

BUSINESS DATA MANAGEMENT PROJECT

Penetrating B2C Market and Optimizing Operations For Cloth Merchant

FINAL SUBMISSION

Submitted by

Name: Tarunpreet Singh

Roll number: 23F1000113



IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

SPECIAL THANKS

Dr. Aaditya Chandel

Postdoctoral Research Associate

IIT Madras' Department of Aerospace Engineering

Dr. Ashwin J Baliga

Assistant Professor of Marketing & Sales

IESEG School of Management, Paris, France



Sahib Embroidery,
Manufacturer of Ladies Suits, Ludhiana
Punjab, India, 141007

“You can have data without information, but you cannot have information without data.”
— Daniel Keys Moran

Table of Contents

I. Executive Summary	III
II. Detailed Explanation of Analysis Process/Method	IV
II.1 Objectives.....	V
II.2 Research Analysis	VI
II.3 Data Collection.....	VII
II.4 Data Cleaning and Preprocessing.....	VIII
III. Results and Findings	IX
III.1 Descriptive statistics	IX
III.2 Time taken at each development stage by all the different SKUs	X
III.3 Comparison of Time taken with expected Time	XI
III.4 Differentiating Stages/Processes.....	XII
III.5 Rejections At a Glance.....	XIII
III.6 Time killed by Rejected Samples.....	XIV
III.7 Customer distribution By City and Distance	XV
III.8 District Wise Analysis	XVI
IV. Interpretation of Results and Recommendation.....	XVII
IV.1 Production Timeline	XVII
IV.2 Sample Manufacturing.....	XVIII
IV.3 B2C Expansion	XIX

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to my Instructors, for their invaluable guidance and support throughout this project. I also extend my thanks to TAs for their encouragement and insights. Special appreciation goes to Sahib Embroidery for providing resources and data. Your contributions have been instrumental in the successful completion of this work.

I. Executive Summary

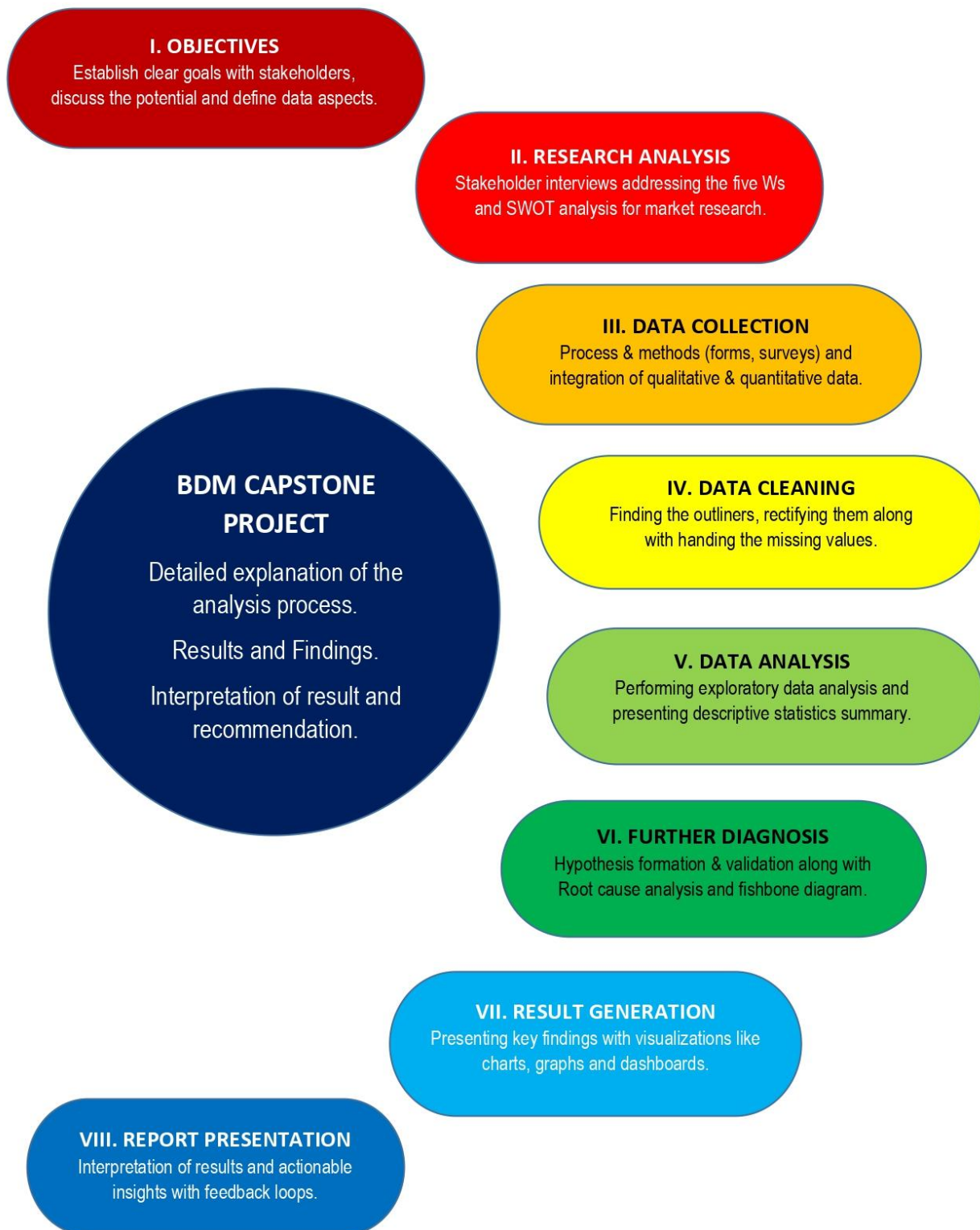
The goal of business data management project is to collaborate with business, collecting, preprocessing and analyzing data to address the objectives. The business venture is **Sahib Embroidery** which a B2B textile firm which manufacture ladies' clothing. The objectives for this study are, to introduce the business into B2C sector without affecting the B2B clients and to optimize the manufacturing process. Multiple visits and thorough observations were carried out to understand the operations and challenges faced by the business.

The analysis process begin with brainstorming and finding the background of the objectives, followed by data collection and data cleaning. The initial descriptive findings were elaborated in the mid term submission and key inferences are also discussed in results. This laid foundation for deeper insights.

The results along with their interpretation and recommendations are presented in the report. The visualizations are presented with each result and its inference is also given along. The visualizations were made using python libraries and excel charts and graphs. The geo map, heat map were done using google maps and other python libraries.

The recommendations were shared with business and were found fruitful by the stakeholders. Both objectives were covered for analysis and successfully executed. The reduction of time in manufacturing process was achieved and the suitable location for B2C retail sector were provided. The key performance indicators were evaluated and found to be optimal. The analysis of data and diagnosis of reasons were done with owner of firm stating true facts adhering to the findings.

II. Detailed Explanation of Analysis Process/Method



II.1 Objectives

Understanding Stakeholder needs: The first step in the analysis process was to define the objectives by understanding the needs of the company officials. A brainstorming session was done with the proprietor of Sahib Embroidery to note the requirements. It came into light that the owner wanted efficient and optimized manufacturing process along with expansion into the retail sector.

Background of the problem: Diving deeper into the conversation, it was found that the company already tried to expand into B2C sales but faced challenges by the existing B2B clients. The owner also highlighted the problem of overstock or understock they face because of unpredicted market nature and delayed manufacturing process.

Key Objectives: In this project, we will examine the manufacturing process thoroughly and try to eliminate the challenges incurred during B2C expansion. The objectives laid for the project are first, to help business penetrate into the retail sector without compromising the existing wholesale business model and second, to analyze the manufacturing process to make it more robust.

Scope and Data Aspects: The scope of the analysis is limited to the operations of the firm. The unpredicted nature of the market for sales will not be covered in this report. There are various aspects for the B2C expansion like costs, infrastructure etc. which will not be covered. The distance and existing clients will be kept in context for evaluation.

Key Performance Indicators (KPIs): The proposed solutions and the extent of study will be evaluated using the performance indicators. The short term KPIs include the number of at risk customers and the long term KPIs cover the overall growth in revenue by sales in retail sector. The manufacturing process will be analyzed by benchmarking existing manufacturing time.

Impact: Successful establishment of B2C component will lead to drastic growth in terms of revenue. The problem of season discrepancy would be decreased by optimizing the manufacturing process thus reducing the overall time taken in demand and supply of goods. Both the objectives hold great potential for further development of the business.

Importance: As per the data from the stakeholders, the net profit margins in B2B is 8 percent whereas for B2C it stands at 25 percent (Image1). Despite the high profit margins, the retail market contributed only two lakhs which stands at 1 percent (Image2) of the total revenue (two crores) generated in FY 2023-2024. Thus, it is crucial for the growth of revenue and profit margins.

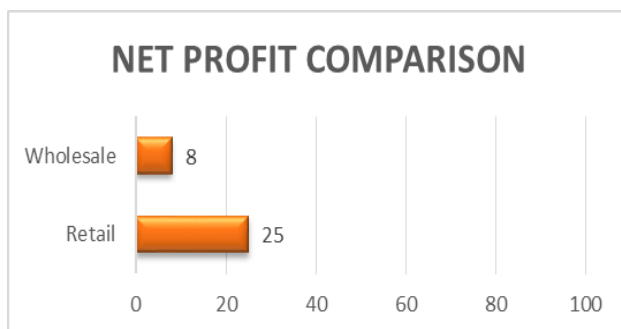


Image1: profit (in percent)

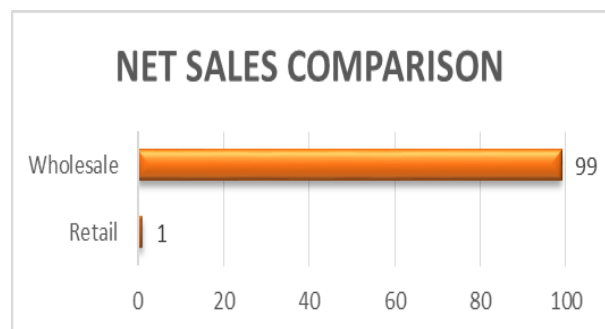


Image2: Sales (in percent)

II.2 Research Analysis

Prior to the data collection, thorough site visits and interviews were done. The 5Ws model was used to dig out the important information about the manufacturing process. In context with B2C expansion, SWOT analysis was used to determine the facts.

What is manufacturing process timeline? The pipeline begins with raw material selection and ordering. The selected raw material is delivered to the firm within 5 days. The unpacking and checking is done to ensure quality and this process takes 1-2 days. This is followed by coloring, drying and finishing which takes another 3 days. The next step is design, cost and threads analysis which is done by professional designers who complete the process in 5 days. Then a sample is produced, inspected and decision making happens taking 12+ hours. The passed samples are move forward with actual machine production taking 3 days followed by a day of inspection and cleaning. Some products need another round of hand embroidery and further cleaning taking 3 and 1 day respectively. The packaging and transportation is done at the last which also takes few hours.

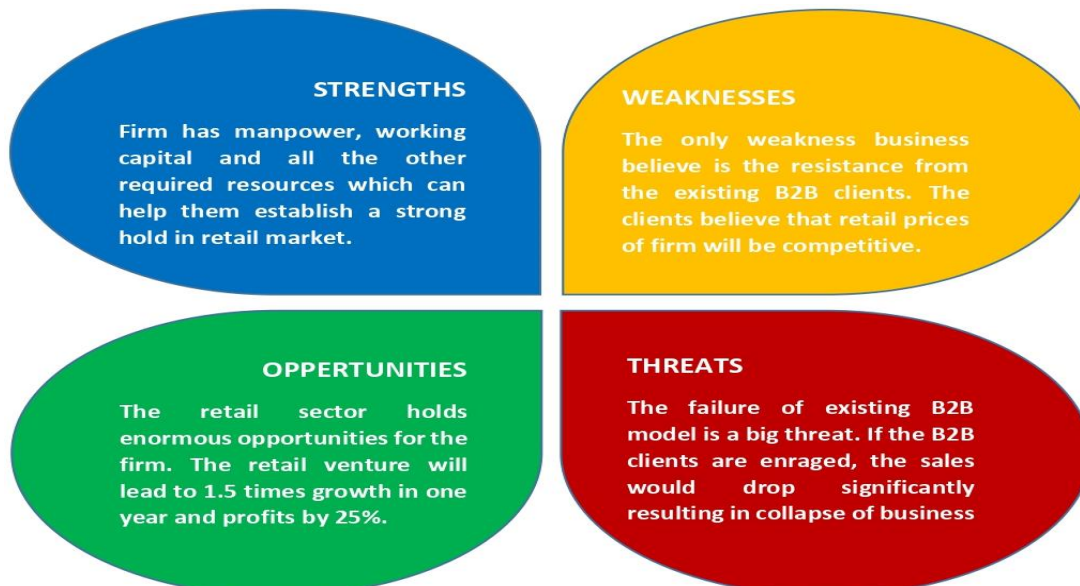
Why is manufacturing data important? The manufacturing process data is important to analyze the time taken at various levels. Time is integral aspect of key performance indicator and will be utilized for in depth analysis.

Where does it makes affect? The production time is directly related to the inventory mismanagement. The time when the product is in demand vs the time when the product is available plays a huge role. The long manufacturing process creates the understock/ overstock situations.

Who is responsible for data generation? The manufacturing team is different for each process. Only the sub teams know the time taken for their processes thus, it was advised to collect data at each process level.

When is data collected during the manufacturing process? The data is present at each manufacturing stage but the company does not keep records. Only the date and quantity of raw material sent for dyeing and final products delivered to shop are checked to ensure consistency.

SWOT Analysis for B2C Expansion



II.3 Data Collection

Data Source: The manufacturing time data was sourced from workers working in the factory while the customer data was sourced from accounts/sales team. Customer data was taken directly from the index of ledger used for accounting by the firm.

Data Collection Period: The customer data has records of all the active clients since the inauguration of business in 2005. The production time data was gathered over a period of three months from 14-07-2024 to 29-09-2024.

Data Collection Frequency: The manufacturing data was collected every week by paying a visit to the manufacturing site. The data was generated every day as various stages take different time for production.

Data Format: The customer data was taken in the form of snapshots in the presence of the owner due to the privacy concerns as same ledger holds all the accounts. The production time data was collected by making a manual data entry form which had columns for design number, all the production stages and final outcome.

Raw Data (Images taken of index of ledger): <https://drive.google.com/drive/folders/1VH8To-MkBnz8d1Nkk73EZmWykkVfjiHJ?usp=sharing>

Data Collection Process: The form created for data entry was taken along at each visit to site. The workers were asked at each stage about the time taken for completion of respective process. All the data collected on weekends was viewed by the owner.

Data Transformation: All the collected data was transformed into excel sheets manually. This process ensured every entry was done carefully into the sheet. The link to the excel sheet is given below:

https://docs.google.com/spreadsheets/d/1j9Od0QdHfmJoo68bSk_IV539wNCepiI9/edit?usp=sharing&ouid=105547014113034604851&rtpof=true&sd=true

Challenges Associated with Data Collection: The main challenge encountered was the data quality and accuracy as there were no formal structured methods for record keeping in manufacturing. Thus, the whole process relied on the workers for data entry. To ensure the accuracy, all the entries were cross verified by the next stage workers as the process ending of one stage is process beginning of another. Hence, the quality and accuracy of data variables was ensured to be intact.

Sampling Strategy: To mitigate the bias in the collection process, random sampling strategy was used. The manufacturing unit has multiple machines for production, each week a different machine and different worker was sourced for data collection.

Tools: Microsoft Excel & Google Spreadsheets for data, Google Drive for storage.

Methods: Interviews, discussions and observations.

II.4 Data Cleaning and Preprocessing

The next big step after data collection is cleaning the data to prepare it for analysis. Data cleaning is crucial process, ensuring accuracy, consistency, and completeness. The following steps were performed to process the data before exploratory data analysis:

Handling Null Values: It was observed that the dataset contained 21 missing values in various rows and columns. Out of those 15 values were related to products that were rejected in the sample manufacturing stage by the firm. Other 6 values were found to be products which had fewer manufacturing stages. All these missing values were substituted with zeroes.

Outlier Detection: To detect the outliers, plotting of each variable was done. The outliers found were not flagged neither removed because they didn't cause much deviations in results. The plotting of outlier available in the excel sheet 3.

Managing Inconsistencies: The customers' data had many inconsistencies in the name column. The entries were taken from index which didn't has consistent name protocols. Some entries were expressed by registered business names like "Raman Traders", some entries were based on clients' first name like "Prabh" or "Rehmaan" while others were represented by shop names like "Tilak di Hatti". These inconsistencies doesn't affect our findings and analysis process.

Standardization: The time taken for some stages which were carried out in single day were given in hours while other processes took days. To standardize the data, all the entries which were in hours were converted to days. This ensured all the data variables had same units.

Validation: There were 30 records for the manufacturing process which were collected. The initial validations by the owner removed 10 percent of the records i.e. 3 records were deleted by the owner. The reason stated by the owner for removal included cross validation failures and mechanical fault in machineries which delayed the process time by few days.

Additional Data Sourcing: The locations were not consistent in terms of district or cities. So a new column "District" was added and the data for the same was based on the location of town or cities. All the records were added with District, State and Distance columns. The approximate distance to each location was calculated in Kilometers using Google maps. These were necessary to ensure that the in depth analysis on locations data was possible.

Tools: Python libraries like pandas and numpy for handling missing values, Google Spreadsheet for outlier detection via plotting graphs and Google Maps for distance calculation.

III. Results and Findings

III.1 Descriptive statistics

VARIABLE	MEAN	MODE	MEDIAN	MIN	MAX	RANGE	STD. DEV
I	5.96	5	6	5	11	6	1.4
II	3.29	3	3	3	5	2	0.7
III	4.22	5	5	3	5	2	0.8
IV	5.48	5	5	5	7	2	0.6
V	1.07	1	1	1	2	1	0.2
VI	4.00	3,4,5	4	0	5	5	1.5
VII	1.00	1	1	0	1	1	0.3
VIII	2.79	3	3	0	4	4	1.3
IX	0.87	1	1	0	1	1	0.4
X	1.03	1	1	0	2	2	0.5
TOTAL DAYS	29.75	30	30	20	35	15	3.4

Table: Descriptive statistics for different production stages

Key statistic features from the production time data:

The mean of all the stages depicts the average time taken for the respective processes. It can be observed that the **stage-1 and IV** are the most time consuming ones.

The values for all the measures of central tendency are identical indicating that all designs took approximately **30 days for completion**.

It is worth mentioning that stage-VI has three modes with 3, 4 and 5 as values.

A **completion rate of 90%** (24 completed out of 27) indicates a strong performance, though the 10% rejection rate suggests areas for improvement.

Key statistic features from the customers data:

Distribution: The customer base is spread among two states, Punjab and Haryana. The client base of Punjab is distributed among 16 districts and 33 cities/towns.

Dominance of Ludhiana: The mode of the data is Ludhiana with maximum number of clients. The Ludhiana city has 14 clients whereas the count of Ludhiana district is 24 which indicated that the clients are regionally distributed around the Ludhiana city. This suggests a strong customer base or significant market activity in this area.

Potential Growth Areas: Places like Ferozpur, Khanna, and Moga have only one or two entries. This might indicate potential for market expansion or under-served areas.

These insights lay the foundation for deeper analysis.

III.2 Time taken at each development stage by all the different SKUs

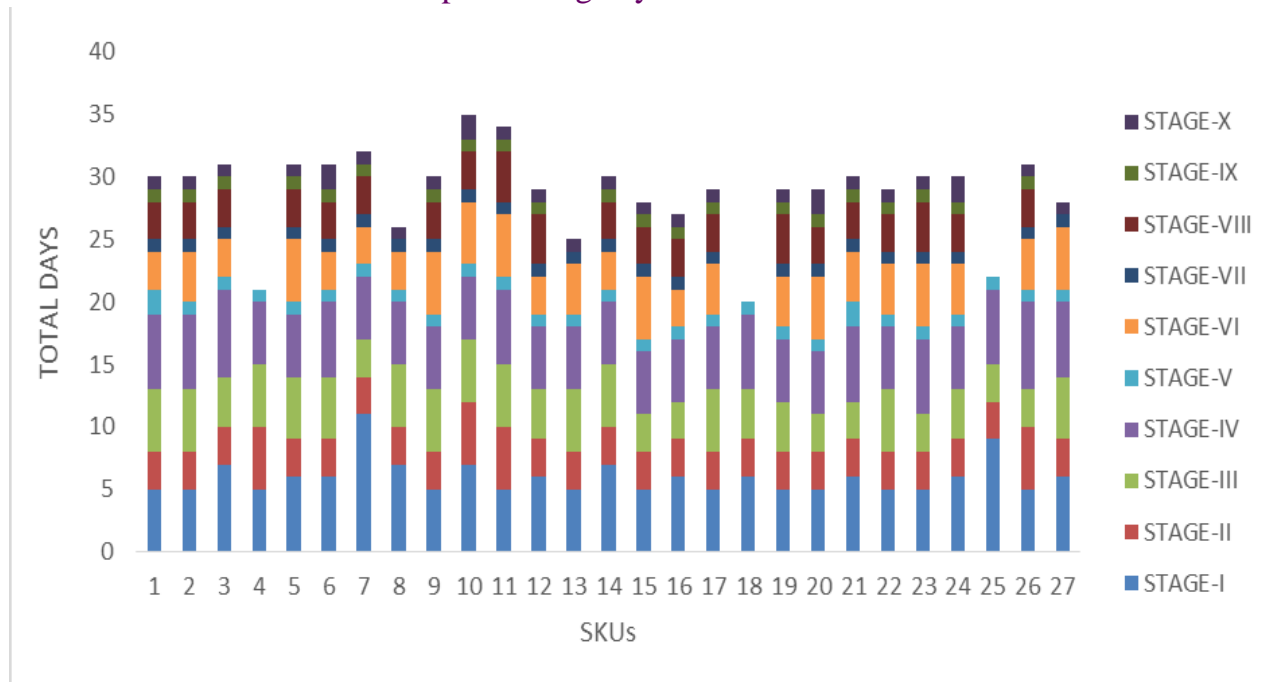


Figure1: Time distribution of all SKUs at various stages of development

Key Findings:

1. Product 4, 18 and 25 end at stage-V showcasing that these are the rejected at the sampling stage.
2. Product 8, 13 and 27 does not have stage –VIII and IX indicating a shorter completion cycle and these stages may not be necessary for all products.
3. Product 10 and 11 took maximum time for production which is around 35 days.
4. Product 13 is the only one with same expected time as actual time.
5. Some pairs had similar manufacturing time indicating their similar design patterns.

III.3 Comparison of Time taken with expected Time

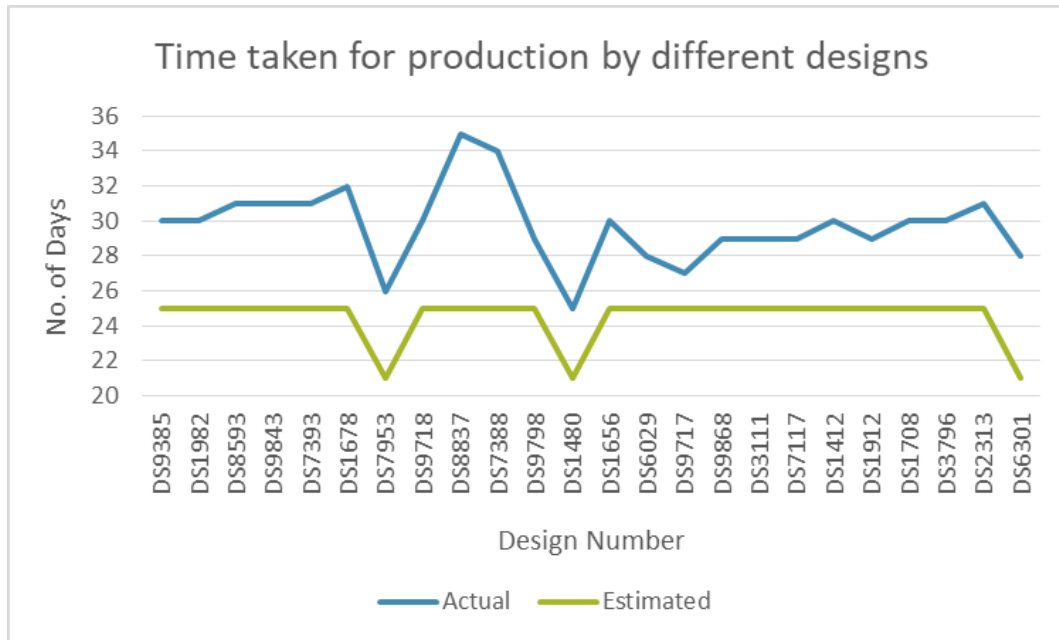


Figure: Designs with actual and estimated production time.

The data shows that there are two types of production timelines. One has estimated time of 21 days while the other had estimated time of 25 days. This was due to the less of processes in their manufacturing. The designs with less no of processes include 7952, 1480 and 6301.

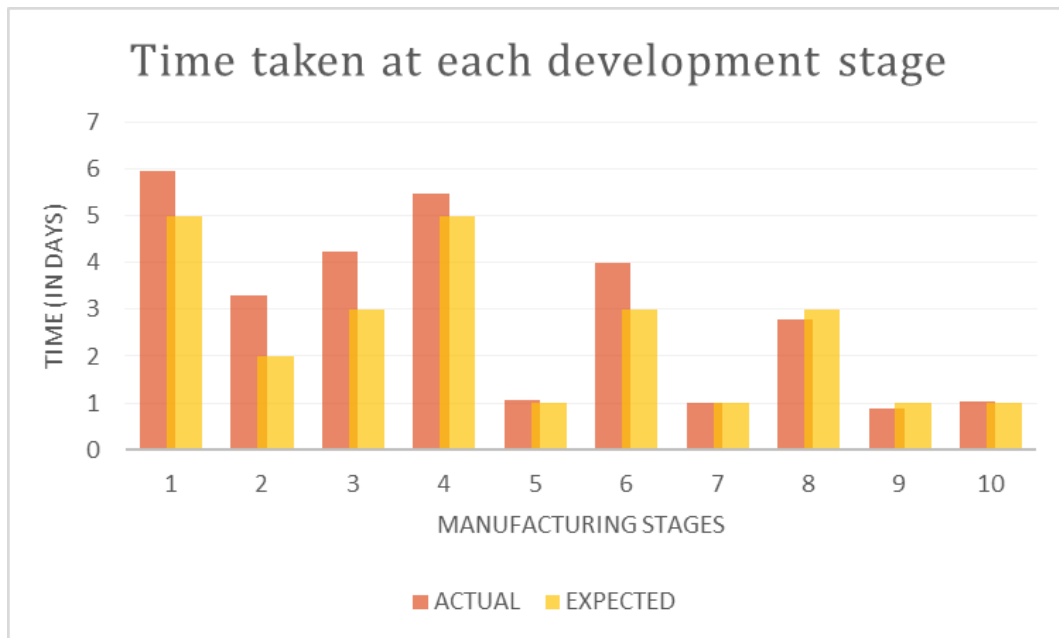


Figure: Each manufacturing stage with actual & expected time.

The stage 1, 2, 3, 4 and 6 take more than expected time due to the following reasons:

Stage1: External factors like late delivery and transportation.

Stage2: Inspection is labor extensive process taking more time.

Stage3: The weather affects the drying time in monsoon. The data collected was over the monsoon period in Punjab.

Stage4: Professional designers take many things into account before finalizing the designs.

Stage6: Higher production orders take higher batch timings.

III.4 Differentiating Stages/Processes

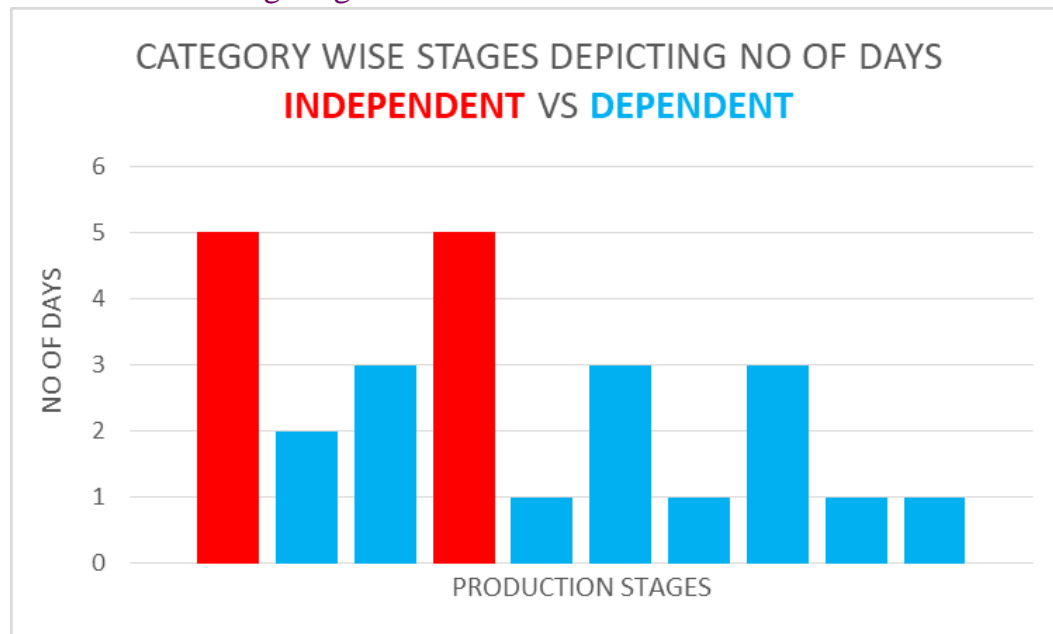


Figure: stages which are independent of previous stages

It was found that 2 stages are independent of outcomes of previous processes. Stage1 and stage4. These include the ordering and selection of raw materials and design, cost and thread analysis which is done by professional designers. This will help redefine the production process.

III.5 Rejections At a Glance

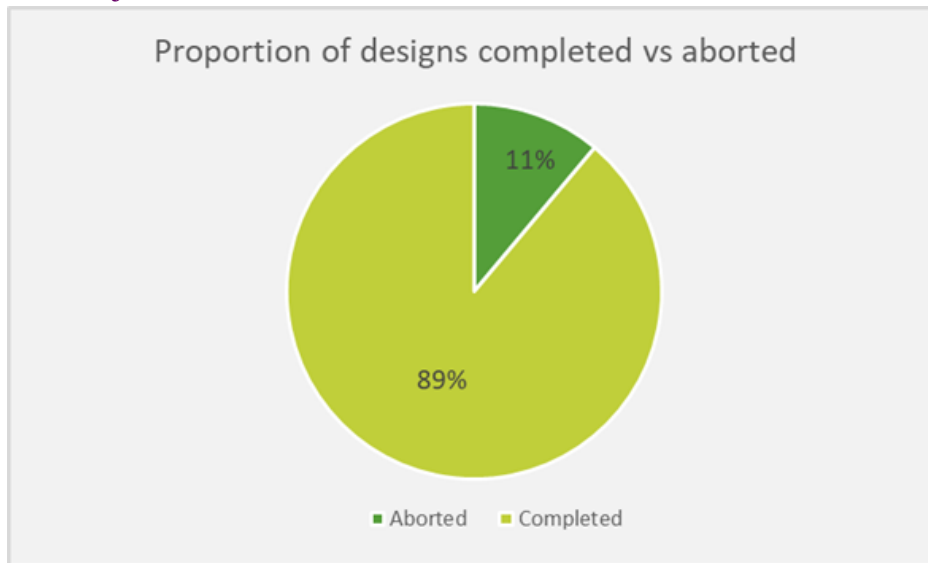


Figure: rejected samples by the inspection team

Out of 27 designs, 3 were rejected by the inspection team. This lead to a big finding that rejected samples make 11% of the actual production. The hypothesis thus formed was confirmed by the owner stating that the actual number stands at 10% which is close to the percentage found during analysis.

III.6 Time killed by Rejected Samples

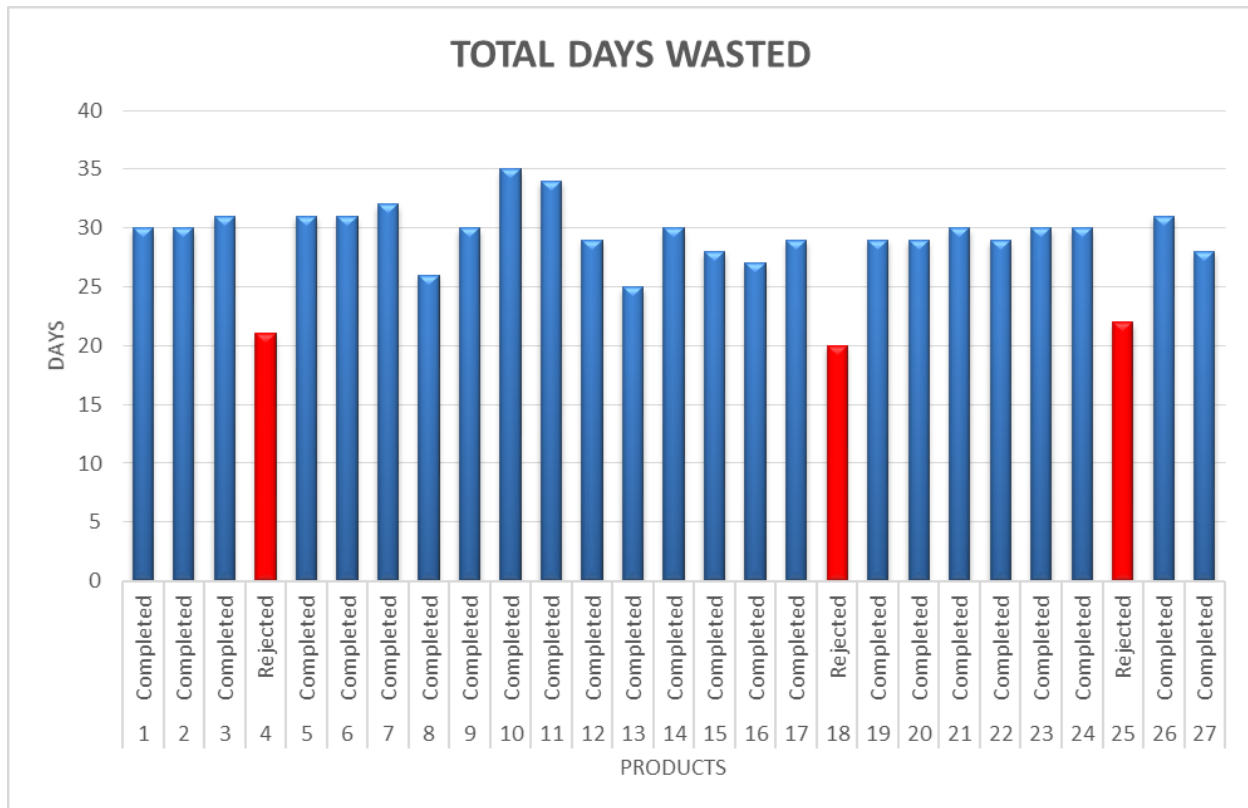


Figure: Time taken to produce the rejected samples along with completed products.

The figure shows that the all samples take 20 days for production. If a sample is rejected it cause loss of resources as well as time to the company. The samples were rejected by the staff due to their quality, designs or finishing levels. There is proper team at the firm which handles this process and judge every product before sending to the production stage.

III.7 Customer distribution By City and Distance

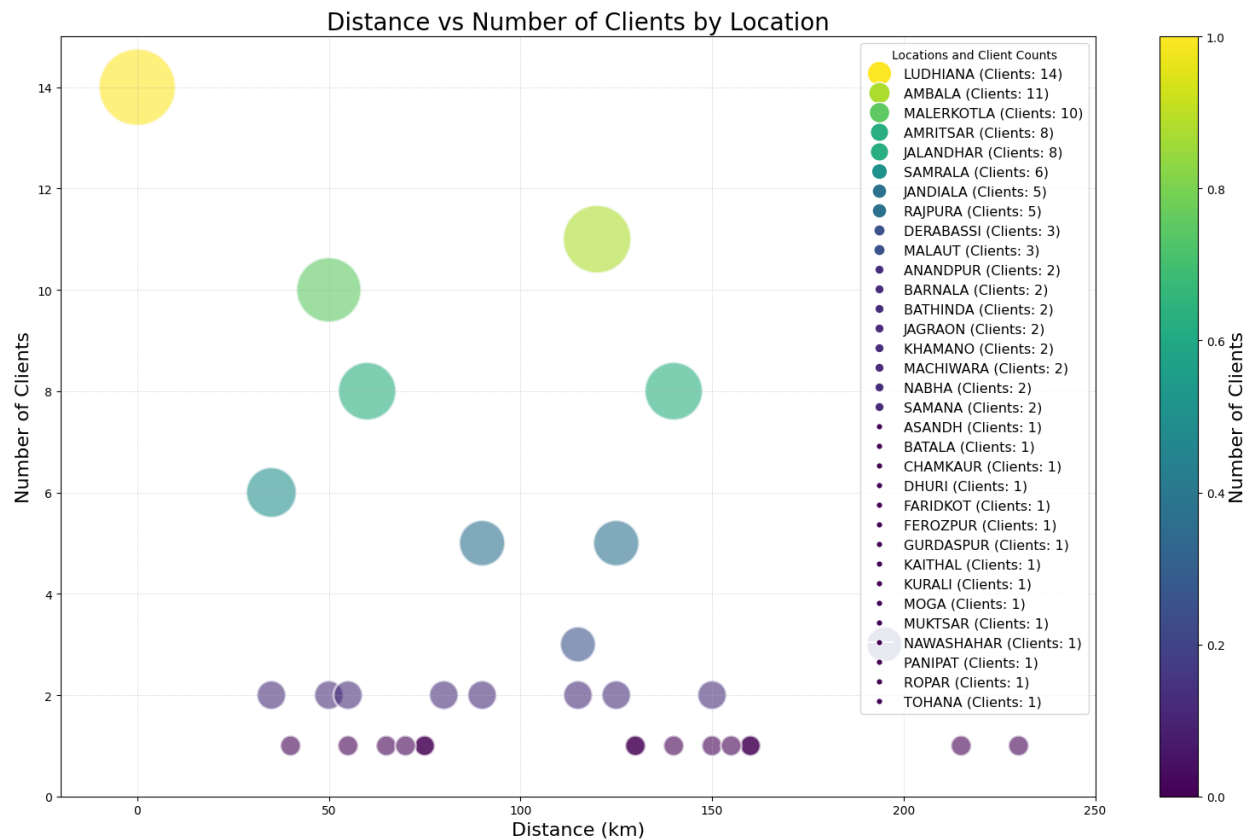


Figure: Scatterplot using distance and customer count for cities.

To find the best region data was plotted on the scatter plot and it was found that the no of clients was very low with some cities. It was found that Ambala is the biggest hotspot from Haryana and its distance is in range of 150 kms. An initiative was done to find and evaluate all the cities that fall under the distance of 100 kms and has less clients. This lead to district wise analysis performed in next section where all the districts were evaluated and categorized based on the count of customers.

III.8 District Wise Analysis

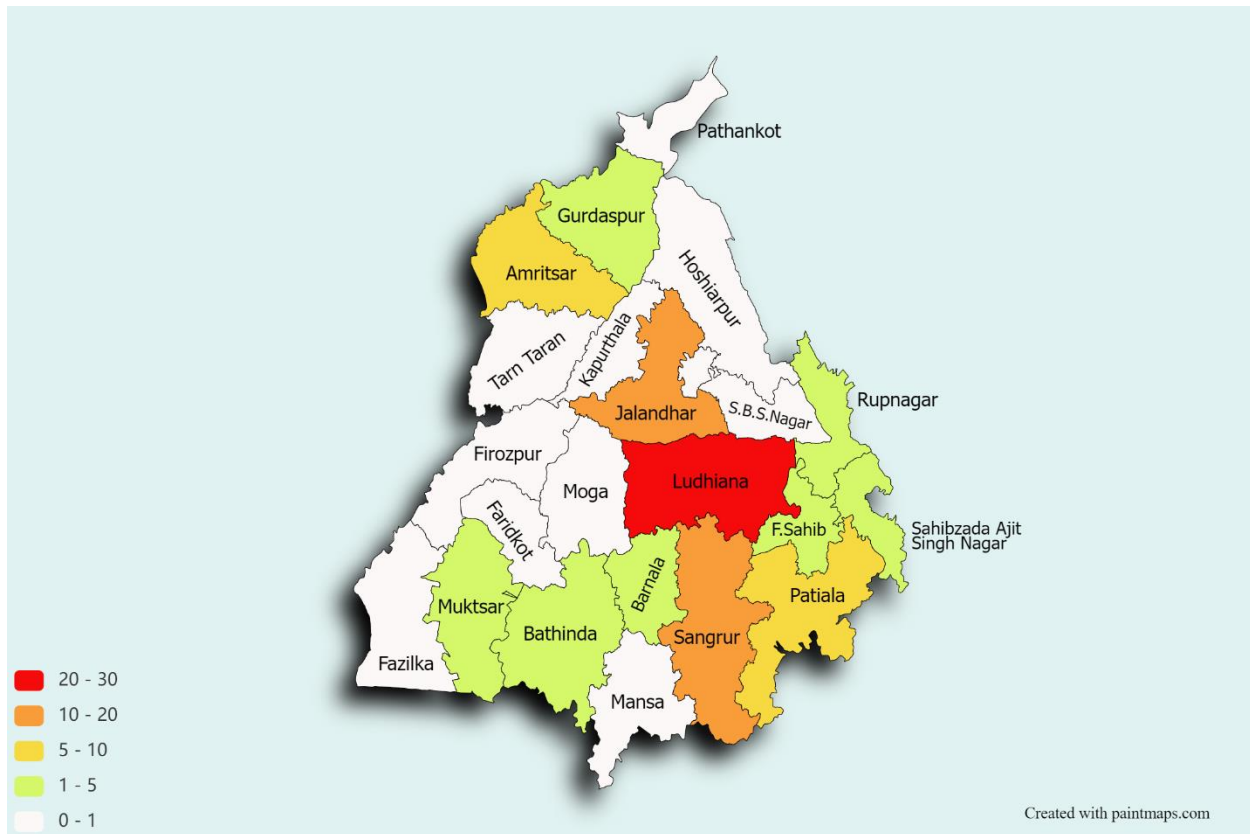


Figure: Map of Punjab showing district wise customers volume

Map along with district wise data reveals a wide disparity in values across different districts. Ludhiana stands out with the highest value of 24, followed by Jalandhar with 13 and Patiala with 9. Several districts, such as Fazilka, Hoshiarpur, Kapurthala, Mansa, Pathankot, and Tarn Taran, have recorded a value of 0, indicating no data or negligible activity. Sangrur also reports a relatively high value of 11, while other districts like Amritsar and Sahibzada Ajit Singh Nagar have moderate values.

IV. Interpretation of Results and Recommendation

IV.1 Production Timeline

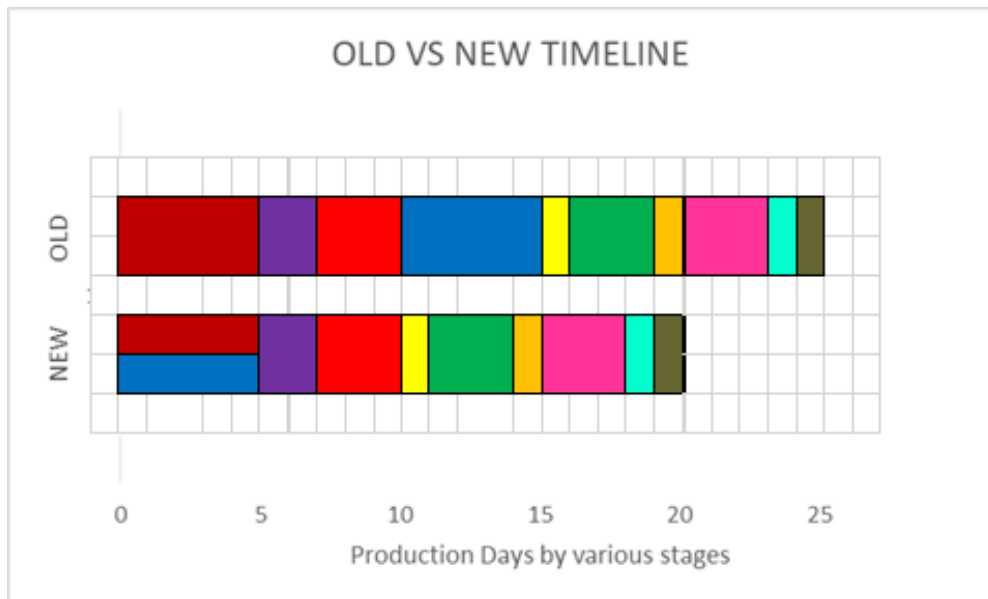


Chart depicting the differences between the new proposed and old production timelines.

Interpretation:

- Both Stage 1 and 4 are most time consuming taking more than 5 days each.
- It was observed that the stage 4 is independent from the results of previous stages.

Recommendation:

- A new production timeline is recommended beginning with stage 1 and 4 running parallel.
- The Results from the stage 4 can be incorporated and utilized when required for stage 5.

How it will impact the business:

- It will decrease the overall expected production time from 25 days to 20 days and 21 days to 16 days for both types of designs.
- Lower production time will help business to generate production faster to meet the demands.

IV.2 Sample Manufacturing

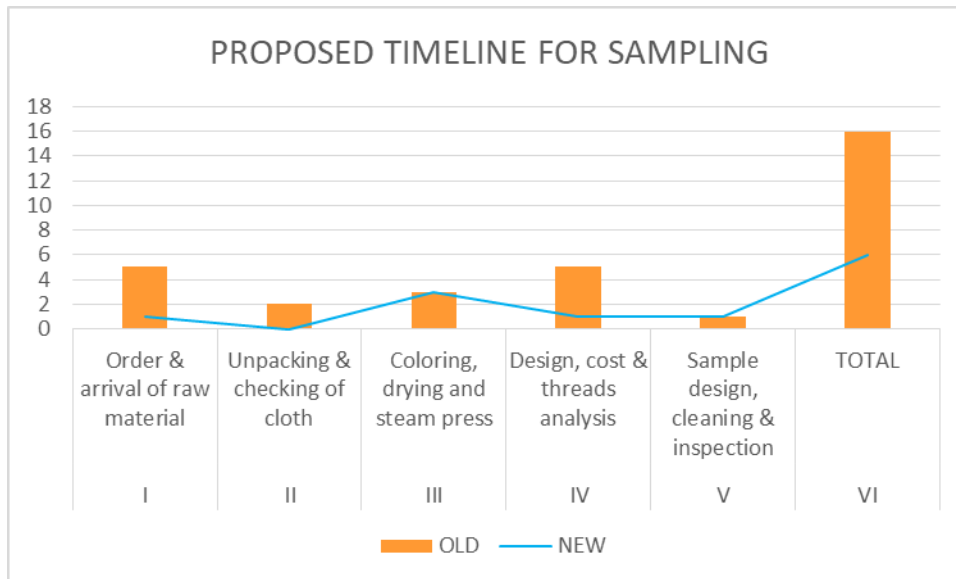


Chart depicting the difference between the new proposed and old sampling timelines.

Interpretation:

- Rejection of products was done at stage 5.
- The rejected products took 20 days minimum.

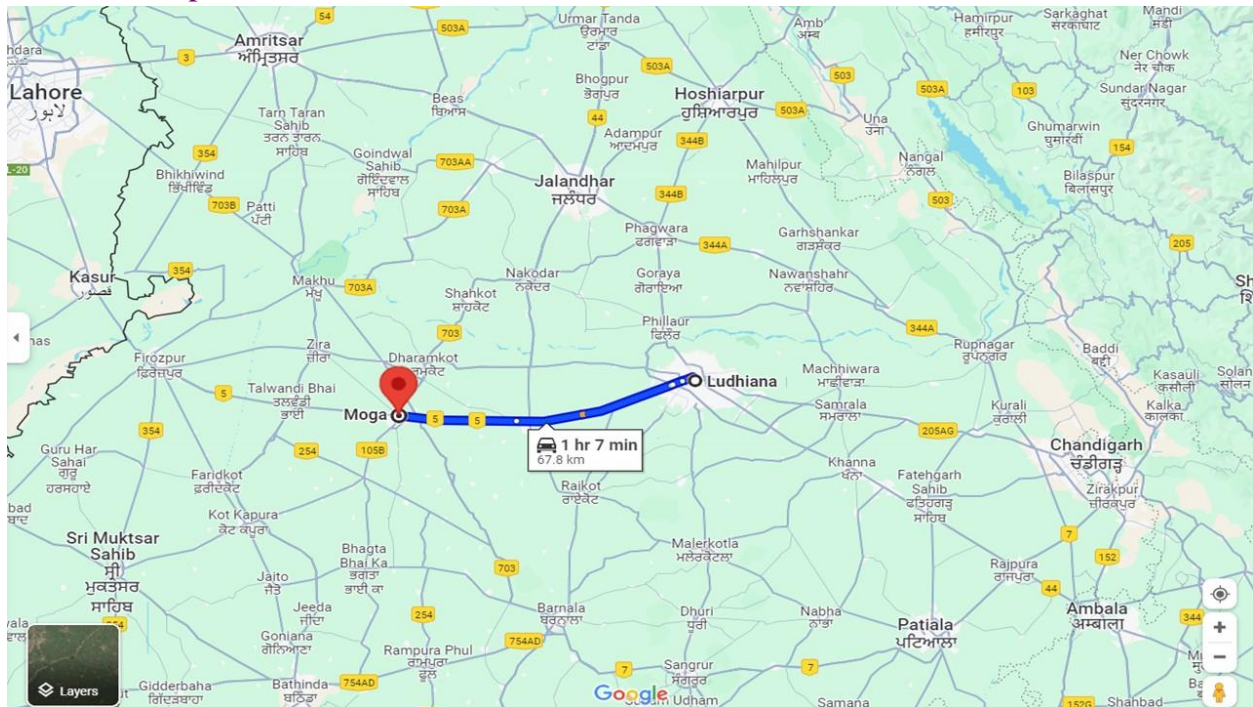
Recommendation:

- A new pipeline is recommended by sourcing raw materials locally for sampling.
- It is combined with recommendation 1 to make a better impact.

How it will impact the business:

- It will decrease the overall expected sample production time from 16 days to 6 days.
- Lower sampling time will help business utilize the time and resources for production of approved samples.

IV.3 B2C Expansion



Map depicting the proposed city for B2C expansion.

Interpretation:

- Moga and SBS Nagar are two neighboring districts of Ludhiana.
- The client base in both the districts was negligible.
- Moga had better scope for development because it adjoins 2 more districts which had no client base.
- The distance from Moga city to Ludhiana can be easily covered in an hour.

Recommendation:

- It is recommended to start retail outlet in Moga City.
- It is also advised to not deal with new B2B clients from Moga district.

How it will impact the business:

- It will help business expand its sales into the retail sector.
- There would not be resistance from existing B2B clients as Moga had only 1 existing client.