DIGITAL TRANSFORMATION IN LIFE SCIENCES

A Project Report

Submitted By:

Kesha Bagadia (AU1841011)
Yashvi Pipaliya (AU1841092)

in partial fulfilment for the award of the degree

Of

BACHELOR OF TECHNOLOGY

IN

INFORMATION AND COMMUNICATION TECHNOLOGY (ICT)

at

Ahmedabad University

School of Engineering and Applied Sciences (SEAS)
Ahmedabad, Gujarat

May, 2022

DECLARATION

We hereby declare that the project entitled "Digital Transformation in Life Sciences" submitted for the B. Tech. (ICT) degree is our original work and the project has not formed the basis for the award of any other degree, diploma, fellowship or any other similar titles.

Yashvi Pipaliya

Kesha Bagadia

Place: Ahmedabad

Date: May 05, 2022

CERTIFICATE

This is to certify that the project titled "Digital Transformation in Life Sciences" is the bona fide work carried out by Kesha Bagadia (AU1841011) and Yashvi Pipaliya (AU1841092), students of BTech (ICT) of School of Engineering and Applied Sciences at Ahmedabad University during the academic year 2021-2022, in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology (Information and Communication Technology) and that the project has not formed the basis for the award previously of any other degree, diploma, fellowship or any other similar title.

Anuraag

Signature of the Guide

Place: Ahmedabad

Date: May 10, 2022

ACKNOWLEDGEMENT

We would like to start by thanking the School of Engineering and Applied Sciences, Ahmedabad University and Eris Lifesciences Ltd. for allowing us this opportunity. It was an immense learning experience, working directly in the industry while receiving constant mentorship. Next, we would like to extend our gratitude to our internal project guide Prof. Anurag Lakhlani, Ahmedabad University, for ceaselessly assisting us in our academic endeavours and guidance in this project.

We would also like to thank the Data Science and Engineering team at Accenture for guiding us with the initial training and helping us with the project handover. We are equally grateful to the full-stack team of EPIC (Eris Pharma Intelligence Centre) for assisting us with the end vision of the project. Thanks are also due to the different teams at Eris who were patient in helping us with the non-technical context of the project. Finally, we would like to thank our friends and colleagues at Ahmedabad University and Eris for their support and camaraderie during our internship and academic journey.

Kesha Bagadia

Yashvi Pipaliya

ABSTRACT

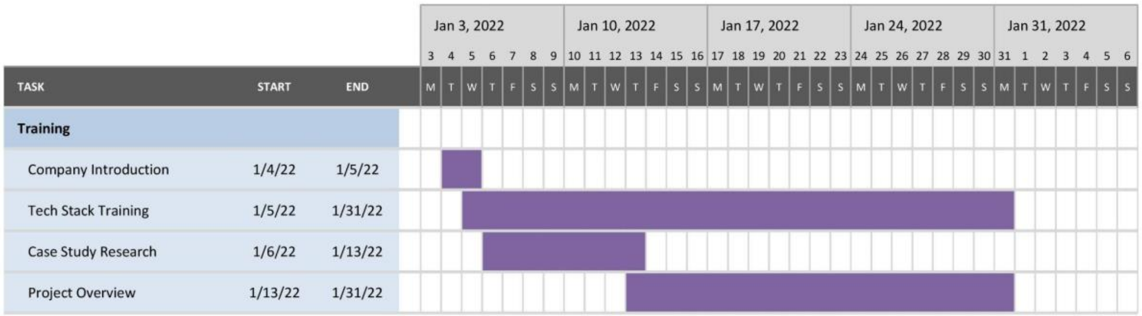
Amidst the people-powered, technology-assisted fourth industrial revolution, digital transformation is the new wave that every enterprise, big and small alike, is gradually getting swept by. With the constantly evolving virtual landscape and the ever-rising customer expectations in tow, it has become essential for businesses to be up for the challenge of changing with the market to continue to remain relevant, and Life Sciences is no exception to the rule. The pharmaceutical industry, in particular, holds immense untapped potential for a real resurgence in the manner that it approaches this development.

As crucial is the need to revamp the direct customer experience, it is also necessary to utilise these systems for better-informed internal data analytics. With this project, we intend to make use of the existing mammoth resource that is data and make it more accessible to draw actionable insights. The relevant data is extracted, transformed and loaded as per requirements and manipulated to visualise the decided Key Performance Indicators in a comprehensive format. This project was built in close collaboration with the target user departments and the process was cyclically validated and revised to ensure optimum user experience. The end goal is to have a system that extracts required raw data from its source and generates data visualisations and numerics that are periodically and automatically updated to work analysis upon. The purpose is to minimise repetitive labour and employ that human resource potential for more creative and critical-thinking oriented tasks.

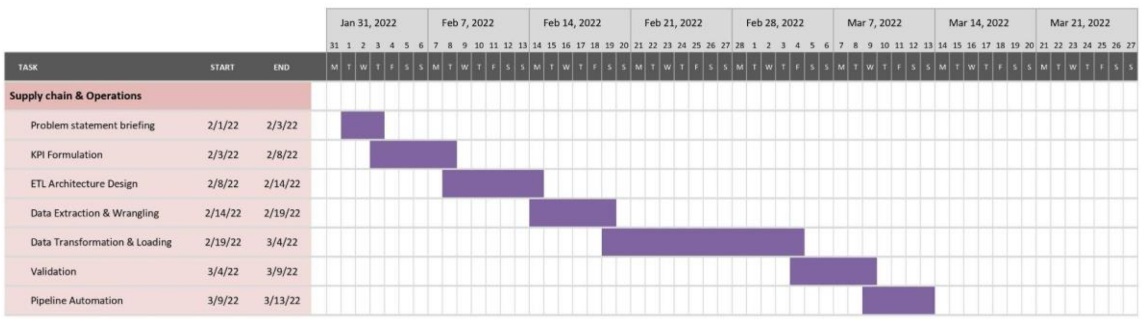
List of Figures

List of Tables

GANTT CHART



Training



Supply Chain & Operations

Finance

Figure 1 Gantt Chart

TABLE OF CONTENTS

1. INTRODUCTION

Digital transformation is the process of leveraging data-driven strategy and digital technologies to fundamentally enhance or modify a company's user perspective, business models, operating processes, and culture. Advancement in tools and services coupled with rising consumerism has made every field one can imagine privy to a fundamental overhaul. Amongst these is the Life Sciences industry.

Despite being in place since 2016, the fourth industrial revolution has seen an unprecedented and accelerated rise during the pandemic. Quite a bit of it has to do with the shift to remote work and the consequential spread of cloud services, but the prevalence of COVID19 also had a domain-specific impact on the pharmaceutical industry. The magnitude of data keeps increasing every day at an exponential rate; raw data generation is facilitated through a multitude of sources like the proliferation of personal medical devices, social media, and electronic medical records. This "big data" can conceivably transform the industry if the organisations can convert it into actionable intelligence. And as innovations continue to transform healthcare firms, pharmaceutical executives strive to align with this growth by investing in digital transformation aided advanced business analytics for every stage in the process.

1.1. Problem Definition

Possessing a substantial share of the global generic drug exports market, India has been an established global pharma player for the past few years. While the market is booming, the system fails to serve the throughput required to sustain the same invariably. The past few years have seen several pharmaceutical industries undergo sharp hikes, but not all have been able to maintain their positions without inconsistent dips. As enterprises venture headfirst into the anticipated growth trajectory, they fail to bridge strategic vision with operational reality. But the optimum approach to such a massive undertaking starts with a strong base and subsequent divided approach, as no amount of data surplus can ensure output without appropriate procedure.

The problem can be succinctly defined as possessing means with no end. That is, while the resource remains abundant yet massively underutilised, the difficulty is in tackling how to make the best use of it. For clarity in addressing multiple business problems, as many teams come together to run a thriving institution seamlessly, we have dealt with different departments as different problem statements. As the company we worked with is in its initial stages of digital transformation, the first problem is to clean, format and engineer the existing gigantic data source. Analytics and predictions can only be built upon a robust data engineering cycle, which is often the most head-wrenching task of all. Once regularly running and functional data pipelines are built to make the data accessible and readable, visual analytical widgets can be built over it.

Surrounding this upcoming transformation is the concern for a functional and scalable infrastructure, as the traditional systems are not equipped to house such copious amounts of data or host expansive analytics. Datasets no longer reside within a single central server, and neither can the rigid spreadsheet visualisations suffice, which is why there is such a surge in demand for data products that can process and store billions of bytes of real-time data with hundreds of thousands of transactions per second.

1.2. Project Overview

We undertook a part of the digital transformation project while interning for Eris Lifesciences Ltd. The name of the data portal we worked on is EPIC, short for Eris Pharma Intelligence Center. It's a collective dashboard that houses applications like customisable analytical reports, live data visualisations, forecasts and summaries. With this project, we intend to remodel every department that touches data by embedding modularity and automation so as to extricate manual operations wherever possible. The goal is to collaboratively work with the end-user teams to engineer data and provide thorough analytics plus visualisations to assist the development of insights with practical calls to action. The ETL pipelines once built are automated and triggered at their set frequency, so they fetch the data they require, transform it into desirable form and dump it onto the data warehouse where it can be accessed to derive Key Performance Indicators. During our time with the company, we undertook Supply Chain and Operations and Finance departments for the same.

1.2.1. Supply Chain and Operations

While pharmaceutical companies fearlessly dive into aspirational goals like low-cost production and market expansion, they face the unavoidable pressure to tighten their end-to-end operations. The cross-interaction of an array of factors; how GST influences the distribution footprint, quality and pricing issues, the simultaneous hike in product proliferation and users' therapeutic needs; sheds light on how there is a desperate need to primarily reexamine the existing supply chain before developing analytics.

To thoroughly investigate supply chain and operations would include extraction and evaluation of all data, including transportation costs and efficiency, manufacturing capital requirements and inventory levels. Forecasting supply and demand, assessing waste management, optimising logistics are a few direct analytical applications to name.

1.2.2. Finance

While the pharmaceutical industry continues to grow, very few companies invest in technology for finance to the same level as their user-facing functionalities. The continual challenges related to expenses, regulations and quality assurance garner the spotlight. Hence there is a need to invest in the appropriate technologies, skills, and processes for finance to play a more strategic role while balancing profitability and risks. With ever-evolving customer expectations and technologies, a performance-driven finance function can be exceedingly beneficial for a digitally transforming enterprise.

Some of the tasks that we cover with our project are: automating the existing systems to extricate the possibility of human error and saving human resources for more notable and creative tasks, assessing budgets and expenditures, building scalable cash flows from balance sheets and so on.

1.3. Hardware Specifications

Processor	2.0 GHz
RAM	4 GB
HDD	40 GB

1.4. Software Specifications

Operating System	ANY
Programming Language	-
Other tools and tech	Apache Spark, Azure Cloud, MariaDB

2. LITERATURE SURVEY

2.1. Existing System

Only two decades ago, the flow of information in the pharma industry was relatively simplistic, and the applications of modern technology were limited[1]. Though now as we move forward into a further integrated society, where technology cannot be extricated from the business procedure, the system of information transfer has grown more convoluted.

Lately, the worldwide medical industry, esteemed at US$9.59 trillion (PwC Report, 2015), has gone through major developments in every section of its functioning[2]. These changes are inclusive of the humongous adoption of electronic wellbeing records (EHR) by governments, medical clinics and doctors to finish the digitalisation of research datasets and billions of patient records belonging to pharmaceutical companies.[3]

There is an industry-wide focus on healthcare compliance as per strict and binding regulations alongside matching the customer expectations for affordable healthcare through innovation in business models. If there is one factor that unifies all these healthcare trends, it is "big data".[4] The industry faces vast gaps in converting unstructured information bytes into meaningful business intelligence. Healthcare data analytics, a market expected to grow to US$18.7 Billion by 2020, sits at the heart of various transformations in the industry[5].

The pharmaceutical sector in India is expected to grow at a CAGR (Compound Annual Growth Rate) of 12 per cent, from $41.7 billion in 2020 to $130 billion by 2030[6]. Pharmaceutical executives require ways to improve efficiency, uncover new business opportunities, and build better relationships with patients and prescribers to align with this growth. The pharmaceutical and life sciences industry is flooding with data, which translates to untapped potential.

Most pharmaceutical firms use a traditional approach for data analysis which often involves manual work and interpretation of data that is slow, expensive and highly subjective[7] (Fayyad, Piatsky Shapiro and Smyth, 1996). Most healthcare institutions lack the appropriate information systems to produce reliable reports with respect to other information than purely financial and volume related statements (Prins & Stegwee, 2000).

2.2. Proposed System

Digital transformation holds tremendous potential to help pharma companies address these challenges. In all industries, companies are improving operations through a set of emerging technologies collectively known as Industry 4.0[8]. Specifically, that term refers to new tools and processes that enable intelligent, decentralised production via intelligent factories, integrated IT systems, the Internet of Things, and flexible, highly integrated manufacturing systems. The process would include:

- integrating and managing the data from various sources,
- further converting the unstructured data into structured information by cleaning,
- transforming and loading it into the appropriate data warehouse.

As the data would be formatted, various visualisation and analytics could be built upon it based on the decided KPIs (Key Performance Indicators).

This resultant analytics can conceivably transform the supply chain operations, by improving the quality of the processes at the core and providing a boost to the overall productivity. The real-time data visibility supports a quicker pace in both decision making and end-to-end supply chain integration. It also additionally assists network scalability and greater production efficiency, improved operational processes and maintenance and much more[9].

Similarly, as the core-driver department of any enterprise, the business analytical digital transformation of finance can provide any institution with an established groundwork for growth. The finance functions being the gatekeepers of all data pivotal to the enterprise, is responsible for forecast generation and strategic planning. This section is where the data for sales, order fulfilment, supply chains, customer demand, regulatory requirement and business performance as well as managing real-time data.[11]

A system with the ability to manage and integrate data generated at all stages and sectors of a business will unleash the industry to a sea of untapped benefits from ever-rising technology trends.

2.3. Feasibility Study

2.3.1. Financial Feasibility

The majority of the project: the data lake, the pipeline runs, the error handling, the linked functions, etc., is hosted on Azure services. Being one of the leading cloud infrastructures today, it can be expected to be expensive. Nevertheless, Azure proves to be economical with its moderate pricing plans and the scalability it provides in terms of add-on licences and agreements.[12] Even so, to cope with the high costs associated with data transfers, we have hosted our data warehouse using the MariaDB server, which is open source and compensates for a possible chunk of expense.

2.3.2. Technical Feasibility

EPIC packs an expansive tech stack, which is to say, it has been curated with utmost thought and consideration to its purpose and requirement. The main tools and services are provided by Azure, which is a budding cloud infrastructure with a variety of resources for the users, including scalable storage for structured and unstructured data, logic apps, alerts etc. It has a 99.995% uptime rate which makes it reliable for the data processing work of a pharmaceutical company with enormous data resources.[13] MariaDB is also equipped to support extensive data operations and has active user support. Overall, the technical feasibility has been successfully factored in the planning of the project.

2.3.3. Risk Feasibility

A part of the project is data collection, structured and unstructured alike, and to format and clean it onto a data warehouse, from where it will be further utilised for insight generation. As the end product will be used by more than 3000 of the company employees, the scale of the project and the traffic on the server are hard to estimate. However, as Azure is a scalable service, the risk associated is not very high. Data leak risks have also been secured by an establishment of a tight virtual private network for access to data, and there is an adequate security company login for access to the EPIC Azure portal. All the technologies utilised are well-received in the market and relevant to the current technical landscape, so stability, process issue and development environment risks are also minimal.

2.3.4. Social/Legal Feasibility

All the data used for analytics for the project is currently ethically sourced by a first party. Furthermore, we intend to continue working on a self-sustained data generation system so as not to cause any legal gaps.

3. SYSTEM ANALYSIS & DESIGN

3.1. Requirement Specification

3.1.1. Purpose

In today's information-driven industrial landscape, the old product-oriented business model is being challenged by patent expirations, increasing customer demand, rising competition, and rising pricing pressures. In the wake of COVID19, the digital transformation of the healthcare industry is more relevant than ever in allowing firms in all sectors to improve performance through better manufacturing productivity, stronger competitive skills, more accurate planning and forecasting, and financial sustainability. And digitalisation with a scalable and dynamic approach is vital for building a robust model that can allow unrestrained horizontal and vertical growth.

The process of digitalisation includes increased use of robotics, automated solutions, and computerisation. This would thereby allow the business to reduce costs, improve efficiency and productivity, and be flexible to changes. This form of pro-action in terms of keeping up with the general market trends and having good foresight can help the company gain an immense edge over its competitors.

3.1.2. Scope

EPIC is the company's first step toward automating their everyday tasks and enhancing their business. This platform is exhaustive of several sub-initiatives like:

- streamlining and automating various internal processes,
- bringing in new client-facing functionalities, like speech-to-text prescriptions through AWS service,
- collecting data obtained from the EGC and health reports of various patients to detect trends, which could say, help patients diagnose heart illness at an early stage,
- developing a one-click solution for MRs to view and update their daily data etc.

Many more project POCs are yet to be discussed and implemented. Currently, we are working on inter and intradepartmental interfaces and pipelines for business intelligence and analytics while other projects are conducted in parallel.

3.1.3. Intended Users

In the current scenario, there are over 3000 employees in the company. And as data is relevant in every sector and team in the enterprise, all employees, from the MRs to the CEO, are directly or indirectly users of the project. These users will constantly or periodically need to view various components of the system as per their requirements for their scheduled tasks. The users are bifurcated as per their departments, and the system is designed to show them only that respective data. The list of the current users is as below:

- CEO

- COO
- CIO
- Sales Department
- Marketing agents
- HR Department
- Finance Department
- Operations Department

3.2. Functional Requirements

3.2.1. Extract Data

Data needs to be extracted from its respective source to the data lake for further processing.

3.2.2. Reformat Data

The format of the data files, like .csv or .xlsx should be checked and converted accordingly for further processing.

3.2.3. Transform Data

Data should be transformed with the required logic to give the required outputs.

3.2.4. Establish server connection

The cloud servers of the data lake and data warehouse should be connected to allow data transfer.

3.2.5. Load Data

Data should be dumped onto the data warehouse after transformation for further access.

3.2.6. Automate pipeline

The built ETL pipeline should be automated for periodic reruns.

3.2.7. Failure notification

If any pipeline fails, there should be a timely notification sent to the respective team member.

3.2.8. Clear cache

The pipeline should clear the cache once it's done with its ETL procedure.

3.2.9. Error handling

There should be a backup procedure for errors or failures so that there is no ripple effect and so that the company doesn't incur any heavy loss due to one blip.

3.3. Non-Functional Requirements

3.3.1. Availability

- Updated data should be available timely to its target users to avoid delays and piled up tasks.
- Pipeline triggers should be set to correct frequency in accordance with the data and its usage.

3.3.2. Security

- Data should be stored in a virtual private network server.
- Limited data should be accessible with abstraction to the ones who only require the end analytics.
- Data transfer across multiple servers should be set in a secure and encrypted manner.

3.3.3. Consistency

- Data should be stored with consistent naming conventions across FTP, data lake and data warehouse.
- Data flows and pipelines should be named uniformly.
- The processing of data should follow a set and defined order, like cleaning followed by transformation before loading.
- Columns should be appropriately named across the datasets because there will be overlaps.

3.3.4. Dynamicity

- Pipelines should be built to fetch and dump data dynamically based on date-based or other feature-based requirements.
- Folder paths should be defined using variables to allow dynamic updates and access.

3.3.5. Scalability

- Pipelines and dataflows should be built with modularity and scalability in mind; they should be able to assist more datasets if needed.
- Data flows should be set such that they allow schema drift and aren't affected by the restructuring of datasets.

3.3.6. Legal and Licensing Issues

- Legality of the cloud and service products should be airtight
- Data should be ethically sourced and not be in conflict with any privacy regulations.

3.3.7. Maintainability

- Pipelines should be easy to edit and rectify.

- Quick manual processing options should be available in case of technical issues.

3.3.8. Data Integrity

- Data should be extracted from the correct source.
- Data should be dumped to the correct location.

3.3.9. Quality

- Data should be cleaned and filtered with care to detail to ensure good quality.
- The pipelines should be optimised and neat.

3.3.10. Extensibility

- Pipelines should be easy to update and build further.
- Schema should be convenient to update.
- Dataset unions and joins should be seamless to add.

3.3.11. Efficiency

- Optimal use of debug and activity run resources should be done.
- More tasks should be fulfilled in the least possible time.

3.4. System Flow

First, the business problem is brainstormed and ideated upon alongside the respective department. The required Key Performance Indicators or KPIs are decided upon and their logic is fleshed out. Next step is the sourcing of data. The required format and frequency of data input is then decided. Now, data is to be fetched from the FTP, where the necessary data will be cyclically transferred.

Once the data is fetched, it is converted to the required file format and stored in the data lake. From the data lake, the data is then cleaned and wrangled. After that, the transformation logic is applied. This is only one step in the process but it's a hefty one. As each transformation is different, there's always room for new mistakes. Once a naive transformation logic is applied, the next stage is optimisation of the logic. After that comes loading the data onto the data warehouse.

The ETL pipeline is then run for the recent and historic datasets. Soon after, there is again a meeting with the respective department for the next stage of data validation. Once the output data is validated, the pipeline is automated and triggered in accordance with the required frequency of the data output required. As soon as the data pipeline is set, the next stage is the development of KPIs and/or derived KPIs. After that is done, the data alongside the KPIs is pushed to the full stack team for the development of the visual widgets.

The flowchart of the explained system is as below.

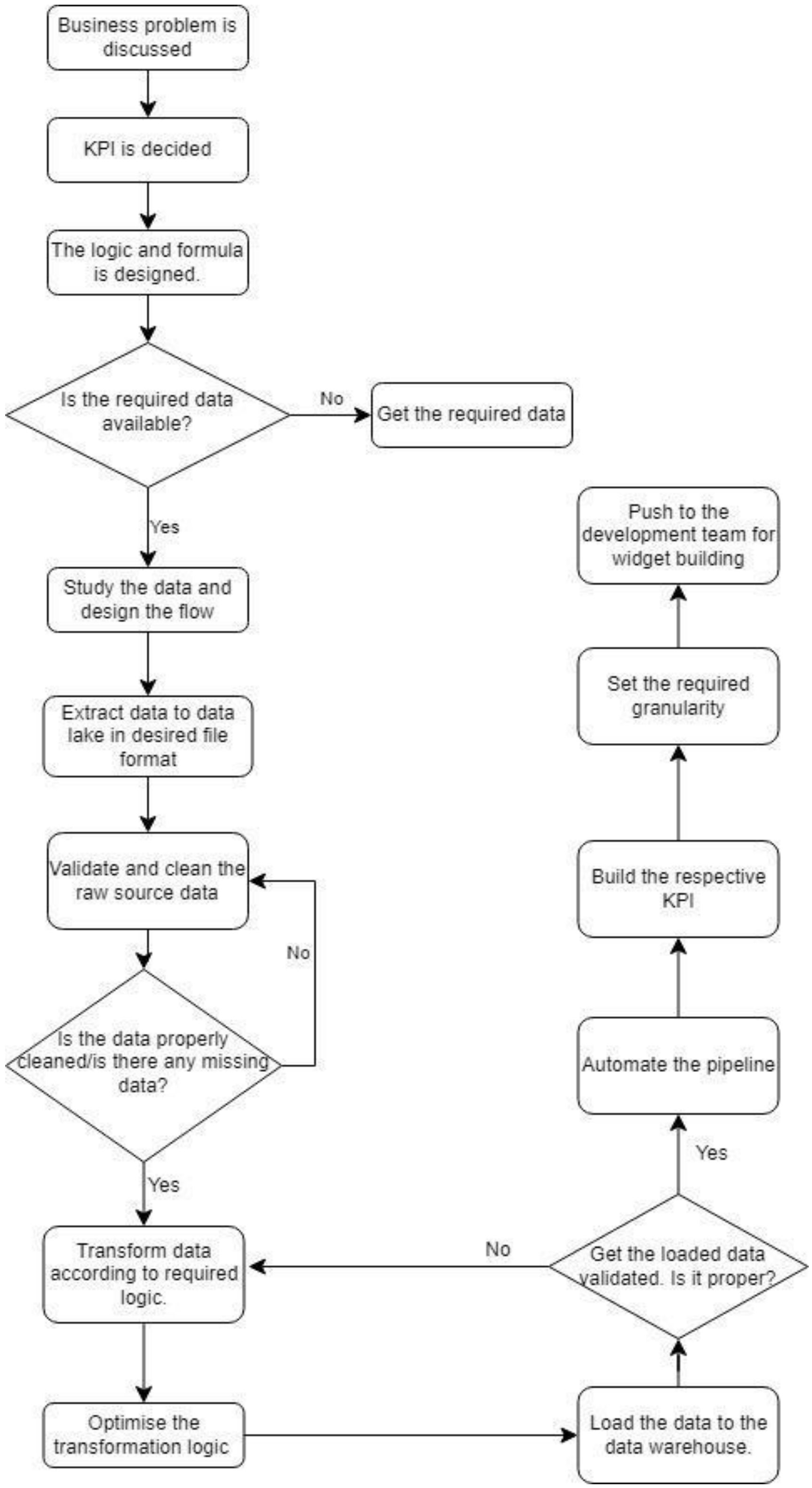


Figure 2 System Flow Chart

3.5. Design and Test Criteria

3.5.1. Design Considerations

3.5.1.1. Assumptions

There are a set of assumptions that have been made in order to efficiently build the system:

- It is assumed that the built KPIs have been successfully translated to visual widgets by the development team.
- It is assumed that the end users have the necessary knowledge and context to consume the data and build analysis upon it.
- It is assumed that the required data is available.
- It is assumed that the received data is accurately and ethically sourced as per agreement.
- It is assumed that the functional and business requirements that are provided to us by the respective stakeholders are logically sound.
- It is assumed that the source data will be periodically updated and will be in the predefined format for automated pipelines.

3.5.1.2. Constraints

While Azure has a 99.5% uptime rate, it is possible that the server is occasionally under internal maintenance and in a non-functioning state. At that time, all the data pipelines would be halted as our entire system has been hosted on the cloud.

3.5.1.3. System Environment

The data warehouse, which is the final stage of the ETL pipelines and the immediate source of data to be pushed onto KPIs and widgets, can only be accessed through the company's tightly secure VPN. Other than that, any computer device with the general specifications can easily access the system as it is entirely hosted on cloud.

3.5.2. Design Process

For any analytical widget that is built, the first step is the vision of the last step. The end goal, that is the final KPI to be developed and its subsequent logic is decided before working the process of how to achieve it.

The transformation process starts with the composition of a practically achievable formula or implementation from the discussed logic and business requirements. For example, for a KPI called Ageing of Advances, the resulting data from the difference of multiple organisations' recorded provision amounts and outstanding balances is to be provided. There is an additional grouping logic and date-wise priority to be allotted. So the technical implementation would be something like:

Provision(x) = max(Provision(x-1) - Outstanding(x-1),0)

agingOfAdvances(x) = max(Outstanding(x) - Provision(x), 0)

After the formula is achieved, an attempt on a naive query logic is made on the basis of the form of data available and the equation derived. Like if we follow up on the previous example, a part of the SQL pseudocode would follow the following steps:

- Apply left join on AmountOutstanding and Provision file on Customer Name
- Select columns CustomerName, AmountOustanding, AccountDate, Closing debit, provision and group
- Check if the provision is Yes then aggregate the columns according to Account Date and Customer Name
- Create temporary variable equal to Closing debit and order the data by AccountDate
- Take the temporary variable and subtract it from AmountOutstanding and save the remaining value in the temporary variable and iterate it until the variable is not zero.

Once a pseudocode or an SQL query is formed, it is optimised for neatness, ease of understanding, and efficient use of resources. Again, following on from the previous example, the optimisation of the section on First In First Out iterative difference logic can be cleanly solved by deriving cumulative summation of the outstanding amount ordered by the date. A part of the query logic code would look something like this:

```
select
    GL_Group,
    Customer_Name,
    Customer_Code,
    City,
    Division,
    Bill_Number,
    Company,
```

```
        Branch,
        Cluster,
        Account_Date,
        Amount_Outstanding,
        Closing debit,
        Provision,
        'group',

      sum(Amount_Outstanding ) over (
      partition by Customer_Name
      order by
        Account_Date
      ) cumulative_billed_amount

from
        erisdenorm_prod.test_amount;
```

Now, we only have to subtract the cumulative billed amount from the Amount_Outstanding, which would provide us the required amount.

```
select
        GL_Group,
        Customer_Name,
        Customer_Code,
        City,
```

```
        Division,
        Bill_Number,
        Company,
        Branch,
        Cluster,
        Account_Date,
        Amount_Outstanding,
        Closing debit,
        Provision,
        'group'
         cumulative_billed_amount ,



       Amount_Outstanding-cumulative_billed_amount as
       balance_amount
from
        erisdenorm_prod.test_amount;
```

After a robust database query logic is formulated, the next step is its translation to Azure Data Factory's data flow. There are usually two dataflows in place for a single KPI, one is the sanitisation: which includes reformatting and renaming of columns to their respective data types and commonly and shortlisting of the relevant data for the transformation, and the other is the transformation: which consists the central logic application

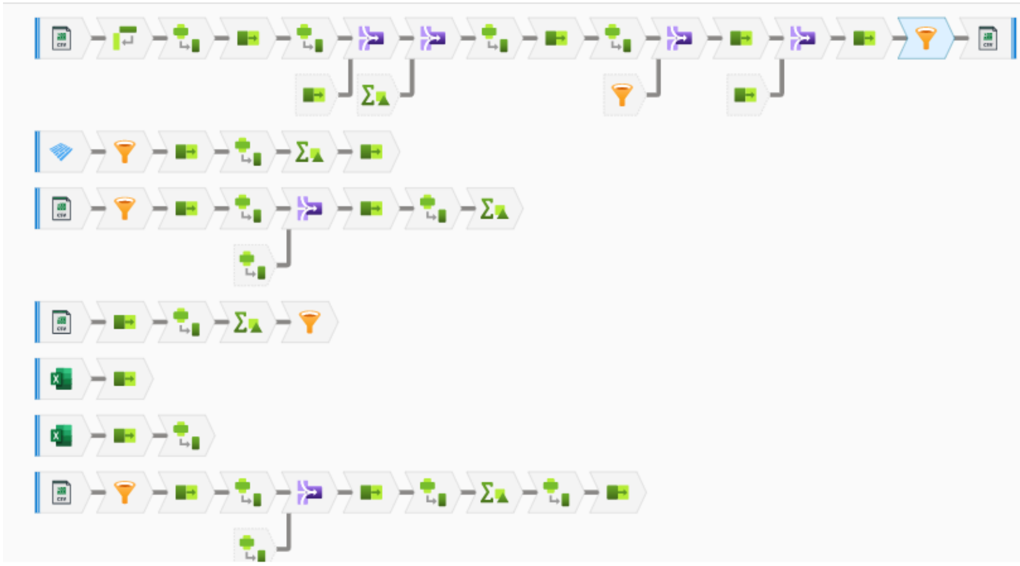


Figure 3 Dataflow Example (Forecast Accuracy)

The next stage would be the optimisation of the pipeline from an Azure service perspective, this would mean picking the options that make the optimal use of the cloud resource and its tools specifically. Some of the tips that we generally keep in mind as protocol are:

- Only select the relevant columns prior to data processing and for data dumping to avoid unnecessary overhead.
- Using wildcard paths for dynamic logic and merge file requirements to avoid multiple copy and data flow activities.
- Reusing dataflows and datasets for data with same schema and processing requirements.
- Perform indexing to bolster search and processing speed.
- Optimise partitioning of files for data sinks where a single partition isn't of importance to avoid high kurtosis.
- Use control flow for parallel processing.

- Avoid using cross joins.
- Choosing optimised partitioning by default for mid-process datasets for parallel processing of data.
- Picking only the required lowest number of rows for every respective dataset in data debug mode.

In some cases where the logic is too complex for ADF's rigid UI to manipulate, an additional view is created upon the transformed data from Azure using MariaDB. An SQL script is written to complete additional transformations according to the logic provided.

An example of view created to obtain Forecast Accuracy:

```
Create or replace
    algorithm = UNDEFINED view
    `erisdenorm_prod`.`denorm_forecast_accuracy_ma
    x_date` as
    select
        `dfa`.`pk_ID` as `pk_ID`,
        `dfa`.`granularity` as `granularity`,
        `dfa`.`rowID` as `rowID`,
        `dfa`.`formatting` as `formatting`,
        `dfa`.`createdDate` as `createdDate`,
        `dfa`.`modifiedDate` as `modifiedDate`,
        `dfa`.`createdBy` as `createdBy`,
        `dfa`.`modifiedBy` as `modifiedBy`,
        `dfa`.`createdIP` as `createdIP`,
        `dfa`.`modifiedIP` as `modifiedIP`,
        `dfa`.`Cluster` as `Cluster`,
        `dfa`.`Revised_SKU` as `Revised_SKU`,
        `dfa`.`Vender_code` as `Vender_code`,
```

```
    `dfa`.`SKU_Name` as `SKU_Name`,
    `dfa`.`Division` as `Division`,
    `dfa`.`SKU_Category` as `SKU_Category`,
    `dfa`.`Status` as `Status`,
    `dfa`.`Date` as `Date`,
    `dfa`.`Quantity` as `Quantity`,
    `dfa`.`As_On_Date` as `As_On_Date`,
    `dfa`.`PTS_Rate` as `PTS_Rate`,
    `dfa`.`PrimarySalesValue` as
`PrimarySalesValue`,
    `dfa`.`ExpiryReturnValue` as
`ExpiryReturnValue`,
    `dfa`.`ForcastValue` as `ForcastValue`,
    `dfa`.`Over_Under_Forecast` as
`Over_Under_Forecast`,
    `dfa`.`TargetQuantity` as
`TargetQuantity`,
    `dfa`.`TargetValue` as `TargetValue`
from
    ((
    select

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
pk_ID` as `pk_ID`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
granularity` as `granularity`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
rowID` as `rowID`,
```

```
`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
formatting` as `formatting`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
createdDate` as `createdDate`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
modifiedDate` as `modifiedDate`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
createdBy` as `createdBy`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
modifiedBy` as `modifiedBy`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
createdIP` as `createdIP`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
modifiedIP` as `modifiedIP`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
Cluster` as `Cluster`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
Revised_SKU` as `Revised_SKU`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
Vender_code` as `Vender_code`,
```

```
`erisdenorm_prod`.`denorm_forcast_accuracy7`.`SKU_Name` as `SKU_Name`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`Division` as `Division`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`SKU_Category` as `SKU_Category`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`Status` as `Status`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`Date` as `Date`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`Quantity` as `Quantity`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`As_On_Date` as `As_On_Date`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`PTS_Rate` as `PTS_Rate`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`PrimarySalesValue` as `PrimarySalesValue`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`ExpiryReturnValue` as `ExpiryReturnValue`,
```

```
`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
ForcastValue` as `ForcastValue`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
Over_Under_Forecast` as `Over_Under_Forecast`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
TargetQuantity` as `TargetQuantity`,

`erisdenorm_prod`.`denorm_forcast_accuracy7`.`
TargetValue` as `TargetValue`
    from

`erisdenorm_prod`.`denorm_forcast_accuracy7`)
`dfa`
join (
    select
        `dfa`.`Date` as `Date`,
        max(`dfa`.`As_On_Date`) as
`As_On_Date`
    from

`erisdenorm_prod`.`denorm_forcast_accuracy7`
`dfa`
    group by
        `dfa`.`Date`) `max_date` on
    (`dfa`.`As_On_Date` =
`max_date`.`As_On_Date`
        and `dfa`.`Date` = `max_date`.`Date`))
```

```
where
    `dfa`.`Date` + interval 1 month <=
curdate();
```

3.5.3. System Architecture

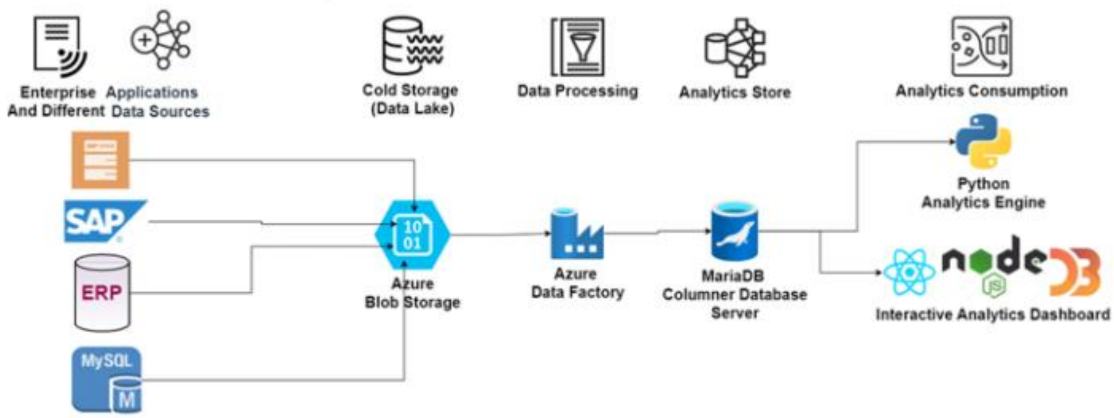


Figure 4 System Flow Chart

3.5.3.1. Data Sources

Eris is divided into multiple departments mainly. There are six functional departments, and they use different cloud-based storage service providers to save, collect, and organise their daily data. The marketing department and sales department use the ERP database. In contrast, the HR department uses Darwinbox, SAP is used by the finance department, and the manufacturing department uses one pharma cloud as their primary database and datastore. All the data from different platforms are first collected on a single platform by API calls or by respective vendors to analyse further.

3.5.3.2. Azure Blob Storage

Azure Blob Storage is the storage solution provision of the Azure cloud infrastructure. It's an optimised storage for massive amounts of multiformat datasets. We utilise this as our centralised data lake, wherein data from our multiple sources is gathered before further manipulation and transformation for KPI development and other analytical usage.

3.5.3.3. Azure Data Factory

Azure Data Factory, or ADF is the ETL service provided by the Azure cloud infrastructure and is built over Spark Scala. It's an intuitive user interface that provides a variety of activities and functionalities which allow us to build the necessary data pipelines.

Here is where the pivotal data processing of the project is done, and it's done through pipelines. Copy activity can conduct the extraction and loading of data to and from various resources like FTP and MariaDB via different linked services, whereas the cleaning, transformation and manipulation part of data processing is performed by the Data Flow activity.

Dynamic automation and triggers can be set for periodic pipeline runs and alerts for errors can also be set up. To summarise, ADF alone handles the creation, hosting, scheduling, debugging and monitoring phase of the ETL pipeline of any large or small scale data project.

3.5.3.3.1. Data pipeline flow

This is the systematic flow of building a basic pipeline in Azure Data Factory once the dataset is in the data lake.

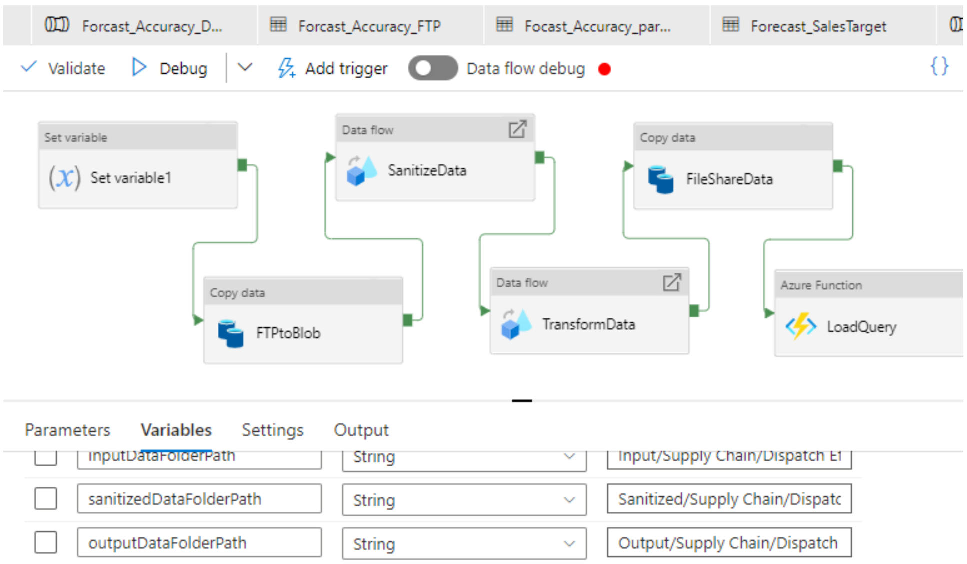


Figure 5 ADF Pipeline

The first step for any data pipeline is defining the datasets. There will be multiple datasets for each process. Starting with the raw data file, to the blob stored data file, to the sanitised data parquet, to the transformed data file, to the output file, all datasets are to be separately defined.

The schemas have to be either imported from connection or set, depending on whether the source is preset or dynamically fetched, but the flexibility is provided. The correct source index has to be set in case it is an excel workbook. Wildcard paths have to be set in case it is a merge file functionality.

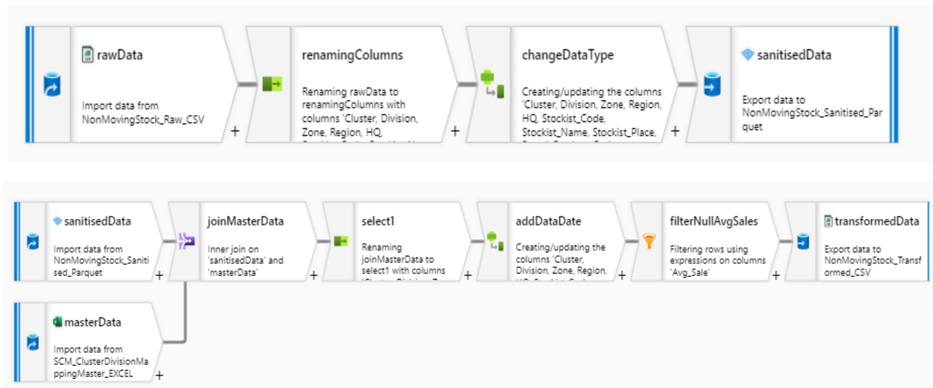


Figure 6 Simple Dataflow

The next comes data flow. This is where data is sanitised and transformed. Data flows include a wide range of functionalities, almost exhaustive of all SQL features: there is join, there is union, there is window, there is aggregation, there are filters, and the list keeps going. Even within these function activities there is an expression building which gives one the access to multiple predefined functions for quick data manipulation or calculations.

Data flows also have flowlets which allow reusability and modularity in processes. There are branches to allow different trajectories for the same mid-resultant data.

Data debug mode allows the user to preview data, view statistics of different columns of the data and export the limited data preview. One can also expand the limit of the data preview to have a better grip of the state of the process. Data flows do most of the heavy lifting of the pipeline as they hold the central transformation logic of the pipeline, so they are very vital to the system.

A data pipeline is where everything comes together. There are a variety of functions, it offers:

- Copy Data: transfers data from a source to a sink.
- Data Flow: Embeds a data flow activity and runs it.
- Set variable: Allows to dynamically define a value for any variable.
- Azure Function: Gives the chance to run any Azure logic function. Usually used to load data in our case.
- For Each: Provides the functionality to iterate a pipeline multiple times

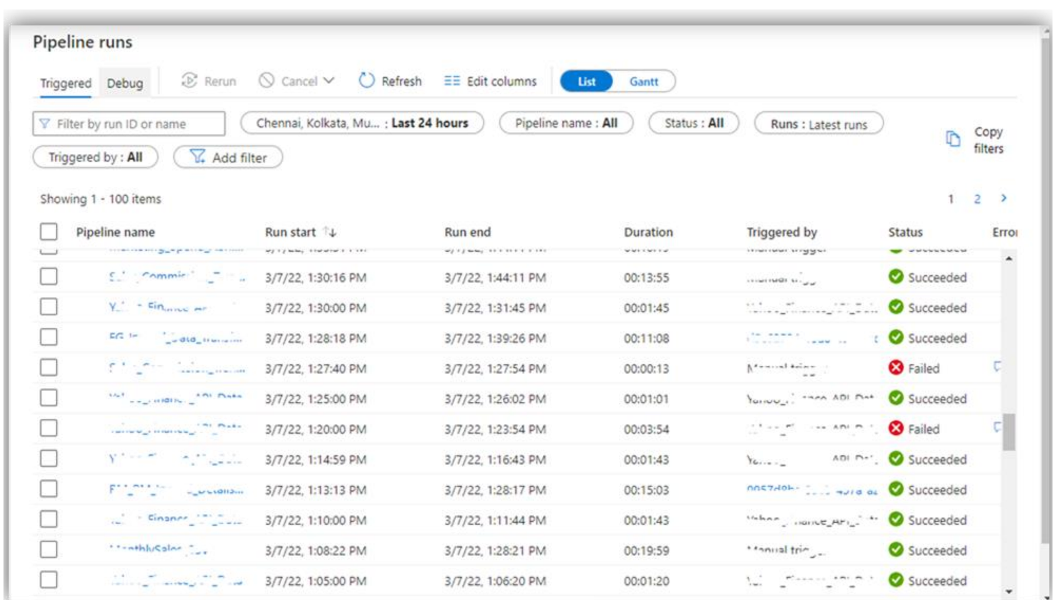


Figure 7 Scheduling a Pipeline

Data pipelines can be automated by setting dynamic variables for folder paths of all relevant datasets and then setting a trigger. A trigger can be run at the business needs' desired frequency.

3.5.3.4. MariaDB Server

MariaDB is an open source database server that we utilise as our data warehouse. We use a columnar database for our purpose as the data is usually in the form of OLAP cubes, and also because it's a cost and time effective alternative. As they generally consist of a large number of columns, those tables are typically the denormalised form of their relational data structure.

The data table design here would be as per the final KPI logic and analysis requirements.

3.5.3.5. Analytics Consumption Tools

Multiple systems are developed with different purposes to view the processed and transformed data. These systems are built on node.js and react using d3.js which displays all the KPIs, numerical report, and figures which are previously defined by the business analyst and further designed by the frontend developers.

The three main systems which are used by the employees and company head are:

- Hawkeye: it's the collective dashboard for numerous greatly insightful data visualisation widgets.

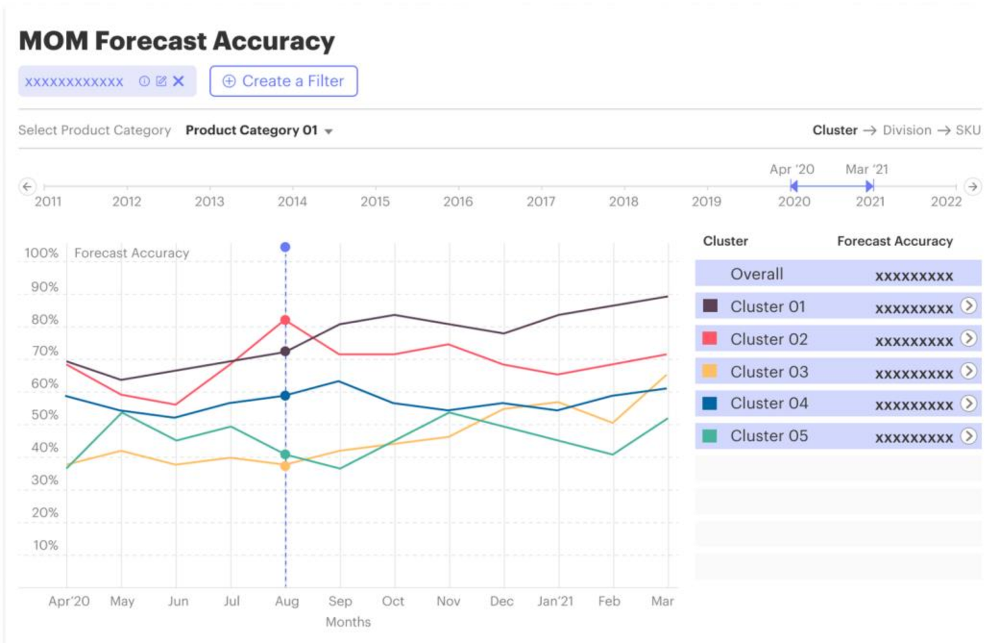


Figure 8 Hawkeye preview

- Monitor First: Gives quantitative summary of required KPIs.

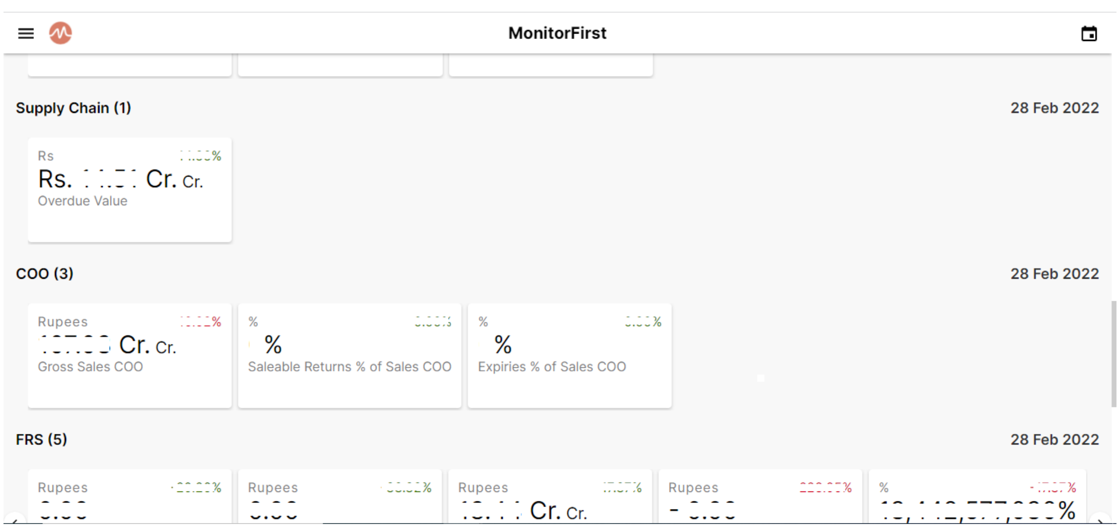


Figure 9 Monitor First preview

- Sherlock: Here users can look into all data, current and historical, to create customised analytical reports.

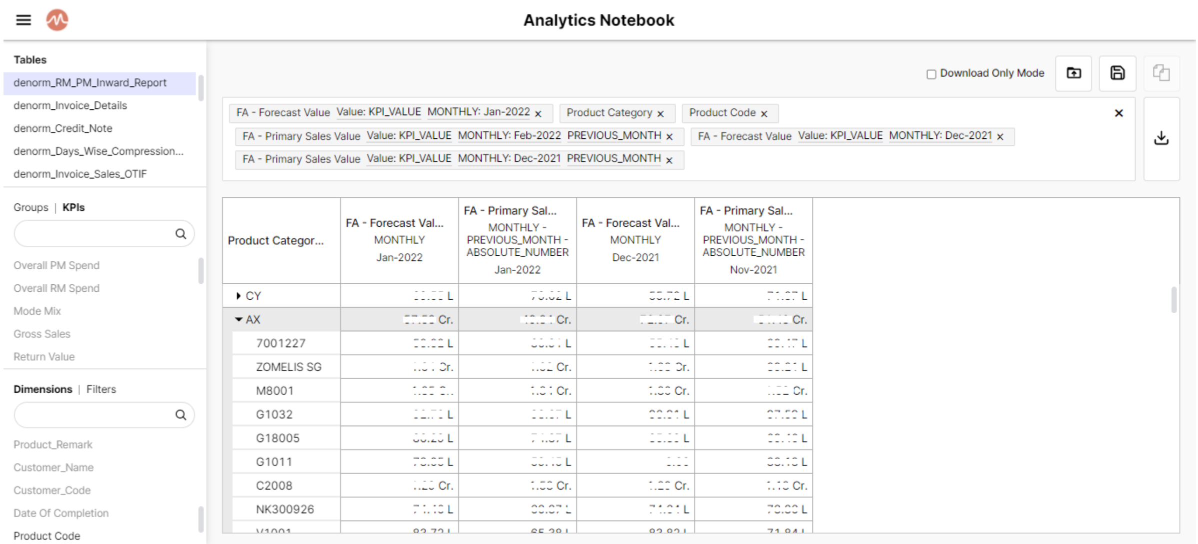


Figure 10 Sherlock preview

3.5.4. Testing and data validation

As data is all related to numbers, and as it is human to err, testing at every stage of data processing is crucial for accurate and valid analytics. There are different ways to check the tightness of the data pipeline and validate the final output in respect to different KPI logics, but some general testing and validation procedures are as follows:

- Checking for any data loss after file format conversion like .xlsx to .csv
- Validating the number of rows and columns in the source data set with the respective department business analyst.
- Validating the number of rows and columns in the source data set once transferred to the data lake.

- Cross-checking the index number or page name for excel workbooks.
- Checking the data preview of the Azure data lake stored dataset.
- Testing the data preview at every stage of data transformation.
- Applying filters to ADF dataflows to test specific data due to row number constraints.
- Utilising ADF column wise data preview statistics for cross-verification.
- Sending the output file to the respective department lead for data validation.
- Randomly selecting row from input file and check if its present in the output file.
- Calculating complex formulas on excel and cross-checking the results using pivots and vlookup.

4. RESULTS

Once the pipeline loads the data onto the data warehouse and once the KPI is coded and inserted into the database, a visual analytical widget is developed over it. These are the final widgets, that is, the outcomes of all the data processing. These widgets come with customisable granularities, dimensions and filters.

4.1. Supply Chain and Operations

As the supply chain and the logistics around it is where the actual legwork of the sales and distribution happens, the visual analytics of the same can be very illuminating in weeding out flaws with the system and gaining assistance working on possible sprouts of growth and development.

4.1.1. Production Efficiency

A study of the operational practices of more than 25 global pharma production firms concluded that the top ones are more than twice as productive as their average counterparts [11], which is a marker of how important a factor it is. Production efficiency is a collective metric of the below mentioned parameters:

- Availability takes into account Availability Loss, which includes any events that stop planned production for an appreciable length of time.
- Performance considers Performance Loss, which represents anything that causes the assembling system to run at not exactly the most extreme conceivable speed when it is running (counting both Slow Cycles and Small Stops).
- Quality takes into account Quality Loss, which accounts for manufactured parts that do not meet quality standards.

- Production yield is a metric derived from the division of the number of good parts produced by the total number of parts started in production.
- Capacity Utilisation represents the production or manufacturing capabilities that the company utilises to generate results. It's the ratio of the potential output to the actual realised output and it gives insight into a company's operational efficiency.
- OEE (Overall Equipment Effectiveness) identifies the percentage of manufacturing time that is truly productive. It is a mix of three factors that let you know how proficient a resource is during the assembling system: resource accessibility, resource execution, and creation quality.

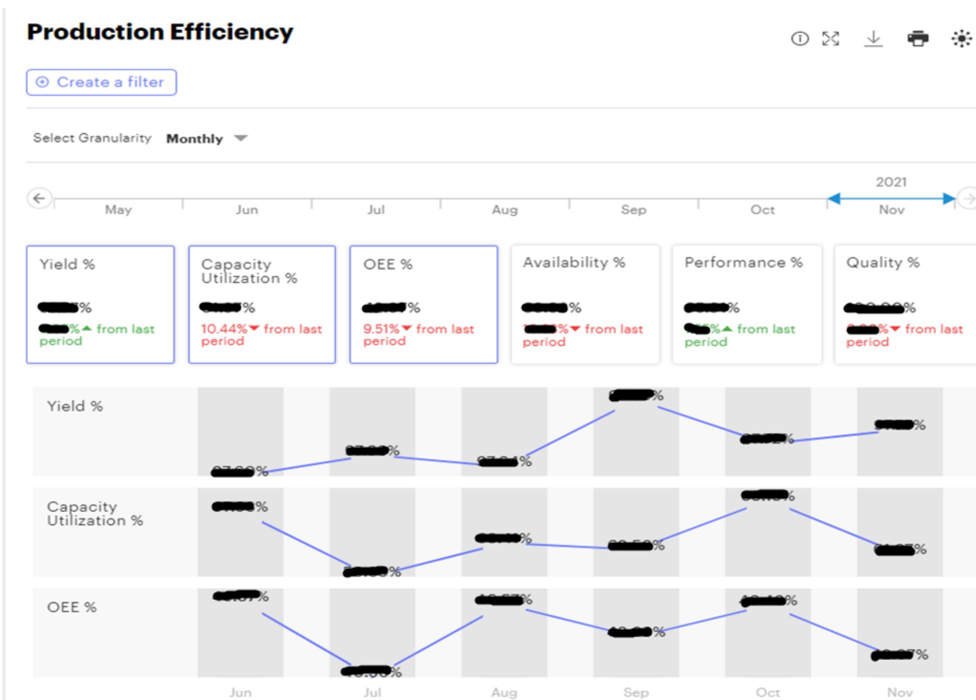


Figure 11 Production Efficiency

4.1.2. Distribution Cost and Distribution Timelines:

Distribution is the most challenging and crucial part of the vast supply chain and it’s pallbearer for the smooth functioning of a business. So it's necessary to assess the cost and the on-time-in-full efficiency of different media of transports and the logistics:

- Logistics cost is the total expenditure of an enterprise for factors like packaging, storage and transport. As it's a factor that is easy to overlook, poor pre-planning often leads to excessive costs.
- Air Cost and Land Cost, as the name would suggest, are the summation of expenses that would be incurred for transport of goods through air and land respectively.
- Per KG costs are the rates of the aforementioned costs for a unit kilogram of goods.
- OTIF or on-time-in-full performance percentage is an efficiency measure for different modes and the Company Warehouse to Clear and Forwarding Agents delivery. It's the calculation of successful OTIF deliveries.

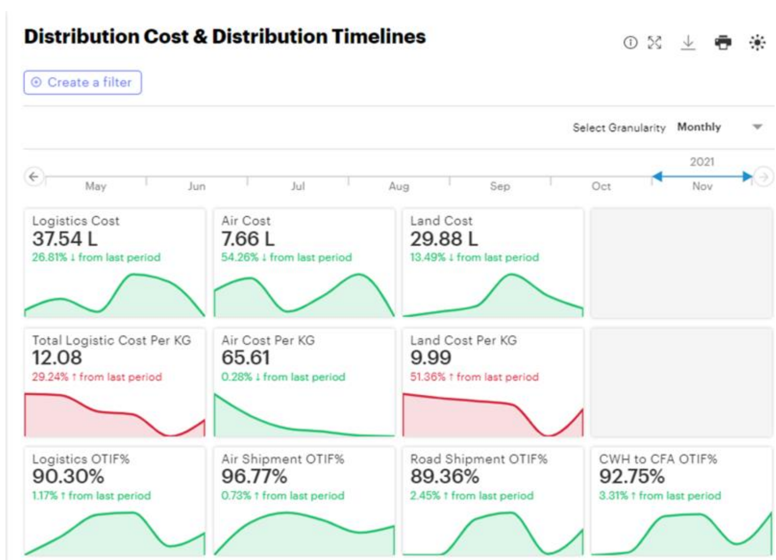


Figure 12 Distribution Cost & Distribution Timelines

4.1.3. Dispatch Efficiency:

Timely dispatch precedes and sets the tone for a timely delivery. Poor dispatch efficiency could cause a ripple effect and accumulated delay in the entire delivery cycle. So while delivery date may be in direct correlation with any transport company SLA, the dispatch date is of equal or even greater importance.

This KPI is a measure of how many goods are dispatched on or before the calculated date for timely delivery against the ones that aren't

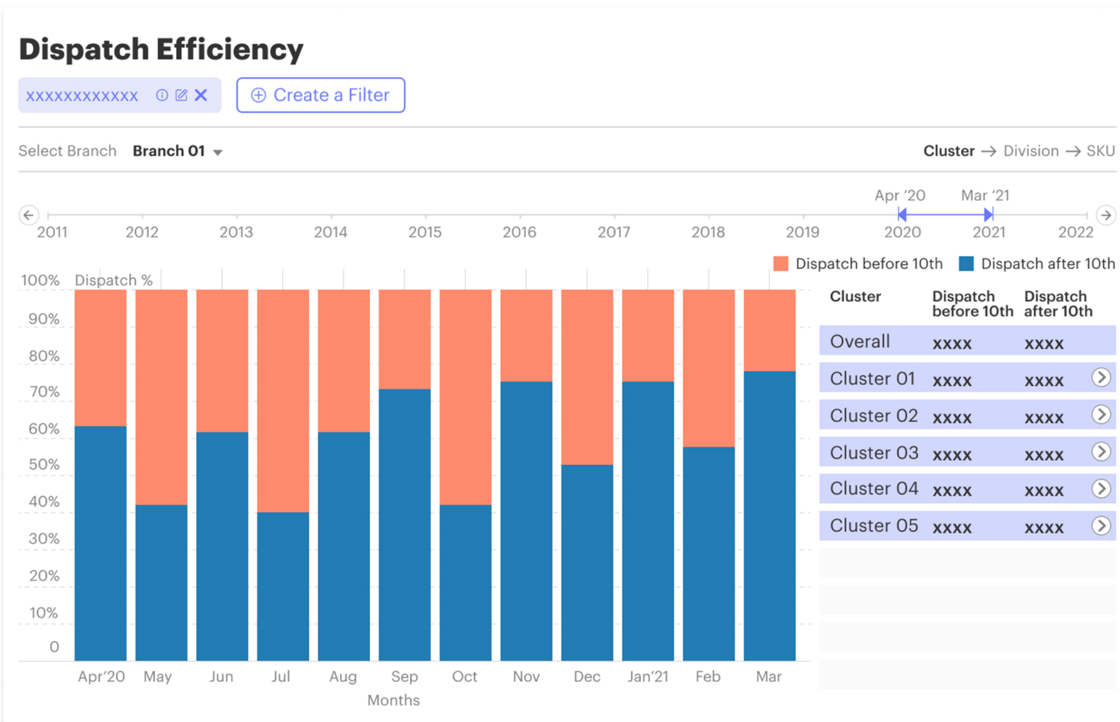


Figure 13 Dispatch Efficiency

4.1.4. Inventory Days

This measurement estimates the typical time between the acquisition of the raw materials and the sale of the finished item to a wholesaler. Utilising this measurement, the company can work out the quantity of days it would take your inventory to run out on the off chance that it is not renewed. This KPI aims to guarantee that one can limit inventory days of supply to diminish the chance of an outdated stock and keep away from overabundant stock that ties up operational cash flow.

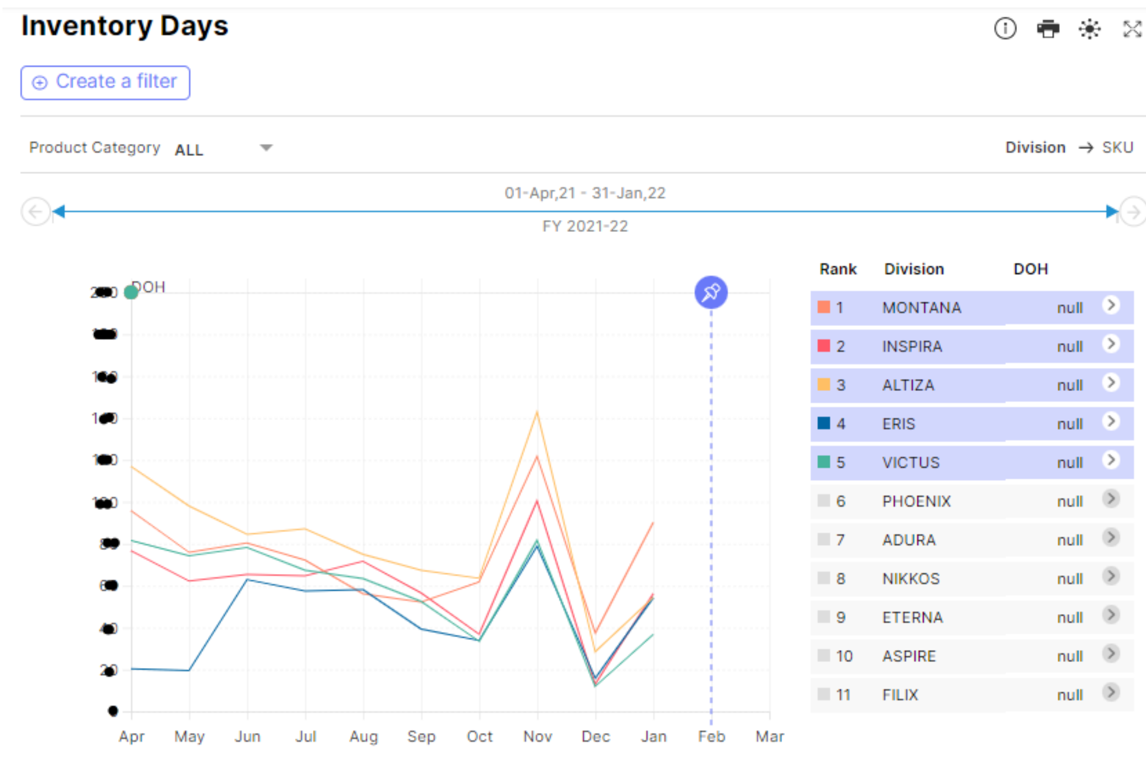


Figure 14 Inventory Days

4.1.5. Near Expiry Liquidation Compliance:

This KPI is a trend of the count of the inventory products that are near expiration date, within 6 months to be precise, and would need to be liquidated soon. This is a good way to assess and cut future losses when the company devises its next cycle of supply and demand.

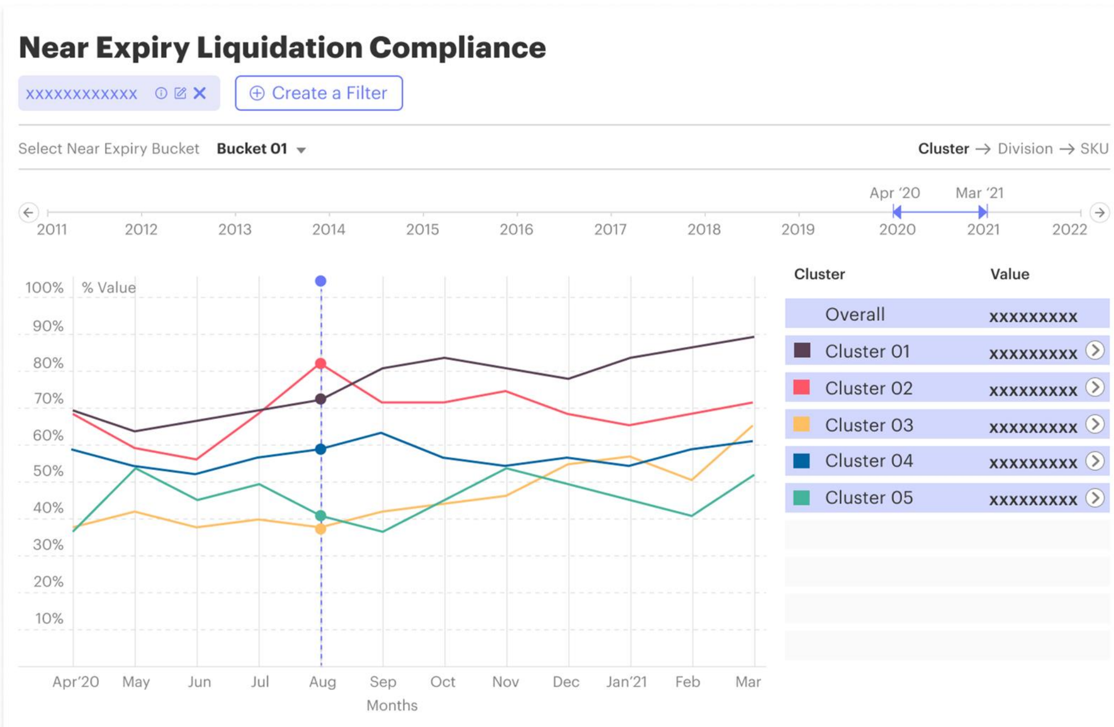


Figure 15 Near Expiry Liquidation Compliance

4.1.6. Non-moving inventory value:

This is the visual representation of the monetary value of the goods that haven't been sold and have been in the inventory for over three months. It's the summation of all the excess value. As this widget can let you manoeuvre dimensions, one can assess which division or brand or cluster incurs the most of the percentage. The awareness of this number can be important to reassess the supply and demand of certain products and can let one define the root of the problem before seeking its solution.

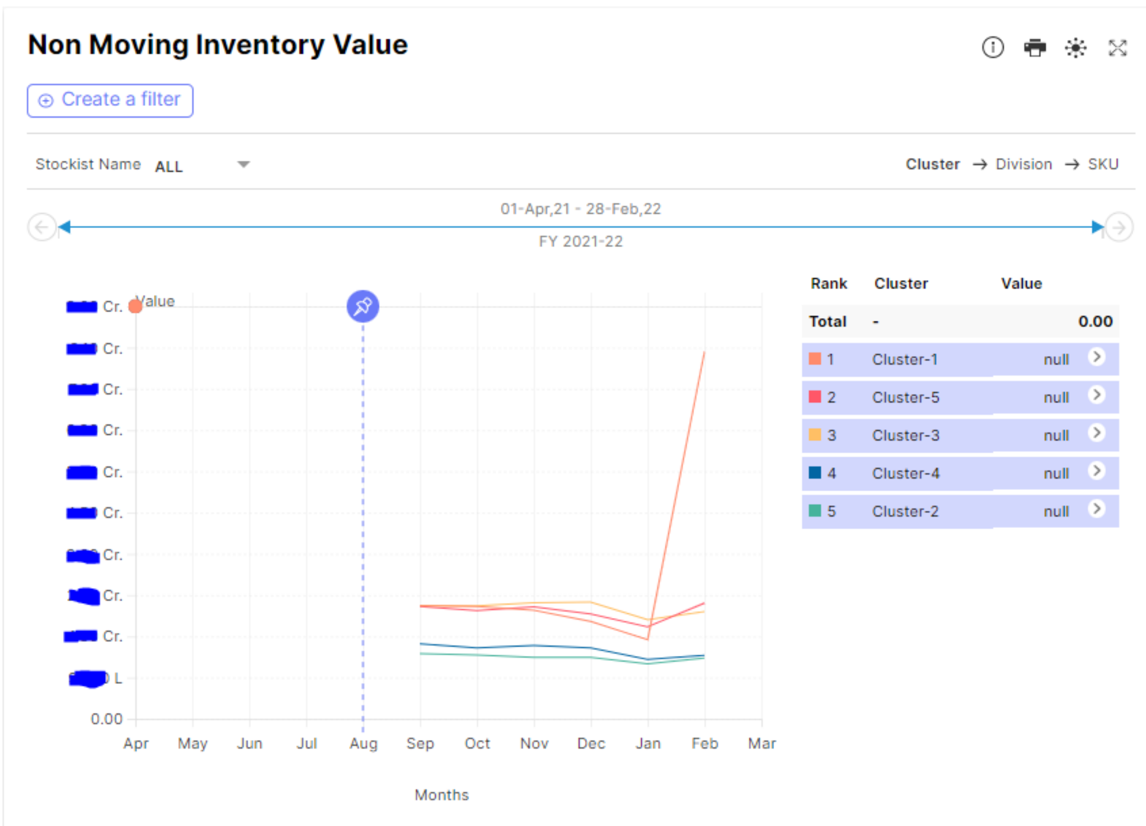


Figure 15 Non-moving inventory value

4.1.7. Forecast vs Target vs Actual

From reducing expenses to keeping purchasers blissful, forecasting is an indispensable part of production and supply chain, assisting organisations with occupying orders on time, staying away from pointless inventory costs and planning for future cost inflation. Thus it is important to understand and analyse the forecast accuracy. The above KPI widget gives a brief comparison of Actual sales, Target sales and Forecast sales. This visualisation helps to identify the most accurate number of units which can be manufactured so sustain the demand

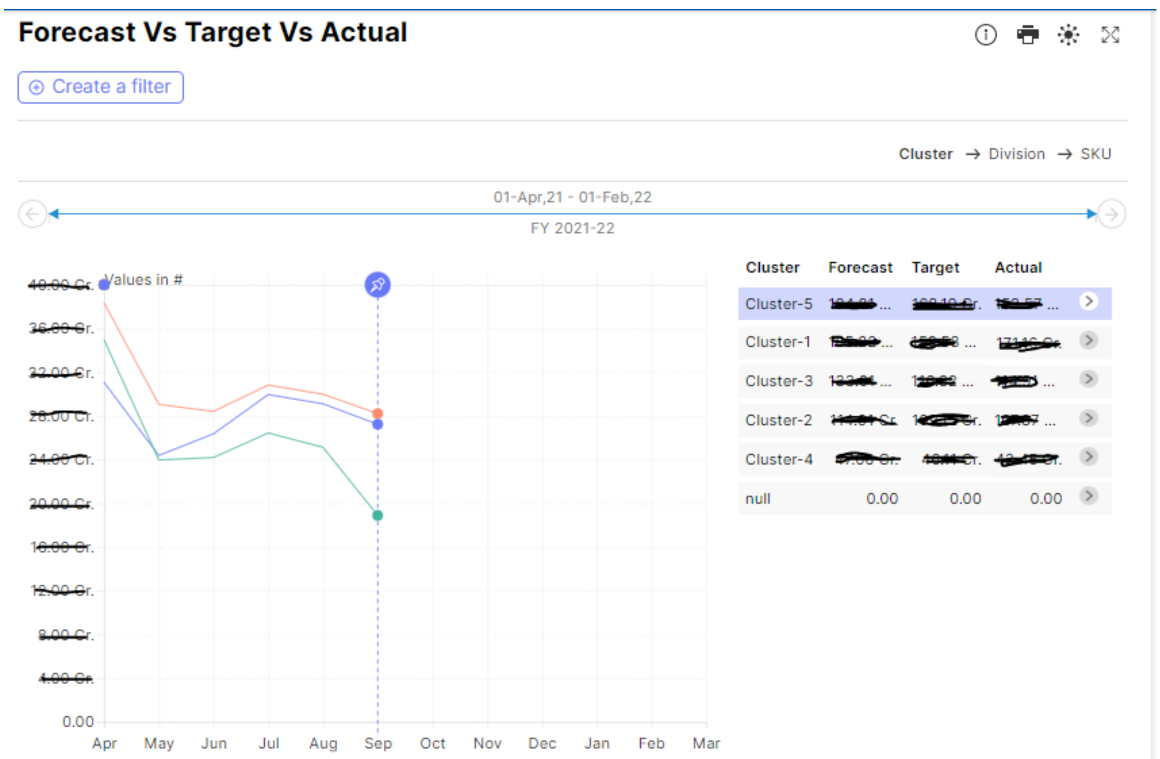


Figure 16 Forecast vs Target vs Actual

4.1.8. Over and Under Forecast

Sometimes the forecasted value could be more compared to the sales n vice versa and hence a comparison graph could be very helpful to understand their difference. By reducing forecast error in each case, one can see improvements in the profits.

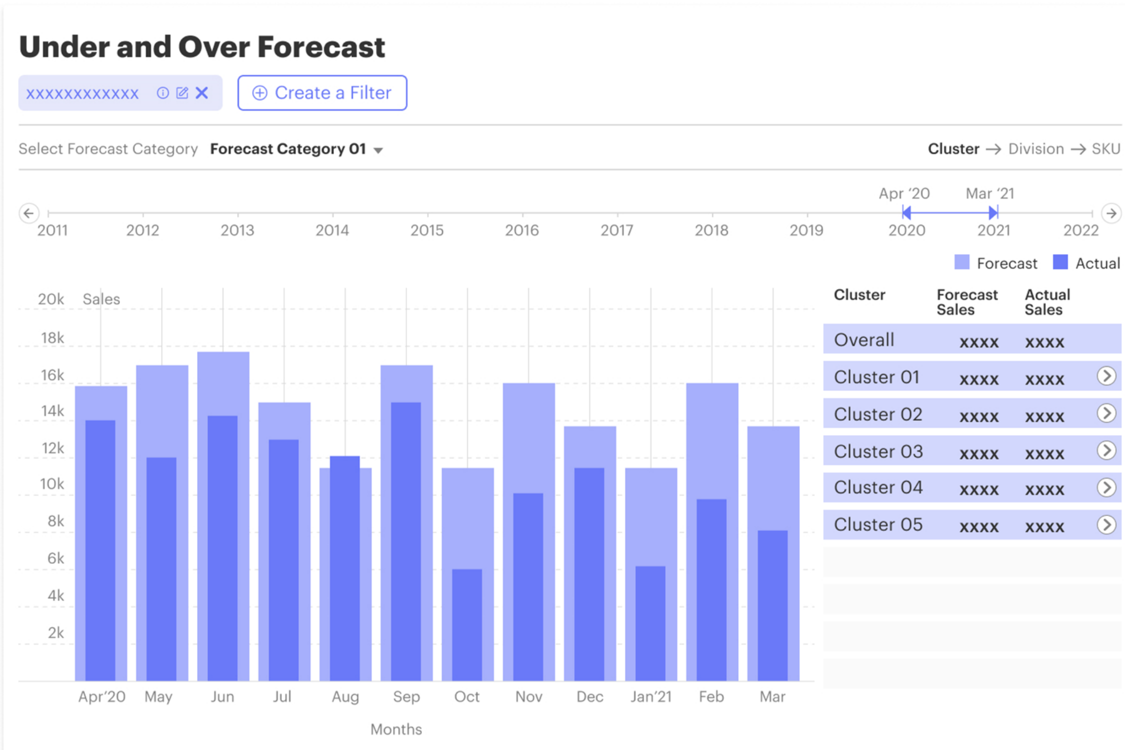


Figure 18 Over and Under Forecast

4.1.9. Requirement vs Actual

This KPI is a direct plot of supply and demand, which can help in weeding out glaring problems if any. If the percentage is too off, it is worthy of deeper investigation. Otherwise, the application remains explanatory. The data helps one assess the production units for the upcoming cycles and identify any sudden trends.

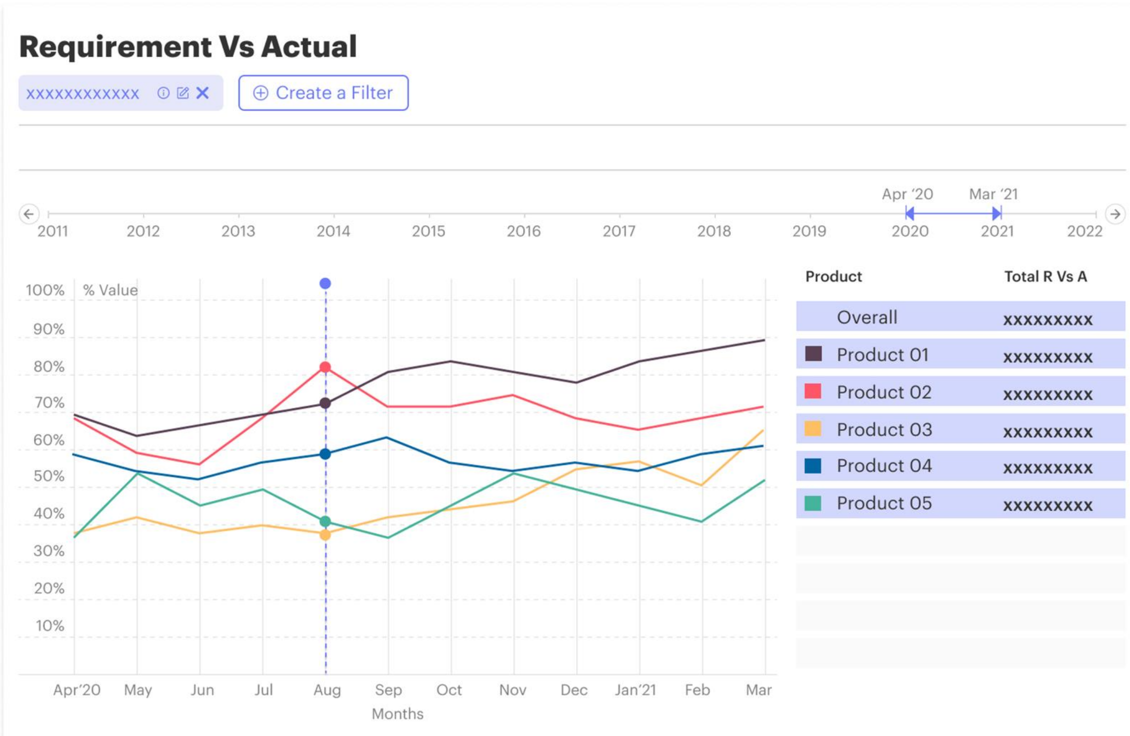


Figure 19 Requirement vs Actual

4.2. Finance

Finance function is the central governing body of the entire monetary flow of the business, so naturally it holds a responsibility far too heavy. Business intelligence and picturisable statistics here therefore, can have an enterprise have a better grip and understanding of the cash flow of the system.

4.2.1. Expense Reported and MIS

Expense shows the operational efficiency of your company by comparing operating expenses (the cost associated with running core operations) to the total revenue. The widget below display monthly expenses incurred.

Figure 20 Expense Reported and MIS

4.2.2. Cash Conversion Cycle

This is the waterfall plot of the cash conversion cycle. This is an important KPI that is a marker of the total efficiency of a business. Keeping its track lets an enterprise visualise the pace of conversion of sales in cash and back into cash. This also provides assistance in gauging a business's cash flow position.

Here, the working capital is the summation of inventory and debtors, minus the creditors.

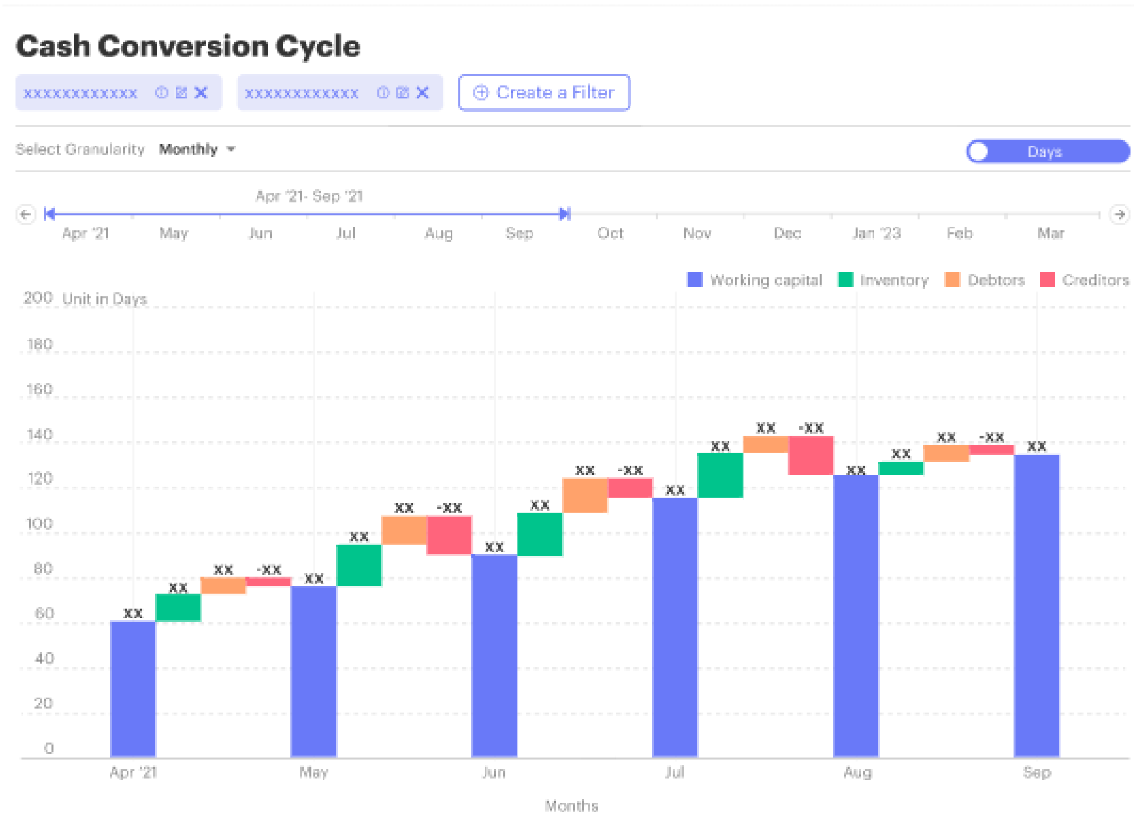


Figure 21 Cash Conversion Cycle

4.2.3. Receivables Ageing

This widget is a heat map that describes the ageing of receivables based on the GL name ie Customer name and GL Code ie Customer Code bifurcated by the intensity of colours. Accounts receivable ageing is the procedure of distinguishing open accounts receivables on the basis of the length of time an invoice has been outstanding. It is a highly efficient metric in determination of the allowance for accounts that raise doubt.

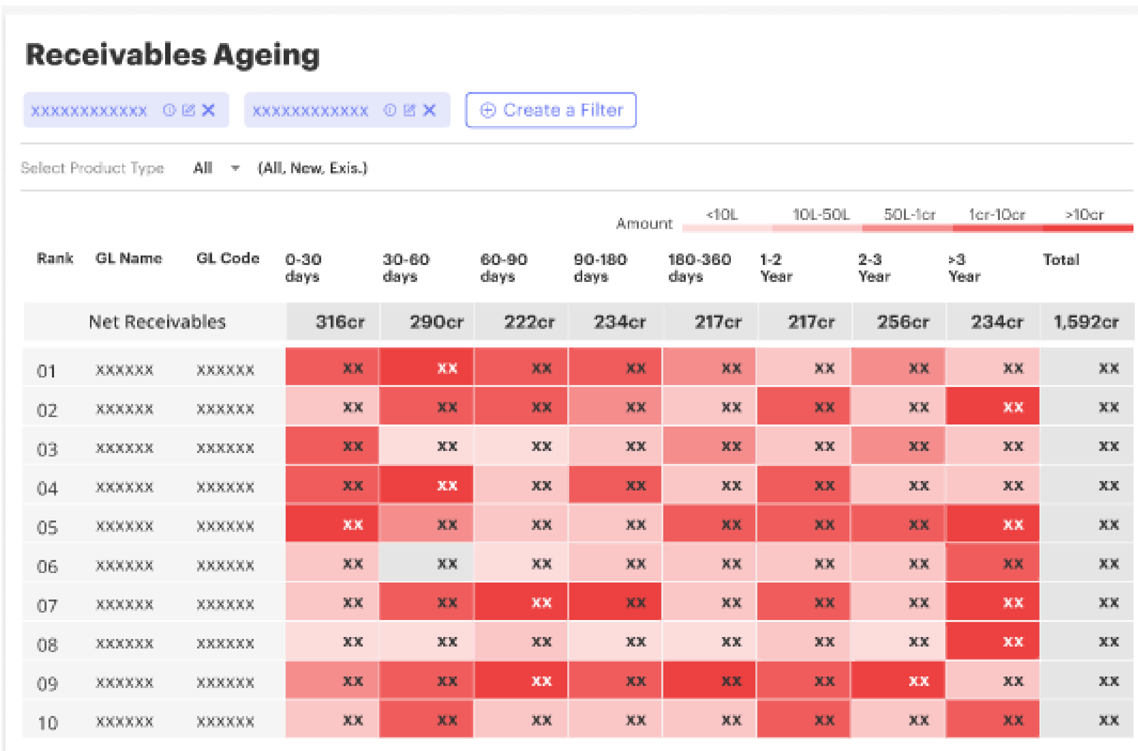


Figure 22 Receivables Ageing

4.2.4. Movement of Advances

Advances are any prepayment made on a future commitment or payment. This widget depicts the current sum of advances per month calculated by subtraction of outstanding amount and provision made

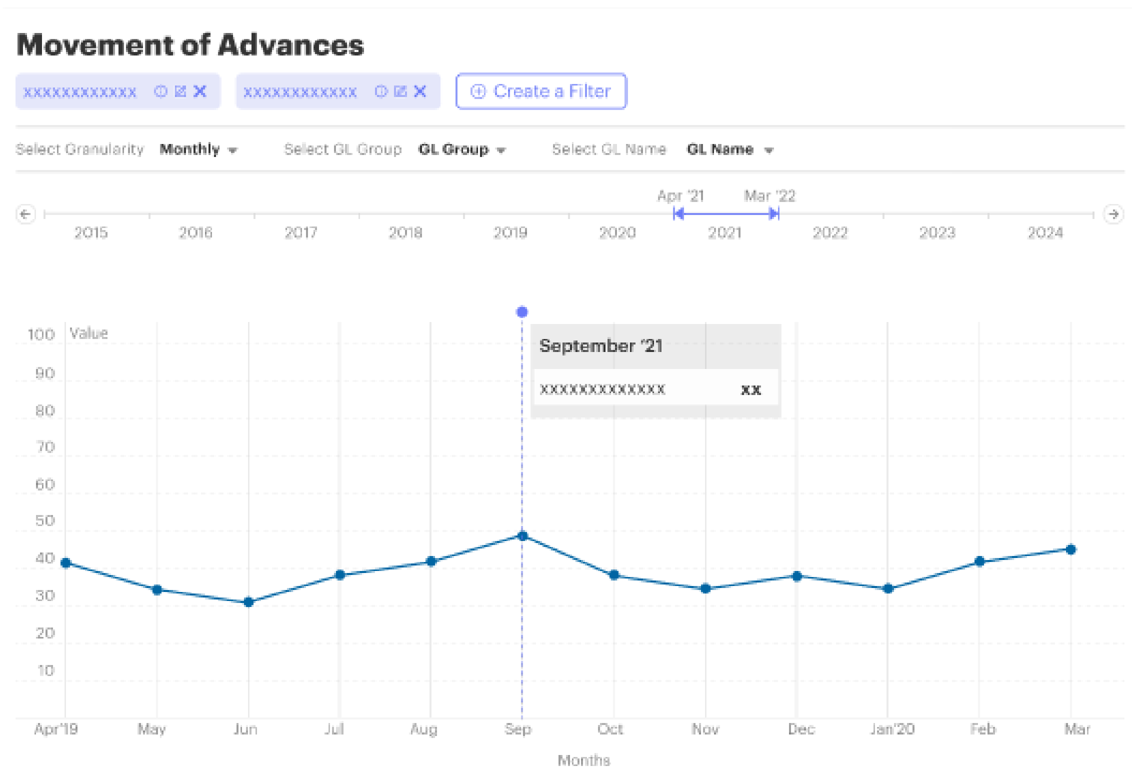
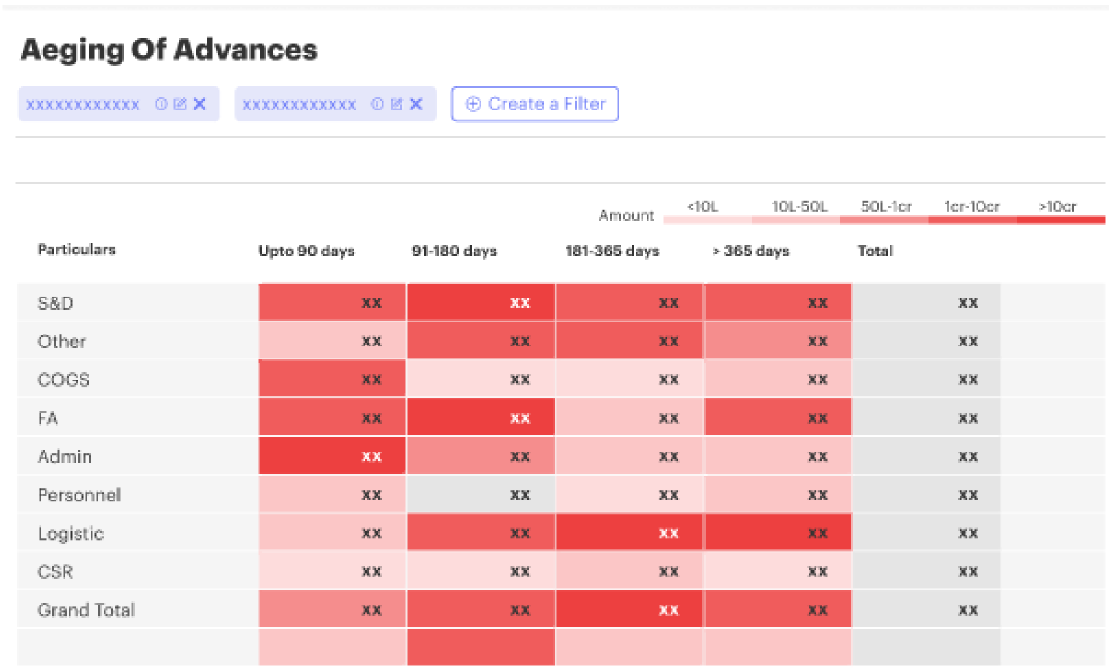


Figure 23 Movement of Advances

4.2.5. Ageing of Advances

Very similar to Receivables ageing, this KPI is used to visualise the accounts/customers with advance payment based on the length of time an invoice has been made. This is useful to understand the transaction of various advances based on a time period.



Particulars	Upto 90 days	91-180 days	181-365 days	> 365 days	Total
S&D	xx	xx	xx	xx	xx
Other	xx	xx	xx	xx	xx
COGS	xx	xx	xx	xx	xx
FA	xx	xx	xx	xx	xx
Admin	xx	xx	xx	xx	xx
Personnel	xx	xx	xx	xx	xx
Logistic	xx	xx	xx	xx	xx
CSR	xx	xx	xx	xx	xx
Grand Total	xx	xx	xx	xx	xx

Figure 24 Ageing of Advances

4.2.6. Valuation Ratios

A valuation proportion shows the connection between the market worth of an organisation or its value and some key financial measurement. Here we have utilise three fundamental measurement i.e:

- Price to Book ratio analyses an organisation's market worth to its book esteem. The market worth of an organisation is its share price increased by the number of outstanding shares. The book esteem is the net assets of an organisation

- Price to earning ratio is the proportion of an organisation's share price to the organisation's income per share. The proportion is utilised for esteeming organisations and to see if they are overvalued or undervalued

- Enterprise value to EBITA (Earnings before interest, taxes, and amortisation) ratio. It is used to decide the fair evaluation of an organisation as it is free of the capital design (for example the combination of debt and equity). Therefore this multiple can be utilised to differentiate companies with different levels of debt. It likewise evades the huge inadequacy of the P/E ratio which can be tangibly impacted by the level of leverage in the organisation

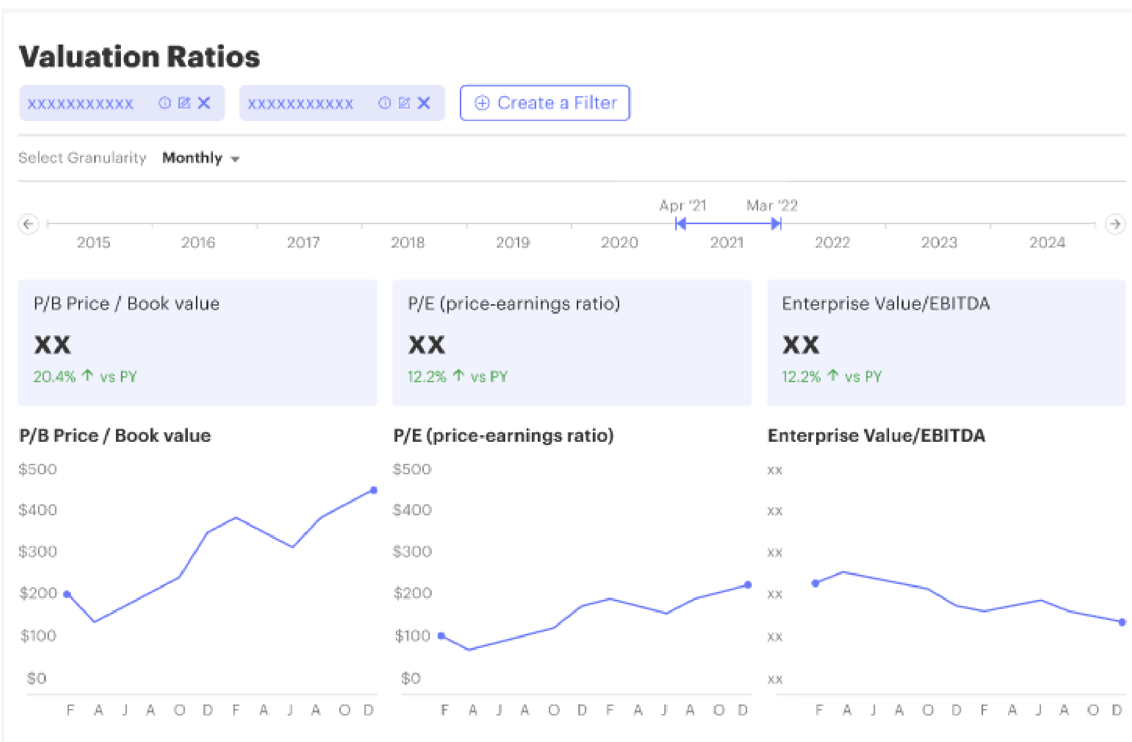


Figure 25 Valuation Ratios

4.2.7. Liquidity Ratios

Liquidity ratios are the measure of an enterprise's ability to pay debt obligations and its margin of safety. It is achieved through the calculation of metrics like:

- The current ratio measures a company's ability to pay off its current liabilities with its total current assets
- The quick ratio measures a company's ability to meet its short-term obligations with its most liquid assets and therefore excludes inventories from its current assets
- Inventory Holding Period is a ratio that depicts the number of days for which an organisation holds inventory before sales
- Interest coverage ratio is a debt and profitability ratio used to determine how easily a company can pay interest on its outstanding debt.

Figure 26 Liquidity Ratios

4.2.8. Cash Flow Ratios

Companies use cash flow ratios to determine their overall financial performance. Cash flow ratios make an examination between cash flows and different components of a financial statement. A more significant level of cash flow, a superior capacity to endure decreases in working performance, as well as a superior capacity to deliver profits to investors.

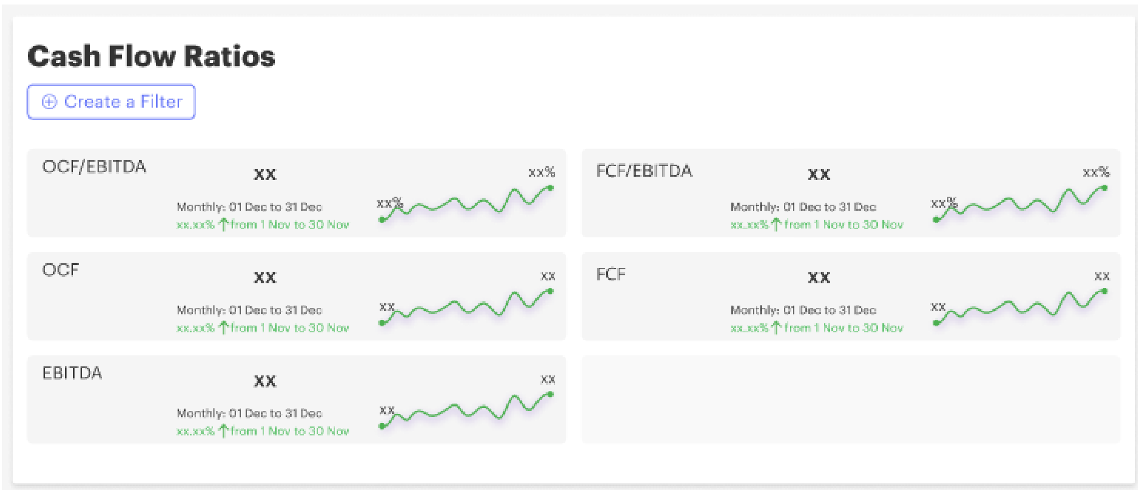


Figure 27 Cash Flow Ratios

4.2.9. Profitability Ratios

Profitability ratios are a class of monetary measurements that are utilised to survey a business' capacity to create profit comparative with its income, working expenses, balance sheet assets, or shareholders' equity over time

- ROE(Return on Equity) :Communicates the percentage of net income comparative with investors' value, or the pace of return on the cash that value investors have placed into the business

- ROCE(Return on Capital Employed): Implies how much income or benefit an organisation creates considering the investors' value as well as the obligation and different sources of assets
- Return on assets: The measurement is usually communicated as a rate by utilising an organisation's net gain and its average assets. A higher ROA implies an organisation is more proficient and useful at dealing with its balance sheet to create benefit

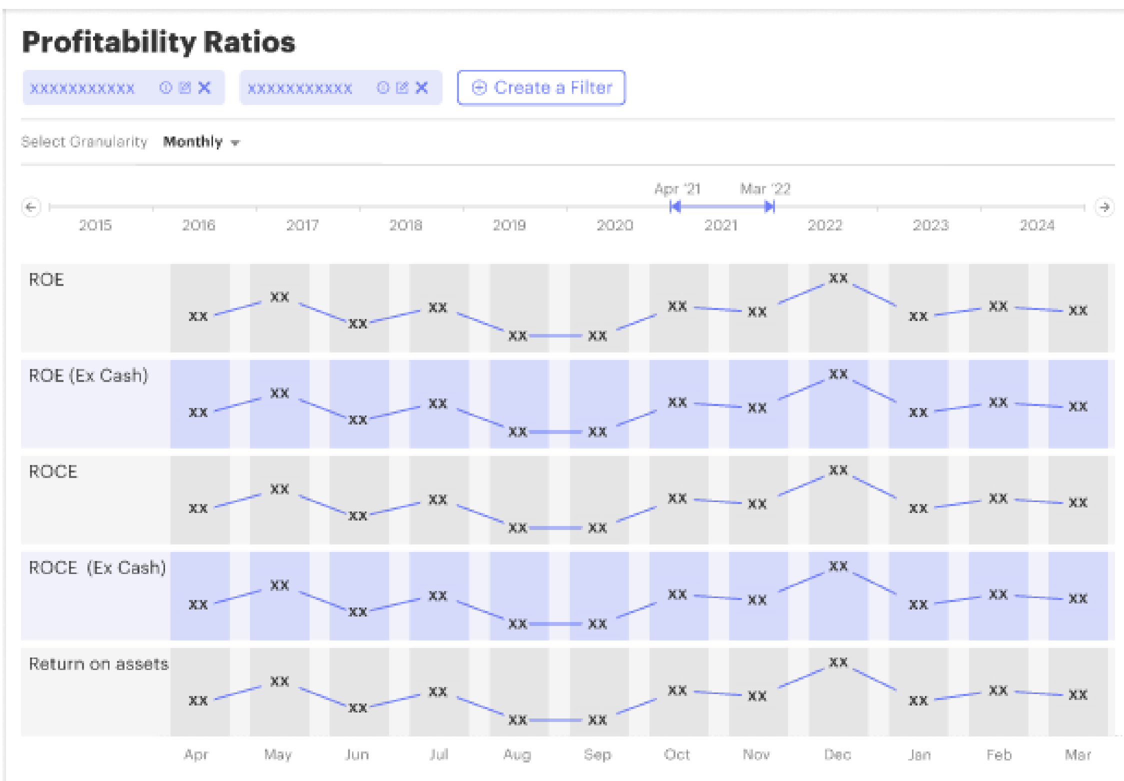


Figure 28 Profitability Ratios

5. CONCLUSION

There's a lot of new industrial jargon evolving daily in correspondence to the developing global techscape and it's hard to keep up when the exponential glow-up doesn't even seem to flicker much less dim. And that fact is telling of how more than building on layered tool stacks for your enterprise, the answer lies with being able to adapt and scale. And bigger things can only be built over a robust base. This is what digital transformation is all about at the crux of it.

With domains so crucial to human existence like Life Sciences, especially after the close call that COVID19, it's important to invest in an environment that doesn't shy from change. Product based companies tend to underinvest in technology and even though the times are changing, it's again a matter of who gets the head start in the race. Competitive market perspective aside, as consumerism continues to stay the epicentre of modern businesses, it's also essential to factor in the waves of advantage a fully equipped to evolve and modern technology powered system could bring in. With, say, state-of-the-art analytics deployed for healthcare or even vigorously nano-networked health alert systems, the healthcare industry cannot even begin to visualise its peak. And it's the way forward.

The way forward is all there is.

6. REFERENCES

[1] Ranjan, Jayanthi. (2007). "Application of data mining techniques in pharmaceutical industry." Journal of Theoretical and Applied Information Technology. 3.

[2] Weinberg, Zach, and Eric Schadt. "Dean for Precision Medicine, Mount Sinai Health System Professor, Genetics and Genomic Sciences at Mount Sinai.".

[3] Raghupathi, Wullianallur, and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." Health information science and systems 2.1 (2014): 1-10.

[4] "Groves, Peter, Basel Kayyali, David Knott, and Steve Van Kuiken. "The 'big data' revolution in healthcare: Accelerating value and innovation." (2016).

[5] Samhan, Bahae, Tara Crampton, and Regina Ruane. "The trajectory of IT in healthcare at HICSS: A literature review, analysis, and future directions." Communications of the Association for Information Systems 43.1 (2018): 41.

[6] Greene, William. "The emergence of India's pharmaceutical industry and implications for the US generic drug market." (2007).

[7] Ranjan, J. (2009), "Data mining in pharma sector: benefits", International Journal of Health Care Quality Assurance, Vol. 22 No. 1, pp. 82-92. https://doi.org/10.1108/09526860910927970

[8] Hole, Glenn, Anastasia S. Hole, and Ian McFalone-Shaw. "Digitalization in pharmaceutical industry: What to focus on under the digital implementation process?." International Journal of Pharmaceutics: X 3 (2021): 100095.

[9] Seyed Ghorban, Zahra, Hossein Tahernejad, Royston Meriton, and Gary Graham. "Supply chain digitalization: past, present and future." Production Planning & Control 31, no. 2-3 (2020): 96-114.

[10] Khetan, S. K., & Sonchhatra, J. (2021, March 23). *Digital Transformation in Finance – insights from EY Survey*. EY US - Home. Retrieved May 7, 2022, from https://www.ey.com/en_in/financial-accounting-advisory-services/digital-transformation-in-finance-insights-from-ey-survey

[11] Cremer, Philipp, M. Losch, and Ulf Schrader. "Maximizing efficiency in pharma operations." New York, NY: The McKinsey Quarterly (2009) from https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/Maximizing%20efficiency%20in%20pharma%20operations/Maximizing%20efficiency%20in%20pharma%20operations.pdf

[12] *Pricing overview-how azure pricing works: Microsoft azure*. Pricing Overview-How Azure Pricing Works | Microsoft Azure. (n.d.). Retrieved May 7, 2022, from https://azure.microsoft.com/en-in/pricing/

[13] *Service level agreements summary: Microsoft Azure*. Service Level Agreements Summary | Microsoft Azure. (n.d.). Retrieved May 7, 2022, from https://azure.microsoft.com/en-in/support/legal/sla/summary/#:~:text=Azure%20Logic%20Apps,available%2099.9%25%20of%20the%20time.

[14] Jianleishen. (n.d.). *Copy activity performance and scalability guide - azure data factory & azure synapse*. Azure Data Factory & Azure Synapse | Microsoft Docs. Retrieved May 7, 2022, from https://docs.microsoft.com/en-us/azure/data-factory/copy-activity-performance

[15] Kromerm. (n.d.). *Optimizing performance of transformations in mapping data flow - azure data factory & azure synapse*. Optimizing performance of transformations in mapping data flow - Azure Data Factory & Azure Synapse | Microsoft Docs. Retrieved May 7, 2022, from https://docs.microsoft.com/en-us/azure/data-factory/concepts-data-flow-performance-transformations

[16] *Azure security: Microsoft Azure*. Azure Security | Microsoft Azure. (n.d.). Retrieved May 7, 2022, from https://azure.microsoft.com/en-in/overview/security/

[17] Finelli, L. A., & Narasimhan, V. (2020). Leading a digital transformation in the pharmaceutical industry: Reimagining the way we work in Global Drug Development. *Clinical Pharmacology & Therapeutics*, *108*(4), 756–761. https://doi.org/10.1002/cpt.1850

[18] Insight, AT Kearney. "India's Pharma Supply Chain: Does the Industry Have What It Takes to Win." (2016), ATK Insight - AT Kearney, Chicago

[19] *Steering pharma to future success with Digital Finance*. Genpact. (n.d.). Retrieved May 7, 2022, from https://www.genpact.com/insight/point-of-view/steering-pharma-to-future-success-with-digital-finance