Forecasting Optic Fiber Deployment for enhanced Revenue Planning

A Mid-term Report for the BDM Capstone Project

Submitted by

Name: Archit Handa

Roll Number: 22f2000744

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

Declaration Statement

I am working on a Project titled “Forecasting Optic Fiber Deployment for enhanced Revenue Planning”. I extend my appreciation to **Sterlite Technologies Limited (STL)**, for providing the necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through analytical procedures.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfillment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.

Signature of Candidate:

Name: Archit Handa

Date: 6th April, 2024

1 Executive Summary and Title

This BDM Capstone Project aims to optimize and streamline the 'Trenching and Ducting' (T&D) process for optic fiber cables at Sterlite Technologies Limited (STL). 'Trenching and ducting' is critical for capacity planning and revenue projections for the business vertical. However, STL currently is unable to accurately predict their monthly output, leading to inefficient planning, delays in turn around time (TAT) and resource wastage. Consequently, this increases costs and indirectly affects profitability.

To address this, the project ultimately aims to utilize regression models and perform statistical analysis to develop a forecasting model that can provide accurate predictions for T&D monthly output. This model can potentially aid the organization to locate and mitigate hindrances, streamline capacity planning, enhance revenue projections, and allocate resources efficiently.

Objectives for the mid-term report:

➔ Identify factors that impact 'Trenching and Ducting'

➔ Collect and clean the data to make it usable for the forecasting model

➔ Identify trends shown by the variables in-consideration, extract any hidden patterns, engineer new composite variables

➔ Analyze why these trends occur, thus derive useful insights for model building

➔ Develop a roadmap for the final model

2 Proof of Originality of the Data

- **Organization Details:**
  - **Organization Name:** Sterlite Technologies Limited
  - **Address**: STL - Gurgaon Office, 15-16th Floor, Capital Cyberscape Building, Sector-59, Gurgaon, Haryana - 122102

- **Person of Contact:**
  - Mr. Pankaj Singh, PMO Lead at STL
  - Mr. Shailendra Kumar, PMO Analytics Engineer at STL

- **Organization and Project Scope Details:**

  Sterlite Technologies Limited (STL) is a global optical and digital solutions company, which aims to deliver end-to-end data network solutions to its clients. The company (originally a copper cable manufacturing plant set up by Mr. Anand Aggarwal in 1988) provides a myriad of products and solutions, and has become a go-to service provider for its business clients, distinguishing itself as one of the market leaders. STL's operations can be primarily categorized into 3 business verticals:

Manufacturing, *Network Services* and *Digital Solutions*. This capstone project principally focuses on the operation of 'Laying Network', in particular the crucial step of 'Trenching and Ducting', which falls under the *Network Services* Business Unit (BU) of the Business-to-Business (B2B) firm. The scope of this project is better explained through the diagram in Figure 1 below.

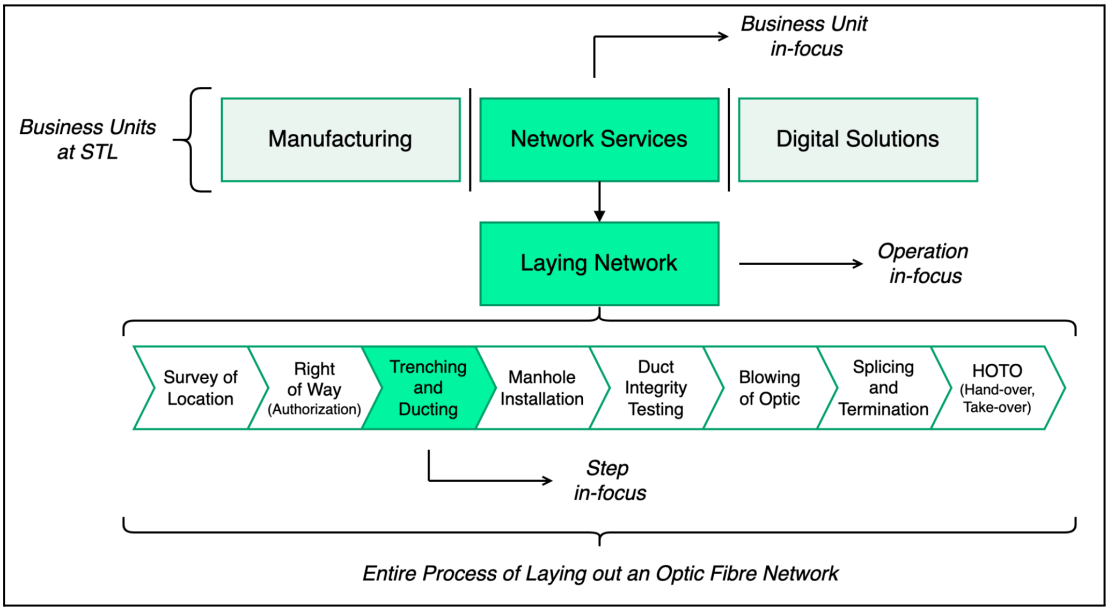


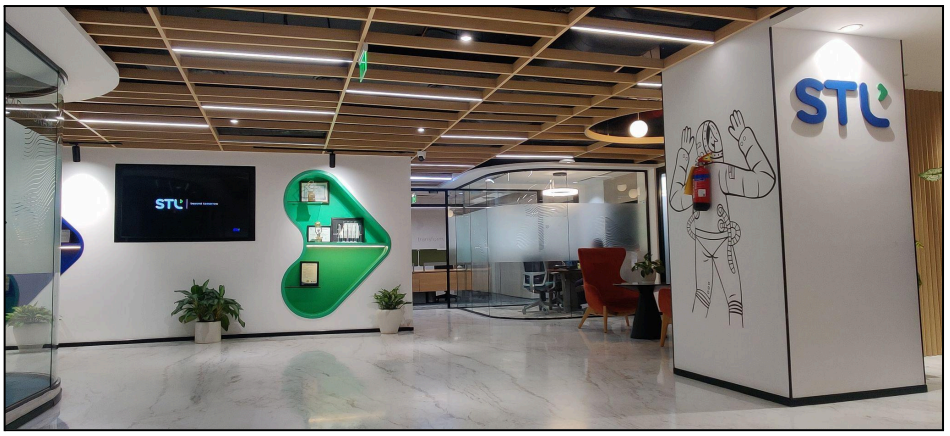
Figure 1: Scope of this BDM Capstone Project

Figure 1 aims to depict the entire process of 'Laying out a Network' and highlight the subprocess of 'Trenching and Ducting', as a part of which narrow underground channels are drilled using excavators and/or Horizontal Directional Drilling (HDD) machines. Ducts are then placed in these channels through which the Optic Fibre is passed later in the 'Blowing' stage. 'Trenching and Ducting' is the most crucial step in the entire process and forms a major portion (about 70%) of the Capacity Planning.

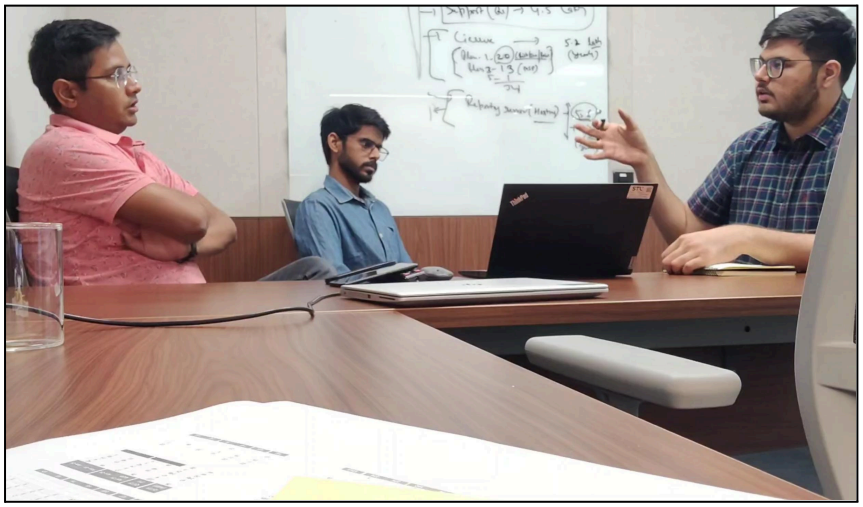
- **Images:**
  - Photographs of the STL Gurgaon Office premises as Proof of Originality

← *Figure 2: STL Gurgaon Office Entrance*

→ *Figure 3: STL Gurgaon Office Common Area*



- **Recorded Video Interaction with Mr. Pankaj Singh and Mr. Shailendra Kumar:**



← *Figure 4: Discussing the overall process and understanding the problem at hand*

(Left to Right: Mr. Pankaj Singh, Mr. Shailendra Kumar, and Archit Handa)

Interaction Video.mp4 https://drive.google.com/file/d/1bG91YSR1iV3p3Xa0sLq_kJU8nUyH9Wg_/view?usp=sharing

- **Letter of Offer from Organisation:**

Letter from Organisation.pdf

https://drive.google.com/file/d/1SeJs6w5svvsEBqPUpV0ilqycQra9rG-X/view?usp=drive_link

STL

www.stl.tech

Date: 14th March, 2024

Ref No. STL/CL/00011

Archit Handa
IIT, Madra

Sub: Internship Offer Letter

We are pleased to offer you an Internship at Sterlite Technologies Ltd., starting **15th March, 2024** at **Gurgaon** location. The Internship assignment will automatically come to an end at the close of working hours of **15th April, 2024**. The project you will be working on will be shared with you at the time of your joining.

For Sterlite Technologies Limited

Authorized Signatory

Sterlite Technologies Limited
4th Floor, Godrej Millenium, Koregaon Road 9, STS 12/1 Pune, Maharashtra - 411 001, India
CIN - L31300PN2000PLC202408

→ *Figure 5: Authorization Letter by STL used to issue a company laptop[1], allowing access to the 'Trenching and Ducting' data*

3 Data

3.1 Data Collection Process

The problem at hand revolves around the 'Trenching and Ducting' that, as discussed before, involves extensive fieldwork. This means there are various on-ground individual features and variables that can

[1] To maintain STL's Data Privacy Laws and prevent any data leakages, I was provided with an official email ID and a company laptop.The tasks of data analysis, modeling, and report generation have all been performed on the same laptop.

significantly impact the overall productivity of the process, and indirectly, have an effect on the overall Capacity Planning. Upon enquiring Mr. Singh and Mr. Kumar about the factors to consider, there were 3 highlighted of utmost importance:

1. **Types** of machines being deployed for a ***span***[2] - JCB, Poclain, or HDD
    - Machine type is dictates its 'Design' and 'Overall' productivity
2. **Machine Count** for each type
    - Ideally, the more the machines the better. However, Mr. Singh noted that situations arise where some machines are kept idle, signifying an 'under-utilization of assets'.
3. **Work Duration** of using the machines
    - A work day is usually divided into 2-3 shifts depending on the site location.
    - Like any equipment-intensive business, certain shifts are dedicated for the maintenance and repair of the machinery

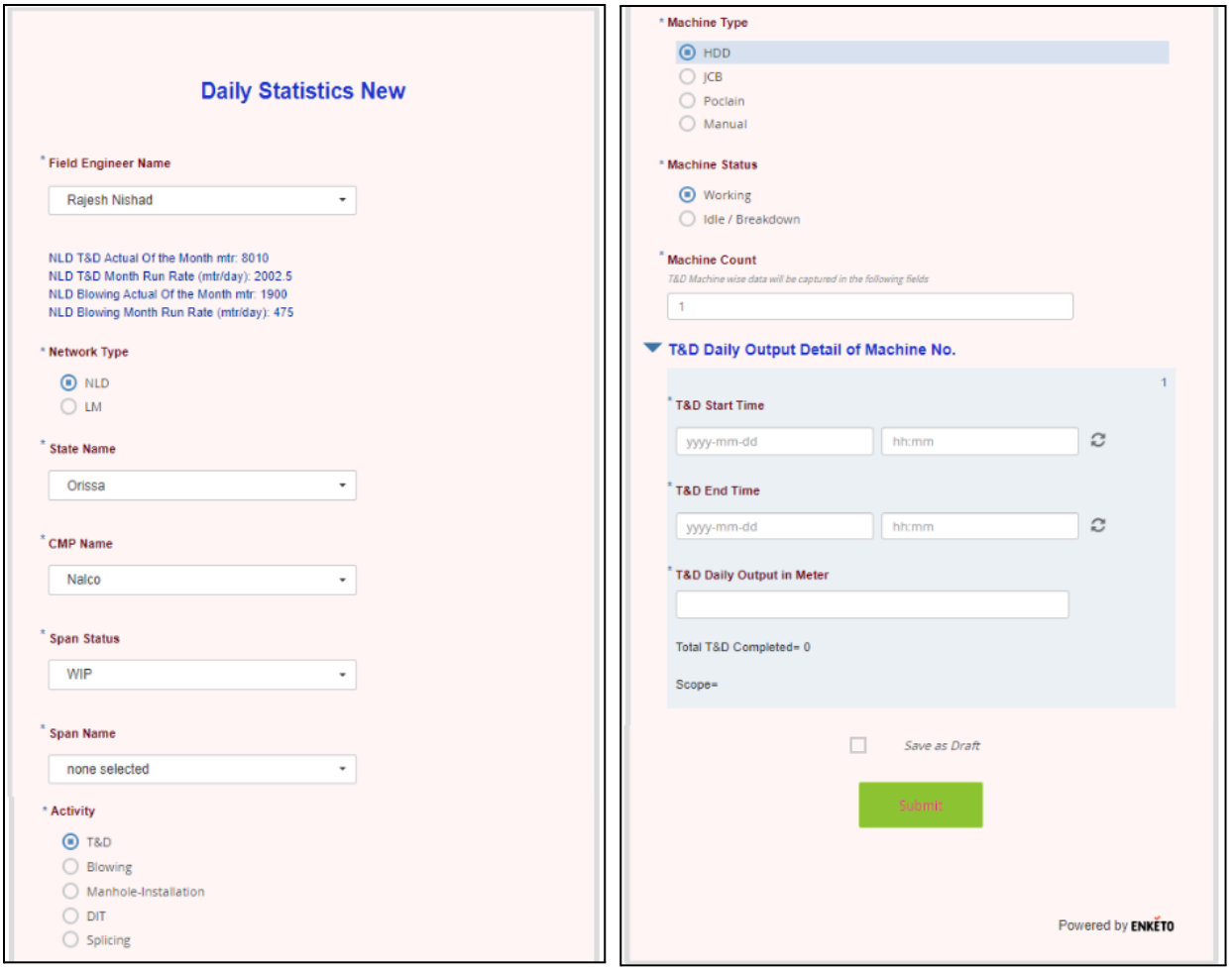
Given the variety and diversity of features, STL has already implemented an in-house open source tool named 'FieldForce' that helps them capture multiple variables for the spans they are working on. 'FieldForce' is a data entry form, developed with ease-of-use in mind, in which the field engineer can conveniently record daily site details and outcome. The factors recorded that are useful for modeling purposes include:

- State Name
- Central Management Plane (CMP)[3] Name
- Span Name
- Activity (for this project, T&D)
- Machine Type, Count, and Status
- Machine Start and End Time to calculate Work Duration
- Daily Output (in meters)

Figure 6 below shows a screenshot of the tool with sample values filled-in. The daily statistics is then sent to a dedicated server through which data can later be retrieved via API calls. Overall, The entire task of data collection is well streamlined and significantly reduces efforts in the model building process.

[2] A **span** refers to the section/segment between two points or nodes in between which optic fiber needs to be laid

[3] A **CMP** region refers to a geographical area where a myriad of network operations and functions are performed. Such a division allows STL to organize and manage different locations in a more optimized manner.



← *Figure 6: A sample 'ForceField' tool screenshot showing the details noted; the field engineer must input the daily data at the end of the work day*

3.2 Metadata, Descriptive Statistics and Data Analysis

The 'FieldForce' tool was deployed in October 2021; however, for the first 3 months the data collected was extremely noisy and unfit for modeling. It had multiple missing values (for example, there were records where machine count was 0 but the daily output was more than 0) and even invalid combinations (for instance, machine is in 'Idle/Breakdown' status yet an output was achieved). For statistical and modeling purposes, data post January 1, 2022 has been used. Upon querying about these anomalies, Mr. Kumar pointed that soon after the tool deployment, a newer version had to be released because of some technical issues, but more importantly, such a situation might have arisen because the staff was not well-versed with how to fill the form, and workshops had to be conducted to educate them regarding the same.

Even the new sample of data had to be cleaned and massaged (tasks including imputation for missing values, handling outliers found using the IQR technique, engineering new variables like work duration from start and end times) to gain more useful insights. All these tasks were performed using the `pandas` library in Python and Google Sheets. Figure 7 below shows a sample snapshot of the cleaned data.

Date	State	HDD_Count	HDD_Hours	HDD_Output	HDD_Productivity	Poclain_Count	Poclain_Hours	Poclain_Output	Poclain_Productivity	JCB_Count	JCB_Hours	JCB_Output	JCB_Productivity	Total_Count	Total_Hours	Total_Output	Total_Productivity
2022-01-01	Orissa	1	7.50	150.00	150.00	0.00	0.00	0.00		2.00	18.28	980.00	490.00	3.00	25.78	1130.00	376.67
2022-01-02	Orissa	3	22.93	1000.00	333.33	0.00	0.00	0.00		2.00	17.62	400.00	200.00	5.00	40.55	1400.00	280.00
2022-01-03	Orissa	2	12.90	355.00	177.50	0.00	0.00	0.00		4.00	34.72	1850.00	462.50	6.00	47.62	2205.00	367.50
2022-01-04	Orissa	3	25.82	780.00	260.00	0.00	0.00	0.00		3.00	23.50	460.00	153.33	6.00	49.32	1240.00	206.67
2022-01-05	Orissa	2	16.02	650.00	325.00	0.00	0.00	0.00		4.00	31.87	1020.00	255.00	6.00	47.88	1670.00	278.33
2022-01-06	Orissa	3	23.00	820.00	273.33	0.00	0.00	0.00		3.00	23.48	810.00	270.00	6.00	46.48	1630.00	271.67
2022-01-07	Orissa	4	30.85	1020.00	255.00	0.00	0.00	0.00		2.00	13.83	730.00	365.00	6.00	44.68	1750.00	291.67
2022-01-08	Orissa	3	23.02	800.00	266.67	0.00	0.00	0.00		2.00	18.80	900.00	450.00	5.00	41.82	1700.00	340.00
2022-01-09	Orissa	5	38.30	1415.00	283.00	0.00	0.00	0.00		2.00	16.05	680.00	340.00	7.00	54.35	2095.00	299.29
2022-01-10	Orissa	3	23.95	1415.00	471.67	0.00	0.00	0.00		3.00	23.42	875.00	291.67	6.00	47.37	2290.00	381.67

Figure 7: A sample Snapshot of the Processed Data

As per request of Mr. Singh, a 'Productivity' value for each machine type and overall for each day was also calculated. Initially, the formula used to calculate the metric was

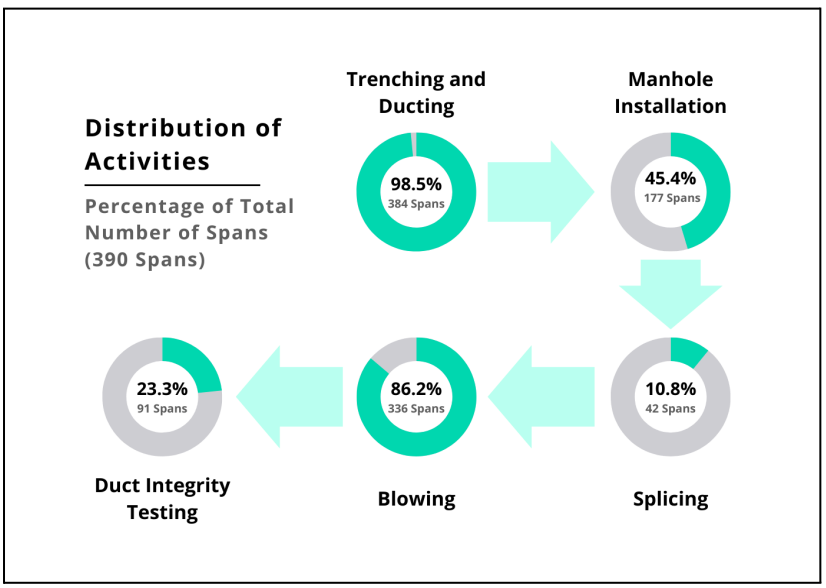
$$Productivity = \frac{Daily\ Output\ (in\ meters)}{Work\ Duration\ (in\ hours) \times Machine\ Count}$$

to arrive at a 'meters per hour per unit' value. After showing him the results, Mr. Singh advised to not consider the 'Work Duration' parameter, and instead compute 'Productivity' as a daily metric (in meters per unit), as seen below.

$$Productivity = \frac{Daily\ Output\ (in\ meters)}{Machine\ Count}$$

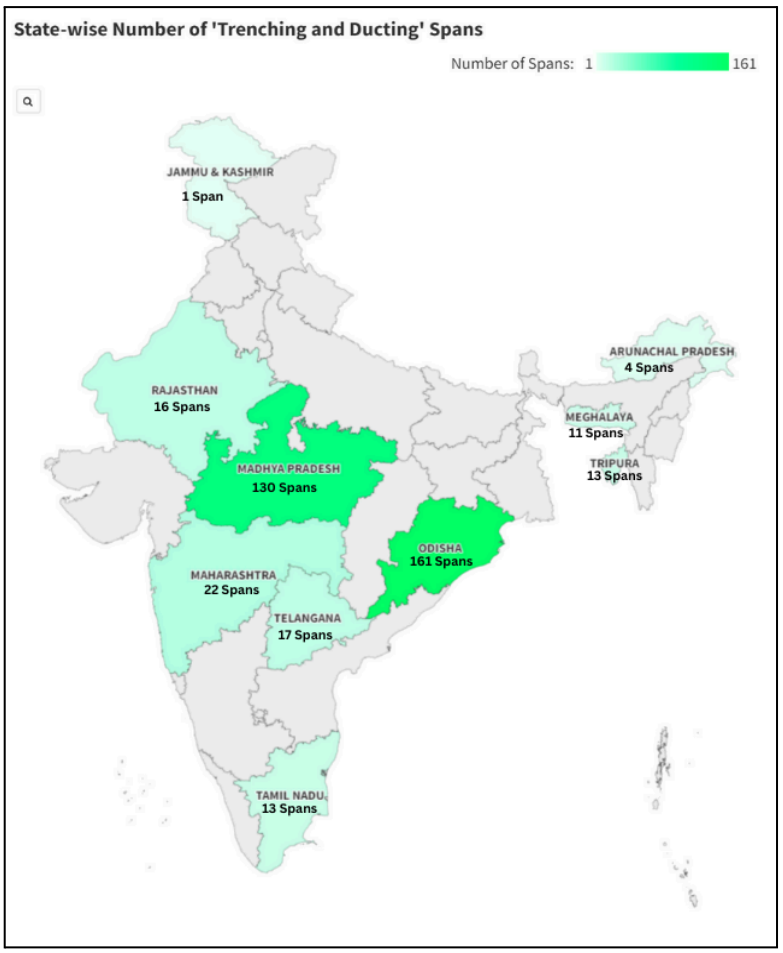
Such a change was suggested since he desired a monthly output prediction as the end result for his Capacity Planning. Using the hour-based formula would mean he would have to multiply the hourly prediction by the number of hours worked in a month, which can vary depending on location and local festivities. On the other hand, using the day-based productivity formula, he would simply have to multiply the daily prediction with a fixed value of 30-31 (28-29 for the month of February) and can use that value.

Post-processing the data, a variety of statistical techniques were employed to gain certain insights about the data. Firstly, it was found that there are currently 390 spans throughout the country where STL has worked or is working. To further analyze what activity is being performed in how many spans, a series of donut progress charts were created as shown in Figure 8 below.



← *Figure 8: Donut Progress Charts depicting in what percentage of Total Spans is each activity being performed*

The green arrows in Figure 8 show the order in which the tasks are performed for a typical span. One can notice that Manhole Installation, Splicing, and Duct Integrity Testing are performed in less than half of the spans. Enquiring Mr. Singh regarding the same, he mentioned that these 3 activities are 'quick' in nature suggesting the low occurrence due to expedited processing. On the other hand, Mr. Kumar specified that T&D and Blowing are simultaneously performed across spans and in-fact go hand-in-hand with a 10-day lag, thus the high percentage, signifying the weight these tasks carry while Capacity Planning.



A geochart was employed to showcase the distribution of the T&D spans throughout India. Madhya Pradesh (130 spans) and Orissa (161 spans) are the states of major concentration, and hence focus for STL.

→ *Figure 9: Geochart of India showing State-wise distribution of the T&D spans*[4]

Next, line charts were graphed state wise to see the trends in Monthly Output. Given the richness of data for Madhya Pradesh and Orissa, their Month-on-Month (MoM) output charts are shown in Figure 10 below.

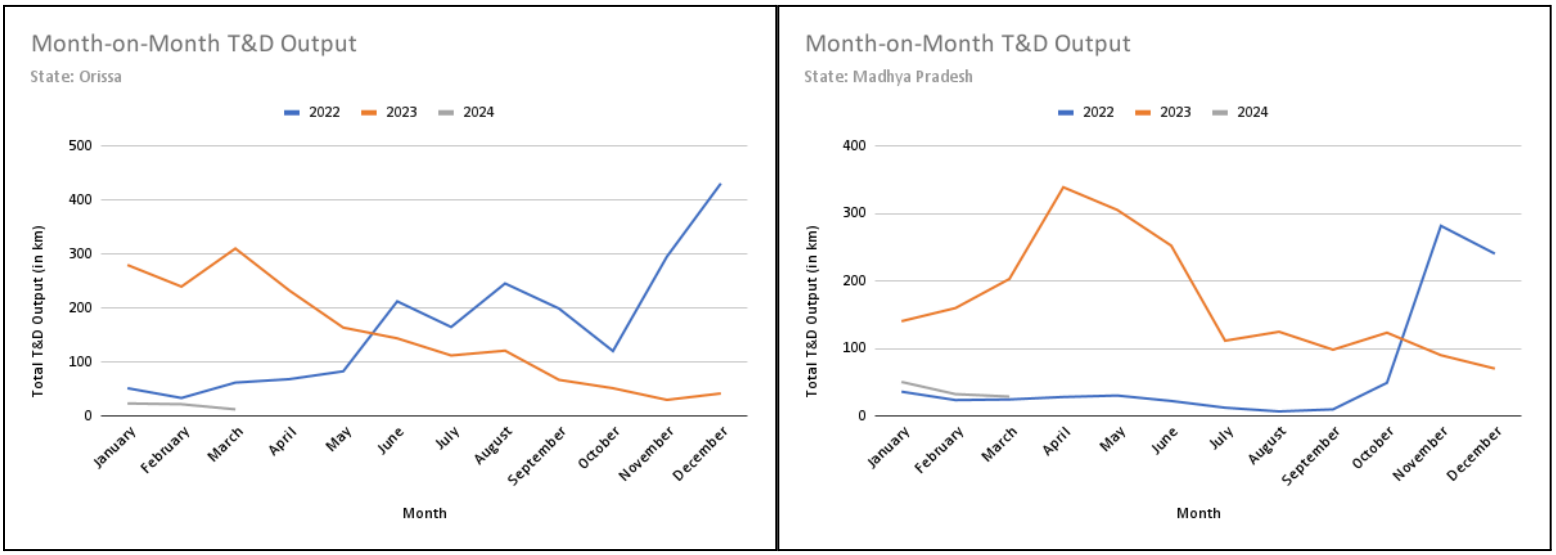
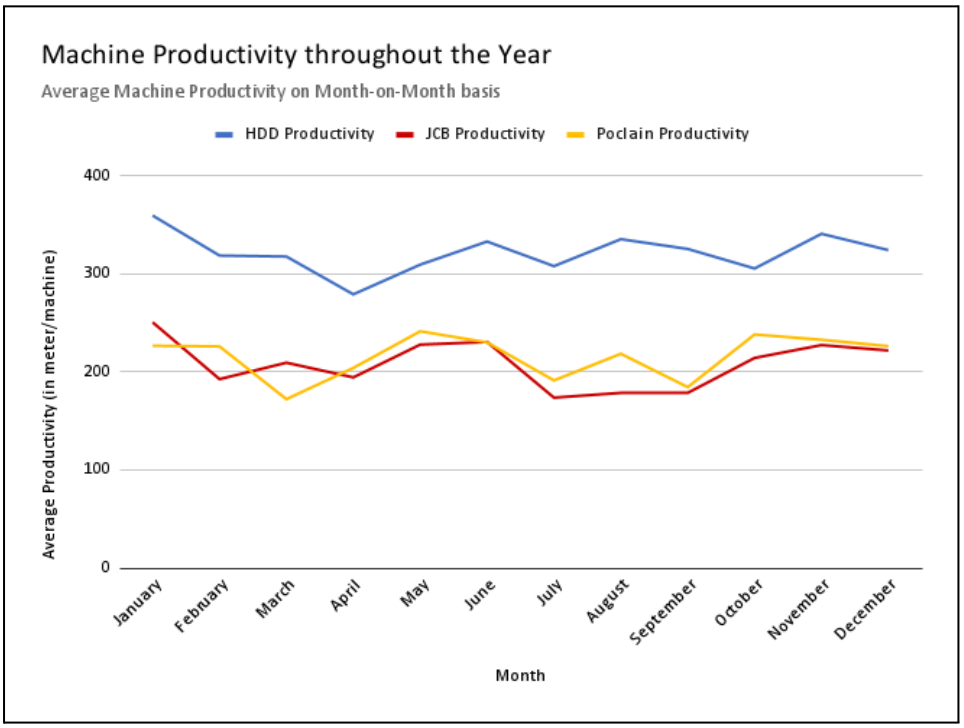


Figure 10: MoM Output for Years 2022-24 for MP and Orissa

[4] Number of spans in Figure 9 add up to 388 but as per Figure 8, T&D is performed on only 384 spans. Upon further inspection, it was found that 4 spans in Northeast India were crossing state boundaries.

Upon showcasing the trends to Mr. Singh and Mr. Kumar and enquiring them about the same, the following results were found:

- For both the states, there is a sharp increase in Monthly Output for November 2022. This is because STL got new orders to lay out Optic Fiber in these regions, specifically around 600 and 400 new spans were added in the month of November 2022 for Orissa and MP respectively.
- For Orissa, there is an upward trend for 2022 yet 2023 sees a decline in output. This is because 2023 was a year with many cyclones (Cyclone Mocha in May 2023 to Cyclone Hamoon in October 2023).
- Mr. Singh identified the trend for MP 2023 as a typical one. He mentioned that productivity tends to be highest during the summer months (March to June), then falls sharply due to the monsoon (July to September) and festive season (October to November), and remains low for the winter months (December to February) before springing back up.



← *Figure 11: MoM productivity for each Machine Type (HDD, JCB, and Poclain)*

Finally, a trend chart was plotted to see how the machine productivity varies throughout the year. It is evident that HDDs have much greater productivity than others. Interestingly, the productivity oscillates around the mean for all machine types. This suggests that changes in Monthly output is most probably caused due to lost working hours and that machine productivity can be assumed as constant for modeling purposes.

4 Summary

Key observations and findings made based on the statistical analysis performed:

- T&D step is extremely essential for STL's capacity planning (as evident from Figure 8).
- While there are multiple field variables, the most significant ones are: Machine Type, Machine Count, and Working Hours to some extent. Other variables like the State (highlights the soil condition) and Month (considers the seasonality effect) also contribute to the T&D output.
- Orissa and MP are the most vital states for STL given the number of spans.
- Productivity for each machine type remains consistent throughout the year. As aforementioned, it can thus be treated as a constant, but more importantly, it highlights the imperative effect of season on the final output.

5 Roadmap for Final Report

As per the requirements of Mr. Singh, the model must take the number of machines, type of machines, state, and month as input parameters and return a monthly T&D output figure that can be used further in capacity planning. The objectives as of now for Final report:

- Train an array of models, including Multiple Linear Regression (MLR) model for regression analysis and Weighted Moving Average model for time-series analysis.
- Determine how good the models perform on unseen data.
- Observe and evaluate any useful insights gained from the model.
- Deploy a final model and propose some recommendations that can help STL streamline their Capacity Planning process.