

# Understanding E-Commerce For Effective CRM

Rathod Harshrajsinh  
github.com/rathore287  
harshraj.rathore@gmail.com,  
Dublin.

**Abstract**—With Jeff Bezos being the richest person on the planet, it doesn't seem impossible a feat as Amazon has been in business for more than 20 years now. In these years, Amazon has become innovative leader in E-commerce market where heart of this innovation is in identifying Hard Trends that are bound to occur, one of which is "Anticipatory Shipping".

As the market saturates with a new technology, internet in this case, customer satisfaction drives the business to be the leader of the market segment. In this era of automation and E-Commerce, unquestionably the modus of operandi heavily relies on convenience. Retailers that don't involve into E-commerce are filing bankruptcy and companies such as Walmart is targeting E-commerce sales to compete with Amazon and ever-growing E-commerce segment. But, where logistics and automobile industry cause delays due to falling short by infrastructure that is relevant for quick delivery, predictive analytics prove useful crystal ball that estimates customer needs before the actual need arises. Using predictive analytics, Amazon can predict where purchase might occur before the actual order generates. So, when the purchase is made, it's already within hours range for delivery, due to having product available within proximity by being "anticipatory".

**Keywords**—*Anticipatory Shipping, Predictive Analysis, E-Commerce.*

## I. INTRODUCTION

The rise of E-Commerce is unavoidable but one of the impediments that it faces, is longer wait time for delivery. Over the years, Amazon has thrived on being one-stop shopping platform that recommends products to the customer over the internet by analysing customer data. Data is at the heart of E-commerce and Amazon gathers great data for analysing customer preferences and patterns for Anticipatory shipping. Hence, Amazon guards it's data very closely and keeps it secured, understandably so. Overall, Amazon reduces delivery time and logistics costs by shipping in bulk rather than air shipment of overnight delivery for ever growing prime members.

Amazon published data [1] containing 142 million user reviews for products, but this data is highly limited in its features and lacks necessary attributes to analyse Anticipatory Shipping. It's more suitable for sentiment analysis or NLP tasks.

In order to analyse E-Commerce efficiently, I have chosen Brazilian E-commerce data [2] which contains quite useful features for analysing E-commerce aptly. This data is published by Olist and contains orders from year 2016 to 2018, ranging from multiple states, product attributes, and order reviews. This is real commercial E-commerce data which is anonymised for public usage but contains useful features for predictive analytics. Olist operates on the E-commerce segment in Brazil and serves as SaaS company

which provides marketplace solutions to shop-owners by dealing as Store that connects consumers with retailers.[3]

### A. Motivation

When it was revealed that Amazon filed and obtained a patent for Anticipatory Shopping, its strategy took predictive analytics to next level for expanding its loyal customer base even further. Also, E-commerce analysis helps in understanding an important market segment that possesses convenience and discounts deals, arrived at the door.

### B. Research Question

What kind of deals and discounts can be presented to customer segment? What customer segments to target for membership? What trends can be interpreted for anticipatory shipping? Overall E-commerce analysis.

### C. Objective

Overall E-commerce analysis for augmenting product delivery and to help sellers do so more effectively for Operational CRM.

## II. METHODOLOGY

Analytical CRM supports daily back-office analytics operations that develop, support and enhance decision making capability of the organization by developing strategies that involve customers, but not directly. In this regard, Analytical CRM differs from Operational CRM. But unlike Operational CRM, where automation of marketing, salesforce and service that achieves customer satisfaction by direct interaction, Analytical CRM analyses customer data and unravels essential behaviours of customers, based on which capitalisation can be made further. Here, Analytical CRM aims to make reliable predictions on consumer behaviour and needs by integrating diverse data, ranging in various attributes from various Operational CRM systems which leads to predicting strong patterns, which in turn helps other CRM systems. [4]

In the following project, deployment is performed in framework which allows flawless execution that accomplish all necessary guidelines for successful project, which is overall analysis of complex issues of E-Commerce for Analytical CRM purposes.

### A. Business Objectives

This is crucial initial stage that focuses on careful goal analysis for understanding E-Commerce and what the business objectives, success-criteria and requirements are, which defines outcome to a successful deployment that is linked closest to success or failure of overall CRM systems. Hence, even before proceeding, it is critical to evaluate strategies based on short-term and long-term goals that

addresses the organizational needs by defining the scope of features leading to successful CRM implementation.

In the traditional brick and mortar store, items that are sold more regularly, are kept on the shelves as well as cellars and warehouses if demand for such products is frequent. Keeping such products within proximity makes sense because it minimizes both delivery time and costs on frequent products. But, doing one such task, on global scale with millions of products, is what Amazon has performed over the years by analysing trends and using the customer behaviour data for predictive analytics. Here, we will perform predictive analysis for case study on Brazilian E-Commerce Olist and see if similar results can be achieved by understanding E-commerce in detail.

## B. Selection

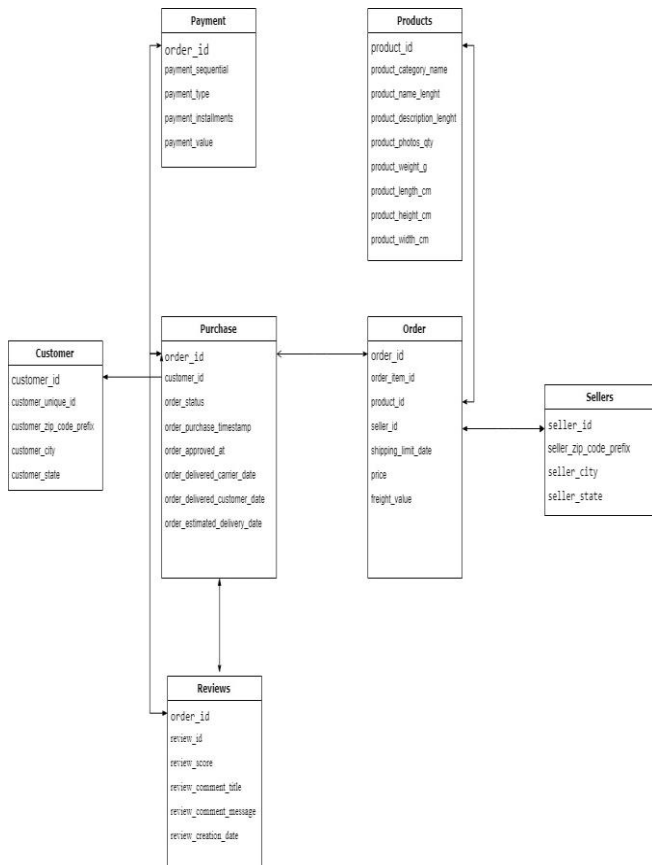


Figure. 1. data schema

The initial data collection pertaining to business objectives, from all the relevant and essential attributes of customers from various channels and sources, is utmost crucial for organisation to perform in-Depth analysis that facilitates adequate overview for the organization to develop hypothesis formations. At this stage, choosing the right data for analysis, that is integrated into a central repository knowledge base, helps to analyse long-term and short-term requirements for the scope of post-implementation service scenario which accounts for marketing, sales and services.

For this project, data for analysing Brazilian E-Commerce exists into multiple datasets with various attributes, which is merged together and displayed in the schema shown in Figure. 2. which provides better data understanding for discovering underlying patterns within.

Olist, one of the largest marketplaces in Brazilian, operates on selling products through the Olist web Store that is accessed directly as front to the customers, and deliveries are handled by Olist's logistics partners. In an example purchase scenario, seller receives notification for the product to fulfil the order, once the order is made from Olist Store. Then, customer receives the product within estimated delivery date, which is rated based on purchase experience. These orders are made of 100k+ orders from 2016 to 2018 with multiple features and dimensions, shape (118315, 39), of real commercial data that is anonymised. Here, references to the companies and partners have been replaced with the names of Game of Thrones great houses.

Such integration of data through various channels, helps determining set of rules for developing analytical methods to scale and optimize relationship with customers that provides benefits to both business and customers, by solving the questions suitable for business, using data. Results from this implementation for deployment, might help to enhance the efficiency of Operational CRM system to improve relationship and interaction with customers.

As mentioned in the schema, the several datasets that comprises the main data, are explained below.

### 1. Order Dataset

This dataset describes orders with the items purchased with 7 features as follows,

order_id	order unique identifier, which is primary key to <b>Purchase</b> dataset.
order_item_id	number of items included in the same order
product_id	product unique identifier
seller_ID	seller unique identifier
shipping_limit_date	seller shipping limit date for getting order to the logistic partner
price	item price
freight value	freight value of an item, with multiple items having freight value splitted between them.

### 2. Purchase Dataset

This dataset describes the actual purchase made with 8 features as follows,

order_id	unique order identifier, which is primary key to <b>Payment, Review</b> dataset.
customer_id	each order with a unique customer_id.
order_status	Reference to order status such as delivered, shipped, etc.

order_purchase_timestamp	Shows actual purchase timestamp.
order_approved_at	Shows timestamp for approval of payment.
order_delivered_carrier_date	Shows the order timestamp for handing to the logistic partner for delivery to customer.
order_delivered_customer_date	Actual delivery timestamp for the order to the customer.
order_estimated_delivery_date	Estimated delivery date used to inform customer at the purchase moment.

### 3. Sellers Dataset

This dataset describes sellers which fulfils orders which are made at Olist and contains 4 features as follows,

seller_id	seller unique identifier, which is primary key to <b>Order</b> dataset.
seller_zip_code_prefix	first 5 digits of seller zip code.
seller_city	-
seller_state	-

### 4. Customers Dataset

This dataset describes customer's information necessary for delivering the orders delivery location. But, as each order is assigned to a unique customer\_id, the same customer will get different id for different orders. Here, customer\_unique\_id allows identification of customers that made repurchases. This dataset contains 5 features as follows,

customer_id	each order with a unique customer_id, which is primary key to <b>Purchase</b> dataset
customer_unique_id	unique identifier of a customer making repurchases.
customer_zip_code_prefix	first five digits of customer zip code
customer_city	-
customer_state	-

### 5. Products Dataset

This dataset describes products information that are sold by Olist, with 9 features as follows,

product_id	unique product identifier, which is primary key to <b>Purchase</b> dataset.
product_category_name	root category for the product in Portuguese
product_name_length	number of characters used that are extracted to describe the product name.
product_description_length	number of characters used that are extracted to describe the product description.
product_photos_qty	Number of photos uploaded for the product.
product_weight_g	-
product_length_cm	-
product_height_cm	-
product_width_cm	-

### 6. Payments Dataset

This dataset describes orders payment information, with 5 features as follows,

order_id	unique order identifier, that is primary key to <b>Purchase</b> dataset.
payment_sequential	customer might pay for order with more than one payment method. In that case, created sequence will accommodate all payments
payment_type	method for payment
payment_installments	number of installments chosen for the purchase
payment_value	transaction value

### 7. Order Reviews Dataset

This dataset describes the reviews made by the customers after purchase, to note purchase experience, sometimes with comments. It contains 7 features as follows,

review_id	unique review identifier
order_id	unique order identifier, which is primary key to <b>Purchase</b> dataset.
review_score	rating given by customer that ranges from 1 to 5 as level of satisfaction

review_comment_title	Comment title of review from customer, in Portuguese
review_comment_message	Comment of the review from customer, in Portuguese
	timestamp for when the satisfaction survey was sent to the customer
review_answer_timestamp	timestamp for when satisfaction survey was answered.

The shape of final data is (118315,39) with some missing data in 12 of the features that mostly includes reviews and comments posted by the customers, followed by some of the product information. These values have to be treated differently as per their correlations and impact on the data.

### C. Findings

Descriptive statistics, with functions such as mean, inter-quartile range, max, min, provides effective summary of the numerical features which enhance insight into distributions that can promote increased sales by adopting analysis that target customer satisfaction.

	count	mean	std	min	25%	50%	75%	max
order_item_id	118315.0	1.20	0.70	1.00	1.00	1.00	1.00	21.00
price	118315.0	120.65	184.11	0.85	39.90	74.90	134.90	6735.00
freight_value	118315.0	20.03	15.84	0.00	13.08	16.28	21.18	409.68
seller_zip_code_prefix	118315.0	24440.79	27571.68	1001.00	6429.00	13660.00	27946.50	99730.00
payment_sequential	118315.0	1.09	0.73	1.00	1.00	1.00	1.00	29.00
payment_installments	118315.0	2.94	2.78	0.00	1.00	2.00	4.00	24.00
payment_value	118315.0	172.58	267.10	0.00	60.85	108.20	189.26	13664.08
review_score	118315.0	4.01	1.40	1.00	3.00	5.00	5.00	5.00
customer_zip_code_prefix	118315.0	35051.68	29824.81	1003.00	11310.00	24310.00	58640.00	99990.00
product_name_lenght	116606.0	48.77	10.03	5.00	42.00	52.00	57.00	76.00
product_description_lenght	116606.0	785.94	652.58	4.00	346.00	600.00	983.00	3992.00
product_photos_qty	118315.0	2.17	1.73	0.00	1.00	1.00	3.00	20.00
product_weight_g	118295.0	2112.33	3786.72	0.00	300.00	700.00	1800.00	40425.00
product_length_cm	118295.0	30.27	16.19	7.00	18.00	25.00	38.00	105.00
product_height_cm	118295.0	16.62	13.45	2.00	8.00	13.00	20.00	105.00
product_width_cm	118295.0	23.08	11.75	6.00	15.00	20.00	30.00	118.00

Figure. 2. Overall descriptive statistics.

The Figure. 2. above, describes the succinct summary of all numerical features within dataset that help provide quick overview of numeric attributes for effective operations.

Within business model, sellers facilitate conception of the products. They also handle distributions to the logistic partner for the produced goods. Hence, assessment of sellers gives crucial knowledge base as inbound logistics for entire operations. Within dataset, there are 3095 unique sellers that are comprised within 23 unique states, from 611 different

cities of Brazil. Here's pie chart for states where most sellers operate from.

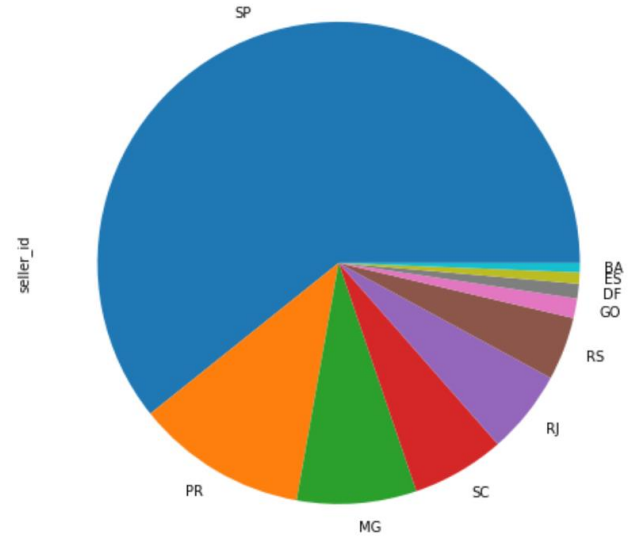


Figure. 3. sellers by state

The pie chart from Figure. 3., displays top 10 states from where most sellers come from. Here, almost 2/3rd of the sellers come from Sao Paulo, followed by Paraná, Minas Gerais, Santa Catarina and Rio de Janeiro, total of which comprise almost 90% of the sellers. Further analysis on the sellers by cities reveal that Sao Paulo makes up less than 25% sellers, followed by Curitiba and Rio De Janeiro, 'The First Capital of Brazil'. But, rest of the sellers are distributed within several cities, almost 611 of them.

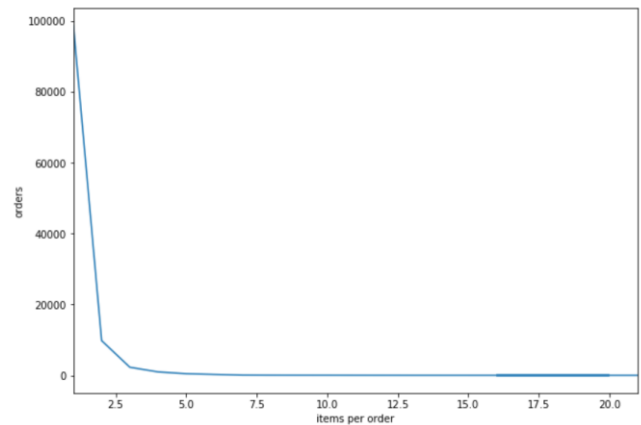


Figure. 4. Items per Order.

Orders stimulate the growth and make up the most primal aspect of business in any domain. For this project, we analyse orders, 98665 of them, based on unique order\_id, but duplicates are removed within dataset to provide accurate analysis. As seen Figure. 4., bulk of the orders are comprised of less than 3 Items and most orders are made for singular item within sales cycle.

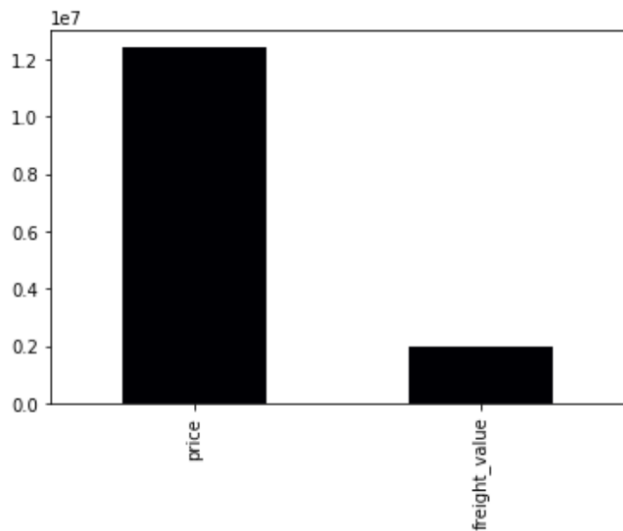


Figure 5. Comparison of Price vs Freight Value of Orders

In the span of years 2016 to 2018, orders made at Olist were priced around 12 million accumulatively, whereas freight value paid for the logistics for them accounted for almost 2 million for the total orders of 98665, for 32126 unique products by 3095 unique sellers.

Out of customer base of 95419 unique customers, only 2673 purchased twice and 240 returned for repeated purchases. Hence, total number of customers lost within time frame of years 2016 to 2018 remained highest, 92506, indicating recurring business value lost as customer churn.

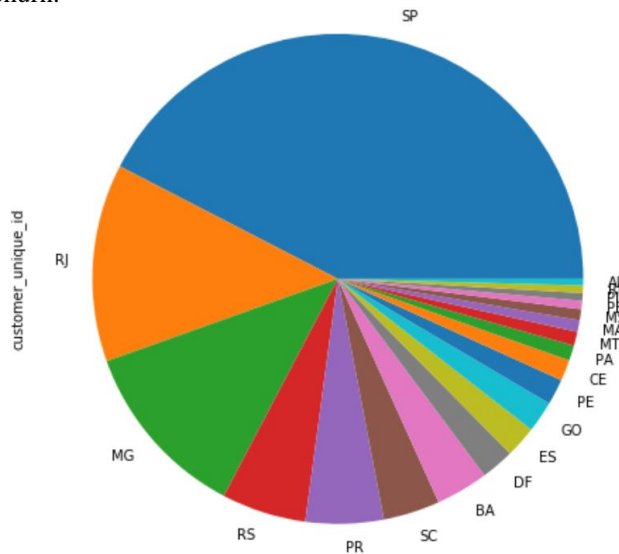


Figure 6. unique Customers by state

The pie chart in Figure 6., describes top 20 states where unique customers are distributed, with more than half of the customers residing in Sao Paulo and Rio.

There are 32126 unique products, within total orders of 98665. These products are distributed in 73 different categories. The pie chart in the Figure 7., describes the top 20 categories of products, where most orders are distributed in.

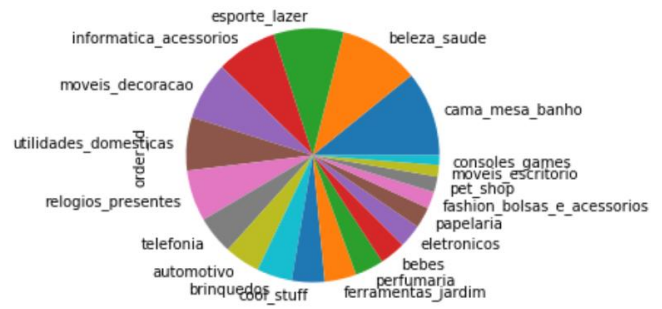


Figure 7. Product categories with higher orders.

The top category of product sold is bed-bath-table, followed by sports-leisure, and home decorations etc. which can help in Operational CRM to make better decision based on customer preferences and whether customers orders are family and home oriented.

order_id		payment_value	
payment_type	payment_type	payment_type	payment_type
boleto	19614	boleto	2842240.16
credit_card	74731	credit_card	12310049.88
debit_card	1521	debit_card	215079.53
voucher	2799	voucher	216183.21

Figure 8. Orders and their value received based on Payment type

Olist facilitates 4 methods of payment which accounts for all the customer payments within the country. As seen in Figure 8., the highest number of orders by far are paid through credit card, and boleto is second most popular cash-based payment method, which allows payment for those who do not have a bank account. In comparison, debit\_card and voucher are less desirable payment methods. Here, Olist can collaborate with credit card companies to advertise and promote their banks for a charged fee.

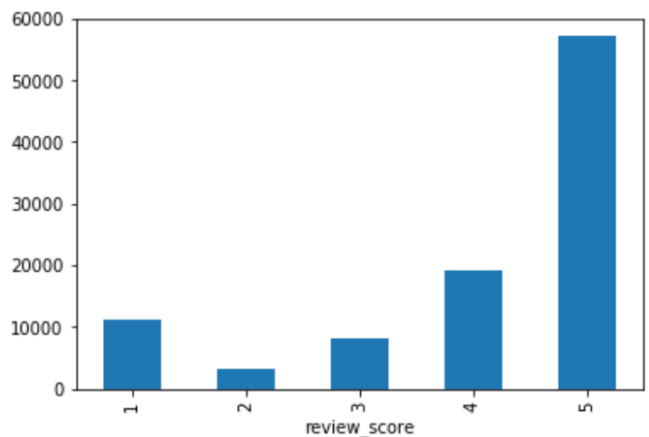


Figure 9. Reviews for the orders.

In the orders, most are given the highest ratings. But there exists no missing data for review\_score, which is highly unusual as this is a real commercial data and most users in real world tend not to submit reviews. Hence, it might not be a wrong assumption to be made that Olist released data mostly consisted of higher ratings by effort and not randomly by chance.

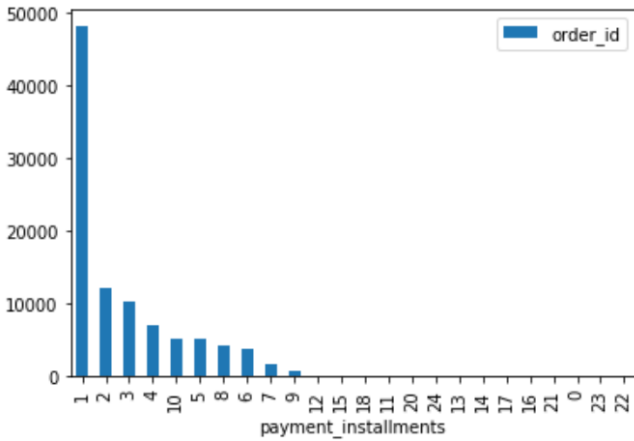


Figure. 10. Number of payment installments taken for orders

As seen in figure. 1., most orders are paid in one installment, possibly because only credit card payments allow payment in more than one installment.

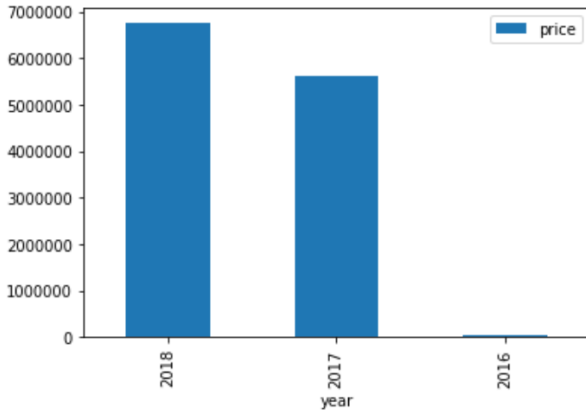


Figure. 11. Accumulated yearly changes in price.

Olist published the data from Sept 2016 to Sept 2018 and as seen in Figure. 11., sales in 2016 seems quite low in comparison. Sales grow higher gradually and it is highest in 2018.



Figure. 12., Total order status(Left) with weekly order status accumulatively(Right) for all orders.

Total deliveries made are significantly higher in comparison to minuscule cancelled orders, as seen in Figure. 12(Left). On the other hand, approved and other processes seem to be ongoing operations reflected. Also, as shown in graph in Figure. 12.(Right), most deliveries are made within weekdays in comparison to weekends.

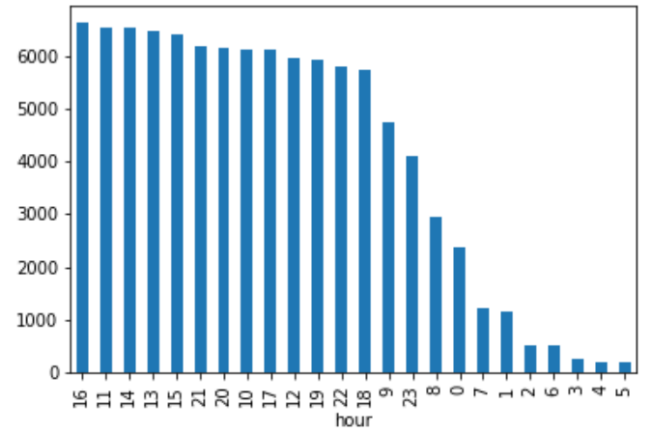


Figure. 13., Hourly orders from highest to lowest.

As seen in Figure. 13., usually from 12:00 a.m. to 8:00 a.m., order size seems lower, in comparison to rest of the day, with peak hour being 16:00 for highest business orders made.

### III. RELATED WORK

From the patent by Amazon [5], predictive analytics and usage of intermediary hubs can be inferred but predictive algorithms or data for such analysis is nowhere to be found. Here, [5] Amazon describes Anticipatory Shopping process as "packaging one or more items for eventual delivery" by shipping the products from Fulfilment Centres, which can be Seller's Distribution Centre or Amazon's own. From such Distribution Centre, product goes to targeted geographical area "Hub" beforehand in advance without specifying delivery address by anticipating eventual order in that geographical area. Though, within supply chain, idea of anticipatory product availability is not new. It's not unusual for manufacturer to ship products beforehand to the retailer's distribution centre. But this DC Bypass strategy requires strong IT-enabled collaboration ties between manufacturer and retailers for accurate forecasting with careful planning and tracing for execution of successful DC Bypass.

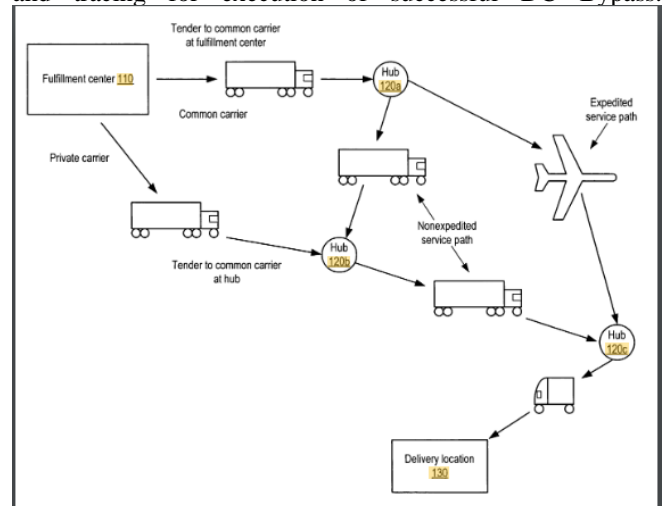


Figure. 14. Anticipatory shipping from Amazon.

Here, the above diagram in Figure. 14., process starts with distribution/fulfilment centre through one or more hubs of geographical area within close proximity where items are kept, till order is matched with the items available. If items don't match with the orders or predictive analytics algorithm

makes error, package gets redirected to next geographical area but that leads to higher logistics costs which can be mitigated by giving away promotional gifts and discounts on such items. Hence, only regression analysis and classification might not be sufficient to forecast shopping trends that play key role in successful Anticipatory Shipping for the E-Commerce.

Lee (2016) provides in-depth analysis of Anticipatory Shipping [7] and facilitates valuable insight into the process. Whereas (Pereira et al., 2018) builds on the findings by paper Lee (2016) gives high level understanding on the subject by demonstrating conceptual modelling and test cases. [8]

Genetic algorithm-based optimisation model involves clustering by association rule mining to accurately predict purchase patterns and if-then prediction rules. But, modelling customer behaviour in this manner requires not only multichannel but omnichannel retailing where seamless convergence between online, offline, and mobile channels is necessary, leading to fourth generation retail revolution, retail 4.0. [7]

When it comes to customer placing order at online or offline store, fast delivery is one of the most critical aspect in omnichannel retailing but it requires both product data and customer data across all the channels, to form market basket analysis and in some cases, perform designed marketing and promotional strategies to lure customers. But, in any case, anticipatory shipping faces a dilemma where improving one factor sabotages the other. Factors, in this case, are cost and distance. For example, when cost required for product shipping is low for DC A and high for DC B. But, what if prediction confidence is higher for DC B. Hence, multiple factors such as transportation cost, travelling time, and prediction confidence come into account for accurately perform successful Anticipatory Shipping.

While performing implementation of Anticipatory Shipping, oscillations in sales volumes for supply-demand require adoption of complex systems that can not only utilize intelligent decision methods that adapt to new information easily with technology, but also minimises uncertainty. By combining machine learning and simulation based optimisation, omnichannel retailing supply chain management can be achieved, that converges global economy with information technology which leads to advancement of E-Commerce that can deal with increasing demand on higher level logistics services that requires agile and dynamic supply chain. In order to handle major transformation in retail logistics, multichannel or even omnichannel supply chains are required, which support data and business process to function in synergy to minimise delays and inefficiencies.

Such model is proposed by (Pereira et al., 2018) which is a conceptual model for predictive and adaptive management approach for omnichannel retailing supply chain that combines machine learning with simulation-based optimisation methods, that handles different stages of supply chain which focuses exclusively at anticipatory shipping. This model uses Agent-based simulation (ABS) model which can be utilized to analyse different scenarios such as delays and distorted information to simulate inefficiencies of supply chain. Here, solution from machine learning gets evaluated

with in simulation model that optimizes the process until stopping criterion is achieved. In this conceptual model, machine learning performs forecasting that reduces uncertainties by anticipating customer behaviour accurately using clustering algorithms and artificial neural networks that analyses demand of each product and feeds the data to simulation based model that anticipates the product distribution activities that lead to lower lead time and costs. This simulation model handles the incompatibility between forecast and demand to adopt with distortions and ensures sales of forecasted products to the customers optimally.

## CONCLUSION

In today's world of innovations and progress, businesses which reliably interact with customers to fulfil their needs before needs arise, such businesses thrive, whereas others don't survive. In this project, we try to understand the E-commerce through various perspectives, which allows sales analytics to strategize selling opportunities by boosting CRM performance to provide for favorable customer experience. An efficient CRM can provide 360-degree customer view based on data from various channels, which help solve lower retention rates, referral opportunities and damage to brand loyalty. But, such pipelines need to be enhanced constantly forevermore as demand and competition increase in tandem.

## REFERENCES

- [1] McAuley, J. (2019). *Amazon review data*. [online] Jmcauley.ucsd.edu. Available at: <http://jmcauley.ucsd.edu/data/amazon/links.html> [Accessed 7 Aug. 2019]
- [2] Kaggle.com. (2019). *Brazilian E-Commerce Public Dataset by Olist*. [online] Available at: <https://www.kaggle.com/olistbr/brazilian-ecommerce> [Accessed 7 Aug. 2019].
- [3] Olist. (2019). *OLIST - ANUNCIAR em MARKETPLACES e ACELERAR suas VENDAS*. [online] Available at: <https://www.olist.com/> [Accessed 7 Aug. 2019].
- [4] En.wikipedia.org. (2019). *Customer relationship management*. [online] Available at: [https://en.wikipedia.org/wiki/Customer\\_relationship\\_management](https://en.wikipedia.org/wiki/Customer_relationship_management) [Accessed 7 Aug. 2019].
- [5] Patents.google.com. (2019). *US8086546B2 - Method and system for anticipatory package shipping - Google Patents*. [online] Available at: <https://patents.google.com/patent/US8086546B2/en> [Accessed 7 Aug. 2019].
- [6] Logistics Viewpoints. (2019). *Bakers Shoes Implements DC Bypass - Logistics Viewpoints: A Blog for Logistics, Supply Chain, and 3PL Executives*. [online] Available at: <https://logisticsviewpoints.com/2013/02/12/guest-commentary-bakers-shoes-implements-dc-bypass/> [Accessed 7 Aug. 2019].
- [7] Lee, C.K.H.. (2016). A GA-based optimisation model for big data analytics supporting anticipatory shipping in Retail 4.0. *International Journal of Production Research*. 55. 1-13. 10.1080/00207543.2016.1221162.
- [8] Pereira, M., de Oliveira, D., Portela Santos, P. and Frazzon, E. (2018). Predictive and Adaptive Management Approach for Omnichannel Retailing Supply Chains. *IFAC-PapersOnLine*, 51(11), pp.1707-1713.