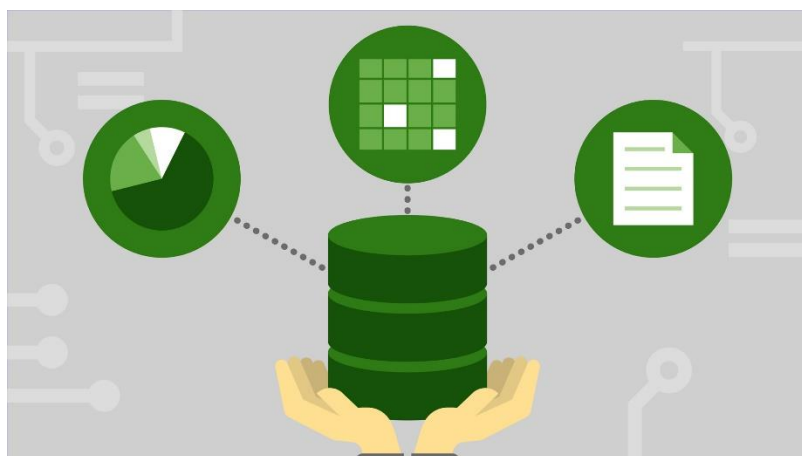


به نام خدا



دانشگاه تهران  
پردیس دانشکده‌های فنی  
دانشکده برق و کامپیوتر



## آزمایشگاه پایگاه داده

دستور کار شماره ۶

مجتبی بنائی

مسئولین آزمایش: امیرعلی رایگان، نگار مرادی

## دستور کار شماره ۶ - کار با الاستیک سرچ

همانطور که در پیش گزارش توضیح داده شد، در کنار این دستور کار چند فایل پایتون برای گردآوری داده های مورد نیاز قرار داده شده است. فایل اول scrap.py نام دارد. در این فایل تکه کدی برای به دست آوردن اطلاعات موجود در صفحه ی سایت خرید و فروش خودروی باما وجود دارد. پیشنهاد می شود ابتدا وارد لینک شوید و با مطابقت کد های این فایل به نحوه ی گرفتن اطلاعات (web scraping) پی ببرید. خروجی این تکه کد یک فایل Json می باشد که حاوی object یا همان دیکشنری هایی است که اطلاعات هر خودرو را در خود جای می دهد.

فایل دیگری نیز در اختیار شما قرار دارد که insert2elastic.py نام دارد. در این فایل یک داده ی نمونه ایجاد شده و به مجموعه داده ای با نام cars که در الاستیک سرچ قابل دسترسی است اضافه شده است. دقت کنید نکات مهم این کد نحوه ی برقراری ارتباط با الاستیک سرچ توسط پایتون و همچنین اضافه کردن داده به آن است.

### گام اول:

شما در ابتدا باید کتابخانه های لازم برای اجرای کد اول را نصب کنید و سپس آن را اجرا کنید تا خروجی json مد نظر را به دست آورید. داده هایی که در این فایل هستند تعدادی از ماشین هایی هست که در سایت باما برای فروش قرار داده شده اند. حال باید فایل کد دوم را باز کنید. در این فایل باید تغییراتی ایجاد کنید تا برنامه ی شما بتواند فایل json خروجی برنامه ی قبلی را باز کند و تک تک داده های داخل آن را در پایگاه داده ی cars اضافه کند تا بتوانیم از آنها در الاستیک سرچ استفاده کنیم. (توجه کنید زمانی که می خواهید کد پایتونی که با الاستیک سرچ ارتباط برقرار می کند را اجرا کنید لازم است الاستیک سرچ از قبل روی پورت ۹۲۰۰ بدون مشکل در حال اجرا باشد.)

### گام دوم:

حال می خواهیم دستوراتی مشابه آنهایی که در پیش گزارش انجام دادیم را بر روی داده های جدیدی که داریم اعمال کنیم. برای این کار کنسول را باز کنید، داده های جدیدی که داریم را لود کنید و دستوراتی را وارد کنید تا خروجی های زیر را چاپ کنید:

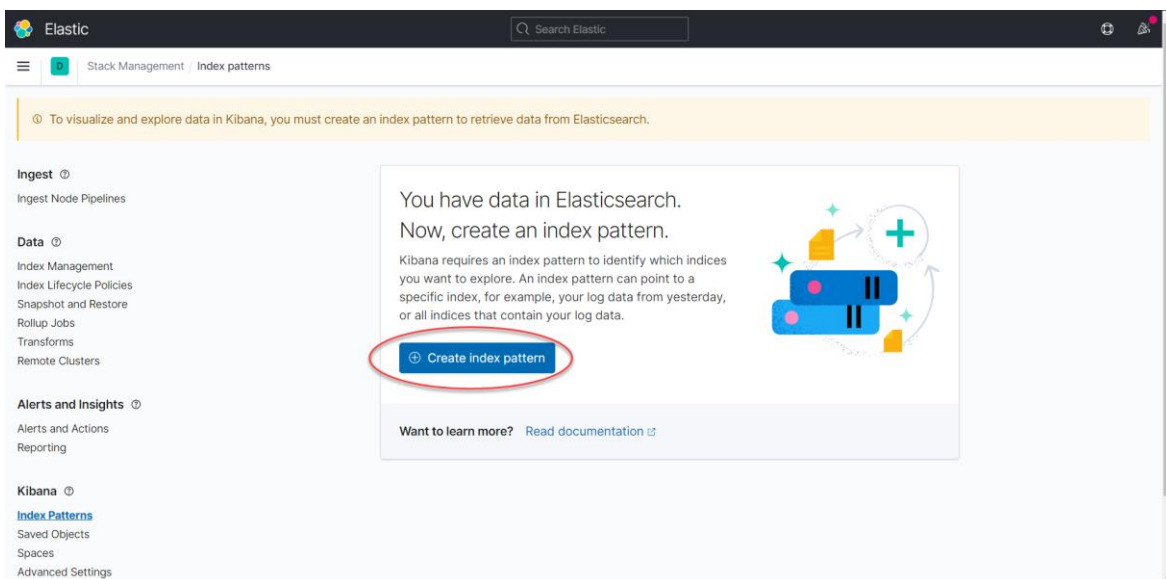
- اطلاعات تمام ماشین هایی که در شیراز به فروش گذاشته شده اند.
- اطلاعات تمام ۲۰۶ هایی که در تهران یا کرج به فروش گذاشته شده اند.
- اطلاعات تمام ماشین هایی که کمتر از صد هزار کیلومتر کار کرده اند و جدید تر از سال ۱۳۹۷ تولید شده اند.
- اطلاعات تمام هیوندای هایی که قدیمی تر از سال ۱۳۹۸ هستند، ارزان تر از ۲ میلیارد تومان هستند و بیشتر از صد هزار کیلومتر کار کرده اند.

## گام سوم:

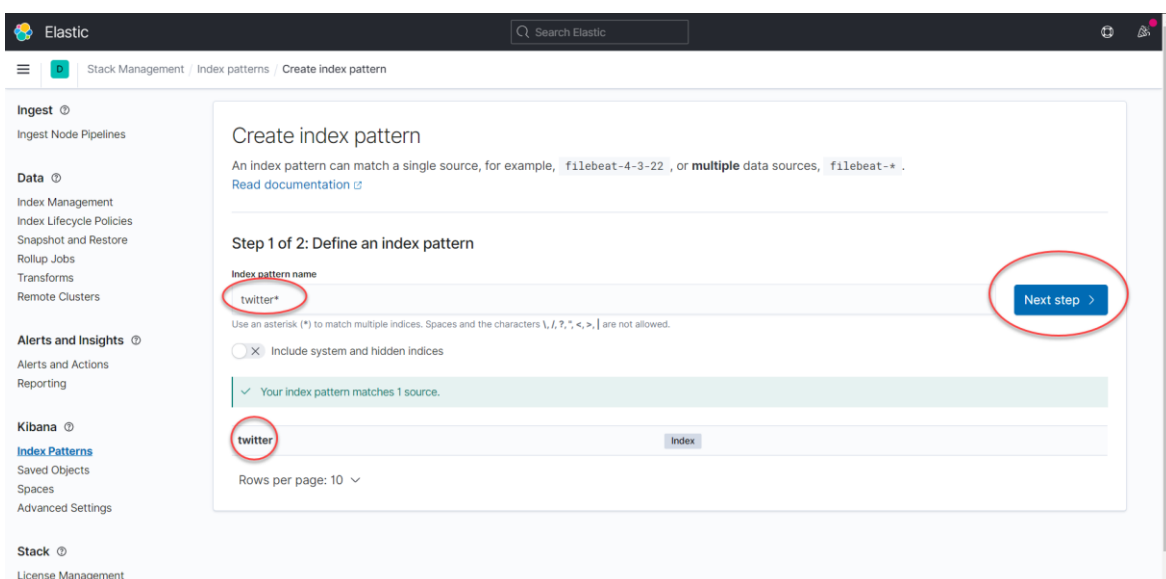
حال می خواهیم بکه تحلیل تصویری روی این داده ها داشته باشیم. برای این کار یک داشبورد با چهار پنل ایجاد کنید. این چهار پنل می توانند مواردی مثل نمودار خطی یا ستونی، ابرکلمات، توزیع انواع ماشین ها، تعداد ماشین ها با ویژگی های خاص و یا مانند آن باشند.

نحوه ایجاد داشبورد هم به صورت زیر است :

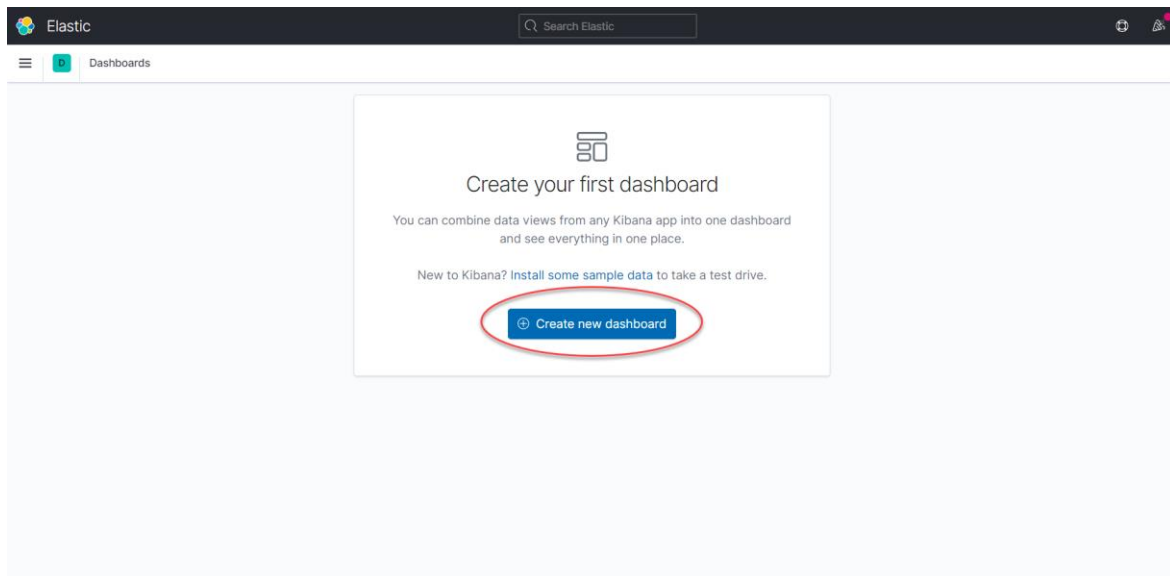
ابتدا یک **index pattern** می سازیم.



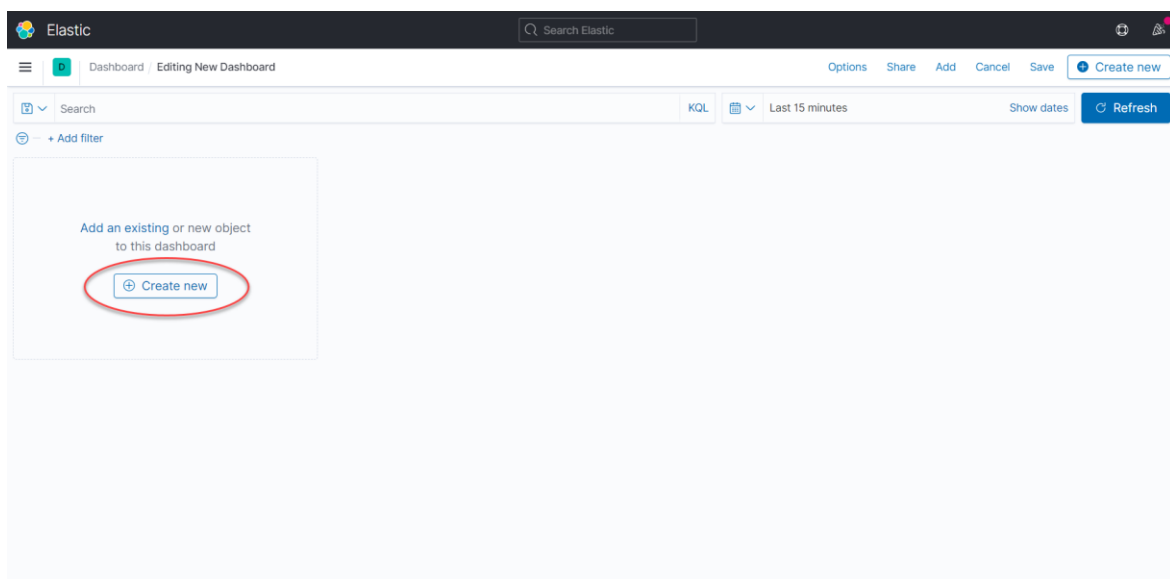
ایندکس با نام **cars** را انتخاب کرده و **next** را می زنیم. (به جای **twitter** شما باید **cars** را انتخاب کنید)



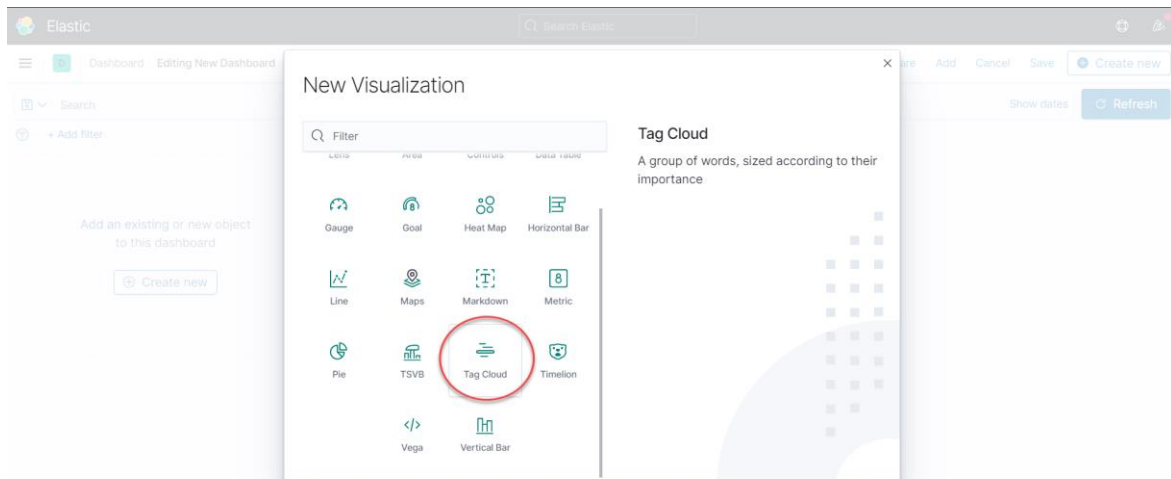
سپس یک داشبورد می سازیم.



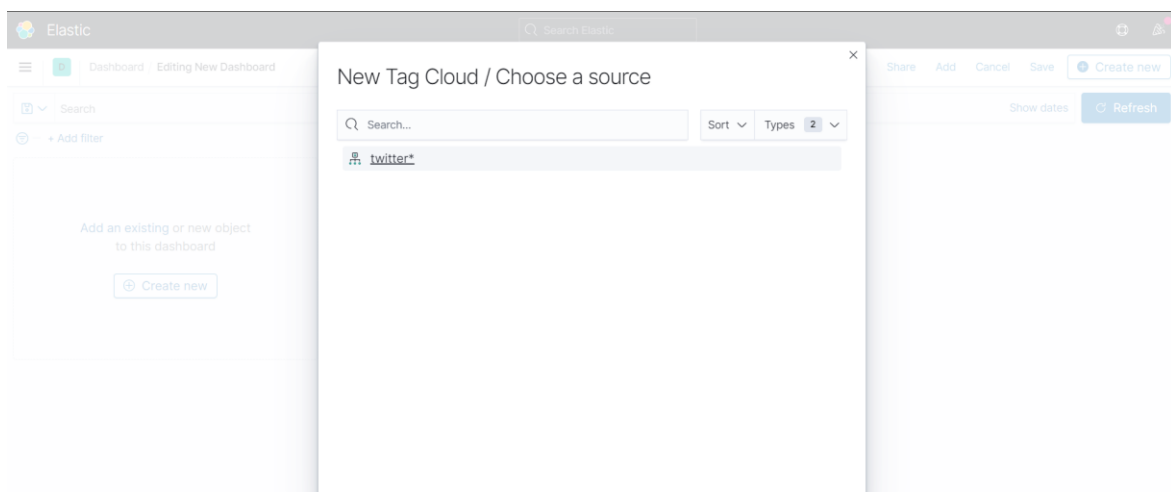
سپس **create new** را می زنیم تا پنل های مختلف را به این داشبورد اضافه کنیم.



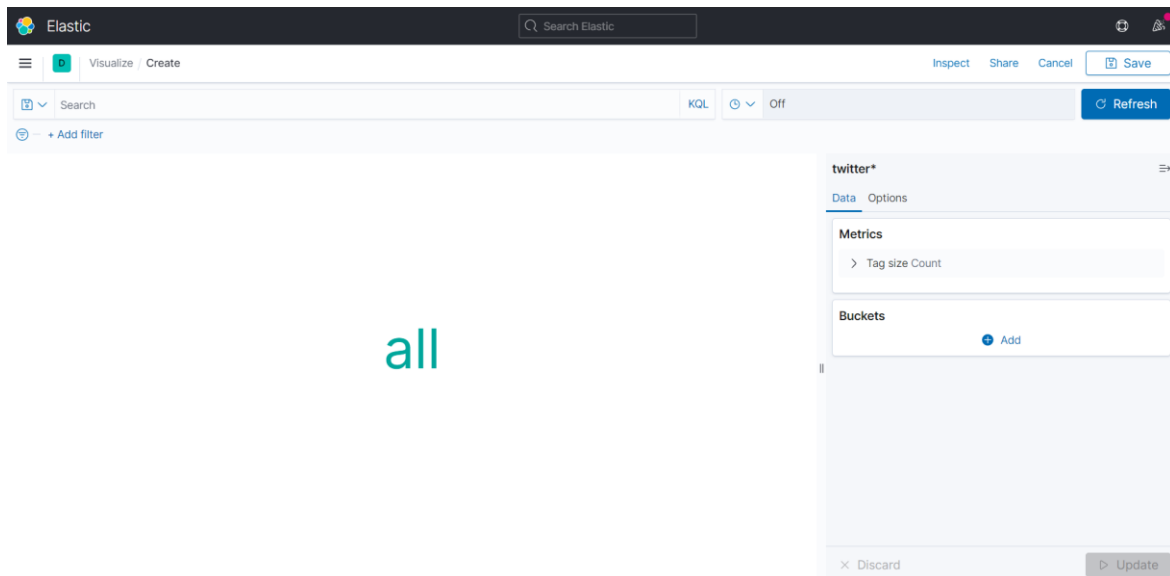
برای **visualization** اول **tag cloud** را انتخاب می کنیم که بسته به تعداد تکرار کلمه، سایز آن ست می کند.



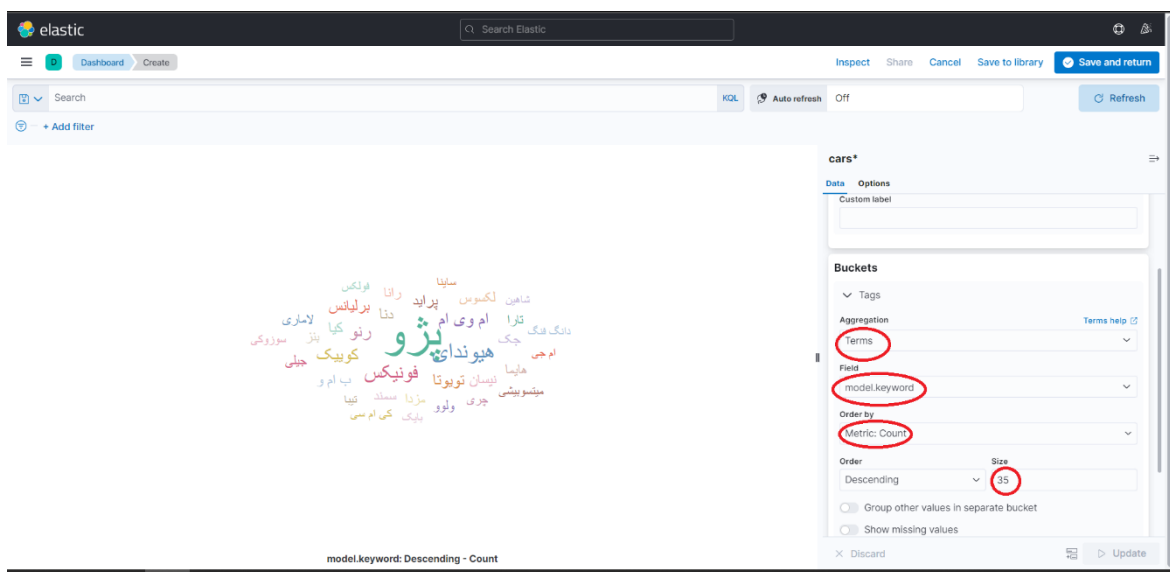
سپس **cars** را انتخاب می کنیم. (به جای **twitter** شما باید **cars** را انتخاب کنید)



در ابتدا تنها **all** نشان داده می شود.

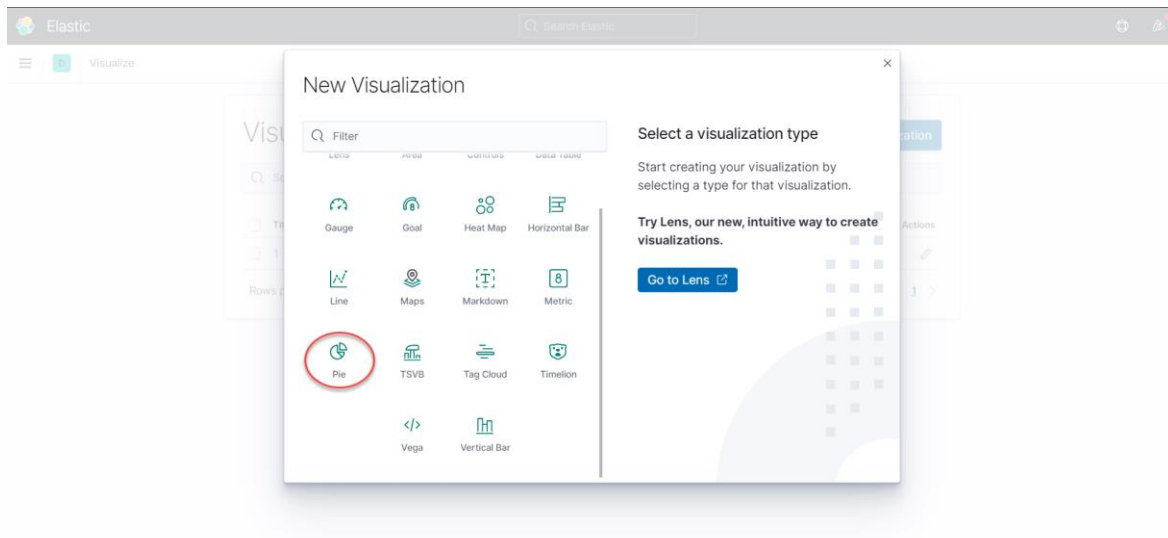


تنظیمات رو در سمت راست تغییر می دهیم تا برای hashtag ها این visualization انجام شود که خروجی به صورت زیر می شود.

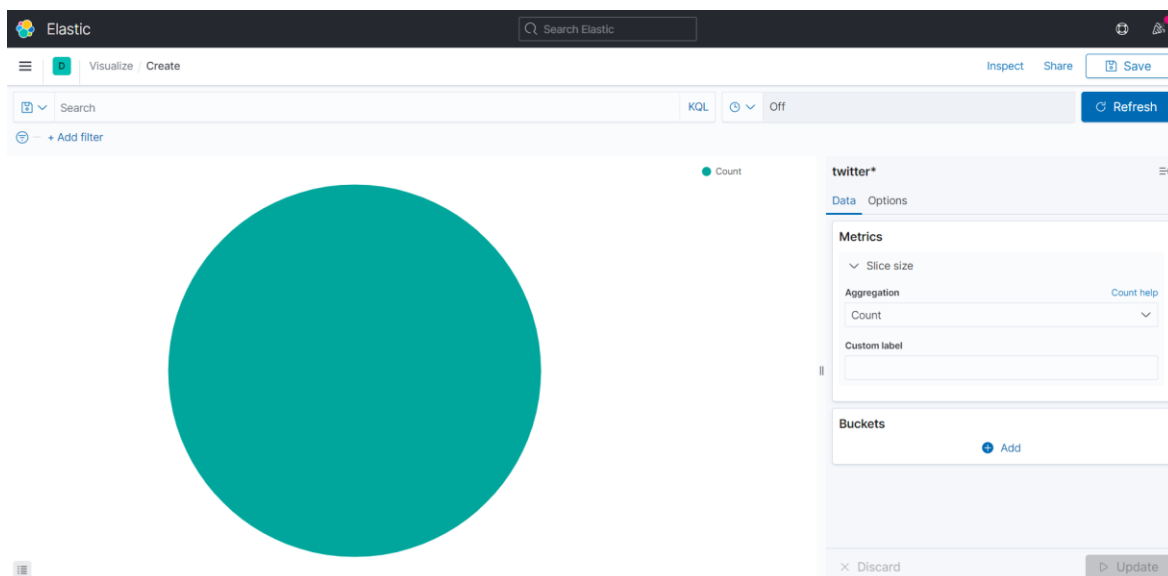


برای نمونه در مثال بالا ۳۵ ماشینی که بیشترین آگهی روی آنها ایجاد شده نشان داده شده اند.

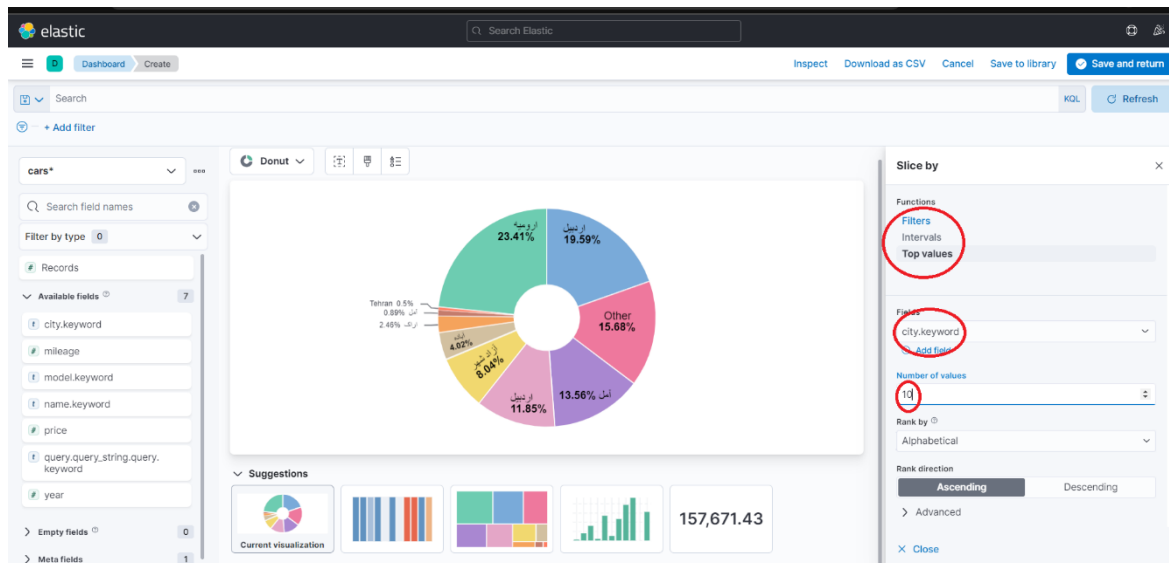
برای **visualization** دوم، **Pie** را انتخاب می کنیم که سایز **slice** دایره بر اساس تعداد تکرار کلمه می شود.



خروجی اولیه به این صورت می شود.



تنظیمات را تغییر می دهیم و خروجی نهایی به این صورت می شود.



در این نمودار هم ۱۰ شهری که بیشترین آگهی از آنها ایجاد شده نشان داده شده اند.

در این گام نیز باید با کنجکاوی و کار با امکانات تحلیل تصویری که در کیبانا وجود دارد ۵ تحلیل تصویری روی داده هایی که داریم به همراه توضیح آن ایجاد کنید.

برای نمونه یک تحلیل دیگر می تواند به صورت زیر باشد که نشان دهنده ی توزیع قیمت های ماشین های فروشی است. میتوان دید حدود ۲۰ درصد آگهی های این سایت را ماشین هایی با قیمت ۱۵ تا ۱۷ میلیارد تومان تشکیل داده اند!!

