Nama : Muhammad Ilham Isfadhillah

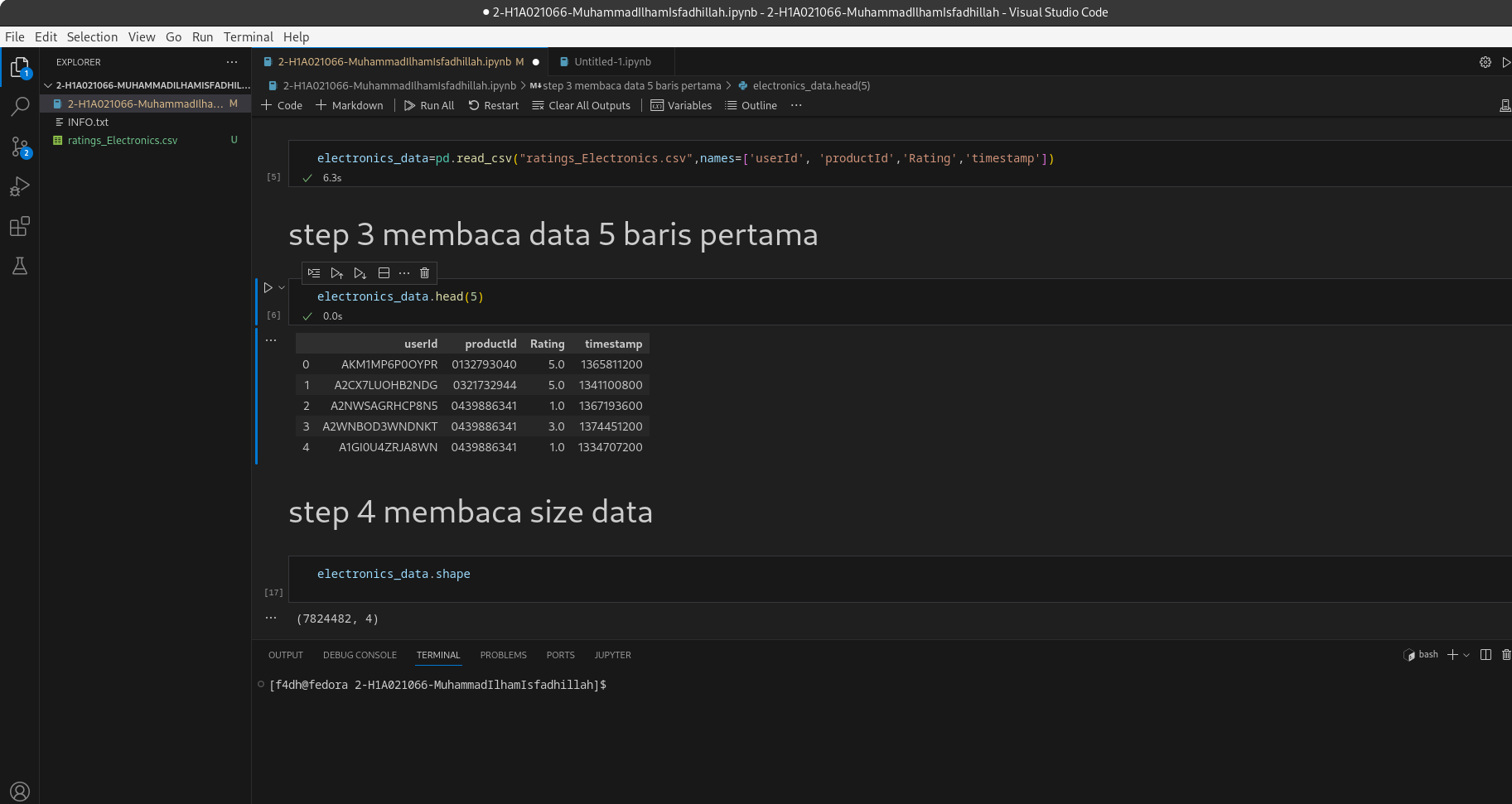
NIM : H1A021066

BAB : 2

# Laporan Real-Time Recommendation Practice for Retail Products

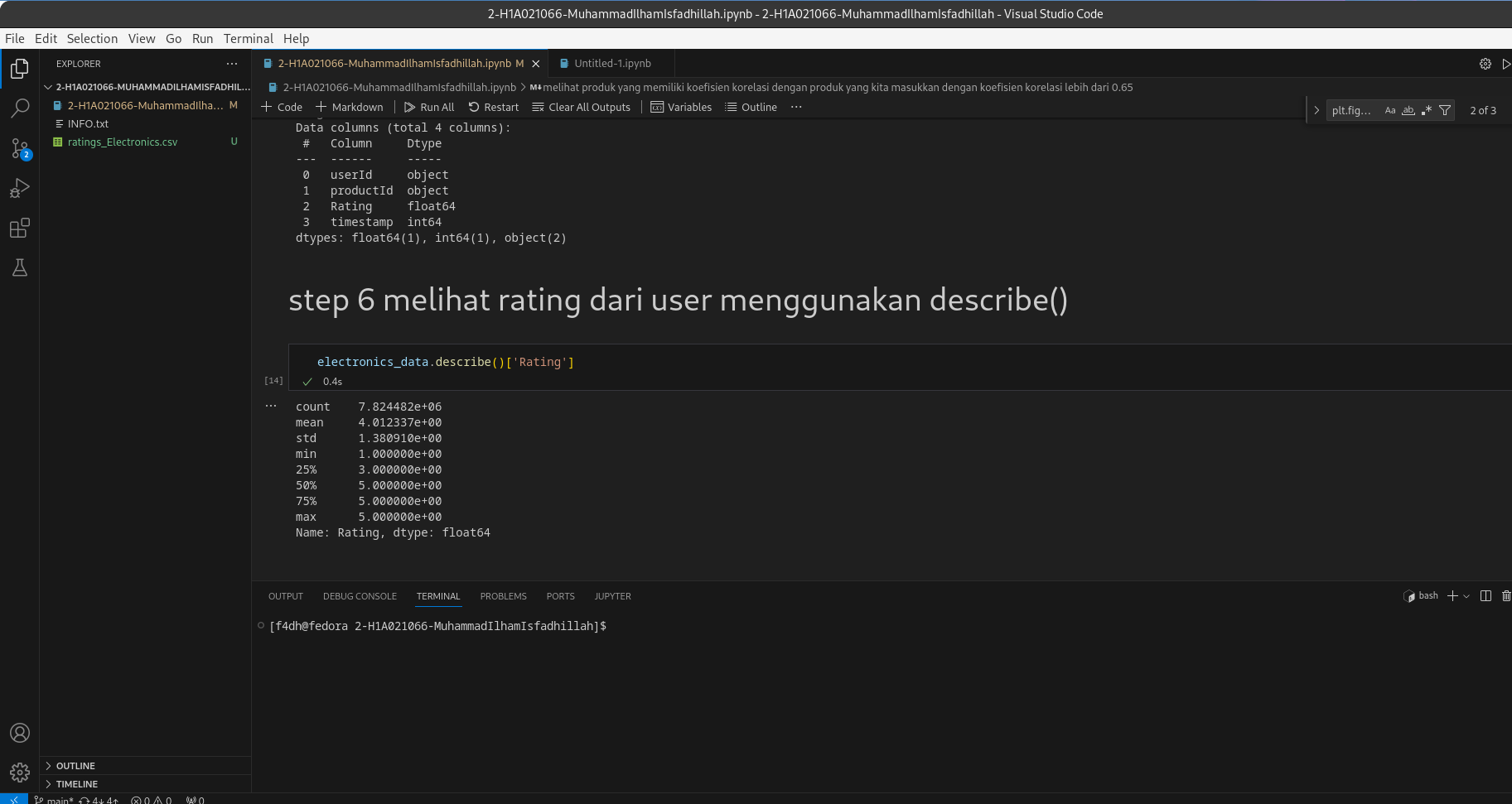
**2.2.1 Preparing E- Commerce Platform Data**

Pada Percobaan ini, disini kami belajar apa saja tahapan awal dalam membuat recommendation practice for Retail Products secara Real Time. Pada percobaan ini berfokus bagaimana kita menyiapkan sebuah platform data yang berada di e-commerce untuk kita olah. Disini kami memanggil beberapa library yang akan digunakan untuk membangun projek ini, diantaranya yaitu NumPy , Pandas , Matplotlib dan Seaborn. Namun, untuk pertama kali kita perlu menginstall library tersebut sebelum melakukan import , yaitu dengan menggunakan syntax pip install <namalibrary> , kemudian setelah itu barulah dapat digunakan. Setelah itu kita perlu membaca dataset yang kita sediakan menggunakan fungsi yang disediakan library dari pandas , yaitu pd.read.csv() yang nantinya akan kita simpan pada sebuah variabel. Kemudian, untuk melihat baris dataset dalam variabel tersebut bisa kita gunakan object method pada variabel tersebut yaitu head(). Maka nanti akan tampil datanya seperti berikut.

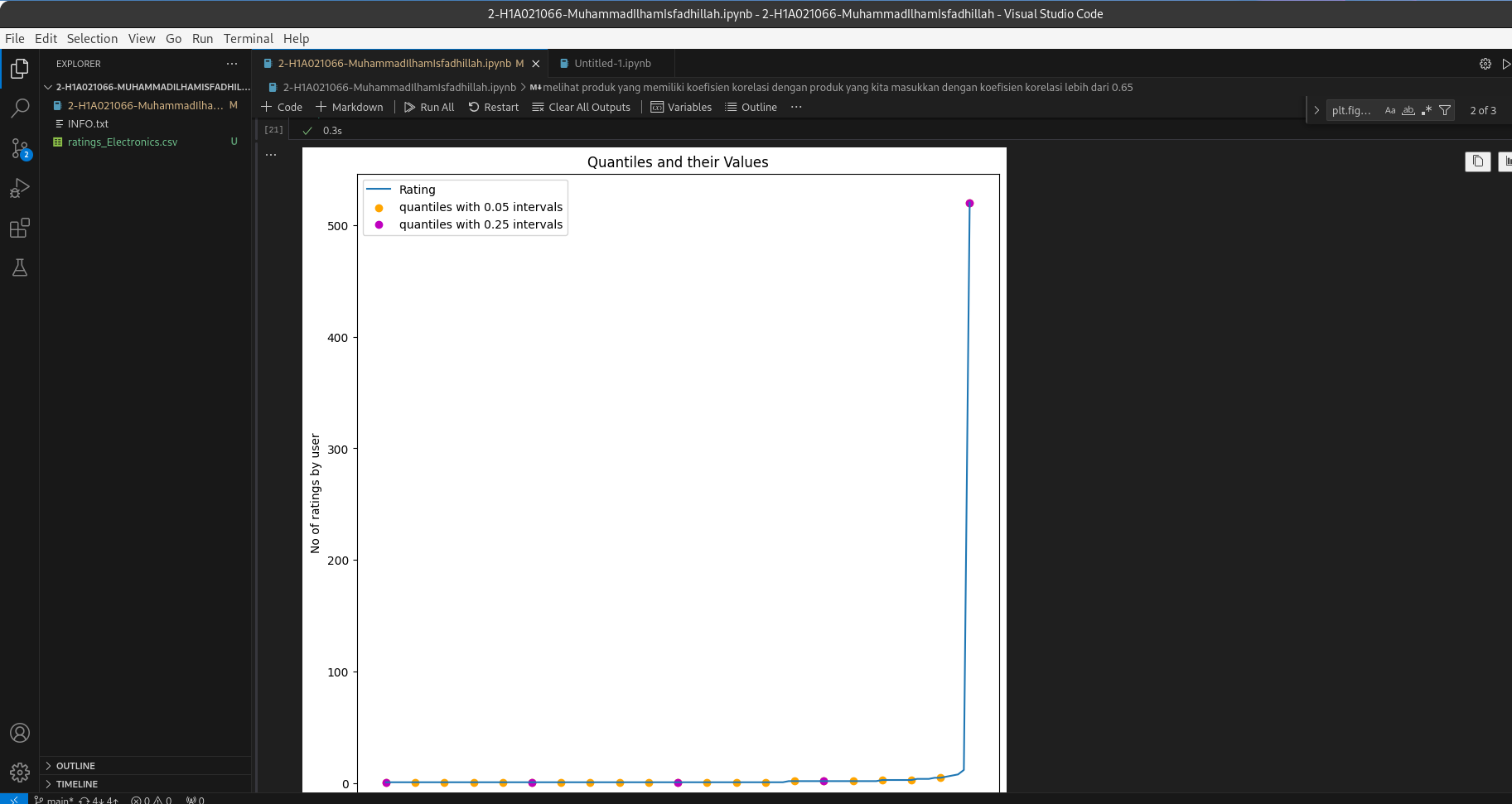


*Gambar 1 menampilkan data frame menggunakan head()*

Kemudian, selain itu juga kita dapat melihat berapa baris dan kolom yang tersedia pada dataset kita, yaitu dengan cara menggunakan method shape() pada variabel dataset kita , dan akan tertampil untuk indeks ke 0 berupa jumlah baris dan indeks ke 1 berupa jumlah kolom. Kemudian kita juga dapat melihat tipe data dari setiap data kolom dengan menggunakan dtypes() , sehingga ditampilkan tipe data dari kolom userID, productID, Rating dan juga timestamp. Kemudian untuk melihat informasi umum tentang variabel dataset kita, kita menggunakan method info(() , sehingga akan dilihat secara detail informasi dataset yang kita yang berada pada DataFrame. Kemudian, jika kita ingin menghapus beberappa informasi dari method info(), kita dapat menggunakan memberikan nilai False dalam parameter mengenai data yang tidak ingin kita tampilkan. Kemudian, dengan bantuan library pandas juga kita dapat melihat informasi lebih detail mengenai kolom tertentu. Misal kita ingin mengetahui tentang kolom Rating, makan dengan method describe() kita dapat melihatnya, yaaitu sebagai berikut

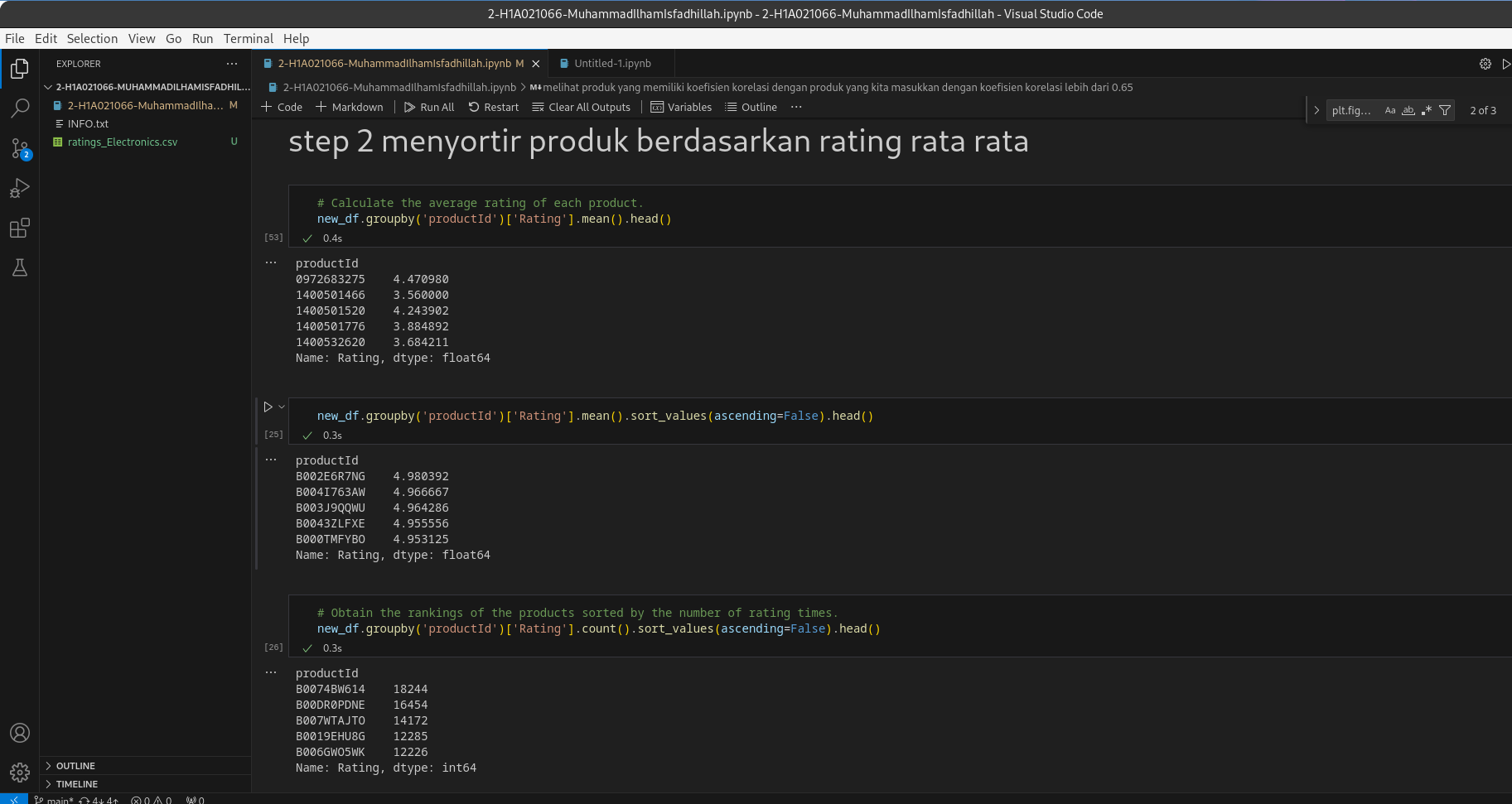
*Gambar 2 menunjukkan informasi dari kolom rating menggunakan describe()*

Kemudian pada tahapaan ini juga kita dapat melihat nilai tertinggi dan nilai terendah dari kolom Rating dengan cara menambahkan method min() dan max() , yang mana ini sangat berguna jika kita ingin mengetahui secara cepat berapakah nilai terbesar dan terkecil dibandingkan kita diharuskan untuk ngescroll dan menceknya satu satu. Kemudian juga , kita dapat mengecek apakah terdapat parameter yang null dan akan kita dapat menggunakan sum() untuk menghitung jumlah yang null jika ada. Ini merupakan tahapan awal kita untuk mengecek apakah ada data yang null sebelum data tersebut kita proses dalam bentuk grafik untuk menghindari error yang terjadi. Kemudian untuk mengecek ada berapa sih data mengenai kolom kolom yang ada di dataset kita , seperti berapa rating yang ada, berapa produk yang ada dan berapa user yang ada di dataset kita, kita dapat menggunakan len(np.unique()) dari kolom yang ingin kita coba cek. Selain itu juga, semisal dalam sebuah dataset terdapat terlalu banyak informasi yang tidak kita butuhkan, mata kolom data tersebut dapat kita harpusnya menggunakan method drop(), seperti pada percobaan ini , kita menghapus kolom timestamp karena kedepannya tidak terlalu berfungsi saat kita coba olah dalam bentuk grafik. Kemudian, kita juga dapat membuat sebuah pengelompokkan dalam datasest kita, semisal kita ingin menampilkan salah satu kolom dalam dataset kita, kita dapat menggunakan method groupby(). Selain itu, disini kita juga dapat menampilkan sebuah dataset yang data kolomnya telah disortir sesuai kemamuan kita dengan menggunakan sort\_values(). Selain itu, kita juga dapat mendapatkan nilai kuantil dari data kolom di dataset kita menggunakan method quantile().

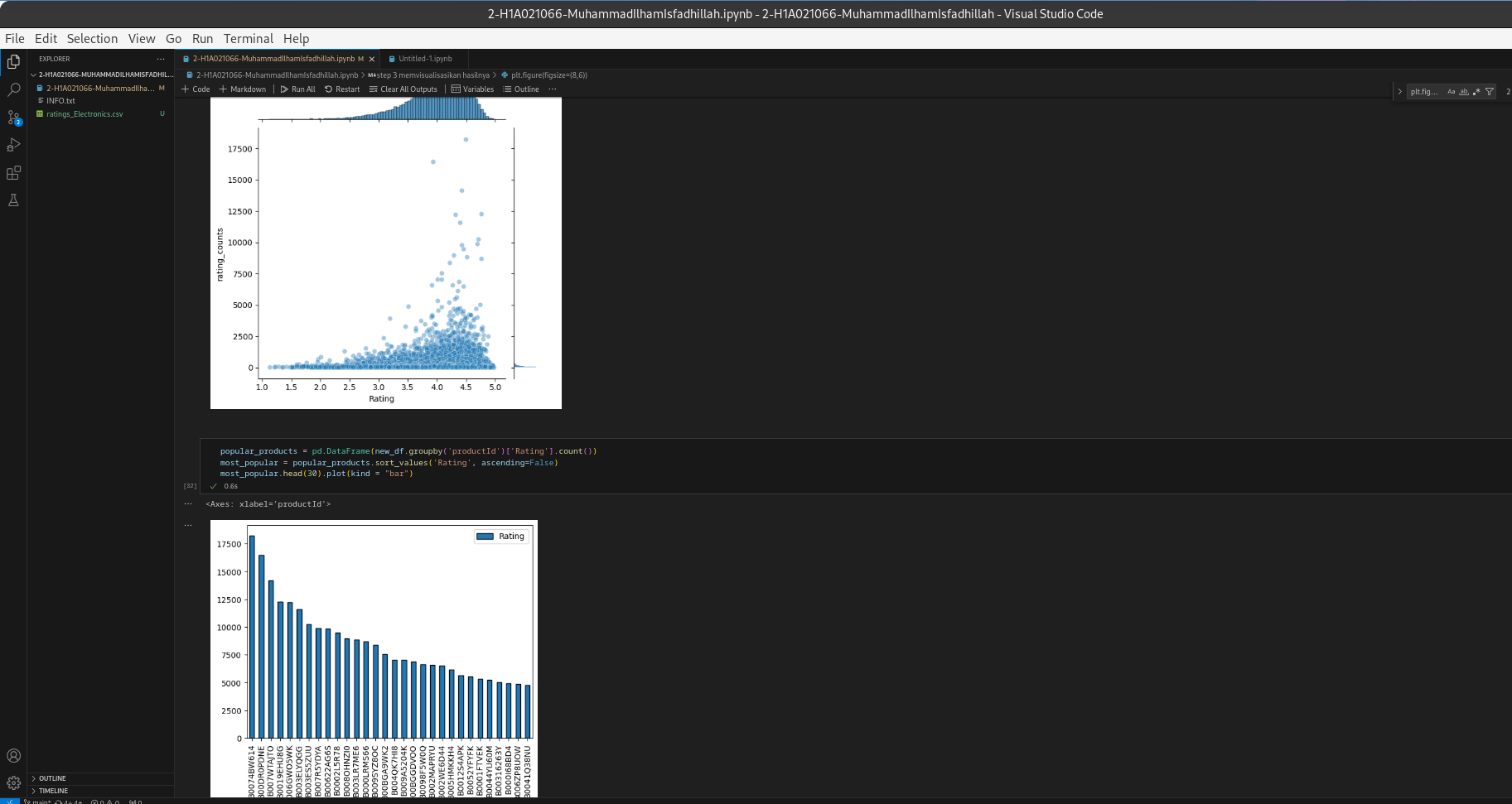
*Gambar 3 Quantiles dari data*

**2.2.2 Recommending Products Based on the Product Popularity**

Kemudian, pada percobaan selanjutnya kita mulai masuk mencoba untuk membuat sebuah rekomendasi produk berdasarkan popularitas produk yang ada di dataset kita. Langkah yang perlu kita lakukan disini adalah melakukan penyortiran produk kita, kita akan menyortir sebuah produk yang memiliki Rating lebih dari dan sama dengan 50 kali dan kemudian menyortirnya berdasarkan Rating terendah terlebih dahulu dan melakukan plotting grafiknya. Kemudian, kita dapat melakukan penyortiran berdasarkan rata rata rating, jadi disini kita membuat sebuah tabel yang terdiri dari sebuah produk dan rata rata ratingnya dengan menggunakan groupby() dan mean() atau menyortir produk berdasarkan produk yang sering dirating yang nanti hasil akhirnya dapat kita sortir untuk menjadi lebih rapih.

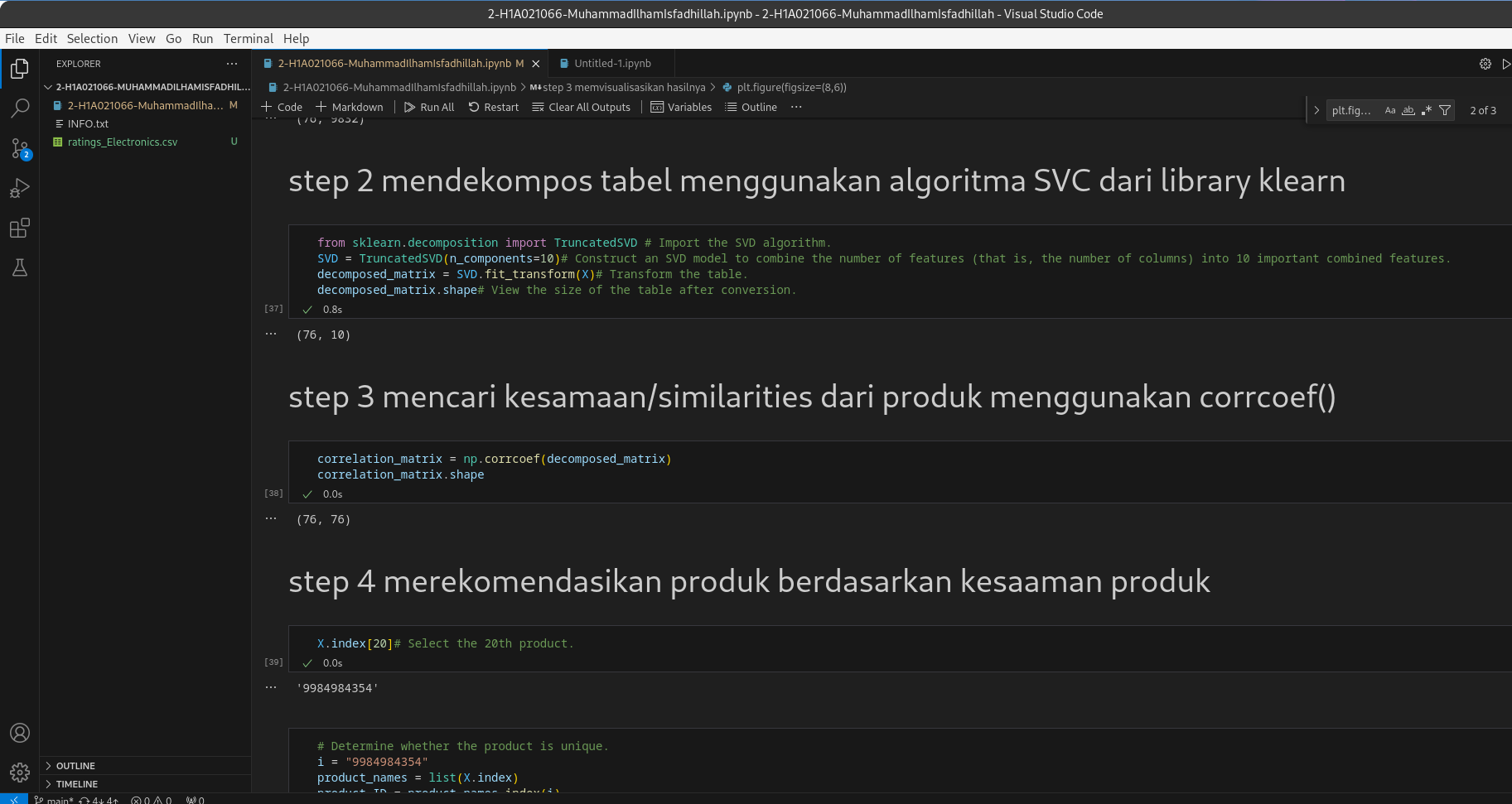
*Gambar 4 pensortiran produk berdasarkan rata rata dari rating*

Kemudian, kita juga dapat menambahkan kolom baru dalam dataset yaitu dengan menambahkan jumlah rating / rating\_counts dalam pada tabel dataset. Setelah itu, kita dapat memvisualisasikan hasilnya menggunakan hist() , yang mana parameter di dalamnya dapat kita ganti nilai bins nya.

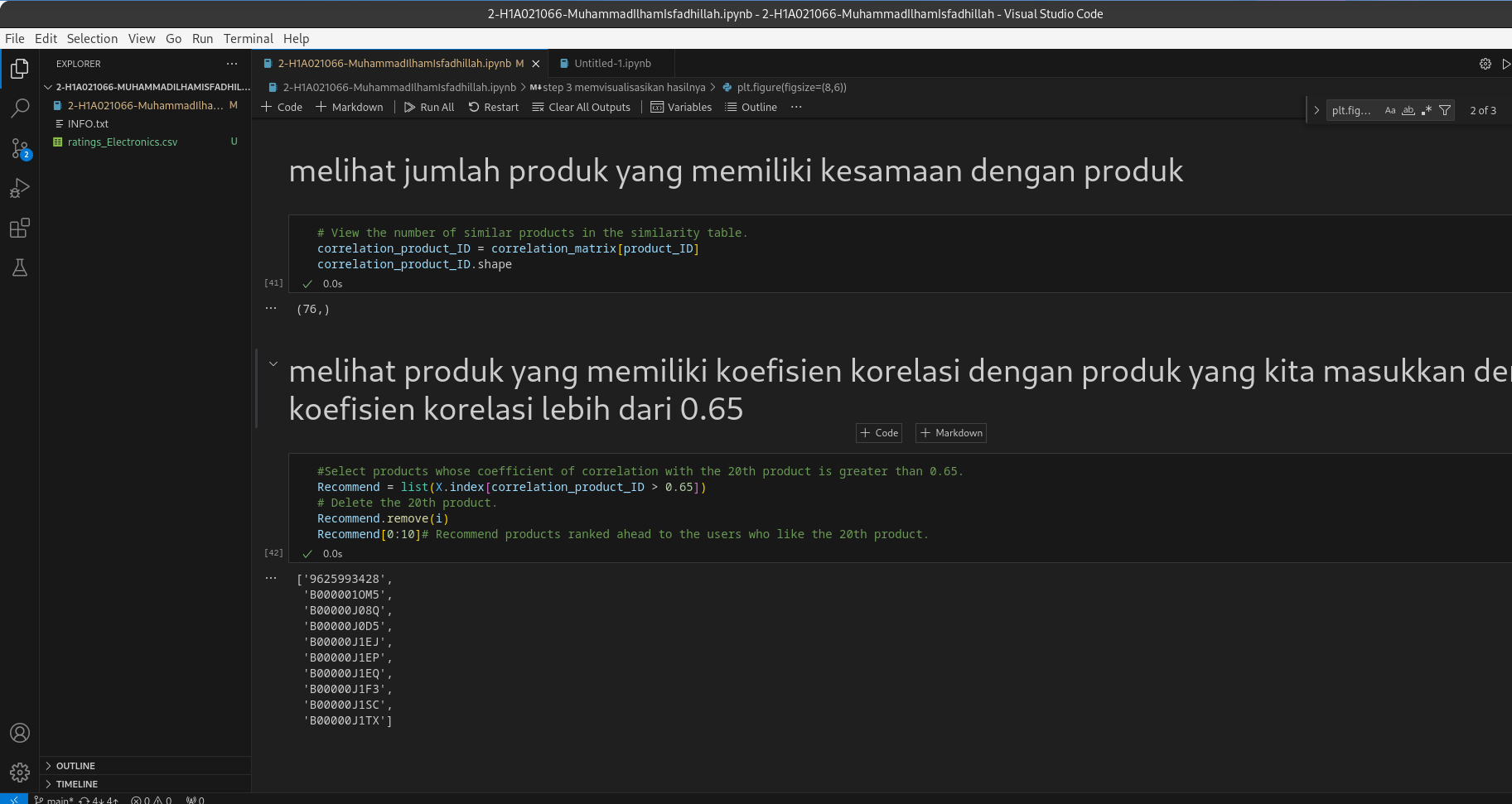
*Gambar 5 membentuk histogram dengan bins = 50 dari data frame*

**2.2.3 Recommending Products Based on Collaborative Filtering**

Kemudian, pada percobaan terakhir barulah kita membuat sebuah rekomendasi produk berdasarkan pada Filtering Kolaboratif , dimana tahapannya disini kita perlu memilih 10000 sampel dan menggunakan method pivot\_table() untuk membuat sebuah hubungan antara productID dan userID dan di dalamnya kita dapat memberikan default value pada kolom yang berupa NaN dengan mengubah nilai fill value pada parameter pivot\_table() seperti pada gambar berikut. Selain itu kita juga dapat menukar kolom dan baris pada tabel dengan menggunakan object attributes T. Selanjutnya, pada percobaan ini digunakan algoritma Singular Value Decomposition (SVD) untuk mereduksi dimensi tabel tersebut dan mengidentifikasi fitur-fitur produk yang paling penting (10 fitur dalam contoh ini). Dengan mereduksi dimensi, data menjadi lebih ringkas namun masih mempertahankan informasi yang relevan. Langkah berikutnya adalah membangun matriks korelasi antara produk dengan corrcoef().

*Gambar 6 mendekompos tabel menggunakan SVD dan corrcoef()*

Matriks ini mengukur kesamaan antara produk berdasarkan pola rating pengguna. Semakin tinggi korelasi antara dua produk, semakin mirip mereka dalam hal preferensi pengguna. Dalam hal ini, kita dapat merekomendasikan produk-produk kepada pengguna berdasarkan kesamaan produk yang telah mereka sukai sebelumnya. Rekomendasi yang dihasilkan didasarkan pada pola rating produk oleh pengguna lain yang memiliki preferensi serupa, seperti pada gambar berikut

*Gambar 7 mencari produk serupa dengan koefisien korelasi lebih dari 0.65*