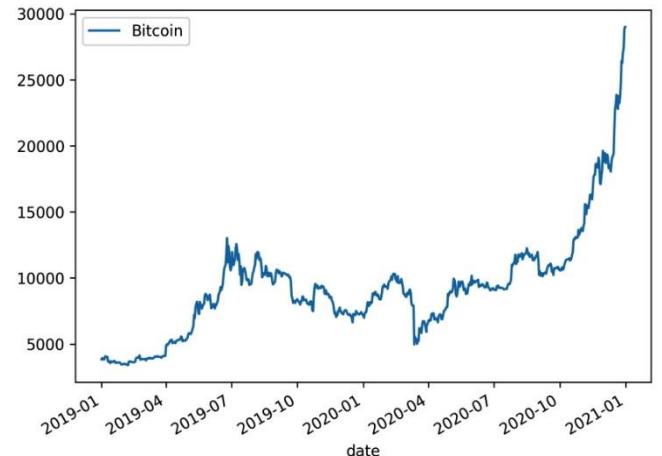
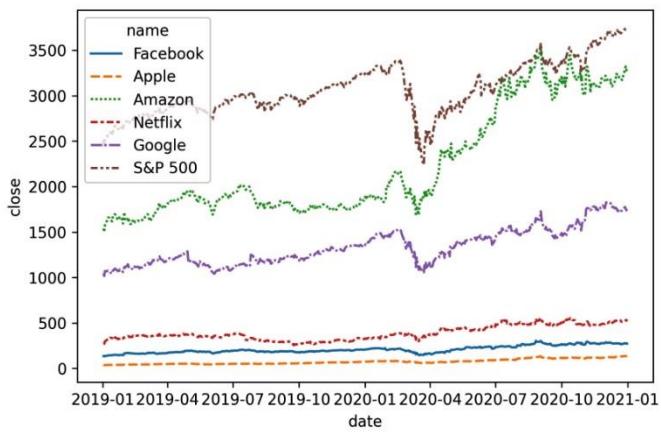
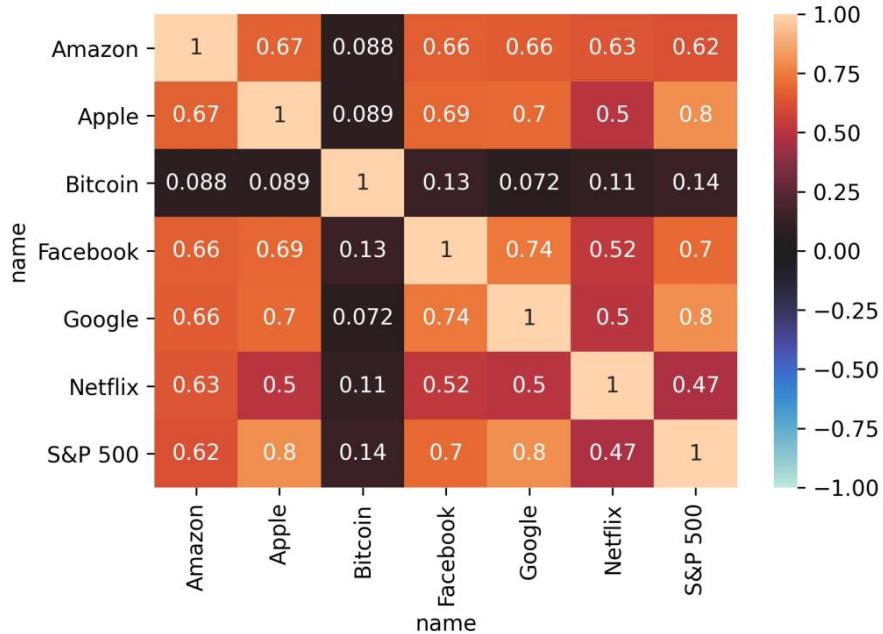
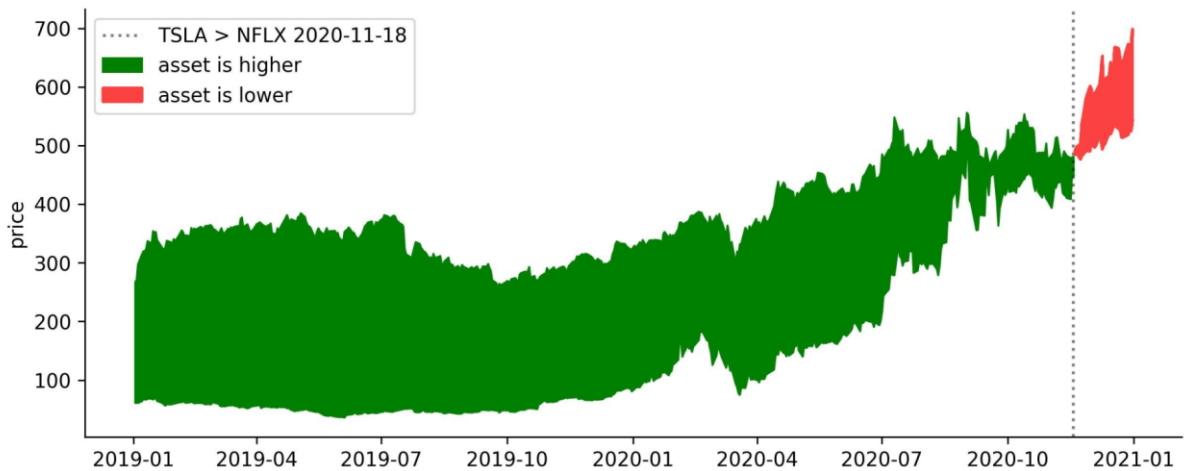


After-hours trading monthly effect

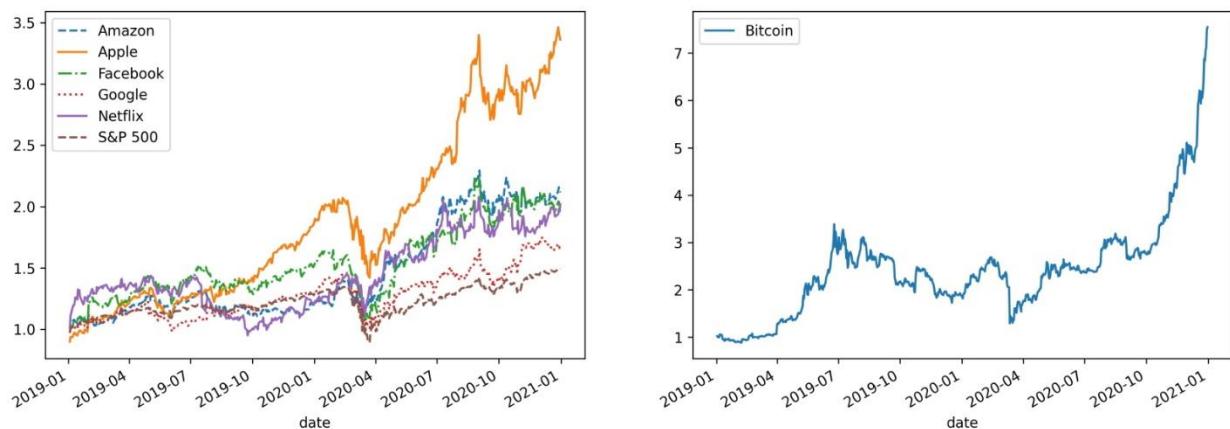




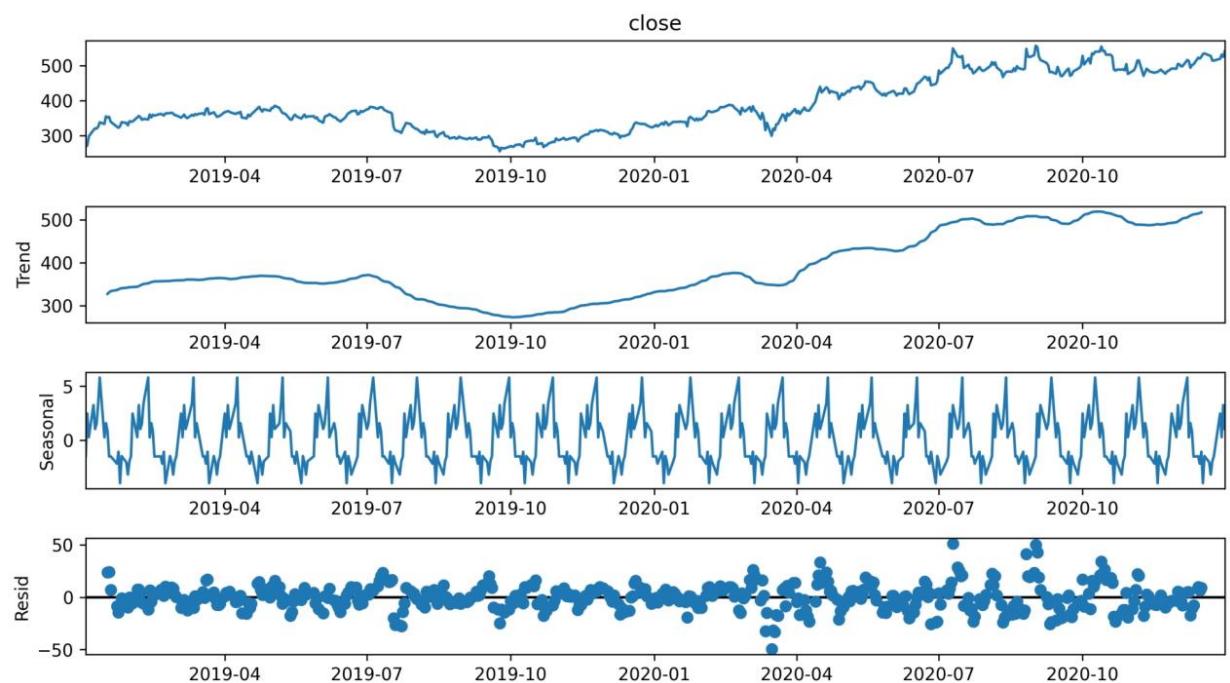
Differential between asset closing price (this - other)



Cumulative Returns



Netflix Stock Price Time Series Decomposition



SARIMAX Results

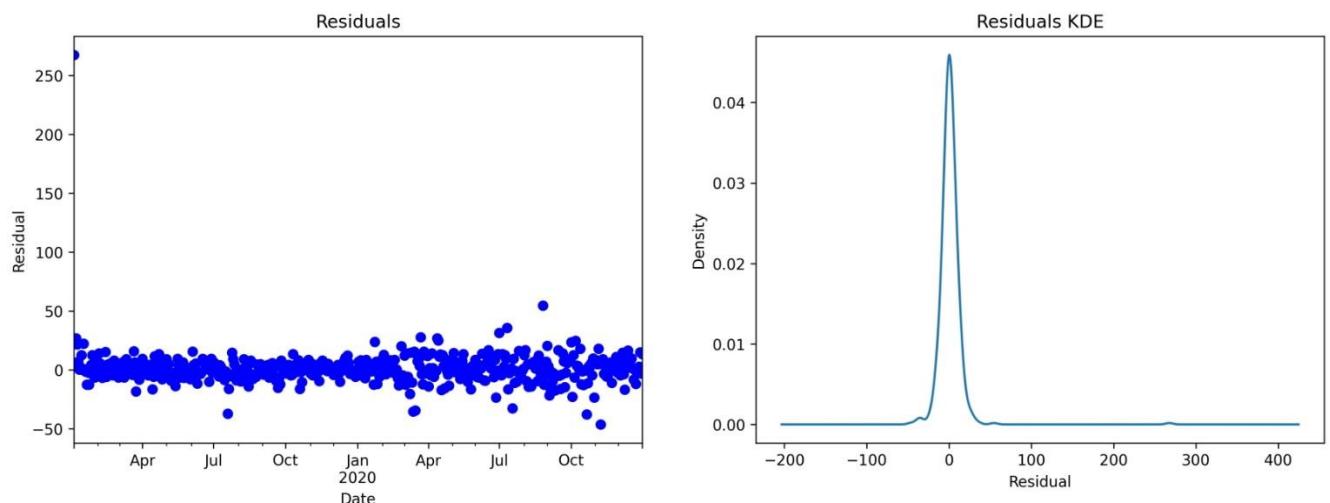
```
=====
Dep. Variable:          close    No. Observations:      522
Model:                 ARIMA(10, 1, 5)   Log Likelihood:   -1925.850
Date: Mon, 18 Jan 2021   AIC:                  3883.700
Time: 19:02:23           BIC:                  3951.792
Sample: 01-02-2019 - 12-31-2020   HQIC:                  3910.372
Covariance Type: opg
```

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.1407	0.254	-0.554	0.580	-0.639	0.358
ar.L2	0.1384	0.178	0.777	0.437	-0.211	0.488
ar.L3	-0.3349	0.165	-2.033	0.042	-0.658	-0.012
ar.L4	0.6575	0.171	3.839	0.000	0.322	0.993
ar.L5	0.5988	0.215	2.787	0.005	0.178	1.020
ar.L6	-0.1005	0.076	-1.315	0.188	-0.250	0.049
ar.L7	0.0555	0.052	1.072	0.284	-0.046	0.157
ar.L8	-0.0522	0.042	-1.256	0.209	-0.134	0.029
ar.L9	-0.0722	0.051	-1.425	0.154	-0.172	0.027
ar.L10	0.1021	0.056	1.813	0.070	-0.008	0.212
ma.L1	-0.0084	0.257	-0.032	0.974	-0.513	0.496
ma.L2	-0.0854	0.196	-0.435	0.663	-0.470	0.299
ma.L3	0.3300	0.184	1.797	0.072	-0.030	0.690
ma.L4	-0.6166	0.174	-3.549	0.000	-0.957	-0.276
ma.L5	-0.5170	0.213	-2.425	0.015	-0.935	-0.099
sigma2	93.0293	3.711	25.071	0.000	85.756	100.302

```
=====
Ljung-Box (Q):            33.12 Jarque-Bera (JB):        373.34
Prob(Q):                  0.77 Prob(JB):                0.00
Heteroskedasticity (H):  2.46 Skew:                   -0.10
Prob(H) (two-sided):     0.00 Kurtosis:               7.14
=====
```

Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

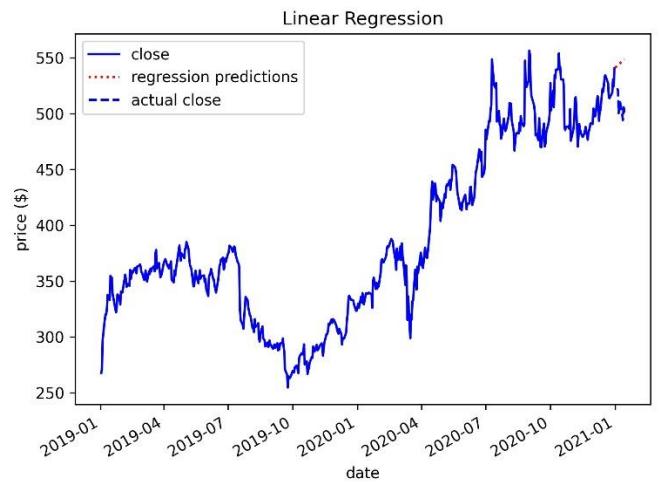
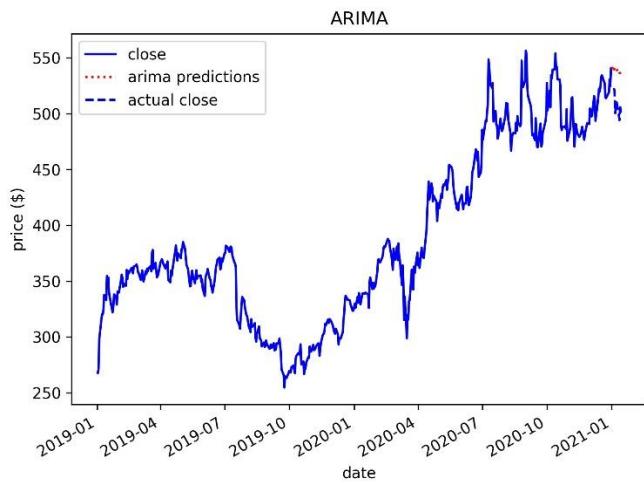
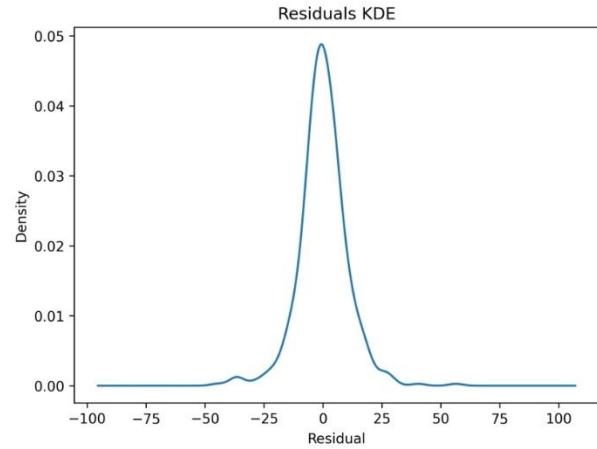
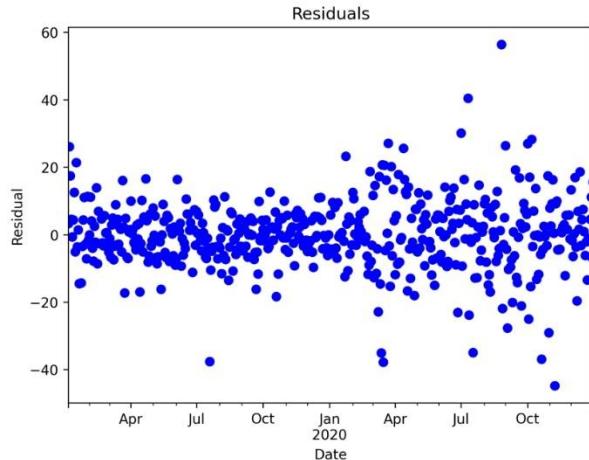


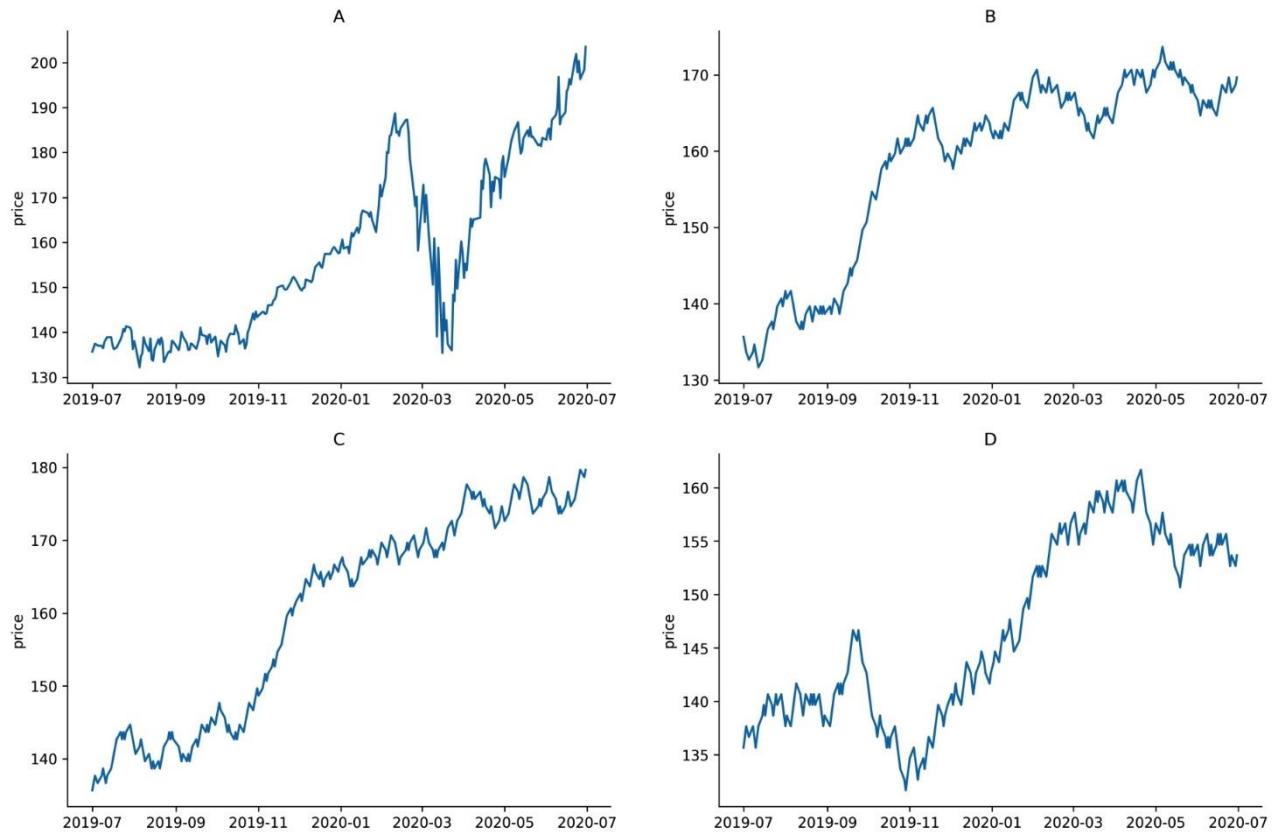
OLS Regression Results

Dep. Variable:	close	R-squared (uncentered):	0.999
Model:	OLS	Adj. R-squared (uncentered):	0.999
Method:	Least Squares	F-statistic:	7.470e+05
Date:	Mon, 18 Jan 2021	Prob (F-statistic):	0.00
Time:	19:15:40	Log-Likelihood:	-1889.3
No. Observations:	504	AIC:	3781.
Df Residuals:	503	BIC:	3785.
Df Model:	1		
Covariance Type:	nonrobust		
coef	std err	t	P> t
close	1.0011	0.001	864.291
			0.000
			[0.025 0.975]
			0.999 1.003
Omnibus:	50.714	Durbin-Watson:	2.317
Prob(Omnibus):	0.000	Jarque-Bera (JB):	307.035
Skew:	-0.014	Prob(JB):	2.13e-67
Kurtosis:	6.824	Cond. No.	1.00

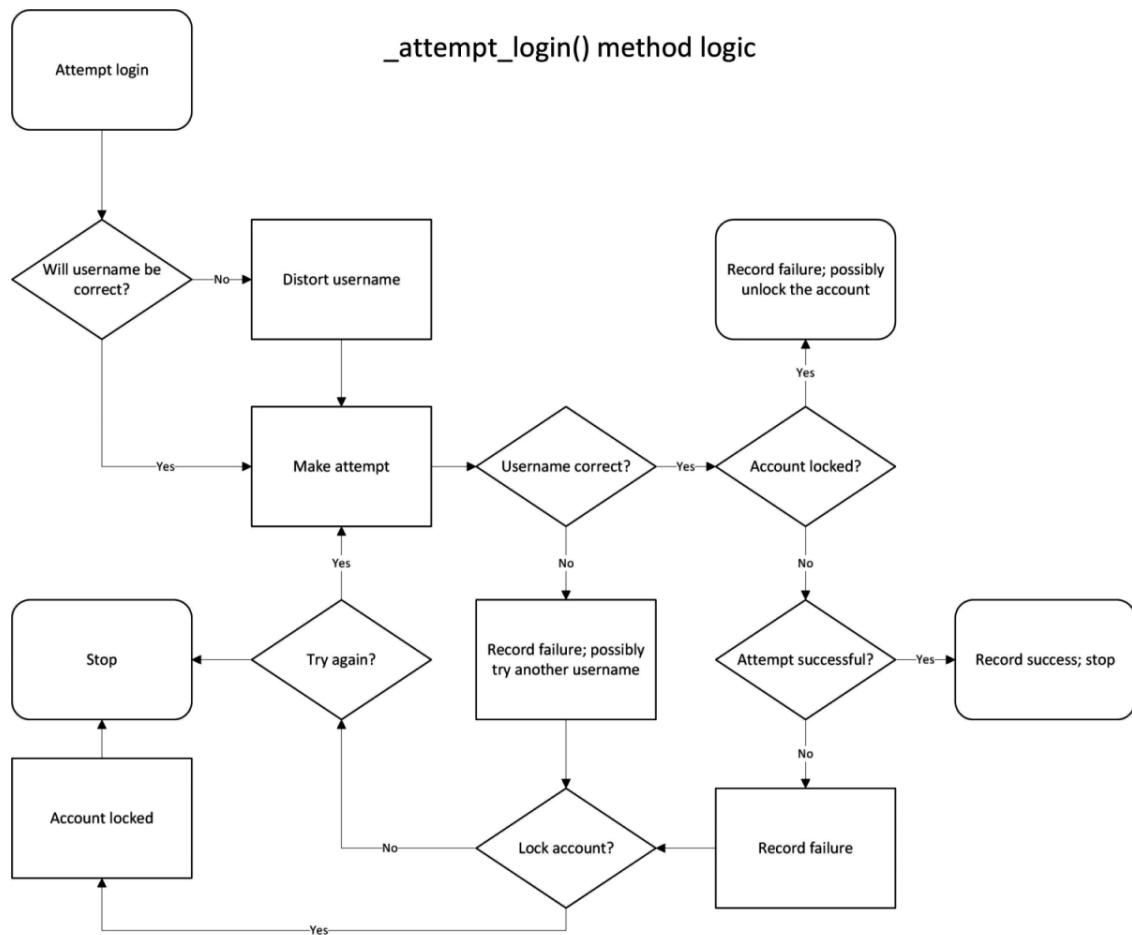
Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

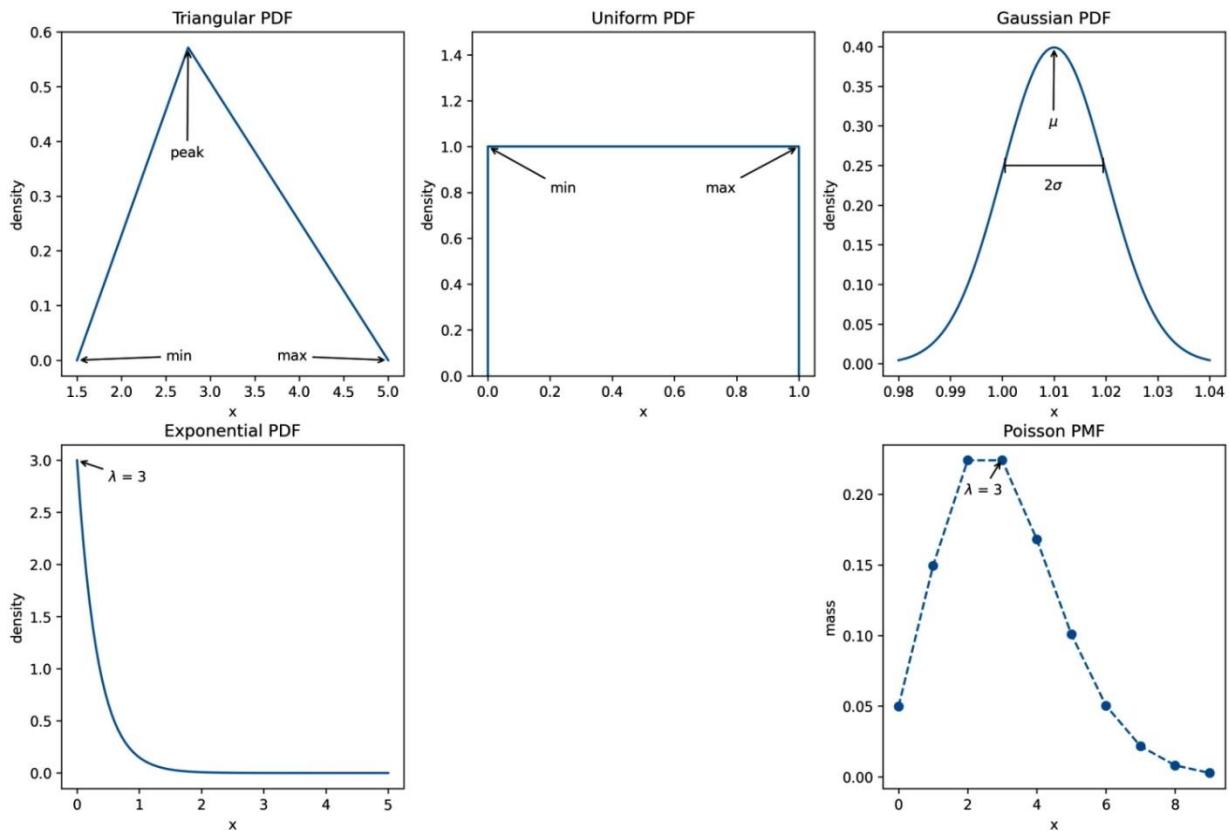




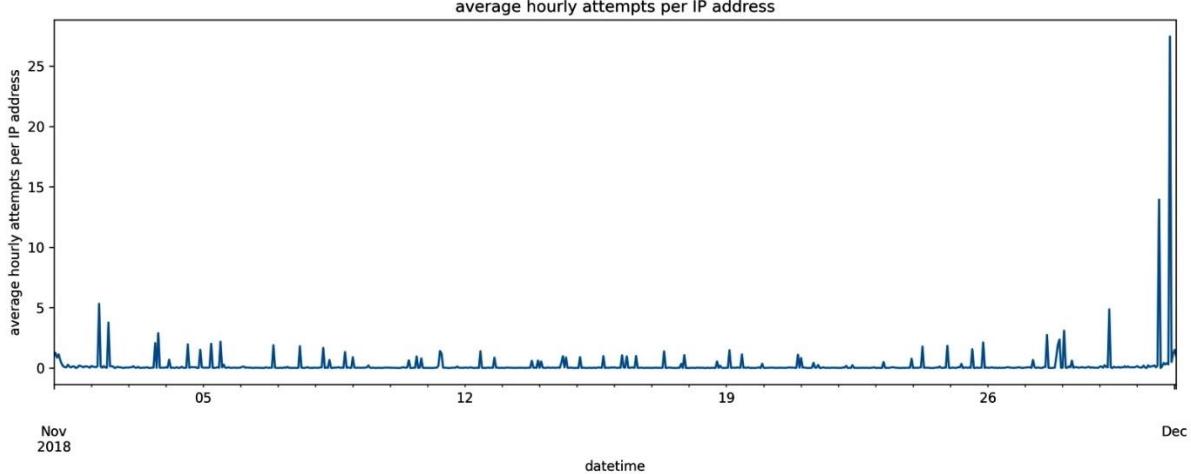
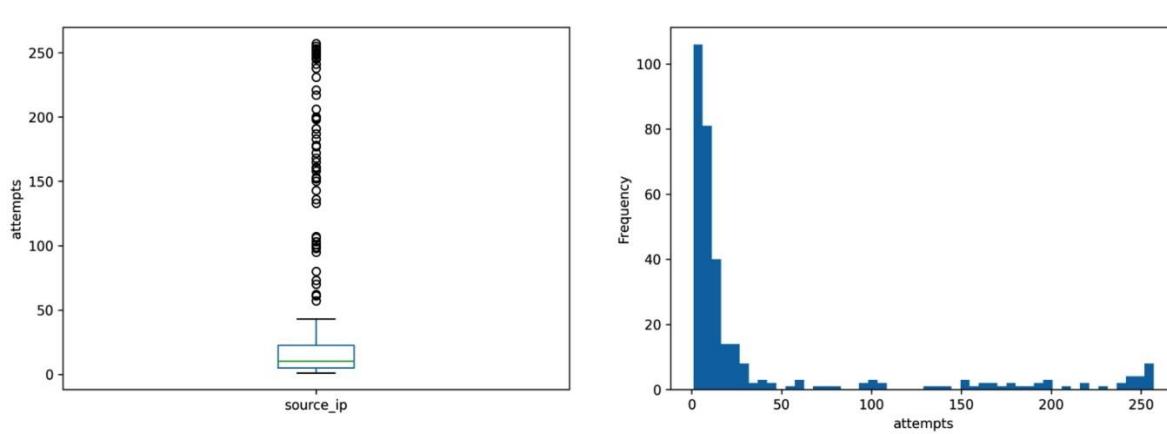
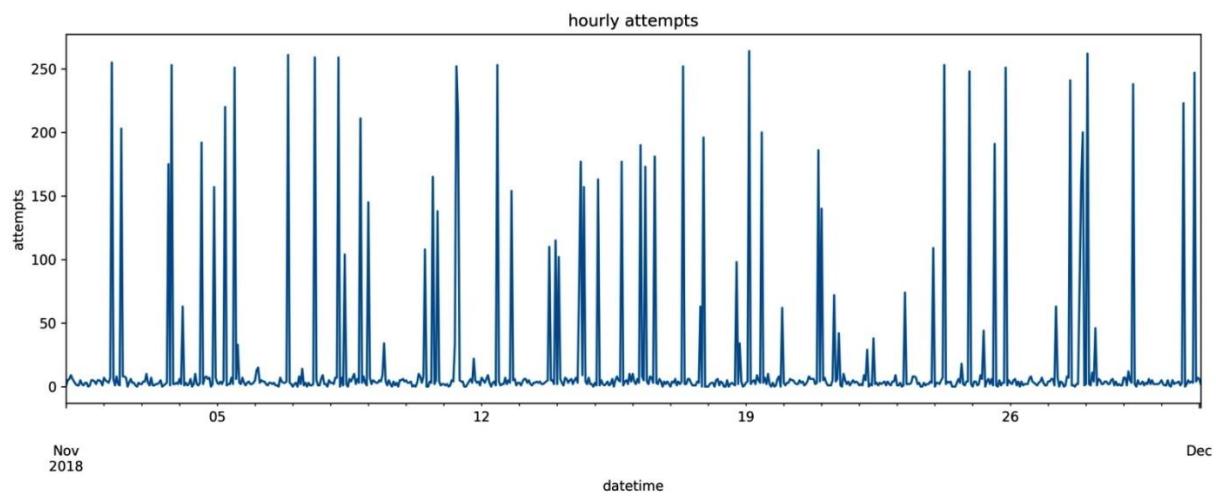
Chapter 8: Rule-Based Anomaly Detection



Understanding the distributions used for the simulation



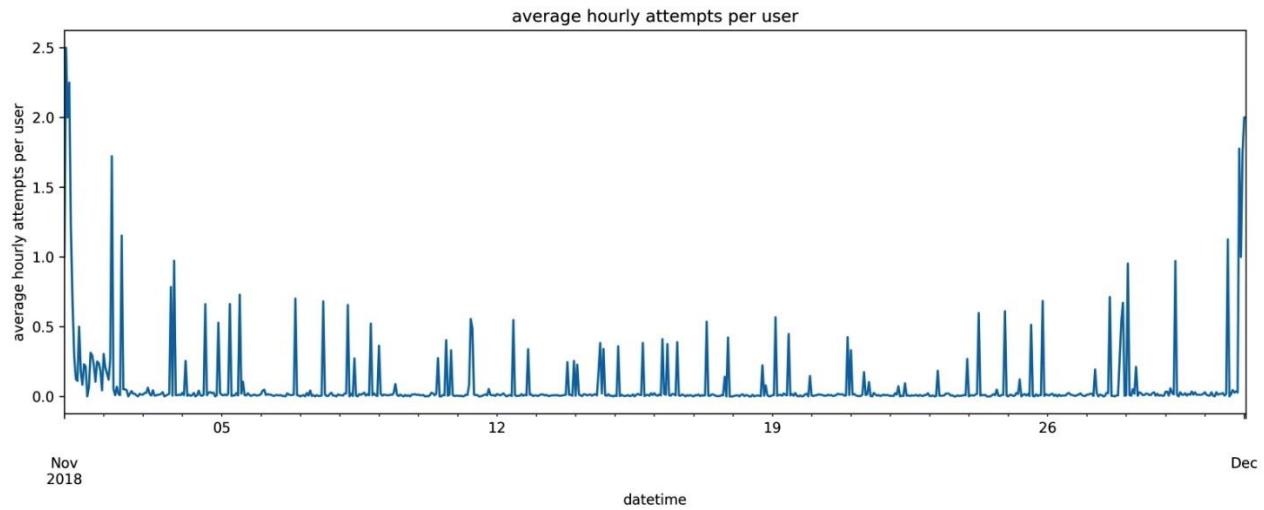
	source_ip	username	success	failure_reason
	datetime			
2018-11-01 00:36:52.617978	142.89.86.32	vkim	True	NaN
2018-11-01 01:00:23.166623	5.118.187.36	kkim	True	NaN
2018-11-01 01:31:50.779608	142.89.86.32	vkim	False	error_wrong_password
2018-11-01 01:31:51.779608	142.89.86.32	vkim	True	NaN
2018-11-01 01:32:44.016230	15.176.178.91	kkim	True	NaN
	start	end	source_ip	
0	2018-11-02 05:06:17.152636	2018-11-02 05:10:30.152636	212.79.15.228	
1	2018-11-02 11:42:38.771415	2018-11-02 11:45:58.771415	44.207.171.119	
2	2018-11-03 17:49:39.023954	2018-11-03 17:52:27.023954	15.223.158.165	
3	2018-11-03 19:45:05.820292	2018-11-03 19:49:11.820292	68.102.121.161	
4	2018-11-04 02:51:07.163402	2018-11-04 02:52:09.163402	103.93.254.233	



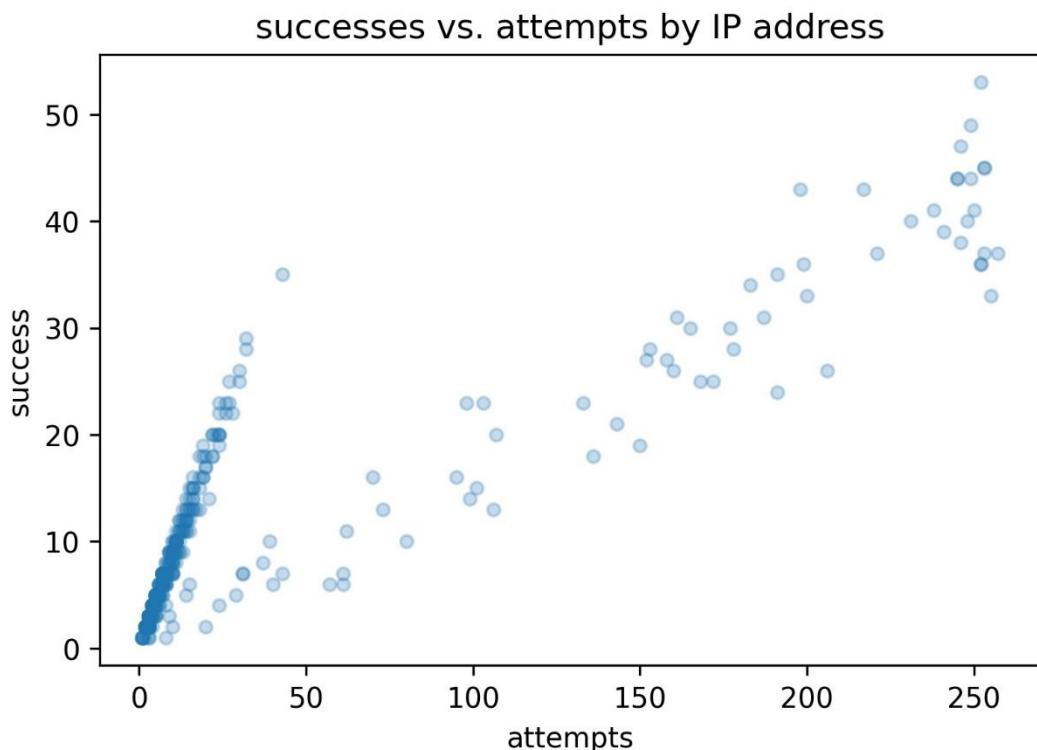
failure_reason error_account_locked error_wrong_password error_wrong_username

is_hacker

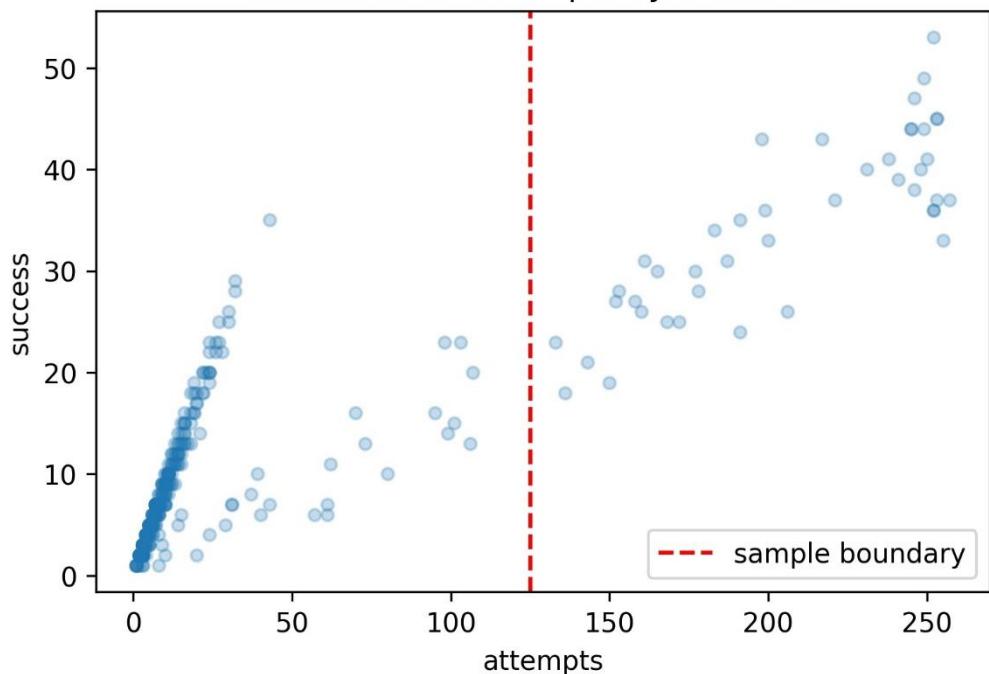
False	1	299	2
True	0	3316	5368



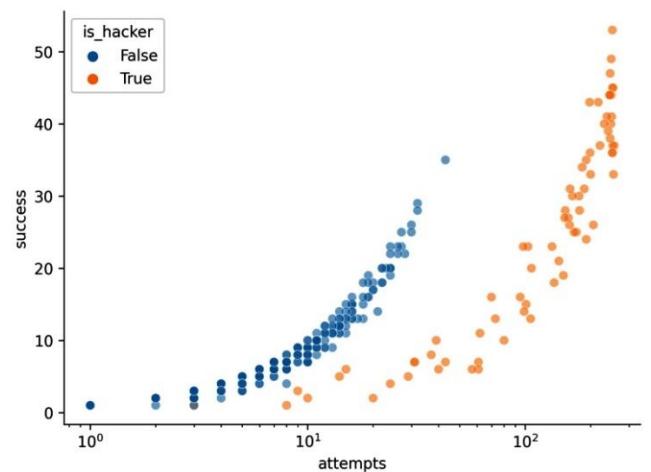
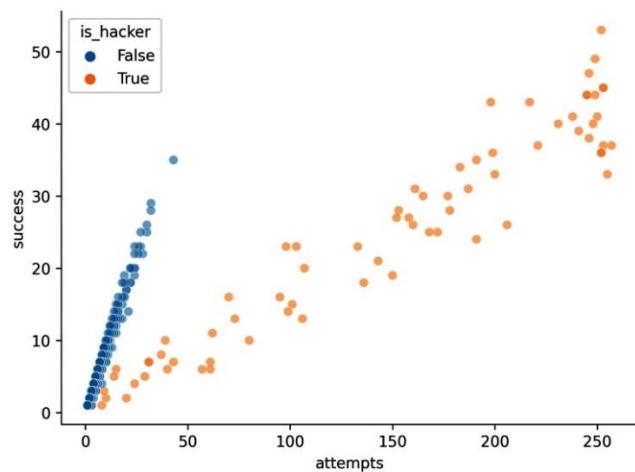
failure_reason	attempts	error_account_locked	error_wrong_password	error_wrong_username	success	success_rate	error_rate
source_ip							
85.1.221.89	257	0	92	128	37	0.143969	0.856031
109.67.154.113	255	0	78	144	33	0.129412	0.870588
212.79.15.228	253	0	89	127	37	0.146245	0.853755
181.217.195.170	253	0	70	138	45	0.177866	0.822134
211.56.212.113	253	0	88	120	45	0.177866	0.822134



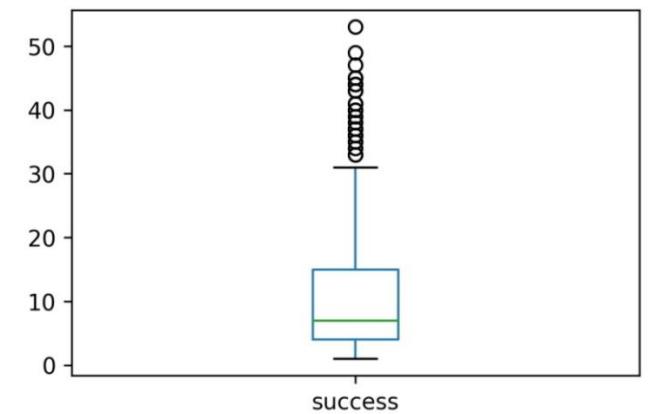
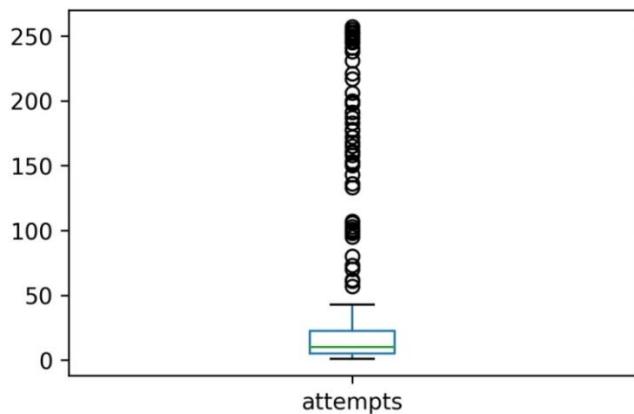
successes vs. attempts by IP address



successes vs. attempts by IP address



stats per IP address



	source_ip	datetime	username	success	failures	attempts	success_rate	failure_rate
0	1.138.149.116	2018-11-01 04:00:00		1	5	1	6	0.833333 0.166667
1	1.138.149.116	2018-11-05 18:00:00		1	1	0	1	1.000000 0.000000
2	1.138.149.116	2018-11-05 19:00:00		1	1	0	1	1.000000 0.000000
3	1.138.149.116	2018-11-06 03:00:00		1	2	0	2	1.000000 0.000000
4	1.138.149.116	2018-11-06 04:00:00		1	2	0	2	1.000000 0.000000

	username	success	failures	attempts	success_rate	failure_rate	hour
hour							
19	14.9	5.5	21.4	26.9	0.736876	0.263124	19.0
23	12.4	3.9	18.7	22.6	0.791195	0.208805	23.0
3	1.0	1.1	0.4	1.5	0.800000	0.200000	3.0
11	1.1	2.0	0.6	2.6	0.816667	0.183333	11.0
14	24.7	8.4	35.5	43.9	0.833401	0.166599	14.0
16	1.0	1.5	0.4	1.9	0.841667	0.158333	16.0

	username	success	failures	attempts	success_rate	failure_rate	hour
hour							
19	1.0	1.4	0.4	1.8	0.871429	0.128571	19.0
23	1.0	2.0	0.1	2.1	0.966667	0.033333	23.0
3	1.0	2.0	0.3	2.3	0.925000	0.075000	3.0
11	1.1	1.9	0.2	2.1	0.933333	0.066667	11.0
14	1.0	1.4	0.2	1.6	0.950000	0.050000	14.0
16	1.0	1.4	0.2	1.6	0.925000	0.075000	16.0

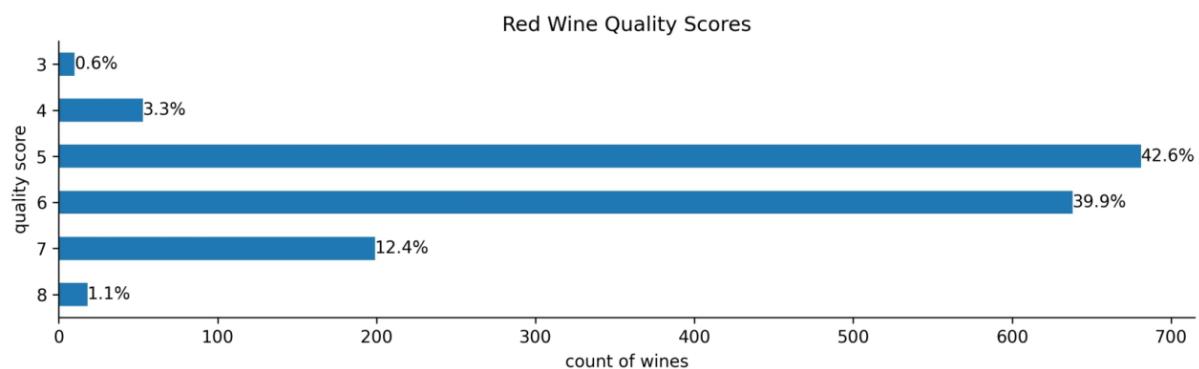
Confusion Matrix

		Predicted	
		True	False
Actual	True	TP (True Positive)	FP (False Positive)
	False	FN (False Negative)	TN (True Negative)

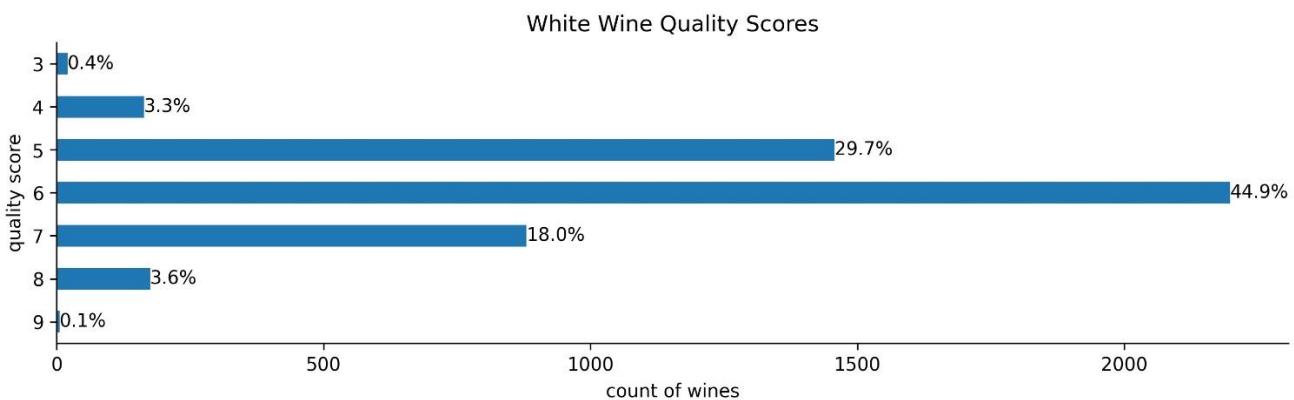
	means	medians	Tukey fence	Z-scores
FPR	0.003922	0.007843	0.078431	0.000000
FDR	0.013699	0.027027	0.240964	0.000000
FNR	0.000000	0.000000	0.125000	0.138889
FOR	0.000000	0.000000	0.036885	0.037736

Chapter 9: Getting Started with Machine Learning in Python

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

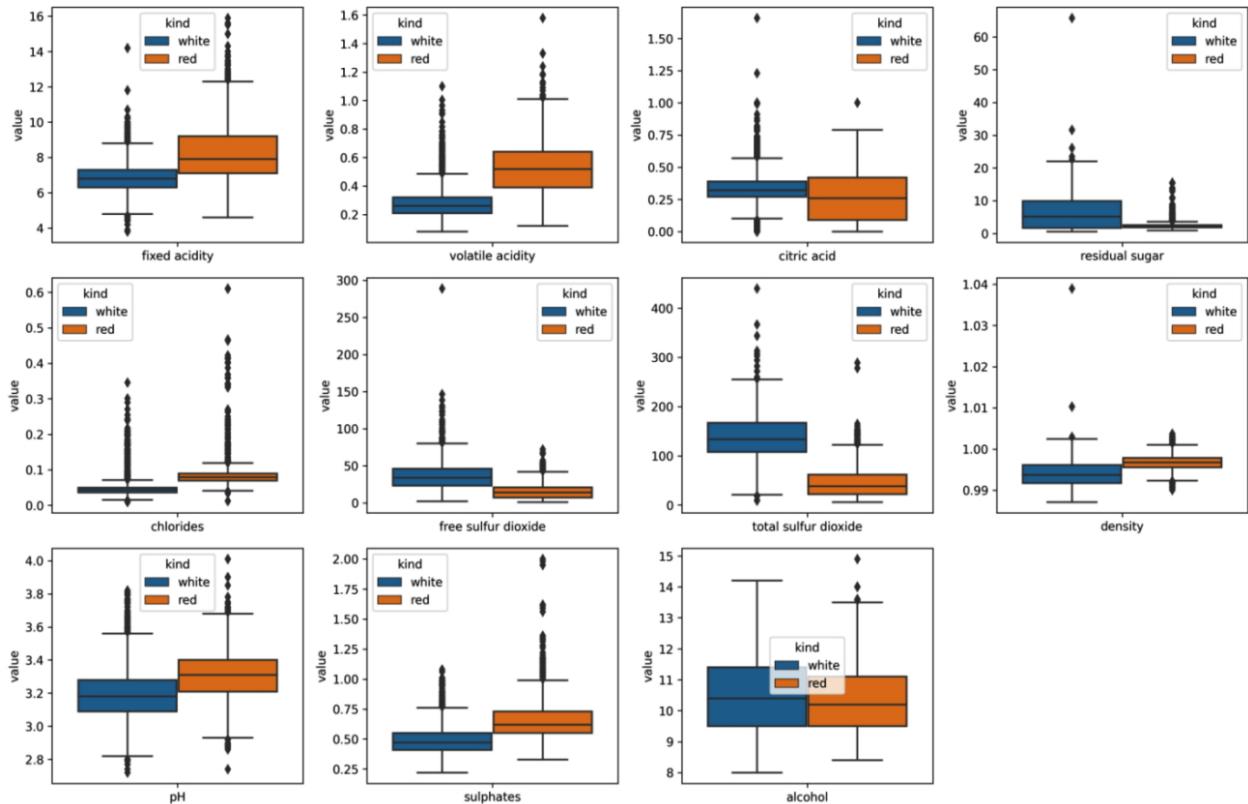


	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	0.996747	3.311113	0.658149	10.422983	5.636023
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001887	0.154386	0.169507	1.065668	0.807569
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	8.400000	3.000000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995600	3.210000	0.550000	9.500000	5.000000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996750	3.310000	0.620000	10.200000	6.000000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997835	3.400000	0.730000	11.100000	6.000000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1.003690	4.010000	2.000000	14.900000	8.000000

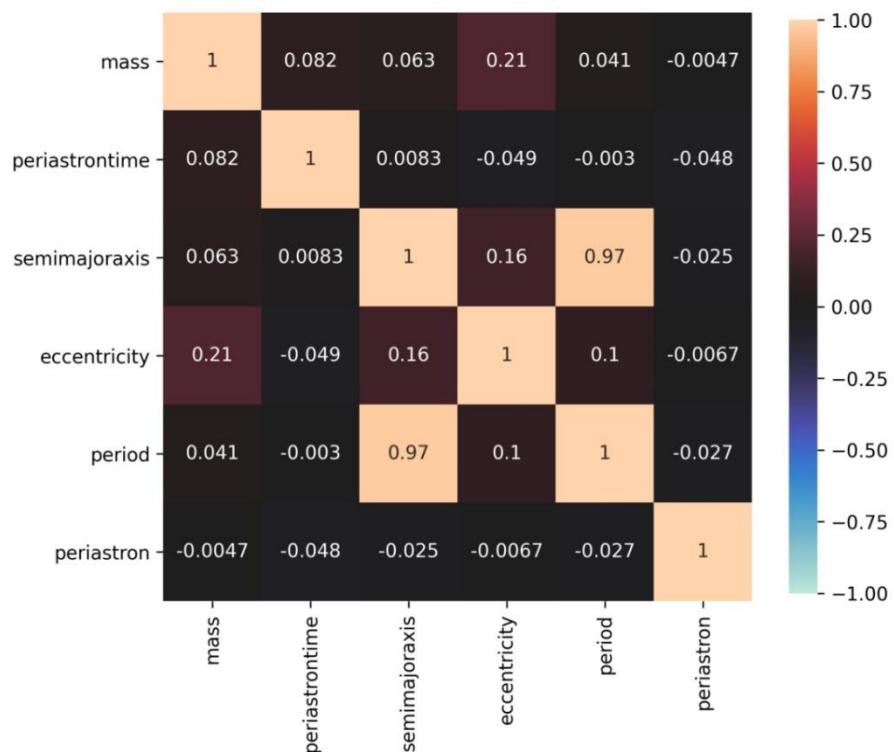


	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	kind
848	6.4	0.64	0.21	1.8	0.081	14.0	31.0	0.99689	3.59	0.66	9.8	5	red
2529	6.6	0.42	0.13	12.8	0.044	26.0	158.0	0.99772	3.24	0.47	9.0	5	white
131	5.6	0.50	0.09	2.3	0.049	17.0	99.0	0.99370	3.63	0.63	13.0	5	red
244	15.0	0.21	0.44	2.2	0.075	10.0	24.0	1.00005	3.07	0.84	9.2	7	red
1551	6.6	0.19	0.99	1.2	0.122	45.0	129.0	0.99360	3.09	0.31	8.7	6	white

Comparing Chemical Properties of Red and White Wines



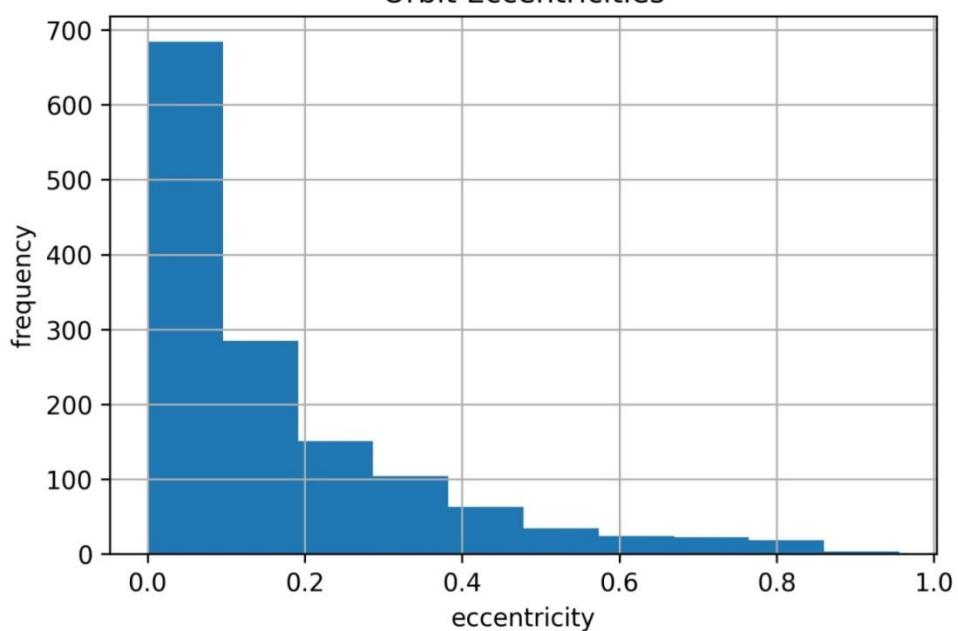
	mass	description	periastron time	semimajoraxis	discovery year	list	eccentricity	period	discovery method	last update	periastron	name
0	19.400	11 Com ...	2452899.60	1.290	2008.0	Confirmed planets	0.231	326.03	RV	15/09/20	94.800	11 Com b
1	11.200	11 Ursa...	2452861.04	1.540	2009.0	Confirmed planets	0.080	516.22	RV	15/09/20	117.630	11 UMi b
2	4.800	14 Andr...	2452861.40	0.830	2008.0	Confirmed planets	0.000	185.84	RV	15/09/20	0.000	14 And b
3	4.975	The sta...	NaN	2.864	2002.0	Confirmed planets	0.359	1766.00	RV	15/09/21	22.230	14 Her b
4	7.679	14 Her ...	NaN	9.037	2006.0	Controversial	0.184	9886.00	RV	15/09/21	189.076	14 Her c

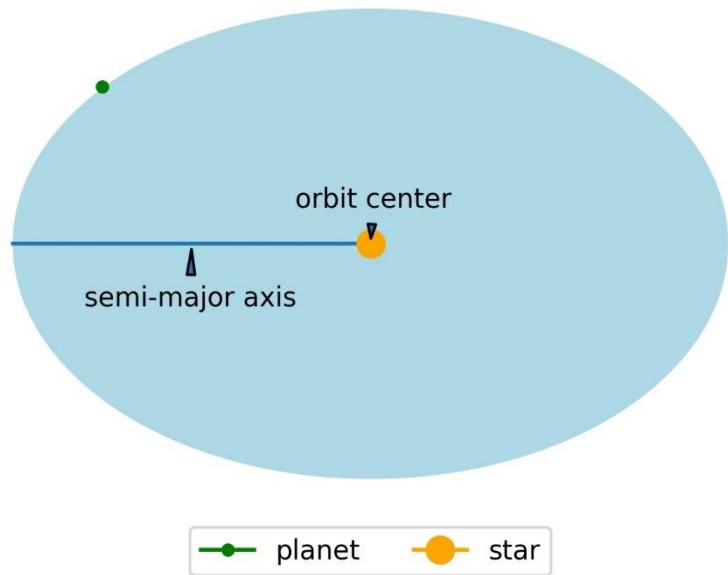


Eccentricity Orbit Shape

0	Circular
(0, 1)	Elliptical
1	Parabolic
> 1	Hyperbolic

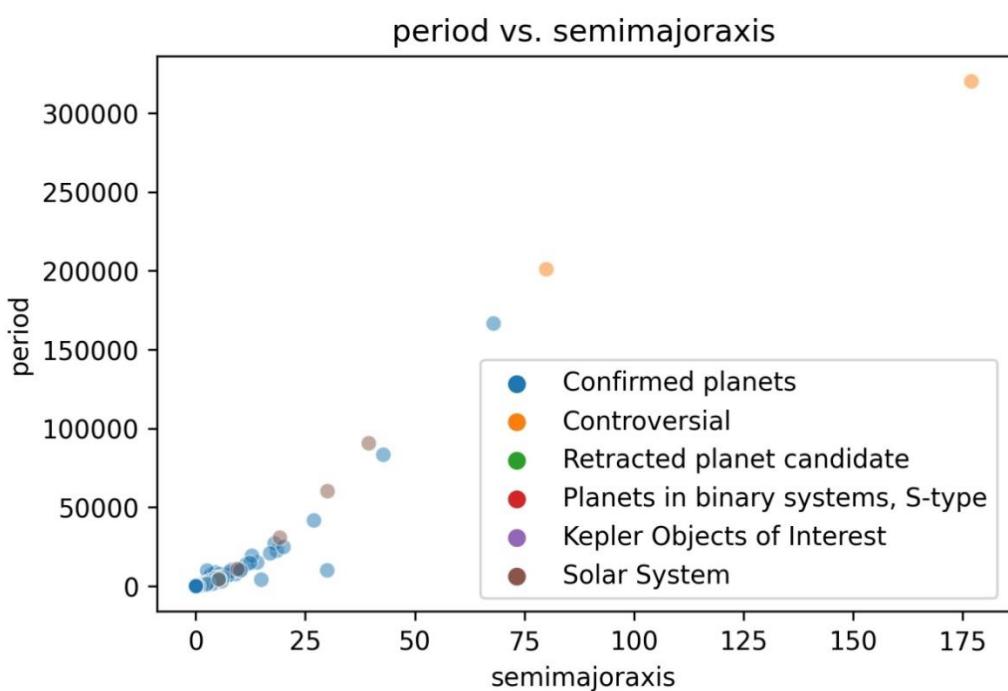
Orbit Eccentricities



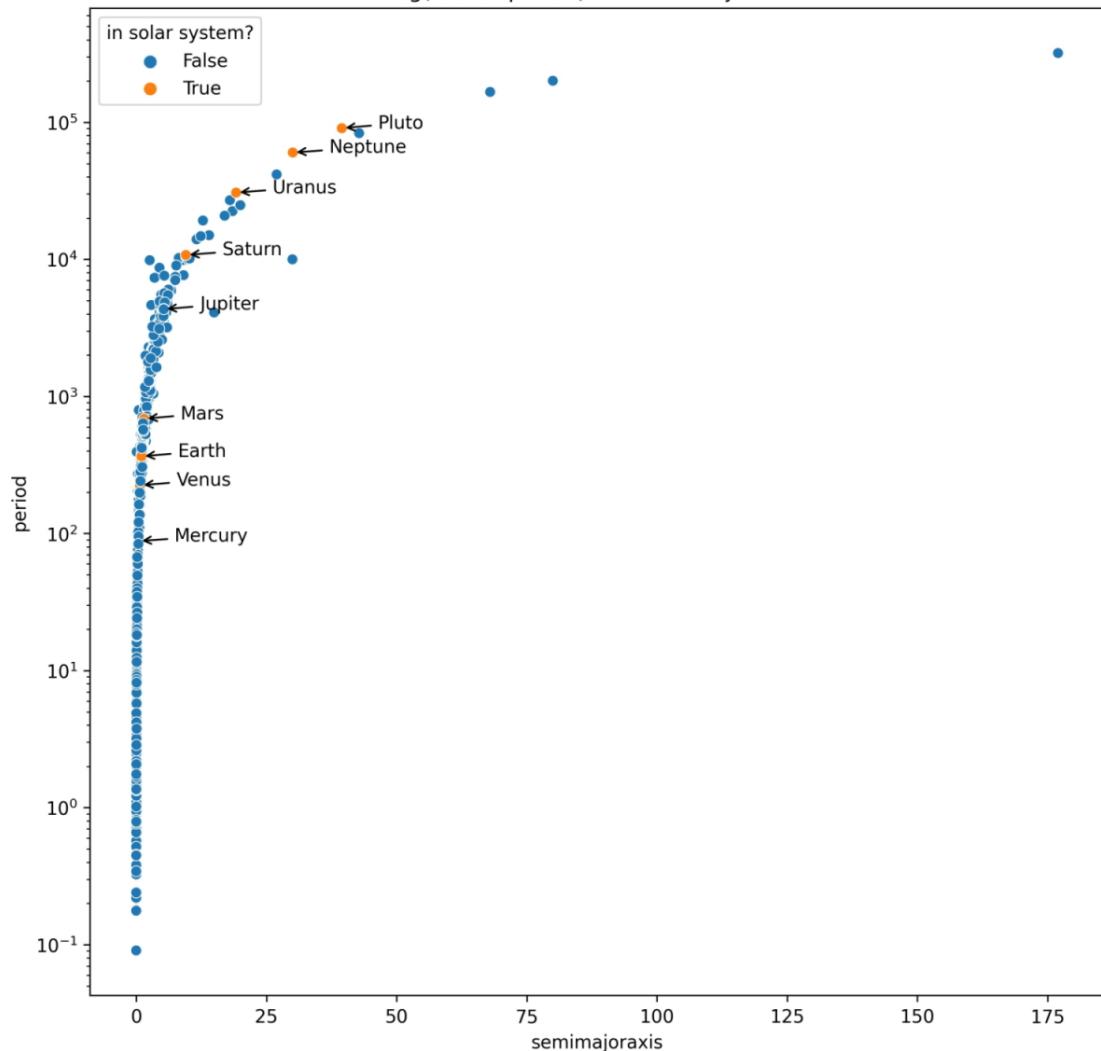


planet star

	period	eccentricity	semimajoraxis	mass
count	3930.000000	1388.000000	1704.000000	1659.000000
mean	524.084969	0.159016	5.837964	2.702061
std	7087.428665	0.185041	110.668743	8.526177
min	0.090706	0.000000	0.004420	0.000008
25%	4.552475	0.013000	0.051575	0.085000
50%	12.364638	0.100000	0.140900	0.830000
75%	46.793136	0.230000	1.190000	2.440000
max	320000.000000	0.956000	3500.000000	263.000000



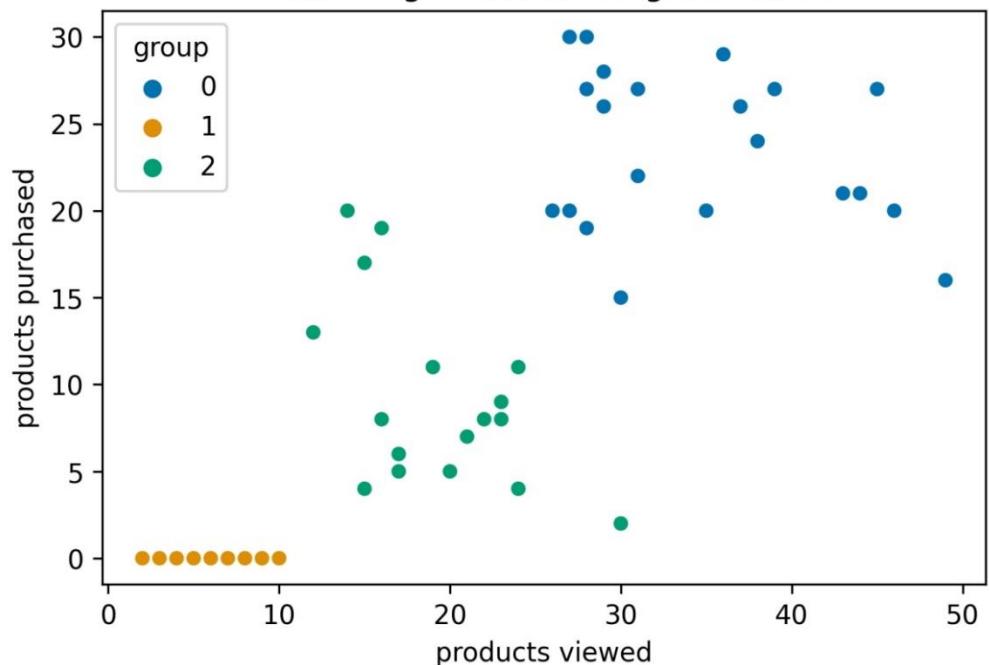
log(orbital period) vs. semi-major axis



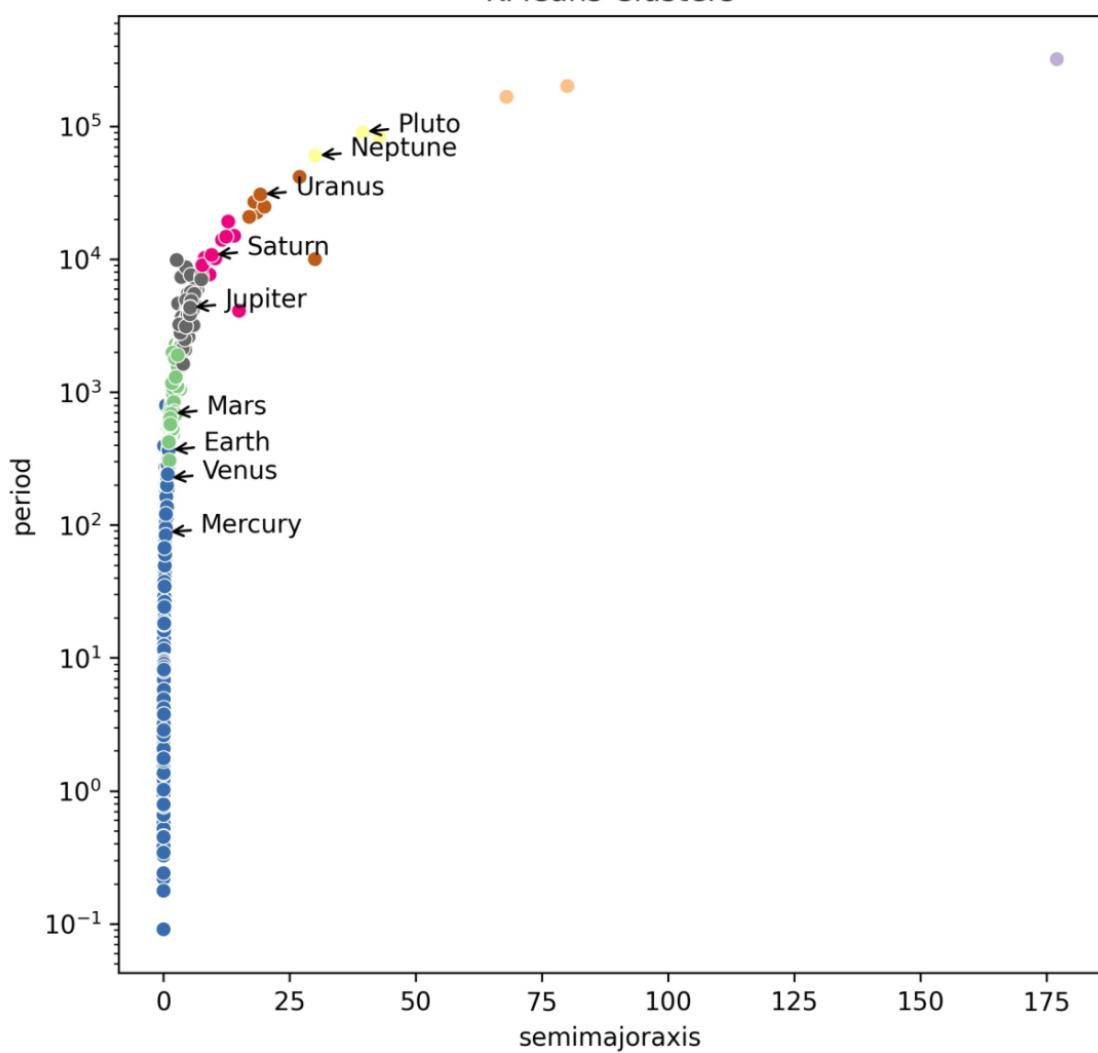
	Confirmed planets	Controversial	Kepler Objects of Interest	Planets in binary systems, S-type	Retracted planet candidate	Solar System
0	1	0	0	0	0	0
1	1	0	0	0	0	0
2	1	0	0	0	0	0
3	1	0	0	0	0	0
4	0	1	0	0	0	0

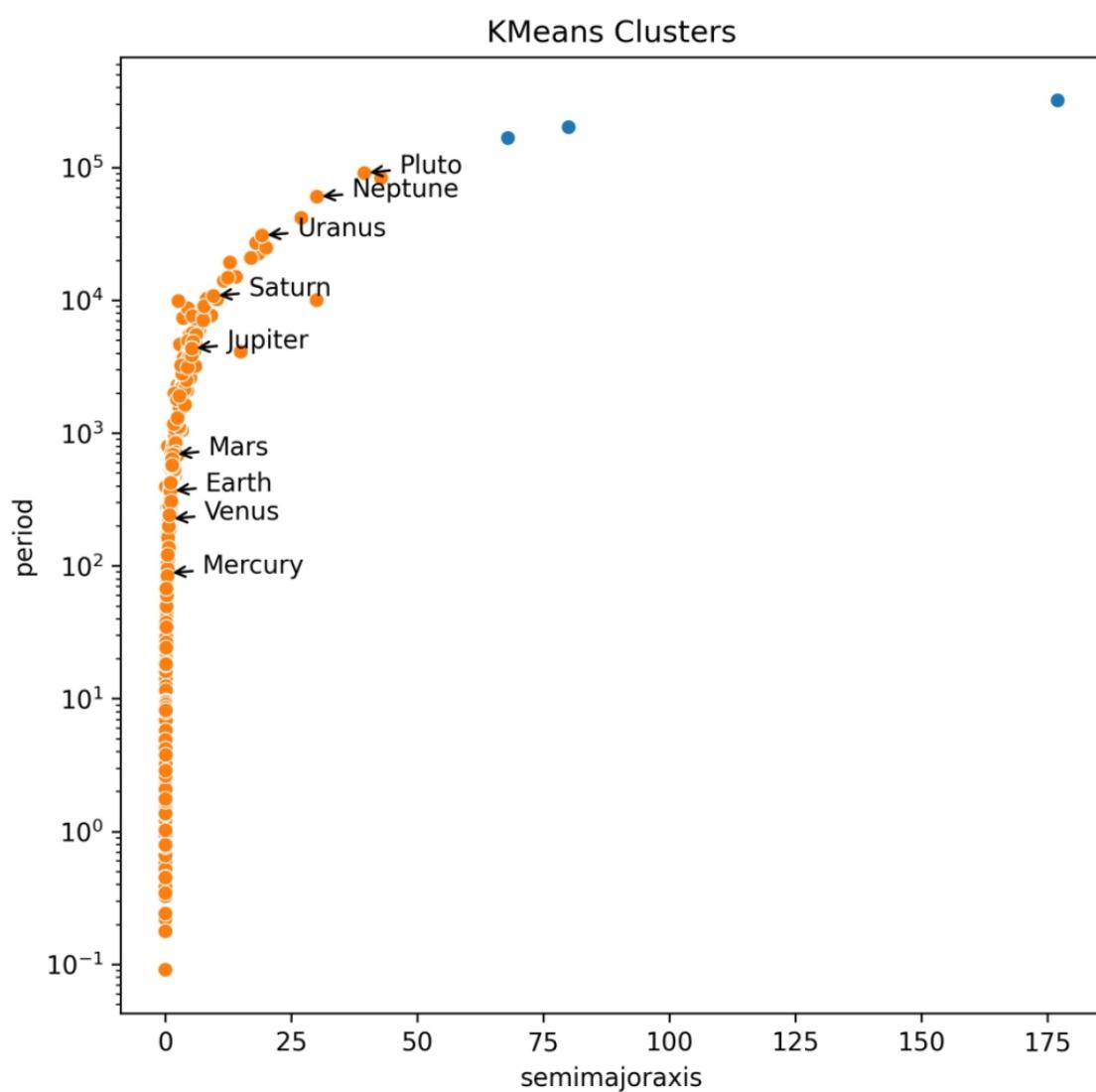
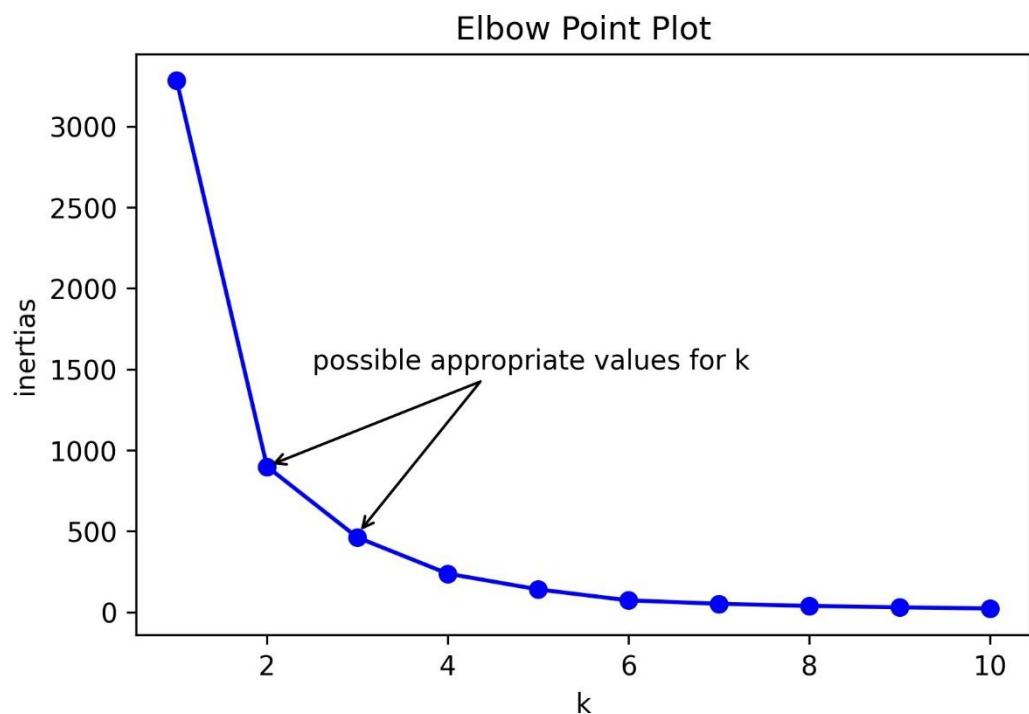
	Controversial	Kepler Objects of Interest	Planets in binary systems, S-type	Retracted planet candidate	Solar System
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	1	0	0	0	0

Clustering for Market Segmentation

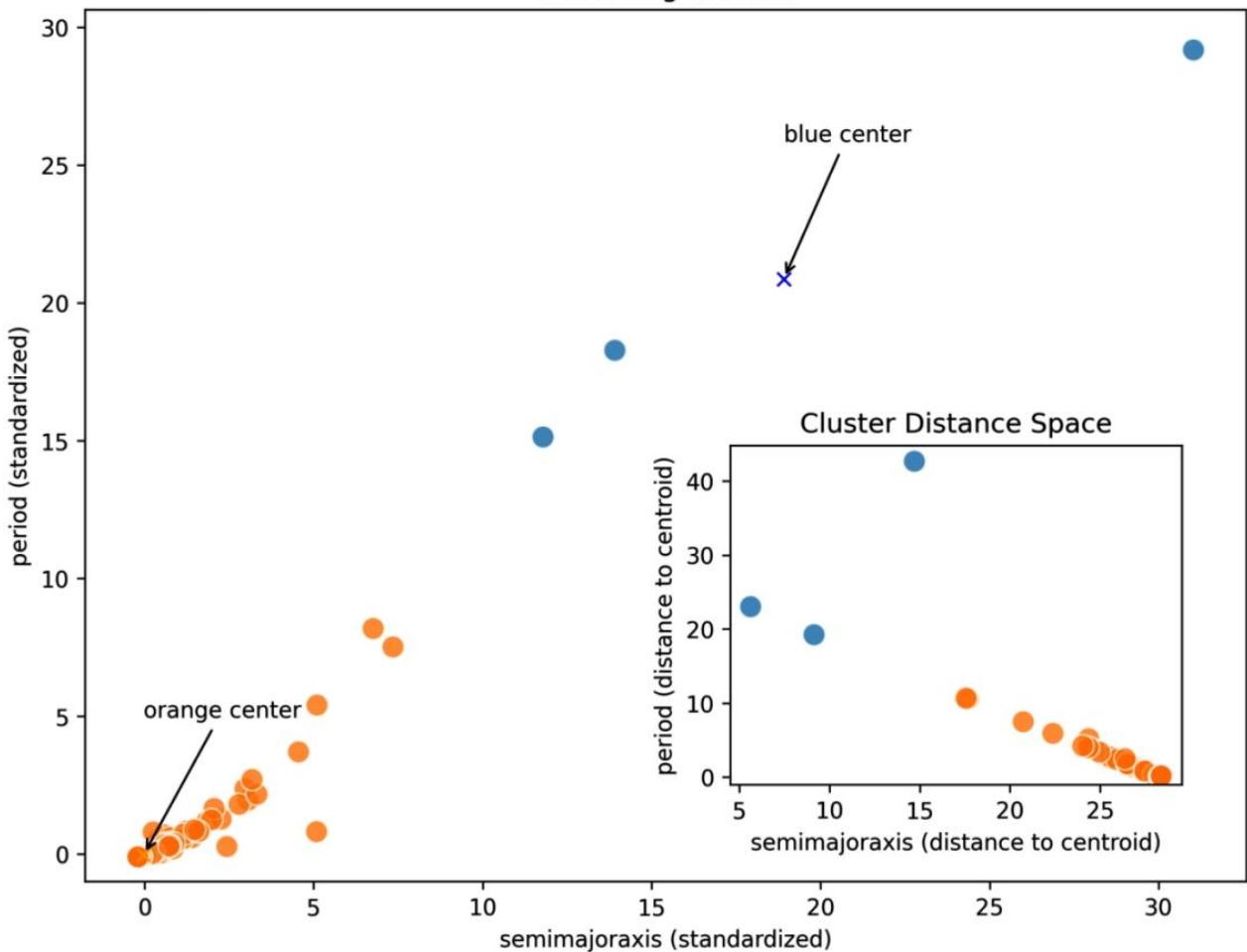


KMeans Clusters



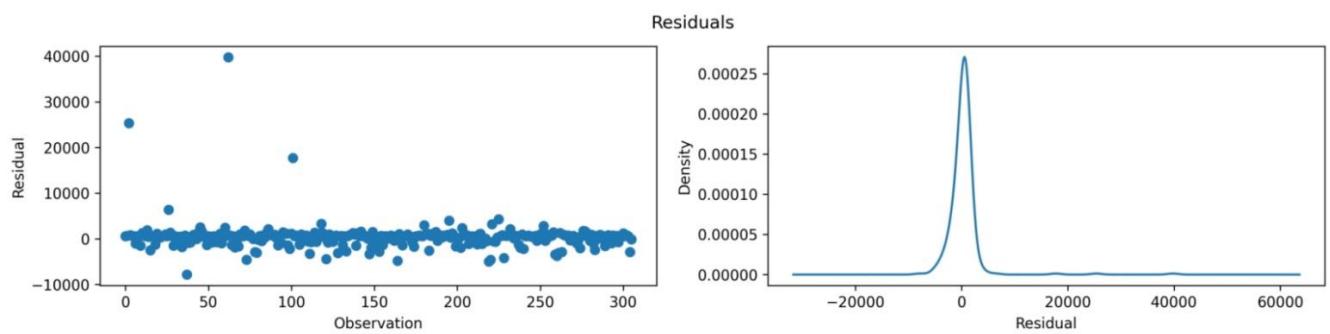
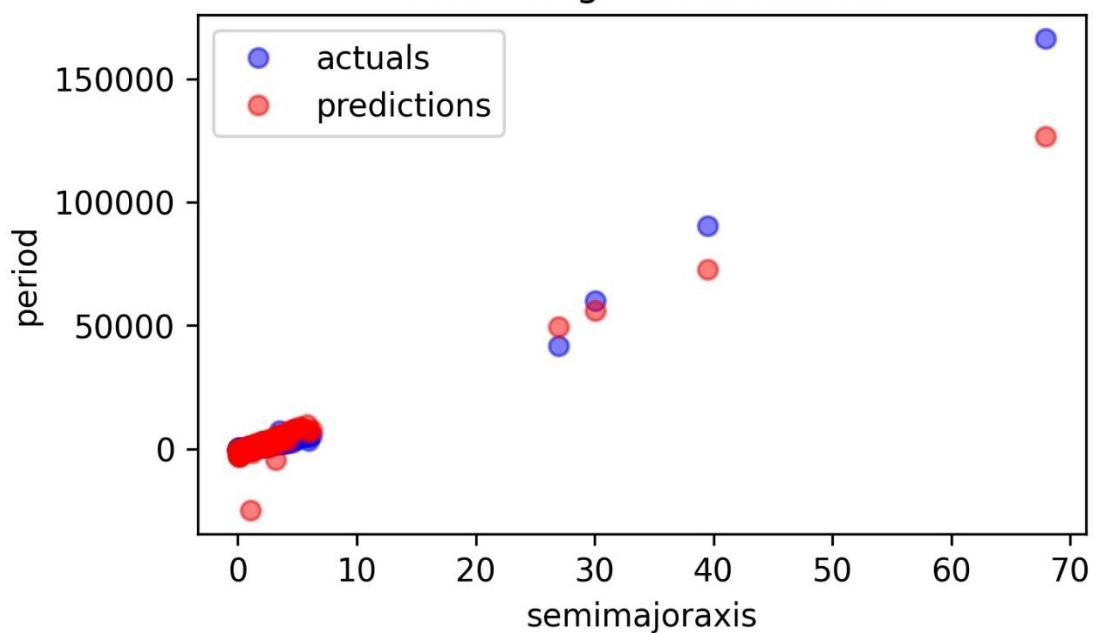


Visualizing Clusters

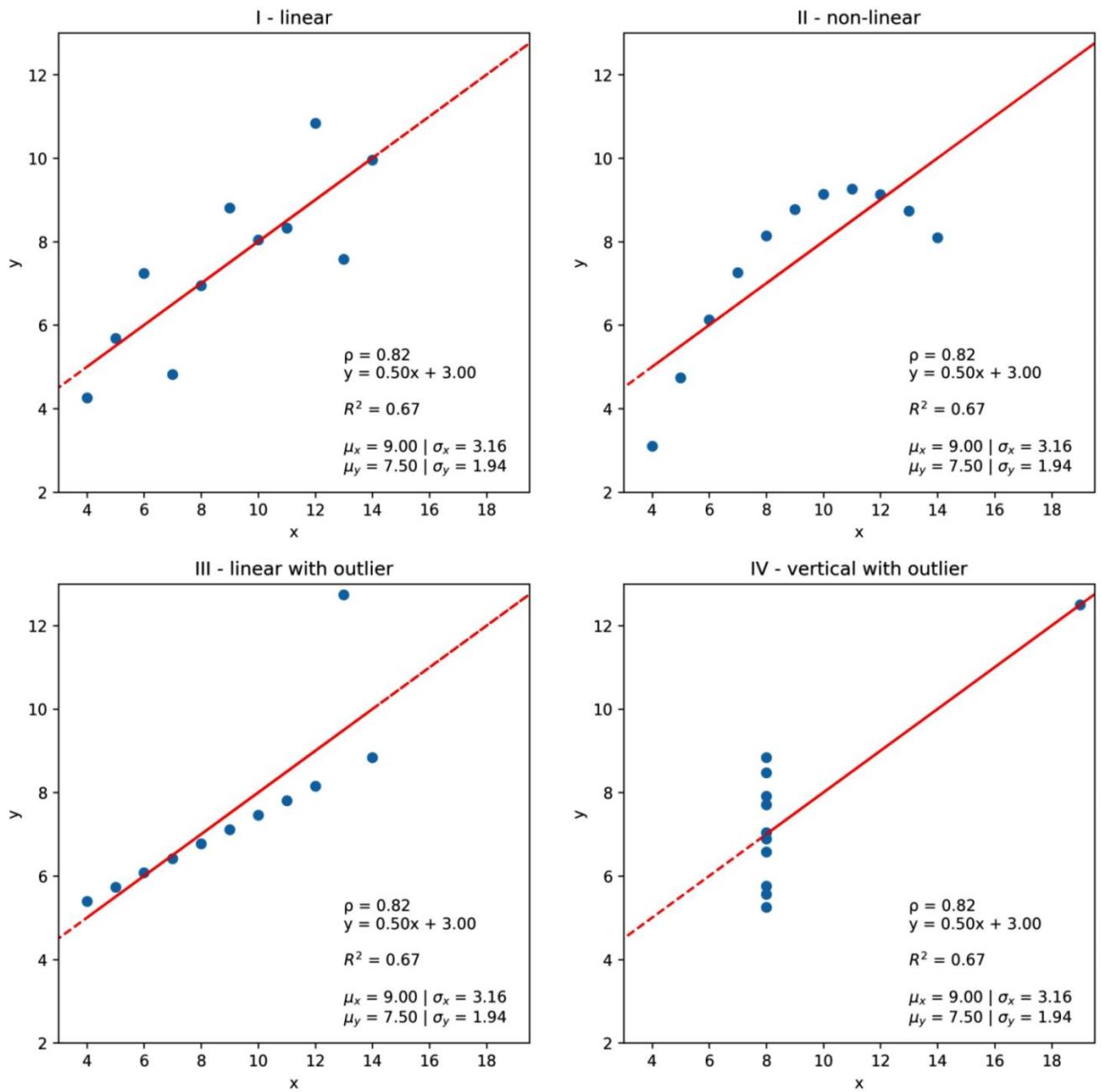


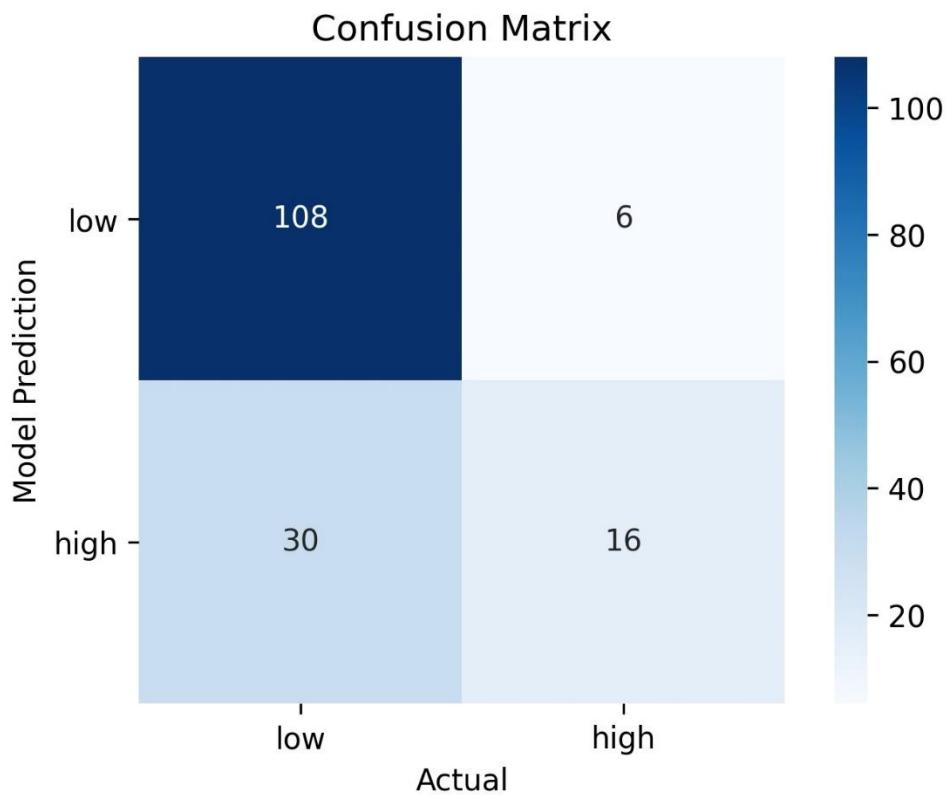
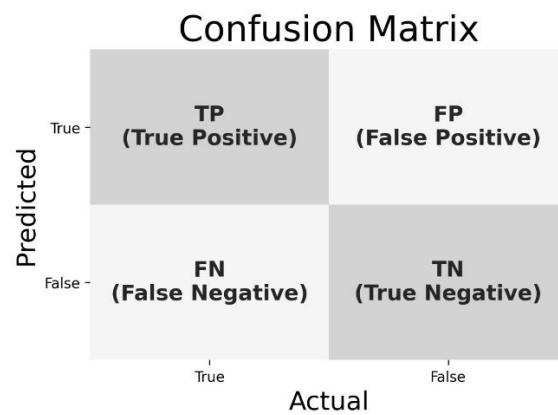
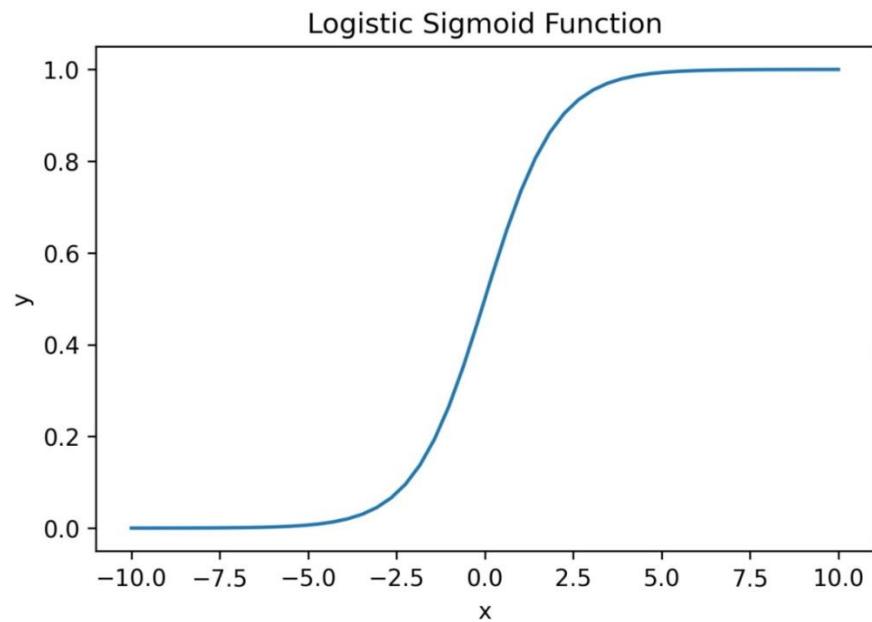
Method	Action	Used when...
<code>fit()</code>	Train the model or preprocessor	Modeling, preprocessing
<code>transform()</code>	Transform the data into the new space	Clustering, preprocessing
<code>fit_transform()</code>	Run <code>fit()</code> , followed by <code>transform()</code>	Clustering, preprocessing
<code>score()</code>	Evaluate the model using the default scoring method	Modeling
<code>predict()</code>	Use model to predict output values for given inputs	Modeling
<code>fit_predict()</code>	Run <code>fit()</code> , followed by <code>predict()</code>	Modeling
<code>predict_proba()</code>	Like <code>predict()</code> , but returns the probability of belonging to each class	Classification

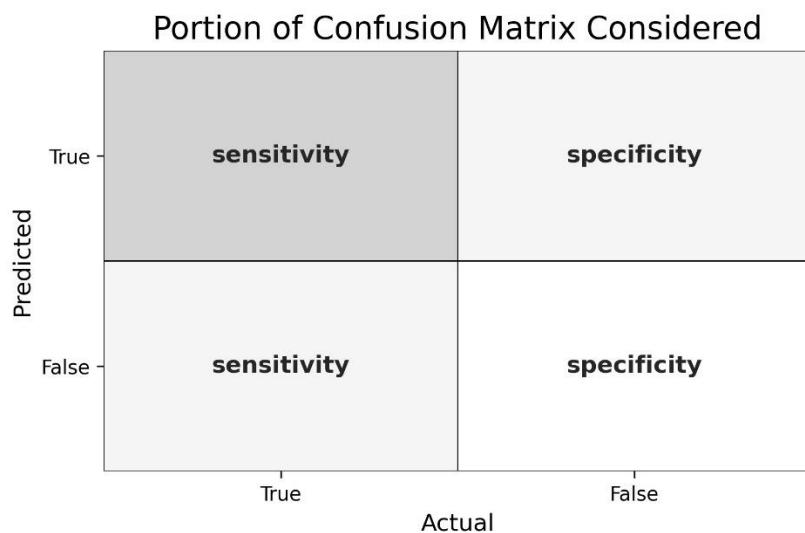
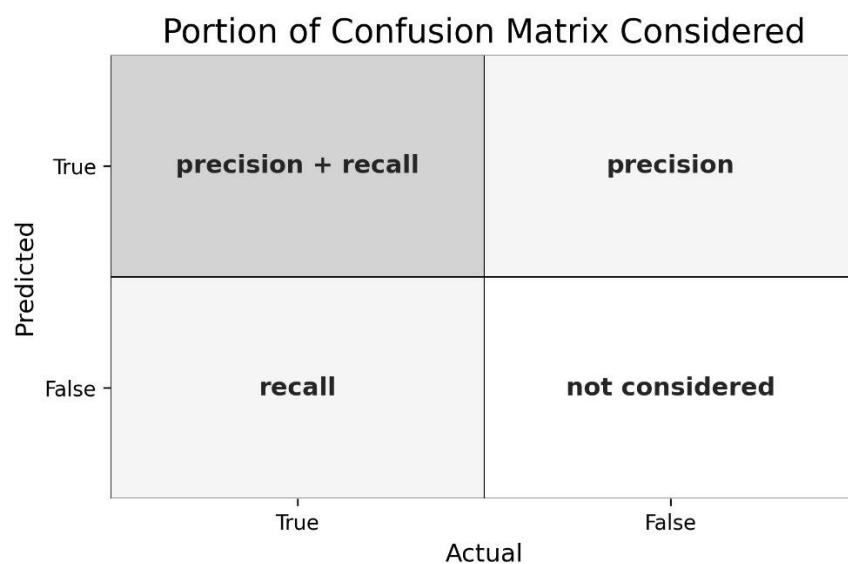
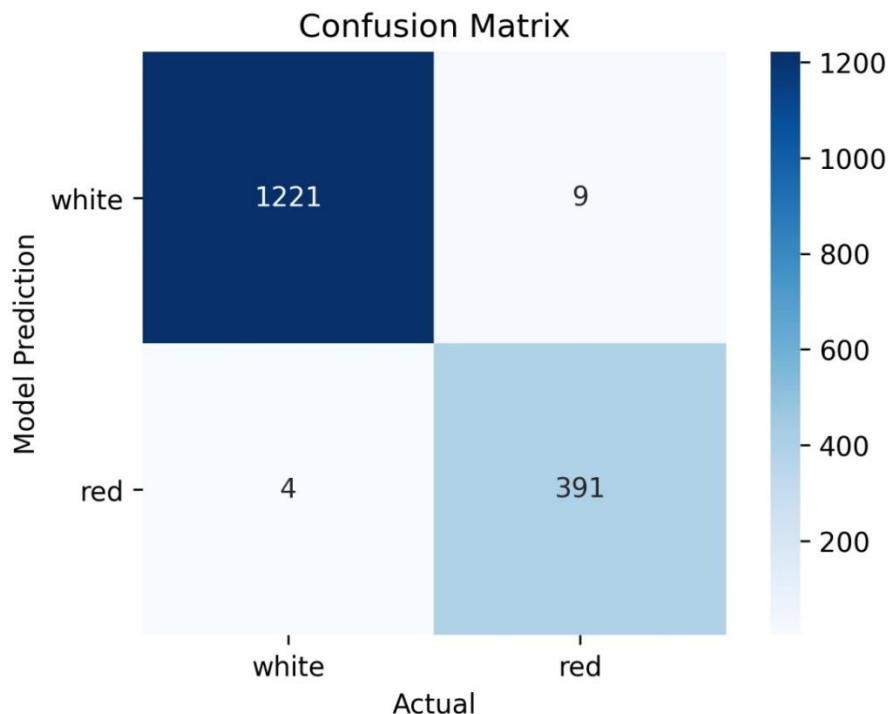
Linear Regression Results



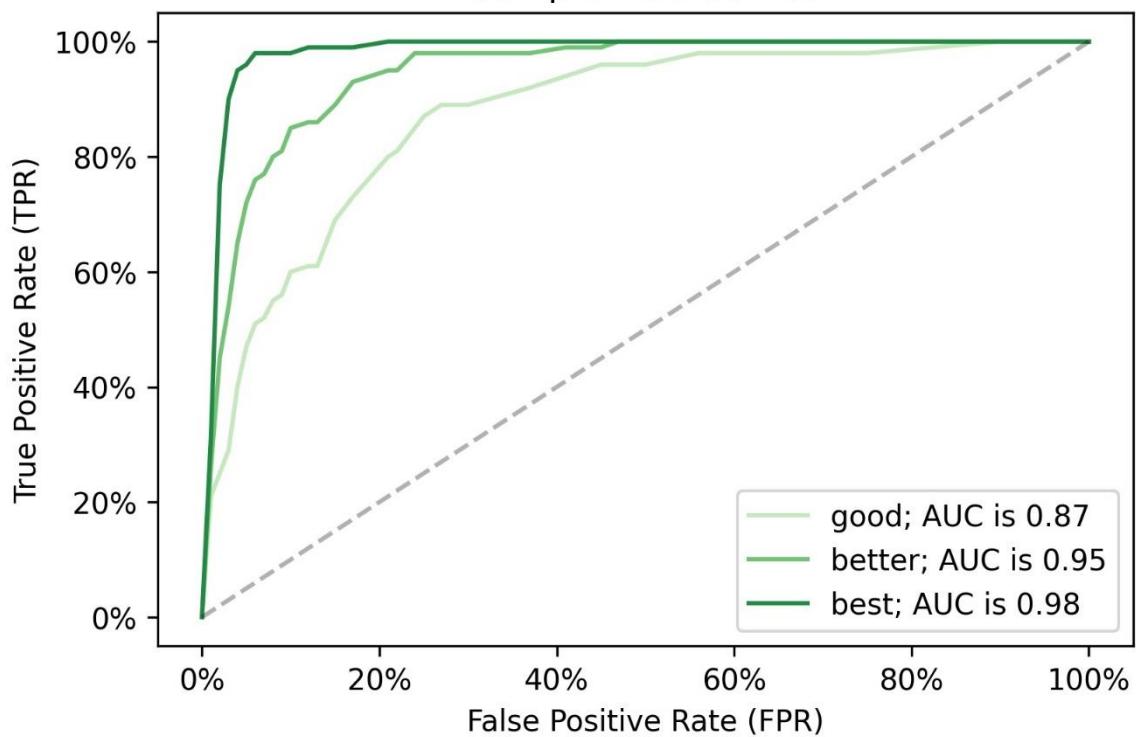
Anscombe's Quartet



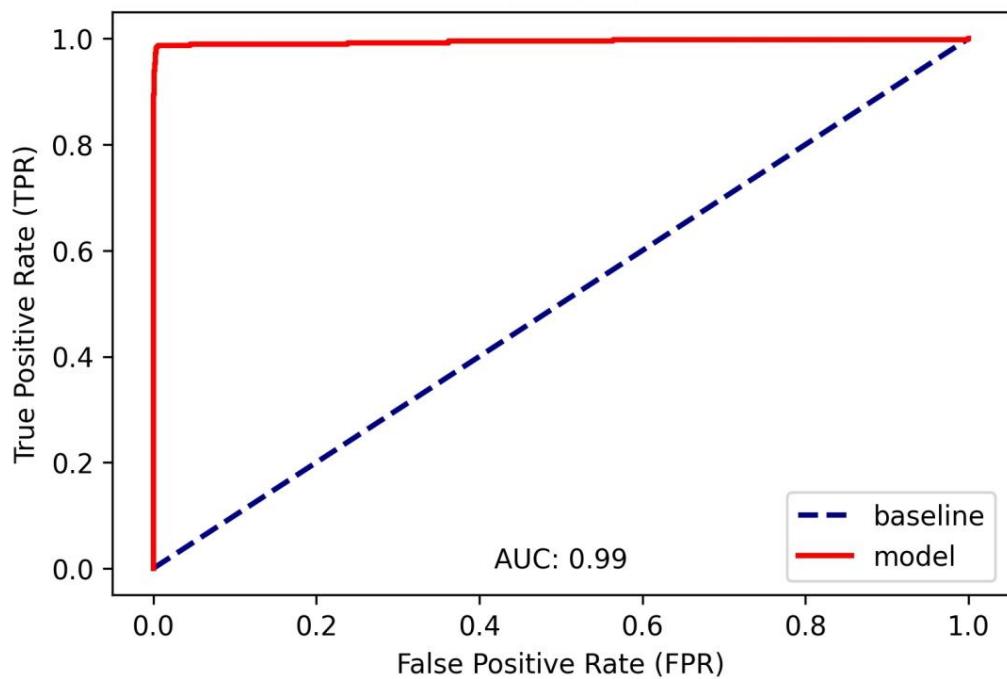


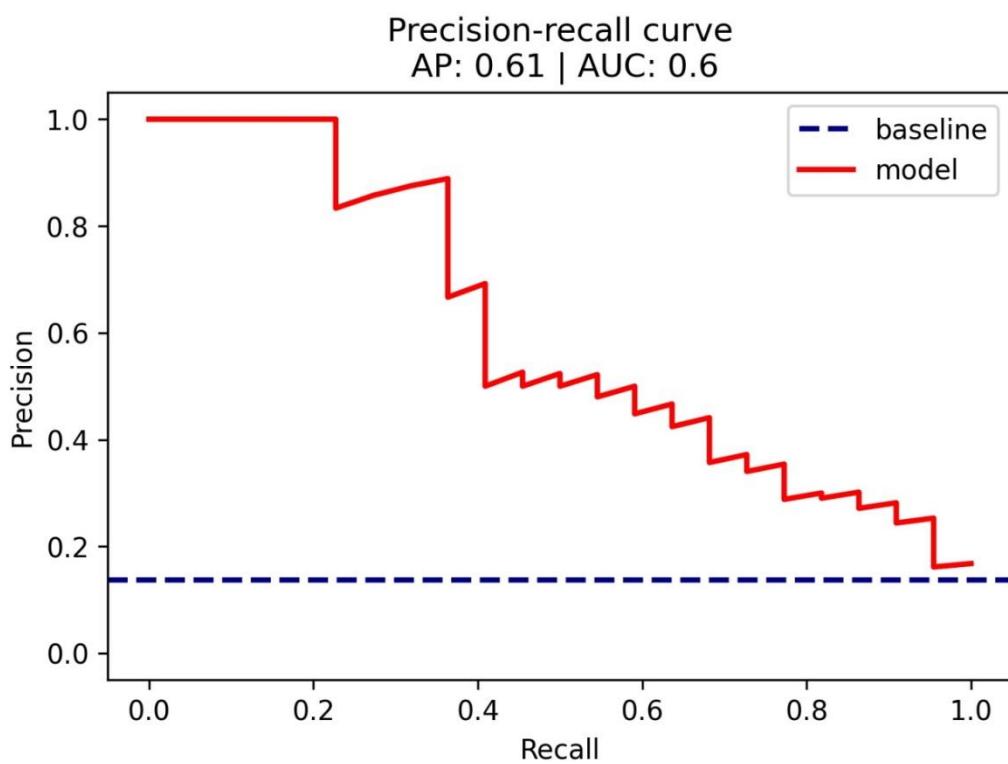
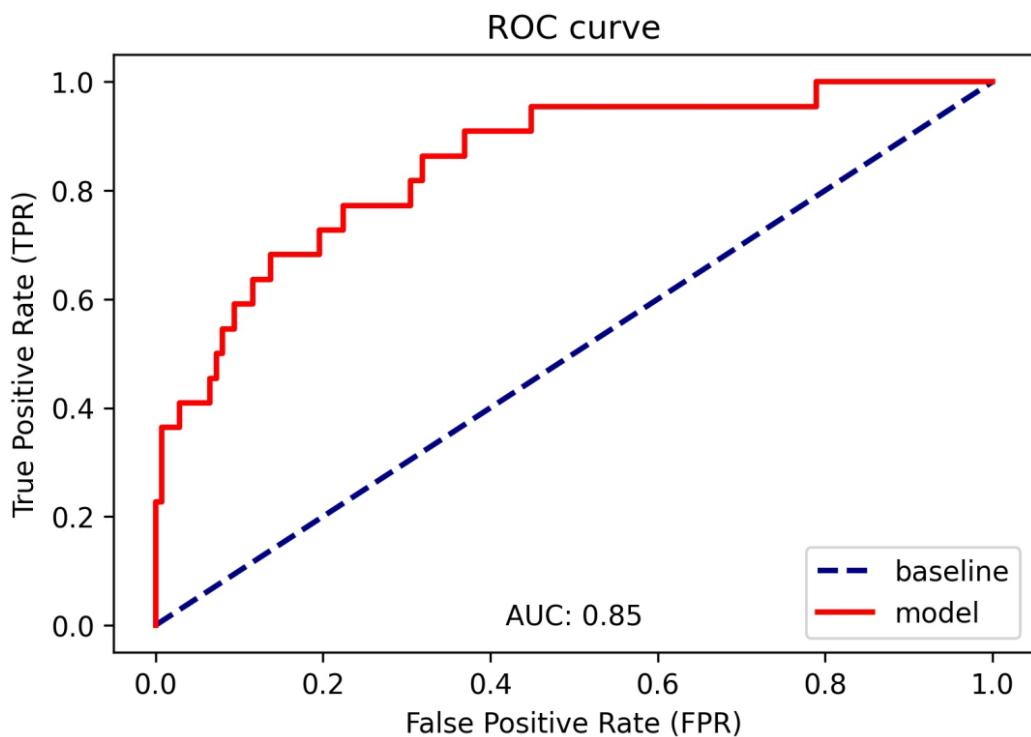


Sample ROC Curves

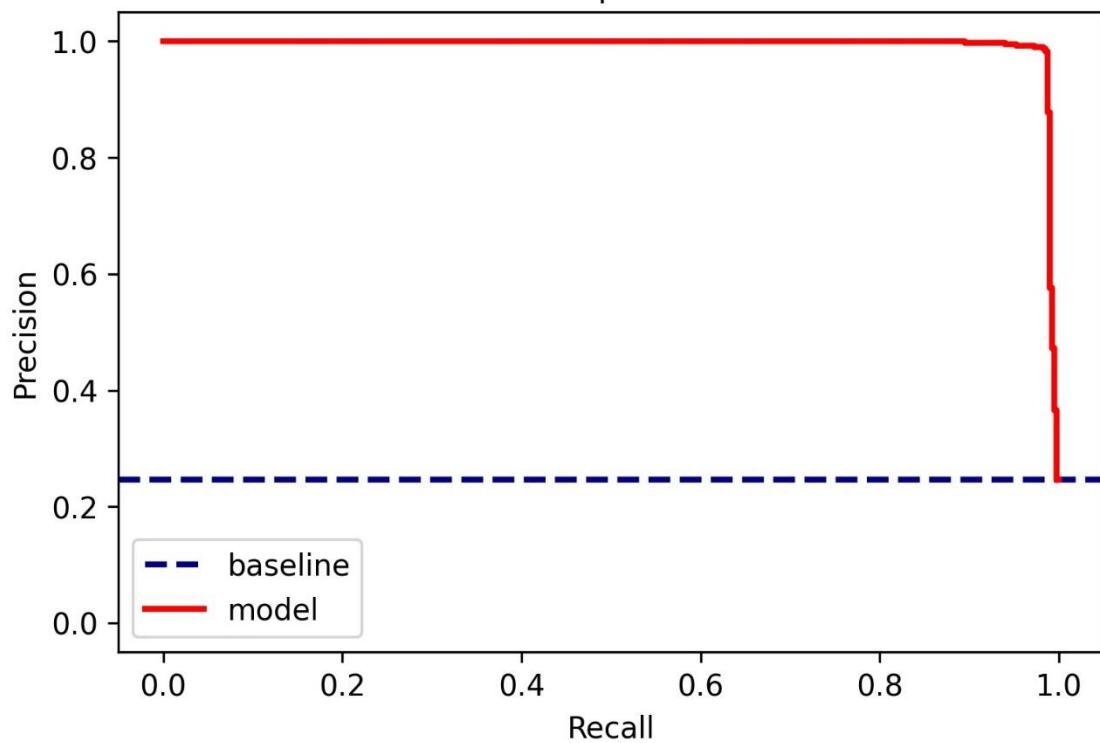


ROC curve

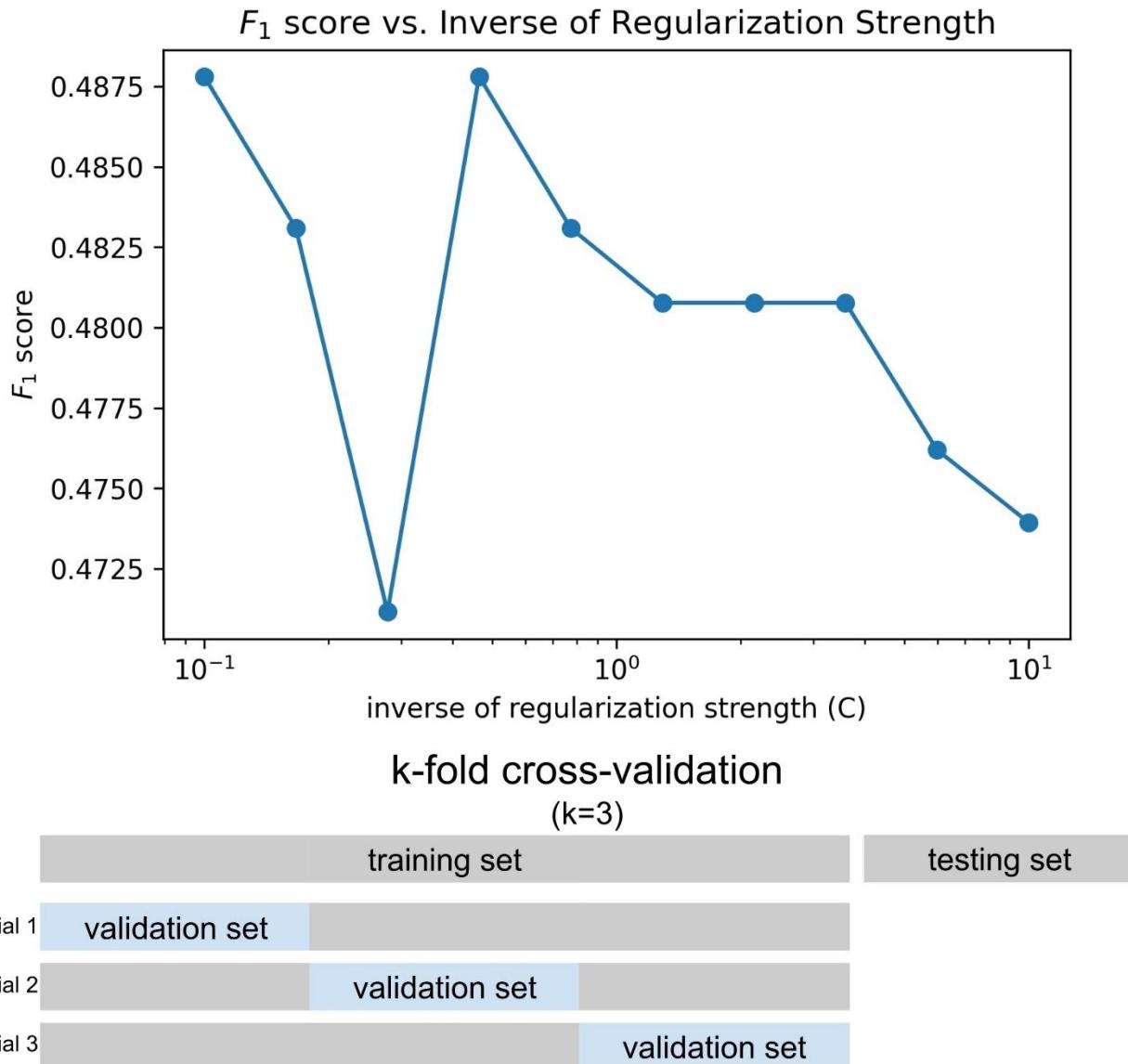




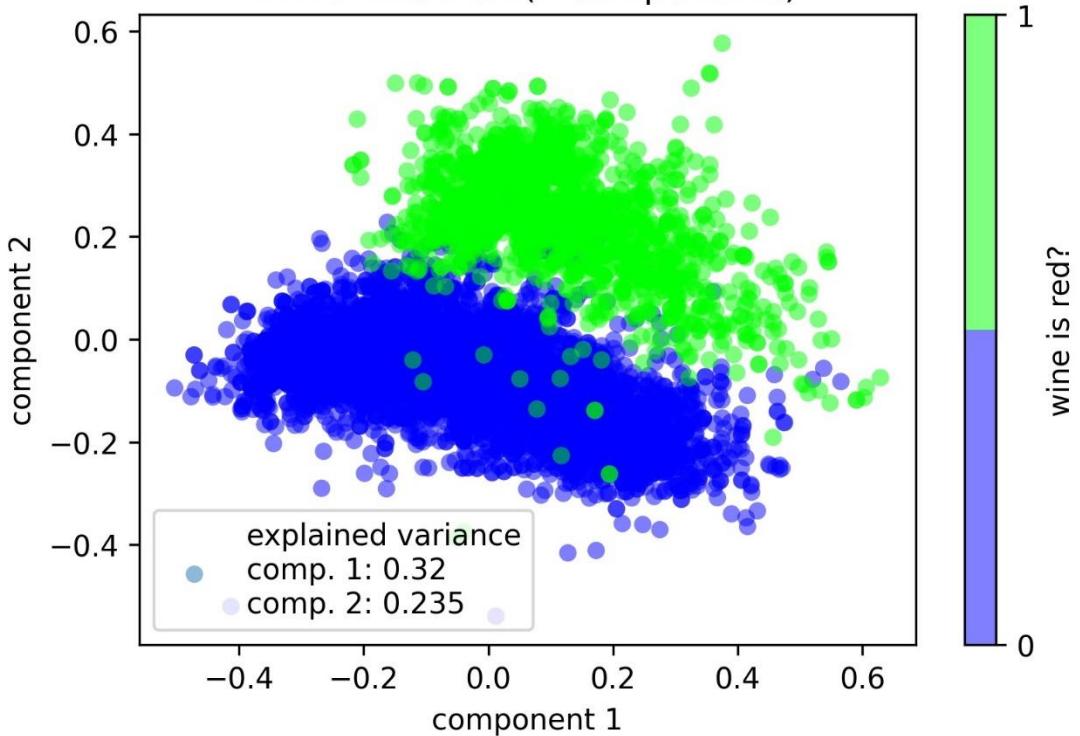
Precision-recall curve
AP: 0.99 | AUC: 0.99



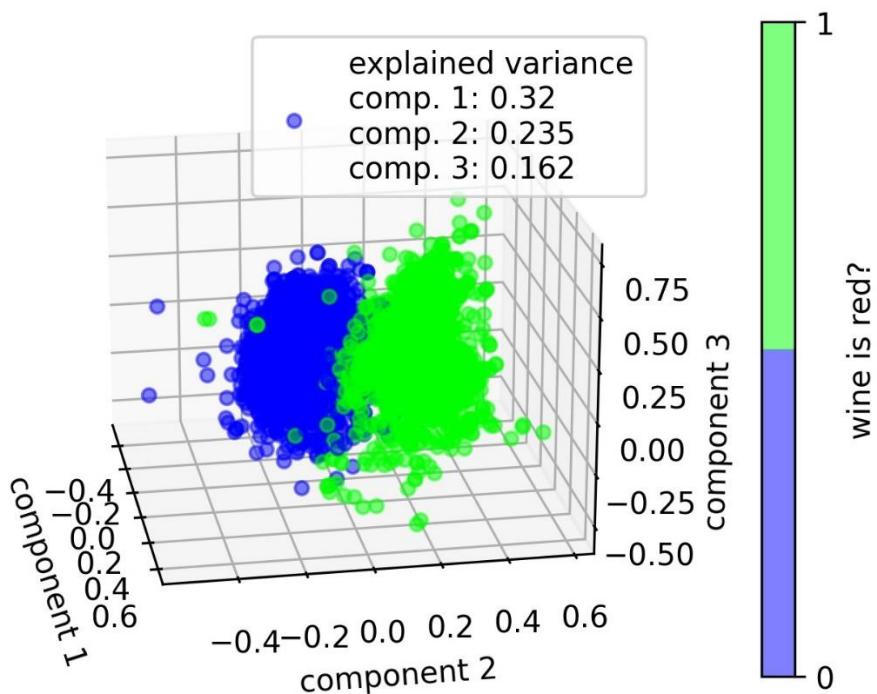
Chapter 10: Making Better Predictions – Optimizing Models

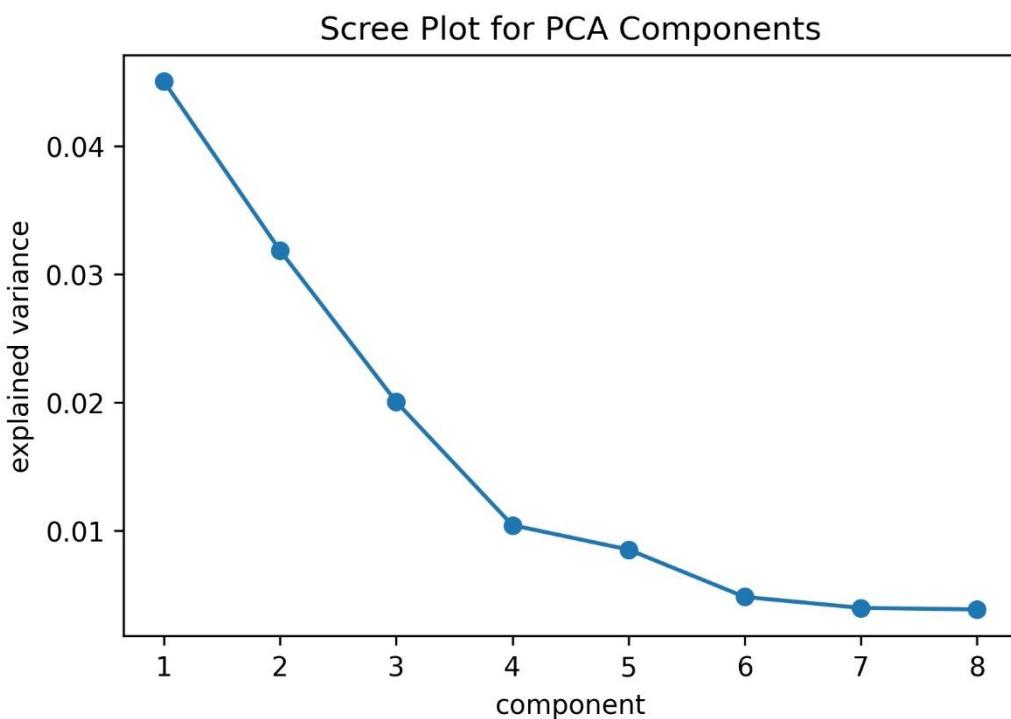
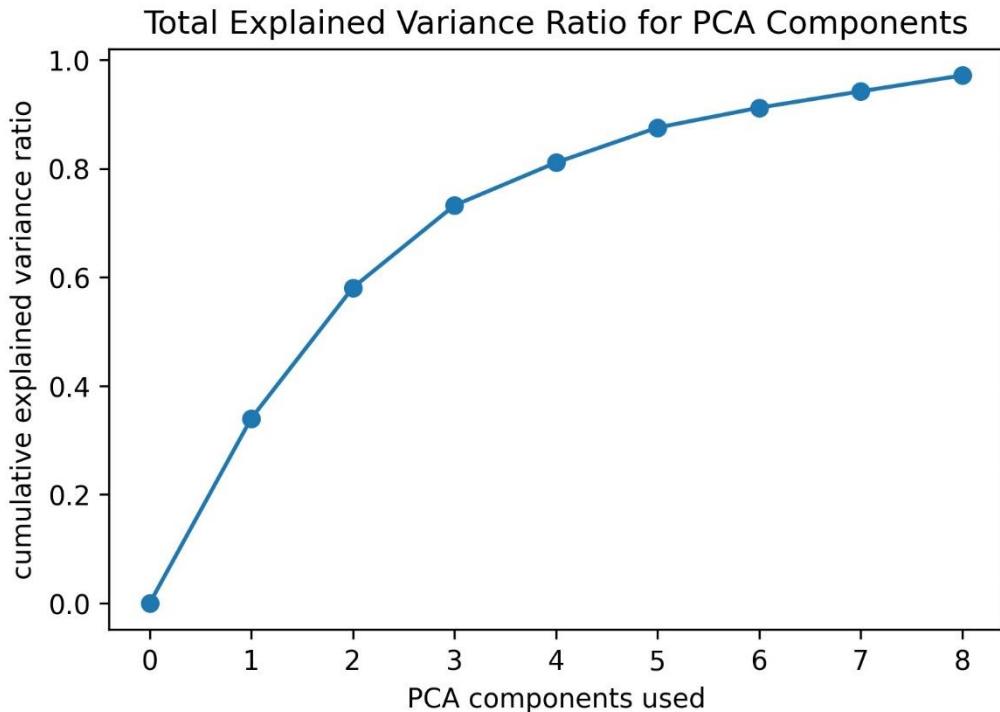


Wine Kind PCA (2 components)

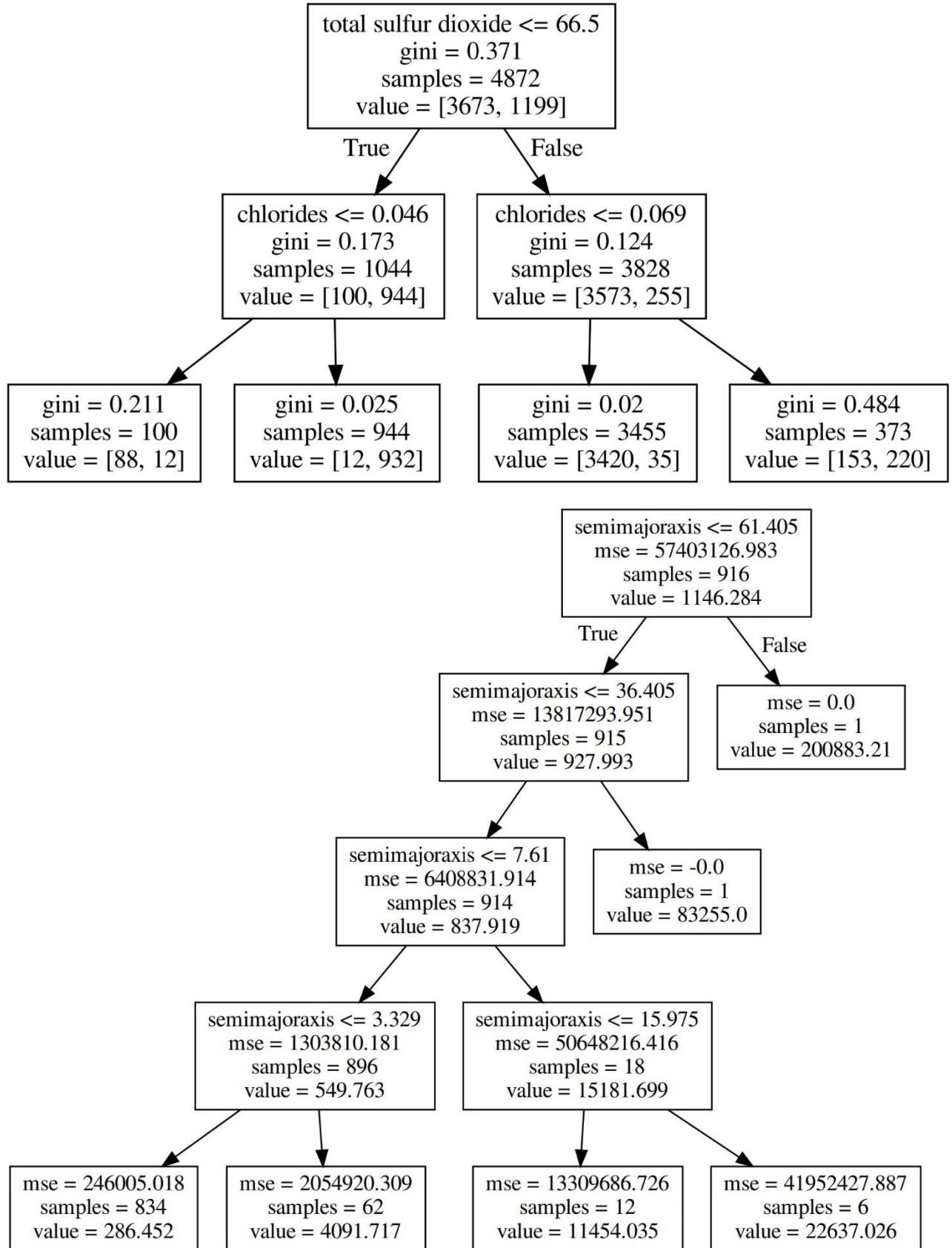


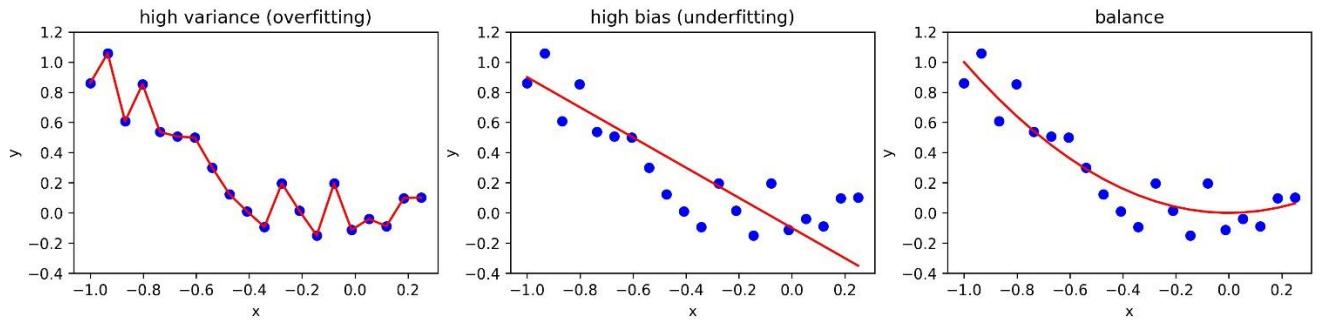
Wine Type PCA (3 components)



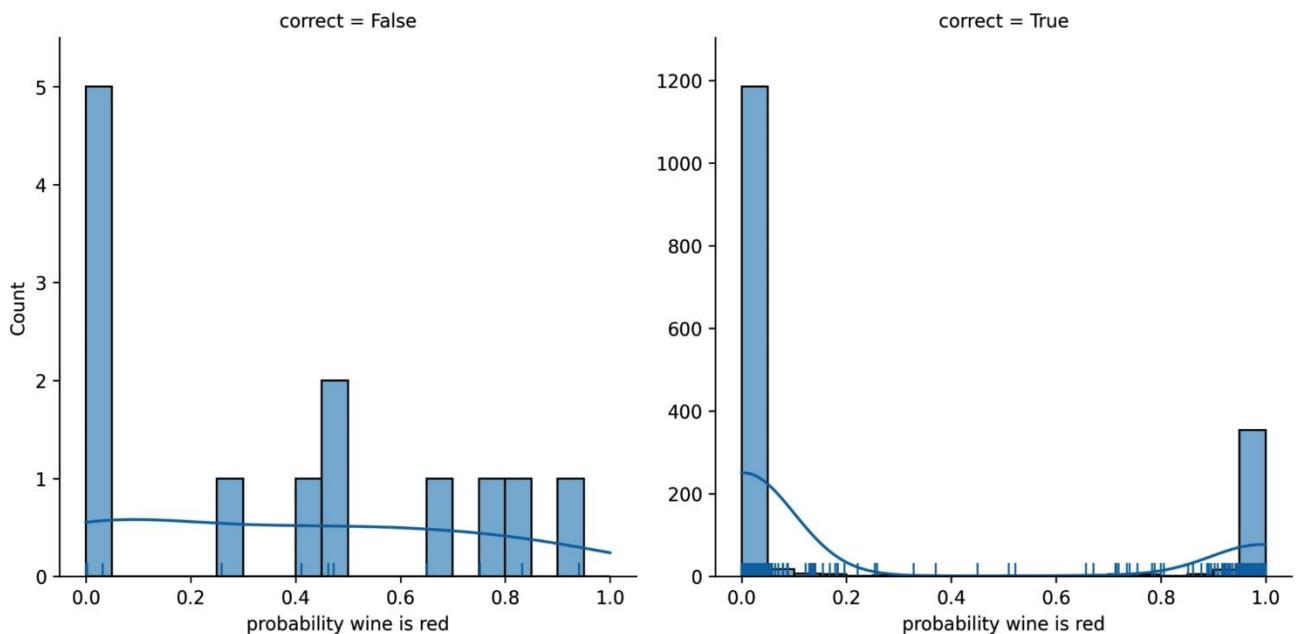


feature	total sulfur dioxide	chlorides	density	volatile acidity	sulphates	pH	residual sugar	alcohol	fixed acidity	citric acid	free sulfur dioxide
importance	0.687236	0.210241	0.050201	0.016196	0.012143	0.01143	0.005513	0.005074	0.001811	0.000113	0.000042

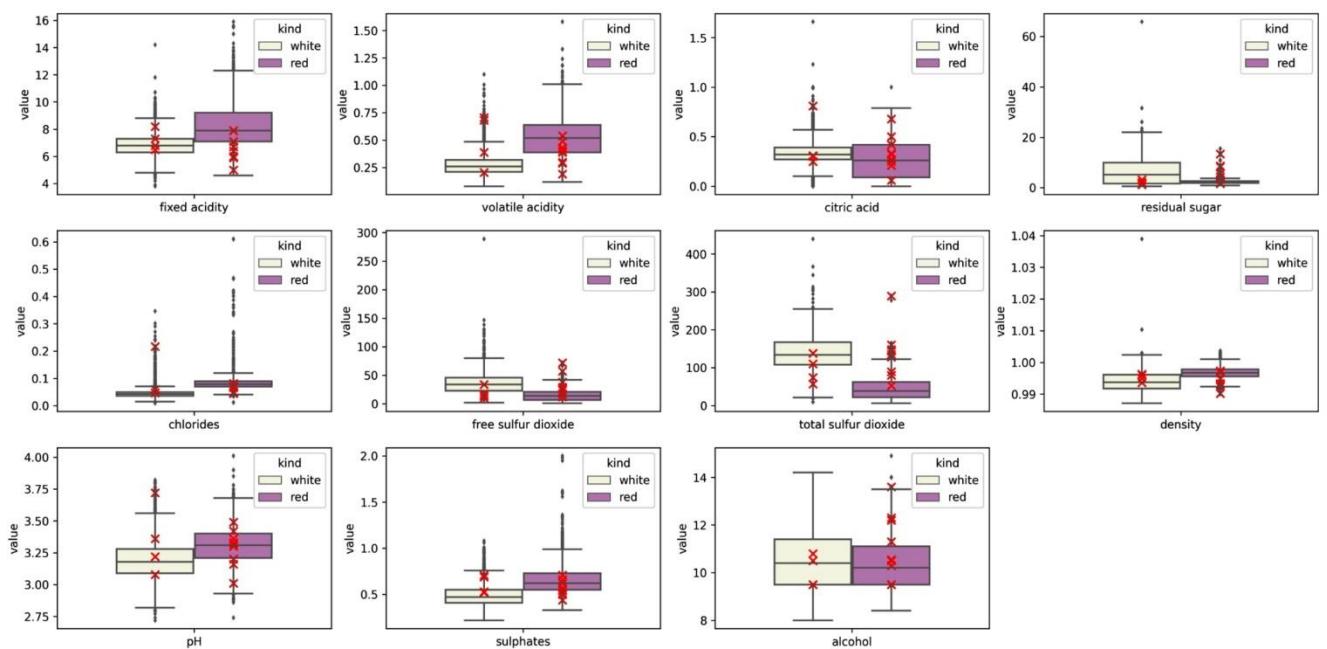




Prediction Confidence



Comparing Chemical Properties of Red and White Wines
(classification errors are red x's)

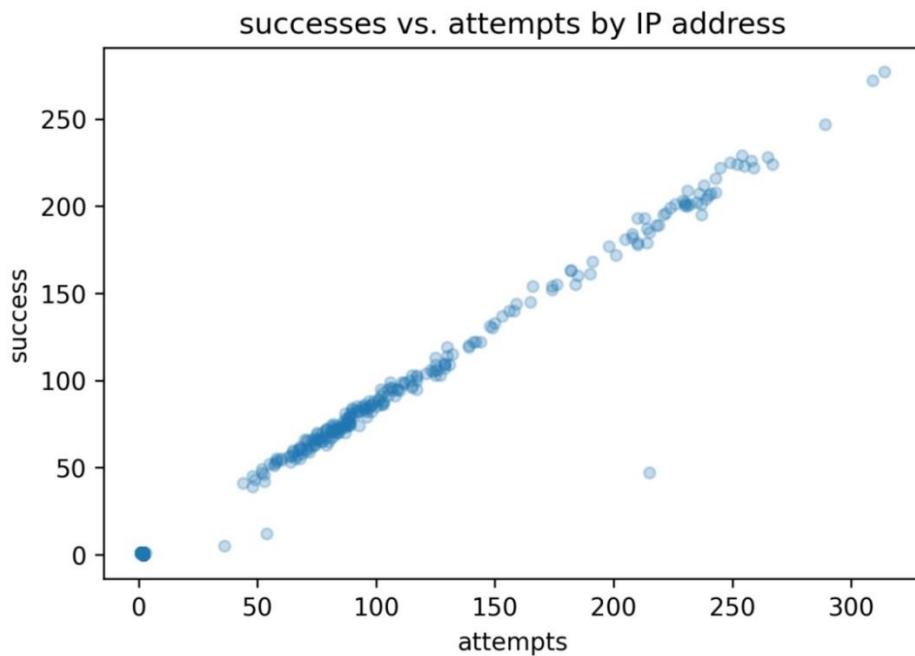


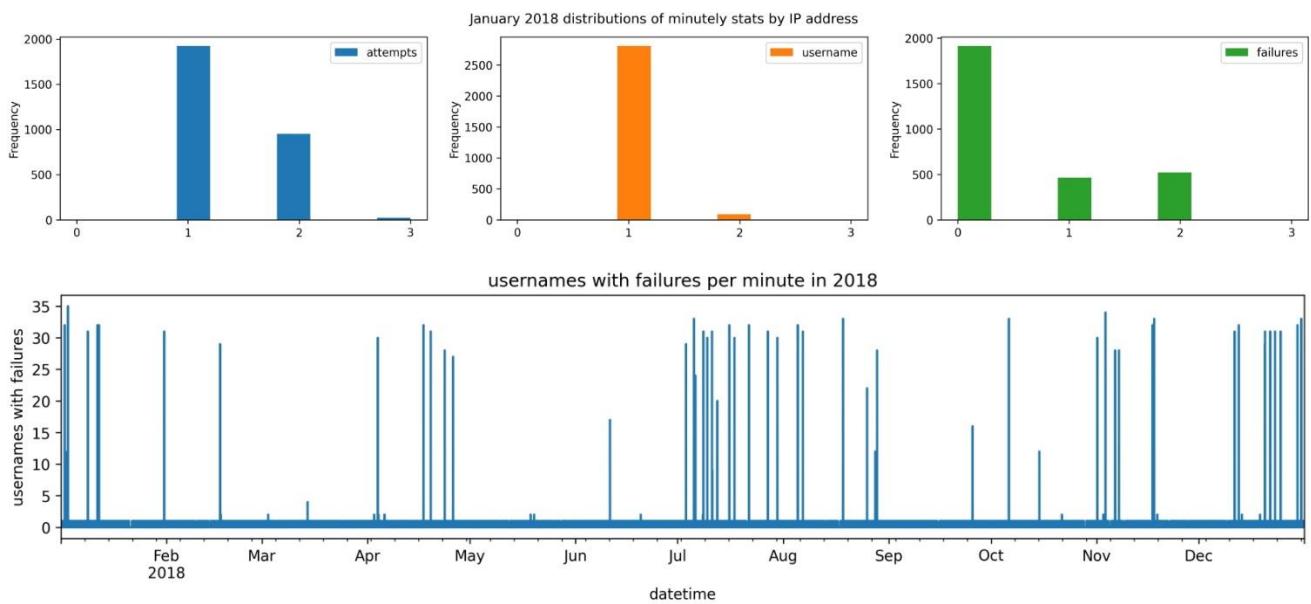
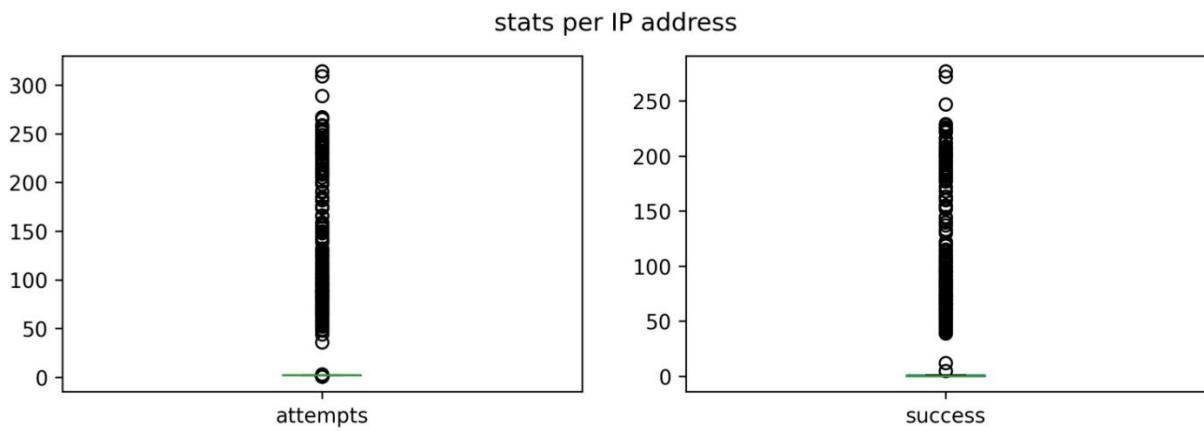
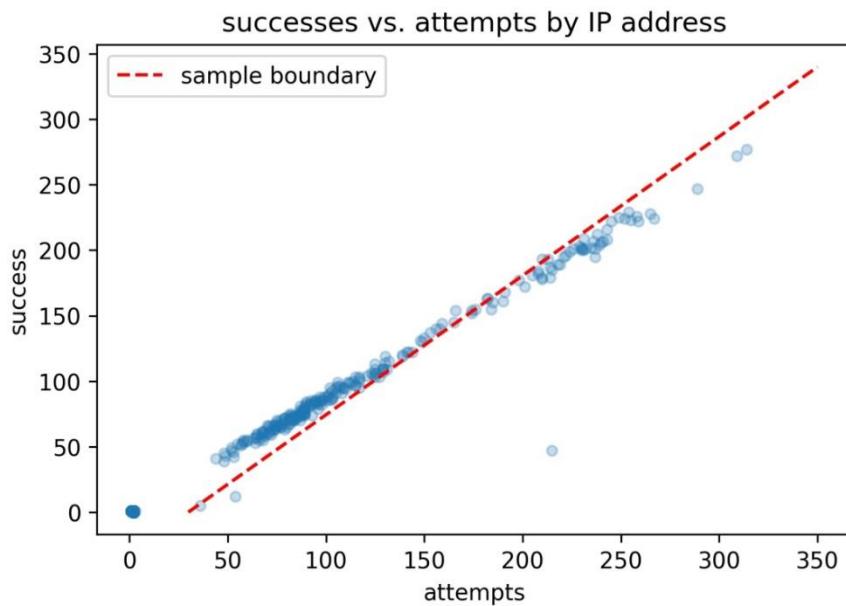
Chapter 11: Machine Learning Anomaly Detection

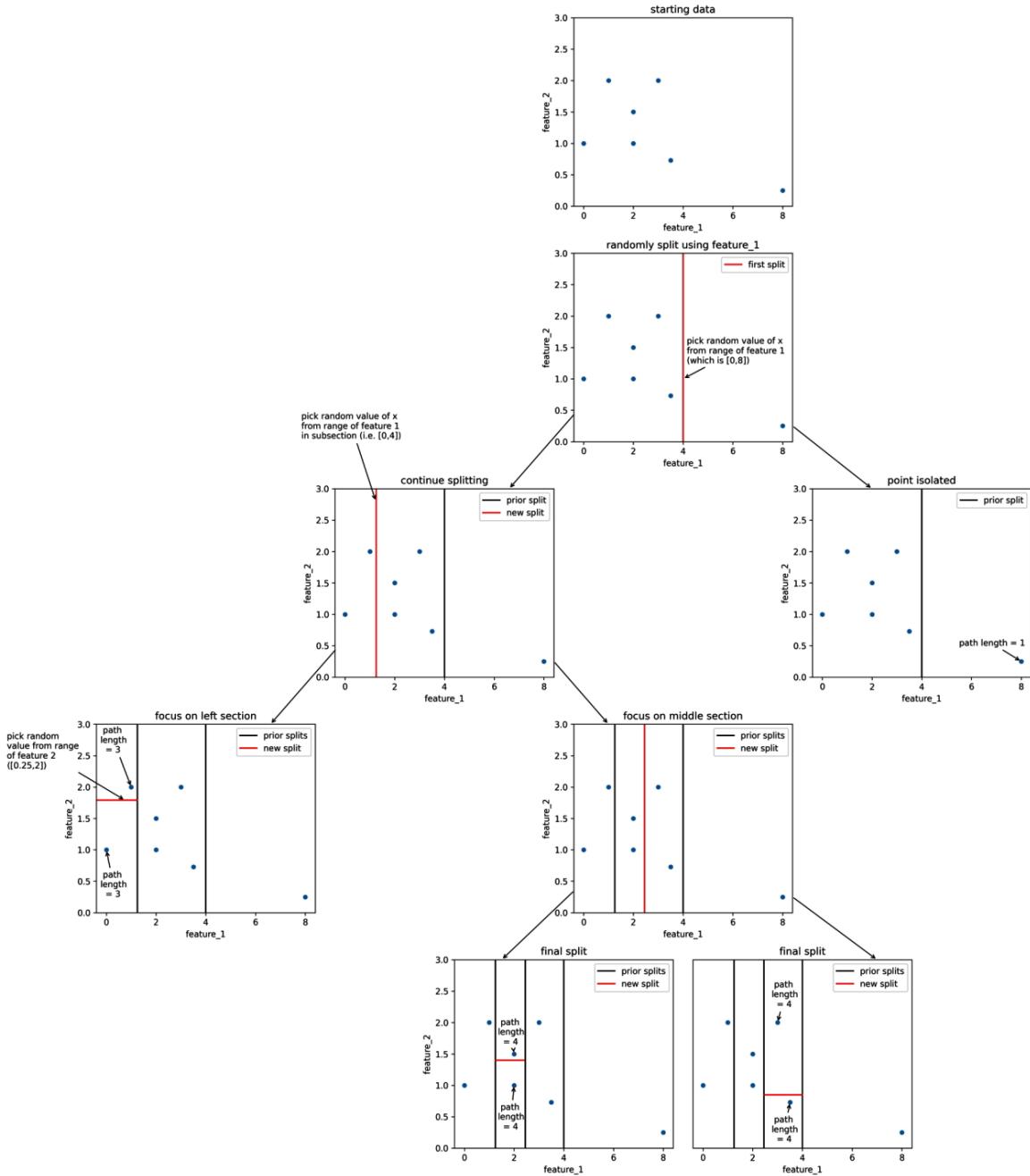
	Jan 2018	Feb 2018	Mar 2018	Apr 2018	May 2018	Jun 2018	Jul 2018	Aug 2018	Sep 2018	Oct 2018	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019
Probability of attack in a given hour	1.00%	0.50%	0.10%	1.00%	0.01%	0.05%	1.00%	0.50%	0.50%	0.20%	0.70%	1.00%	0.80%	0.20%	1.00%
Probability of trying entire user base	50%	25%	10%	65%	5%	5%	15%	10%	10%	12%	17%	88%	8%	18%	18%
Vary IP addresses?	Yes	No	No	Yes	Yes	Yes	Yes	Yes							

	source_ip	username	success	failure_reason
	datetime			
2018-01-01 00:05:32.988414	223.178.55.3	djones	1	None
2018-01-01 00:08:00.343636	223.178.55.3	djones	0	error_wrong_password
2018-01-01 00:08:01.343636	223.178.55.3	djones	1	None
2018-01-01 01:06:59.640823	208.101.11.88	wbrown	1	None
2018-01-01 02:40:47.769630	11.76.99.35	tkim	1	None

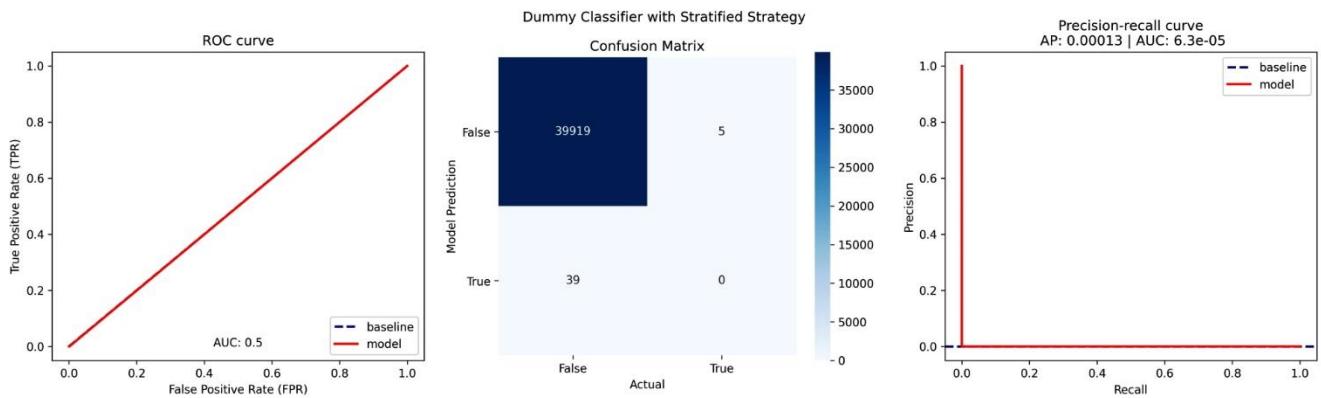
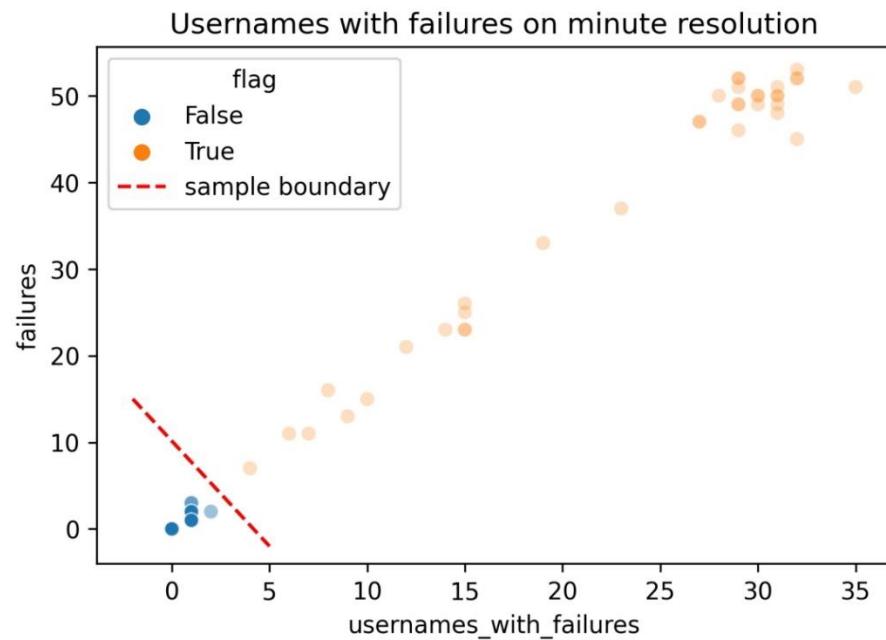
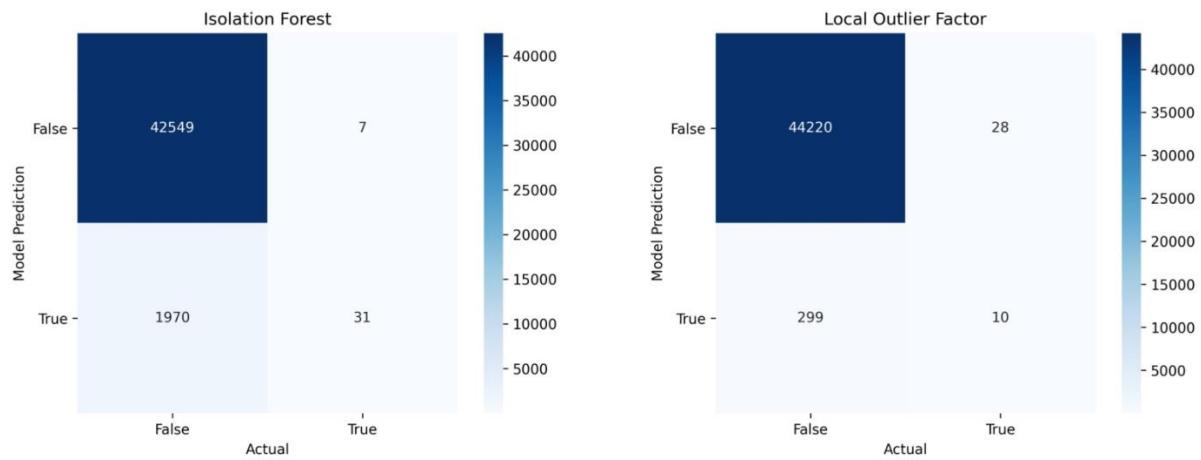
failure_reason	attempts	error_account_locked	error_wrong_password	error_wrong_username	success	success_rate	error_rate
source_ip							
168.123.156.81	314	0	37		0	277	0.882166
24.112.17.125	309	0	37		0	272	0.880259
16.118.156.50	289	0	41		1	247	0.854671
25.246.225.197	267	0	43		0	224	0.838951
30.67.241.95	265	0	37		0	228	0.860377

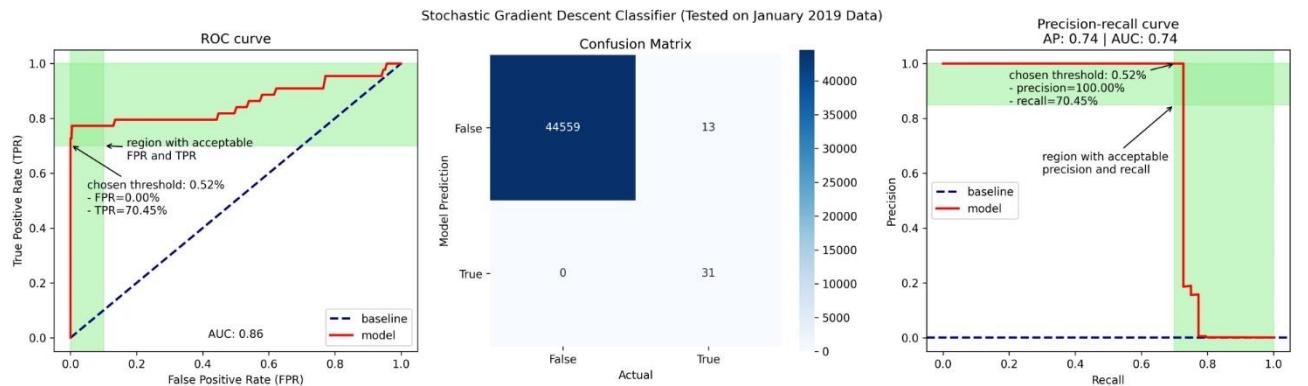
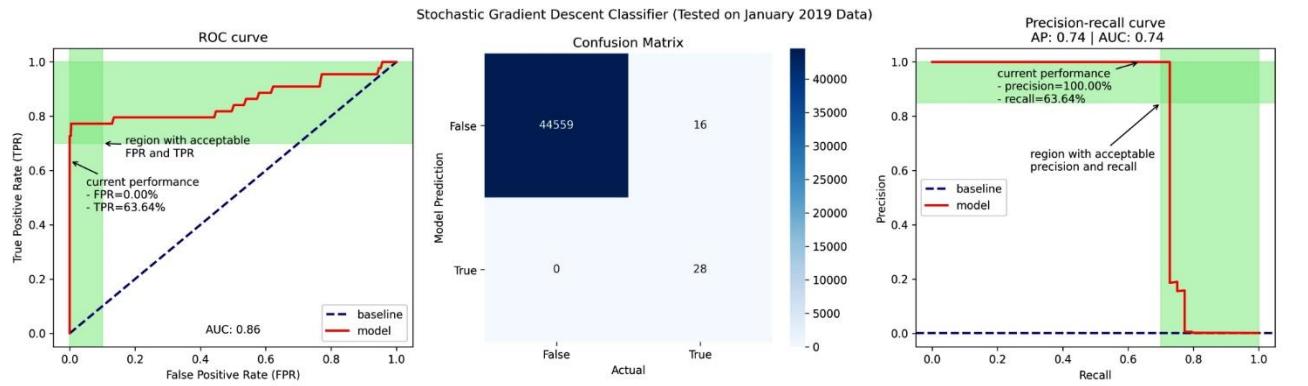
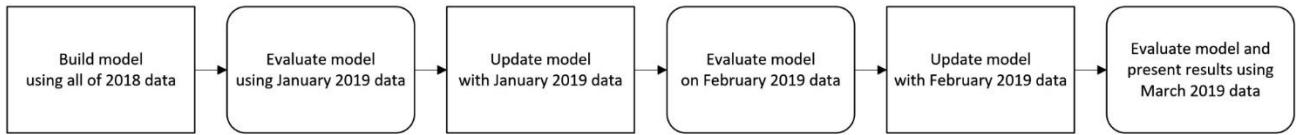
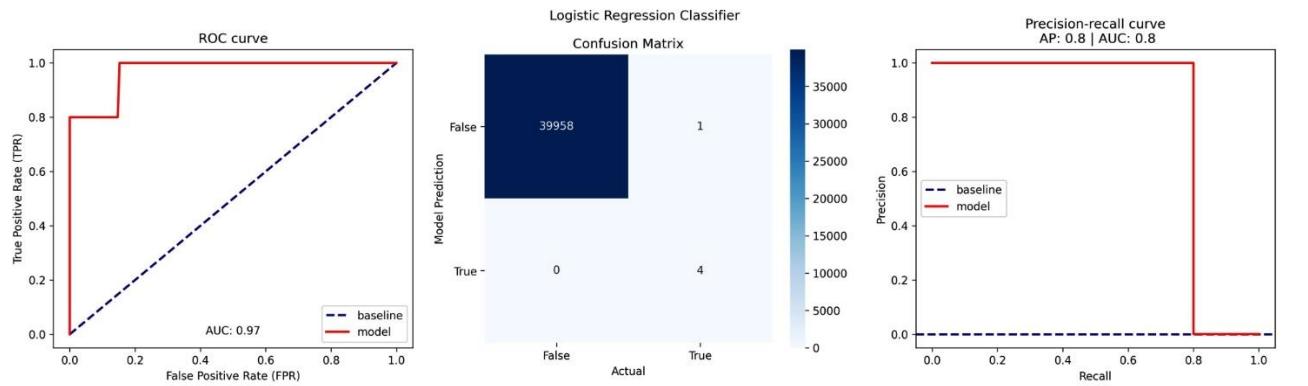
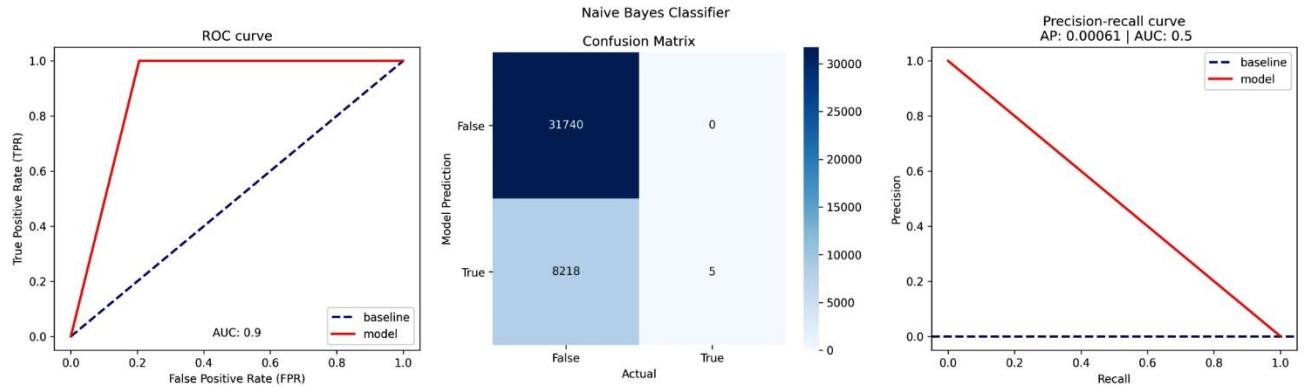


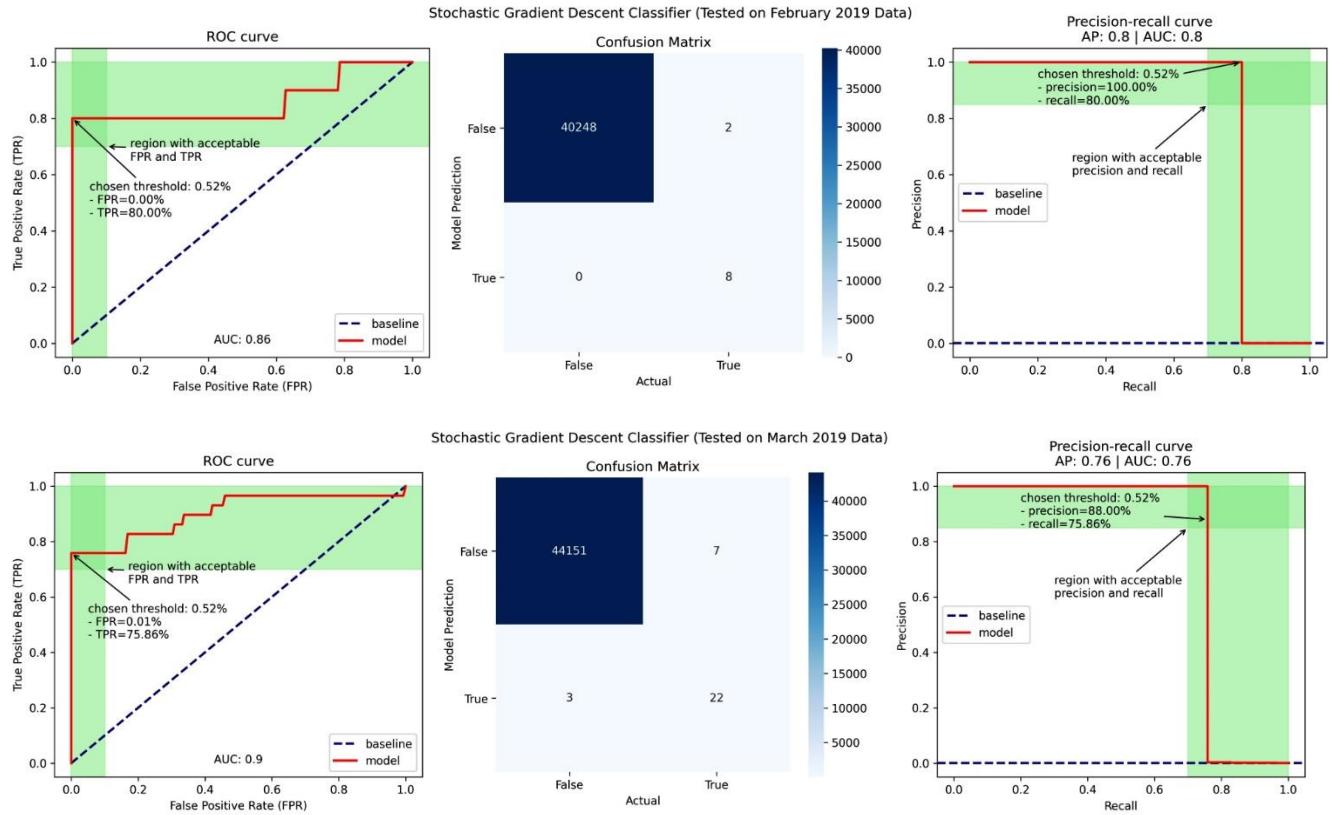




	start	end	source_ip	duration	start_floor	end.ceil
0	2018-01-02 02:31:43.326264	2018-01-02 02:35:16.326264	102.139.159.128	0 days 00:03:33	2018-01-02 02:31:00	2018-01-02 02:36:00
1	2018-01-02 20:14:02.279476	2018-01-02 20:14:28.279476	119.218.239.234	0 days 00:00:26	2018-01-02 20:14:00	2018-01-02 20:15:00
2	2018-01-03 01:25:48.667114	2018-01-03 01:29:13.667114	151.93.164.203	0 days 00:03:25	2018-01-03 01:25:00	2018-01-03 01:30:00
3	2018-01-08 21:41:43.985324	2018-01-08 21:45:56.985324	226.98.192.152	0 days 00:04:13	2018-01-08 21:41:00	2018-01-08 21:46:00
4	2018-01-11 17:38:30.974748	2018-01-11 17:42:33.974748	23.81.78.129	0 days 00:04:03	2018-01-11 17:38:00	2018-01-11 17:43:00
5	2018-01-12 03:32:20.284167	2018-01-12 03:36:29.284167	74.90.28.4	0 days 00:04:09	2018-01-12 03:32:00	2018-01-12 03:37:00
6	2018-01-31 07:39:17.514901	2018-01-31 07:43:29.514901	236.174.156.247	0 days 00:04:12	2018-01-31 07:39:00	2018-01-31 07:44:00







Chapter 12: The Road Ahead

