Business Data Management Project

Mid Term Submission Report

OPTIMIZING CASH FLOW AND STAGGERED CREDIT SALES SYSTEM FOR A B2B TEXTILE COMPANY

Submitted by

NAME : SHRIRAM G

ROLL NO : 21f2000668

DIPLOMA LEVEL

SEPTEMBER TERM

CYCLE 3

B.S. DATA SCIENCE AND APPLICATIONS

2021-2024

INDIAN INSTITUTE OF TECHNOLOGY, MADRAS

EXECUTIVE SUMMARY:

This report consists of the proof of originality, metadata and descriptive statistics, description of the analysis process and method, and some initial analysis results of our business data management project. The proof of originality includes a video recording of a short interaction with the business owner, a written letter expressing his support, and a link to the data that has been collected and analyzed.

The metadata and descriptive statistics provide a detailed description of the data, including the size, format, summary of the key variables and statistical parameters.

Prior to conducting our analysis, preprocessing was done to ensure the accuracy and reliability of our findings. This included cleaning and transforming the data to remove any errors or inconsistencies, as well as normalizing and standardizing the data to account for differences in scale and units. Missing or incomplete data was imputed where appropriate.

The initial analysis results show that there is a significant erratic change in the quantity of yarn purchased versus the turnover of each month, while the price of yarn follows a neat trend. The visualization charts generated from Tableau have been attached in this report.

Further, the choice of state variables, control input and target variables have been finalized and illustrated for the development of the feedback control system model design.

PROOF OF ORIGINALITY:

To prove authenticity of this business data management project data, we have included the following:

- A written letter from Mr. R.V. Devasenan, with the company's official letterhead expressing his support for the project, with his contact details and the company's GSTN enclosed within.
- Source data collected for analysis, supplied by Mr. R.V. Devasenan.

- A video recording of our interview with Mr. R.V. Devasenan, who is the proprietor at R.G. Nagappa Mudaliar & Sons, describing in brief the background of his business and the problems faced by his business. (Kindly excuse the audio issues from my side)

The above information can be accessed via this Google Drive link made accessible to the IITM Domain - BDM Proofs

METADATA & DESCRIPTIVE STATISTICS:

Our source data consists of six spreadsheets. The span of time for which the data is for the financial year 2021-2022 (April 2021-March 2022).

The first sheet, titled "Bsheet," is the balance sheet that provides an overview of the company's assets and liabilities, including capital accounts, fixed assets, loans/liabilities, and sundry debtors. The second sheet titled “PandL” is a profit and loss statement, with the primary expenses, opening and closing cash stock details.

A third sheet, titled ‘GST Purchase of Yarn’, contains details of the purchase of yarn by the company. Specifically, it contains the date of purchase, particulars of the yarn supplier, type of yarn purchased (WEFT Yarn/WARP Yarn), the quantity of yarn in kilograms, rate at time of purchase, VCH (Verification Chart Of Accounts) number and the total amount paid with GST of five percent.

A fourth sheet, titled ‘GST Sales of Poplin’, contains details of the sale of poplin cloth by the company. The details in this sheet include, the name of the customer(business that purchased cloth), length in meters of cloth sold in that transaction, VCH Type (GST Sales), VCH number, rate at which it was sold, GST and told value of transaction.

The ‘Debtors Return’ sheet provided by the business consists of the list of businesses that have defaulted in full payment of cloth sold. These businesses have either been completely unable to return the value in cash owed for the cloth purchased on credit or partially paid the company in installments and the details of those transactions are details in this sheet, including name of the company, amount returned and date of return. The ‘Sales Monthwise’ sheet contains a

month-wise tally sheet to check consistency and integrity of the data and correct for accounting errors in sales.

The data is found to be highly variable upon inspection and hence the central tendency measures Median and Mode values of the data do not yield much insight into the structure of the data as the rate, price and values of sales in meters are highly variable continuous data. The mean of the yarn purchase price 1372 rupees with a standard deviation of 669 rupees. Similarly, the mean and standard deviation of sales data is 4068 and 24771. This illustrates the high variability and erratic nature of the data and the sales trend of this business.

The yarn purchase rate however, follows a linear trend across a financial year. Since the current financial year has not yet come to an end and the financial year prior to our analysis has data that is far too sparse for conducting analysis due to lockdowns across the country due to Covid, the financial year 2021-2022 was chosen for analysis. The data of customers and suppliers is vastly different from the pre-Covid business atmosphere and hence it is unreliable to perform predictive analytics on such data.

Among all the businesses, there are 5 businesses that have defaulted on payments. Two of those businesses have returned more than 75% of their payments, and two have returned payment of at least 60% of the original quoted purchase price. One business, which has made a very small purchase, has incurred a complete loss and there have been no further transactions with that retailer.

One data point consists of a business that has made a purchase by the end of March 2022. Further data about this business' purchases, debts etc., as the financial year had come to end and the subsequent accounts data will not be closed until March of 2023.

Other statistical parameters have also been calculated using the Pandas library of Python, presenting count of records, percentile records, etc. on both purchase and sales sheets. It is attached below for reference.

	KILO GRAMS	RATE	VALUE	+ GST
count	28.000000	28.000000	28.000000	28.000000
mean	1372.705714	253.785714	349907.500000	364055.392857
std	668.919077	32.981877	180127.661819	184146.893497
min	300.000000	195.000000	69552.000000	73030.000000
25%	862.500000	231.500000	204750.000000	214987.500000
50%	1375.000000	247.500000	350250.000000	365825.000000
75%	1800.000000	276.000000	444600.000000	466830.000000
max	2721.600000	315.000000	745200.000000	745200.000000

Sheet : GST Purchase Of Yarn

	MTRS	RATE	Value	GST-5%	INCL OF GST
count	150.000000	149.000000	1.500000e+02	149.000000	149.000000
mean	4068.190667	47.828859	1.930713e+05	4859.177584	102042.729262
std	24771.326384	2.922288	1.175599e+06	2684.738452	56379.507490
min	165.800000	44.000000	7.792600e+03	389.630000	8182.230000
25%	1047.275000	46.000000	5.401678e+04	2380.125000	49982.625000
50%	1799.750000	47.000000	8.892950e+04	4436.025000	93156.525000
75%	2885.125000	47.000000	1.424691e+05	7068.170000	148431.570000
max	305114.300000	58.000000	1.448035e+07	12616.760000	264951.960000

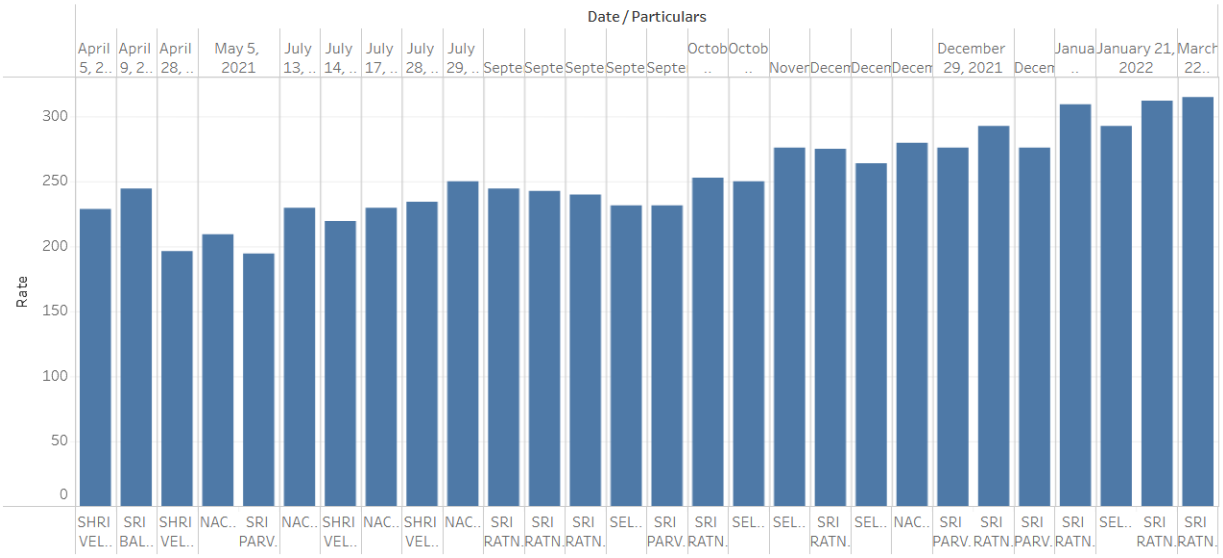
Sheet : GST Sales Of Poplin

METHOD OF ANALYSIS:

Upon surveying the data, it was found that the price of yarn follows a linear pattern. A bar chart was generated with Tableau to visualize the price trendline and it is attached below.

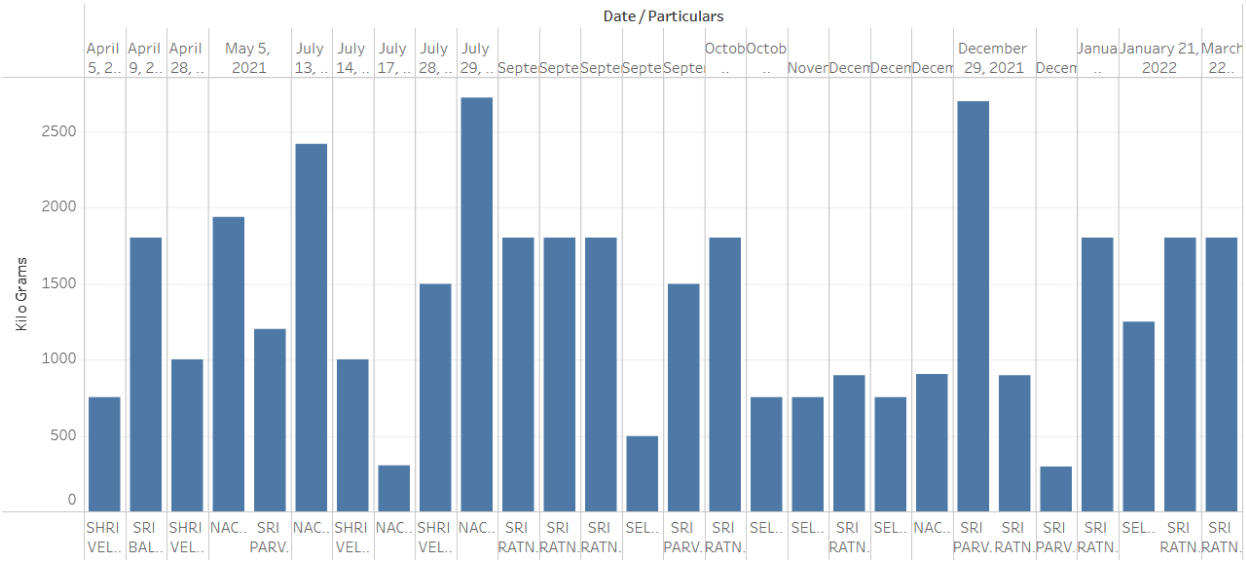
Upon correspondence with the business person, the primary cause of this trend in yarn price was inferred. The initial lower price was due to slower movement of cash and cash after initial removal of lockdown. The continued rise was the return of the price to pre-pandemic levels. However, the same rise is continued to be expected due to rapid inflation and hence as suggested

in the proposal, regression analysis would be sufficient to predict yarn price and determine appropriate pricing, with a slightly higher positive adjustment of the slope.



Price of Yarn

The total purchase of yarn is erratic however depending on the sales and return of payment of the debtors. This is easily visualized as the purchase of weights over time.



Weight of Yarn Purchased

Superimposing the above Weight Purchased Bar Chart with the Price Bar Chart at each transaction, it is clear that a feedback system design is a must based on which the decision tree would be constructed for the business person using which he may make purchase and sales decisions.

State-Space Analysis : State Space Analysis is ideal to design a feedback system with discrete-time data, although the value ranges themselves are continuous. To model a system, a state space equation is given by -

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$$
$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)$$

Here, x(t) is the state vector, which contains parameters that describe the state of the system (in this case, the business) at the given time ‘t’. Eg) Current Accounts Receivable, Stock, Losses incurred thus far, etc.

x’(t) is the vector of the first derivative of the state variables of the state vector. u(t) is the control input, which in this case is the investment of the business at that time. y(t) is the output vector, which in the case of a business is the profit or loss at the the instance of time ‘t’.

A, B, C and D are the matrices of parameters that describe the system. A is the system matrix, B is the input matrix, C is the output Matrix and D is the feedforward matrix. As this represents a system of N linear equations, where N is the number of state variables we choose to model the system.

Polynomial curve fitting or the perceptron algorithm must be sufficient to train the model and derive the values of the A, B, C and D matrices. Subsequently, with the knowledge of the values of the state vector (state of the business variables as mentioned above) that we have at any given time, we may adjust our control input (investment and expenditure) to get our desired output, calculated by solving the equation.

On analyzing the final, trained, state space equation, we may generate graphs to determine thresholds on investment, spending, credit sale to the customer etc., based on the response of the

equation to a continuous range of inputs. This can allow us to break the graph into decision regions to construct the decision tree.

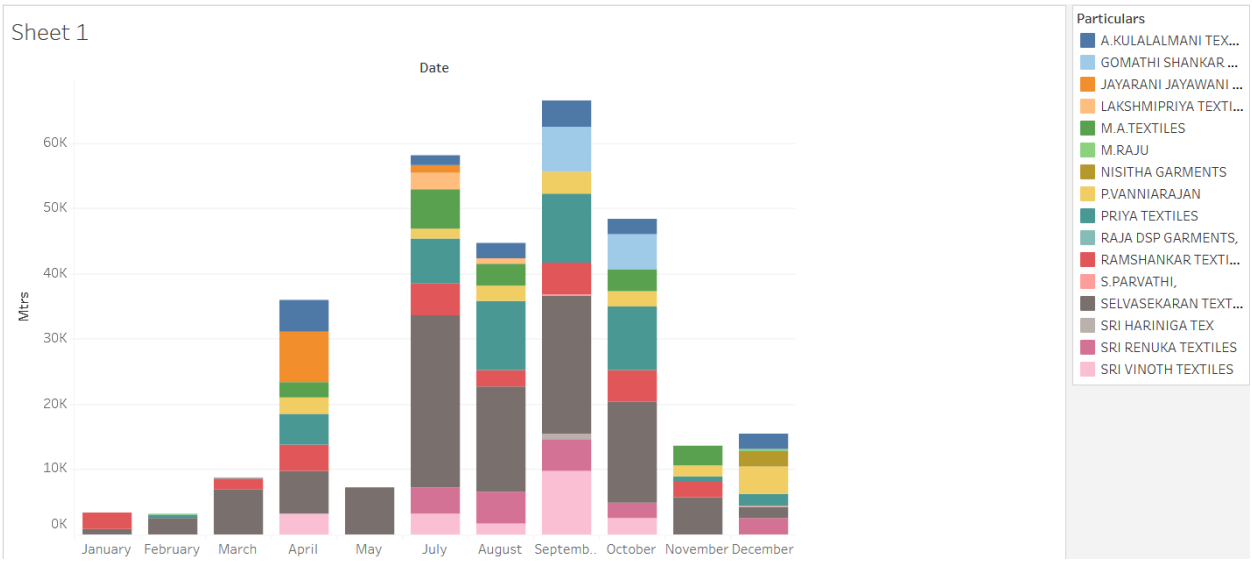
The impulse response and unit step response of the system will give a clean picture of the nature of the system. Custom input functions may also be designed using Scipy and Numpy and fed to the system to determine the response of the system over time to a pattern of inputs.

This will aid in visualizing the long term results of a decision making process and make arrangements to capitalize or mitigate the respective long term positive and negative consequences of current decisions.

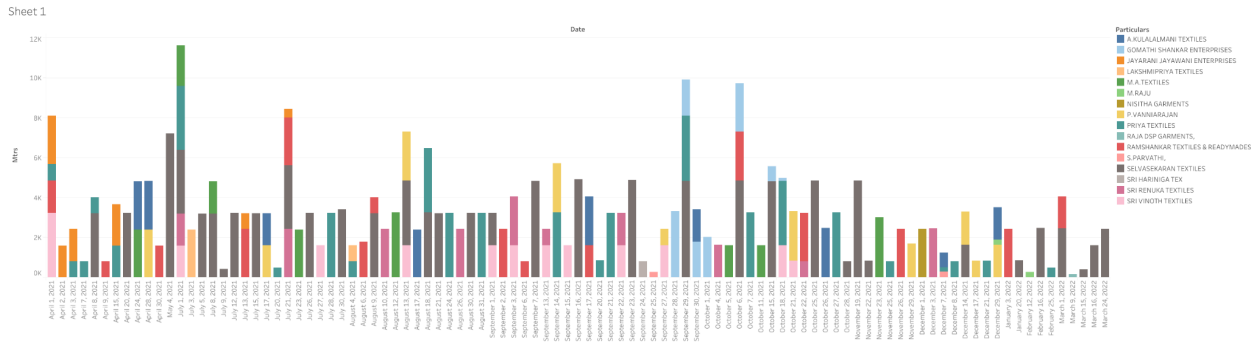
Note : References/citations are provided at the end of the report.

INITIAL FINDINGS AND RESULTS:

A month wise analysis of the sales data is provided below. The 'viz' were designed in Tableau and exported to image format to attach to the report. It is clear there are no outlier months in the cumulative data at first glance. However, the day wise sales data allows us to identify periodic spikes in demand that are not in line with the monthly aggregate sales data.



Month-wise Sales Data



Day-wise Sales Data

Same legend is utilized in both charts

Peak demand is met during festive seasons in end-customer facing business. Since those business have peak demand from October to December, our business of concern has peak demand in the months prior to those seasons where retail stores need to restock. Hence it is prudent to postulate that the control variables will be most affected by data in these months.

While data is shuffled in each epoch of training a model, it may be sagacious to not shuffle the data while training our model as learning these cycles is necessary to our approach as classic algorithmic learning approaches go. These cycles needed to be learned by the model and it is important to not shuffle the data in every epoch as is the norm by usual machine learning conventions.

LIST OF REFERENCES:

1) State Space Analysis State-Space Approaches for Modelling and Control in Financial Engineering: Systems Theory and Machine Learning Methods - by Gerasimos Rigatos
2) Control Systems Engineering - by Norman Nise
3) Optimal Control Theory, Applications to Management Science and Economics - by Suresh P. Sethi
4) Systems Science: Theory, Analysis, Modeling, and Design - by George E. Mobus
5) Near Optimal Control in Ride Hailing Platforms with Strategic Servers - Sushil Mahavir Vera *et al. (Reference paper for a similar method previously used in a different business)*