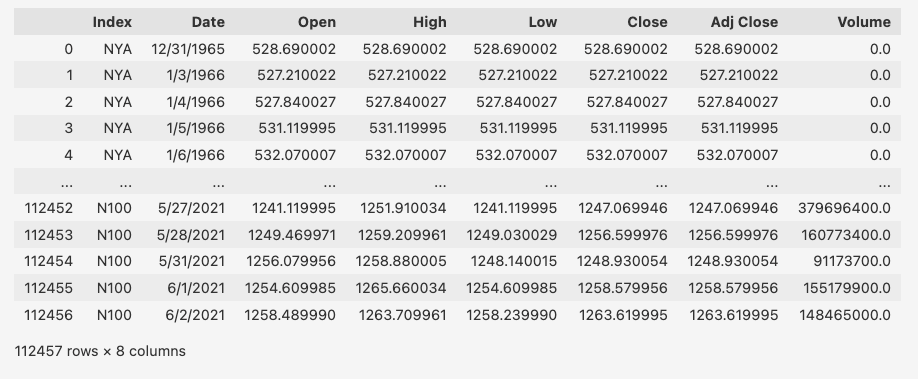
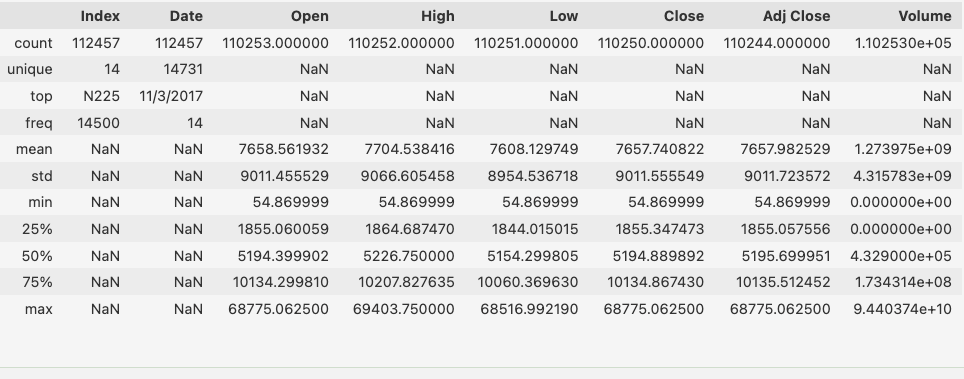
بررسی دیتاست بازار مالی

Milad Hashemi

داده هایی به شرح زیر۱۱۲۴۵۷× ۸دریافت شده است:



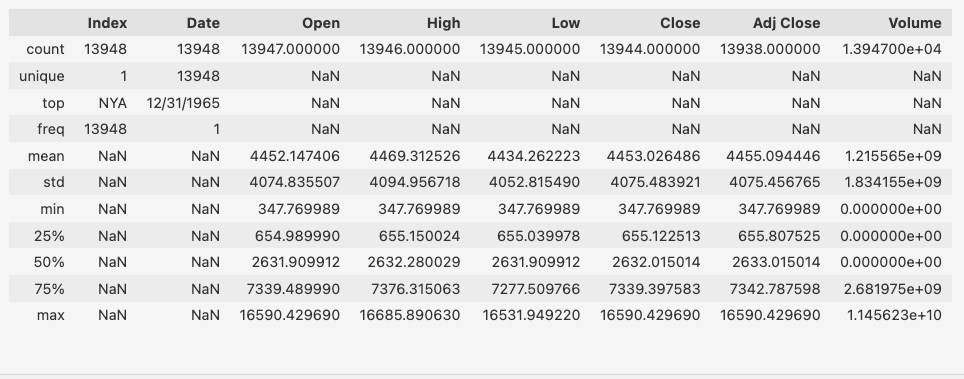
بعد از بررسی , اطلاعات اولیه به شرح زیر از داده دریافت شده است:



از انجایی که کارفرما فقط ایندکس NYA رو مدنظر داشت برای همین این ایندکس از کل دیتاست جدا شد:



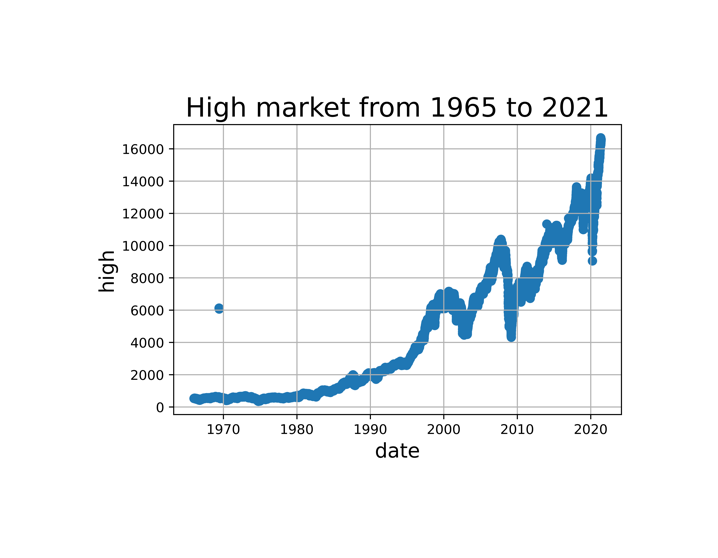
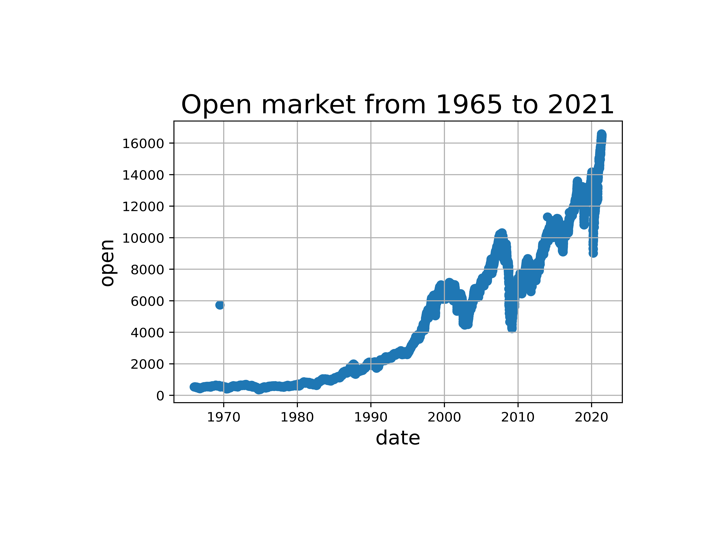
بعد از بررسی , اطلاعات اولیه به شرح زیر از داده دریافت شده است که missing value از یک تا ده تا در این دیتاست وجود دارد:

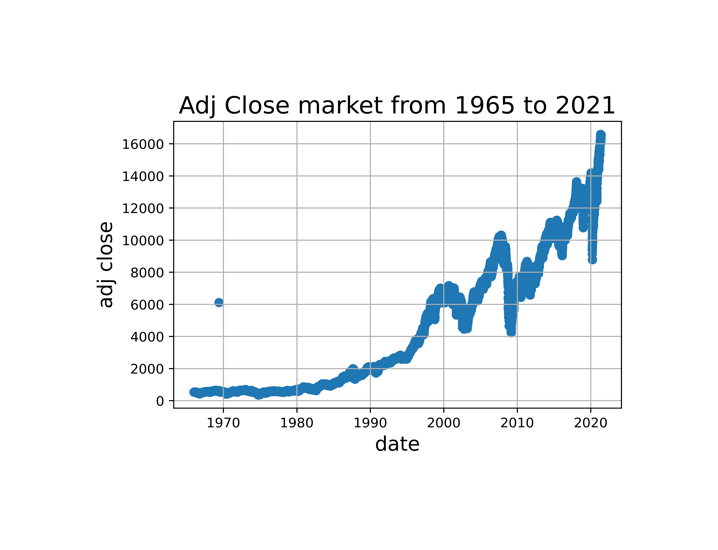
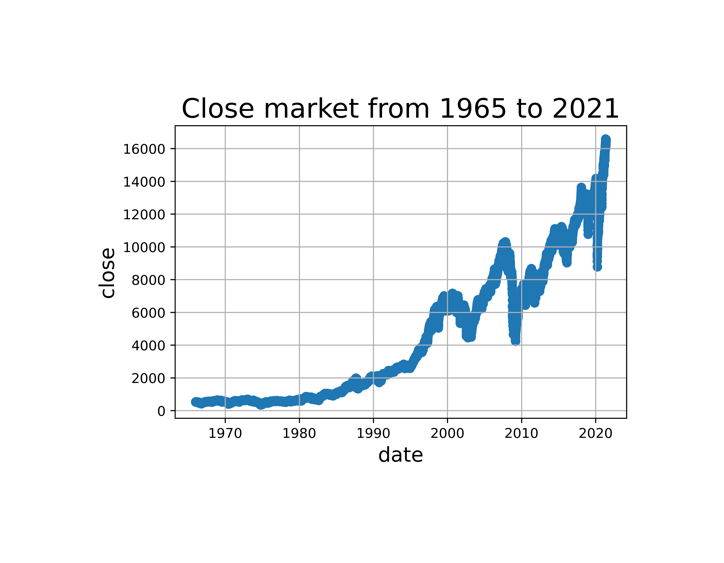


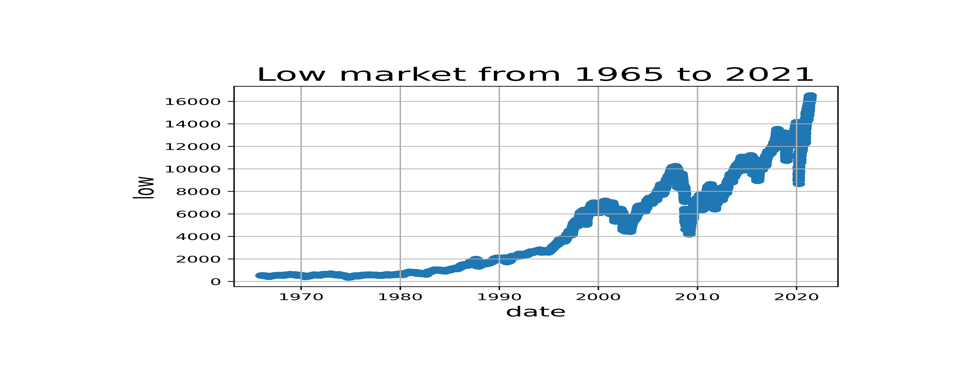
کل دیتاست بر اساس تاریخ (قدیم به جدید) مرتب شد تا نمودار واضح و با کیفیتی رسم شود:



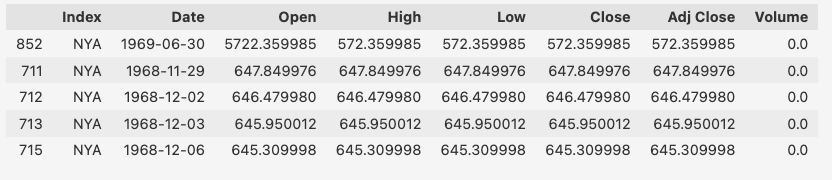
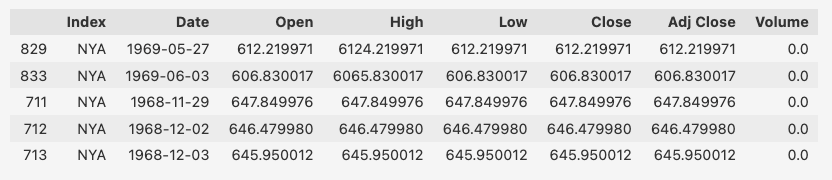
برای هر فیچر به طور جدا گانه scatter رسم شد تا داده های نویز شناسایی شود که در اینجا طبق پلات های زیر ما در فیچر های (Open,High,Adj close) داده نویز داشتیم:

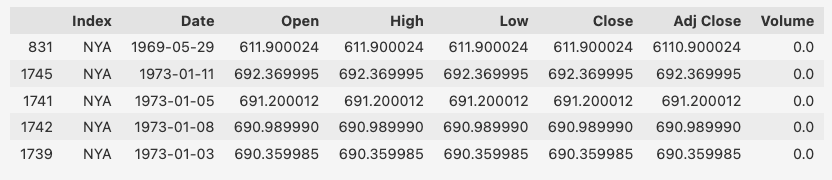






داده های نویز در دیتاست پیدا شد و به صاحب دیتاست گزارش داده شد تا در صورت اصلاح, داده یی حذف نشود ولی با صحبت با اقای مومنی تصمیم به حذف نویز گرفته شد نویز ها یافت شده به شرح زیر است:



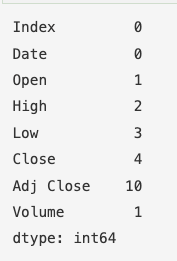
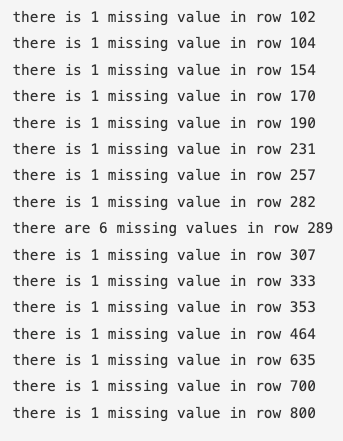
در ایندکس های ۸۵۲ و۸۲۹ و۸۳۱ داده نویز وجود داشته و در سطر مربوطه هیچ missing values وجود ندارد! بعد ار حذف نویز ها ما جدولی ار دیتاست ۱۳۹۴۴×۸ به شرح زیر داریم:



با بررسی داده طبق دیتاست زیر از یک تا ۱۰, missing values در جدول زیر مشاهده شده است:



بعد از صحبت با کارفرما و بررسی , تعداد missing values در هر سطر و ایندکس آن که به شرح زیر است تصمیم به حذف آن گرفته شد:

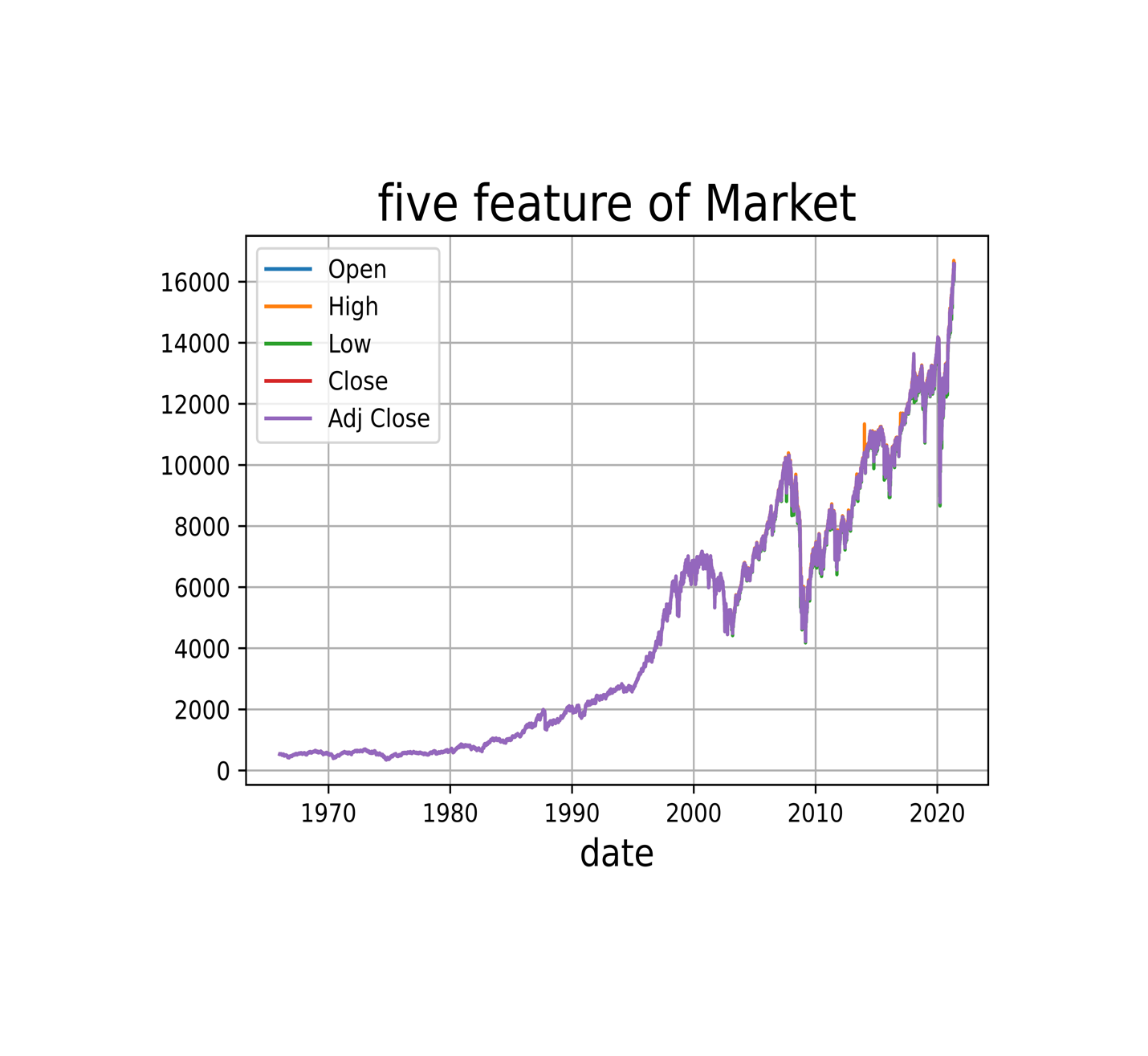


بعد از حذف missing values دیتاست۱۳۹۲۸×۸ مربوطه به شرح زیر است:

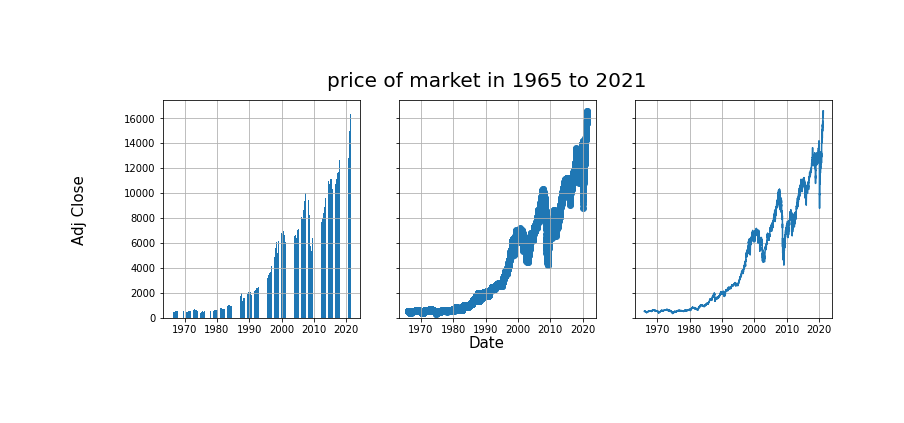


بعد از بررسی و کسب اطلاعات به توجه به دیتاست زیر در سطر count پی میبریم دیتاست به درستی پاکسازی شده است:



با رسم نمودار خطی و بررسی ۵ فیچرها بر اساس تاریخ از سال ۱۹۶۵ تا ۱۹۸۳ روند نمودار تقریبا ثابت بوده و در حدود قیمتی ۵۰۰ بوده از ۱۹۸۵تا۱۹۹۵ نمودار رشد صعودی داشته است و از تقریبا ۵۰۰ به ۳۰۰۰ رسیده است و از ۱۹۹۵ تا ۲۰۰۰ شیب تندی را تجربه کرده است و از حدود قیمتی ۲۰۰۰ به حدود ۷۰۰۰ رسیده است از سال ۲۰۰۰ تا تقریبا ۲۰۰۳ نمودار نزولی بوده است و قیمت به حدود ۴۵۰۰ رسیده است و از اخرای سال ۲۰۰۳ تا نزدیک ۲۰۰۸ دوباره نمودار رشد صعودی را داشته است و قیمت به ۱۰۰۰۰ رسیده است دوباره از نزدیک سال ۲۰۰۸ تا ۲۰۰۹ شیب نزولی را تجربه کرده است و قیمت به ۴۰۰۰ رسیده است از انتهای ۲۰۰۹ نمودار رشد صعودی خود را شروع کرده و با روند صعودی و نزولی در کل رشد قلبل توجه ای را تجربه کرده و به قیمت ۱۴۰۰۰ رسیده است در ۲۰۲۰ دوباره نمودار نزولی بوده و به قیمت حدوده ۸۵۰۰ رسیده و در نهایت از ۲۰۲۰ تا ۲۰۲۱ به بالا ترین قیمت یعنی بالاتر از ۱۷۰۰۰ رسیده است!

علاوه بر نمودار قبلی ستون Adj Close از float به int تبدیل شد و سه پلات از نوع مختلف رسم شد(bar-scatter-plot) که انالیزنمودار نقطه ای و خطی در بالا توضیح داده شد اما نکته جالب در barplot به ما روند رو به رشد کلی بازار را نشان میدهد که با تمام بالا و پایین ها به طور کلی با گذشت زمان روند افزایشی بوده است. 



البته برای انالیز داده های مالی باید از نمودار مخصوص داده مالی استفاده کرد که صعودی بودن و نزولی بودن را با کندل ها رنگی مشخص کرد که کندل سبز سعودی و قرمز نزولی است و در این نمودار میشود به درستی بازار را زیر نظر داشت و از اصلاح بازار روند کلی آن با خبر شد. در اینجا تا سال ۲۰۰۴ روند سعودی بوده دقیقا شبیه عرضه اولیه از ۲۰۰۴ تقریبا با کندل های صعودی و نزولی در کل رشد داشته تقریبا از ۲۰۰۸ ریزش را تجربه کرده است و از ۱۰۰۰۰ به قیمت ۴۰۰۰ رسیده و همینطور اصلاحاتی را در روند خود تجربه کرده است

