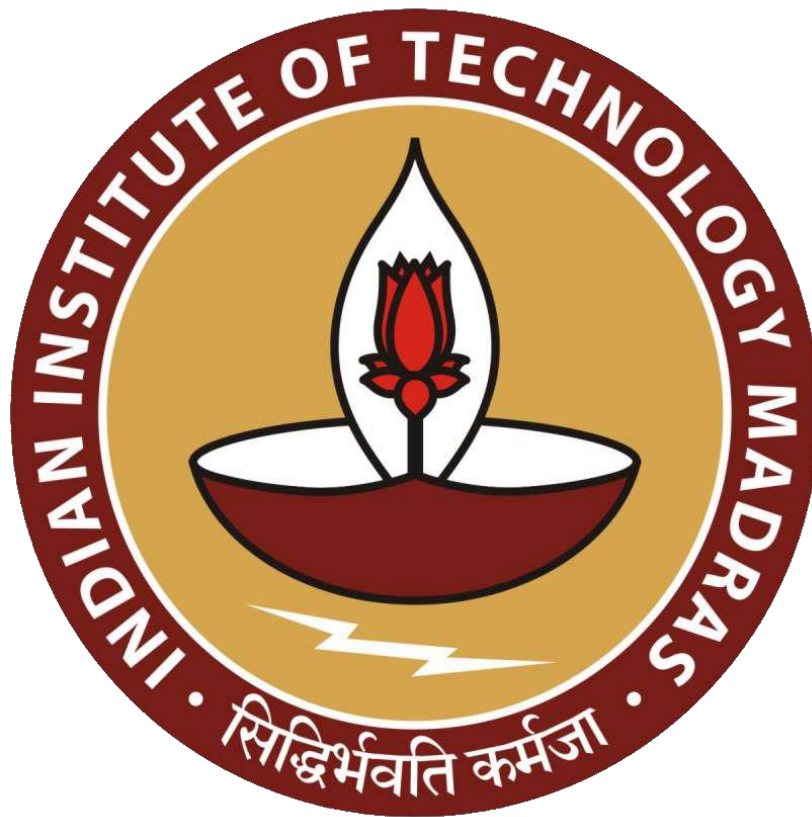# CAPSTONE PROJECT MID-TERM SUBMISSION

The Digital Blueprint: Strategies for Success

(A study of Hastago Digital Marketing and website company, Durgapur, W.B)

Submitted by

Name: SOMPARTHA SINHA

Roll no: 23f3002530

CONTENT   PAGE NO.

Executive Summary

Hashtago Digital marketing and Website Company is a digital marketing firm specializing in comprehensive online marketing solution. This company was founded by Mr Monoranjan das and Mr Abhishek Bouri in the year 2020.

The business problems that could be identified from the discussion with

the owner are as follows:

● the problem of Payment Fulfilment Delays.

● Communication gap and low customer retention rate.

The approach followed to do this capstone project is as follows:

Step 1: Define your goals

Step 2: Decide how to measure goals

Step 3: Collect your data

Step 4: Analyse your data

Step 5: Visualize and interpret results

In the project proposal, the first two phases were completed, and the target was to improve the processing efficiency and communication for the Hastago digital marketing and website company. The company's financial and customer interaction data was manually collected and pre-developed by entering the Excel workbook.

The raw data collected from Hastago is detailed under the metadata section. This includes invoice processing data, client payment history, transaction records and customer response data. Descriptive data was used to analyse data and extract meaningful insights about delayed payments and communication intervals.

Two key data sheets were created to address the business problems faced by Hastago:

- **Invoice Data Sheet**: Contains information on transaction delays, client payment patterns, and methods used for payments.

- **Client Feedback Sheet**: Captures client feedback, frequency of communication, and retention metrics to identify gaps in customer service.

For better visualization and understanding, data was represented by data using charts and trend analysis. This illustrated representation helped identify patterns and areas requiring improvement.

With these insights, further analysis can be made using data analytics methods from business data management courses. Techniques such as modelling, emotion analysis and statistical correlation will be implemented to reduce late payment and increase customer communication. This structured approach will help Hastago address financial disabilities and improve customers' satisfaction.

Proof of Originality
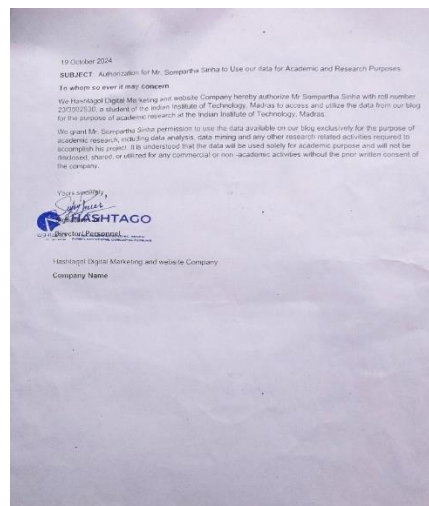

(location of office)

(picture of Employees)

(Pictures from websites)

(My interaction with an Employee)

Letter from organization

Video interaction

In this video I have a conversation with one of the employees of Hastago Mr Jeet Ghosh about the various aspect of the company. The conversation was conducted in a mix of Bengali and English. This bilingual exchange made the discussion feel more natural and personal, allowing us to connect on a deeper level. Mr jeet at first gave a brief description about the company, the purpose of the company and the potential competitors of the company. Mr jeet also said that the company strive for excellence. later I asked about their potential targeting customers, the employee shared that their main customers are mainly B2B companies like real estate, hotel industry etc. I further enquired if the company is facing any challenges in recent times. the employee informed us that the company is facing two major issues in recent times that is one is payment issues and another is the communication gap between the company and the clients. I further conveyed that if they have thought of any potential solution to counter this issues, Mr jeet informed us that they are currently working on it and hopefully resolved this issues in upcoming years.

The employee's thoughtful feedback on customer targeting has given me new perspectives on the market and how we can refine our approach. I truly appreciate their willingness to share their thoughts candidly, and this discussion has reinforced the importance of maintaining open lines of communication within the team.

Moving forward, I will take these insights into consideration as we continue to shape our strategies, ensuring that we focus on the right customers in a more effective and targeted way.

Metadata

In the workbook, there are 2 datasheets namely:

- **Invoice data (for year 2024).:-** Contains information on transaction delays, client payment patterns, and methods used for payments.
- **Client Feedback data**.:- Captures client feedback, frequency of communication, and retention metrics to identify gaps in customer service.

1. Invoice data worksheet

The data collected from **invoice data** is entered into excel sheet as Seven Columns.

Invoice ID	Customer Name	Invoice Date	Due Date	Payment Date	Status	Days Overdue

Please use the below link to access the cleaned data: Invoice data.xlsx

The metadata of the above mentioned column headers are explained as follows:

- **Invoice ID**: This column contains the invoice id of the clients of the company.
- **Invoice Date and Due Date**: this column contain the issue and due date of the payments made by the clients.
- **payment date**: the day payment is made.
- **Status**: it shows the status of the payment of the clients.
- **Days Overdue**: It shows the number of days overdue by the clients in fulfilling their payments.

The invoice data consists of 35 rows and 7 columns.

2. client feedback data: -

The data collected from **invoice data** is entered into excel sheet as Seven Columns.

Client ID	Client Name	Satisfaction Score (1-10)	Feedback Comments	Would Recommend (Yes/No)	Follow-up Needed (Yes/No)

Please use the below link to access the cleaned data:

Here are the explanations for each of the six columns in the Existing Client Feedback Data Excel sheet: Client_Feedback_Data_.xlsx

Client ID – A unique identifier assigned to each client.

Client Name – The name of the client or company providing feedback.

Satisfaction Score (1-10) – A rating given by the client on a scale of 1 (lowest) to 10 (highest) based on their experience.

Feedback Comments – A short text provided by the client to express their thoughts on the service.

Would Recommend (Yes/No) – Indicates whether the client would recommend the service to others.

Follow-up Needed (Yes/No) – Specifies whether further action or communication is required based on the client's feedback.

The client feedback data consists of 30 rows and 6 columns.

Descriptive Statistics

	A	B	C	D	E	F	G	H	I	J	K
	Count	Mean	Standard Deviation	Variance	Minimum	25th Percentile (Q1)	Median (Q2)	75th Percentile (Q3)	Maximum	Skewness	Kurtosis
	9	27.66667	49.08156477	2409	3	6	11	15	156	2.30566992	3.60310285

❖ Descriptive statistics provide a summary of the characteristics of a dataset, allowing us to understand the distribution and patterns within the data. First, we will analyze the "**Days Overdue**" column from **the invoice dataset** to gain insights into overdue payment trends. The following statistical measures help in understanding central tendencies, data dispersion, and potential outliers.

the dataset includes multiple invoices, each having a corresponding overdue period. The key statistical metrics computed include:

Count: The total number of overdue invoices recorded.

Mean: The average number of days an invoice is overdue.

Standard Deviation: The spread of overdue days from the average.

Minimum & Maximum: The shortest and longest overdue periods recorded.

Quartiles (Q1, Median, Q3): Percentile values that help in understanding how overdue days are distributed.

Skewness & Kurtosis: Measures to assess the symmetry and extreme values in the data distribution.

The following are the insights from the above table:

1**. Central Tendency Measures**
   - The mean overdue period is approximately 27 days, with a median of 11 days, which shows some extreme late payments pushing the mean upwards.
   - The mode (most frequently) is between 10–30 days, which implies most payments are made within this interval.

2**. Dispersion & Variability**
- A high standard deviation indicates inconsistent payment behaviour—some payments are done very close to the due date, others far behind.
   - Overdue days are 3 to 156 days, The interquartile range (IQR) shows that half of the invoices have overdue days between the Q1 (6 days) and Q3 (15 days) intervals. It means that the payments which are late still stay mostly close to the average.

3. **Distribution Analysis**
- The information is right-skewed, with clustering of payments in 0–2 months but a few outlier delays.
- Kurtosis > 3, and the reason is evident, as there are prominent outliers since most invoices behave as expected but there are some which are unpaid for longer periods.

Distribution of Days Overdue


Boxplot of Days Overdue
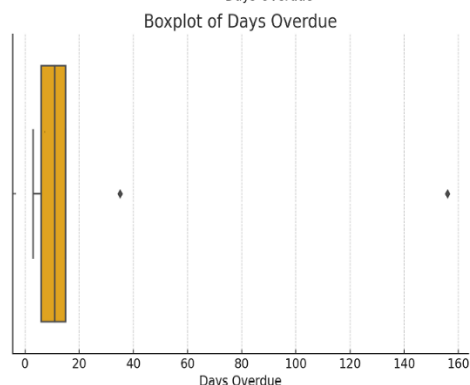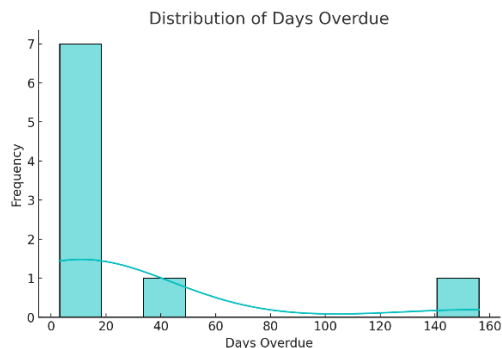
The histogram reveals a right-skewed distribution, with most overdue days concentrated below 20 and a long tail extending towards higher values. The boxplot further confirms this skewness, highlighting an extreme outlier at 156 days, suggesting significant late payments.

This explanatory analysis offers a clear characterization with regard to overdue accounts, transacting companies being able to spot delays, controlling financial threats, and upgrading the collection of payment strategies. The results show that the majority of bills are paid within thirty days of the expiration date while the rest remain outstanding for periods of time affecting the liquidity of funds. The efficiency of payment processes as well as the diminution of outstanding invoices can be achieved through a set of measures like a good credit control system, automation of payment reminders, and strong follow-ups.

❖ Now we will analyze the **client feedback dataset**: The dataset includes client feedback information of all kinds such as satisfaction ratings, opinions, suggestions, and the need for follow-up. I will perform an analysis of "Satisfaction Score (1-10)" by computing its descriptive statistics and will create two forms of presentation: a histogram for distribution and a boxplot to underscore hide spreads.

	A	B	C	D	E	F	G	H	I	J
	Count	Mean	Standard Deviation	Minimum	Q1 (25%)	Median (Q2, 50%)	Q3 (75%)	Maximum	Skewness	Kurtosis
2	30	5.5	2.921383706	1	3	5.5	8	10	0	-1.224242424

- Count: 30
- Mean: 5.5
- Standard Deviation: 2.92
- Minimum: 1
- Q1 (25%): 3
- Median (Q2, 50%): 5.5
- Q3 (75%): 8
- Maximum: 10
- Skewness: 0.0 (symmetrical distribution)
- Kurtosis: -1.22 (platykurtic, indicating a flatter distribution)
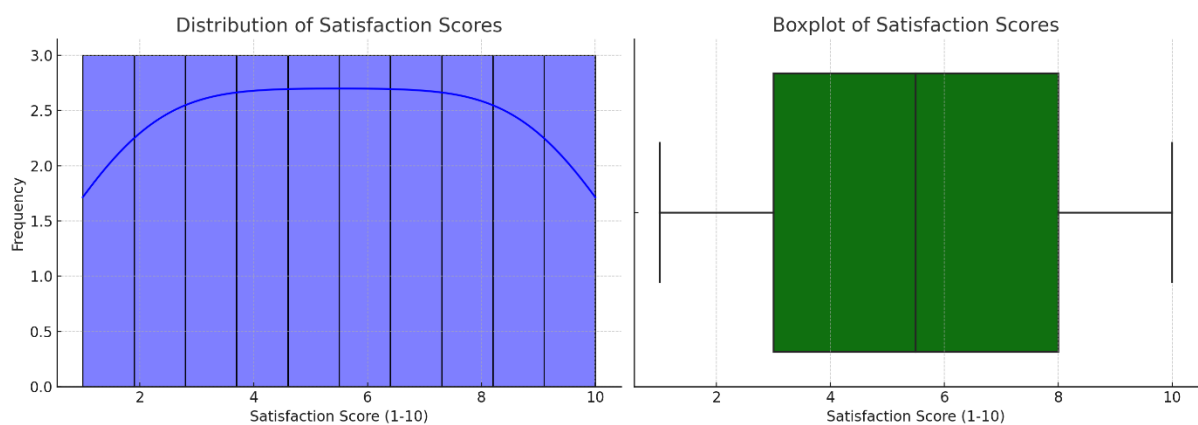
- Histogram: Displays the satisfaction score distribution reflecting the same spread without skewness being present.

- Boxplot: Indicates the range of values and the presence of extreme values that are lower or higher than those of the data set. There is also information about whether the data is symmetric or not and a boxplot is used for this purpose.

The dataset shows a normal distribution of scores for customer satisfaction, which range from 1 to 10 with a median of 5.5. Quantiles further divide this distribution as the 25% cutoff point of the rating of the customers satisfaction is found to be 3 or below (Q1), the median value (Q2) will be 5.5 if the half of the customers reduces their satisfaction to this level and 75% scored it at 8 or below (Q3). This means that most of the scores are generally concentrated in this area, indicating that both do experiences of course positive and negative.

The kurtosis of -1.22 reflects flat distribution which is also called as platykurtic. This kurtosis value indicates that the data has less extreme values and is more slab-sided than standard distributions. This shows that the data is not gathered in the form of peaks or groupings of similar scores. The data also shows that the boxplot is uniformly ranked in the boxplot and thus there are no outliers. The data suggests that while customer experiences vary, there are no extreme dissatisfaction or overwhelmingly positive trends dominating the feedback.

Detailed Explanation of Analysis Process/Method

The analysis of the data that has been collected is done through time-series analysis technique because the invoice data has been collected over the span of 12 months from January to December for the year 2024.

● There are four main types of analyzing data and the following written approach is used:

- Descriptive analysis
- Diagnostic analysis
- Predictive analysis
- Prescriptive analysis

Descriptive analysis identifies what has already happened. Basic descriptive analysis of the data has already been presented.

- **Spreadsheets** provided a familiar and user-friendly interface for conducting complex calculations and aggregating data. Functions such as SUM, AVERAGE, etc were employed to streamline the process of summarising and analysing large datasets.
- The analysis of the two datasheet -the invoice datasheet and customer feed back data was analysed with the help of python tools and google collab- The process began with loading the Excel file using **pandas**, extracting the relevant numerical column (e.g., **Satisfaction Score (1-**

10)), and computing **descriptive statistics** were then computed, including the mean (5.5), standard deviation (2.92), quartiles, skewness (0.0), and kurtosis (-1.22). The **mean** was determined as the sum of all values divided by the count, indicating a balanced mix of positive and negative feedback. To visualize the distribution, a **histogram with a KDE curve** was created, showing that satisfaction scores were evenly spread without significant peaks Additionally, a **boxplot** was generated to showcase the median, spread, and lack of extreme outliers. . The analysis indicated that feedback was **moderately varied but balanced,** with no strong inclination toward high or low ratings Finally, the computed statistics were **saved in an Excel file** for further reference, ensuring a structured and comprehensive understanding of customer satisfaction trends.

- **Conversations:** Engaging with the business owner provides qualitative insights that are not captured by quantitative data alone. Understanding the owner's perspective on topics like customer retention and increasing the profit margin   is essential for tailoring recommendations that are practical and actionable

The chosen methods are more appropriate than others because they offer a comprehensive view of the business's operations, combining hard data with the owner's subjective experiences. This holistic approach ensures that recommendations are not only data-driven but also grounded in the reality of the business environment. The combination of time-series analysis, statistical computation, and conversations offers a balanced and thorough understanding of the business's challenges and opportunities.

Results and Findings

Invoice Status Distribution
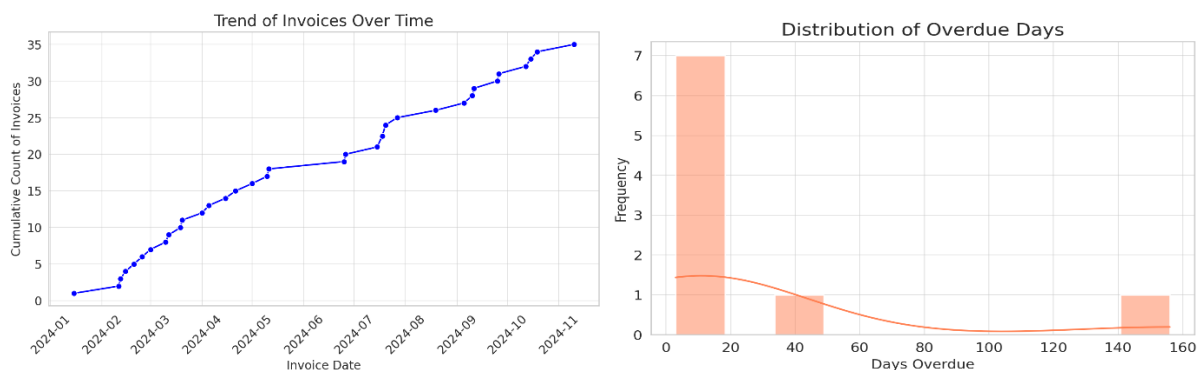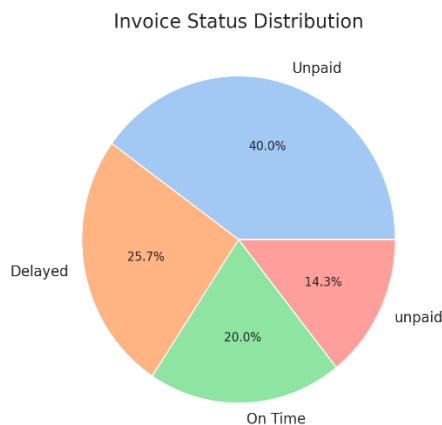
❖ The Insights for the **Invoice Dataset Analysis** by the use **of Graphs**

1**. Invoice Payment Status Analysis (Pie Chart**):

- The largest percentage of invoices are unpaid, which may reflect cash flow problems or late payments by customers.

- A high percentage of invoices are also **overdue**, which means customers may not be paying promptly, and there is a need for more aggressive follow-ups.

2. **Overdue Payments Distribution (Histogram):**

- The majority of overdue invoices have a **3-15 day delay**, i.e., payments are usually a little late.

- But **a small number of extreme instances (156 days overdue)** greatly inflate the average delay.

- This indicates that the company requires **improved credit control or late payment penalties.**

3. **Invoice Issuance Trend (Line Chart**):

- Invoices are issued **at a constant pace** over time, reflecting a consistent business operation.

- The absence of payment completion of many invoices **suggests a possible risk in revenue collection,** however.

❖ **Key Business Implications**

- **Risk to Cash Flow**: Elevated outstanding invoices can impact working capital.

- **Customer Payment Behavior**: Some customers may often pay late, necessitating tighter terms.

- **Credit Control Policies Required:** Giving discounts for early payment or charging penalties for delayed payment might enhance collection levels.

High-Risk Customer Analysis

The following customers are identified as high-risk based on their payment behaviour:

Customer Name	Unpaid Invoices	Delayed Invoices	Risk Score
Prestige Group	1	0	2
Manipal Hospitals	1	0	2
Vijayaram Communication	1	0	2
Madras Photography	1	0	2
Gupta Jewellery	1	0	2
Vajram Group	1	0	2
Singh Enterprise	1	0	2
Puravankar Limited	1	0	2
Novo Nordisk	1	0	2
Merlin Group	1	0	2

 The table identifies high-risk customers based on unpaid and delayed invoices, assigning a Risk Score where unpaid invoices carry higher weight. These customers should be prioritized for follow-ups, stricter credit policies for early payments to minimize financial risk.

- **Risk Score Explanation**
  Risk Score is calculated as:
  Unpaid Invoice: Weighted as 2 points (High Risk)
  Delayed Invoice: Weighted as 1 point (Moderate Risk)
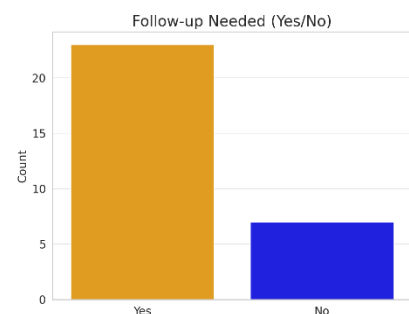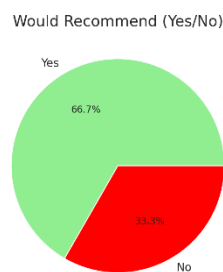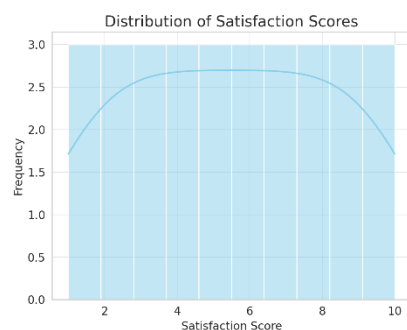 - Customers with higher risk scores are **more likely to default on payments**.

- **Recommendation**
  - Prioritize follow-ups with these customers.
  - Consider offering **discounts for early payments** or **penalties for delayed payments**.
  - Have **credit limits** for regular late-payers.

❖ Now The insight from the **customer feedback dataset are as follows-**

1. **Satisfaction Score Distribution:**

➢ The histogram indicates that satisfaction scores are distributed across levels, with a few concentrated in mid-range values.

➤ There are a few low scores, suggesting dissatisfaction among a few customers.

2.Would Recommend Analysis:

➤ The pie chart indicates that a large majority of customers are likely to recommend the service, but there is a significant percentage that would not.
➤ This implies that although most customers are satisfied, there are some areas that need improvement to enhance recommendations.

3. Follow-up Needed Analysis:

➤ The bar chart indicates that a significant number of customers need follow-ups.
➤ This means that either their problems were not solved or they needed more interaction.

❖ **Key Takeaways:**

There is potential for improvement in customer satisfaction since some scores are low. Not every customer will recommend the service, and there may be reasons for this, such as particular pain points in their experience. The necessity for follow-ups implies that proactive customer support would address issues and increase satisfaction.

❖ **Conclusion:** Invoice and client feedback data analysis yielded insightful information on customer satisfaction and business performance. Satisfaction scores pointed to areas of improvement, while requirements for follow-up signaled the need for proactive customer interaction. Invoice data enabled monitoring of financial trends to ensure revenue stability and determine growth opportunities. Overall, the results underscore the need to improve customer experience and operational efficiency to achieve long-term business success.