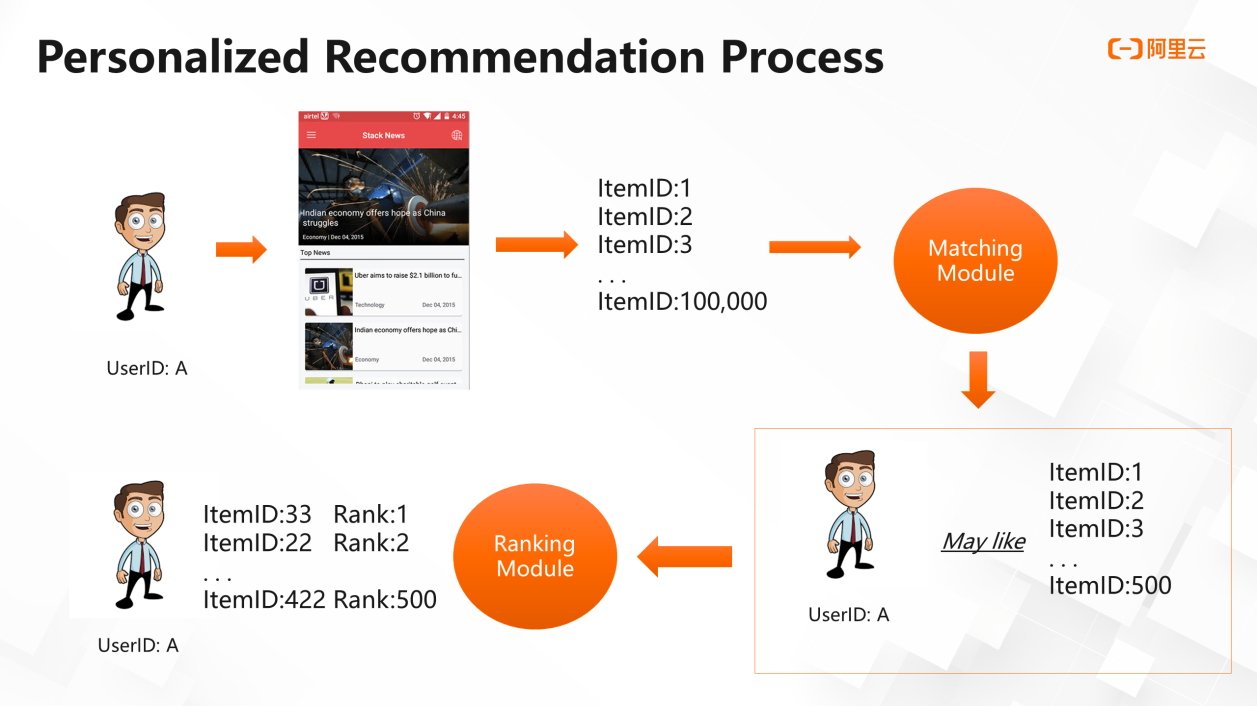


|  |
| --- |
| Personalised Prediction 2022 |
|  |
| September 15  Bz Analytics  Authored by: Nisha Nandal |

# Personalised Prediction

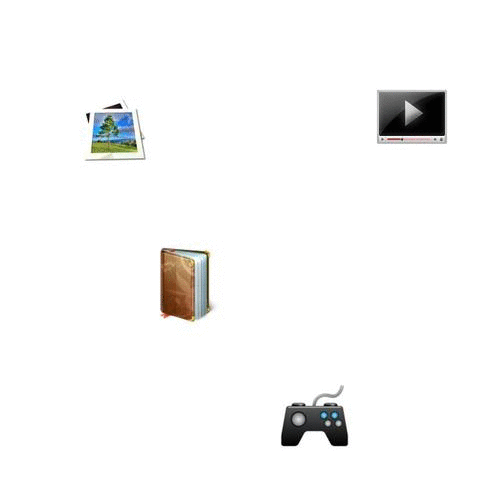
Recommendation engines are a subclass of machine learning that generally rank or rate products/users.

Good recommendations for example Video on Demand (VOD) platforms can increase revenue for long tail content by surfacing it in recommendations based on consumers’ behavior



Personalized Prediction is a subclass of [systems](https://en.wikipedia.org/wiki/Information_filtering_system) that provide suggestions for items that are most pertinent to a particular user.

Recommender systems usually make use of either or both [collaborative filtering](https://en.wikipedia.org/wiki/Collaborative_filtering) and content-based filtering (also known as the personality-based approach), as well as other systems such as [knowledge-based systems](https://en.wikipedia.org/wiki/Knowledge-based_systems). Collaborative filtering approaches build a model from a user's past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in. Content-based filtering approaches utilize a series of discrete, pre-tagged characteristics of an item in order to recommend additional items with similar properties.



Objective

To recommend items according to the user interest based on user rating taking other factors into consideration.

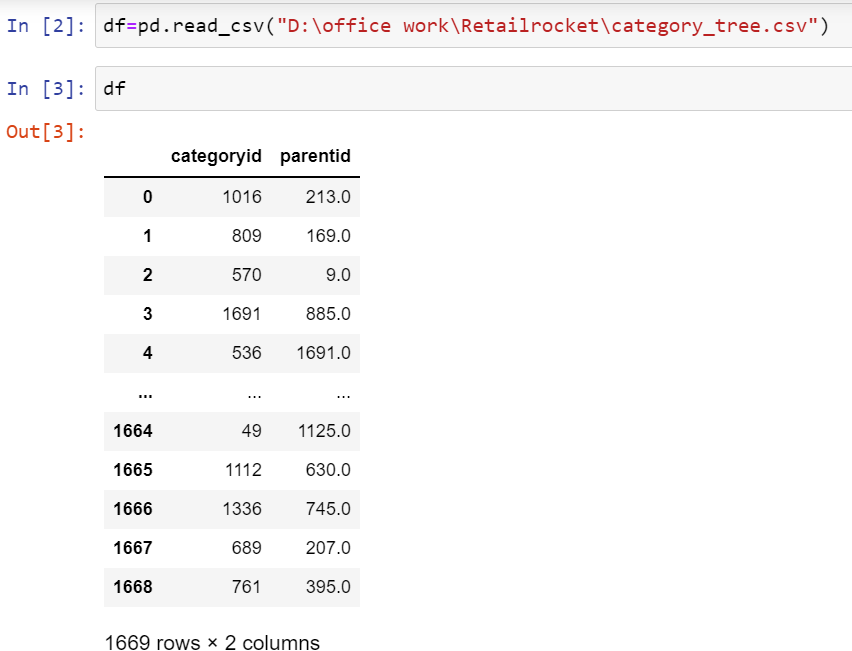
Data link:

<https://www.kaggle.com/retailrocket/ecommerce-dataset/home>

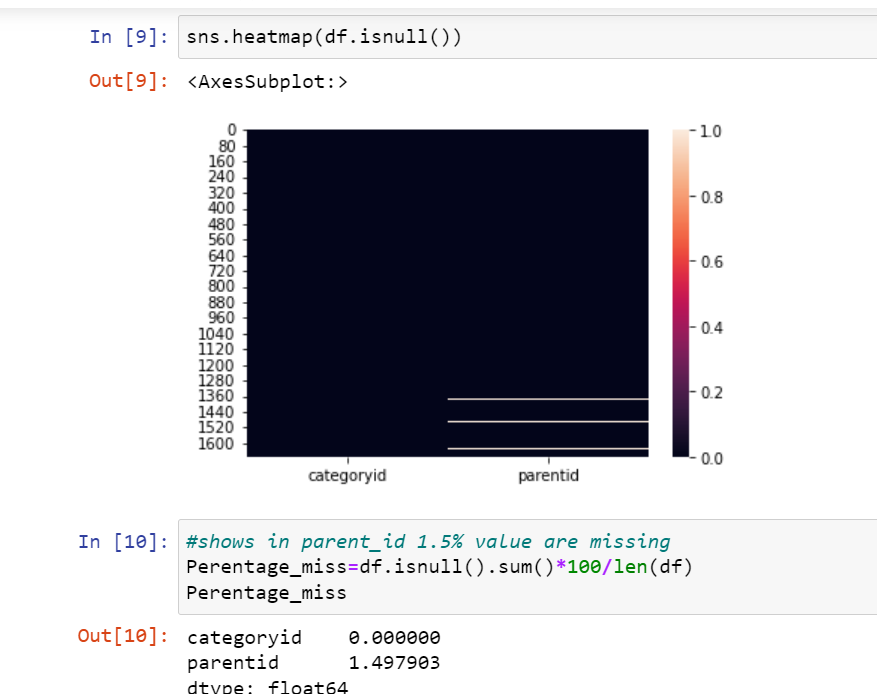
The dataset contains basically three files

* Category.csv which identifies category and parent ID of different products given in the dataset.
* Events.csv shows the different different events like view ,addtocart and transaction.
* Itemproperties.csv shows

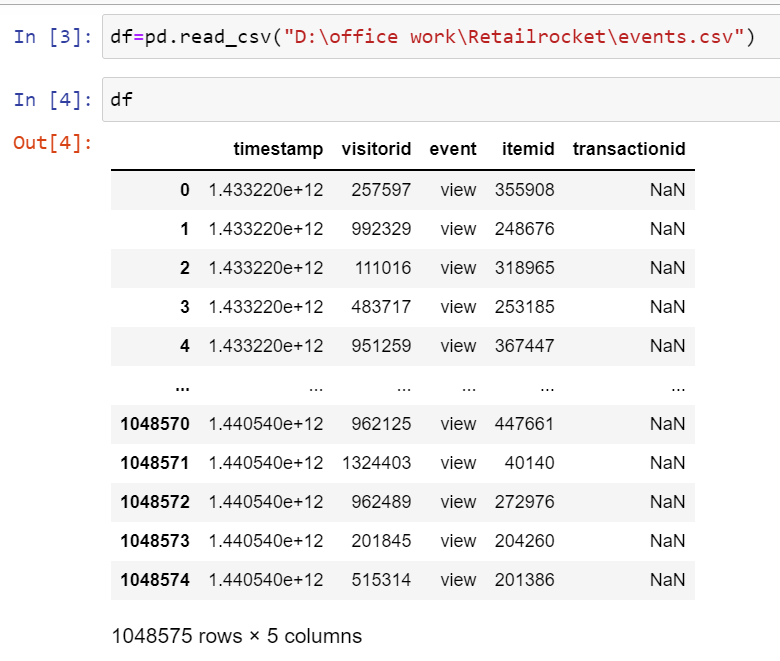
The category file is:



The data filteration of category file:

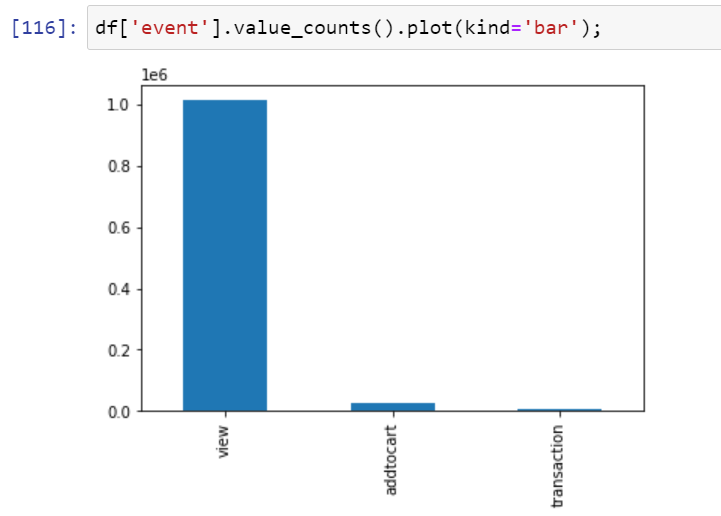


The event file :



The event file is showing different stages of data. The data file shows the timeframe within which a particular item is viewed, addedtocart or transacted. Further calculating with the unique “visitorid” filters the data according to customer’s interes and likes.

By understanding the user’s behaviours we can find out how many items are viewed, addedtocart and being transacted. The user behaviour transaction completes the process classying what items user finally interested in. These items will be used later to be recommended at the time of next purchase.



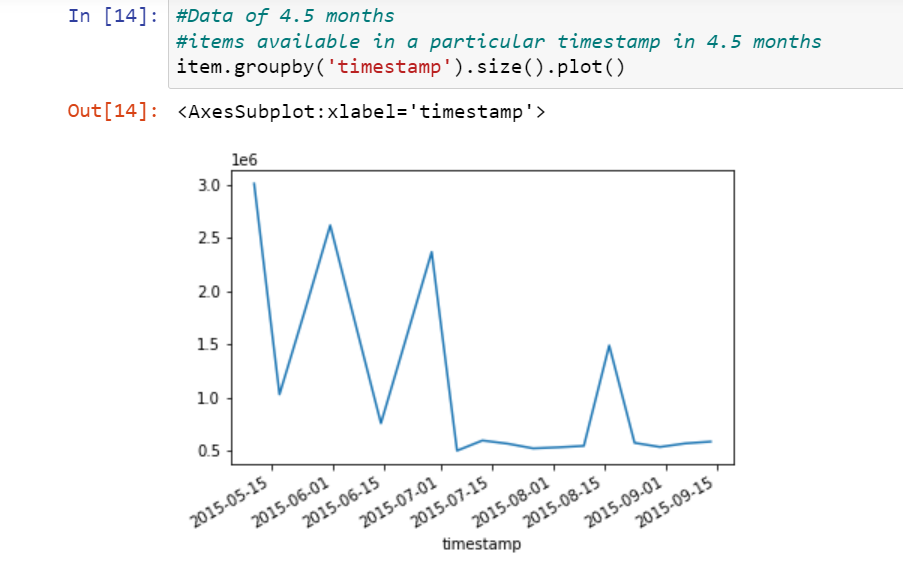
Further, we need to find “timespent” for each of the items:

* This help in classifying user behaviour more clearly
* Shows total time spent on items which predicts the customer’s likes dislikes
* The more time spent means increasing chance of transaction

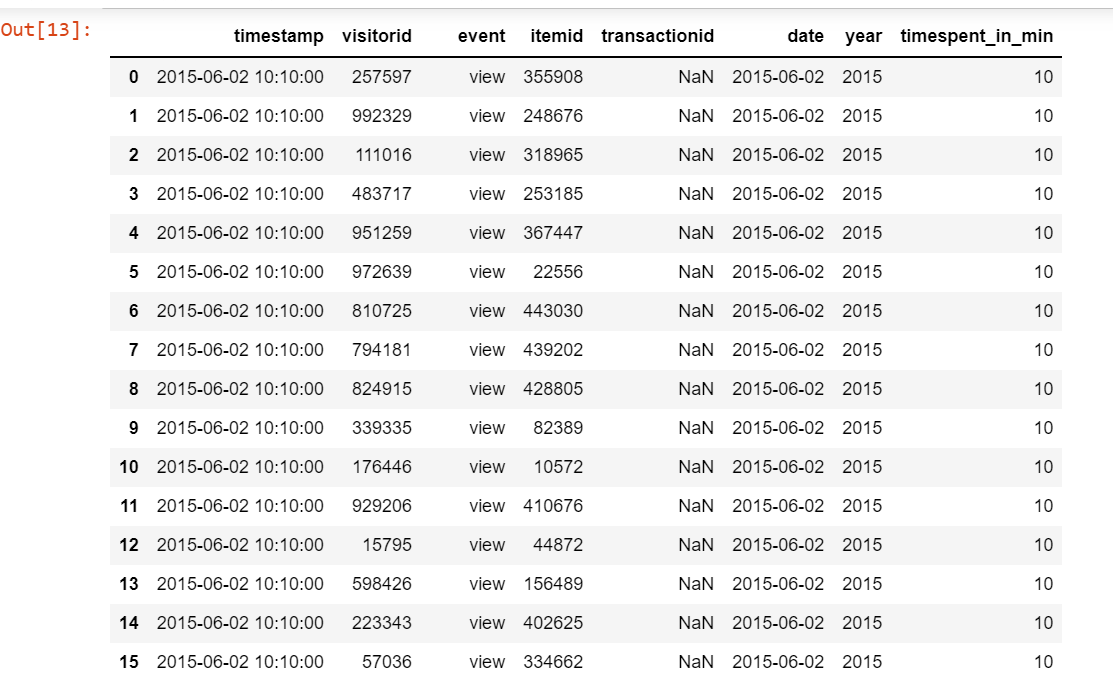
Data of 4.5 months with timestamp

Visualisation of dataset given which showing timeperiod and items available.

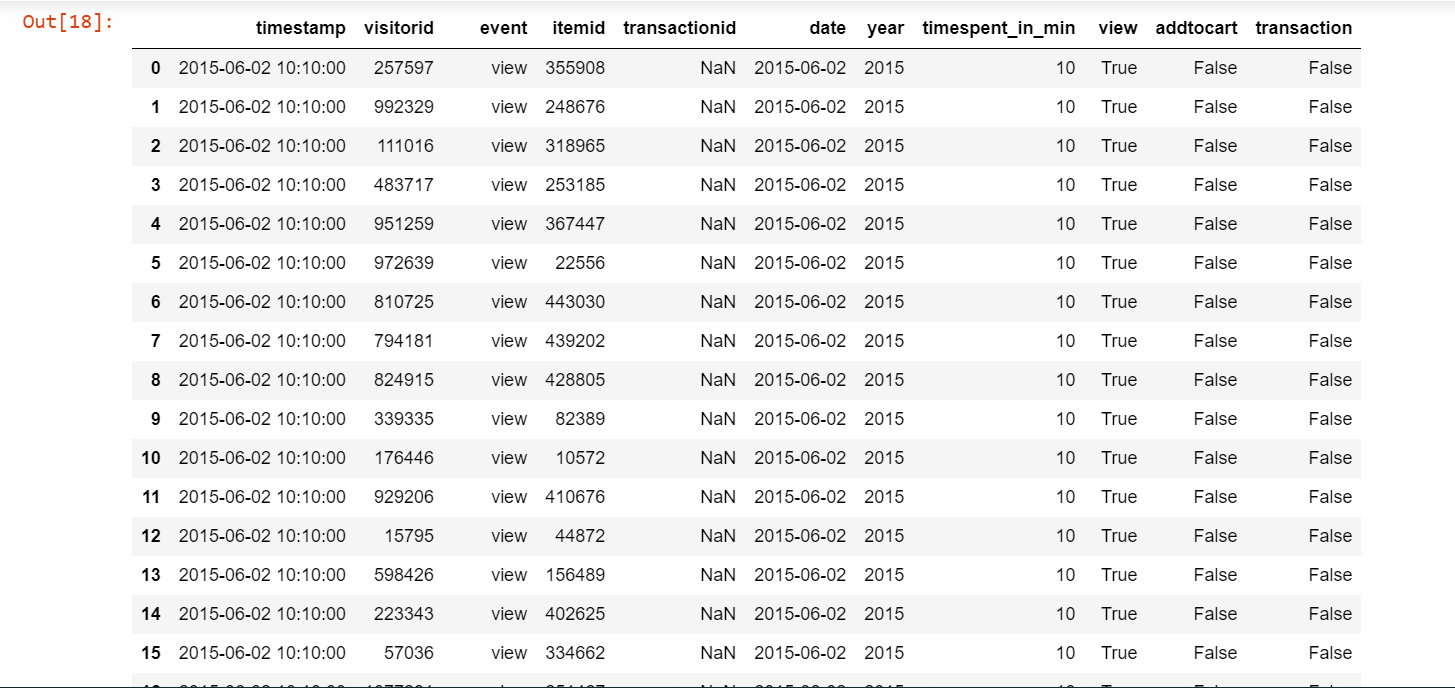
And then converting the timespent in readable format and further finding time spent to buy a certain item.



Timespent on a particular event with all visito’s ID

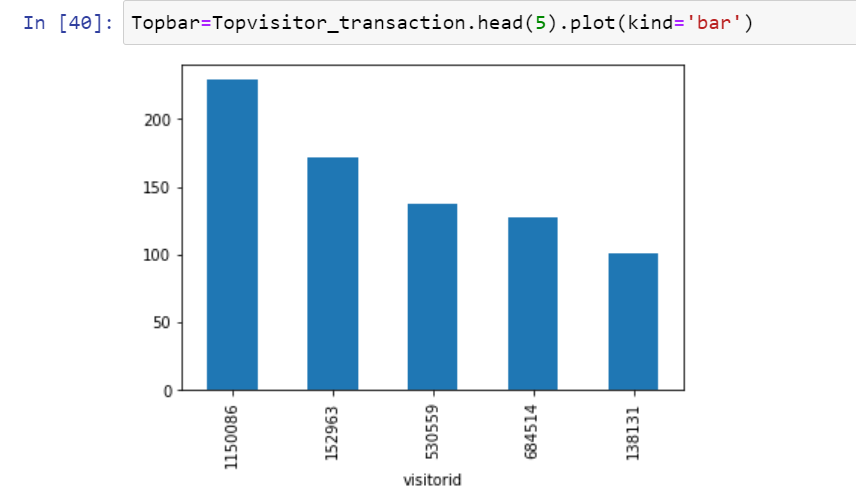


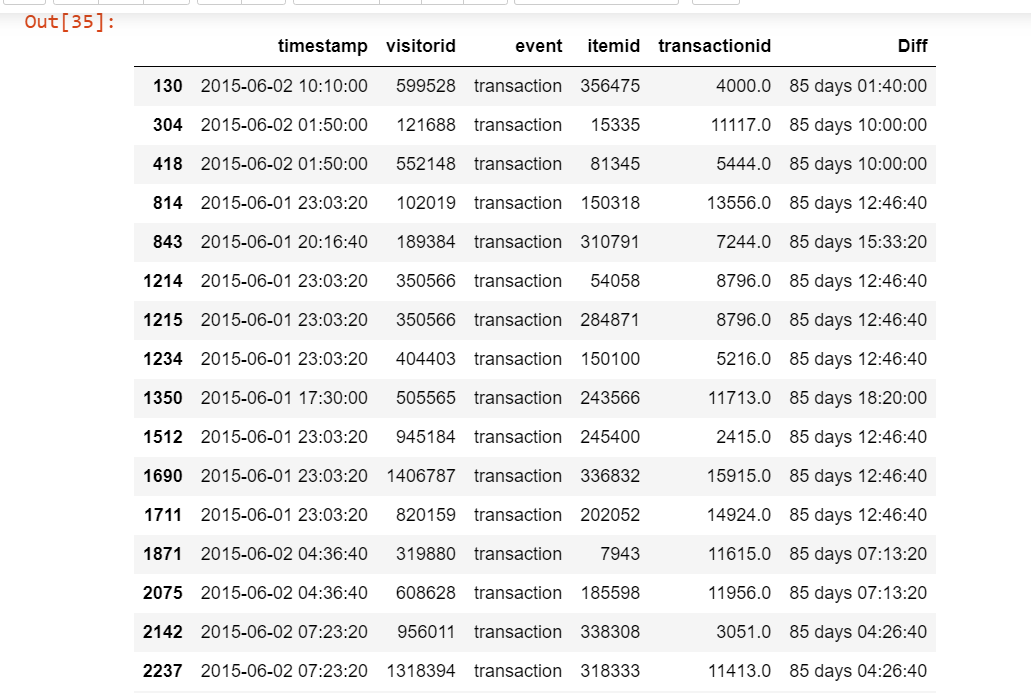
The following data shows the active stage of event performed by the user and classfying user id accordingly.



Now, essentially what is need is how much duration is there between the recent purchase and the last purchase.

1. Classifying and Tracking the transacted items
2. The transacted items will become part of recommendation system
3. For every unique visitor Id and ItemId The difference of purchase has been shown.
4. The top transactions can be shown as:





Grouping of data

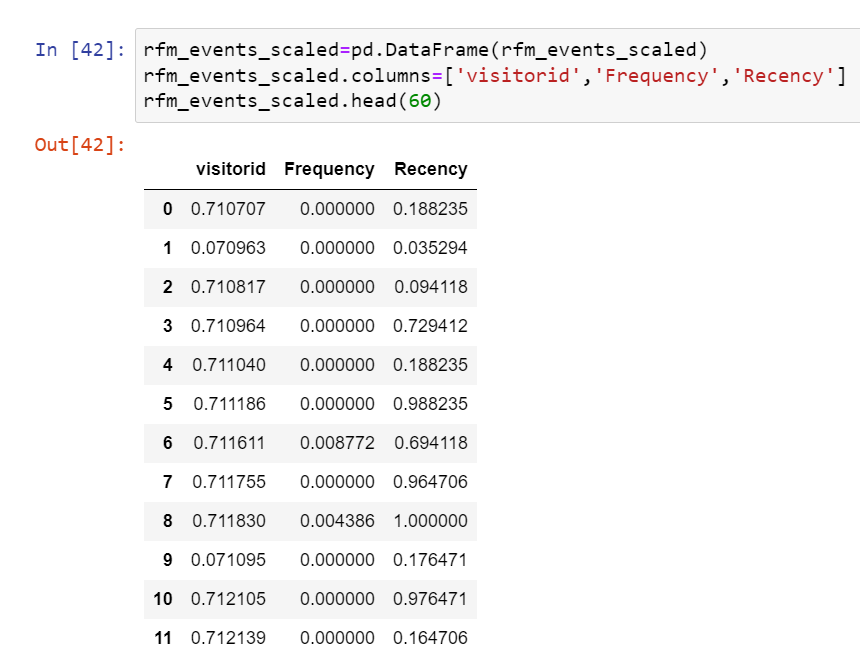
1. After finding the difference it’s important to calculate the retention and frequency of customers.
2. Then Grouping the data as per visitor’s ID to get collective data related to a particular customer.
3. With grouping all the details of a particular customers can be easily seen collectively.
4. Showing difference in the time of visiting to time spent in buying the items
5. Also, shoeing whether the purchase done is within 24 hours?



Plotting with scaled data

The following graph shows the data of frequency and recency of visitors:

* The data is scaled first to obtain a graph within range
* The two clusters here represent the data with visitors showing different characteristics.
* The frequency graph identifies the clusters based on customer arrival
* The recency graph shows the clusters based on how many times customers are visiting.

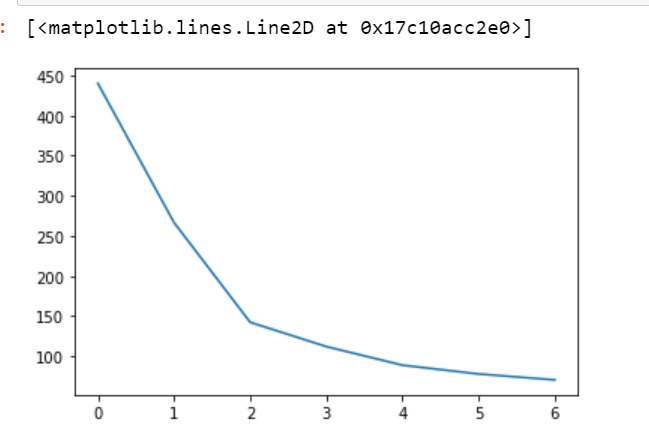


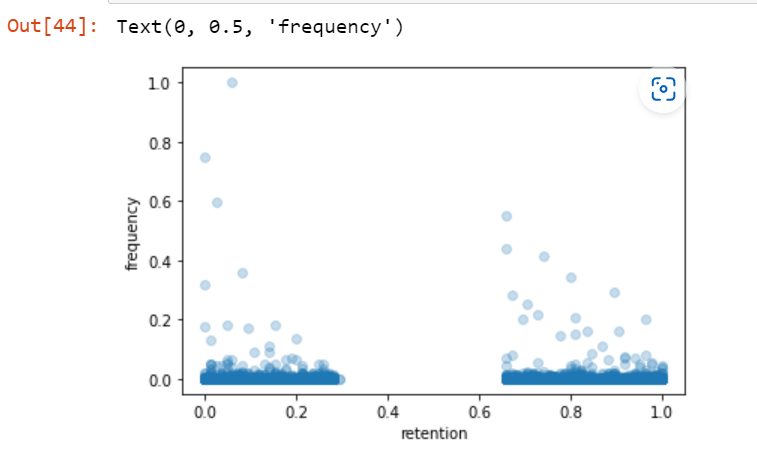
K-means algorithm:

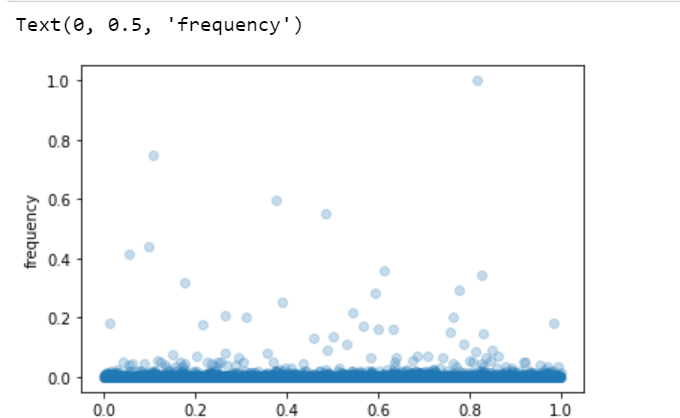
K-means clustering is an intuitive method to find clusters. See: <http://en.wikipedia.org/wiki/K-means_clustering>. It can also be formally stated as an optimization problem where we simultaneously optimize for the location of the clusters 1..K and the assignment of point 1..N to these clusters.

For model creation I applied K-means algorithm for the best fit the below graph shows 2 clusters.

* K shows the number of clusters
* Scaled data according to recency and frequency will be used







Personalised prediction data research

Analysis of data include:

* Item Brand
* Product description
* Session duration
* Item/product similarity(automobile)
* User data -age, location, email
* **Association rule** -----frequently brought together ----eg. A charger or phone case with a phone, a helmet with a motorcycle----ultimately increases the order value
* Similar item---help user refine their needs
* Related content----increases the time spent on site
* Trends----logic and engage the customer from the first second

The above-mentioned data is the one that represents the vitality of recommendation systems.

BtoB

How from one business another business thrives

* Includes essential list of item eg. Want to buy car but what comes with it is essential—need brand, material,usage, no. of people
* Association rule applies here
* Association of technology

Media and entertainment

* Includes some video related to an item category eg. Sustainability, reusability
* Related to manufacturing of item
* Related to workforce

Financial services

* Flexibility in payment structure
* Including various options