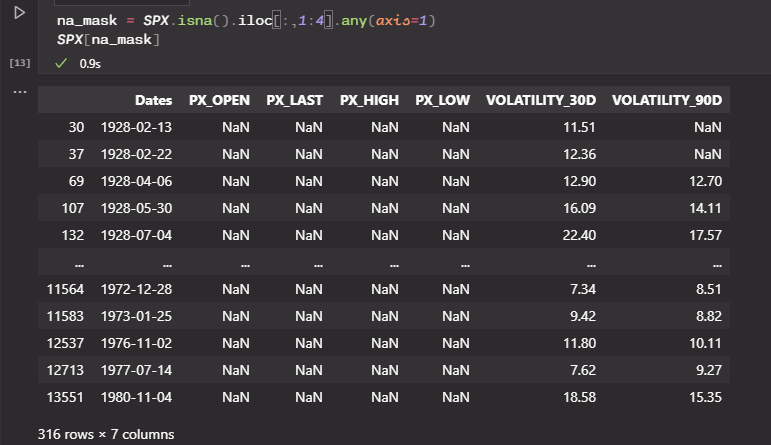
1

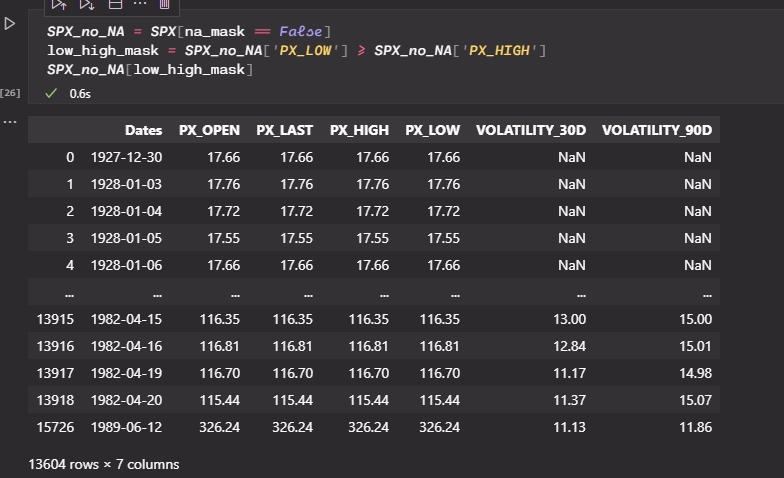
(a) My data range from 1927-12-30 to 2021-09-16. We will test the data completeness after analyzing data quality and integrity in (2)

(b) Integrity Check:

1) non-missing values for OPEN, CLOSE, HIGH, LOW

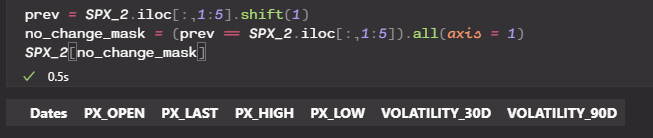


2) HIGH should be strictly larger than LOW, otherwise the market has same price all day

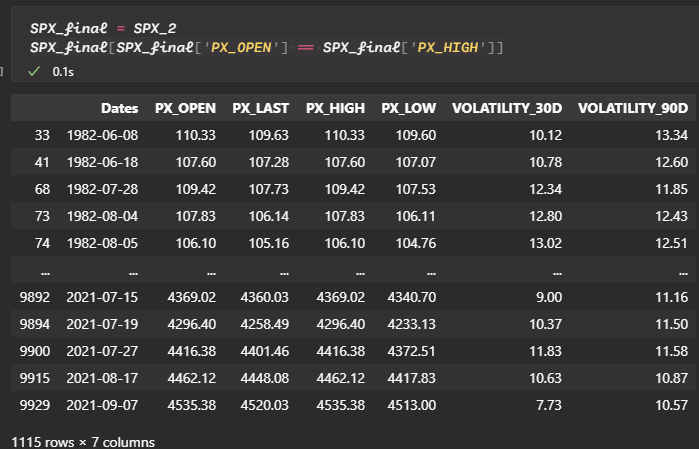


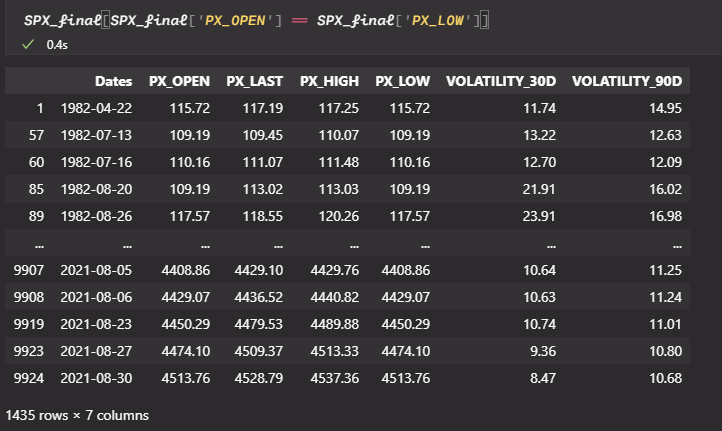
There are fairly a lot of data showing this issue. To finish the task in next question correctly we have to delete these data.

3) at least one observation of {OPEN, CLOSE, HIGH, LOW} should be different from the previous day



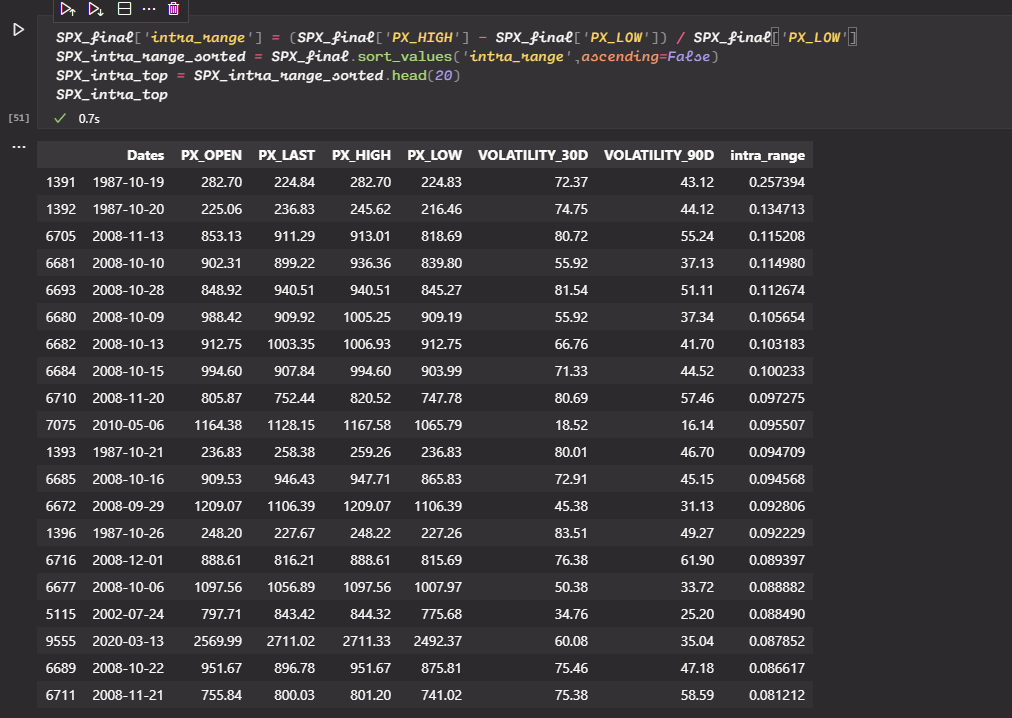
No such observations exist

c) 



These result can be used to reject random walk hypothesis: if random walk hypothesis holds, both probability should be zero

(d)



14 of the top 20 are in the sub-period from 9/1/2008 to 8/30/2011

(e)The largest three-year period is the recent three year from 2018Graphical user interface

Description automatically generated

(f) Only one of the top 20 jumps is in the three year periodGraphical user interface, text

Description automatically generated

(f) based on intra-day change, the voloatility in 2020 is not distinguishable from other years; however based on overnight changes, the volatility in 2020 is a lot larger than other years.

Chart, histogram

Description automatically generated

Chart

Description automatically generated

2.

From Bloomberg: the open price was the lowest price and the close price was the highest price



From Yahoo finance: although OPEN == LOW and CLOSE == HIGH still holds, the close price was off by 1 compared with Bloomberg data

Table

Description automatically generated

The difference of 1 was roughly 1%, and if there was $1 billion dollar worth of SP500 ETF, that would result in $100 million discrepancy, not mentioning the indirect effect on portfolio benchmarking. It could be resolved by calculating the weighted close prices of SP500 components.

3.

(a) As of Aug 24 2020, the value of new basket of equities were $4587.78, and given DJIA index value of $28308.46, the new divisor will be the ratio of the two which would be 0.1621

(b) 3.46%

(c) Their weight decrease from 96.54% to 86.85%

(d) The divisors would be 0.2706 and 11.41539 respectively

(e) It is observable that AAPL took too much weight as of Aug 24 2020: because the DJIA takes one share of the 30 component companies and AAPL had the highest price per share. By 4-for-1 split and decrease the price per share, it created a more “balanced” weighting for DJIA.