Cost Progression and Forecasting from Stage-Wise Billing Data in Public Infrastructure Projects

A Proposal report for the BDM capstone Project

Submitted By :

Name: Manas Taneja

Roll No.: 23f1002121

Mail:
23f1002121@ds.study.iitm.ac.in

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Content

1. Declaration Statement :

I am working on a Project titled "**Cost Progression and Forecasting from Stage-Wise Billing Data in a Public Infrastructure Project"**. I extend my appreciation to Taneja Vidyut Control Pvt. Ltd., 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 thorough 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: (Digital Signature)
Name: Manas Taneja
Date: 02-06-2025

2. Executive Summary:

This project is based on a private infrastructure contracting firm specializing in turnkey electrical and civil projects for government institutions. The organization follows a milestone-based billing mechanism, where contractors submit RA (Running Account) bills at successive stages of project completion. These bills record executed quantities and associated costs for individual line items, reflecting the project's evolving financial footprint. Despite having systematic records across multiple RA stages, the company lacks a structured analytical approach to extract insights from them.

The core problem lies in the absence of data-driven visibility into how costs build up across project stages or domains. As a result, the firm cannot proactively identify cost spikes, detect overruns, or forecast final expenditures based on early billing trends. This affects budgeting accuracy, procurement scheduling, and resource allocation.

To address this, the project will collect and structure data from eleven RA bills, six from civil works and five from electrical works, of a single completed project. Using Excel and Python (Pandas, Plotly), we will conduct exploratory data analysis, visualize trends, and apply regression models to forecast final project costs from early-stage data. The output will include cost progression charts, domain-wise cost concentration, and a predictive framework that enhances financial planning and decision-making for future projects.

3. Organization Background:

Taneja Vidyut Control Pvt. Ltd. (TVCPL) was incorporated on 25 September 1999 and is headquartered in Malviya Nagar, New Delhi. With an authorized share capital of ₹50 lakh and a paid-up capital of ₹22.84 lakh, the firm operates as a privately held electro-mechanical EPC contractor. The company employs between 11 to 50 professionals, including project engineers and field staff. TVCPL specializes in turnkey civil and electrical solutions for government and institutional clients, executing work such as HT/LT installations, lift systems, switchgear, wiring, and structural civil components like RCC, flooring, and excavation. They follow a Running Account (RA) billing process, submitting itemized invoices at multiple project milestones. Despite maintaining systematic billing documentation, the company lacks a formal analytics function to translate these records into actionable insights.

This project directly supports their mission of delivering reliable, high-quality infrastructure by introducing data-driven analysis of billing data. By applying quantitative methods to spend trends, TVCPL can enhance cost control, procurement timing, and project planning, addressing a current organizational gap.

4. Problem Statement:

- **Analyze stage-wise cost accumulation using RA bills:** Identify how cost builds up across different billing stages and domains (civil/electrical) to detect high-expense periods.

- **Develop predictive models to forecast total project cost:** Use early-stage RA bills (e.g., RA1–RA3) to estimate the final project cost using regression models.
- **Visualize domain-wise and item-wise spend trends:** Generate time-based visualizations to highlight categories with highest cost contributions and unusual spikes.

5. Problem Background:

In infrastructure contracting, cost control and forecasting are critical to project success. Contractors submit RA (Running Account) bills at various stages, detailing quantities and costs of executed work. While these bills serve an accounting purpose, their analytical value often goes untapped. In the case of TVCPL, multiple RA bills are systematically prepared, yet there is no mechanism to leverage them for insight into cost buildup patterns, domain-specific burn rates, or early signs of project overruns.

The absence of historical analysis means the firm cannot benchmark new projects, identify consistently high-cost categories, or proactively manage cash flow and procurement. Without a structured review of previous project billing data, each new project starts with limited predictive guidance. This lack of feedback loop hampers both planning efficiency and financial discipline.

For instance, if electrical lift installations tend to generate disproportionate costs in late-stage RA bills, early detection could allow for better resource scheduling or supplier negotiations. Similarly, item-level irregularities, such as delayed billing or unexpectedly sharp cost jumps, could signal execution issues, if analyzed properly. In TVCPL's case, such issues remain latent due to the lack of analytical tools.

This project aims to address this gap by systematically extracting and analyzing data from eleven RA bills across one project's lifecycle, thereby converting static billing documents into actionable insights for improved financial planning and project control.

6. Problem Solving Approach:

6.1 Methods:

This project will employ structured quantitative analysis of RA bills, centering on exploratory data analysis (EDA), regression modeling, and trend tracking. EDA will be used to visualize and summarize cost accumulation patterns, identify peak spending periods, and detect anomalies across RA stages. To forecast final project cost from early-stage bills, regression models will be implemented. While several forecasting models are commonly used in construction analytics—including Linear Regression, Ridge Regression, Random Forest, and ARIMA—this project will focus on Linear Regression and Ridge Regression due to the moderate sample size and the need for interpretability. These models are well-suited for extracting actionable trends from structured numeric data and are commonly used for cost estimation and trend analysis in project management.

6.2 Data Collection:

Data will be collected from eleven RA bills (six civil, five electrical) of a single infrastructure project. The dataset will include the following features:

- Bill Metadata: RA bill number (stage), billing date, and project domain (civil/electrical)

- Item Description: Name of work/component (e.g., wiring, RCC, lifts)
- Units & Quantities: Unit of measurement, executed quantity at each stage.
- Rates & Costs: Contract unit rate, line item total cost, cumulative stage-wise cost.
- Category Grouping: Each line item mapped to broader work categories (e.g., electrical wiring, structural concrete).

  These features enable tracking cost build-up, identifying high-expenditure items, and analyzing stage-wise cost progression. Data will be manually extracted, validated in Excel, and imported into Python for further processing.

6.3 Analysis Tools:

- Microsoft Excel will be used for initial data cleaning, structuring, and cross-checking cost calculations.
- Python (Pandas, NumPy, Matplotlib, Plotly) will support data preprocessing, aggregation, and visualization. Pandas is ideal for handling tabular billing data, while Matplotlib and Plotly enable interactive cost trend plots and category breakdowns.
- Regression Modeling (Linear and Ridge Regression) will be applied to early-stage RA bill data to forecast the project's total cost at completion, providing quantitative, explainable predictions to support budgeting and risk mitigation.

All methods and tools have been chosen to ensure results are transparent, replicable, and directly relevant for business decision-making. The final deliverables will include cost progression charts, domain/category breakdowns, and model-based forecasts grounded in quantitative analysis.

7. Expected Timeline - Work Breakdown Structure :

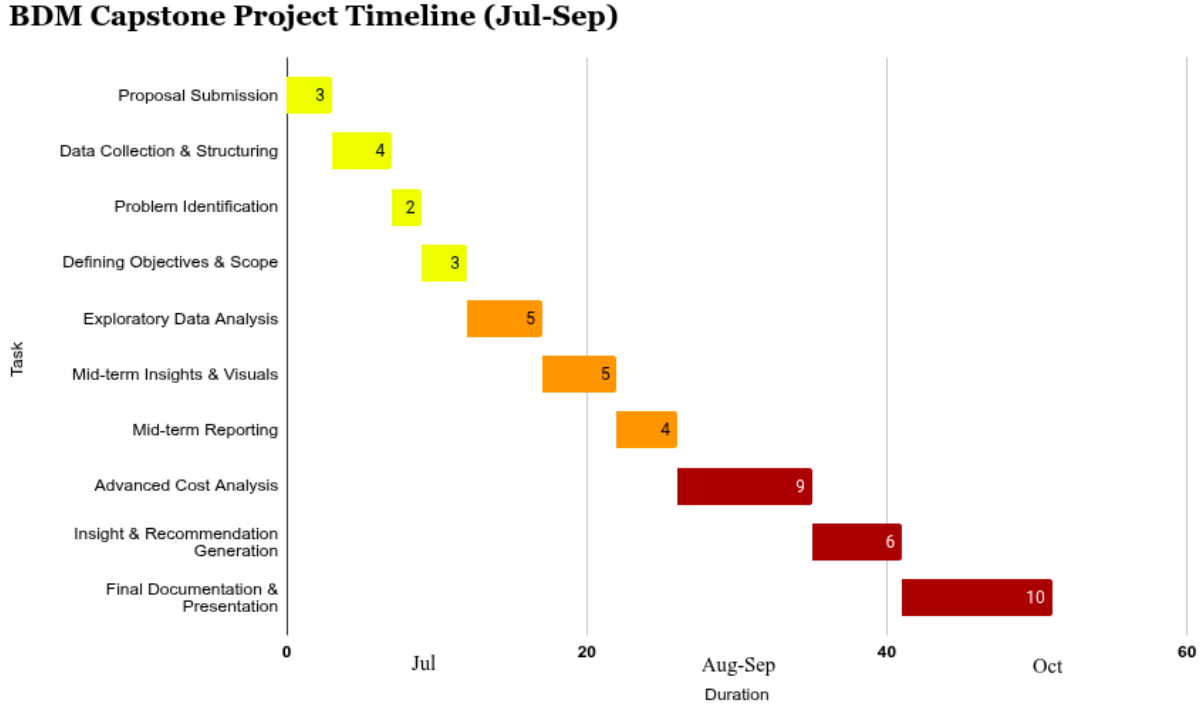

Image 1 :Gantt Chart

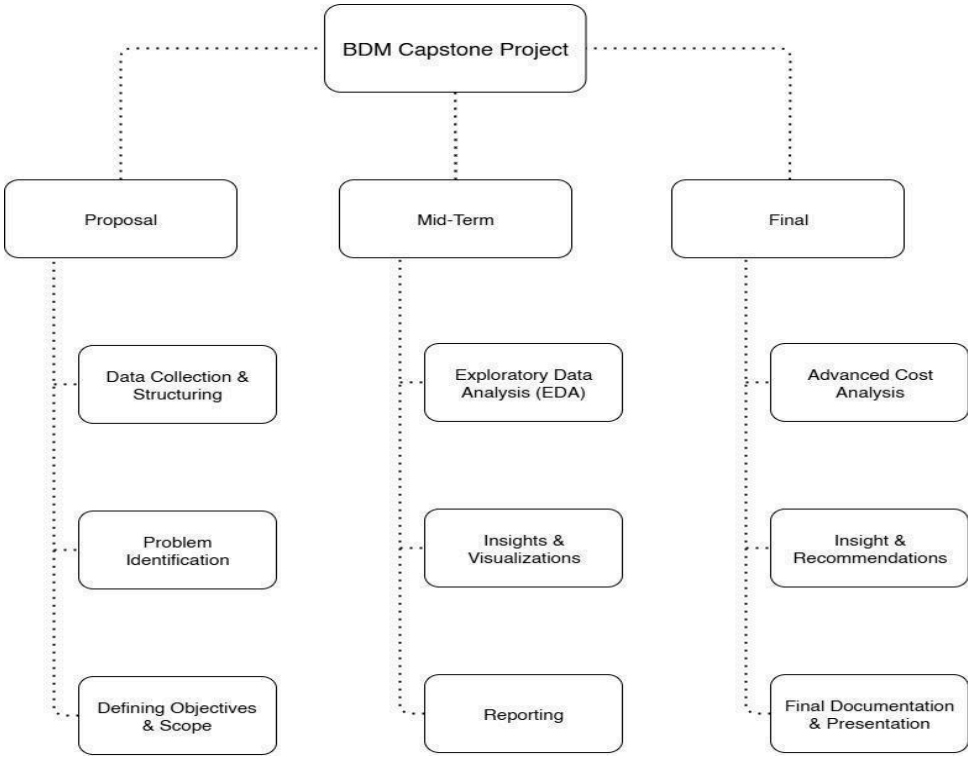

Image 2: Work Breakdown Structure

8. Expected Outcomes:

This project is expected to yield multiple data-backed outcomes aimed at improving TVCPL's project planning and financial oversight. The primary deliverables will include:

1. A cleaned and structured dataset compiled from eleven RA bills, labeled with stage-wise and domain-specific metadata.
2. Interactive visualizations depicting cumulative cost buildup, item-wise expenditure concentration, and temporal cost spikes.
3. Regression-based cost forecasting models using early-stage RA data to estimate total project expenditure.
4. Insights highlighting which categories consistently consume the highest cost and how their patterns differ across civil and electrical domains.

These outcomes will empower TVCPL to benchmark future projects, enhance budget planning, and proactively mitigate financial risks. By identifying cost escalation patterns early in the execution cycle, the company can take corrective action sooner—avoiding delays and overruns.

Crucially, all insights will be grounded in quantitative analysis rather than intuition, making them transparent, explainable, and replicable. This data-driven approach marks a shift toward evidence-based project management and aligns with the firm's broader goals of operational excellence and efficiency.