

# Project Brief

## [Product Recommendation System]

**Team Brie:** Devarshi D Pancholi, Gaurav Singh, Kanak Khanna, Krunal Patel, Rishabh Jamar, Shreyaskumar Kathiriya

### Project Summary

Amazon's recommendation system generates 35% of its total revenue and they use item-to-item collaborative filtering algorithm. The system matches a customer's purchased and rated items to similar products in their list, which are then selected to be part of their recommendations. According to statista, 197 million people around the world order from amazon and 10% of the total population actively reviews. This accounts to 19.7 million reviews going unnoticed.

Customer reviews are a great source of "Voice of the customer" and could offer tremendous insights into what customers like and dislike about a product or service. As a team we thought why not take into account this 19.7 million population and take user experience to the next level. Whether you are watching a movie or listening to songs or simply shopping only recommendation technology helps customer centric companies increase their ROI by fruitfully leveraging personalization and recommendations.

We will be using Amazon dataset provided by Stanford professor J.Leskov which has 568,454 instances and 9 columns. Our project will include a descriptive analysis followed by sentimental analysis using NLP and Text analysis and then we'll conclude our analysis by recommending products for the customer. The impact of the project is to increase revenue as well as enhance customer satisfaction by personalizing the shopping experience.

### Overview of Needs

While researching about the current recommendation system of Amazon, we learnt that it has integrated many recommendation "entry points" into its online experience to maximize cart value. For example, users can click on the "Your Recommendation" link and "Frequently Bought Together" section. They rely heavily on their recommendation engine.

Reviews are actual use cases of the product. Sale of the product does not imply that it is being liked by all the customers. A certain product may be selling well because it has some good offers associated with it, or because it gets delivered fast or maybe just because it is cheap. Reviews should be an integral part of the recommendation engine because only after reading the reviews, one gets to know the real quality and best suited purpose of any particular product. There may be cases when a product is selling very well but the customers are not happy and they are only purchasing it because there are fewer choices available in the market. Reviews also confirm whether the product details mentioned by the seller align with the actual specification of the product. For eg, if you are planning to buy a sweater that is being sold for \$5, it will be selling very well because it is cheap but only reviews can tell you whether the sweater is actually warm or is it just suited for fashion purposes.

For companies, this will act like one more feature to include in the existing recommendation system and for customers, it is enhancing their experience.

## **Project Deliverables**

The main goal of the project is to build a better recommendation system for Amazon on top of their already existing system.

The following are the proposed deliverables for this project:

1. To improve and build a layer of user personalised product recommendations on top of the current recommendation system used by Amazon.
2. Incorporating customer reviews about products into the recommendation system.
3. Performing a text analysis of product reviews to find similarities between different products like two different products can be portable and compact.
4. The objective is to match these product characteristics extracted from the reviews our users have given and match them with the products that have similar characteristics and user who wrote the review. For example say user 1 reviews “loved the product as it is very portable” on water bottle and user 2 reviews “Amazing, portable lunch box” here the review suggest that these both users like products that are portable and chances are high that they will like the products that are portable.
5. To increase revenue as well as enhance customer satisfaction by personalizing the shopping experience.

The following are “reach”, or “nice-to-have” deliverables for this project:

1. The model that we have proposed to personalize the recommendations for each customer is a feature on top of Amazon's existing recommendation system. Integration of our system along with existing system will be a nice-to-have deliverable.
2. Actually increasing customer retention and customer satisfaction based on our recommendation system.

## Data Summary

We have Amazon's Fine food department's product reviews and ratings for more than 568,454 instances and 9 columns like overall rating, reviews, product ID (ASIN). The data was taken from a Stanford professor J.Leskov. It contained 550,000 instances which we filtered to products which had at least 100 reviews, so that we can add reviews as a factor too in our recommendation system. This data lacked the previous search history of users or a record of what they bought before thus we build a recommendation system based on the reviews.

## Proposed Methodology

- *Data Extraction:* Data was obtained in zip file and from that we got a json file which was further converted into csv.
- *Exploratory Data Analysis:* Checking for missing values and general trend of the data. By visualizing various attributes.
- *Data Cleaning:* Removed the duplicates and filtered the data. Before creating corpus stopwords were removed using libraries like "SMART". All the punctuations were removed and data was converted into lowercase.
- *Data filtering:* After grouping the data by Amazon's standard identification number included products with review size more than 100 and created dataset with three columns that is asin, overall and summary.
- *Sentimental Analysis:* Created corpus for food summary and sentimental analysis is performed on predicting the helpfulness of the reviews. Also, created document term matrix and term document matrix. Created word clouds and performed feature extraction and implemented classification model.
- *Recommendation System:* K-Means Clustering is used to perform item-based collaborative filtering to find the 2 most similar items.

## Project Success

The success of this project could be determined by making sure the below deliverables are met:

1. Build a recommendation system based on customer reviews.
2. Build a K-means clustering model for grouping products.

3. Perform a text analysis of user reviews.
4. Increase revenue by providing better recommendations.

## **User-Stories**

Brainstorm specific ways that this project could have an impact positive or negative. Think about how the project might impact each stakeholder or group of stakeholders through specific user stories.

*Customers:* Providing the customers with a better recommendation system that just not only considers their past search history but also product reviews left behind by similar minded users. This will give them a very personalized and accurate recommendations which will save them a plethora of time. Also, this will enrich the user experience giving them a feeling of satisfaction and by saving a lot of valuable time.

*Distributors:* Distributors won't have to go to the hassle of delivering products again and again to the customers. They get more sales due to the better recommendation system. Demand will increase thus leading to faster inventory exhaustion leading to more sales.

*Amazon:* Happy customers and distributors will increase the customer retention ratio. Also, the revenue will increase as more sales are happening. It will increase its good will among the customers and differentiate its service from other such online retail stores. Customer satisfaction increases.

*Manufacturers:* Companies that manufacture a variety of products knitted together will profit a lot as their product will be featured more with the new recommendation system which suggests similar products or products which most go with it.

*Society:* It gives the people and the customers a better way than before. They are able to find products now easily for which they had to spend hours reading reviews, comparing products etc.

## Project Plan and Milestones

Task	Deliverables/Milestone	Lead	Completion Date
Defining Business Problem	Objective of the Project, Project Deliverables and Good Dataset	Kanak, Krunal, Gaurav	Dec 03, 2019
Data Cleaning	Cleaned Dataset	Devarshi, Shreyaskumar	Dec 05, 2019
Analyzing variables	Comes up with the variables that are important to the model	Kanak, Gaurav	Dec 05, 2019
Text mining for review	Sentimental analysis of reviews left by users on products	Shreyaskumar, Rishabh	Dec 07, 2019
Data Analysis Discussion	Business recommendation for further improvement	Krunal, Rishabh	Dec 10, 2019
Running clustering	Improve clustering method so as to recommend better and relevant products	Devarshi, Gaurav	Dec 07, 2019
Create Presentation	Presentation with a good data visualization	Krunal, Rishabh	Dec 10, 2019
Preparing Presentation	Best way to communicate the data	Shreyaskumar, Kanak, Devarshi.	Dec 11, 2019

## Project Roles and Responsibilities

Name	Role	Responsibilities
Kanak Khanna	Project Manager	<ul style="list-style-type: none"><li>- Create a timeline</li><li>- Create a project scope</li><li>- Lead the discussion every meeting &amp; divide work to every members</li></ul>
Krunal Patel	Data Visualization Manager	<ul style="list-style-type: none"><li>- Create a good data visualization</li><li>- Create a good business presentation</li></ul>
Rishabh Jamar	Data Visualization Manager	<ul style="list-style-type: none"><li>- Create a good data visualization</li><li>- Create a good business presentation</li></ul>
Gaurav Singh	Data Researcher	<ul style="list-style-type: none"><li>- Find reliable data that needed - Lead the data selection and data exploratory analysis that is proposed by other members</li></ul>
Shreyaskumar Kathiriya	Data Scientist	<ul style="list-style-type: none"><li>- Clean the data</li><li>- Divide the task to everyone for running different code</li></ul>
Devarshi Pancholi	Business Analyst	<ul style="list-style-type: none"><li>- Lead the discussion of business implications from data that has been mining</li><li>- Research any possible business actions related to the data</li></ul>

