REPORT ON APPLICATIONS OF DATA SCIENCE IN E-COMMERCE

Data-Driven Insights for e-Commerce Success

By: Aayush Baviskar

ayushbaviskar44@gmail.com

Introduction:

Data science has become a crucial component for modern companies and organizations, helping them to improve consumer experiences, streamline processes, and drive decision-making. Data science is radically changing a variety of industries, from providing personalized product recommendations in e-commerce to personalizing healthcare treatments.

India, known to have the second-largest population in the world, has a sizable e-market. Online businesses have numerous opportunities for growth. Indian e-commerce is presently dominated by

many online retail giants, including Amazon, Flipkart, Snapdeal, eBay, Paytm, Tata CLiQ, Ajio, and others. Every month, these e-commerce portals host somewhere between 25 and 30 million potential consumers. According to studies, the e-commerce business in India is projected to increase at an annual pace of 11.45% to reach INR 7,591.94 billion by 2029 from INR 4,416.68 billion in 2024. With over 800 million users, UPI transactions reached INR 125.94 trillion in 2022.

How is data science used in e-commerce?

Online e-commerce businesses are already utilizing data science, machine learning, natural language processing (NLP), and big data analytics for their growth.

Big data analytics, machine learning, natural language processing (NLP), and data science are already being used by these e-commerce companies to support their expansion. The majority of them focus on using specific applications of data science, such as inventory management and optimization, pricing optimization, market basket analysis, recommendation engines using matrix factorization models, and customer segmentation models. These applications help them make more effective business choices by providing a deeper understanding of customer behavior and industry fluctuations, as well as by improving sales, marketing, and merchandising aspects of their operations.

Case Study Report:

The major online retailer eBay also uses data analytics to improve user experiences. Their recommendation algorithms increase user engagement and sales by making relevant product suggestions and optimizing search results. eBay has been able to stay competitive in the always changing e-commerce market because to its data-driven strategy.

Data science has also benefited eBay in the following ways:

- By optimizing search results, eBay has reduced bounce rates by 20%, a sign that users are finding what they're looking for more successfully.
- eBay's recommendation algorithms have increased average order value by 12% as customers are more likely to find and purchase complementary products.
- eBay has seen a rise in sales and revenue thanks to its data-driven tailored marketing strategies, which have an 18% better conversion rate than generic advertising.
- Because of their data-driven improvements to the user experience, eBay has outperformed many of its competitors with a 10% increase in revenue over the last year.

What are the common business questions tackled by data scientists in e-commerce?

Data scientists answer a variety of business questions in the e-commerce domain, such as:

1. **Customer Segmentation**: Who are our different customer groups, and what do they buy, and what are their purchasing behaviors?

2. **Personalization:** How can we provide personalized product recommendations to the customers that are to be purchased by them?
3. **Pricing Strategy**: How can we implement pricing to maintain competitiveness while maximizing profits?
4. **Fraud Detection:** How can electronic & online transaction fraud be found and stopped?
5. **User Experience Optimization:** How e-commerce Website/app can be optimized to improve user experience, increase customer happiness and conversion rates?
6. **Marketing ROI**: Determine the most effective marketing campaigns that yield significant returns on investment.
7. **Customer Lifetime Value (CLV)**: What is the anticipated lifetime value for different customer segments?
8. **Churn Prediction:** How can we keep consumers who are most likely to discontinue buying from us? **9. Sales Forecasting**: What are the future sales trends, and how can we prepare for them.
10. **Inventory Management**: based on demand predictions, which products should be stocked more or less?

Use Cases of Data Science in Retail
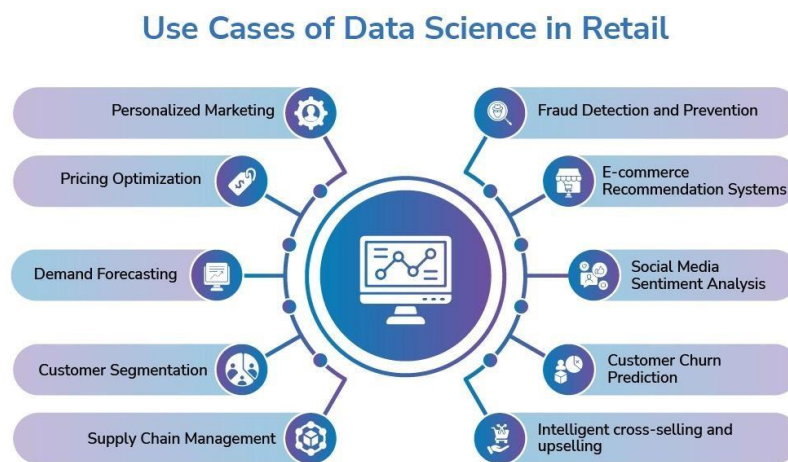
What is K-means clustering and how do Data Scientists use it in e-commerce?

Marketing campaigns have traditionally been targeted by using customer segmentation. Customers can be grouped together into categories like Loyal Customers, High Spenders, etc. based on the similarities between them (such as geographic resemblance, purchasing power, etc.). In other words, customer segmentation divides the customer base into more manageable and precisely targeted subgroups. Segments can be found using machine learning algorithms or by using reasoning.
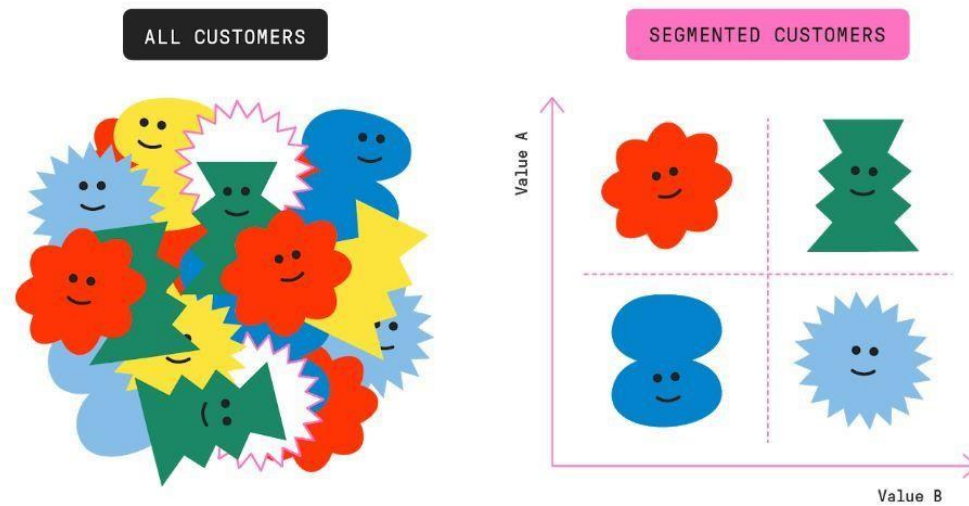
Customer segmentation models are produced using clustering algorithms. There are many ways available, but the K-means Clustering Technique is the most widely used method.

K-means Clustering is an Unsupervised Machine Learning algorithm, which groups the unlabeled, unclassified dataset into different clusters based on their similarities, patterns, and variations.

K-means is a distance- or centroid-based algorithm that uses distance measurements to determine how to allocate a point to a cluster. Each cluster in K-Means has a centroid attached to it.

In K-Means algorithm, our primary goal is to reduce the total distances between each location and the cluster centroid.

A key component of the k-means clustering approach is optimization. Finding the optimal set of centroids to minimize the sum of squared distances between each data point and its closest centroid is the aim of the optimization process. To define the number of clusters i.e. (K), the elbow method is used, it is a graphical method for finding the optimal K value in a k-means clustering algorithm. The **elbow graph** displays the yaxis values of the within-cluster-sum-of-squares

(WCSS) in relation to the various values of K (shown on the x-axis). The graph's elbow-shaped intersection is the ideal K value.
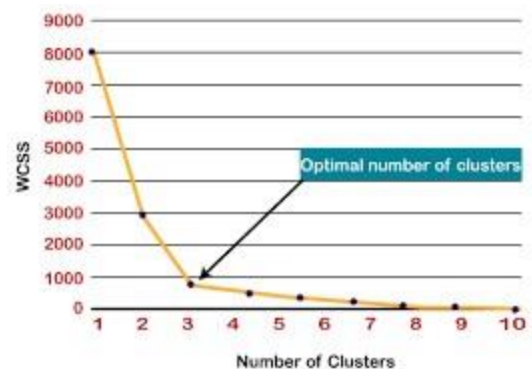
We use the **Euclidean distance measuring method** to calculate the K-means. The Euclidean distance is the length of the line segment in Euclidean space that separates two points. The number of dimensions affects the formula for Euclidean distance.

For Two dimensions:

$$D = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

Where **D** is the Euclidean distance and $(x1, y1)$ & $(x2, y2)$ are the cartesian coordinates of the two points.

For n-dimensions:

$$(x, y) = \sqrt{\sum_{i=}^{n} (x_i - y_i)} \qquad D \qquad 2$$

D is the Euclidean distance

**x** is Array or Vector X,

y is Array or vector Y

x_i is Values of horizontal axis in the coordinates plane

y_i is Values of vertical axis in the coordinates plane

n is the Number of observations

How K-Means Clustering Works:

1. **Initialization:** Choose K points at random from the dataset to begin. These will serve as the original centroids of the cluster.
2. **Assignment:** Determine the distance between each of the K centroids and each data point in the dataset. The cluster with the closest centroid to the data point should be assigned to it. K clusters are essentially formed in this step.
3. **Update centroids:** After every data point has been assigned to a cluster, take the mean of all the data points assigned to each cluster to recalculate the centroids of the clusters.
4. **Repeat:** Continue 2 and 3 until they converge. When a predetermined number of repetitions is reached or the centroids no longer exhibit considerable variation, convergence takes place.
5. **Final Result:** The algorithm outputs the final cluster centroids and assigns each data point to a cluster after convergence is reached.
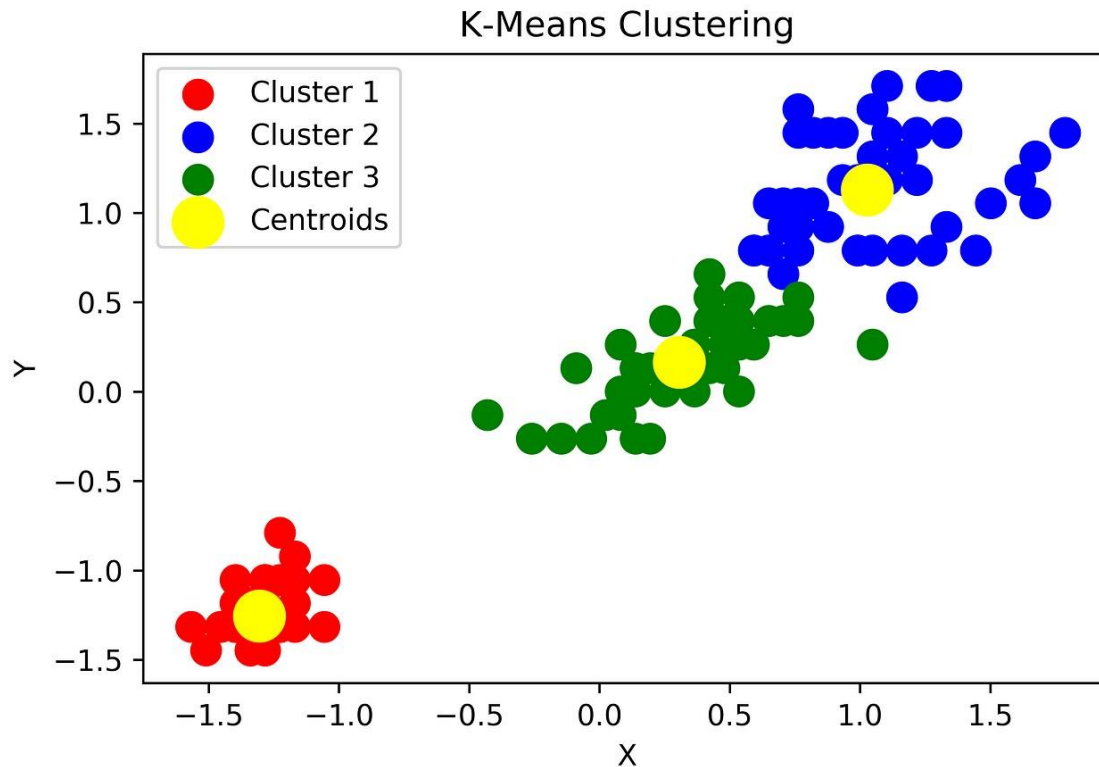
Conclusion:

Data scientists use K-means clustering in e-commerce businesses to categorize customers into buying behavior, age groups and online activities. By grouping customers who are alike, businesses can develop targeted marketing efforts, adjust suggestions according to individual tastes and preferences as well as increase and maintain their customers' loyalty. There are various customer segments identified through K-means that encompass high-end clients, occasional purchasers and discount seekers. Furthermore, it helps with inventory management by clustering items based on performance indicators such as sales and preferences, making prices best suited for all customers and improving website interaction experience overall. Using this strategy will make the e-commerce industry smart choices hence rise in revenues considering shoppers happiness.