Contents

[Workforce Optimization Analytics 2](#_Toc418505570)

[Initial Project Goals 2](#_Toc418505571)

[Project Goals Achieved 2](#_Toc418505572)

[Project Implementation Approach & Details 2](#_Toc418505573)

[Understanding the Available Data 2](#_Toc418505574)

[Data Generation Process 3](#_Toc418505575)

[Technology Stack 4](#_Toc418505576)

[Assumptions 4](#_Toc418505577)

# Workforce Optimization Analytics

Problem statement: Customer has random income order flow but fixed number of employees. Presently customer does not have any control over the number of orders received or their complexity. Each order has its own complexity that requires special skill sets from the employees. To meet the demand, presently customer is cross-training their employees randomly. They would like to minimize their training costs and better prepared for the incoming order demands.

Failure to meet deadlines or unable to cope up with incoming orders could happen due to multiple reasons, one of them being in all departments there is a shortage of resources not enough to meet even the lowest order volume. In such case, training will not help at all - as moving people will only increase the speed of that particular dept. while still halting the other parts of the full pipeline. Hence our goal is to first understand the pattern of the incoming orders, their resource demands and gaps and where further optimization can lead to a quicker turnaround and delivery of the orders

## Initial Project Goals

The initial project goals were as follows:

1. Once an order is approved what is the typical delay for an order to actually start"
2. Is the delay varying with complexity- didn’t do this
3. What is the behaviour of this delay - is it varying w.r.to resource availability ?"
4. At any point of time, how many orders are being processed in parallel
5. What are the no. of parallel orders being executed
6. What is my lowest no. of orders at any given time
7. What is the highest no. of orders at any given time
8. What is my lowest no. of parallel orders for dept. X below which the order count never go low, and the highest"
9. Whether there is any relationship between Production Delay and Production Time

## Project Goals Achieved

In addition to all the above goals were able to drill down to the function in each order and answer the queries in much more detail.

1. We can add additional business insights that we were able to extract here…(can use the shiny app for that)
2. Production Delay details and to understand the Production Delay pattern

## Project Implementation Approach & Details

To understand how we were able to achieve the mentioned goals…please read on for more details

### Understanding the Available Data

The data that we used as the source for the project had the following information captured:

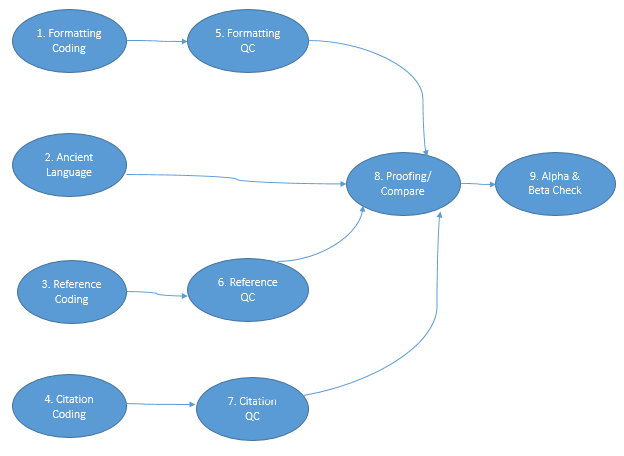
* Series ID
* Resource ID : It is a unique ID using which an order can be identified.
* Project Type: There are 5 types of orders that are being considered here.
  + Regular, Special+Regular, ePub+Regular,Standard Recycle, Limited Recycle
* Source Type: There are 5 types of activities within an project that can take up resources
  + Exporting, Keying, Keying/OCR,L\_Recycle,S\_Recycle.
* Pages: The no of pages estimated to complete the order
* Category: There are 4 categories – 1,2,3,4 that are being considered here.
* Total Estimated Character Count: It is the total characters in kb/s that are estimated for the order to be completed
* Delivery Character Count: It is the total characters in kb/s that are estimated for the order to be completed
* Production Status: It informs whether the project/order is delivered or is still a work in progress.
* Production Approval Date: It is the date when the project/order was approved
* Production Start Date: It is the date when the project/order implementation started.
* Actual Delivery Date: It is the date when the project/order delivery took place
* Estimated Hours required for: All the below are individual activities, some/all of which might be needed to complete the project/order. It basically is an estimated figure in terms of hours.
* Formatting Coding
* Ancient Language
* Reference Coding
* Citation Coding
* Formatting QC
* Reference QC
* Citation QC
* Proofing/ Compare
* Alpha & Beta Check

### Data Generation Process

The available data was not lending itself to extract the business insights that we are looking for so we had to do some initial pre-processing. Some of these steps were as follows

#### Order Execution Flow

We do know that broadly there are 9 activities that an order at max can use to complete the order. Based on the type of order there can be four independent streams (as shown below) that can be used to complete the order. We can also see that activities coded 1,2,3 & 4 can be executed in parallel, activities 5,6 & 7 can be executed in parallel, but the 8th activity can only start after 5, 6, & 7 are all completed.



This order of activities might be prone to change and so to be able to enable our process to lend itself to this change we have used the “DIMACS” format to represent this dynamism. This format is widely accepted as the standard format for boolean formulas in CNF. To represent our process flows using this format we can say as

c process schedule start

p edge 9 8

e 1 5

e 2 8

e 3 6

e 4 7

e 5 8

e 6 8

e 7 8

e 8 9

c process schedule end

Note: each of the line in the dimacs template can start with a ‘c’, ‘p’ or an ‘e’. i.e.

* c: it is just to add a comment to add more information of the graph.
* p: it summarizes the total no of edges and vertices used in the entire order
* e: there is a line for each node where each line informs the starting and ending vertex of the edge.

#### Data Processing

The available data has been checked for completeness and suitableness for the analysis process. Some basic pre-processing steps like datatype conversion, derivation of new variables is done.

#### Order generation per day basis.

As the analysis has to be done to the day-level, the day level status of the active order & activity has been derived i.e. for each activity in the order “its Start Date, End Date, its Slack Days and a Flag to depict whether the activity was active on the given date.

### Technology Stack

To extract the targeted business insights we had shortlisted on the below tools as part of our technology Stack

* R: Backend Data preprocessing
* NoSQL: ElasticSearch as storage for the processed and “analysis-ready” data
* Shiny: To generate the customer Interface and Dashboard Application

### Assumptions

1. All activities might not be needed to complete an order. The process flow as documented above in the Data Generation Process -> Graphs section is what we have assumed for the order execution.
2. We have considered only the delivered projects/orders as the exact “current” status of the work in progress orders is not known.
3. From the estimated hours info provided for each activity only the hours part has been considered i.e. if available estimates is “166:12:45” only 166 hrs has been considered.