

تمرین شماره ۱۴ | پروژه شماره ۲ دیتاساینس

گزارش کار

موارد خواسته شده:

1. دیتای NYA
 2. تبدیل اعداد FLOAT به INTEGER
 3. آنالیز خطی Date و Adj Close
- فایل کدنویسی نیز ضمیمه شده است.

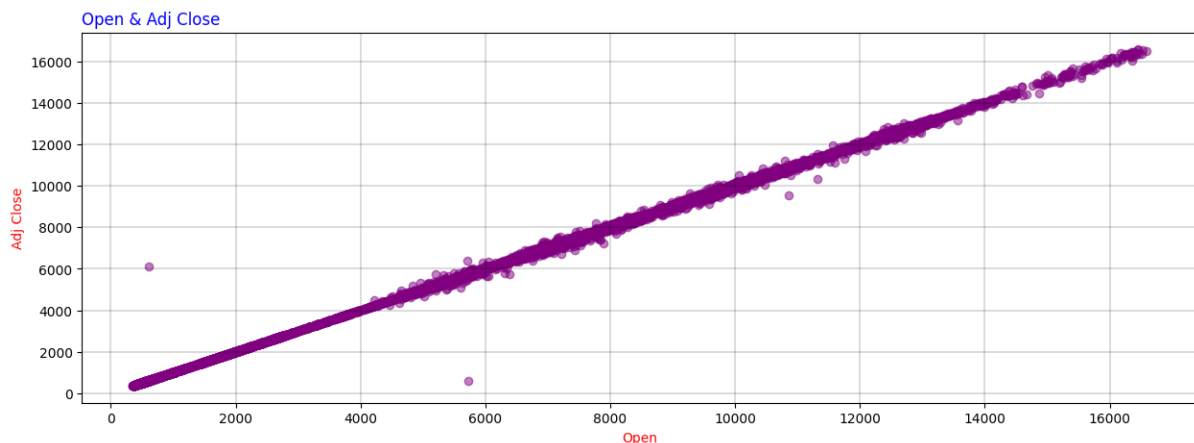
قسمت اول:

در قسمت اول دیتاست را ایمپورت کرده و قسمت مورد نظر - داده‌های NYA - را انتخاب می‌کنیم. در حال حاضر ما یک دیتای ۱۳۹۴۸ در ۸ داریم. این دیتا شامل تاریخ، قیمت آغازین، قیمت پایانی، بالاترین قیمت، پائین ترین قیمت و قیمت میانگین می‌باشد. تارگت ما در اینجا Adj Close می‌باشد.

این دیتاست دارای ۲۱ مورد میسینگ ولیو می‌باشد. این در مقابل دیتای ما مقدار کمی می‌باشد ولی قبل از انجام هر گونه عملی بر روی میسینگ ولیوها ابتدا به سراغ نویزها می‌رویم.

قسمت دوم:

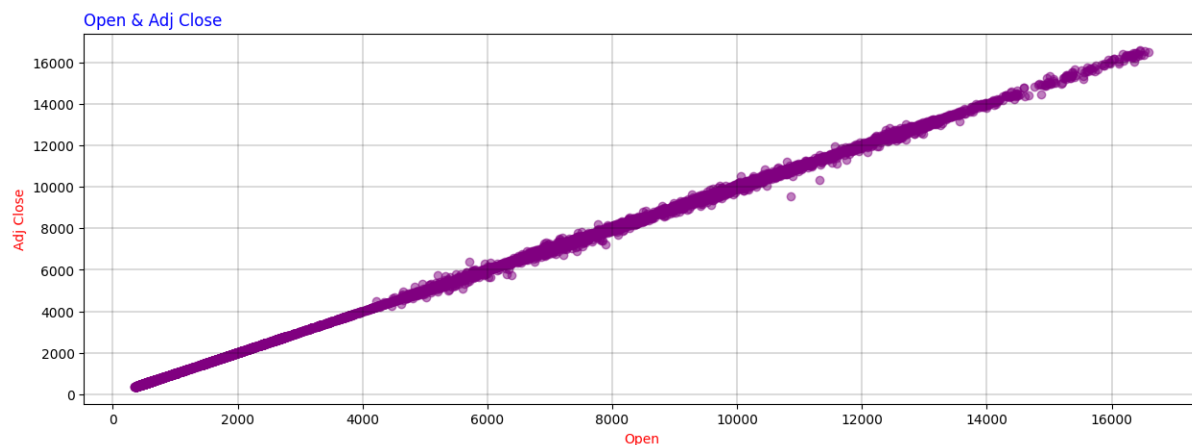
اولین نمودار، نمودار Open و Adj Close می‌باشد.



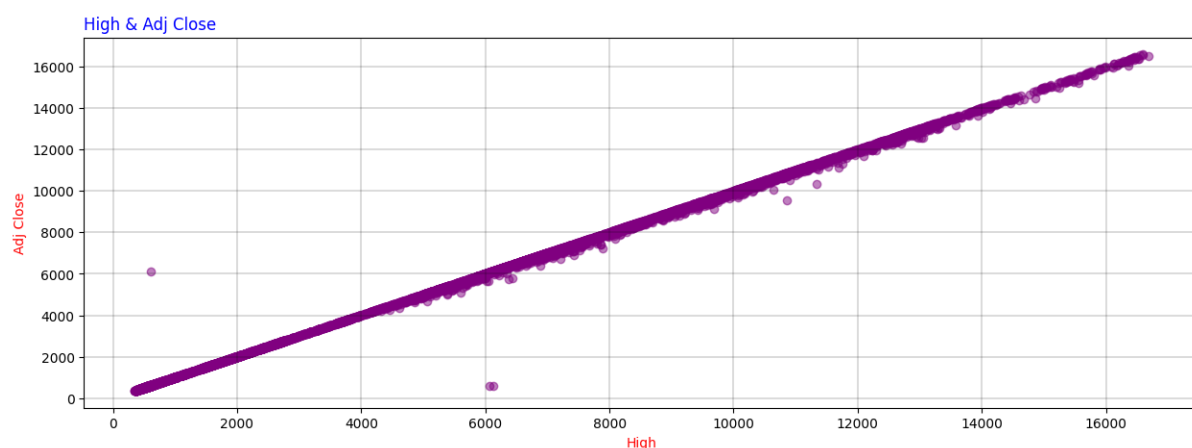
این نمودار یک خط صعودی است. دو دیتا در سمت چپ نمودار نویز محسوب می‌شوند. با دستورات ایندکس را پیدا کرده و حذف می‌کنیم. (در مورد نویز و میسینگ ولیوها می‌توان با

استفاده از روند قابل پیشبینی دیتا، با میانگین گیری از دیتای قبل و بعد آن‌ها را پر کرد ولی با توجه به تعداد کم آنها در این دیتاست تصمیم به حذف مقادیر موجود گرفتم)

ایندکس‌های ۸۵۲ و ۸۳۱ حذف گردید.

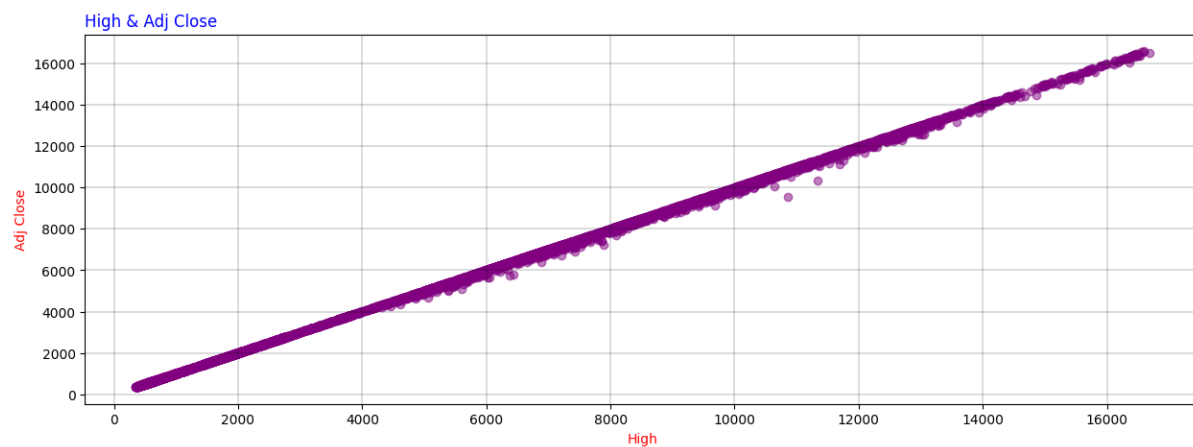


نمودار بعدی نمودار High and Adj Close می‌باشد.

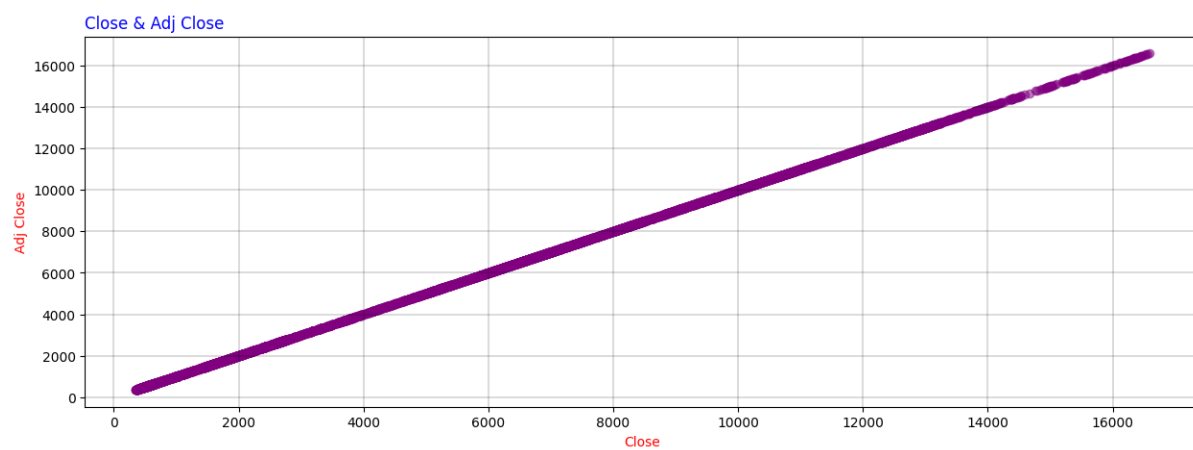
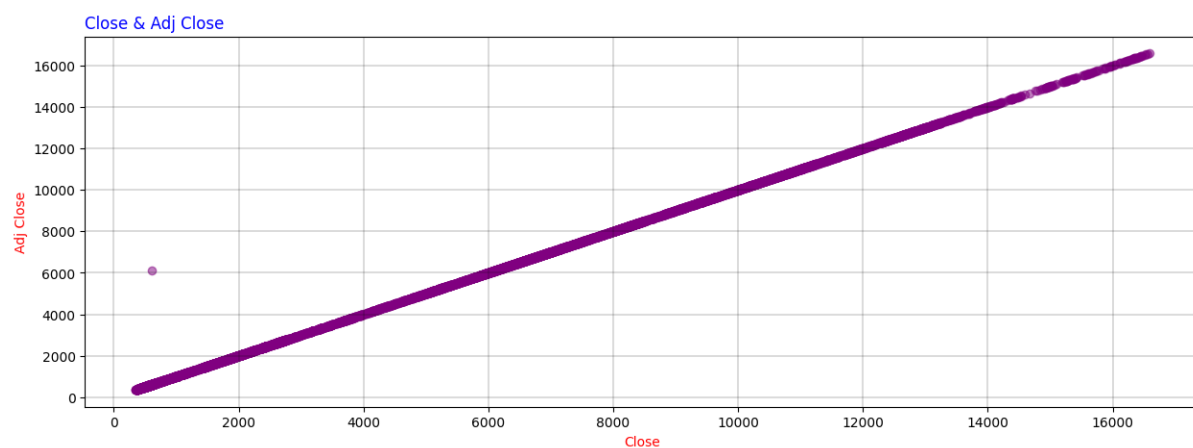


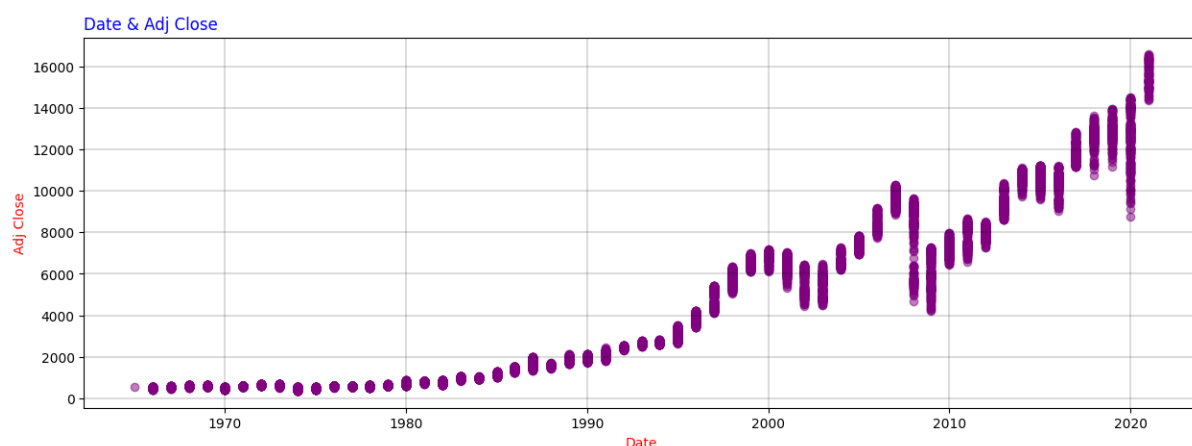
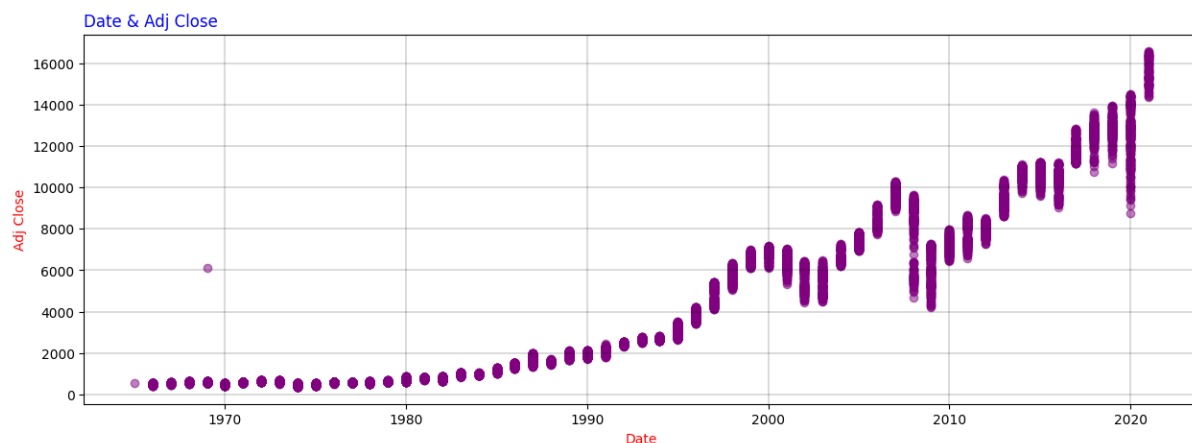
در این نمودار ۳ نویز دیده می‌شود که یکی از آنها در ایندکس ۸۳۱ هندل کرده‌ایم. این نویزها را نیز مکان‌یابی کرده و حذف می‌کنیم.

ایندکس‌های ۸۲۹ و ۸۳۳ نیز حذف می‌شوند.



نمودارهای بعدی نمودارهای Close and Date & Adj Close می‌باشند که نویز آنها قبلاً همدل شده است.





در این قسمت دیتای Date را بصورت زمانی درآورده‌ایم.

قسمت سوم:

در این قسمت به میسینگ ولیوها باز می‌گردیم. تعداد میسینگ ولیوها همچنان ۲۱ مورد است که نشان دهنده این است که دیتای نویز دارای میسینگ ولیو نبوده است. چون تعداد میسینگ ولیوها زیاد نیست، آنها را حذف می‌کنیم. قبلاً هم گفته شد که می‌توانستیم دیتاها را مثلاً با میانگین‌گیری جایگزین کنیم، چون روند کلی دیتاها صعودیست.

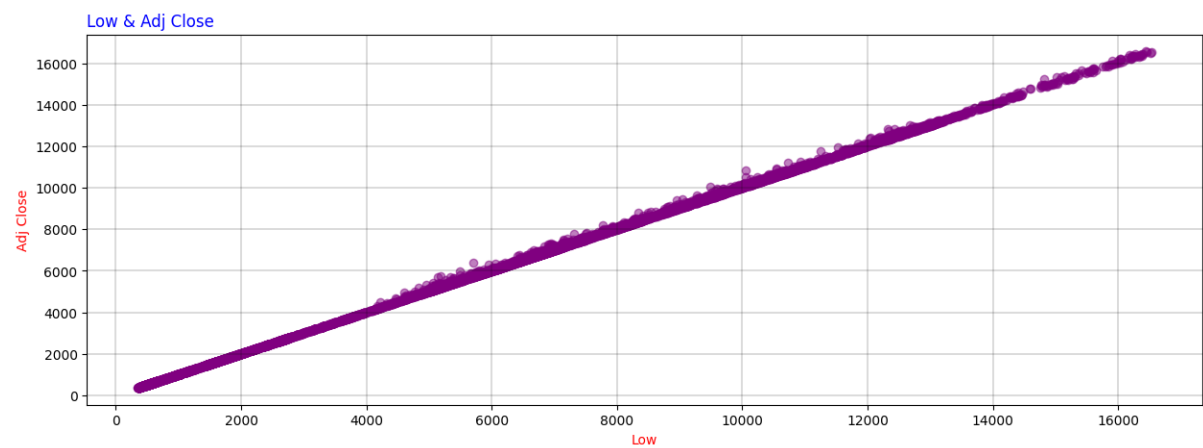
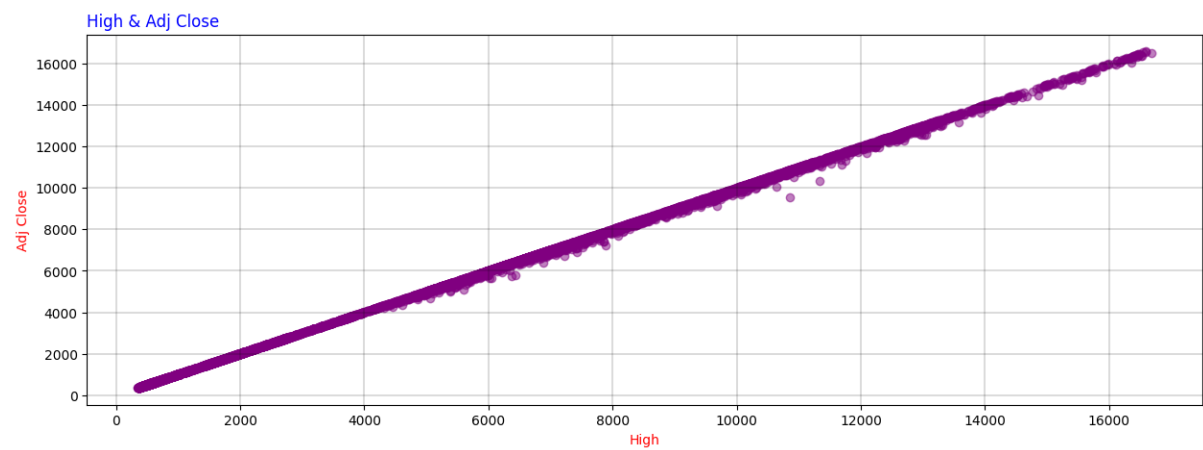
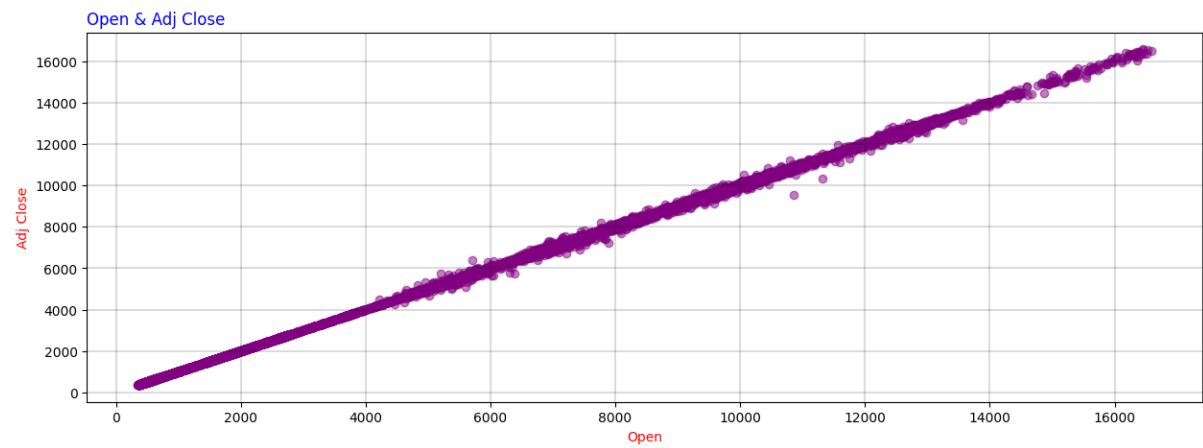
حالا دیتای ما تمیز است که شامل یک دیتاست ۱۹۲۸ در ۸ می‌باشد. حدوداً ۲۰ دیتا حذف شده است.

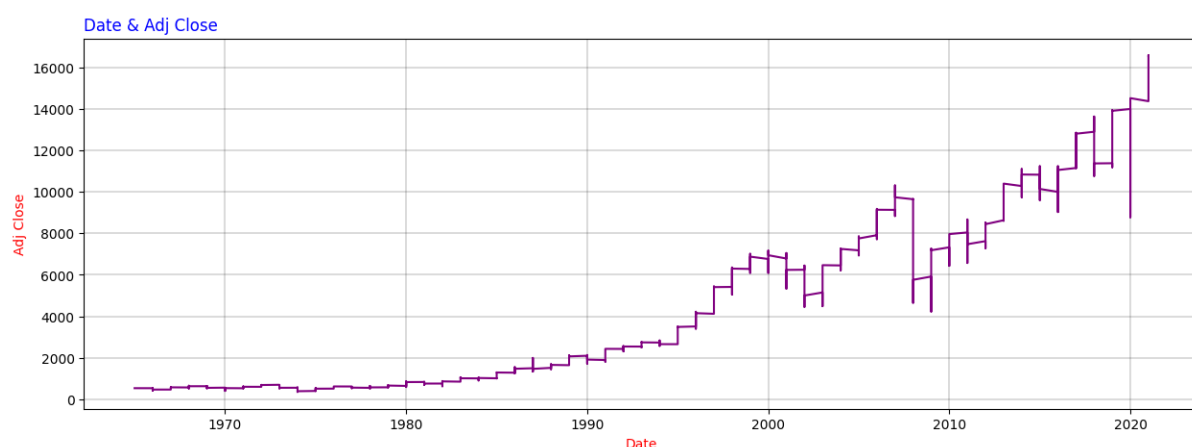
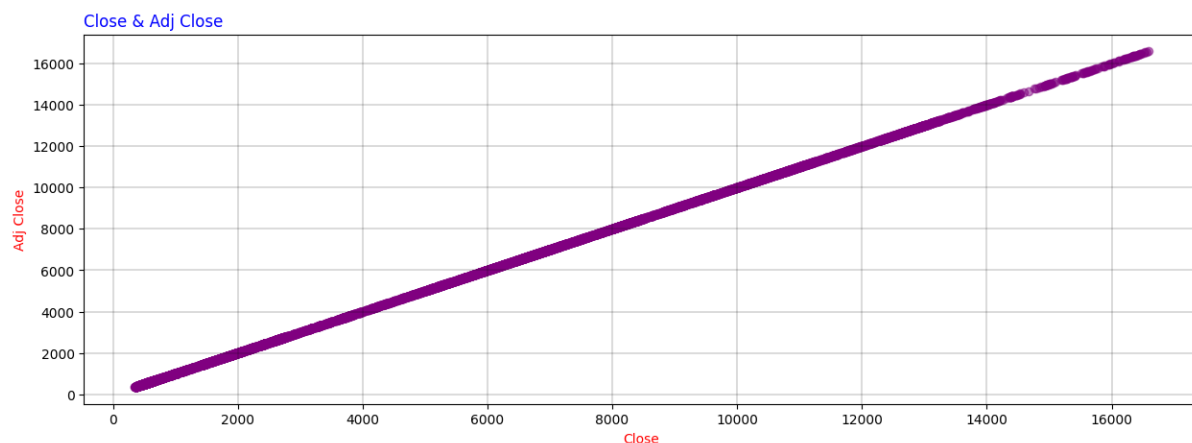
قسمت چهارم:

در این قسمت اعداد float را به اعداد integer تبدیل کردیم.

قسمت ۵:

در این قسمت نمودارهای پایانی کدنویسی شد و نمودار خطی Date & Adj Close تحلیل شد.





1. نمودار حالت صعودی دارد. و با گذشت زمان مقدار Adj Close هم افزایش داشته است.
2. شیب صعودی نمودار از ابتدا تا ۱۹۹۰ کمتر از ۱۹۹۰ به بعد است. و بعد از ۱۹۹۰ شیب نمودار بیشتر شده است.
3. نزول نمودار بعد از سال ۲۰۰۰ بیشتر شده است.
4. بیشترین نزول نمودار در سالهای ۲۰۰۸ و ۲۰۲۰ بوده است که به ترتیب با بحران مالی ۲۰۰۸ و کووید ۱۹ در ارتباط است.
5. نمودار بعد از سالهای ۲۰۰۸ و ۲۰۲۰ از بیشترین مقدار قبل از نزول نیز بیشتر شده است.
6. می‌توان گفت که حتی با وجود نزول در نمودار همیشه بازگشت نمودار به بالاتر امکان دارد.
7. سیر صعودی نمودار با تورم نیز در ارتباط است (پول در گردش بیشتر تورم زاست و با افزایش پول میل به سرمایه‌گذاری نیز بیشتر می‌شود)