Session – Summary

Introduction to Ecommerce

In this session, you understood the business models of ecommerce firms and learnt about the ways in which these firms earn revenue. You also saw the various steps and processes involved in the functioning of the ecommerce firms. In each of these steps, there are major problems faced by ecommerce firms and data analytics finds application in each of them.

Understanding the Ecommerce Business

The two most common types of ecommerce business models are the marketplace and the inventory-based models. Companies like Snapdeal follow the marketplace model. The items that you purchase on Snapdeal are not owned by Snapdeal. Snapdeal only acts as a platform where buyers and sellers interact. Buyers discover products on the website which are put up by various sellers, and although the prices and discounts are offered by Snapdeal, they are largely decided by the sellers. Such a business model is called the marketplace model. In marketplace models, the individual functions such as inventory, stock management, logistics, etc. are not actively done by the company itself. The company serves as a technology provider for a free but regulated marketplace.

Firms like Big Basket follow what is called the inventory model. The main feature of inventory model is that the company owns the product it is selling. Such firms manage an inventory, interface with customers, run logistics, and are involved in every aspect of business.

The major sources of revenue for an ecommerce firm are:

- Commission on transactions
- Shipping services
- Subscription-based services
- Value-added services
- Payment gateways
- Penalty fees

The following figure shows the different stages and steps in the ecommerce business.

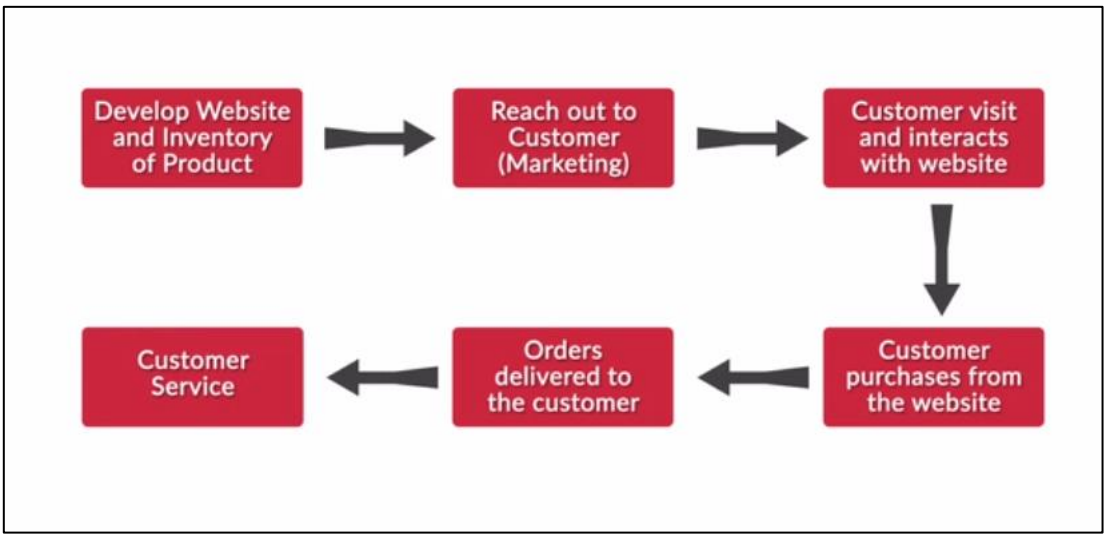

Figure 1: Ecommerce Value Chain

Data Analytics in Inventory Management

The first step in the ecommerce value chain is to plan and manage inventory stocks. You saw that there are two basic analytics applications in inventory management:

1. Demand forecast
2. Inventory management for the forecasted demand

While time series analysis is used for forecasting demand, this is not done on each product item individually. Products showing a similar demand curve are generally clustered together and a time series analysis is carried out on each of these clusters.

So, as you saw in the example, products like ACs, coolers, and refrigerators show a similar demand cycle, so they can be clubbed together in a cluster. Now, let's say, a shop has sold a total of 100 units from this cluster in a year. In addition to this data, you collect some past data as well, and after the time series analysis, you can predict that about 115 units will be sold next year. But now there is another problem — how many units each of ACs, coolers and refrigerators will you store with you?

To find the answer to this, you will again have to use past data to find out what percentage of the cluster is constituted by each of the sub-categories. Let's say that, out of all the cluster units sold, 20% are ACs, 45% are coolers, and 35% are refrigerators. So, you can estimate that, for the forecasted figure of 115 units, each of the sub-categories of AC, cooler and the refrigerator will be present in the same ratio of 20-45-35.

The next step is to manage the inventory for the forecasted demand. So, here again, you can optimise your cost and profit by acquiring, or getting rid of, extra inventory space depending on your forecasted demand.

Marketing in Ecommerce

The roles that analytics can play in marketing are as follows:

1) Mass marketing
   a) Channel analysis: Hypothesis testing to figure out the most cost-effective channel
   b) Content analysis: Hypothesis testing to figure out which content works the best

1) Personalised marketing
   a) Target customers: Using classification algorithms to target customers who are most likely to respond positively
   b) Personalised messaging: Using clustering to group customers as per their buying preferences, and sending personalised recommendations to them

Improving User Experience

Once you are able to attract your customers on to the platform using marketing, it is imperative to provide them a good user experience to increase their recency and frequency. Two tools to ensure this are recommendation systems and A/B testing. While developing an ecommerce website, there may be conflicting opinions about various elements, such as the shape of buttons, the colour of various UI elements, the text on the website, and numerous such things.

Some of these can be extremely crucial decisions — for example, you may have to choose features on the payment page which maximise the chances of a customer completing the payment. We have all come across buggy websites and apps, running away as soon as they ask us to fill some form or have an untidy payment gateway.

To resolve such conflicts, you can use A/B testing. A/B testing provides a way for you to test two different versions of the same element. Let's see how exactly this works.

Let's say that you're working on a website. A is the existing design of your website. This design is called the control. B is the newly designed version, which is called the variation. To conduct the A/B test, you have to divide the traffic between the control and the variation, i.e. A and B, and measure the performance of each design using the metrics which are important to your business. At the end of the test, the data collected will help you select the better performing version for your website.

When you go to an ecommerce platform, one of the most important factors that induces you to make the actual purchase is that how quickly you are able to locate the item that you want to buy. One way in which this is achieved is through a recommendation system, e.g. the 'customers who bought this item also bought…' section on Amazon.

Fraud Detection

As you have seen earlier, the profit made by an ecommerce firm depends on how much it can increase its revenue and decrease its losses. One way to decrease losses is to cut down the instances of fraud transactions. Fraud transactions cost a lot of money to ecommerce companies and also hurt their brand image. Mainly, there are two types of fraud:

1. **Payment fraud from customer:** These is the type of fraud in which a customer makes a fraud payment for a purchase. In this case, the customer bypasses the payment step pretending that he/she has made the payment when actually he/she has not paid.

   These types of frauds are generally solved by supervised machine learning algorithms.

   Now, as you have learnt, for a classified machine learning algorithm, you first need a training dataset. Thus, the first step is to generate labelled data for a large number of transactions on whether they are fraud or not. To do this, human raters are employed who confirm whether each transaction is fraud or not. Once you have a good enough data set, you have to extract the data on various independent variables which can be helpful in detecting fraudulent transactions.

   Some attributes that may be interested in detecting fraud are:
   - Buyer rating: You can analyse a buyer's past purchase history and rate whether the customer is a loyal one — loyal customers are less likely to commit frauds
   - Number of credit cards used by buyer: A high number of credit cards would signal suspicion and thus this is also an important feature
   - Number of previous fraudulent purchases by the buyer: Any past attempt of fraud is cause for alarm
   - Comparison of transaction amount with the previous average transaction amount of the buyer: If the new purchase amount is extremely large compared to the past average amount, it should again raise red flags

   Using a bunch of these variables, you can use a supervised learning model like neural networks, logistic regression, etc. to detect frauds. Each model gives you a probability of fraud.

   Now, since fraud is a sensitive issue both for the company as well as the customer, don't use a single probability cut-off to decide whether a transaction is fraudulent. There are two thresholds when a model is deployed. For example, any transaction with a fraud probability of greater than 0.7 is classified as fraud, and any transaction with a probability of less than 0.3 is classified as not fraud.

   However, all transactions which lie between 0.3 and 0.7 predicted probability are passed on to human raters to decide whether they are fraud or not. This way, the load on human raters is reduced and the system uses additional manually rated data to improve predictions in the future.

2. **Fulfilment frauds:** The second type of fraud is fulfilment fraud. These is a fraud which is committed when an order is being delivered to the customer. Some customers abuse the system by giving wrong personal information, cancelling a lot of orders after they have been shipped or simply returning all the shipments once they have been shipped to them. There are many reasons for which a customer may be prone to such activities:
    - Sabotaging — where competitor companies intentionally commit such frauds
    - People can also commit frauds to qualify for free shipping
    - They may sometimes try out new products and then return the items
    - Some people also misuse the cash on delivery option, i.e. they order products and then change their minds just before the delivery is made

    Generally, for these types of cases, data is analysed on a per customer basis.

    Some metrics that should be observed for these types of frauds are:
    - Percentage cancellation of total orders placed in the system
    - Percentage returns of total orders placed in the system

    The action points from these analyses can be that you may ban certain customers, or raise flags on suspicious orders and have them verified manually.

    Thus, you can see that analytics plays an extremely important role in identifying fraudulent payments and orders.

Shipment Delivery

Delivery problems can be of various types. One important problem is to figure out the exact latitude and longitude of the delivery address. This is not a problem when the address is typed correctly and can be easily found on Google Maps. But it often happens that the customer has not entered the address in a proper format or has made typos and errors in it. This is known as the address geocoding problem.

Another problem is the delivery of shipments from the manufacturer to the warehouse, and transportation of shipments between warehouses. Many orders are shipped across cities by flights, trains and road. The aim is to design better systems to reduce the cost of transport and increase efficiency.

There are two major legs in the delivery process:

1. Transit mile delivery: Transit mile deliveries are generally between warehouse/seller and another warehouse. This is necessitated because the inventory stock in different warehouses would be different. In this leg of delivery, many shipments are transported in bulk but are dropped at one single warehouse.

    The major challenge in this scenario is to figure out the optimal grouping of deliveries and efficient selection of the transport carrier.

2. Last mile delivery: Last mile deliveries are from warehouses to the end customers. Generally, the number of shipments delivered together in this case are not as high as in first mile deliveries, but there are multiple drop points.

   Generally last mile deliveries which are closer to each other are clubbed together, so that one delivery person can make maximum deliveries in minimal time and cost. In these scenarios, if we know the lat-long coordinates of drop off locations, we can use algorithms like clustering to group the nearby shipments together. Each cluster is then assigned to one delivery person to deliver.

Customer Feedback

Customer service is an integral part of customer experience. Customer service not only helps in addressing customer grievances, but it also serves as a feedback loop to the product.

Some metrics to track for customer feedback are:

1. Number of phone calls or emails on a weekly basis: A higher number of complaint calls reflect that some part of the process or product is broken and needs further attention
2. Average rating by customers: The customer feedback ratings that you provide on your purchased items, cab trips, restaurant outings or even UpGrad session ratings, all reflect how satisfied you are with the services offered

Apart from rating, a more granular understanding of customer feedback can be obtained through an analysis of the sentiment of text in the reviews. For example, in food delivery, you may have rated a restaurant low because the food was delivered late although it was good in taste. It is difficult to figure out the exact issue through the rating score only. But text analysis may reveal that people have used the words ‘late’ or ‘delay’ in their reviews, indicating that there’s a problem on the delivery front rather than food quality.