الگو‌های پرتکرار وقوانین انجمنی

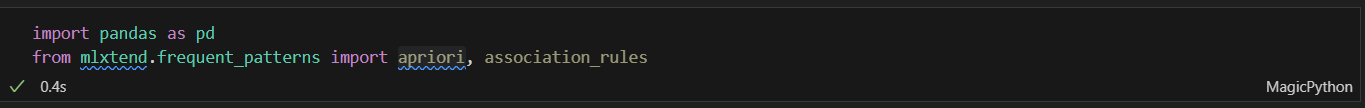
برای درس داده کاوی

# مقدمه

در این قسمت از پروژه سعی بر آن است که هفت الگوی پرتکرار را از داده هایی که در قسمت های قبل پروژه پاکسازی نموده به دست بیاوریم. این امر توسط الگوریتم Apriori انجام میشود.

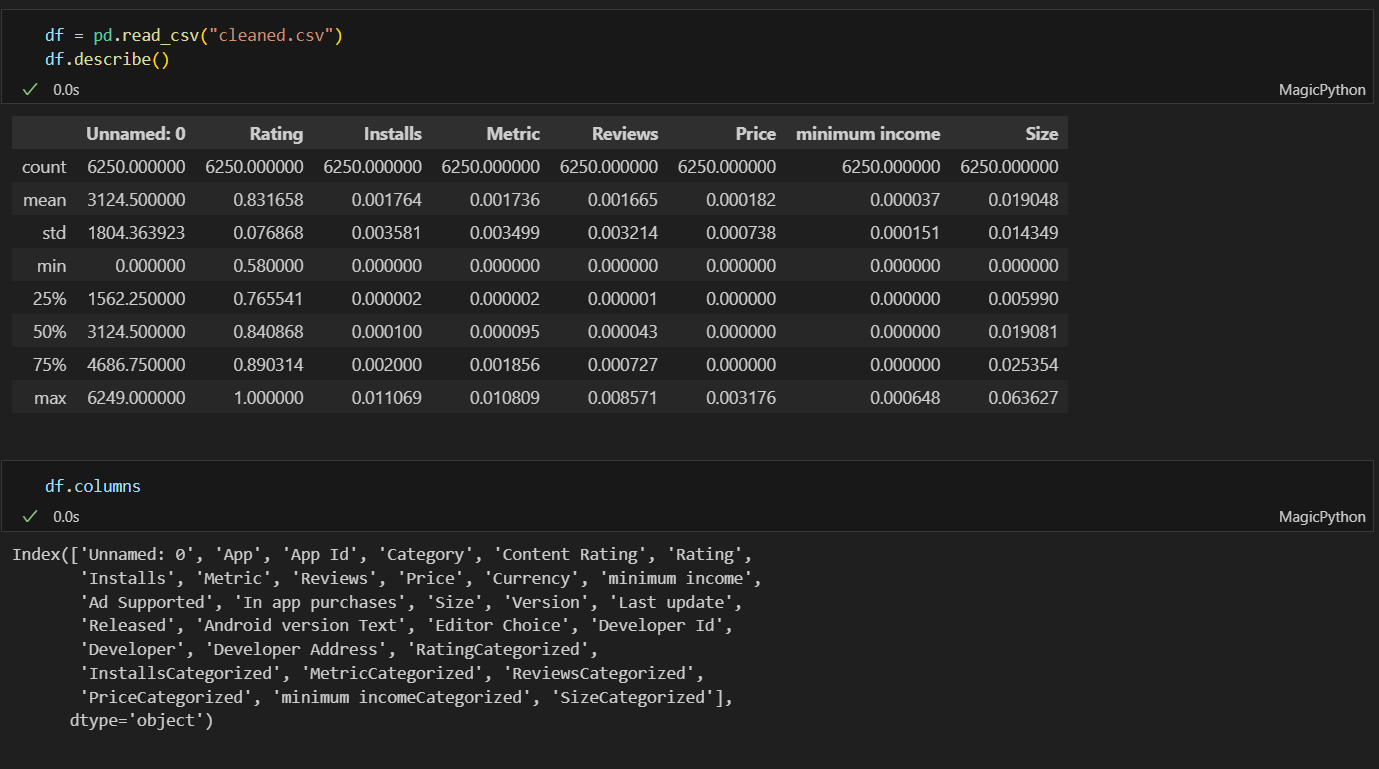
# توضیحات کد

## کتابخانه های مورد نیاز

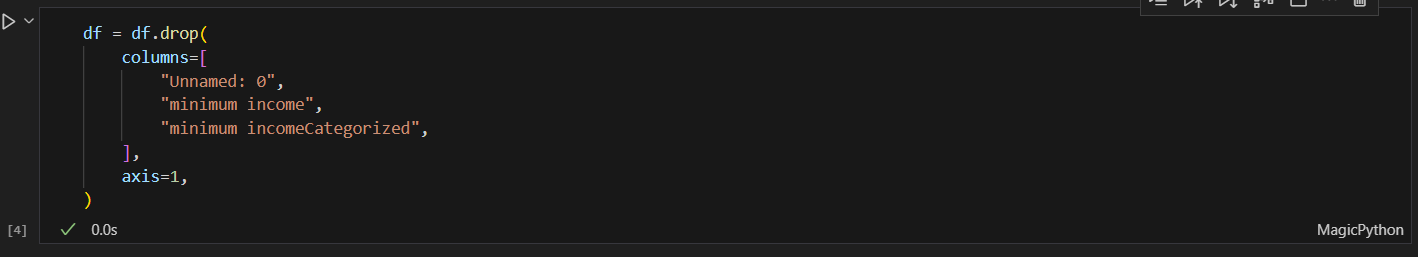


از کتابخانه pandas برای دستکاری داده ها و از کتابخانه mlxtend برای دسترسی با الگوریتم های Aprioriو association rules استفاده میکنیم.

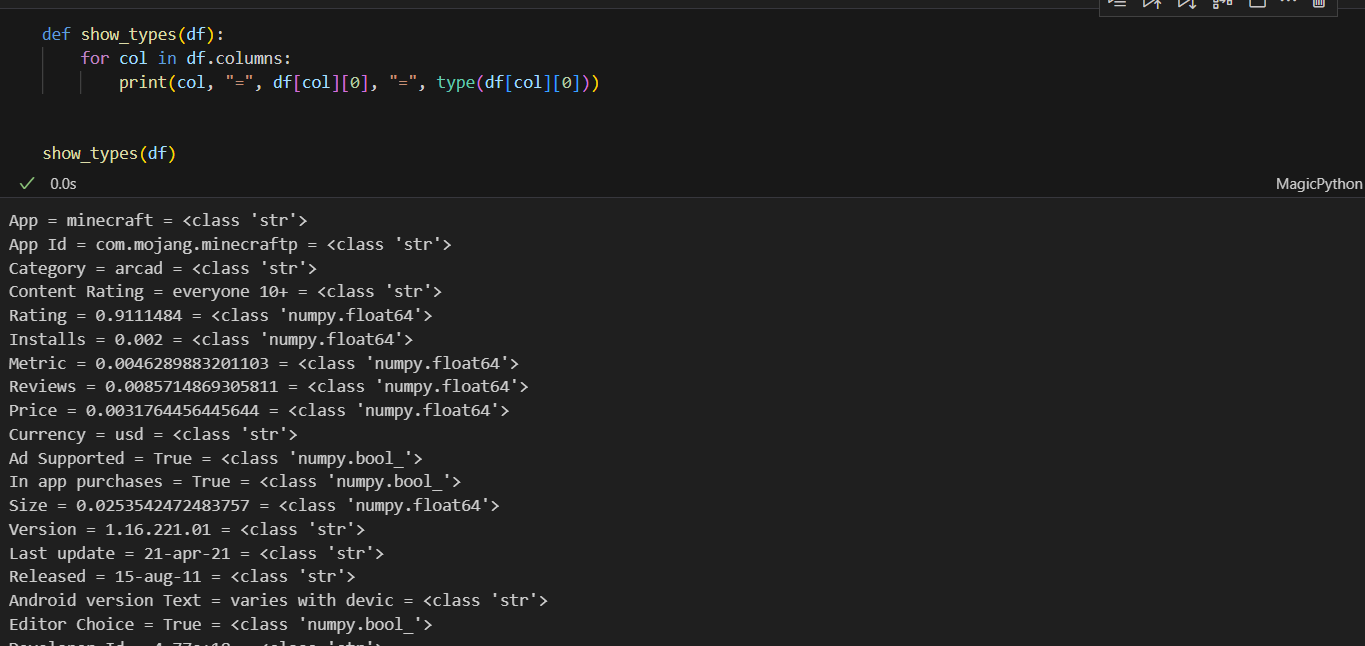
## بارگزاری داده ها در حافظه

با استفاده از کتابخانه pandas داده های پاکسازی شده در قسمت قبلی پروژه را در حافظه بارگذاری میکنیم. بعد از آن توضیحات کلی در مورد داده ها نمایش داده شده اند.

## انتخاب صفات مورد بررسی



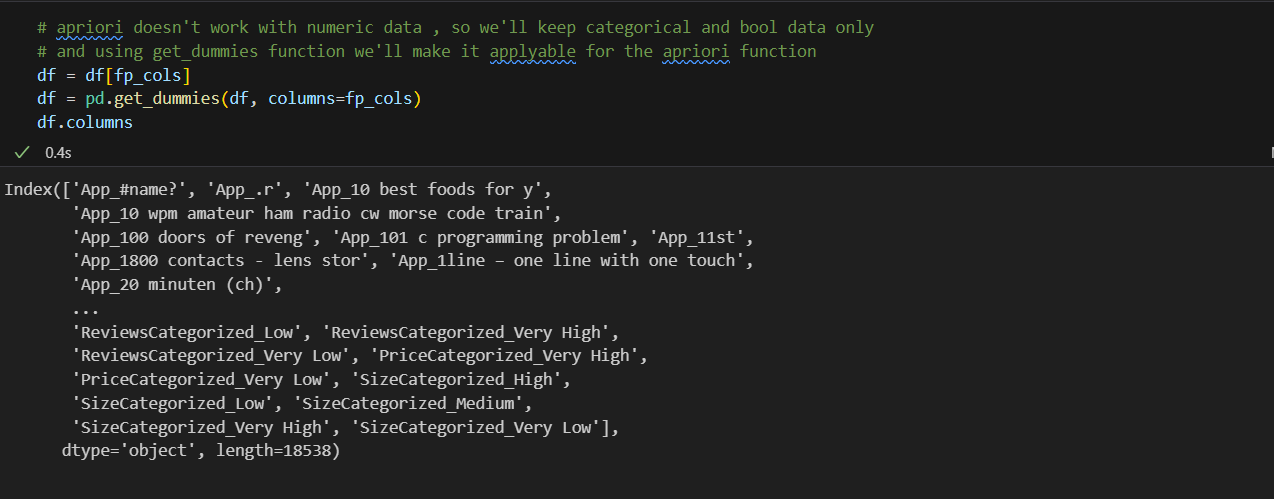
با توجه به اینکه برخی ستون ها مانند Unnamed: 0 و minimum income مشخصه خاصی نیستند ویا از طریق دیگر صفات قابل دست یابی اند آن ها را حذف میکنیم.



حال صفاتی را که نوع مورد نظر آنها از نوع str یا bool است را برای تحلیل انتخاب میکنیم(این انواع داده‌ای ملزومات تابع Apriori میباشند).

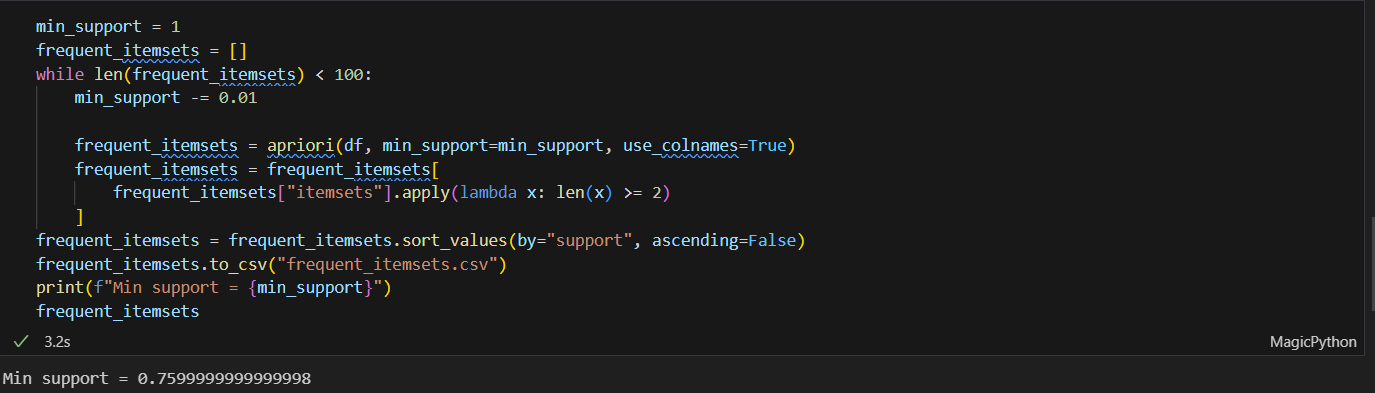


## تبدیل به فرمت مناسب



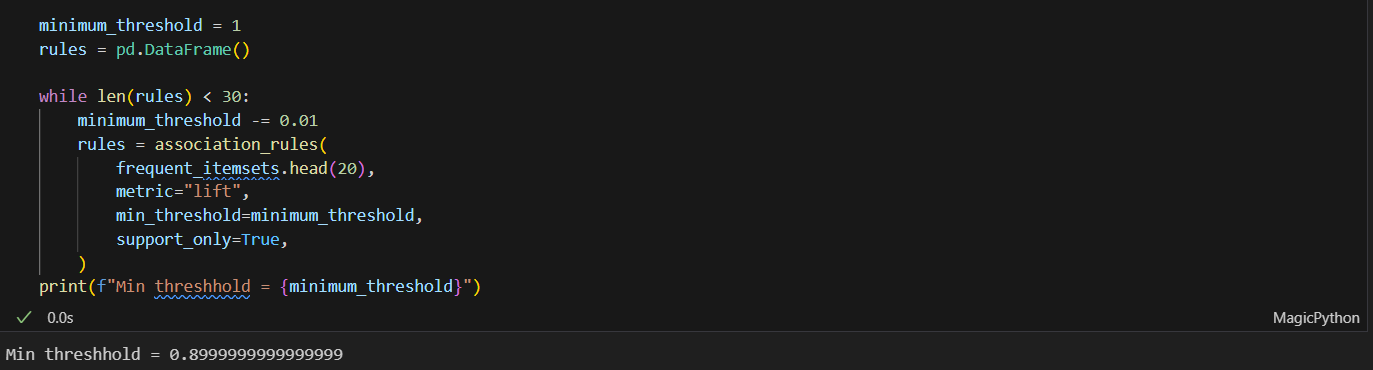
حال یک دیتافریم با ستون ها مورد نظر اختیار میکنیم و با استفاده از دستور get\_dummies داده ها را به صورتی انکود میکنیم که ترکیب ها متفاوت را نشان دهند تا در الگوریتم Apriori قابل استفاده باشد.

## پیدا کردن ایتم ست های پرتکرار



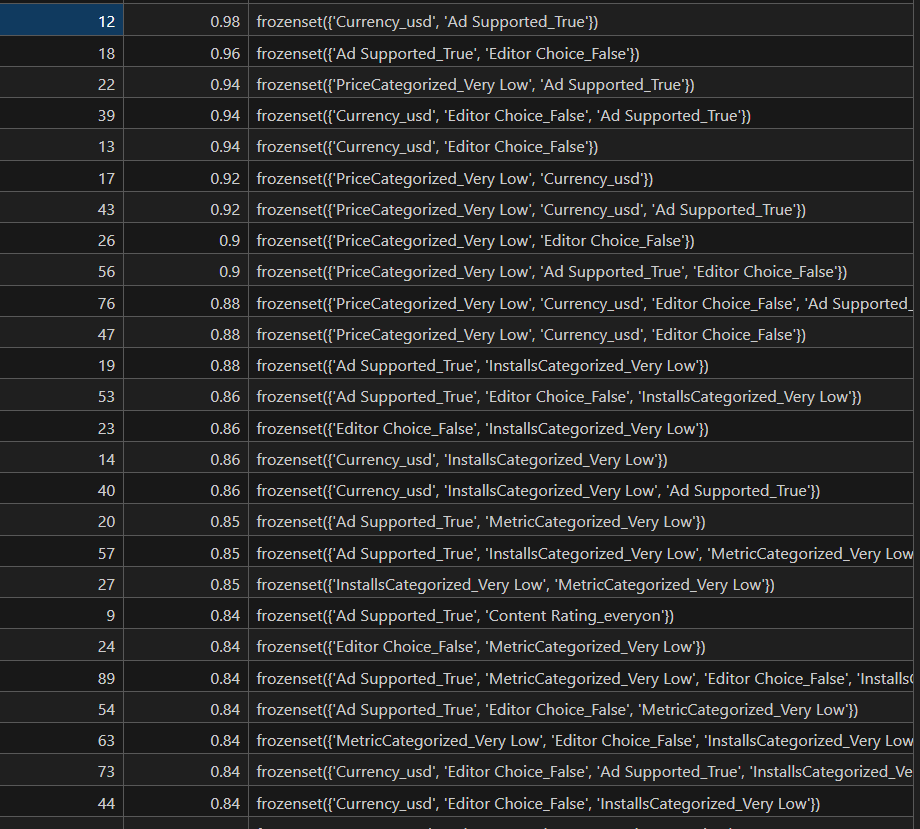
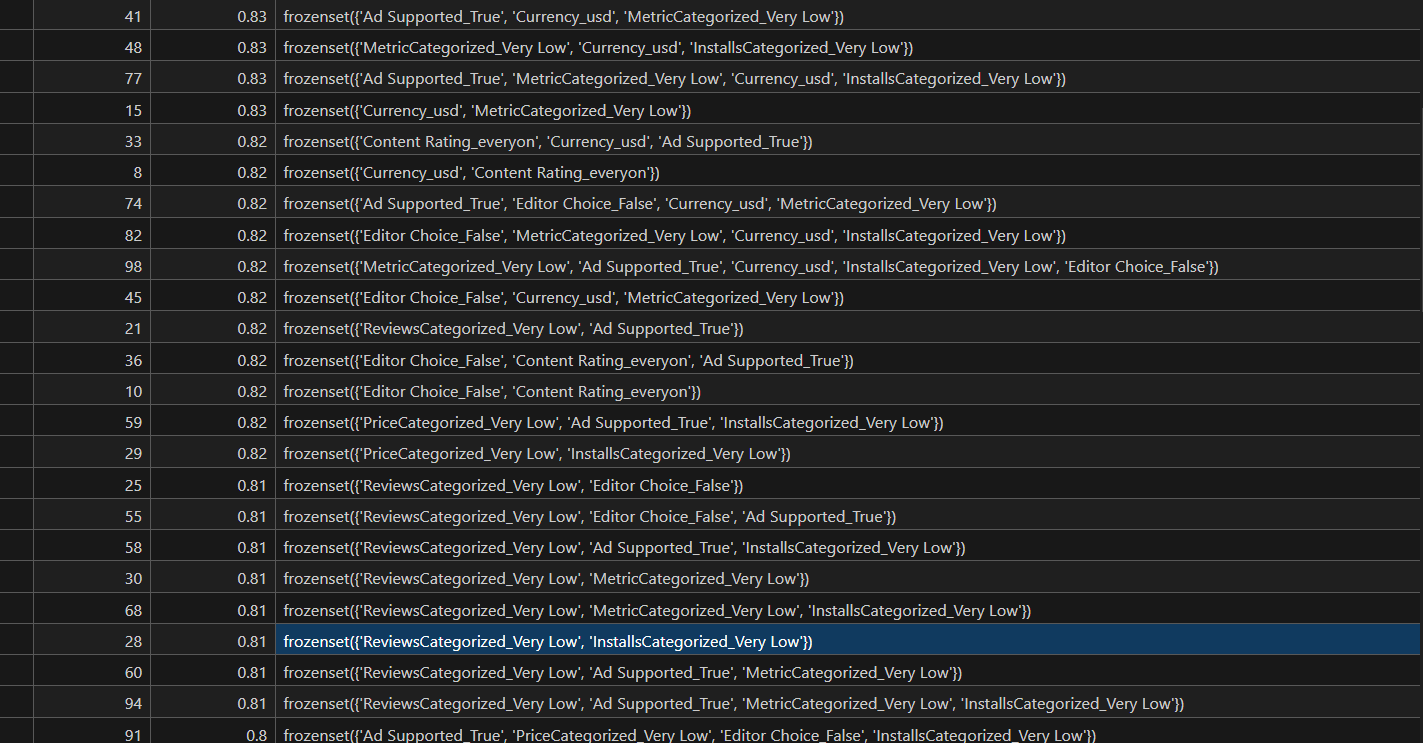
چونکه در صورت پروژه خواسته شده است ساپورت حداقلی باشد و همچنین بسیاری از ایتم های پرتکرار یافت شده دانش خاصی اضافه نمی کنند و بدیهیات اند، در ابتدا min\_support را برابر یک قرار داده و سپس در یک حلقه از مقدار آن با گام های 0.01 میکاهیم تا جایی که ایتم ست های پرتکرار به دست آمده به بیش از 100 عدد برسند تا مطمين باشیم از بدیهیات گذر کرده ایم.

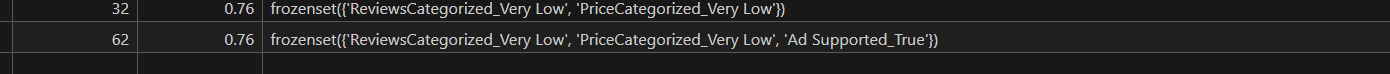
## قوانین انجمنی



در این قسمت نیز مثل قسمت قبل ، minimum\_threshold از 1 شروع و با گام های 0.01 کاهش میابد تا قوانین بدیهی گذر شود.

# تحلیل داده های نهایی



1. مشخص است که 98٪ تبلیغات درون برنامه‌ای اپ های گوگل پلی از طریق دلار برای صاحب اپ درامدزایی میکنند. این مورد احتمالا به خاطر سیتسم تبلیغات درون برنامه‌ای است که خود گوگل پلی ارائه میکند و یا دیفالت بودن این ارز در سیستم.  
   مورد 12.
2. نتیجه ها نشان میدهد که اپلیکیشن هایی که در آن ها از تبلیغات درون برنامه ای استفاده میشود، معمولا توسط تیم گوگل پلی به عنوان Editor’s Choice انتخاب نمیشوند که نشان دهنده کاهش کیفیت محصول شامل تبلیغات است.96٪  
   مورد 18.
3. در 88 ٪ موارد اپ هایی که تبلیغ درون برنامه ای دارند، تعداد نسب بسیار پایینی دارند.  
   مورد 19.
4. 90٪ مواقع اپ هایی که قیمت پایینی دارند یا مجانی هستند، جزو Editor’s Choice گوگل پلی قرار نمیگیرند.  
   مورد 26.
5. در 84٪ اپ ها، رده بندی سنی اپلیکیشن برای عموم است و از طریق تبلیغات درون برنامه ای درامد کسب میکند  
   مورد 9.
6. در 76٪ اپلیکیشن ها که قیمت بسیار پایینی دارند ،بازخورد کاربران از ان ها نیز بسیار ضعیف بوده  
   مورد 32
7. در 81٪ موارد ،زمانی که بازخورد های یک اپ بسیار ضعیف بوده، تعداد نسب های ان نیز بسیار پایین بوده است.  
   مورد 28