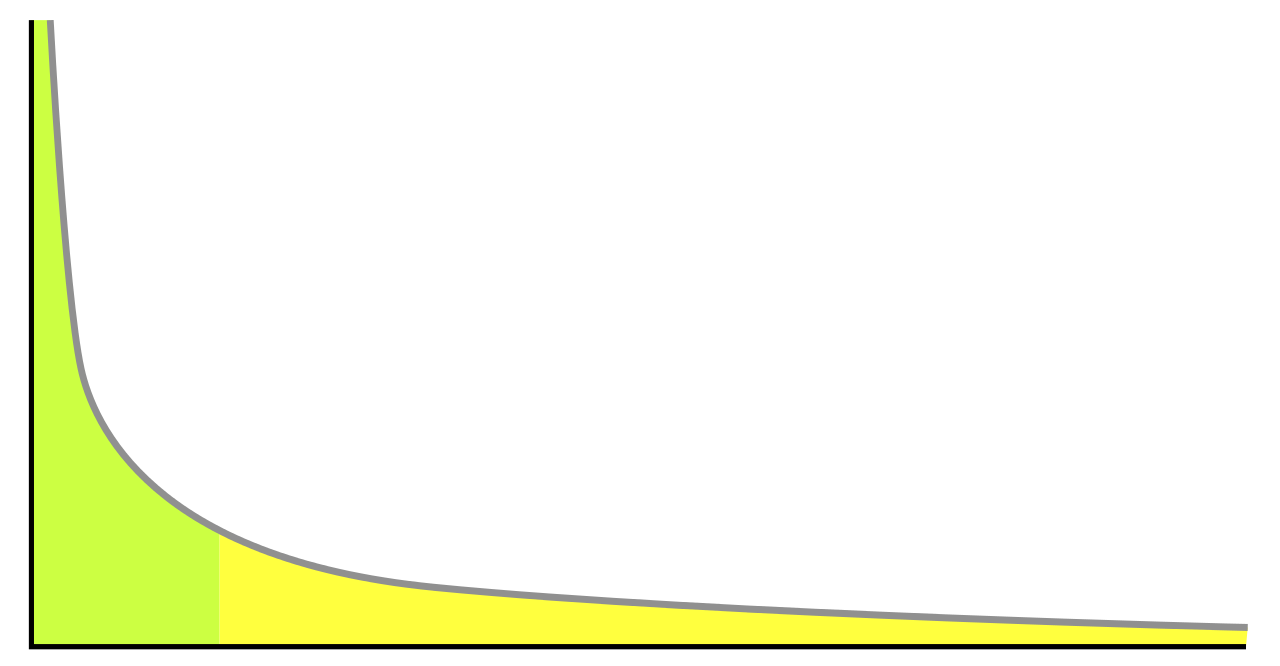
**Background - What is a recommendation system**

Allows the prediction of user responses to options. The ability for machines to predict things for users in a personal level.

**Long Tail Effect**

Physical delivery systems are characterized by a scarcity of resources. It is not possible to tailor the store to each individual customer .Online stores can make anything that exists available to the customer. The distinction between the physical and online worlds has been called the long tail phenomenon which  makes recommendation systems necessary.

****

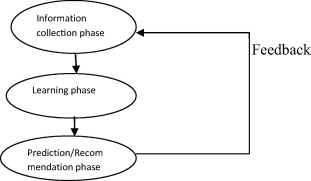
The head is the highest point of sales. In terms of movies and music, we're talking about blockbuster hits (Titanic, Star Wars, Harry Potter, The Beatles, Lady Gaga, etc.)...

The body is where sales begin to slope downward. Think of moderately successful movies or albums by big name artists that didn't fare so well. The body is longer than the head.

Now the tail is much, much lower than the head and the body. But it is much, much, MUCH longer. We're talking now about short films, student films, low budget slasher films, garage bands, your friend's band, etc.... Some might say the tail is actually infinite - there really is no way to quantify how many indie films and songs have been released over the years.

**Phases of recommendation process-**

1. Information Collection
   1. Gathering user profile or model for prediction. Also attributes, behavior and content of anything user has done.
2. Explicit Feedback
   1. Prompts the user to provide ratings for items to improve model.
   2. Accuracy depends on the quantity of ratings the user gives.
   3. Provides a higher quality and more confidence of recommendations
3. Implicit Feedback
   1. Monitoring different actions of users like purchase and navigation history.
   2. Positive is that there is no interference from the user.
   3. Negative is that it is less accurate
4. Hybrid Feedback
   1. Combination of both explicit and implicit feedback.
   2. Can be used to check on explicit rating based on implicit data.
5. Learning phase
   1. Applying learning algorithms to filter the user features.
6. Prediction/recommendation phase
   1. Recommends or predicts what kind of items the user may prefer



**Types of Recommender Systems**

1. **Content based system** - looks at the properties of item being recommended. Similarities between items is determined by the similarities in their properties.

Take Netflix for instance, if you have watched many thriller movies, then they would be given the recommendations to watch other movies based on the genre property of thriller.

* 1. For items where the value of features/characteristics are clear:
     1. Create a collection of records representing important characteristics of the item.  For example, a recommendation system for a movie would have actors, director, year, and genre of movie. This would help with giving users movies with their preference for certain directors or actors.
  2. Items where it is not immediately apparent what values of the features are not clear: Use document collections and images.
     1. Document collection such as news articles. Instead of obvious features, identify words that characterize the topic of the document.
        1. Remove all unwanted and common words and then compute the TF.IDF score for each word in the document. The higher the score for the words the more that words characterizes the document.
        2. Then the recommendation system can suggest other articles based on if all the words have TF.IDF score above a certain threshold.
        3. Measurements to check similarities between two documents:
           1. Jaccard distance - distance between words.
           2. Cosine distance -distance between the set treated as vector. Each words with high TF.IDF score is in vector. If words is in the set then give it a 1 else 0. To find the intersection of the two sets of words and the square vector length = square root of number of words in each set. This allows to compute the cosine angle.

Can estimate degree to which a user would prefer an item by computing distance between user’s vector and item’s vector.

* + 1. Using tags to get item features: if user bookmarks many pages with certain tags, then we can recommend pages with same tags.
       1. Con: only works if user actually creates tags.
  1. Goal of Content-based system: create both an item profile with feature-value pairs and user profile summarizing preference of user.

1. Collaborative Filtering - look at similarity of the user rating for two items,
   1. Users are similar if their vectors are close.
   2. neighborhood-based approach a number of users is selected based on their similarity to the active user. A prediction for the active user is made by calculating a weighted average of the ratings from several users.
   3. Higher the similarity the closer they are together
      1. Cosine Similarity: evaluate the similarity between two vectors based on the angle between them. The smaller the angle, the more similar the two vectors are.

**Utility Matrix**

A recommendation system has users and items. These are represented in a form of a matrix. An example is: user rating movie on scale 1-5 and blank means user has not rated.  Goal of the recommender system is to predict these blanks. The recommendation system might take into account properties of movies, such as their producer, director, stars, or even the similarity of their names. Based on this it might look at the user’s rating for certain movies and use that to make a prediction of how another person might rate a similar movie.

This is needed to recommend items.  Two approaches to discover the value a user places on items:

1. Based on user ratings. Problem with this is that data will be biased by only people who provide ratings.
2. Can infer from user behavior. For example, user buys products at Amazon. The rating system used is 1 (like) or 0.

**Applications of Recommender Systems**

1. Product Recommendations: Amazon for example. Giving users suggestions of products they might buy.
2. Movie recommendations: Netflix. Based on ratings provided by user. Fun fact: Netflix offered a prize of 1 million for the first algorithm that could beat its own recommendation system y 10%.
3. News Articles: identify articles of interest to readers based on articles they have read in the past. Might be based on similarity of important words.

**Algorithms used by Recommender systems**