PATTERN ASSIGNMENT 1 AND 2

Submitted by: Khushi Gupta 500086849
Saif Anjum Khan 500086391

S. No.	Name	Type	Description
0	aPr	String	Ask Price that seller willing to take for the scrip
1	bPr	String	Bid Price that buyer willing to pay for the scrip
2	aSz	String	Ask size/quantity for trading
3	bSz	String	Bid size/quantity for trading
4	sym	String	Actual symbol name of the scrip
5	avgPr	String	Average trading price of the equity or derivative
6	C	String	Close value of market snapshot
7	h	String	High value of market snapshot
8	l	String	Low value of market snapshot
9	o	String	Opening price of a market snapshot
10	ol	String	Open interest is the total number of outstanding derivative contracts that have not been settled
11	olChg	String	open interest changed value
12	ch	String	Change value is the difference between the current value and the previous day's market close
13	chPer	String	Percentage of change between the current value and the previous day's market close
14	lTrdT	String	Time of the last transaction
15	ltp	String	Price at which last transaction / trade is done
16	ltq	String	Quantity of last transaction
17	ltt	String	Last transaction time in milliseconds
18	lttUTC	String	Last transaction time in UTC time zone format
19	tBQ	String	Total quantity of BUY transaction
20	tSQ	String	Total quantity of SELL transaction
21	ttv	String	Total volume of trading done
22	vol	String	Total amount of a security traded Today
23	yH	String	52 week high
24	yL	String	52 week low

Item 2 ->

aPr - 31681.95

PATTERN ASSIGNMENT 1 AND 2

Submitted by: Khushi Gupta 500086849
Saif Anjum Khan 500086391

aSZ - 50
avgPr - 31842.25
bPr - 31562.30
bSz - 50
c – 31873.20
ch - -322.00
chPer - -1.01
h - 32133.30
l - 31551.20
lTrdT - 23 Apr 2021, 09:15:00 AM
ltp - 31551.20
ltq - 25
ltt - 23 Apr 2021, 09:15:00 AM
1ttUTC - 23 Apr 2021, 03:45:00 AM
o -321 33530
oI - 193850
oIChg - -175.00
sym - 61195_NFO
tBQ - 4450
tSQ - 4725
ttv – 1592112.50
vol - 50
yH - 32133.30
yL - 0.00

EXPLAINATION:

This is a sample item extracted from crawling the data. This contains attributes as mentioned above and by exploring further, we concluded that there are only 4-5 attributes whose value is changing, other's are same.

PATTERN ASSIGNMENT 1 AND 2

Submitted by: Khushi Gupta 500086849
Saif Anjum Khan 500086391

(' ', 'start_array', None)
('item', 'start_map', None)
('item', 'map_key', '_id')
('item._id', 'start_map', None)
('item._id'', 'map_key', '$oid')
('item._id.$oid', 'string', '608242bd22ec5cd64382c299')
('item._id', 'end_map', None)
('item', 'map_key', 'aPr')
('item.aPr', 'start_array', None)
('item.aPr.item', 'string', '0.00')
('item.aPr', 'end_array', None)
('item', 'map_key', 'aSz')
('item.aSz', 'start_array', None)
('item.aSz.item',’string’, '@'')
('item.aSz', 'end_array', None)
('item', 'map_key', 'avgPr')
('item.avgPr', 'start_array', None)
('item.avgPr.item', 'string’, '31481.44')
('item.avgPr', 'end_array', None)
(‘item', 'map_key', 'bPr')
('item.bPr', 'start_array', None)
('item.bPr.item', 'string ', '0.00')
('item.bPr', 'end_array', None)
('item', 'map_key', 'bSz')
('item.bSz', 'start_array', None)

Explain:
It appears that the provided data is in JSON format, which consists of key-value pairs organized in a hierarchical structure. The example shows the start of parsing the JSON data, where each key-value pair is represented by a sequence of messages.

The first message is (' ', 'start_array', None), indicating the start of an array.

PATTERN ASSIGNMENT 1 AND 2

Submitted by: Khushi Gupta 500086849
Saif Anjum Khan 500086391

The second message ('item', 'start_map', None) indicates the start of a new object in the array, where the "item" is the key and "start_map" indicates that the value is a new object.
The third message ('item', 'map_key', '_id') indicates that the current key is "_id" within the current object.
The fourth message ('item._id', 'start_map', None) indicates the start of a new object within the "_id" key.
The fifth message ('item._id.$oid', 'string', '608242bd22ec5cd64382c299') indicates that the value of the "_id.$oid" key is the string '608242bd22ec5cd64382c299'.
The sixth message ('item._id', 'end_map', None) indicates the end of the "_id" object.
The seventh message ('item', 'map_key', 'aPr') indicates that the current key is "aPr" within the current object.
The eighth message ('item.aPr', 'start_array', None) indicates the start of an array within the "aPr" key.
The ninth message ('item.aPr.item', 'string', '0.00') indicates that the first item in the "aPr" array is the string '0.00'.
The tenth message ('item.aPr', 'end_array', None) indicates the end of the "aPr" array.
The eleventh message ('item', 'map_key', 'aSz') indicates that the current key is "aSz" within the current object.
The twelfth message ('item.aSz', 'start_array', None) indicates the start of an array within the "aSz" key.
The thirteenth message ('item.aSz.item', 'string', '@') indicates that the first item in the "aSz" array is the string '@'.
The fourteenth message ('item.aSz', 'end_array', None) indicates the end of the "aSz" array.
The fifteenth message ('item', 'map_key', 'avgPr') indicates that the current key is "avgPr" within the current object, and so on.

ASSIGNMENT-2

OUTLIER DETECTION

PATTERN ASSIGNMENT 1 AND 2

Submitted by: Khushi Gupta 500086849
Saif Anjum Khan 500086391

We used Local Outlier Factor and used hyperparameters n-neighbours as 20 and contamination as 0.005.

To calculate the LOF for a given data point, we first determine its k-nearest neighbors (i.e., the k data points closest to it). Then, we calculate the local reachability density (LRD) for the data point, which is a measure of its local density relative to its neighbors. The LRD is calculated as the inverse of the average reachability distance between the data point and its k-nearest neighbors.
Finally, we calculate the LOF for the data point as the ratio of its LRD to the LRDs of its k-nearest neighbors. A high LOF value indicates that the data point is an outlier, while a low LOF value indicates that it is an inlier.

We took 2 attributes (average price and volume) as features and ran different instances of LOF on them. Below given are the plots for the same.
After that we combined the results and classified the points where both instances detected it as an outlier then, it is a confirmed outlier. If it is detected by any one, then it is classified as maybe outlier and if none of them detected it as an outlier then it is classified as a normal point.

png image for the plots are attached in the github repository.

PATTERN ASSIGNMENT 1 AND 2

Submitted by: Khushi Gupta 500086849
Saif Anjum Khan 500086391