فاطمه قنادی ۹۶۳۶۱۳۰۷۵

در کلاس Arules سه تابع explorarity\_analysis , get\_frequent\_item\_sets , get\_arules پیاده سازی شده اند.

همچنین این کلاس دارای ۳ متغیر freq\_1, total , transactions است که total پس از اجرای تابع get\_frequent\_item\_set ارایه ای میشود از دیکشنری‌ها که هر دیکشنری شامل مجموعه‌های پرتکرار i تایی است که i برابر ایندکس همان دیکشنری است و freq\_1 برابر لیست آیتم‌های پرتکرار است.

در تابع explorarity\_analysis کمترین مقدار ساپورت گرفته می‌شود و آیتم‌هایی که فروششان از ان مقدار بالاتر است را خروجی میدهد.

به این صورت که یک بار کل فایل را میخواند و هر ایتمی را که دید:

اگر قبلا ندیده بود، ‌ان را با تکرار ۱ به دیکشنری اضافه میکند.

اگر قبلا دیده بود، به مقدار value اش یکی اضافه می‌کند.

در تابع get\_frequent\_item\_sets مقدار ساپورت از کاربر گرفته می‌شود و تمام مجموعه خریدهایی که مقدار ساپورتشان از این مقدار بیشتر است را برمیگرداند. بدیهی است که هر چه یک مجموعه بزرگتر باشد، احتمال رخدادش کمتر می‌شود.

در این تابع همانند تابع قبلی یک بار کل دیتابیس خوانده میشود و مجموعه‌های ۱تایی که تکرارشان بیشتر از مینیمم ساپورت است در دیکشنری large به همراه تعداد تکرارش ذخیره میشوند.

مقدار key های این دیکشنری در ارایه arr ریخته میشود، سپس سورت میشود و از روی این ارایه تمام مجموعه‌های دوتایی ساخته میشود و به عنوان کاندیدها در دیکشنری candinates ذخیره میشوند.(مطابق اصول apriori)

سپس اعضای candidate بررسی میشوند و هر کدام که پرتکرار بود در دیکشنری large ذخیره میشود.

همینطور این کار در یک حلقه بی‌نهایت انجام میشود تا جایی که large خالی باشد(یعنی نتوانسته مجموعه‌ بزرگتر بسازد.) در این شرط ارایه total را بازمیگرداند.

در تابع get\_arules سعی میکنیم قوانین را از روی ارایه total استخراج کنیم. به این صورت که از روی کلیدهای هر دیکشنری ذخیره شده در total زیر مجموعه‌هایش را میسازد و با توجه به مقادیر min support , min confidence , min lift قوانین تایید و ذخیره میشوند.

هر قانون را به این صورت بررسی کردیم:

A, B, … -> X

سپس در دیکشنری LF هر قانون تایید شده با مقدار LIFT اش ذخیره میشود.

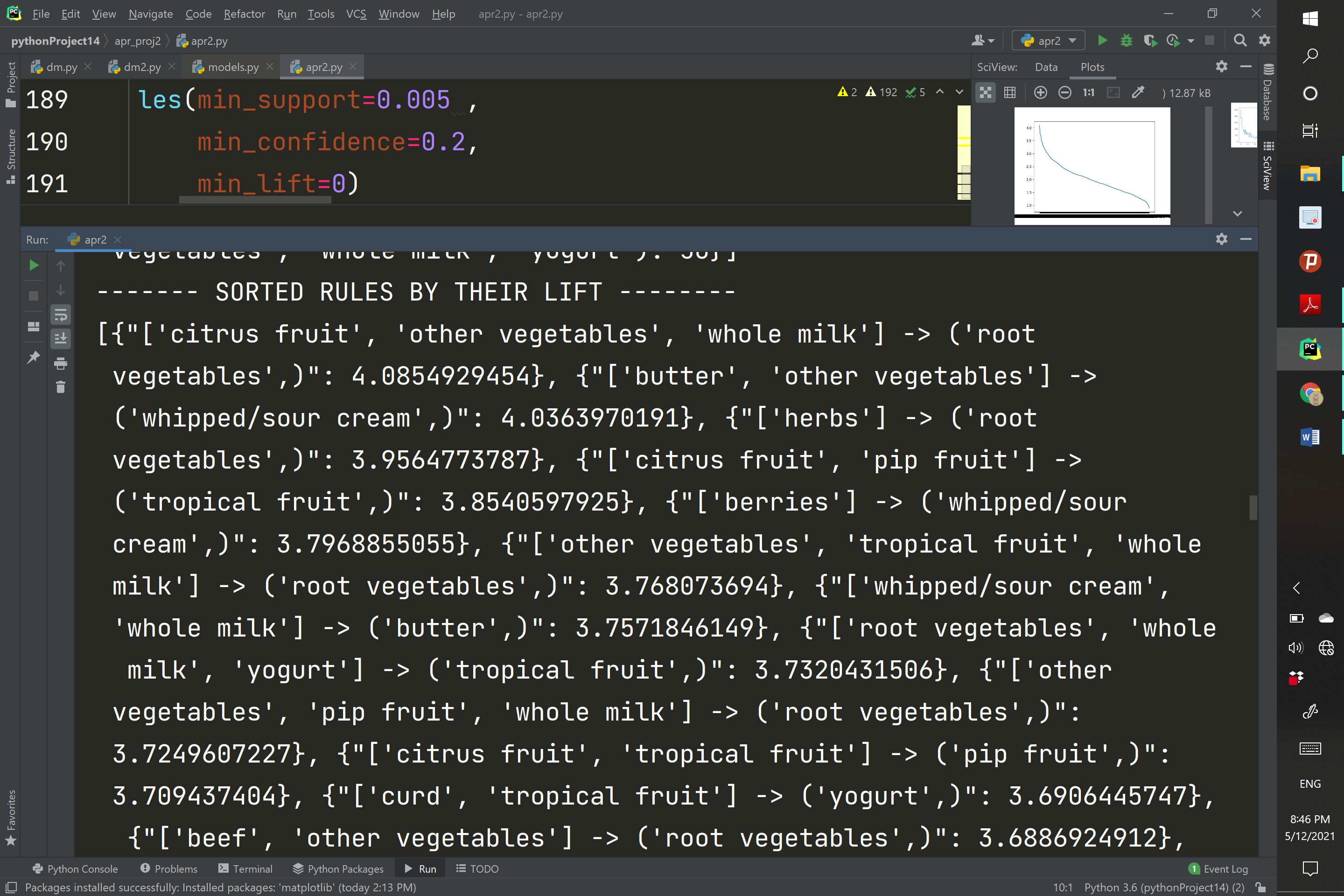
در دیکشنری CONF هر قانون تایید شده با مقدار CONF ذخیره میشود.

و در دیکشنری SUP هر قانون تایید شده با مقدار SUPPORT اش ذخیره میشود.

خروجی ها در کنسول شامل: لیست آیتم‌های پرتکرار به همراه تعداد تکرارشان، لیست مجموعه‌های پرتکرار به همراه تعداد تکرارشان و لیست قوانین استخراج شده با شرایط ‌min\_lift , min\_conficence , min\_support به همراه مقدار لیفشان به صورت نزولی.

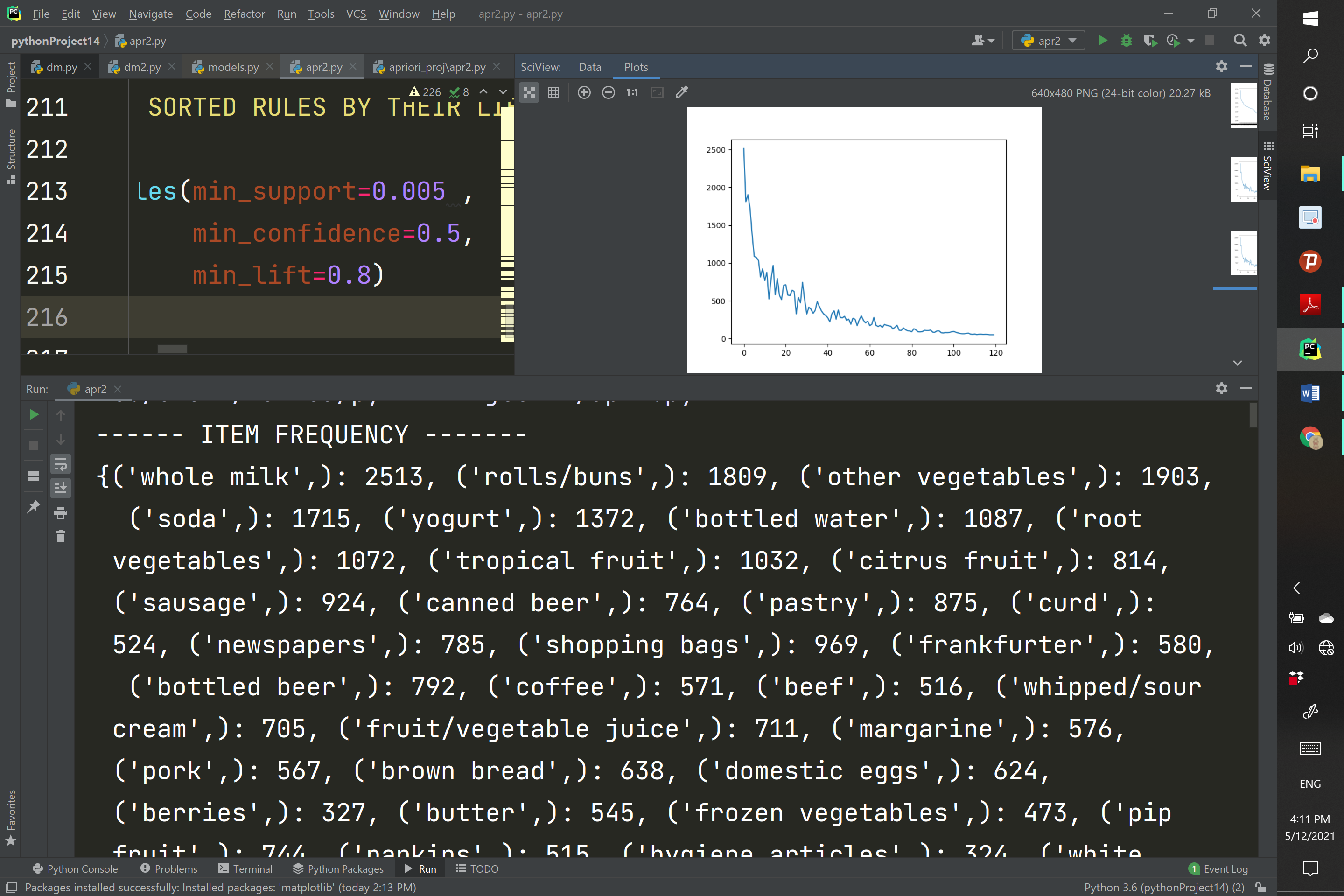
تحلیل:

خروجی خواسته شده در داک)

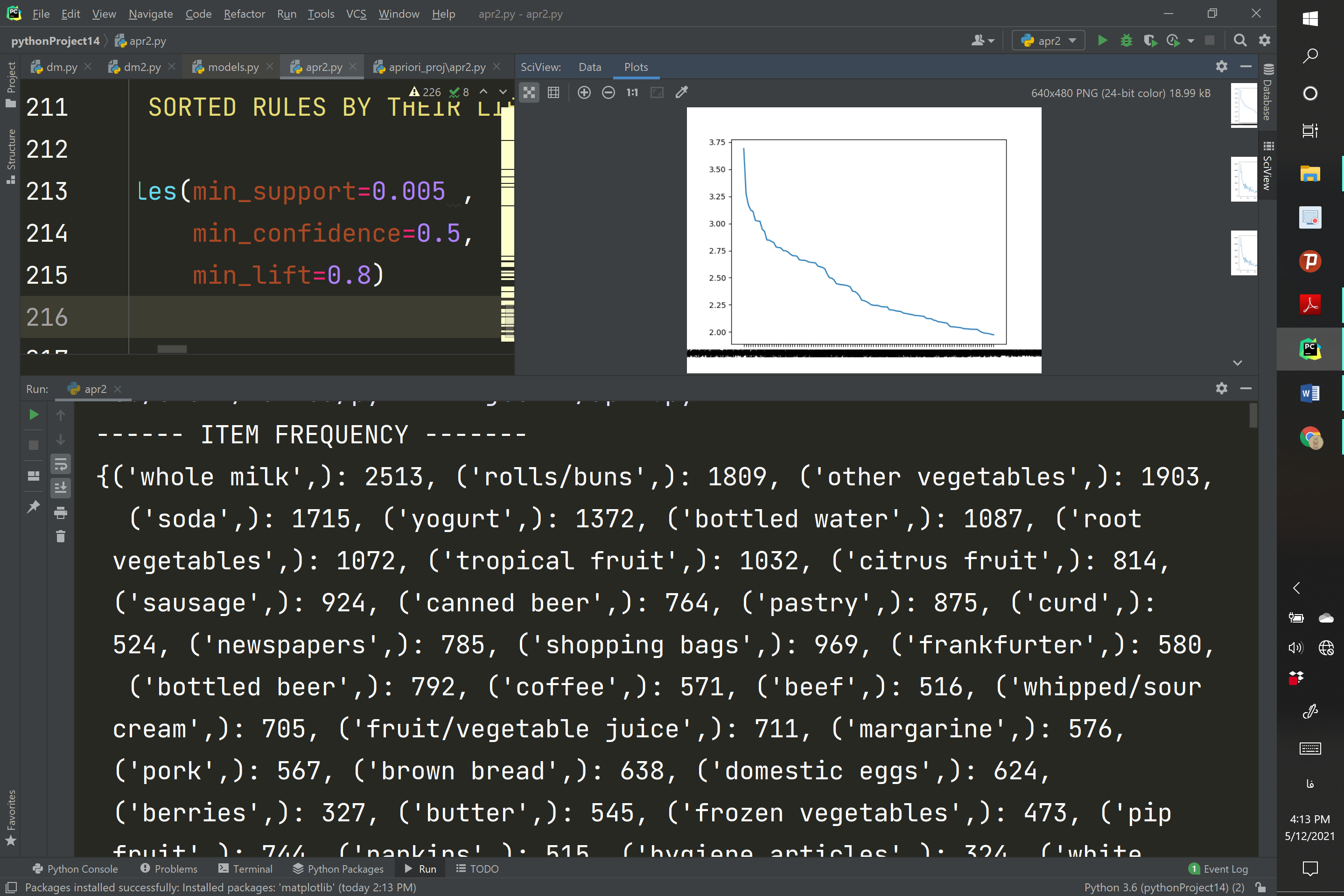


تست ۱ ) در این تست مقادیر ورودی مشابه کد زیر داده شده.

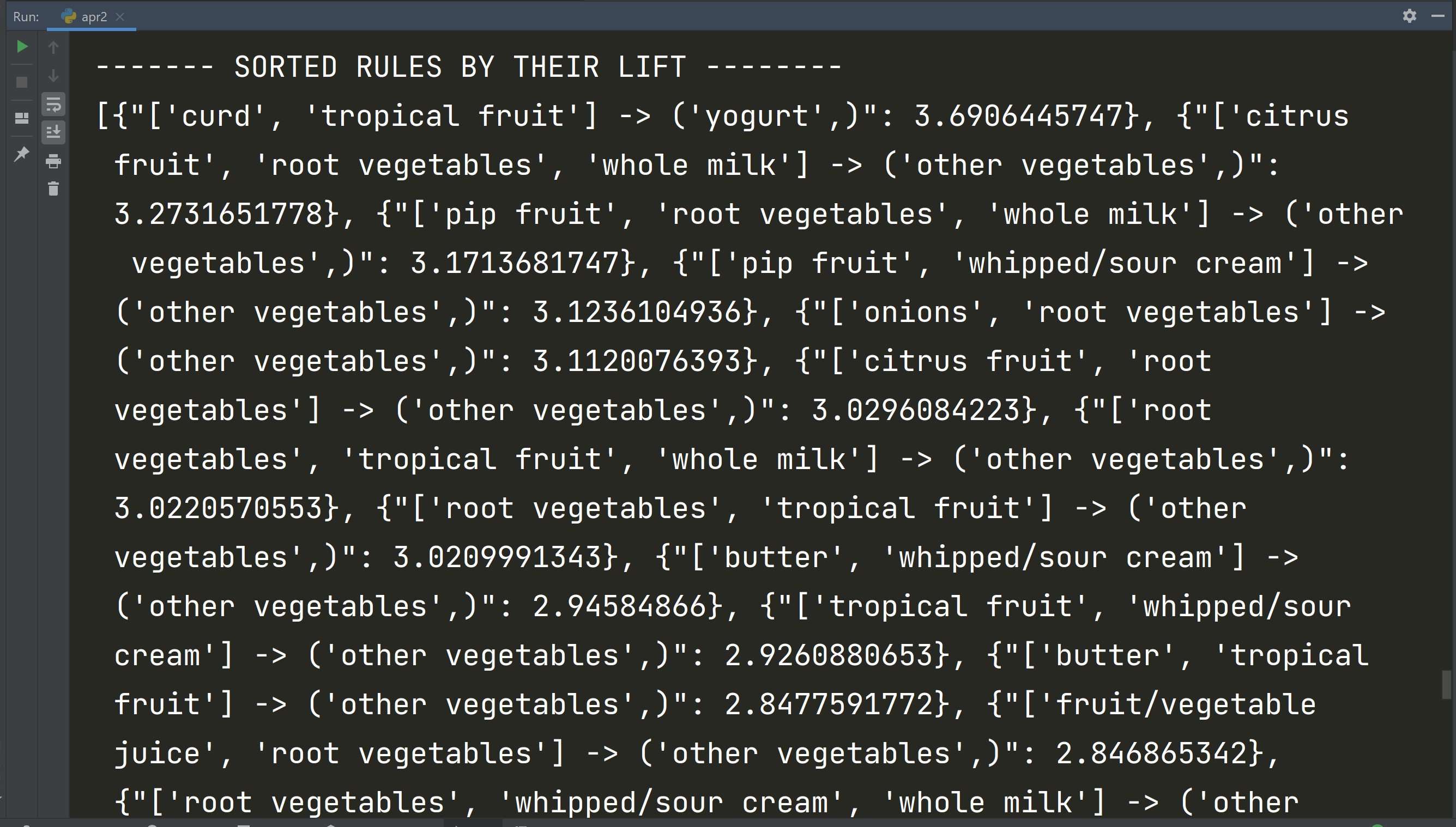
و در نمودار اولیه مقدار تکرار هر آیتم امده.



در نمودار ثانوی، مشابه شکل زیر مقدار لیفت قوانین استخراج شده نمایش داده شده است.



نتیجه بدست امده از تست بالا این است که قوانین استخراج شده با شرایط فوق دارای حداقل ساپورت ۱.۹ شدند که به این معناست که قوانین قوی هستند. در زیر قسمتی از این قوانین به همراه مقدار لیفتشان نشان داده شده است.

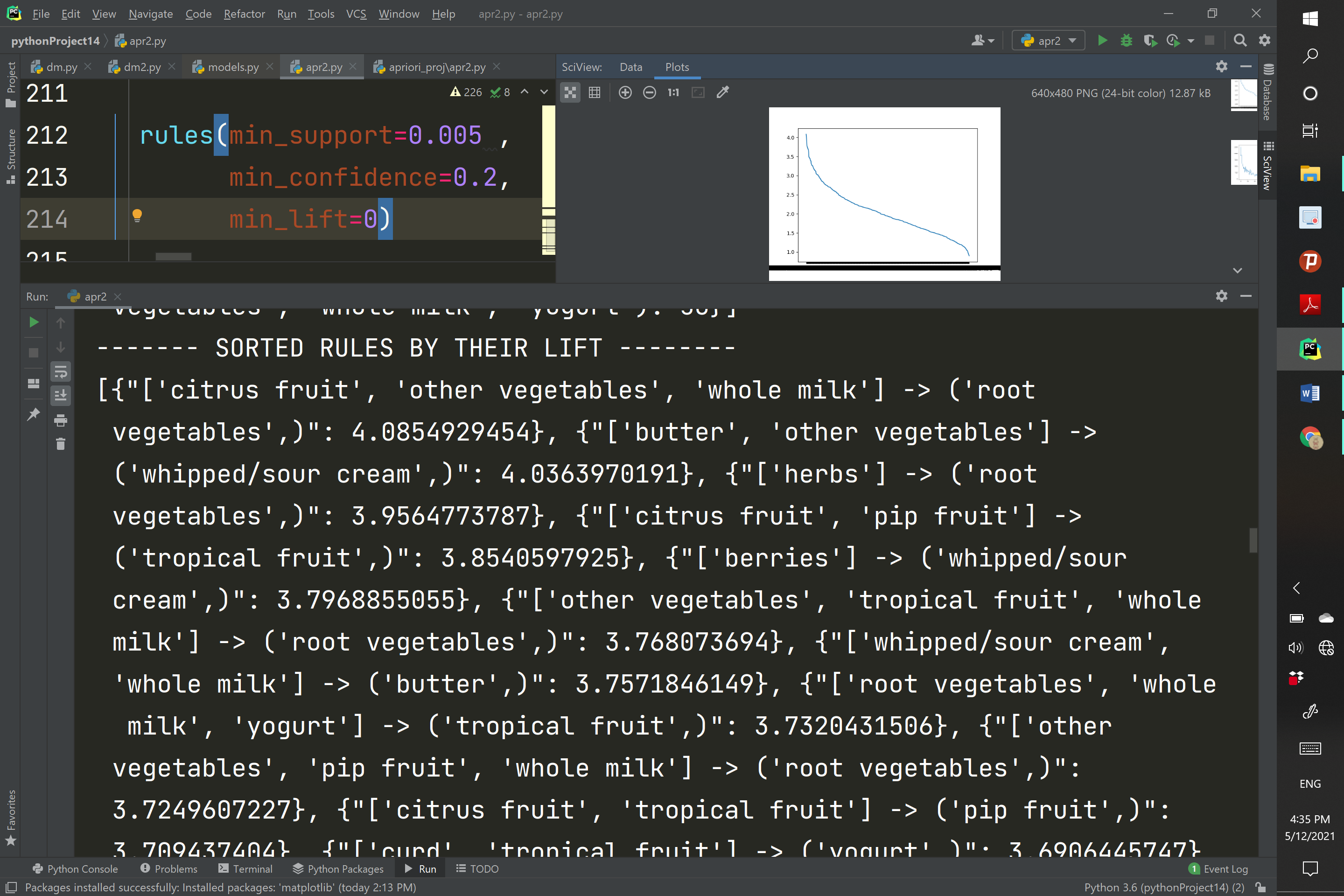


قوی ترین این قوانین :

Curd , tropical fruit -> yogurt

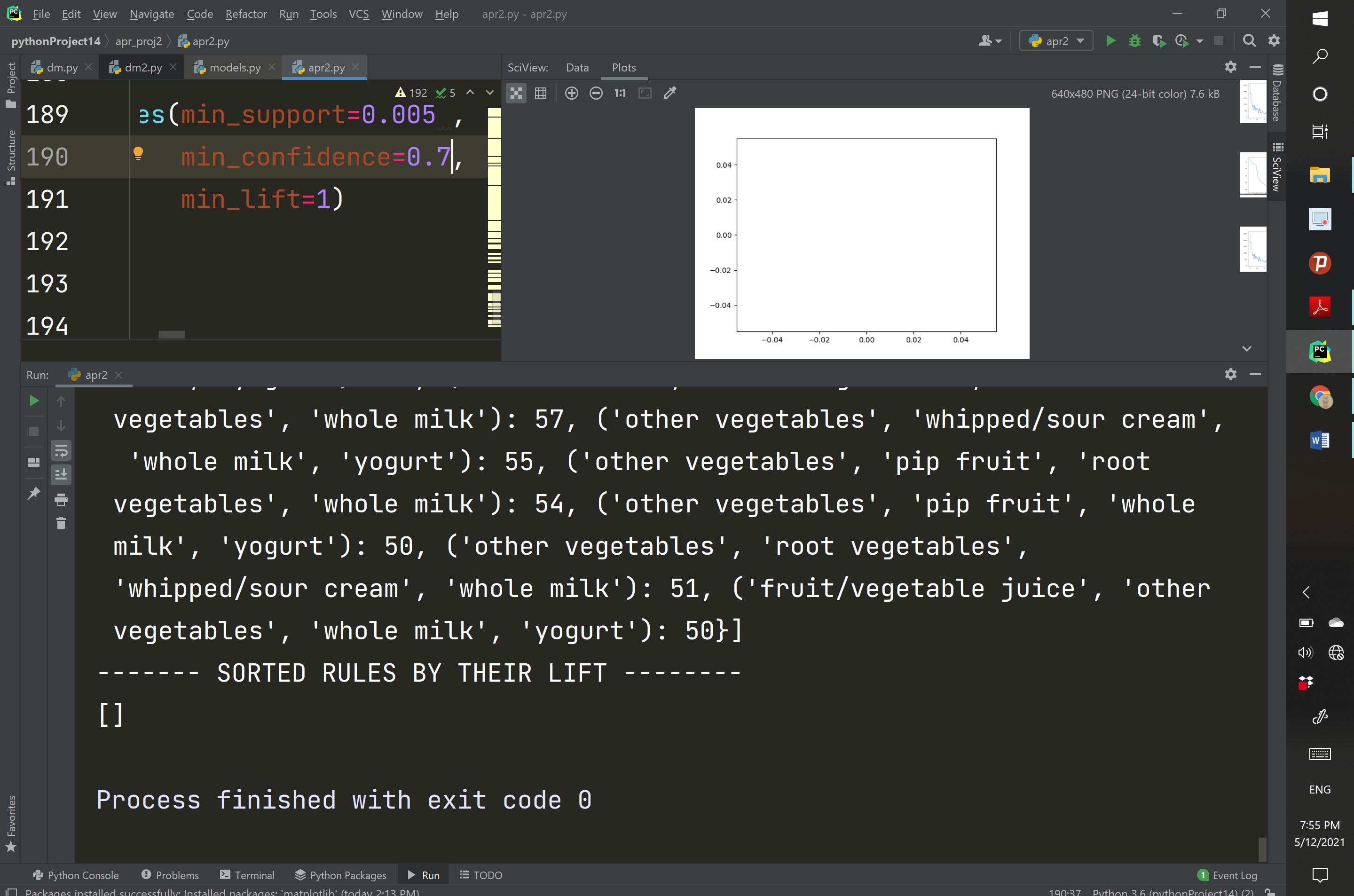
که لیفتش برابر ۳.۶۹۰ میباشد.

تست ۲)



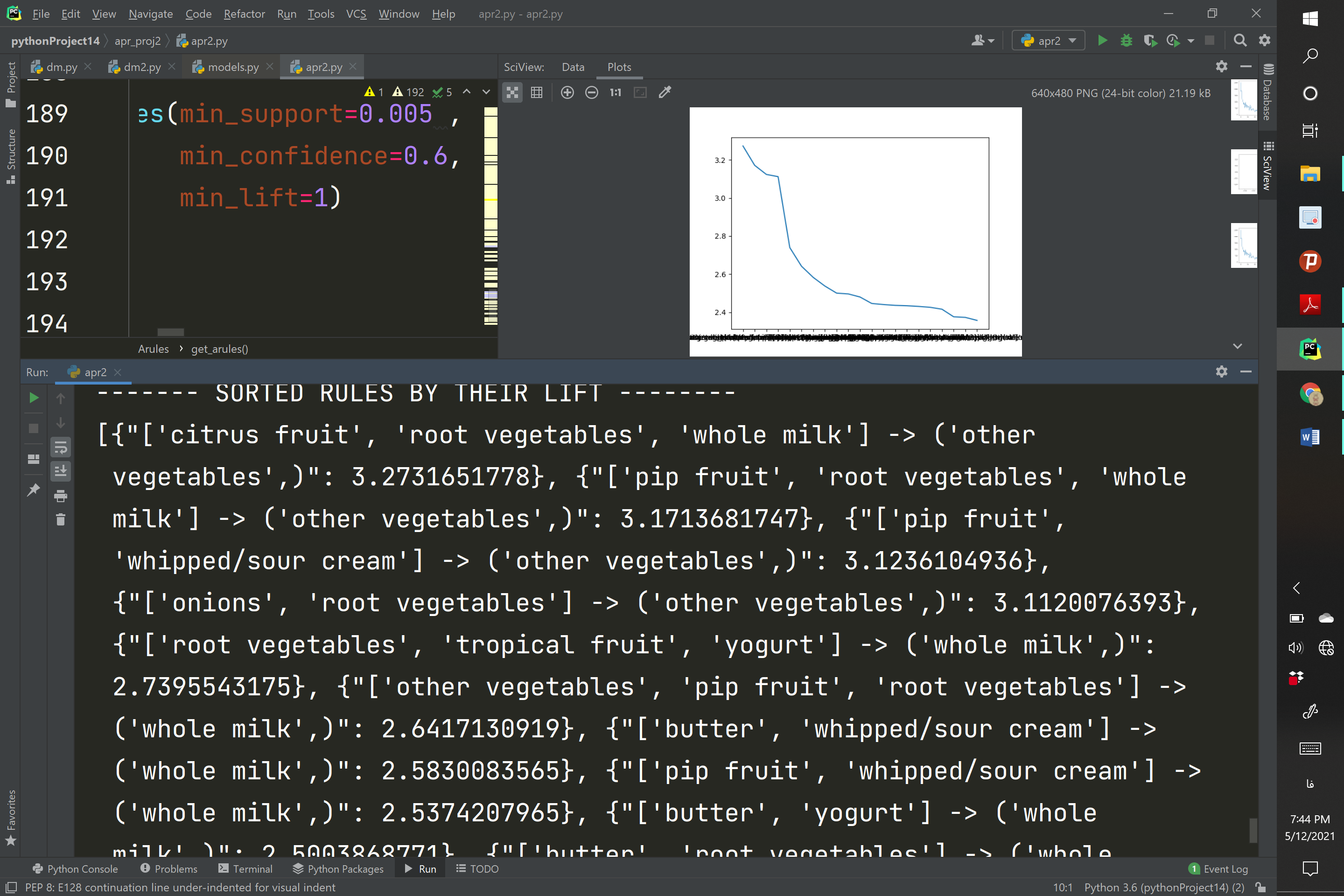
در مثال بالا با شرایط داده شده در ورودی بالا. بیشترین مقدار لیفت برابر ۴ شده است.

تست ۳)

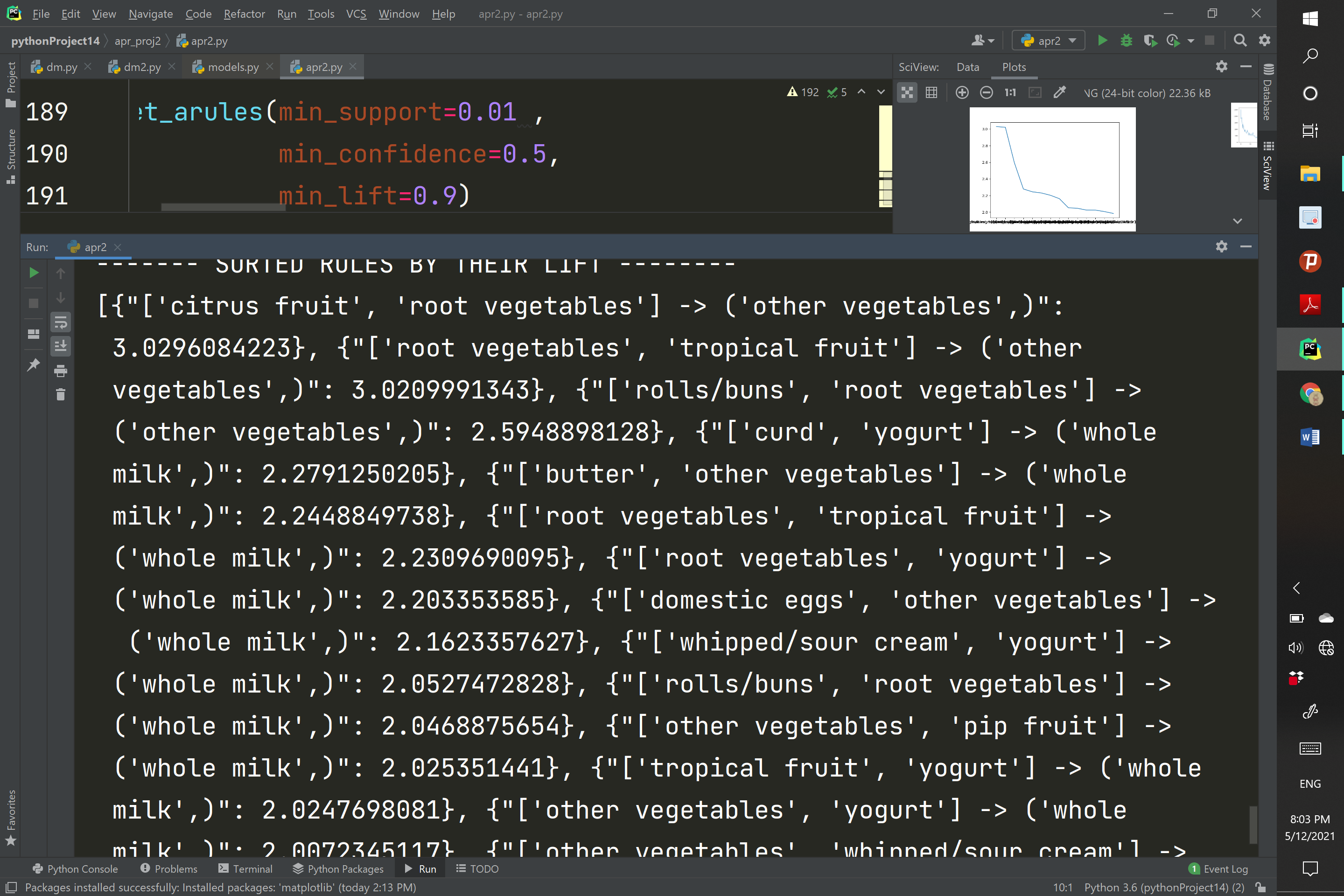


در شرایط این تست هیچ قانونی یافت نشده.

تست ۴)



تست ۵)



"['citrus fruit', 'root vegetables'] -> ('other vegetables',)": 3.0296084223

"['root vegetables', 'tropical fruit'] -> ('other vegetables',)": 3.0209991343

"['rolls/buns', 'root vegetables'] -> ('other vegetables',)": 2.5948898128

"['curd', 'yogurt'] -> ('whole milk',)": 2.2791250205

"['butter', 'other vegetables'] -> ('whole milk',)": 2.2448849738}

"['root vegetables', 'tropical fruit'] -> ('whole milk',)": 2.2309690095

"['root vegetables', 'yogurt'] -> ('whole milk',)": 2.203353585

"['domestic eggs', 'other vegetables'] -> ('whole milk',)": 2.1623357627

"['whipped/sour cream', 'yogurt'] -> ('whole milk',)": 2.0527472828

"['rolls/buns', 'root vegetables'] -> ('whole milk',)": 2.0468875654

"['other vegetables', 'pip fruit'] -> ('whole milk',)": 2.025351441

"['tropical fruit', 'yogurt'] -> ('whole milk',)": 2.0247698081

"['other vegetables', 'yogurt'] -> ('whole milk',)": 2.0072345117

"['other vegetables', 'whipped/sour cream'] -> ('whole milk',)": 1.9843854212

قوانین بالا به همراه تکرارشان، قوانینی هستند که با شرایط این تست استخراج شده اند و قوانینی هستند که هم لیفت بالایی دارند و هم مینیمم ساپورتشان برابر ۰.۰۱ است و min\_conf=0.5 است.

در نهایت میتوان قوانین بالارا به عنوان قوانین قوی در این مجموعه سبد خرید ۹۸۰۰ تایی بحساب اورد.

این قوانین منجر به خرید whole milk یا other vegetable میشوند.