

A Project Report on

**Customer Base Analytics In E-Commerce**

Submitted in partial fulfilment for award of degree of

**Master in Business Administration**

In

**Business Analytics**

Submitted by

**Sanjeev Kumar Jha**

**R19MBA06**

Under the Guidance of

**Dr. J.B.Simha**

Chief Mentor

REVA Academy for Corporate Excellence

**REVA University**

Rukmini Knowledge Park, Kattigenahalli,

Yelahanka, Bangalore – 560064

**September, 2020**



# Candidate’s Declaration

I, Sanjeev Kumar Jha hereby declare that I have completed the project work towards the Master in Business Administration at, REVA University on the topic entitled **Customer Base Analytics In E-Commerce** under the supervision of Dr.J.B.Simha, Chief Mentor. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year 2020.

Sanjeev Kumar Jha

Place: Bengaluru Name of the Student:

Date: 13th Sep 2020 Signature of Student



# Certificate

This is to Certify that the PROJECT work entitled Customer Base Analytics in E-Commerce carried out by Sanjeev Kumar Jha with SRN **R19MBA06**, is a bonafide student of REVA University, is submitting the project report in fulfilment for the award of Master in Business Administration in Business Analytics during the academic year 2020. The Project report has been tested for plagiarism, and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of PROJECT work prescribed for the said Degree.

<Signature of the Guide> <Signature of the Director>

<Name of the Guide> <Name of the Director>

Guide Director

External Viva

Names of the Examiners

1. <Name> <Designation> <Signature>
2. <Name> <Designation> <Signature>

Place: Bengaluru

Date: 13th Sep 2020

# List of Abbreviations

|  |  |  |
| --- | --- | --- |
| **Sl. No** | **Abbreviation** | **Long Form** |
| 1 | RFM | Recency, Frequency, Monetary |
| 2 | RFMT | Recency, Frequency, Monetary, Tenure |

# Abstract

The Customer Base Analysis (Segmentation) of E-commerce customer into multiple categories can contribute to a better understanding and characterization of purchasing behaviour in the E-commerce market. E-commerce shopping databases consist of multiple kinds of data on customer purchasing activity and demographic characteristics, as well as consumption. Information about customers uncovered by segmentation enables company administrators to establish good customer relations and refine their marketing strategies to match customer expectations. To achieve optimal segmentation, we developed a RFM and Clustering approach to classify E-commerce customers based on their purchasing data across categories. A technique derived from the K-Means Clustering is used to create the customer segments. Variational approximation is leveraged to generate estimates from the segmentation in a computationally-efficient manner. The proposed Clustering method yields more promising results than RFM Base.

**Contents**

[Candidate’s Declaration 2](#_Toc50912622)

[Certificate 3](#_Toc50912623)

[List of Abbreviations 4](#_Toc50912624)

[List of Figures 4](#_Toc50912625)

[List of Tables 4](#_Toc50912626)

[Abstract 5](#_Toc50912627)

[Chapter 1: Introduction 7](#_Toc50912628)

[1.1: Introduction 7](#_Toc50912629)

[Chapter 2: Literature Review 9](#_Toc50912630)

[Chapter 3: Problem Statement 12](#_Toc50912631)

[Chapter 4: Objectives of the Study 13](#_Toc50912632)

[Chapter 5: Project Methodology 14](#_Toc50912633)

[Chapter 6: Business Understanding 16](#_Toc50912634)

[Chapter 7: Data Understanding 18](#_Toc50912635)

[Chapter 8: Data Preparation 20](#_Toc50912636)

[Chapter 8: Data Modeling 26](#_Toc50912637)

[Chapter 9: Data Evaluation 28](#_Toc50912638)

[Chapter 10: Deployment 28](#_Toc50912639)

[Chapter 11: Analysis and Results 29](#_Toc50912640)

[Chapter 12: Conclusions and Recommendations for future work 33](#_Toc50912641)

[Bibliography 34](#_Toc50912642)

[Appendix 36](#_Toc50912643)

[Plagiarism Report 36](#_Toc50912644)

[Publications in a Journal/Conference Presented/White Paper 36](#_Toc50912645)

[Any Additional Details 36](#_Toc50912646)

# Chapter 1: Introduction

# 1.1: Introduction

The Customer Base Analysis (Segmentation) of E-commerce customer can be described as a set of customers who have similar characteristics of demography, behaviours, values. The segmentation techniques have become very popular with the developments in information and E-commerce business, especially database management systems and data mining have changed the way of marketing. The vast availability of data and the inefficient performance of traditional statistical techniques (or statistics-oriented segmentation tools) on such voluminous data have stimulated researchers to find effective segmentation tools in order to discover useful information about their customers. Thus, knowledge discovery (KD) and data mining (DM) have been seen as a solution to this problem. Disciplines such as machine learning, statistics, artificial intelligence (soft and hard computing techniques), expert systems, data and knowledge management technologies are incorporated with KD and DM by making use of their theories and algorithms. Marketing researchers are interested in the application of these technologies in marketing-related problems, such as forecasting, segmentation, knowledge-based marketing decision support systems, and so forth, especially in the frame of DM. Soft computing, as a family of data mining techniques, has been recently started to be exploited in the area of segmentation and it stands out as a potential area that may be able to shape the future of segmentation research. The significant usage of soft computing techniques in business-related problems, particularly in segmentation, makes segmentation problems more attractive, since these techniques are very effective and applicable. From this perspective, the objective of this article is to find out where the future of segmentation is heading towards in terms of being able to obtain effective segmentation results.

Segmentation technique is often used to discover profitable quotient. RFM (recency, frequency, and monetary) model is the most frequently adopted segmentation technique and has a long history in the application of direct marketing. Decision makers can effectively identify valuable customers and then develop effective marketing strategy by adopting RFM model. Recent studies suggest that the predictability of RFM models can be improved via adding additional variables when predicting customer behaviour’s, limited to the use of RFM model, firms cannot effectively distinguish between the short-life and long-life customers. In light to the above, extends RFM model to RFMT model by adding an additional variable – Tenure (T) in a E-commerce Business to effectively identify valuable customers marketing. Before using RFMT model, the number of clusters must be determined. The K-Means clustering has been widely used to segment customers when applying RFM model for lesser data. K-Means clustering can help the market managers easily recognize the customer segments precisely and compare market maps over time and monitor market responses of every segment, the data set is clustered by K-Means clustering to determine the number of clusters. The transaction data consist of 8671 unique customers who visited a E-commerce January 1, 2019 to Dec 31, 2019. The profile for each customer includes the Order number, Order Date, City, SKU Code, Size, Color, Quantity, Invoiced, Gross Margin, the days from the first visit date to the last visit date, the last visit date and visit frequency.

# Chapter 2: Literature Review

* 1. **. RFM and RFMT models**

Data mining techniques have been widely applied to different domains. As the transactions of an organization becomes much larger in size, data mining techniques, particularly the clustering/segmentation technique, can be used to divide all customers into appropriate number of clusters based on some similarities in these customers (Huang, Chang, & Wu, 2009).

RFM model is a well-known customer value analysis method widely applied to segment customers (Chang, Huang, & Wu, 2010; Cheng & Chen, 2009). It is a behaviour-based model to analyse the behaviour of a customer and then make predictions according to the behaviour in the database (Hughes, 1996; Yeh et al., 2008). It consists of three measures – recency, frequency, and monetary and are combined into a three-digit RFM cell code. Recency measures the number of periods since the last purchase (i.e., days or months). Frequency measures the number of purchases made in a given time period. Monetary measures the total amount of money spent during a given period of time, or the average rupees amount per purchase or all purchases to date. The general way to use RFM model in customer behaviour analysis is to sort the customer data by each dimension of RFM and then divide them into Six equal quintiles. For recency, the customer database is sorted by purchase dates by descending order. The top segment is given a value of 6 and the others are discerningly assigned of 5, 4, 3, 2 and 1. For frequency and monetary, to sort customer visiting frequency data and the customer data related to the amount of the money spent in descending order, respectively. The top 20% is also assigned the value of 6. The value of 5 is also given to the next 20% and so on (Hughes, 1996; Kahan, 1998; Tsai & Chiu, 2004). Some literature has attempted to develop new RFM models to test whether they perform better than the traditional RFM models by taking additional variables into account (Hosseini et al., 2010; Yeh et al., 2008). For example, Yeh et al. (2008) selected targets for direct marketing from a database using a modified RFM model, namely RFMT, by adding Tenure parameters, i.e., time since first purchase (T). RFM model cannot segment which customers have long-term or short-term relationship with the company (Reinartz & Kumar, 2000). The customer loyalty depends on the relationship between a firm and customers (Chow & Holden, 1997), revealing that the key of customer loyalty is built from a long customer relationship management (Chang & Tsay, 2004). In this regard, RFM model is extended as RFMT model by taking Tenure (T) into account.

* 1. **. K-Means/ Hierarchical Clustering models**

Clustering/Segmentation technique is used to identify a set of groups that both minimize within-group variation and maximize between-group variation according to a distance or dissimilarity function. Clustering for problem solving involving tasks such as clustering techniques of K-means and Hierarchical.

For segmentation problems, previous research suggests that hierarchical approaches do not perform very well with large data sets (Kuo et al., 2002a). Due to the fact that hierarchical methods build a tree structure using a dendrogram, they are not able to provide a unique clustering because partitioning to cut the dendrogram above a certain level becomes imprecise The process of cutting the dendrogram is usually done by visualising the dendrogram through taking into account the distance between cluster centres, which can be considered as an arbitrary process. Moreover, non-hierarchical or partitional methods work based on the assumption that the number of clusters and initial cluster points (not necessarily) are pre-defined, and this affects the final cluster solution (Lee, Lee, & Wicks, 2004). However, integration of hierarchical and partitional methods makes the clustering result powerful, especially in large databases (Kuo et al., 2002b). In customer segmentation problem, there are only a few studies that combined two clustering methods together. Punj and Steward (1983) first introduced a two-stage clustering concept by combining a hierarchical (Ward’s minimum variance) and a non-hierarchical technique (K-means).

# Chapter 3: Problem Statement

The purpose of this Analysis is to identify profitable customers in a E-Commerce Business. This E-Commerce Business begins its operation since March 1, 2018. In order to identify most profitable Customers, it is necessary to consider the unique order number and SKU code in this E-Commerce. Thus, the time period taken in this study to measure the frequency and Monetary becomes essential as well. This study collects the data set with 8671 unique customers who visited a E-commerce site from January 1, 2019 to Dec 31, 2019.

* E-Commerce Business is a fashion category multinational company with a turnover of more then one billion dollars. It has 40 factories across the globe.
* E-Commerce Business is a profitable company which makes money using their E-commerce website and few other marketplace like Amazon and Flipkart.
* E-Commerce Business products are sold through online and offline channels both company owned and with partners.
* Their clients includes - Victoria’s Secrets, Marks & Spencer, Triumph, Speedo, Calvin Klein among others.

Order Number is unique for all the purchase. The SKU Code is classified into many categories based on cloth type. The SKU Code end with “XL” Means extra-large, “L” Means large, “M” Means medium, “S” Means small and etc..

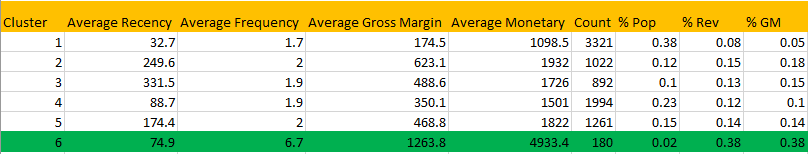
# Chapter 4: Objectives of the Study

Here we are trying to find out the Customers Segmentation of E-commerce customer can be described as a set of customers who have similar characteristics of demography, behaviours, values.

Clustering/Segmentation technique is used to identify a set of groups that both minimize within-group variation and maximize between-group variation according to a distance or dissimilarity function. Clustering for problem solving involving tasks such as clustering techniques of K-means and Hierarchical.

Which will help the marketing team to Upsell cross sell Potential in Medium value segments and Targeted campaigns and offers.

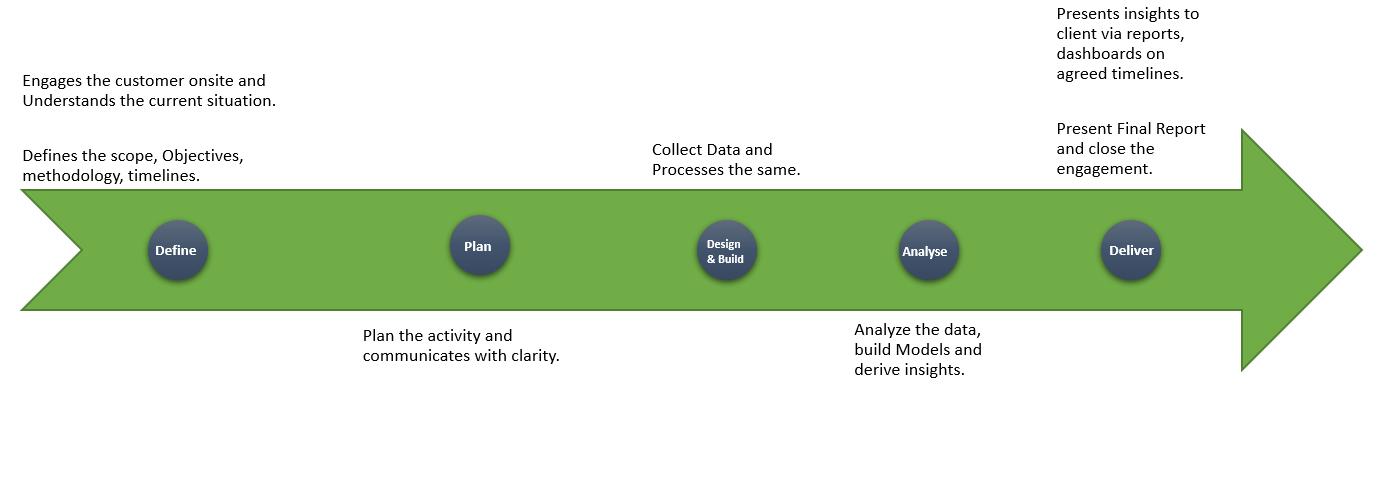
Customer Value Matrix to depict different clusters of customers to provide unique marketing strategies for each cluster of customers. The segmentation process created by Customer Value Matrix requires the calculation of the average values for frequency (F) and monetary (M). The length of customer relationship measures the number of days from the first visit date to the last visit date (T) in this paper.



# Chapter 5: Project Methodology

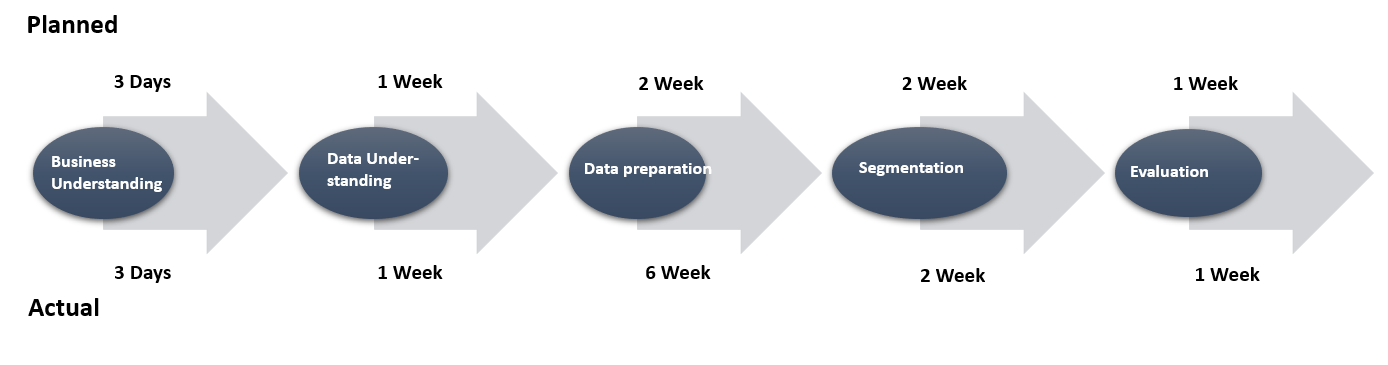
The project methodology comprises of below point: -

* **Define**: - Engages the customer onsite and Understands the current situation. Defines the scope, Objectives, methodology, timelines.
* **Plan**: - Plan the activity and communicates with clarity.
* **Design & Built**: - Collect Data and Processes the same.
* **Analyse**: - Analyse the data, build Models and derive insights
* **Deliver**: - Presents insights to client via reports, dashboards on agreed timelines. Present Final Report and close the engagement.



**Project Pipeline**

The project pipeline describes the project roadmap over the time. Basically, the planned roadmap vs actual roadmap (as shown in the image below)



The above image shows how the project has be planned using CRISP-DM methodology which included

Business Understanding: - Focuses on understanding the project objectives and requirements from a business perspective. The analyst formulates this knowledge as a data mining problem and develops preliminary plan

Data Understanding: - Starting with initial data collection, the analyst proceeds with activities to get familiar with the data, identify data quality problems & discover first insights into the data. In this phase, the analyst might also detect interesting subsets to form hypotheses for hidden information

Data Preparation: - The data preparation phase covers all activities to construct the final dataset from the initial raw data

Segmentation/Clustering : - The analyst evaluates, selects & applies the appropriate modelling techniques. Since some techniques like neural nets have specific requirements regarding the form of the dat. There can be a loop back here to data prep.

Evaluation: - The analyst builds & chooses models that appear to have high quality and performance. The analyst them tests them to ensure that they can generalise the models against unseen data. Subsequently, the analyst also validates that the models sufficiently cover all key business issues. The end result is the selection of the champion model(s).

Once the project roadmap has been evaluated based on planed vs actual, we found that data preparation took around one month extra from planned timeline. This is because, there are various fields which need to be pre-processed before going to modelling stage. The pre-processing steps are explained under data preparation heading in this document.

# Chapter 6: Business Understanding

ABC is a fashion category multinational company with a turnover of over one billion dollars. It has 40 factories across the globe.

ABC is a profitable company which makes money using their E-commerce website and few other marketplaces like Amazon and Flipkart.

They have also tie up with few brands for which they are manufacturing the products.

Their clients include - Victoria’s Secrets, Marks & Spencer, Triumph, Speedo, Calvin Klein among others.

The first and foremost challenge for any new entrant in the market is attaining customer’s trust, especially when you are in the e-commerce sphere. With so many competitors already present and more and more entrants coming into the market almost everyday, gaining visibility is increasingly becoming an uphill task. Identifying a niche and working actively towards it helps largely. It’s easier and more practical to start from a niche and expand than the other way around.

The second thing that is a major challenge is gaining a potential customer’s trust and establishing credibility. When the customer can’t touch or feel your product, it’s rather difficult to convince them of its quality, which in our case, is the USP.

Logistics and payment gateways charge high commissions which ultimately influences the price of the product. The logistics of Cash on Delivery as a service is another problem that is often faced by e-commerce setups. Last but not the least, as per our law, the customer can reject or deny an order just in case he changes his mind. These increase return rates and overall costs and are most common for COD orders.

Being not so familiar with the complex arena of the ever-growing digital space, we try to have a brainstorming session daily to plan and execute creative initiatives and new capacities. Building a digital strategy to adjust marketing focus from brand and services/products to the end customer is something that we always try to inculcate in our system.

Implementation of well-established payment platform and consumer authentication was another big hurdle we experienced. Digitisation comes with fraud and cyber risks which can easily target companies to be a victim of malware. Be it the data theft or malicious customer targeting brands, we ensure extra attention to try and be safe from these malicious activities. They are a threat not only in terms of time consumption and financial losses, but also in terms of market credibility.

Getting customer support-oriented resources is one of the most challenging parts for any start-up. Reassuring customers of a situation to keep them from hitting the panic button and coming up with negative feedback on social media is something that start-up have to aim for. Customers come first and everything we do has to be as per their needs and convenience. There will be difficult situations that one may face with a few customers but taking them as lessons is the best way forward. [10]

In this project, we will be focusing on the customer churn and retain analysis. For past few months they have found that customers who have bought products in past few months are not coming back.

Here we are trying to find out the customers who are not coming for sometime with customer repeat purchase visit modeling which will gradually focus on regaining them by giving some recommendation (future work).

# Chapter 7: Data Understanding

The dataset contains the sales data from Mar 2018 to Feb 2020.

For Repeat Purchase Modeling, we have taken 1year data i.e. Jan-Dec\_2019

Data contains 35 features, which are - Order No, External Order No, Order Date, Order Type, Status, Customer Name, Country, State, City, Email, SKU Code, Style, SKU Desc, Category1, Sub Category, Size, Color, Type, Quantity, Return Qty, Order Currency, Price, Ship Cost, Packing Cost, Discount, Discount Code, Tax, Invoiced, COGS, Invoiced In Base Currency, Gross Margin, GM Percent, Primary Vendor, On Hold Status, Replacement Order.

Data Dictionary below:-

* Order No: - Auto generated Order number in order management system
* External Order No: - Shopify order number sent to the customer
* Order Date: - Exact time of the order
* Order Type: - Pre paid & Cash of Delivery
* Status: - Shipped complete - order shipped to the customer | Cancelled - Cancelled ordered before shipping
* Customer Name: - name of the customer
* Country: - Orders are shipped only in India
* State: - Order placing state
* City: - Order placing city
* Email: - Email id which has placed the order
* SKU Code: - The SKU code (product code) ordered by the customer
* Style: - Style/Colour & size details of the SKU
* SKU Desc: - Style/Colour & size details of the SKU
* Category1: - Main category
* Sub Category: - Silhouette
* Size: - Product Size
* Color : - colour name given by the designer
* Type: - Basic colour
* Quantity: - Quantity ordered per SKU
* Return Qty: - Quantity returned
* Order Currency: - in INR
* Price: - price after product discounts & before cart level discounts
* Ship Cost: - Cost of shipping the product
* Packing Cost: - cost of packaging the product
* Discount: - Discount amount cart level
* Discount Code: - Discount code applied at cart level
* Tax: - tax amount
* Invoiced: - invoiced to the customer
* COGS : - cost of goods sold - approx.
* Invoiced In Base Currency: - invoiced to the customer
* Gross Margin: - Profit made on invoiced amount
* GM Percent: - percentage of profit made on invoiced amount
* Primary Vendor: - Primary vendor who is selling the product
* On Hold Status: - Product’s status
* Replacement Order: -Orders which are getting replaced

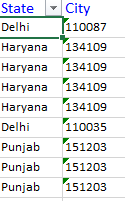
Few important features in the data are Order number, Order Date, City, SKU Code, Size, Colour, Quantity, Invoiced, Gross Margin. These features will be used while modelling repeat visit for the business.

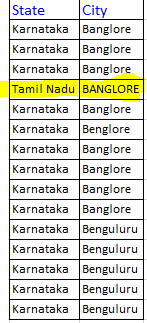
# Chapter 8: Data Preparation

For any business, data preparation is the most important step as it improves the data quality and increases the overall productivity

Below are the few techniques which has been used for the data cleaning:

* Creating Unique Order No.: - Dataset contain multiple Order number for same product purchase (as it’s taking one row for every unit purchase) Solution to this is creating a new order number by matching Order No and SKU Code
* Removing Outlier: - Based on Invoiced amount removed the outlier (these are retailer. Will treat them separately)
* Cleaning State column
* Huge amount of time has been spent on cleaning the city. Prominent issues with the city columns are: -
  + Similar names for same city. E.g. Bangalore, Bengaluru
  + Whole address or just the pin code in city column
  + Some cities like Bangalore is tagged to Tamil Nadu, Delhi etc.



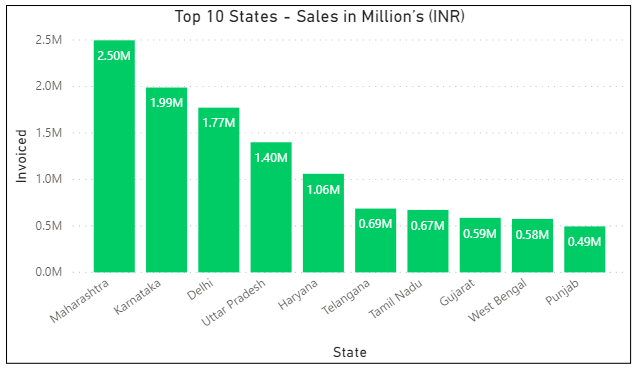


* **Invoiced feature has some negative values**: - Upon check with business, we found that these are few dissatisfied customers who has been rewarded with free gift items.

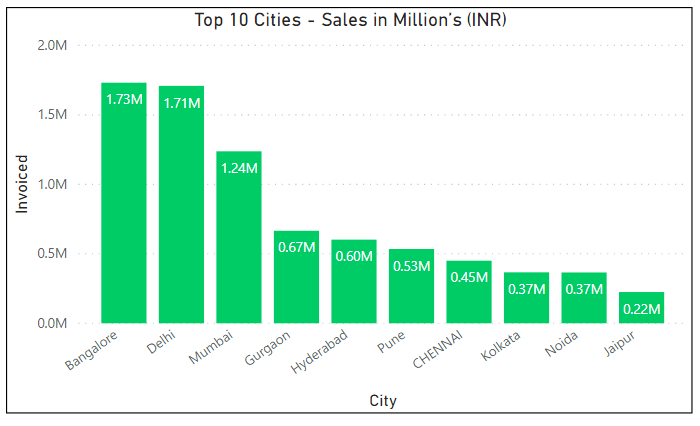
Removing negative value from invoiced has been suggested.

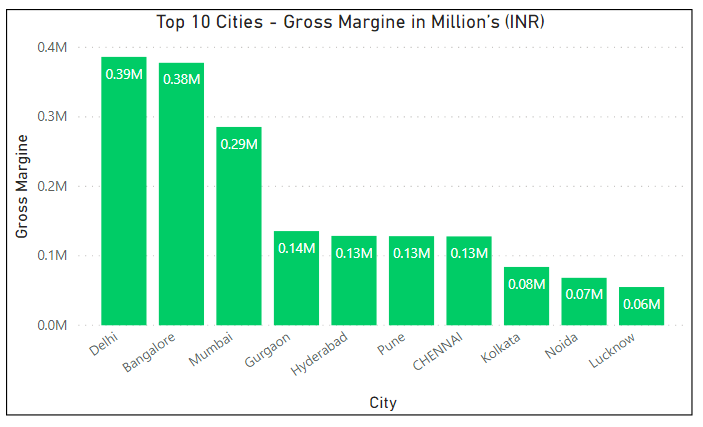


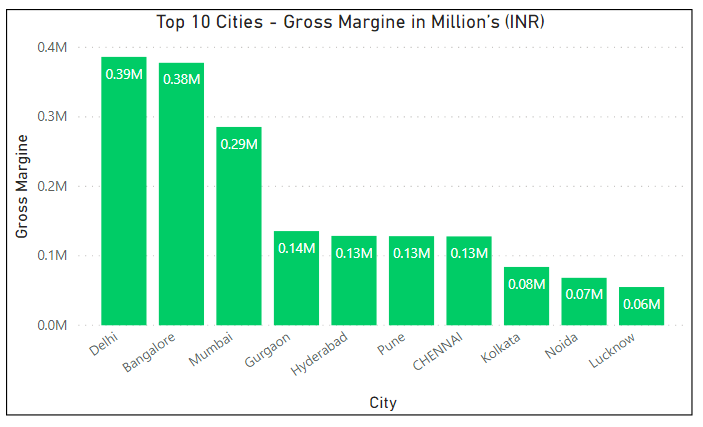
* Exploratory Data Analysis: - Finding out the correlation between the features using Heatmap

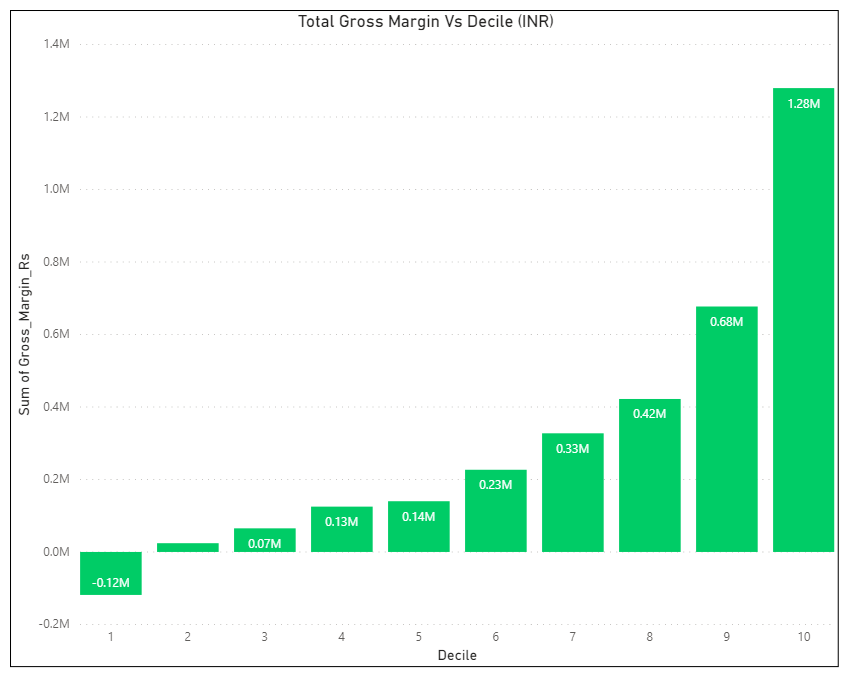


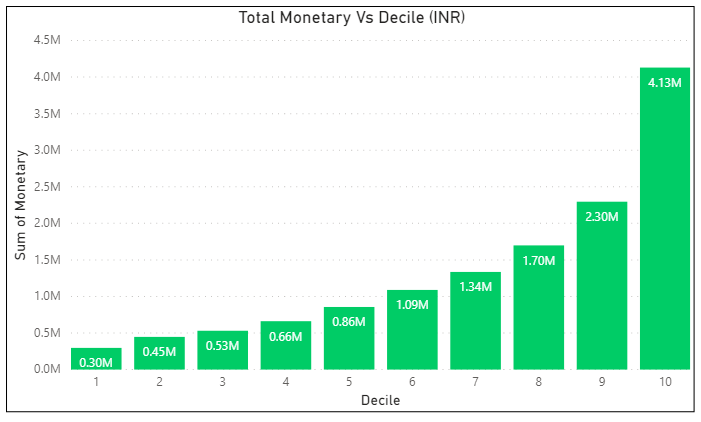












# Chapter 8: Data Modeling

In order to establish a more effective and accurate segmentation model, three data modeling techniques have been considered. The aim of data modelling is to Segment customer can be described as a set of customers who have similar characteristics of demography, behaviours, values. Segmentation technique is used to identify a set of groups that both minimize within-group variation and maximize between-group variation according to a distance or dissimilarity function. It is a process of discovering various models, summaries, and derived values from a given collection of data. The following describes several major techniques which can achieve the goals of this research, including RFM, K-Means and Hierarchical Clustering.

**Segmentation/ Cluster Analysis Technique** (Post-hoc approach)**:**

* **RFM Analysis:** [RFM](https://clevertap.com/rfm/) stands for Recency, Frequency, and Monetary value, each corresponding to some key customer trait. These RFM metrics are important indicators of a customer’s behaviour because frequency and monetary value affects a [customer’s lifetime value](https://clevertap.com/blog/customer-lifetime-value/), and recency affects retention, a measure of engagement.
* **Hierarchical algorithm:** In [data mining](https://en.wikipedia.org/wiki/Data_mining) and [statistics](https://en.wikipedia.org/wiki/Statistics), hierarchical clustering (also called hierarchical cluster analysis or HCA) is a method of [cluster analysis](https://en.wikipedia.org/wiki/Cluster_analysis) which seeks to build a [hierarchy](https://en.wikipedia.org/wiki/Hierarchy) of clusters. Strategies for hierarchical clustering generally fall into two types.
* **Divisive approach** (**Top-down):** All observations start in one cluster, and splits are performed recursively as one moves down the hierarchy. All data points are treated as one big cluster and the clustering process involves dividing the one big cluster into several small.
* **Agglomerative approach** **(Bottom-up):** Each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. All data points are clustered using a bottom-up approach starting with individual data points.
* **Steps to create the hierarchical clustering using Agglomerative:**
* At the start, treat each data point as one cluster. Therefore, the number of clusters at the start will be K, while K is an integer representing the number of data points.
* Form a cluster by joining the two closest data points resulting in K-1 clusters.
* Form more clusters by joining the two closest clusters resulting in K-2 clusters.
* Repeat the above three steps until one big cluster is formed.
* Once a single cluster is formed, [dendrograms](https://en.wikipedia.org/wiki/Dendrogram) are used to divide into multiple clusters depending upon the problem.
* **Centroid-based clustering(K-means)**
* It’s the task to create the group with the same group data share the same feature and it will be a member of the same cluster.
  + Analyses and find patterns/clusters within data
  + Measures distance between inter/ intra clusters
  + Clusters data by trying to separate samples in n groups of equal variances
  + By shifting the mean in inter-cluster, we can minimize a criterion known as the inertia or within-cluster sum-of-squares error.
  + We need to find the required number of clusters to be specified. We are using a very popular technique called elbow method or silhouette analysis
* **Steps to create the K-Meansclustering:**
* Divides a set of samples into disjoint clusters
* Each described by the mean of the samples in the cluster.
* The means are commonly called the cluster “centroids”
* Note that the centroids are not, in general, points from, although they live in the same space.
* The K-means algorithm aims to choose centroids that minimize the inertia, or within-cluster sum of squared criterion
* **The disadvantage of K-means clustering:**
* The globally optimal result may not be achieved
* The number of clusters must be selected beforehand either using elbow method or silhouette analysis
* k-means is limited to linear cluster boundaries
* k-means is the distanced based so it can be slow for large numbers of samples
* **Elbow Method:**
* Elbow method used to determine the optimal cluster in K-means
* Use intrinsic metrics
* As we know if K will increase then average distortion will decrease
* An example for this is the within-cluster Sums of Squared Error, it’s also called as inertia.

# Chapter 9: Data Evaluation

Three machine learning algorithms which are, RFM Analysis, Hierarchical and K-means clustering are built for Customer segmentation model.

# Chapter 10: Deployment

In this project, we have built up a customer segmentation model. The model has been built and tested across multiple algorithms like RFM Analysis, Hierarchical and K-means to find most accurate solution. For the time being, we haven’t deployed the model. ­­

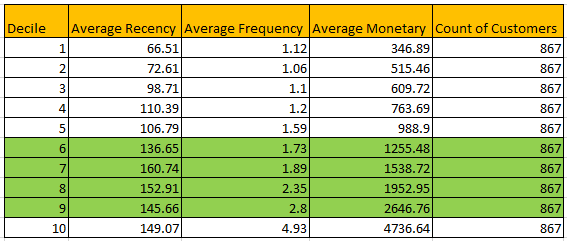
# Chapter 11: Analysis and Results

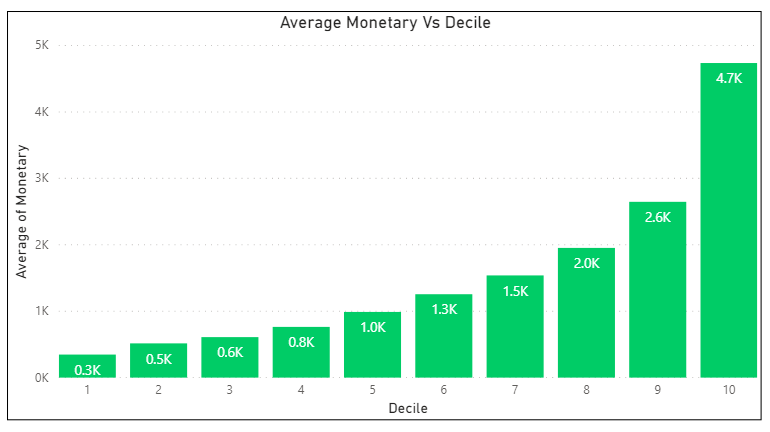
The result of project and analysis are presented in this section. This section includes the outcome of the methodology; i.e., the numerical or descriptive work that was performed.

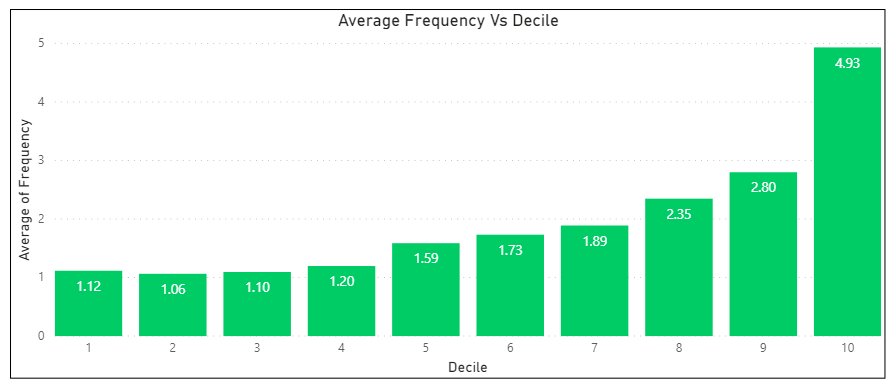
Based on the model and data comparison we found that K-means clustering is able to perform better for customer segmentation.

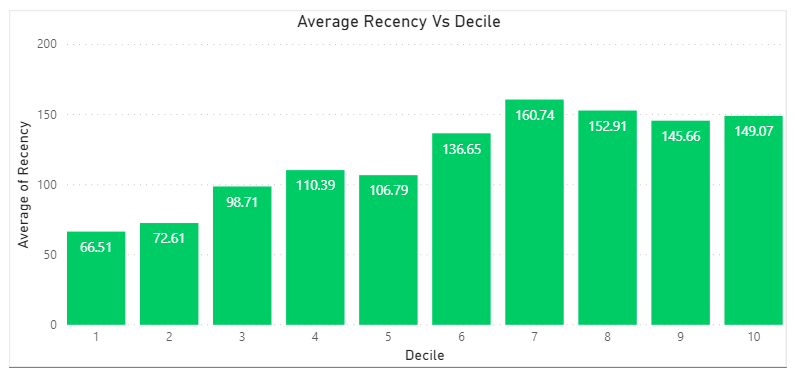
**Decile Analysis:** 30% of the Sales revenue is from 20% of Customers

* Top decile is 13 times better than worst decile and Twice better than the second-best decile
* 10% of sub base will give 30% revenue
* 3 segments as suggested by customer base analysis
* Top decile is not recent!
* Recent acquisitions/visits are low value transactions



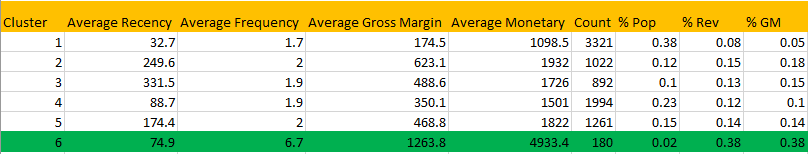


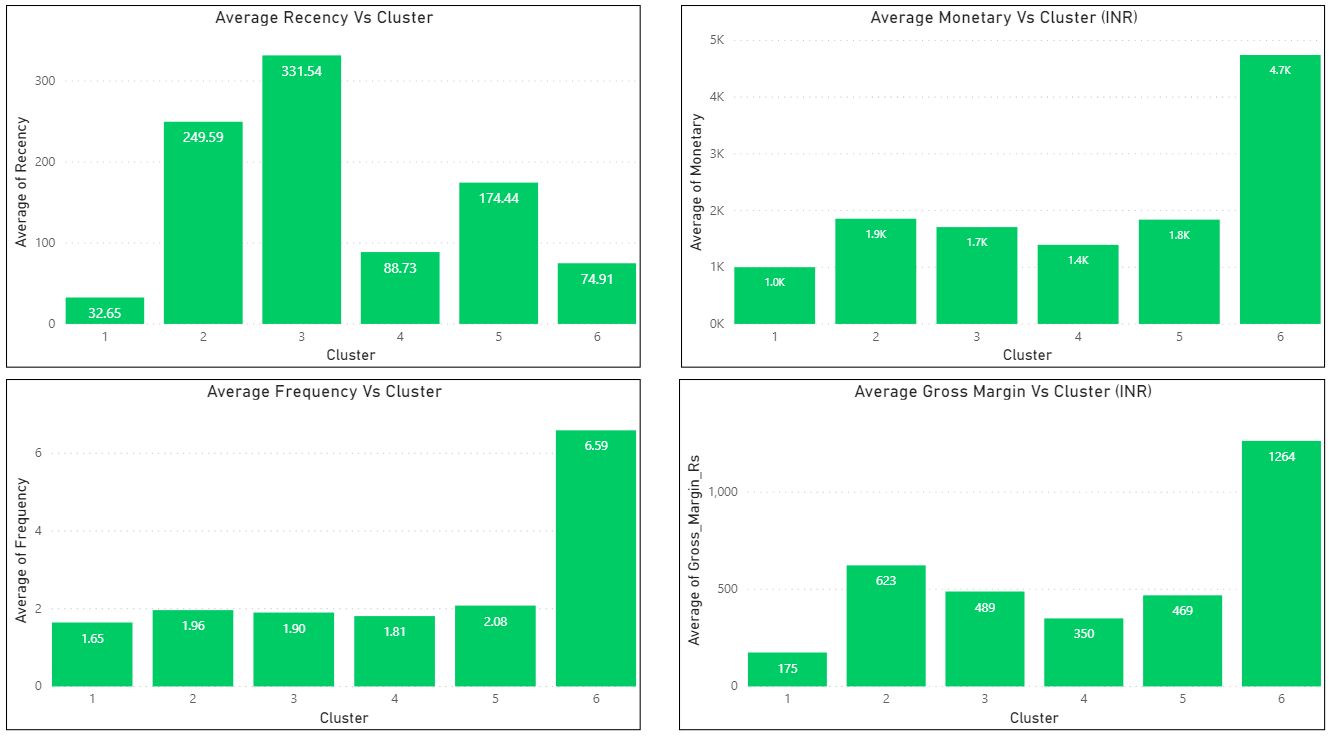




**Customer Clusters and Profiles:**

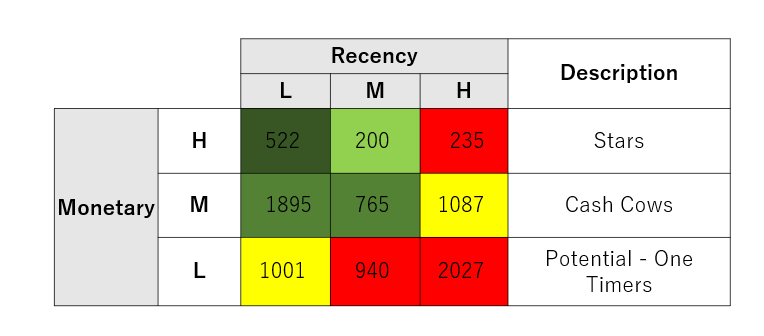
* Average Recency 123 days
* Total No. of Customers 8670
* Average Frequency 1.97 Times/Year
* No. of Products Bought (Avg) 02
* Cluster 6 is the Most Profitable Cluster But, not recent





**Customer Base Analysis**

* Recency – Indicates Behavior
* Monetary – Customer Value in INR



# Chapter 12: Conclusions and Recommendations for future work

* Customer growth is not proportional to revenue growth Need to Focus on Customer Acquisition
* Top 10% give 30% revenue – retain
* 3 segments possible based on revenue
* Low Recency of high value segments: Need early intervention
* Upsell Potential in Medium value segments
* Targeted campaigns and offers

# Bibliography

Akay, M. F., Abasıkeles, I., & Oral, M. (2010). Application of self organizing maps for investigating network latency on a broadcast-based distributed shared memory multiprocessor. Expert Systems with Applications, 37, 2937–2942. Barnes, N. G., & Mowatt, D. (1986). An examination of patient attitudes and their implications for dental service marketing. Journal of Health Caring Marketing, 6(3), 60–63. Capon, N. (1982). Institutional change in delivery of dental service: A marketing perspective. American Journal of Public Health, 72, 679–683. Chang, E. C., Huang, H. C., & Wu, H. H. (2010). Using K-means method and spectral clustering technique in an outfitter’s value analysis. Quality & Quantity, 44(4), 807–815. Chang, H. H., & Tsay, S. F. (2004). Integrating of SOM and K-man in data mining clustering: An empirical study of CRM and profitability evaluation. Journal of Information Management, 11(4), 161–203. Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications, 36, 4176–4184. Chiu, C. Y., Chen, Y. F., Kuo, I. T., & Ku, H. C. (2009). An intelligent market segmentation system using K-means and particle swarm optimization. Expert Systems with Applications, 36, 4558–4565. Chow, S., & Holden, R. (1997). Toward an understanding of loyalty: The moderating role of trust. Journal of Management issues, 9(3), 275–298. Churilov, L., Bagirov, A., Schwartz, D., Smith, K., & Michael, D. (2005). Data mining with combined use of optimization techniques and self-organizing maps for improving risk grouping rules: Application to prostate cancer patients. Journal of Management Information System, 21(4), 85–100. Clow, K. E., Stevens, R. E., McConkey, C. W., & Loudon, D. L. (2007). Attitudes of dentists and dental patients toward advertising. Health Marketing Quarterly, 24(1/2), 23–34. Fish, K. E., & Ruby, P. (2009). An artificial intelligence foreign market screening method for small businesses. International Journal of Entrepreneurship, 13, 65–81. Hanafizadeh, P., & Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and fuzzy Delphi method – ADSL market of a telecommunication company. Expert Systems with Applications, 38, 198–205. Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Systems with Applications, 37, 5259–5264. Hsu, S. H., Hsieh, P. A., Chih, T. C., & Hsu, K. C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. Expert Systems with Applications, 36, 7947–7951. Huang, S., Chang, E. C., & Wu, H. H. (2009). A case study of applying data mining techniques in an outfitter’s customer value analysis. Expert Systems with Applications, 36, 5909–5915. Hughes, A. M. (1996). Boosting response with RFM. Marketing Tools, 5, 4–10. Hung, C., & Tsai, C. F. (2008). Market segmentation based on hierarchical selforganizing map for markets of multimedia on demand. Expert Systems with Applications, 34, 780–787. Kahan, R. (1998). Using database marketing techniques to enhance your one-to-one marketing initiatives. Journal of Consumer Marketing, 15(5), 491–493. Kiang, M. Y., & Fisher, D. M. (2008). Selecting the right MBA schools – An application of self-organizing map networks. Expert Systems with Applications, 35, 946–955. Lee, W. I., & Shih, B. Y. (2009). Application of neural networks to recognize profitable customers for dental services marketing – a case of dental clinics in Taiwan. Expert Systems with Applications, 36, 199–208. Lee, W. I., Shih, B. Y., & Chung, Y. S. (2008). The exploration of consumers’ behavior in choosing hospital by the application of neural network. Expert Systems with Applications, 34, 806–816. Marcus, C. (1998). A practical yet meaningful approach to customer segmentation. Journal of Consumer Marketing, 15(5), 494–504. Reinartz, W. J., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. Journal of Marketing, 64(4), 17–35. Rutsohn, P., & Ibrahim, N. A. (2000). An analysis of the demand for regular dental health care: Implications for marketing. Health Marketing Quarterly, 17(3), 26–37. Shieh, J. I., Wu, H. H., & Huang, K. K. (2010). A DEMATEL method in identifying key success factors of hospital service quality. Knowledge-Based Systems, 23(3), 277–282. Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology. Expert Systems with Applications, 27, 265–276. Wang, S. (2001). Cluster analysis using a validated self-organizing method: Cases of problem identification. International Journal of Intelligent Systems in Accounting, Finance and Management, 10(2), 127–138. Wang, Z., Janda, S., & Rao, C. P. (1996). Dental services marketing: Do marketing segments based on usage rate differ in terms of determinant attributes? The Journal of Services Marketing, 10(4), 41–55. Yeh, I. C., Yang, K. J., & Ting, T. M. (2008). Knowledge discovery on RFM model using Bernoulli sequence. Expert Systems with Applications, 36, 5866–5871. Zhang, K., Chai, Y., & Yang, S. X. (2010). Self-organizing feature map for cluster analysis in multi-disease diagnosis. Expert Systems with Applications, 37, 6359–6367.

# Appendix

## Plagiarism Report[[1]](#footnote-1)

## Publications in a Journal/Conference Presented/White Paper[[2]](#footnote-2)

## Any Additional Details

1. Turnitn report to be attached from the University. [↑](#footnote-ref-1)
2. URL of the white paper/Paper published in a Journal/Paper presented in a Conference/Certificates to be provided. [↑](#footnote-ref-2)