Introduction:

The advent of the e-commerce markets facilitates the process of shopping without the need for physical interactions with products. However, an appealing aspect of physical retail stores is that customers who are undecided on the products they desire to purchase have the ability to browse and receive recommendations from shelf displays and salespeople. The e-commerce industry utilizes recommendation models to satisfy this objective . A personalized recommendation model aims to identify products that are of most relevance to a customer based on his or her past interactions. This enhances a user's intention to browse more products and makes them more likely to buy these products, effectively increasing e-commerce revenue . Thus, the evaluation of recommendation algorithms for a range of properties is essential in order to select the best algorithm from a set of candidates]. Performance is either measured by offline evaluations, by conducting user studies, or by using online evaluations when a recommendation model is live on an online platform. Online evaluations are unique in that they allow direct measurement of overall system goals, such as long-term profit or user retention.

Errors which can reduce the performace of the model:

- Data entry errors: data entries recorded by humans from speech or written context, where mostoften there are fields of data stored that might not always have a value present; hence an improper default value may be assigned to this field without much consideration to whether or not thedefault value is a possible outcome of real data.

- Distillation errors: data that should be preprocessed before storing them onto a database toreduce complexity and noise of raw data, which if not considered, may impede final analysis.For example, incremental revenue or the conversion metrics calculated may be invalid if the pricesutilized for each product varies in the currency they are stored against. It can then be beneficialto assert this when recording data for e-commerce businesses who have a global dominance,to standardize the currency used against one system (such as the USD) for all customer interactions.This notion serves as the background to Section Uniform Data Management.

- Data Integration errors: when the data collected stems from various servers or sources where themerging of records into a single database may cause inconsistencies. For example, it is crucialthat the timestamps of each user interaction on the database are stored against a standard timezone, such as using Coordinated Universal Time (UTC), and not against the time zone of the IPregistering the interaction. This is discussed in Section 4

Data Preparation:¶

I am going to build two recommendations with help of Nearest Neighbor. I will use the collaborative filtering technique. I have two datasets one is about movies rating and the other one is about amazon product reviews.

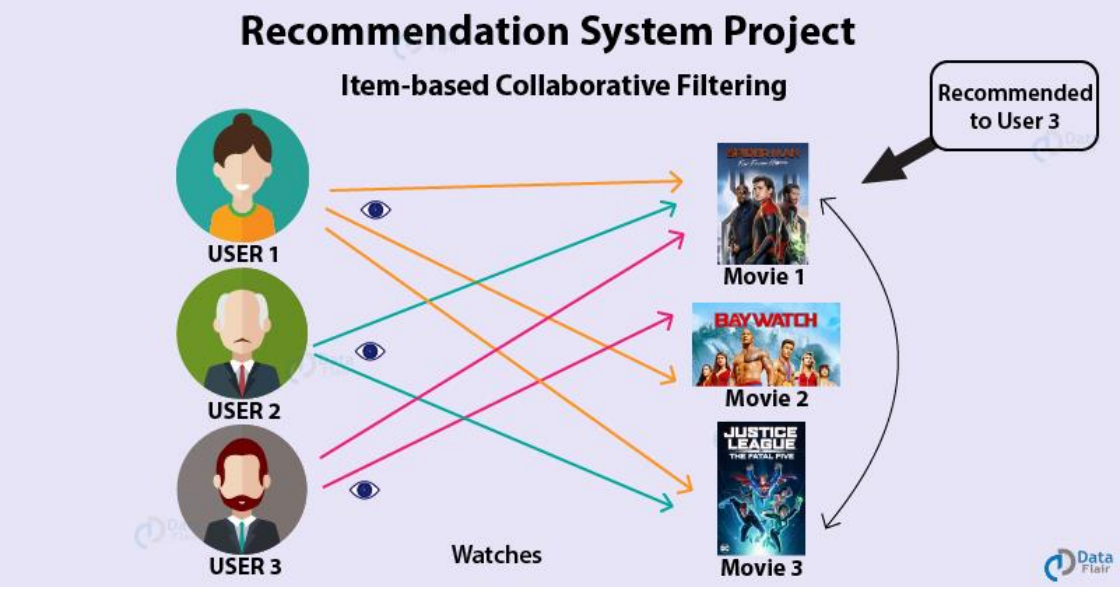
Movie's rating dataset features:

1) UserId
2) MovieId
3) Rating (Target column)
4) Title

Movie Recommendation System:

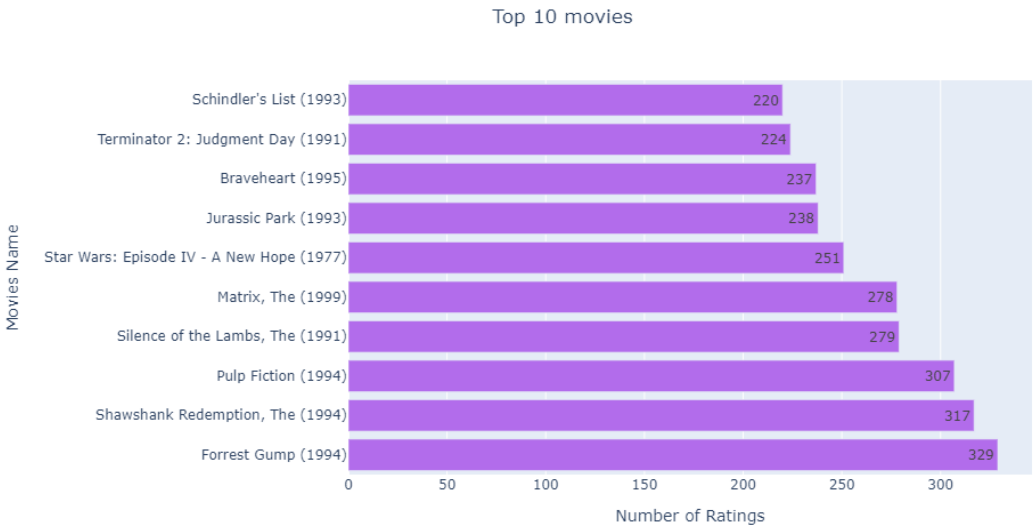
I have a dataset of movies reviews. This dataset contains features. The dataset is pretty so I don't have to do much work for cleaning.

Exploratory data Analysis of Movies dataset:



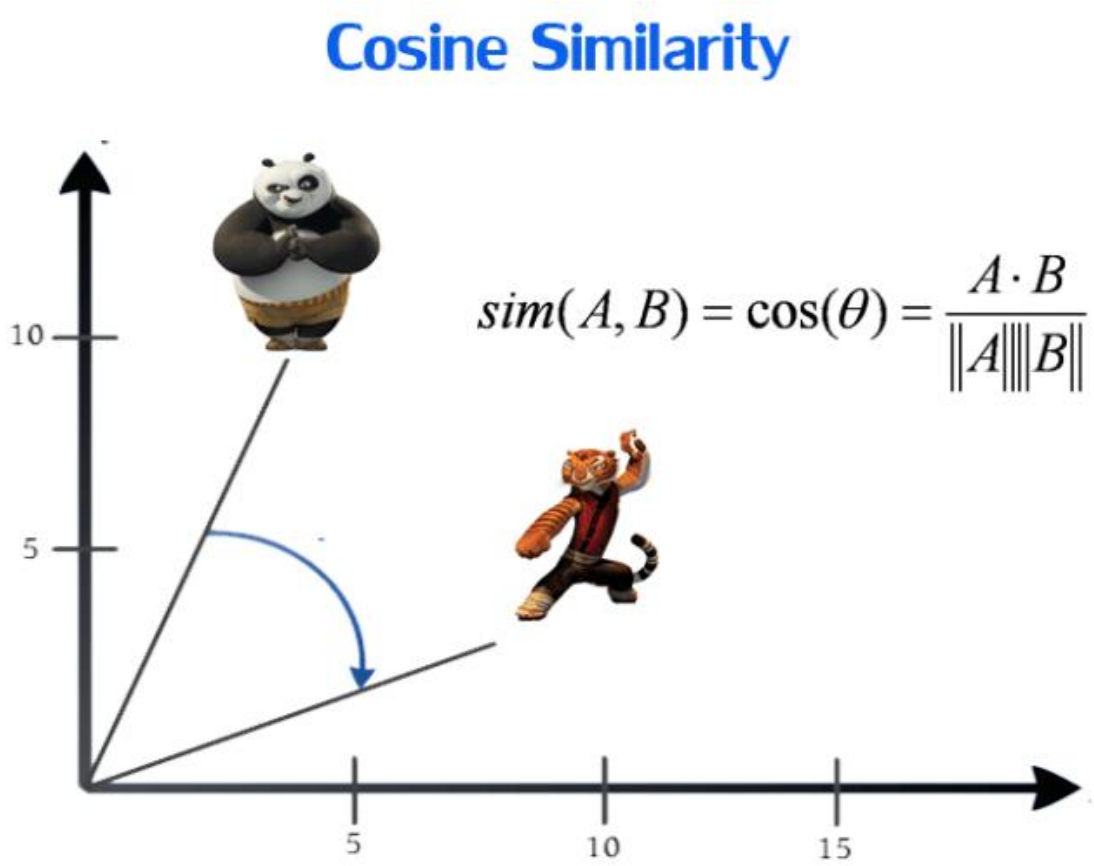
Exploratory data Analysis on Movies dataset:

- According to the dataset these are the best movies and top rated movies.



Nearest Neighbors:

kNN is a machine learning algorithm to find clusters of similar users based on common book ratings, and make predictions using the average rating of top-k nearest neighbors.



We are going to use the Nearest neighbors to do collaborative filtering. I will use cosine similarity to calculate the distance between the points. Cosine similarity is a metric used to measure how similar two items are. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space. The output value ranges from 0–to 1. 0 means no similarity, whereas 1 means that both the items are 100% similar.

Product Recommendation System:

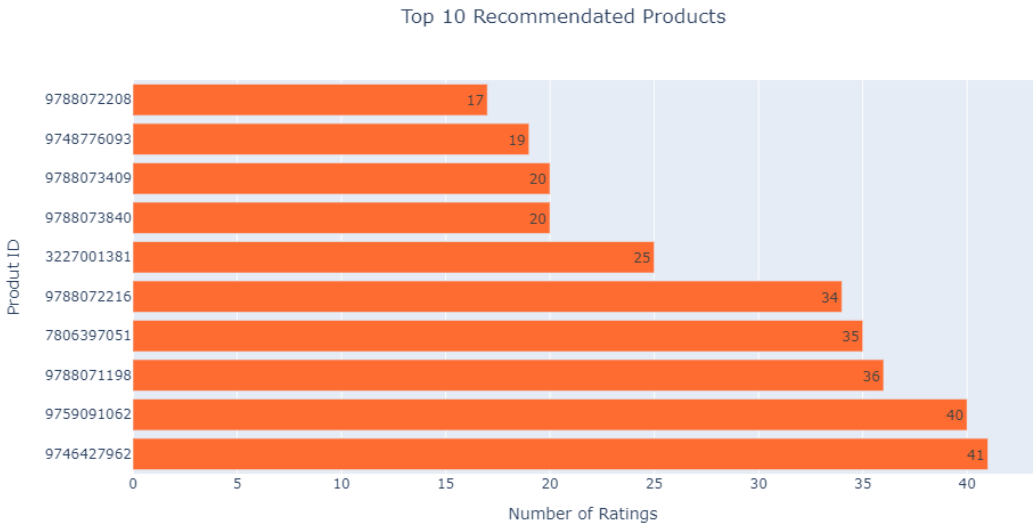
For product Recommendation I am going to use the amazon reviews dataset. the dataset contains 4 features but it does not contain the product title feature. we only have got product id. so our recommendation system will only return product id instead of title.

Products rating dataset features:

1) UserId
2) ProductId
3) Rating (Target column)
4) Timestamp

	UserId	ProductId	Rating	Timestamp
0	A39HTATAQ9V7YF	0205616461	5.0	1369699200
1	A3JM6GV9MNOF9X	0558925278	3.0	1355443200
2	A1Z513UWSAAO0F	0558925278	5.0	1404691200
3	A1WMRR494NWEWV	0733001998	4.0	1382572800
4	A3IAAVS479H7M7	0737104473	1.0	1274227200

Results of Product Recommendation Engine:



Product:

- Eyeliner Pen.

Suggestions:

- Kms California Hair Conditioner.
- Anti Aging face cream.
- Women's Perfume.
- Peanuts: A Charlie Brown Christmas.