**Problem Statement :**

An ecommerce company is struggling to increase its sales revenue and retain customers due to a lack of personalization in their product recommendations. Develop a model that can predict the most relevant products for a user, based on their past purchase behaviour, browsing history, and other relevant characteristics.

**Data Collection:**

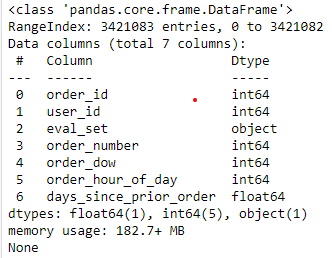
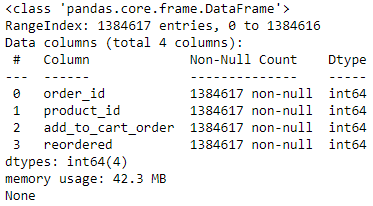
Collected a dataset which suits the above problem statement.

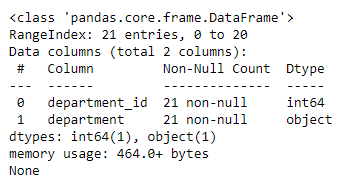
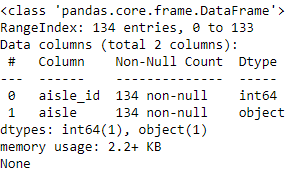
The data set collected was from Kaggle Instacart market basket analysis.

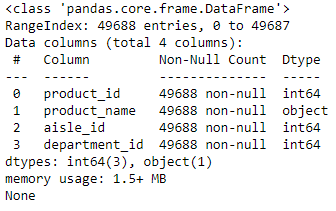
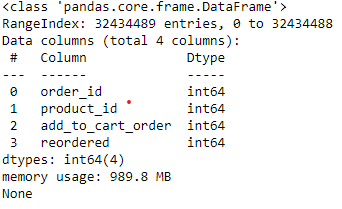
Which consist of 6 different table :

Aisle,department,order\_products\_train,order\_products\_prior,products,orders.

The info of different tables are below.This helps us to understand the contents of the table like number of row’s ,the data types of different columns, and memory utilization.

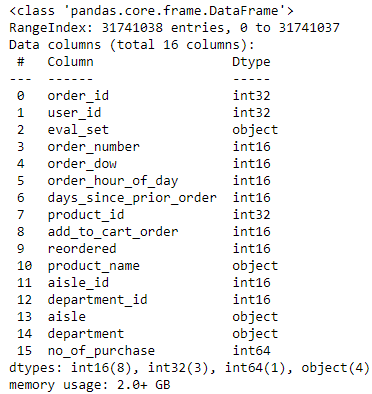
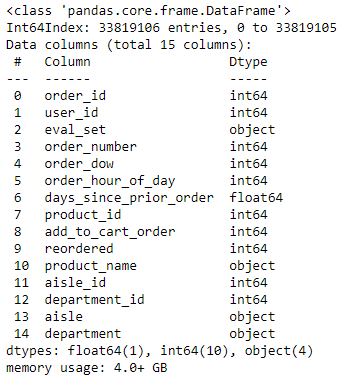
 

**Merging The dataset:**

Using merge method I joined all the tables using the inner.and after compliation there were about 3,17,41,038 records.



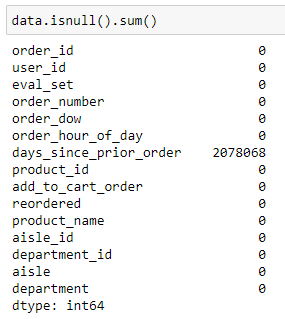
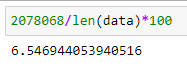
The merged data Utilises almost 4gb of memory size.So inorder to reduce the load on the RAM. I had to perform optimization by down sizing the data types to the required amount.

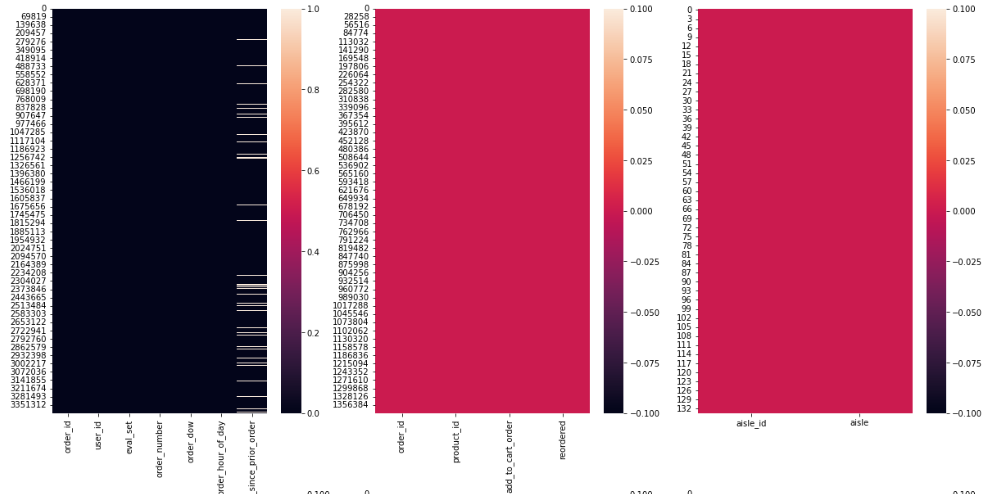
The max value of each column was identified using max function.and corresponding data types were selected.

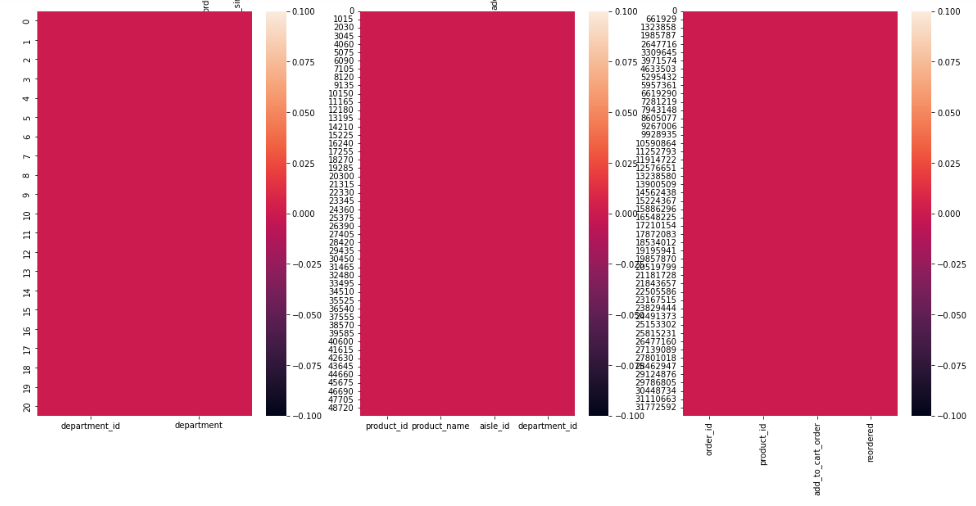
After optimization the memory size was reduced to 2GB.

**Exploratory Data Analysis:**

**Check for missing values:**

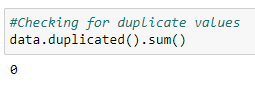
** **

****



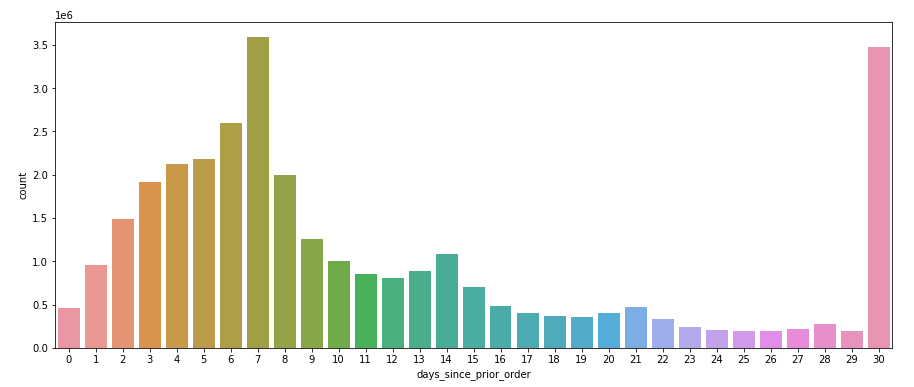
Since only 6% of the data was missing dropping was the technique which I implemented.

**Duplicates:**

 There weren’t any duplicates in the dataset.

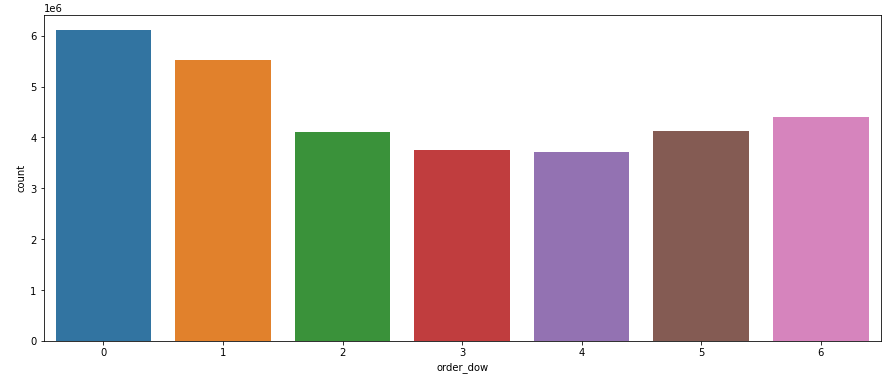
**Insights:**

**Days since Prior order**



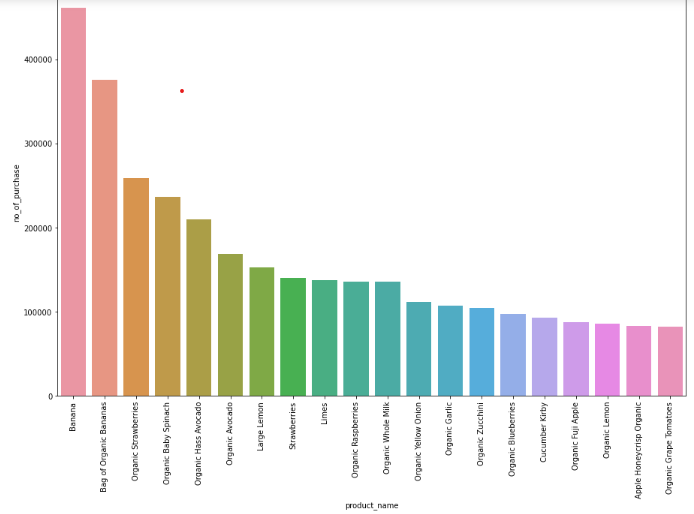
From the above bar plot we can understand that majority of the people visits only once in 31 days. And some on 7th or 8th day after their prior purchase.

**Week wise Purchased pattern:**

****

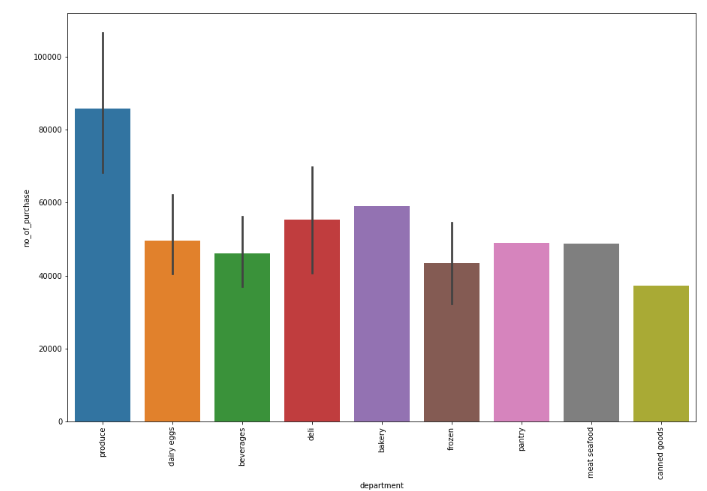
From the above plot we can understand that most of the orders come during the week ends.and with the least number of orders during the middle of the week.

**Most Popular Products**

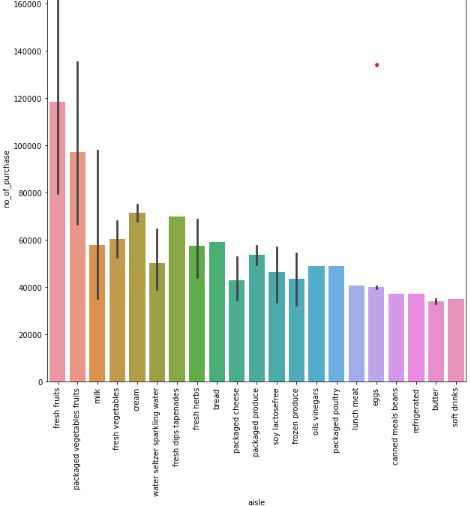


From this chart we can say that banana is the most popularly ordered item in the store.

**Most Busy department:**

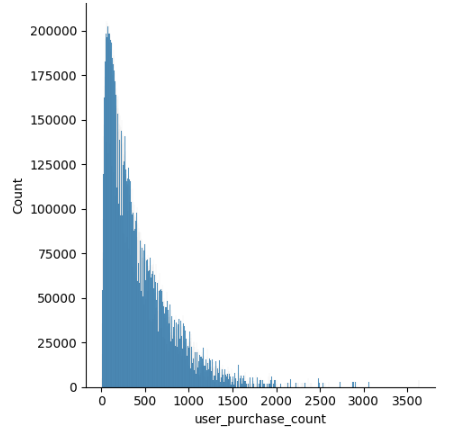
****

The plots tells us that the produce department is one the busiest department of them all.

**Most Busy aisle: **

Here from the plot we can observe that fresh fruits is the most popular aisle.

**User’s Purchase Pattern:**

Here from the plot we can see most of the majority of the customers has made around 300 purchase on an avg.

**Feature Engineering:**

Generated Feature such as:

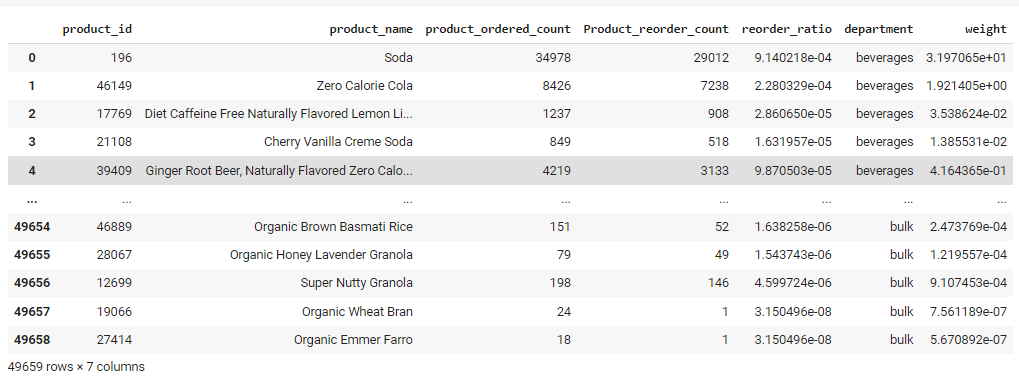
1.**User\_purchase\_count** :- This feature counts the number of purchases each users has made from this E-commerce website. This feature will be used during model building To filter out customer who has made less than 250 orders from this website.

2.**Product\_ordered\_count**:- This feature counts the number of time/fequency of a each specific product has been ordered.This feature is also used to filter out less common product from the website.only products which has been purchased min of 300 orders were taken into the sample data for model building.

3.**Product\_reordered\_count:-** This feature helps us to identify how many time that each product has been re-ordered.

4.**Reorder\_ratio:-** This feature gives us a ratio of number of time that product has been re-ordered divided by total number of re-orders

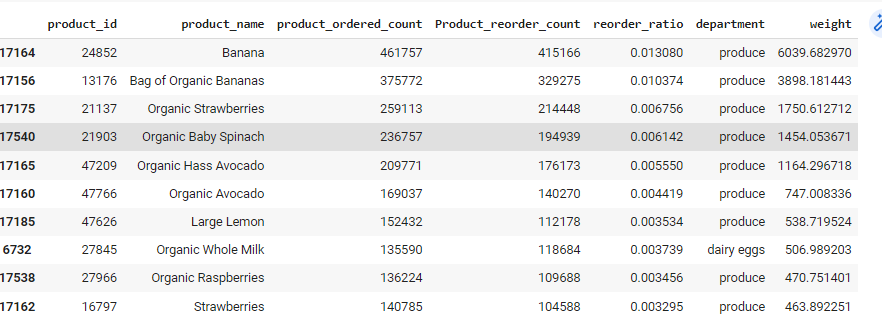
5.**Weight:-** This was used in popularity based model.the product with the highest weight was the most popular product.this weight taken into account of the number of time a product was ordered and re-ordered

****

**Model Building**

**1.Popularity Based model:**

This model made used of the weight feature in the dataframe and top 10 products with the highest weight was named as the most popular products.

****

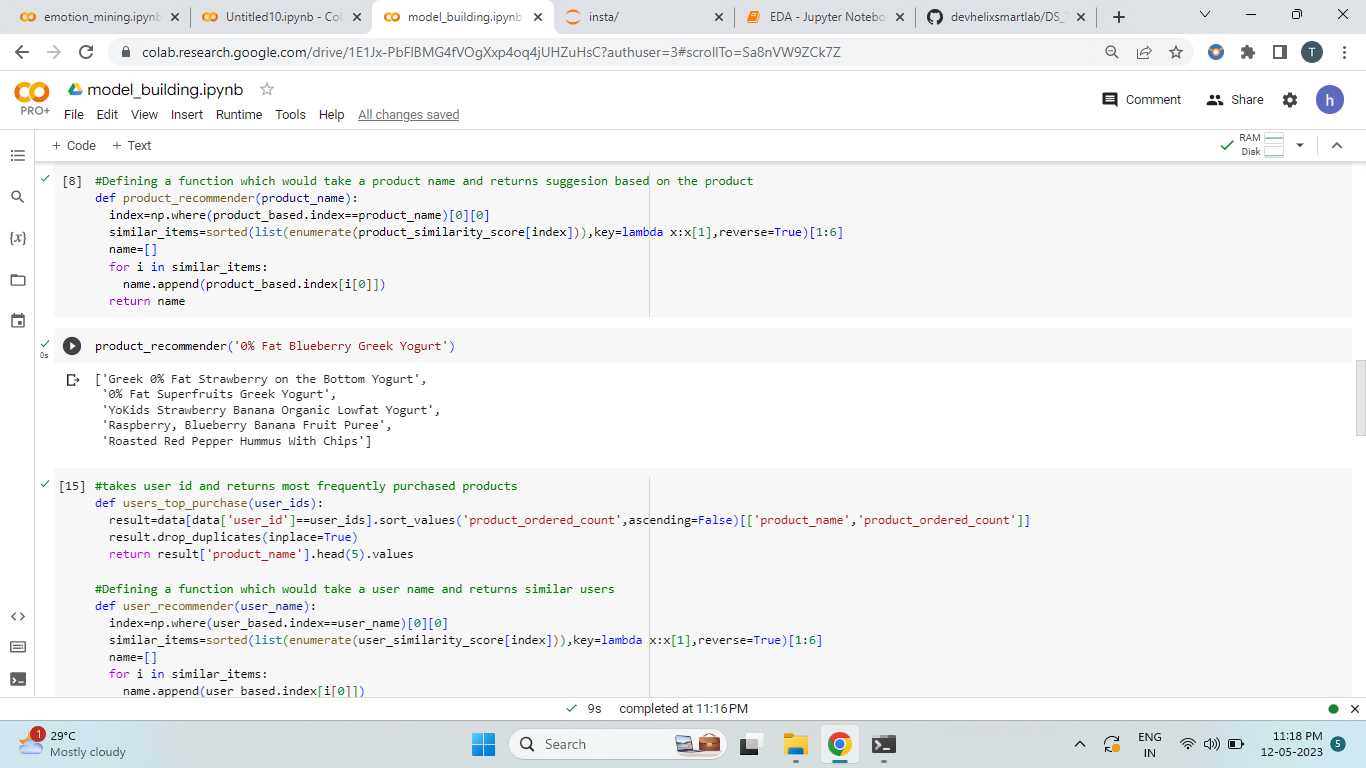
**2.Collaborative Filtering:**

Here user’s who has made a min of 100 order’s and all the products which were purchased min of 250 times and the products which were reordered were taken into account for training the model after filtering there were about 11249 unique products and 87568 unique customer data.

And pivot table was made with index having all the products.while the columns were holding the user-id’s and the values tell wheather that product is purchased (1-for showing that the product was purchased and 0-for not purchasing).and using cosin similarities similar pairs were identified.

Top 5 similar products will be recommended.

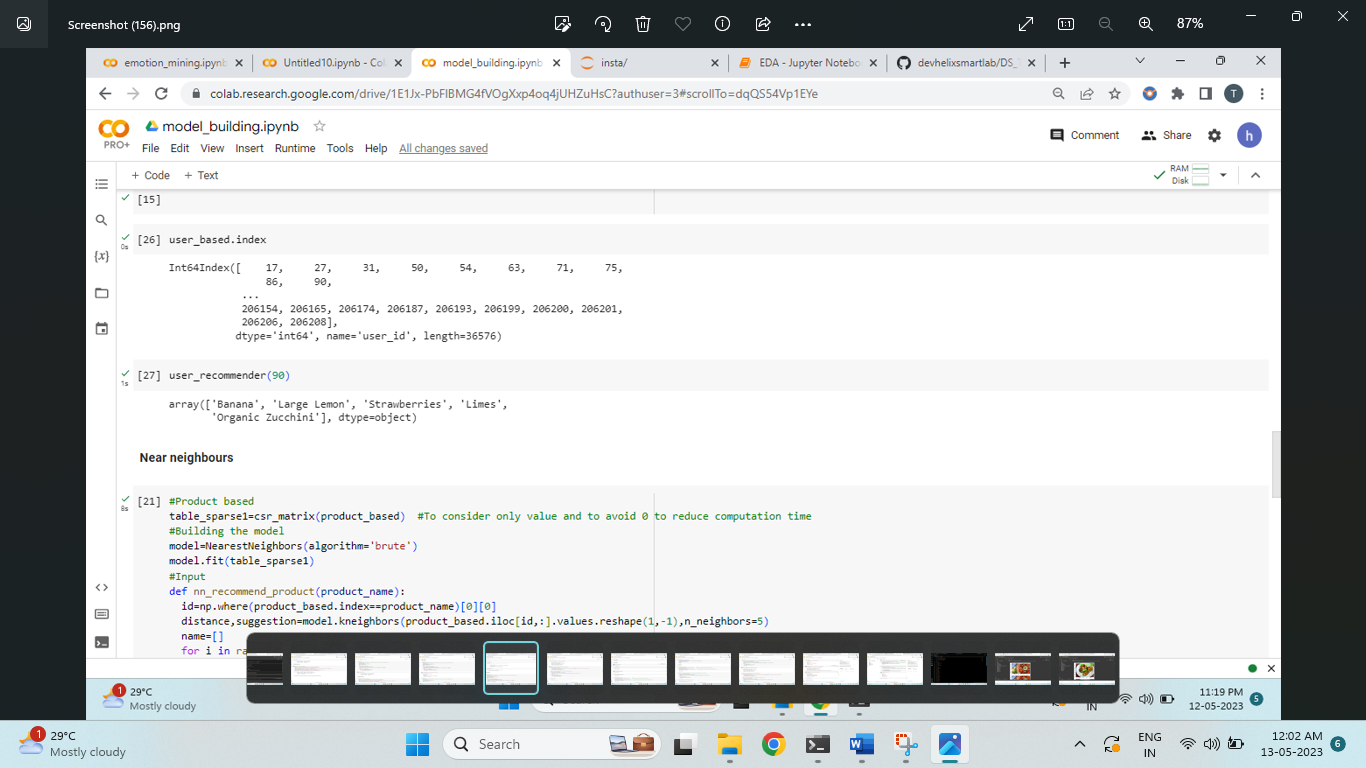
*Output of product based recommendation system*:



*Output of user based recommendation system*:

Here index of the pivot table holds the user -ids and columns holds the product name.

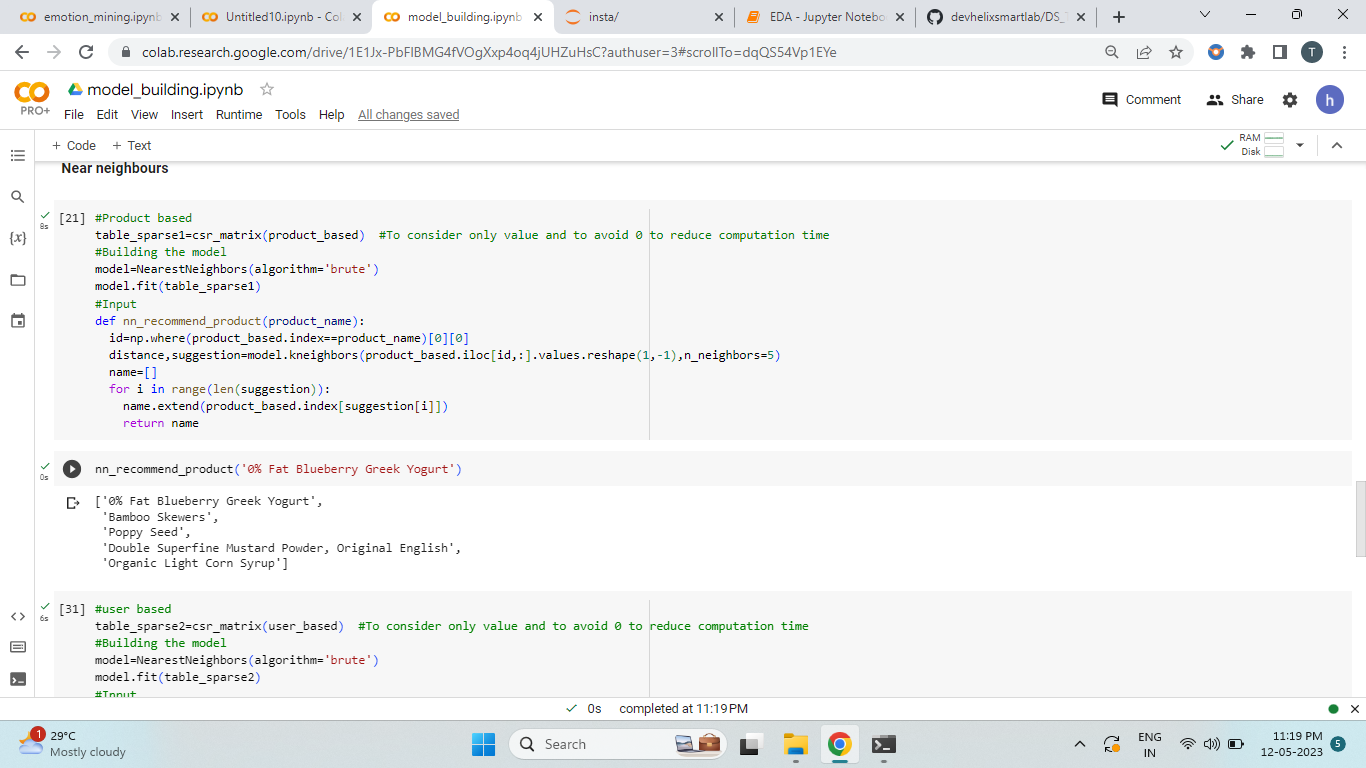
And using cosin similarities similar user’s were identified and the most frequently purchased product of the similar user were recommended.



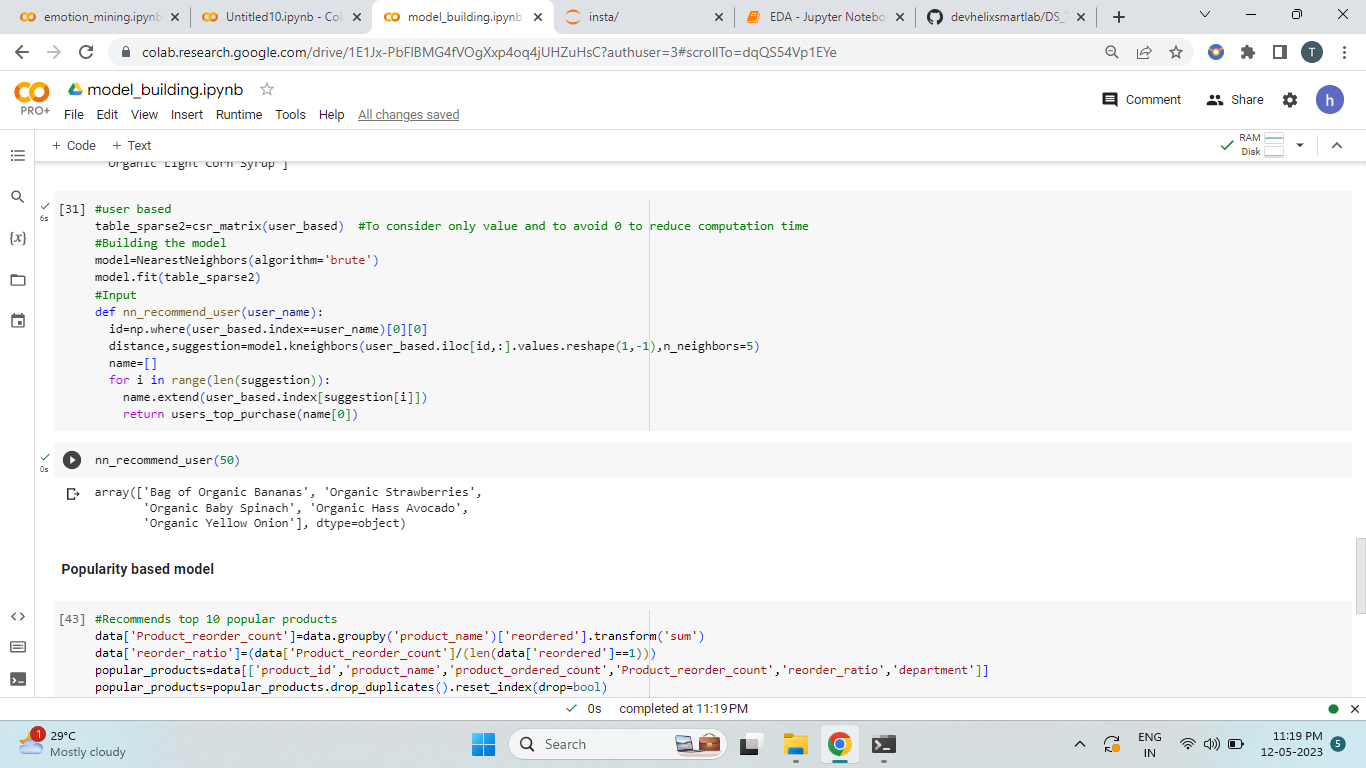
**3. Near Neighbour algorithm:**

Similar to the above pivot table spares table were create in a similar fashion and they were converted into sparse table and was feed into near neighbour algorith with number of neighbour=5 and top 5 products were recommended.

*Output for product based Recommendation:*



*Output for user based Recommendation:*

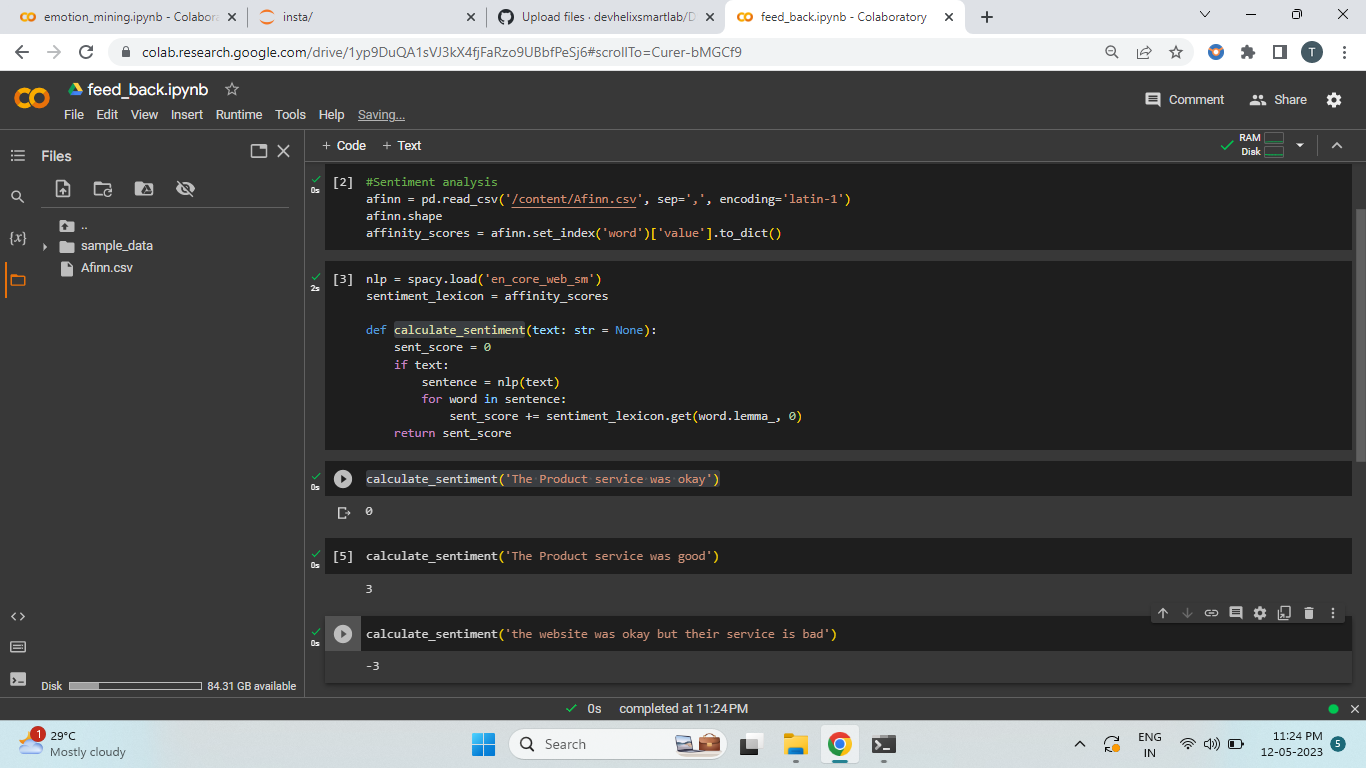


Out of all the models the model with cosine similarity was giving relevant recommendations.

**Feedback system:**

This system helps the company to identify whether their customers has happy or satisfied with their products and services. here I used spacy libraries in combination with affin data set to interpret the user inputs/comments/feedback as good or bad.in a scale of -3 to +3.

Where -3 indicates that the user is least satisfied and +3 is a happy customer and 0 indicating a neutral reaction.



**Evaluation:**

In model based method the metric which was used for evaluating the performance of was rmse and mae and the svd model was giving 27 percent accuracy.and which altering the parameters and data was able to push the accuracy to 37% but was still too low so we for the final model I created a hybrid model which combines both collaborative and popularity based models.

**Final Output:**

The final model is a combination of collaborative filtering and popularity based model and feed back system has be implemented along with it the user will be asked to choose between the type of recommendation system and takes the relavant information such as user id for user based recommendations and product name for product based recommendations system and in return will give you recommendations and will asked the user if he is satisfied with the recommendations if the recommendation (sentimental score fall below 0) the the model will suggest u with and alternative suggestions(popularity based model). If the score is more than 0 then the model will return Thank you for the feedback message.

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

This model also deals with cold start issue by suggesting the user with popular products.