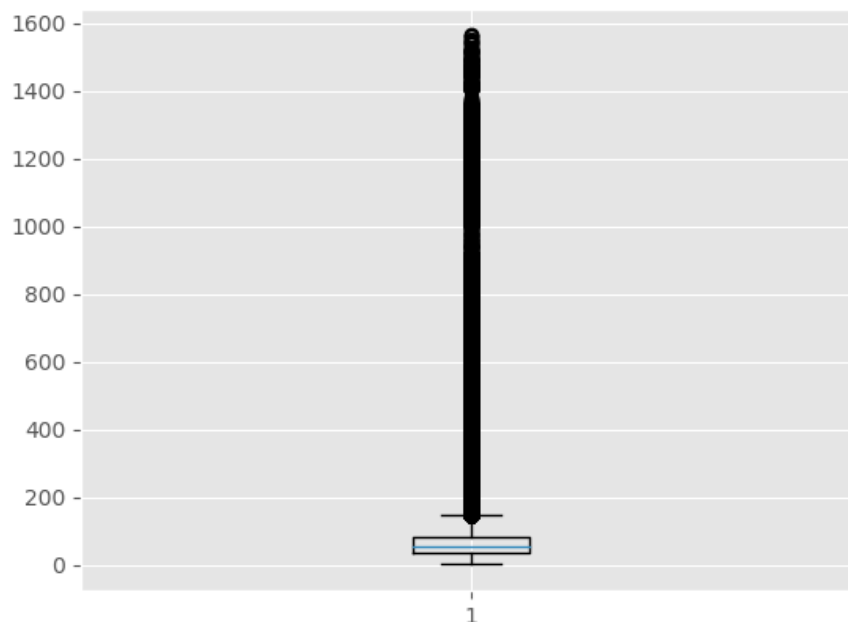
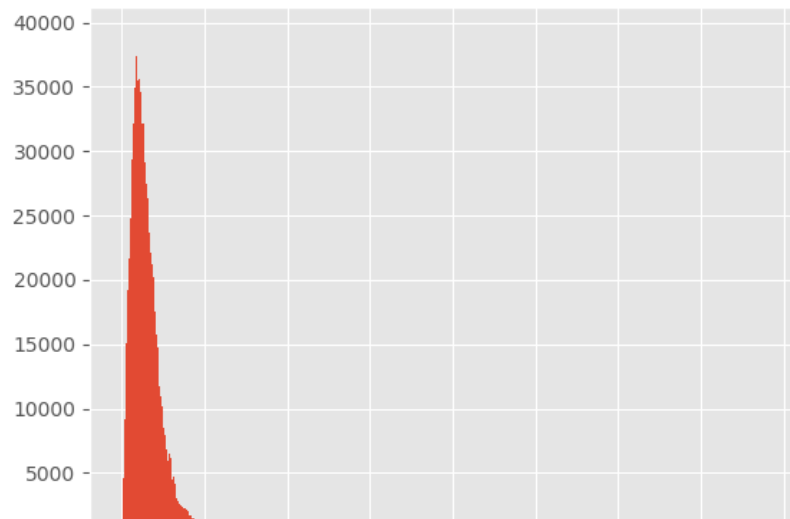


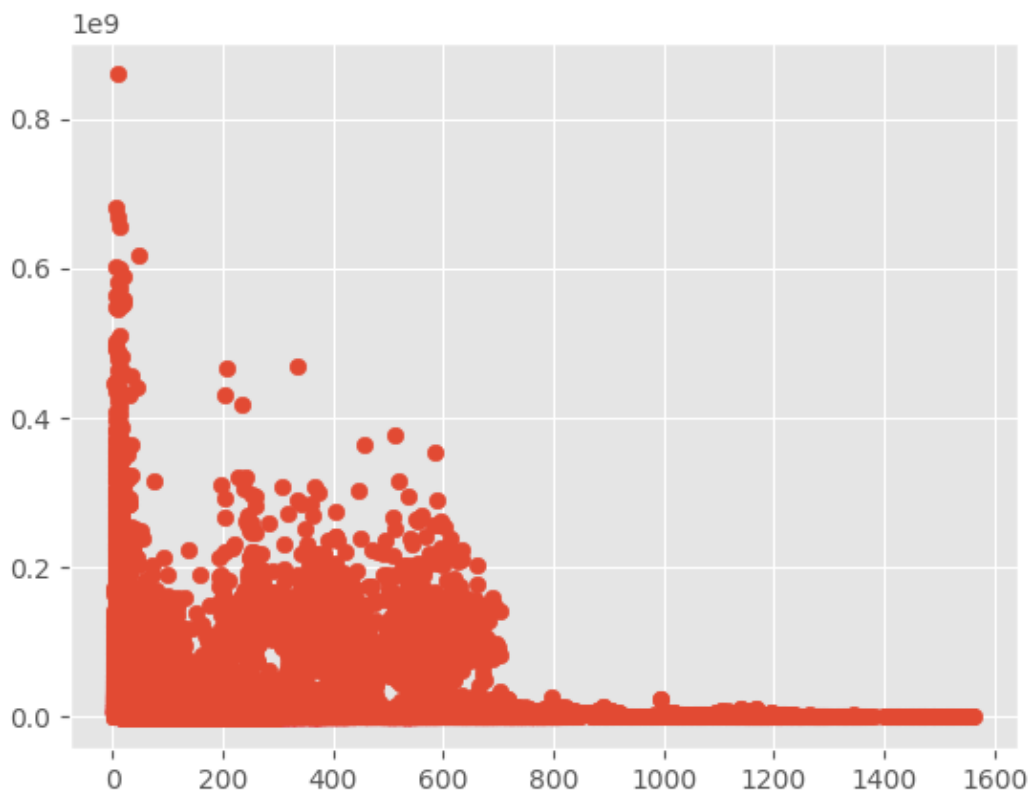
به نام خدا

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 ۱ 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 ۶/۶۵۴۶۲۵۳۰۹۰۲۷۴۳۴ 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 unimodal آن modality است و یک بیشترین مقدار دارد و این نشانه ثبوت بورس است.

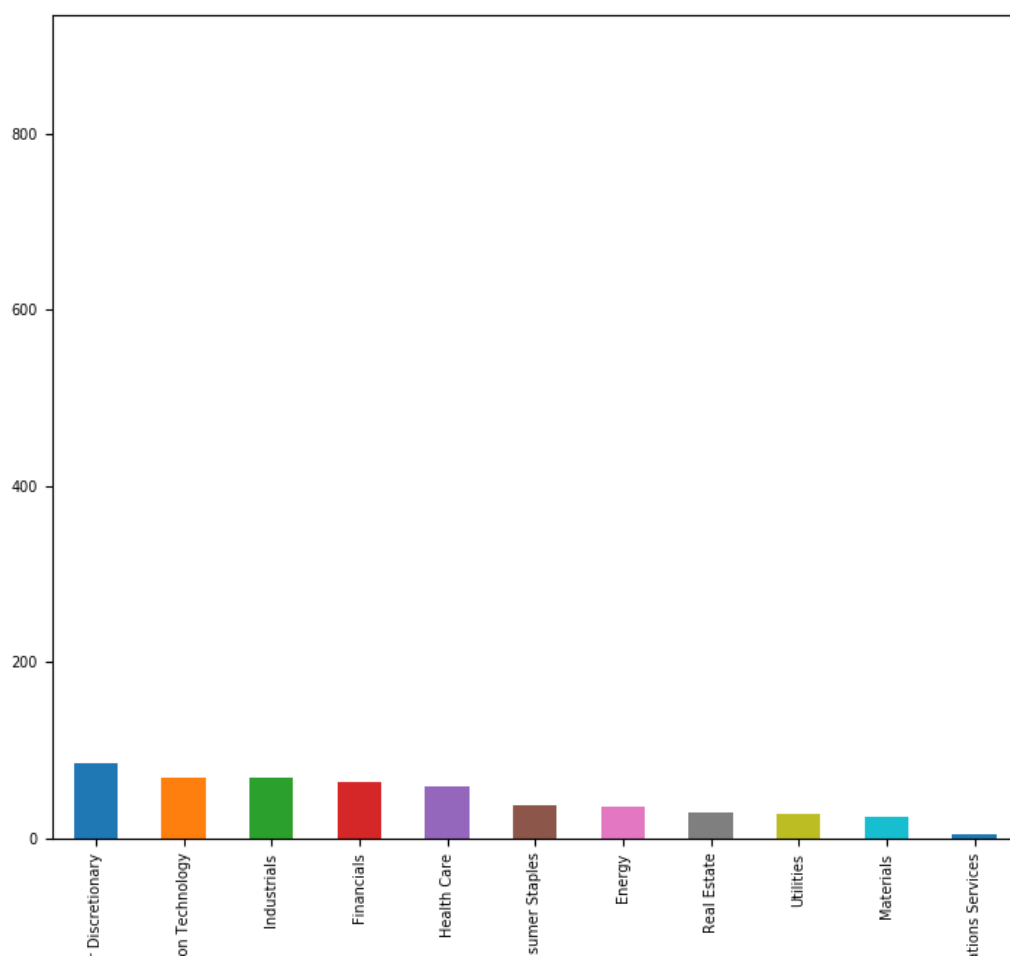


2. در سوال ۲ scatterplot کشیده شده بین حجم معاملات و میانگین قیمت معاملات کوواریانس $(-۶۲۸۳۱۳۶۷/۶۹۴۱۳۲۵۳)$ منفی آن که علاوه بر نشان دادن پراکندگی از میانگین نشان دهنده هم جهت نبودن تغییرات دو متغیر است. همچنین کورولیشن منفی $(-۰/۰۶۰۱۰۱۵۹۰۸۹۱۲۷۹۷۹۶)$ نزدیک به یک نشان دهنده وابستگی البته از نوع معکوس است. در این دیتاست ۴۳۸۴۴ outlier وجود داشت که مثلاً به $[۲۱۳/۴۳۹۹۹۸۵]$ $[۲۳۶/۵۶۰۰۰۵]$ $[۳۴۰/۴۹۹۹۹۲۵]$ $[۲۴۸/۹۲۵۰۰۳]$ $[۲۵۵/۸۸۹۹۹۹۵]$ $[۱۸۲/۵۷۹۹۹۴]$ ها میتونیم اشاره کنیم.



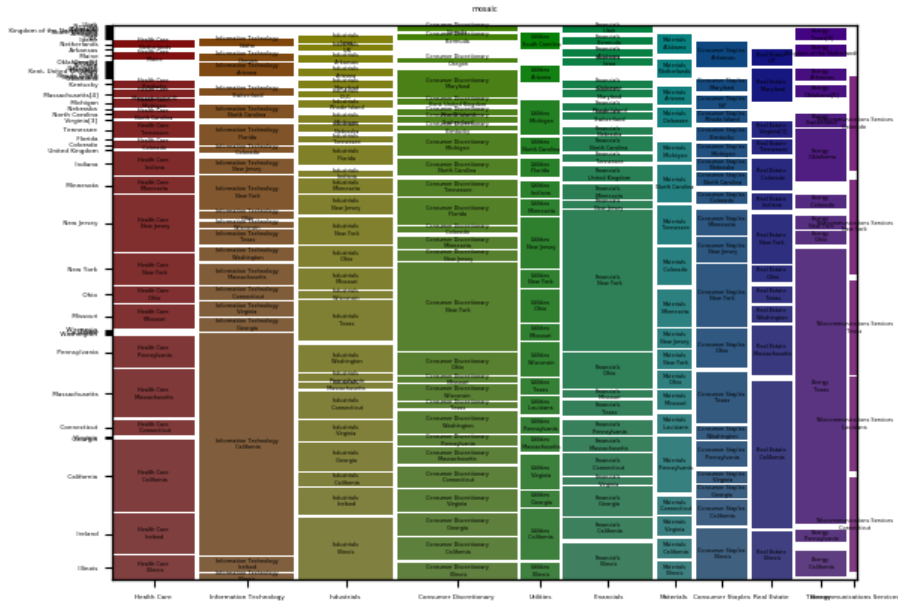
3. در سوال ۳ حوزه کاری شرکت هایی که سهام آنها در بورس وجود دارد به صورت نمودار میله ای درآمد
همچنین جدول فرکانسی آن درج شده.

Consumer Discretionary	85
Information Technology	68
Industrials	68
Financials	64
Health Care	59
Consumer Staples	37
Energy	36
Real Estate	29
Utilities	28
Materials	25
Telecommunications Services	5

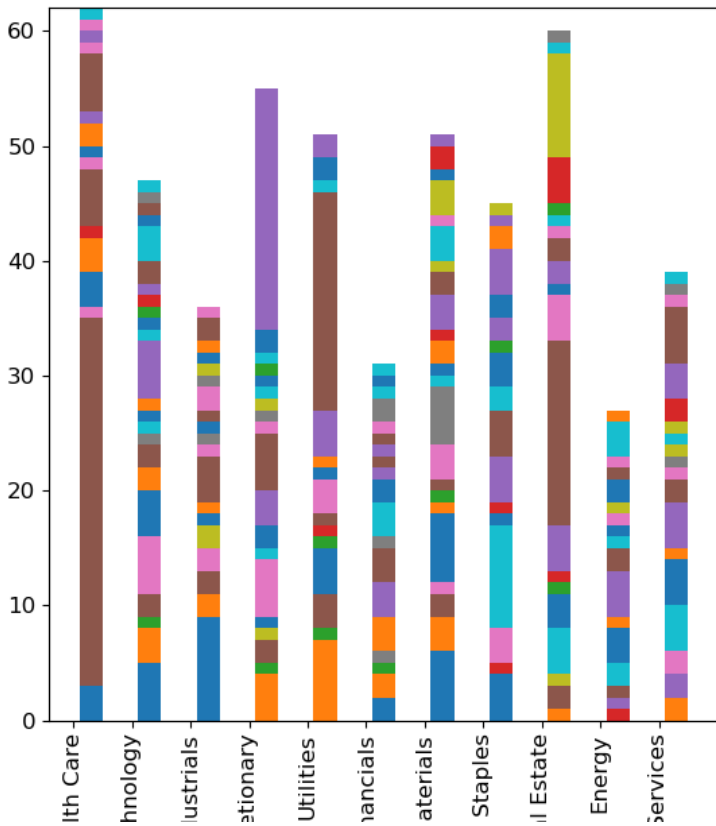


4. در سوال ۴ نیز contingency table برای متغیرهای حوزه کاری و ایالت یا کشور محل اقامت آمده.

ستون های این جدول حوزه کاری و سطرهای 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 .



```
[[3 1 9 3 1 5 1 4 3 0 0]
[5 2 4 0 0 0 0 0 0 0 0]
[9 3 2 2 4 3 3 1 2 9 2 0]
[0 2 4 4 1 4 0 1 0 0 0]
[0 2 3 4 2 1 1 1 0 0 0]
[2 2 4 4 0 3 1 0 0 0 1]
[6 3 1 3 1 2 0 0 3 0 0]
[4 0 1 2 1 2 3 2 0 1 0]
[0 2 4 4 0 0 0 1 1 0 0]
[0 0 0 0 1 0 1 0 0 0 1]
[0 2 6 1 1 2 0 4 1 2 1 1]
[0 1 1 3 2 0 0 0 0 0 0]
[3 0 3 1 1 1 1 0 0 0 0]
[2 1 3 4 2 5 1 3 1 1 0]
[4 5 4 1 6 1 1 9 1 5 3 1 1]
[7 2 3 2 3 1 1 2 0 0 0]
[2 0 2 2 1 2 2 2 0 0 0]
[3 0 1 0 1 0 0 0 1 0 0]
[0 0 0 0 0 2 0 0 0 0 0]
[1 1 0 1 0 0 2 1 2 1 1]
[0 3 3 5 1 0 0 0 0 0 0]
[2 0 1 3 0 1 2 0 1 0 0]
[0 0 0 0 0 0 0 0 1 0 0]
[1 2 0 3 1 2 2 1 0 0 0]
[0 0 1 0 0 1 0 1 0 0 0]
[1 0 1 4 2 0 1 1 0 0 0]
[1 0 0 0 0 0 0 0 0 0 0]
[1 0 0 1 0 0 0 1 0 0 0]
[0 0 0 0 0 0 0 0 0 4 0]
[0 1 0 1 0 1 0 0 0 1 0]
[0 0 1 1 0 1 0 1 0 0 0]
[0 0 0 0 0 0 0 1 0 0 0]
[0 0 1 0 0 0 0 0 0 0 0]
[0 0 0 1 0 0 0 0 0 0 0]
[0 0 1 5 0 1 0 1 2 0 0]
[0 0 0 0 0 1 1 0 0 0 0]
[0 0 0 0 0 0 0 0 1 0 0]
[0 2 1 0 1 0 1 0 0 0 0]
[0 1 0 1 0 0 0 0 0 0 0]
[0 0 0 0 0 0 0 0 0 1 0]
[1 0 0 0 0 0 0 0 0 0 0]
[0 0 1 0 0 0 0 2 0 1 0]
[1 0 0 0 0 0 1 0 0 0 0]
[0 1 0 0 0 0 0 0 0 0 0]
[0 0 1 0 0 0 0 0 0 0 0]
[0 0 1 0 0 1 0 0 0 0 0]
[0 0 0 0 0 1 1 0 0 0 0]
[0 0 0 0 1 0 0 0 0 0 0]
[0 0 0 0 0 0 0 0 0 1 0]
[0 0 0 1 0 1 0 0 0 0 0]
[0 0 0 0 0 0 0 0 0 1 0]
[0 0 0 1 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0]]
```



5. در سوال ۵ نیز میانگین داده های زیاد و کم به عنوان داده در نظر گرفته شده . $70/83094499341804$. میانگین کل داده ها است و $(65320632077656/60868366605953,70/65320632077656)$ بازه ایست که حوزه اطمینان ۹۹ درصد را برای ما فراهم می آورد .

6. در سوال ۶ 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 ۱ p-value صفر و برای مقدار ۶۸ مقدار آن $123 - 15393e - 24200606015393$ است . این بدین معناست که ما بنابر محاسبات مطمئنیم که مقدار میانگین این دیتاست ۱ نیست . همچنین چون مقدار p-value برای حدس ۶۸ از ۵ درصد کمتر است مطمئن هستیم که میانگین ۶۸ نیز نیست . اما برای حدس ۷۰ ، مقدار p-value به دست آمده $0/1874181109963041$ است که از ۵ درصد بیشتر است . این نشان دهنده آن است که ما مدرکی برای اثبات میانگین نبودن عدد ۷۰ نداریم .

7. 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 رگرسیون خط اول به صورت $y = 3.028803735919617 * x + 3639012504.139398$ و خط دوم به صورت $y = 5.251411675864058 * x + 18452118716.404438$ 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 .