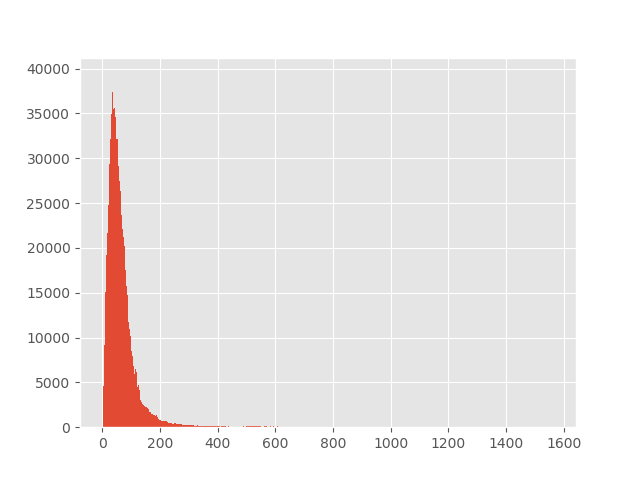
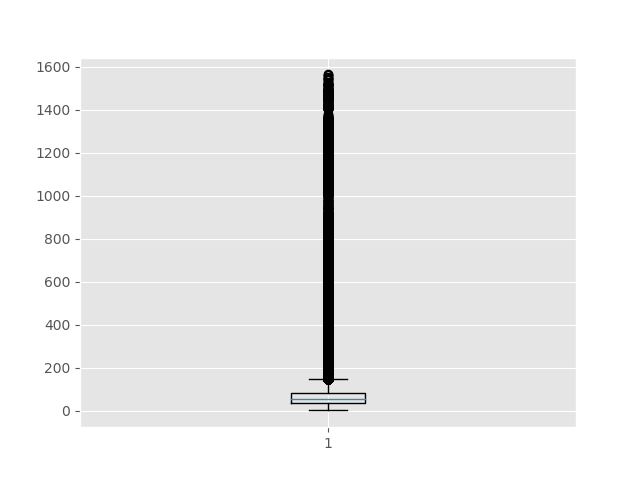
به نام خدا

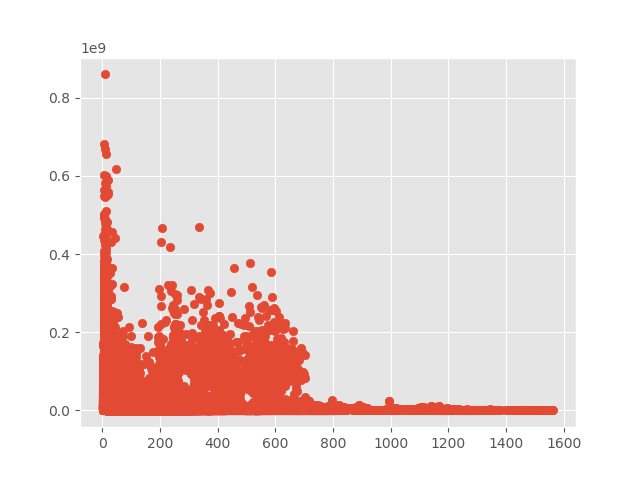
چون به مباحث آماری در زمینه اقتصاد علاقه مند بوده و اهمیت آن نیز بسیار بالا بود ، دیتابیس prices از NY Stock و برای متغیر های رشته ای از دیتابیس securities استفاده کردم.

1. در سوال 1 پس از رسم هیستوگرام و باکس پلات میانگین قیمت معاملات ( که به دلیل ازدیاد مقادیر کمی در هم به نظر می آید) به محاسبه skewness پرداخته شده که مثبت بودن آن (مقدار آن 6.654625309027434 است ) در شکل پیداست. با توجه به این که skewness مثبت و داشتن tail طولانی تر نشانه سود بیشتر است ، نشان می دهد بورس نیویورک در حالت بهینه ای قرار دارد. همچنین modality آن unimodal است و یک بیشترین مقدار دارد و این نشانه ثبوت بورس است.





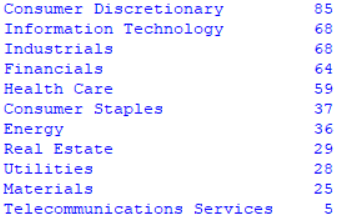
1. در سوال 2 scatterplot کشیده شده بین حجم معاملات و میانگین قیمت معاملات کوواریانس (62831367.69413253-) منفی آن که علاوه بر نشان دادن پراکندگی از میانگین نشان دهنده هم جهت نبودن تغییرات دو متغیر است. همچنین کورولیشن منفی (0.060101590891279796-) نزدیک به یک نشان دهنده وابستگی البته از نوع معکوس است.در این دیتاست 43844 outlier وجود داشت که مثلا به [213.4399985] [236.560005 ] [340.4999925][248.925003 ][255.8899995] [182.579994 ] ها میتونیم اشاره کنیم .

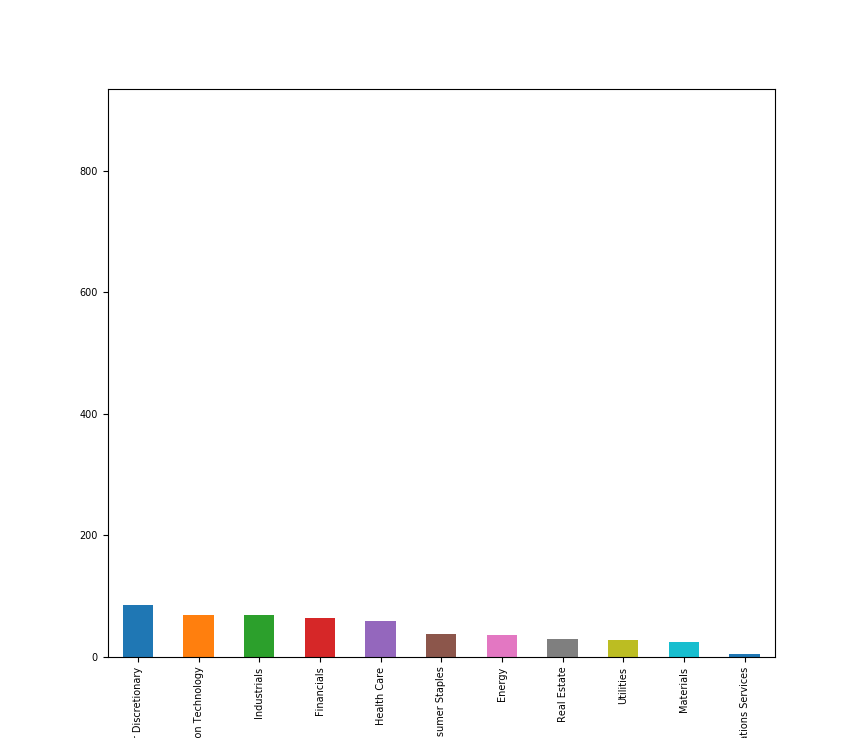


1. در سوال 3 حوزه کاری شرکت هایی که سهام آنها در بورس وجود دارد به صورت نمودار میله ای درآمده

همچنین جدول فرکانسی آن درج شده.

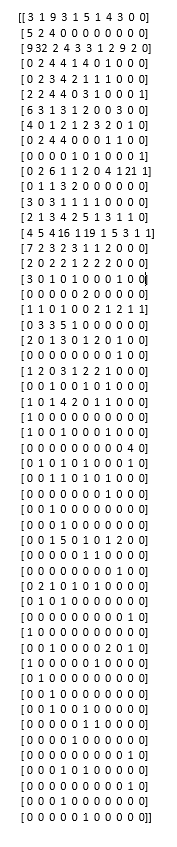
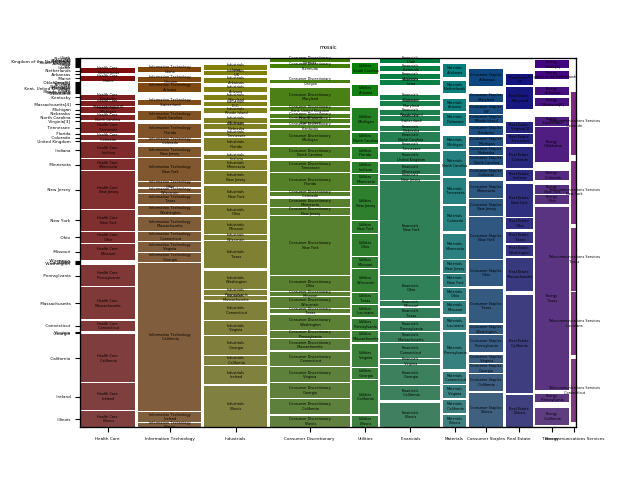
.

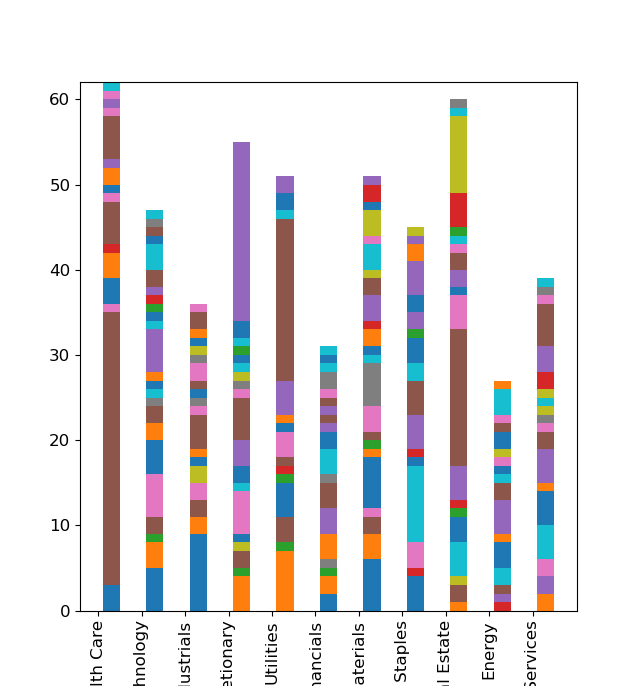




1. در سوال 4 نیز contingency table برای متغیر های حوزه کاری و ایالت یا کشور محل اقامت آمده.

ستون های این جدول حوزه کاری و سطر های آن کشور محل اقامت است.همچنین موزاییک پلات برای نشان دادن رابطه بین حوزه کاری و محل شرکت علاوه بر segmented bar plot کشیده شده .





1. در سوال 5 نیز میانگین داده های زیاد و کم به عنوان داده در نظر گرفته شده . 70.83094499341804 میانگین کل داده ها است و (70.65320632077656, 71.00868366605953) بازه ایست که حوزه اطمینان 99 درصد را برای ما فراهم می آورد .
2. در سوال 6 مقدار p-value برای متغیر فرضی مورد نظر است . متغیر فرضی بیشترین قیمت در نظر گرفته شده و برای مقدار اصلی میانگین (طبق انتظار ) مقدار یک است . برای مقدار 1 p-value صفر و برای مقدار 68مقدار آن 123 e -6.24200606015393 است. این بدین معناست که ما بنابر محاسبات مطمئنیم که مقدار میانگین این دیتاست 1 نیست . همچنین چون مقدار p-value برای حدس 68 از 5 درصد کمتر است مطمئن هستیم که میانگین 68 نیز نیست.اما برای حدس 70 ، مقدار p-value به دست آمده 0.1874181109963041 است که از 5 درصد بیشتر است . این نشان دهنده آن است که ما مدرکی برای اثبات میانگین نبودن عدد 70 نداریم.
3. در این سوال برای پیدا کردن معادله خط رگرسیون از متغیر پاسخ هزینه صرف شده برای تحقیقات و متغیر های توضیحی کل دارایی در حال حاضر و سود خالص استفاده شده.معادله رگراسیون خط اول به صورت و خط دوم به صورت است.

نمودار رسم شده نشان دهنده این است ارتباط مناسبی بین متغیر های توضیحی و پاسخ برقرار است و از روی هزینه صرف شده برای تحقیقات میتوان به طور خوبی دارایی های حال حاضر و سود خالص را حدس زد.