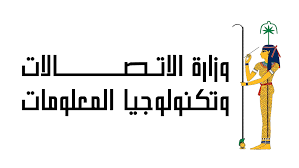
****

Manufacturing Downtime Analysis

Project Report

Digital Egypt Pioneers Initiative (DEPI)

Google Data Analysis Specialist Track

CAI2\_DAT1G6

Contributors

Abd Al-Rahman Mohamed Amin

Ahmed Mohamed

Mohamed Shebl Azab

Omar Mohamed Shebl

November 2024 – April 2025

Table of Contents

[Executive Summary 2](#_Toc194959499)

[Phase I: Data Cleaning and Preprocessing 3](#_Toc194959500)

[Data Overview 3](#_Toc194959501)

[Data Tables Normalization 4](#_Toc194959502)

[The Data Cleaning Process 4](#_Toc194959503)

[Data Modelling 5](#_Toc194959504)

[Tools and Technologies Used 6](#_Toc194959505)

[Deliverables 6](#_Toc194959506)

[Phase II: Exploratory Data Analysis 7](#_Toc194959507)

[Analysis Questions 7](#_Toc194959508)

[Key Findings 7](#_Toc194959509)

[Deliverables 8](#_Toc194959510)

[Phase III: Forecasting Analysis 9](#_Toc194959511)

[Forecasting Questions 9](#_Toc194959512)

[Key Findings 9](#_Toc194959513)

[Deliverables 9](#_Toc194959514)

[Phase IV: Visualization and Final Presentation 10](#_Toc194959515)

[Deliverables 10](#_Toc194959516)

[Recommendations 11](#_Toc194959517)

[Contributors 12](#_Toc194959518)

[Appendices 13](#_Toc194959519)

[Appendix I: Exploratory Analysis Visualizations 13](#_Toc194959520)

[Appendix II: Tableau Dashboard Screenshots 16](#_Toc194959521)

# Executive Summary

This project investigates the factors causing production downtime in a soda drinks company. Using data cleaning, modeling, exploratory analysis, and forecasting, we identify the major contributors to downtime and predict future trends. The findings provide actionable insights to optimize operations and minimize downtime.

The project follows a structured data analysis approach. First, raw operational data is cleaned and preprocessed to ensure accuracy. A data model is then built to establish relationships between key variables affecting downtime. Exploratory analysis is conducted to uncover patterns, followed by forecasting models that predict future downtime trends. Finally, a Tableau dashboard visualizes the findings for decision-making.

The dataset consists of downtime records collected over a period of 6 days. The dataset included categories for causing factors, details on the batches produced such as production date, and details on the products produced by the company such as product flavors and package sizes. This allowed analysis of downtime by product features and causing factors.

The analysis found that five out of possible twelve causing factors contributed to 80% of the downtime. Machine adjustments were the primary cause behind downtime accounting for 23.92% of the downtime during the observed period. Forecasting analysis revealed that downtime is expected to drop by 33.22% within 7 days.

To mitigate production downtime, it is recommended that the company improves staff training on operating the machinery and troubleshooting, and implement a preemptive maintenance schedule. The Tableau dashboard provides real-time insights to help management track downtime patterns and take proactive measures.

# Phase I: Data Cleaning and Preprocessing

Data cleaning is a crucial step before analysis because raw data often contains errors, inconsistencies, and missing values that can lead to inaccurate insights. By removing duplicates, handling missing data, correcting errors, and standardizing formats, data cleaning ensures the dataset is reliable and consistent.

## Data Overview

The raw dataset contains records on produced batches and downtime caused by different factors. The table below describes the dataset in details:

|  |  |  |
| --- | --- | --- |
| **Entity** | **Entity Type** | **Description** |
| **Downtime Factors** | **Dimension Table** | **Includes details on each downtime factor.** |
| Factor | Attribute | Unique identifier for each downtime factor. |
| Description | Attribute | Descriptive text for each factor. |
| Operator Error | Attribute | Whether a factor involves a human error. |
| **Line Downtime** | **Fact Table** | **Records downtime caused by each factor in production batches.** |
| Batch | Attribute | Unique identifier for produced batches. |
| Factor | Attribute | Downtime minutes for each factor in a wide format (12 columns). |
| **Line Productivity** | **Fact Table** | **Records details on each of the batches produced.** |
| Batch | Attribute | Unique identifier for the batch. |
| Date | Attribute | Production date. |
| Product | Attribute | Unique identifier for products. |
| Operator | Attribute | Name of the operator in charge during batch production. |
| Start Time | Attribute | Time at which batch production started. |
| End Time | Attribute | Time at which batch production ended. |
| **Products** | **Dimension Table** | **Includes details on the products produced by the company.** |
| Product | Attribute | Unique identifier for products. |
| Flavor | Attribute | Soda flavor of the product e.g., cola, lemon …etc. |
| Size | Attribute | Pack volume in milli liters. |
| Min Batch Time | Attribute | Minimum time required to produce a single batch without delay. |

## Data Tables Normalization

Data table normalization is the process of organizing a database to reduce redundancy and improve data integrity. It involves structuring tables according to a series of normal forms (NF), such as First Normal Form (1NF), Second Normal Form (2NF), and so on. The process eliminates duplicate data, ensures logical dependencies, and minimizes anomalies in data insertion, updating, and deletion. Normalization was applied to each of the provided tables to the third normal form (3NF):

* First Normal Form (1NF): ensure no repeating groups or multivalued attributes and that each column contains atomic values.
* Second Normal Form (1NF): ensure all non-key attributes are fully functionally dependent on the entire primary key.
* Third Normal Form (3NF): eliminate transitive dependencies so non-key attributes depend only on the primary key.

**Findings:**

1. The Line Downtime table violated 1NF because downtime factors are stored as column headers. Microsoft Excel Power Query’s unpivot columns feature was used to convert the table into a long format with only 3 columns: batch, factor, and downtime.
2. The Products table violated 3NF and was found to have a transitive dependency where Min Batch Time depended on Size rather than the primary key. Min Batch Time and Size were separated in a new table (Products Details) to resolve this issue.

## The Data Cleaning Process

Python was used to ensure that each table was clean and ready for analysis:

1. Check for missing values: No value was found to be missing in any of the tables.

1. # Check for null values in each column

2. line\_prod.isnull().sum()

1. Check that the number of unique IDs is equal to the number of rows in the table: all rows were found to have unique IDs in all the tables.

1. # Count the number of unique batch IDs

2. # The number of unique batch IDs should be equal to the number of rows in the data frame

3. line\_prod.Batch.nunique()

1. Check for duplicate records: no duplicate records were found in any of the tables.

1. # Find the number of duplicate records

2. line\_prod.duplicated().sum()

1. Check data type and formatting consistency in each column

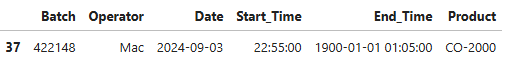
1. # Check that all End Time values are of the type datetime.time

2. line\_prod["End\_Time"].apply(lambda x: isinstance(x, datetime.time)).all()

The data types of values in the End Time column were not consistent. The following code was used to isolate rows where the type was not datetime.time where only one row was found to have a datetime.datetime data type:

1. # Find values in End Time that are not of the type datetime.time

2. line\_prod[line\_prod["End\_Time"].apply(lambda x: not(isinstance(x, datetime.time)))]



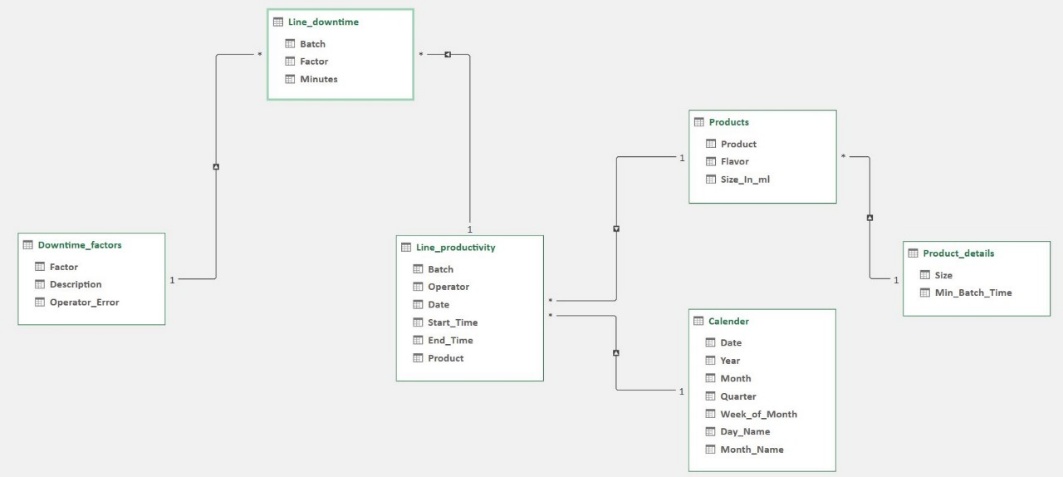
The time component of was extracted and used to overwrite the value at index 37:

1. # Extract the time component and rewrite the value

2. line\_prod.loc[37, "End\_Time"] = line\_prod.loc[37, "End\_Time"].time()

## Data Modelling

Microsoft Excel Power Pivot was used to construct a calendar table, and build a relational model to ensure efficient querying and visualization. The following entity relationship diagram (ERD) summarizes all the relations in the model:



## Tools and Technologies Used

|  |  |  |
| --- | --- | --- |
| 1. Microsoft Excel. | 1. Power Query. | 1. Power Pivot. |
| 1. Python (packages: Pandas, NumPy, and datetime). | | 1. Microsoft SQL Server. |

## Deliverables

After data cleaning, the dataset was analysis-ready, with standardized variables and improved data integrity. The preprocessing steps are documented in a Jupyter Notebook for reproducibility.

# Phase II: Exploratory Data Analysis

The exploratory analysis was conducted to identify patterns and underlying factors contributing to production downtime. This phase helped uncover the most frequent failure type, and time-based trends.

## Analysis Questions

**Analysis of Downtime by Causes**

1. Which Factors Contribute the Most to Downtime?
2. Which Downtime Factors Occur More Frequently?
3. How Much of the Downtime is Associated with Operator Errors?
4. Which Operator Causes More Downtime?
5. Does Downtime Occur at Higher Rates for in Products with Certain Flavors?
6. How Does Product Size Impact the Occurrence of Downtime?
7. Does Product Size Cause More Downtime Due to Product Spill?
8. Which Products Experience More Downtime?

**Time-Series Analysis**

1. How Much Production Time Was Lost to Downtime in Total?
2. How Does Downtime Vary by Work Shifts?
3. How Does Total Downtime Vary over Time?

**Impact of Downtime on Production Rates**

1. How does downtime affect batch production rate?
2. How would production have performed under optimum conditions (with no downtime)?

## Key Findings

1. Five factors contributed to 80% of the total downtime, and these are:

* Machine adjustment.
* Machine failure.
* Inventory shortage.
* Batch change.
* Batch coding error.

1. Most operators require training on machine adjustments.
2. Approximately 80% of the downtime due to inventory shortage occurred with cola-flavored products.
3. Among all operators, only Mac requires training on batch changing.
4. Approximately 87.5% of the downtime due to batch change occurs with products that are not cola-flavored. This might be due to a larger number of batches produced of cola products (62.86% of the total number of batches) which means production switches to cola less often compared to other flavors.
5. All operators contribute to downtime due to batch coding errors.
6. Downtime was highest during morning shifts.

## Deliverables

By the end of this phase, all analysis questions have been formulated, and key insights have been noted. The exploratory analysis was documented in a Jupyter notebook for reproducibility. Also, an initial design for the final dashboard has been drafted.

# Phase III: Forecasting Analysis

Forecasting analysis is the process of using historical data, statistical models, and machine learning techniques to predict future outcomes. In this phase, changes in downtime trends over time was forecasted using Prophet, a python package for predictive modelling. The model was used to predict downtime and the number of produced batches in the next day of operation. Data from the last day of operation was excluded from the historical data frame that was fed to the model as its low total downtime skews the results.

## Forecasting Questions

1. How is downtime expected to change over the next day of operation?
2. Can changes in downtime be predicted with certainty based on the current data?

## Key Findings

1. Downtime was expected to decrease by -2.25% within the next day of operation.
2. The number of produced batches was expected to drop by -0.79% within the next day of operation.
3. The model predicted with a very low level of certainty probably because the size of the historical data is very small.

## Deliverables

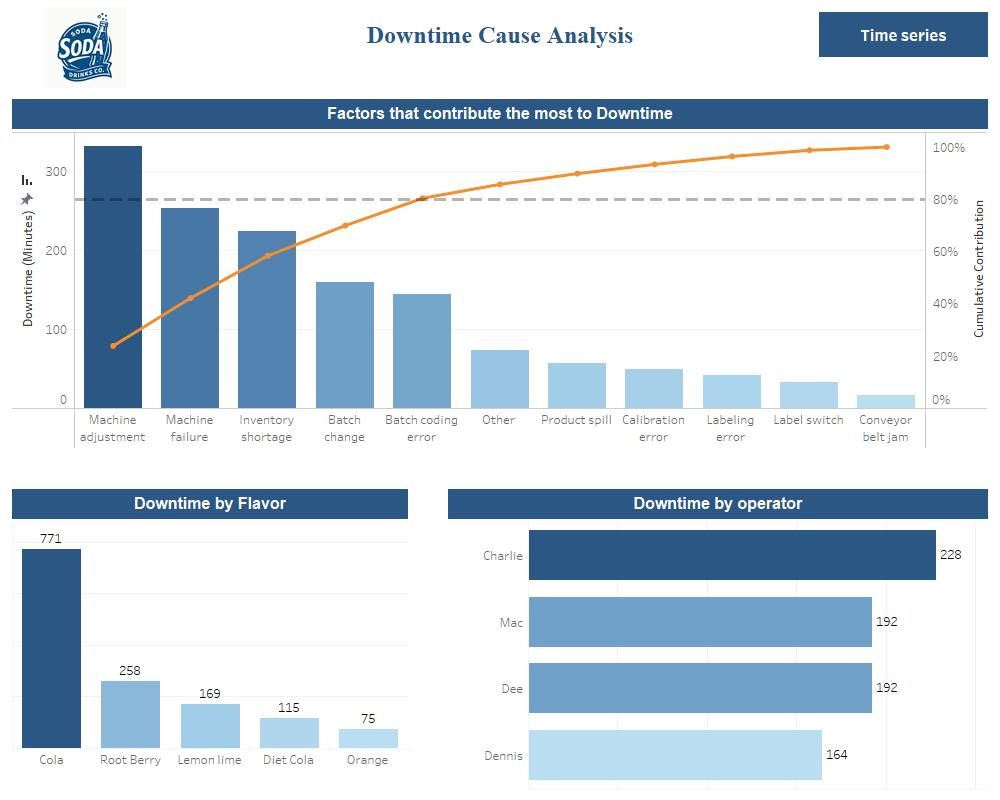
Changes in downtime were forecasted but with a low level of certainty. The forecasting analysis was documented in a Jupyter notebook for reproducibility.

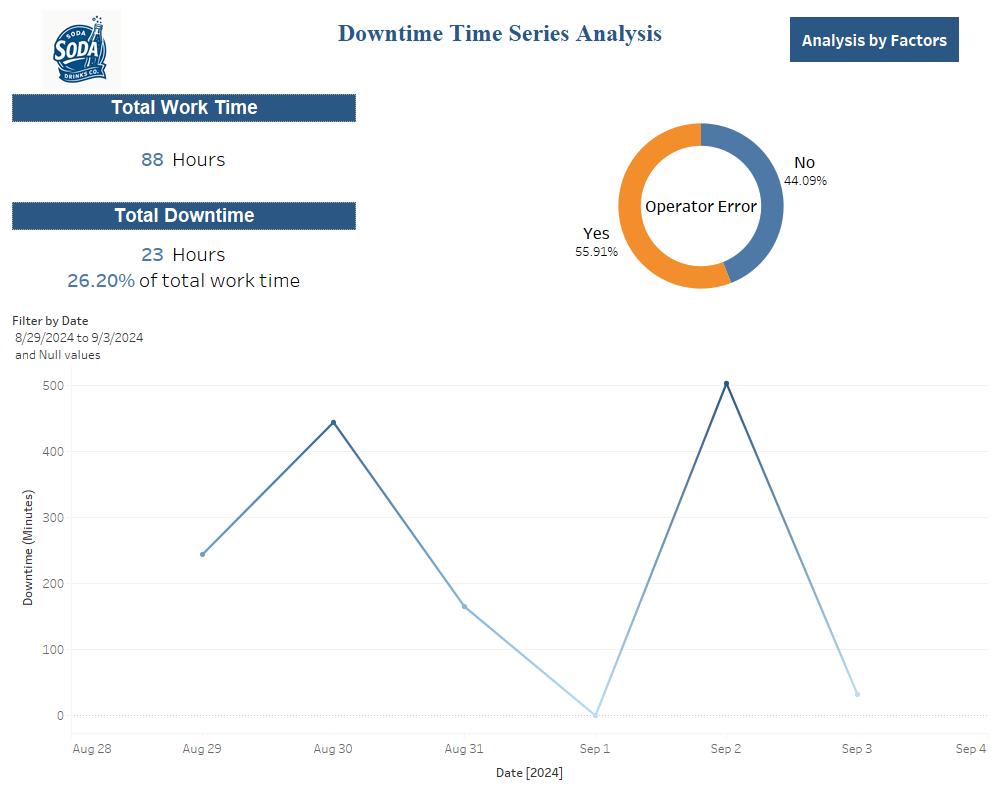
# Phase IV: Visualization and Final Presentation

Data visualization is the graphical representation of data and information using charts, graphs, maps, and other visual tools to help communicate complex insights clearly and effectively. It simplifies the understanding of large datasets, making patterns, trends, and outliers more visible and easier to interpret. In this project’s context, data visualization is important to monitor downtime in real time which allows proactive handling of future downtime events.

## Deliverables

Tableau was used to graphically present key insights discovered in the exploratory analysis phase. The graphs were organized in a dashboard that allows users to easily monitor the causes behind downtime, and the types of products most associated with downtime. Tableau was also used to build a dashboard to monitor how downtime changes over time. Moreover, a Microsoft PowerPoint presentation was built to present an overview of all the efforts made in the project.





# Recommendations

* The company should provide regular training sessions to operators to improve operator familiarity with the machines they operate on, and to equip them with the basic troubleshooting skills that can help them resolve machine failure issues. Additionally, the training should equip operators with enough knowledge about the batch coding process and the machines involved e.g., printers or labelling machines.
* The company produces cola-flavored soda at a higher rate compared to other flavors (63.16% of the produced batches were cola-flavored products). This makes the ingredients of the cola formula more prone to shortages and explains why 80% of the downtime due to inventory shortage occurred in batches of cola-flavored products. The company should re-visit its purchasing strategy and refine it to match its production schedule.
* Downtime from batch changes was noted to be longer for products that are not cola-flavored, suggesting that it occurs when switching production to products that are less frequently produced. The company should invest in obtaining spare parts for parts of the production line that require cleaning before switching to different flavors to avoid contamination e.g., mixers, to eliminate such delay.
* The company should consider upgrading the machinery software to introduce features like preset production settings. Settings can be set by a production expert which reduces operator involvement with machine adjustments. Preset settings can also help with reducing the time taken to change machine settings when switching production between different products.

# Contributors

Abd Al-Rahman Mohamed Amin

|  |  |
| --- | --- |
| Email with solid fill | [abdalrahman.mohamed.amin@gmail.com](mailto:abdalrahman.mohamed.amin@gmail.com) |
|  | [www.linkedin.com/in/a-mohamed-amin/](http://www.linkedin.com/in/a-mohamed-amin/) |
|  | [github.com/AbdAlRahman-M](https://github.com/AbdAlRahman-M) |

Ahmed Mohamed

|  |  |
| --- | --- |
| Email with solid fill |  |
|  |  |
|  |  |

Mohamed Shebl Azab

|  |  |
| --- | --- |
| Email with solid fill | [mohamed.shebl1983@gmail.com](mailto:mohamed.shebl1983@gmail.com) |
|  | [www.linkedin.com/in/mohamed-shebl-/](https://www.linkedin.com/in/mohamed-shebl-/) |
|  | [github.com/Shebl83](https://github.com/Shebl83) |

Omar Mohamed Shebl

|  |  |
| --- | --- |
| Email with solid fill |  |
|  |  |
|  |  |

To access the data files, the dashboard, and the Jupyter notebooks for exploratory analysis and forecasting, please visit the project’s GitHub repo! [(click here)](https://github.com/AbdAlRahman-M/manufacturing-downtime-analysis)

# Appendices

## Appendix I: Exploratory Analysis Visualizations

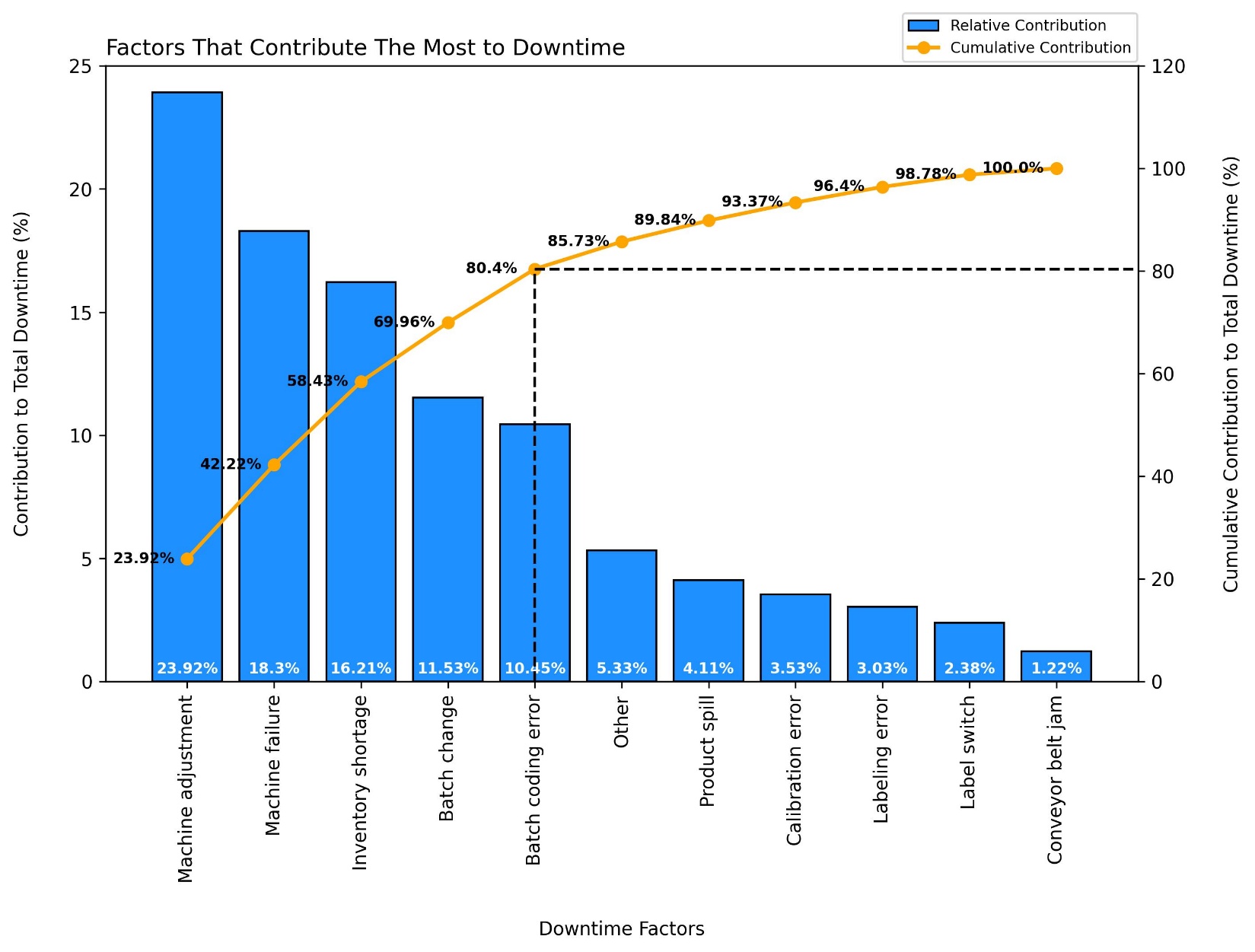


Figure : Pareto diagram depicting factors that contribute to downtime the most (generated using Python)

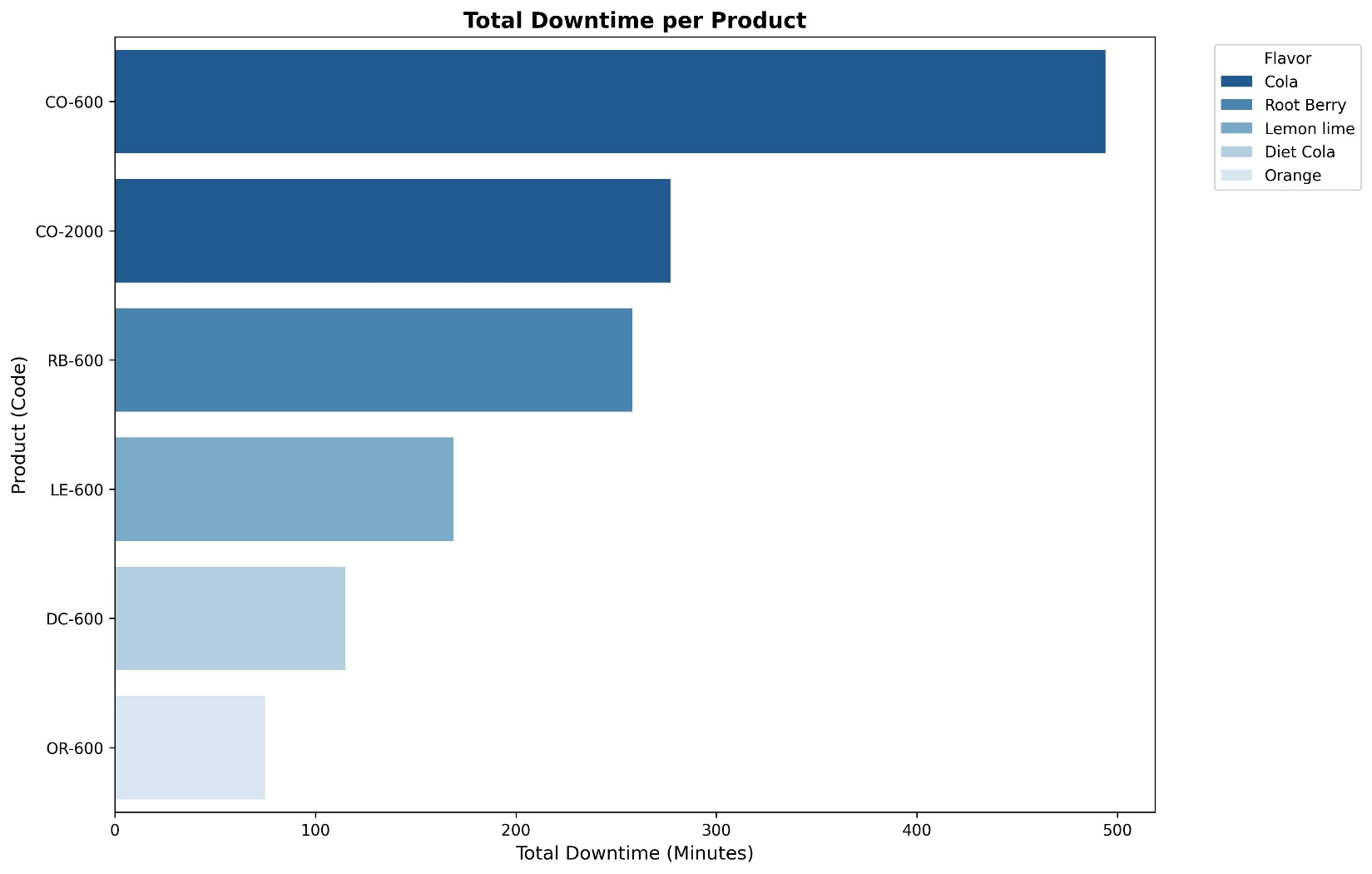


Figure : A horizontal bar depicting downtime by different products (generated using Python)

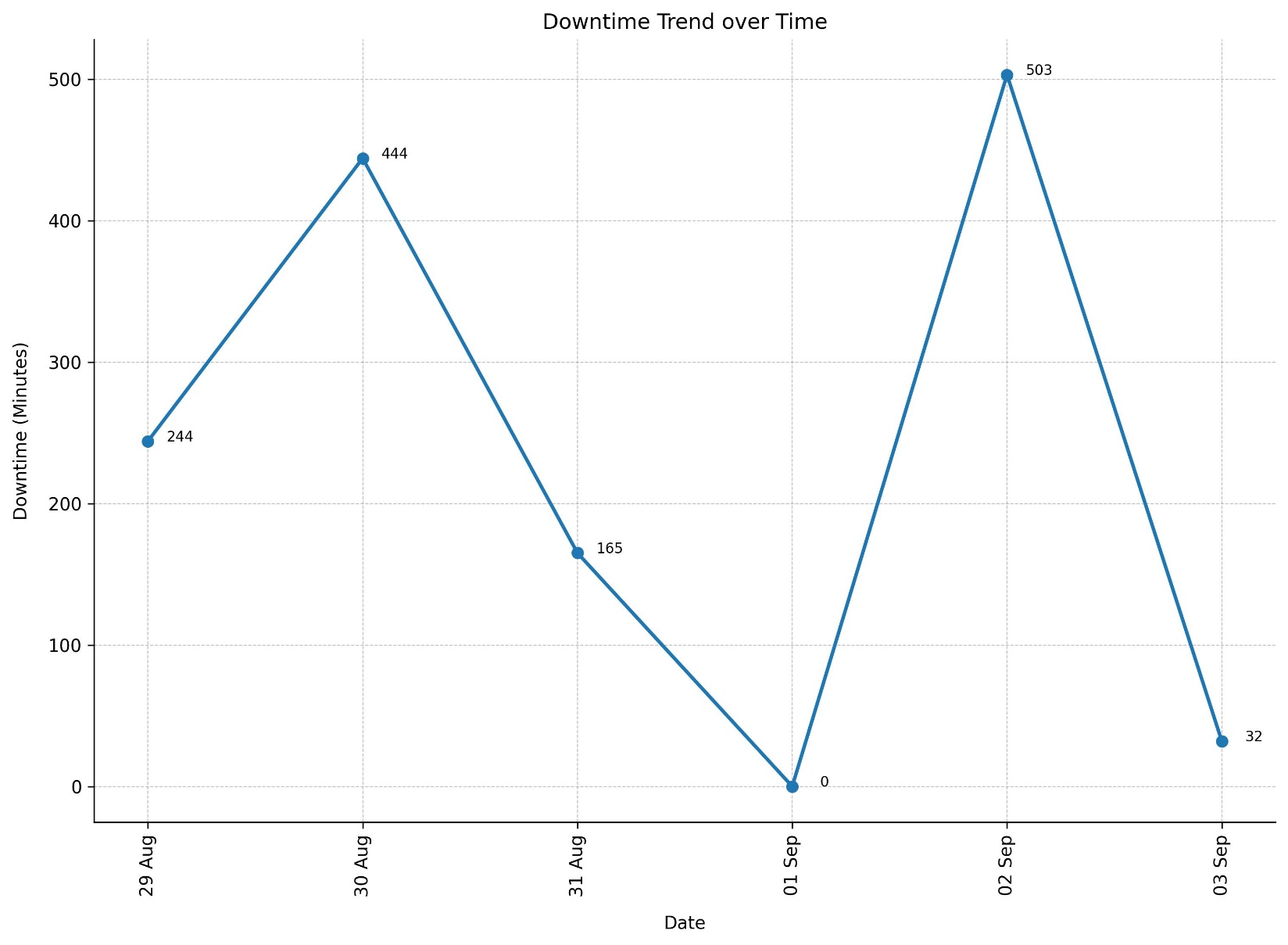


Figure : A line chart depicting time trend in downtime (generated using Python)

## Appendix II: Tableau Dashboard Screenshots

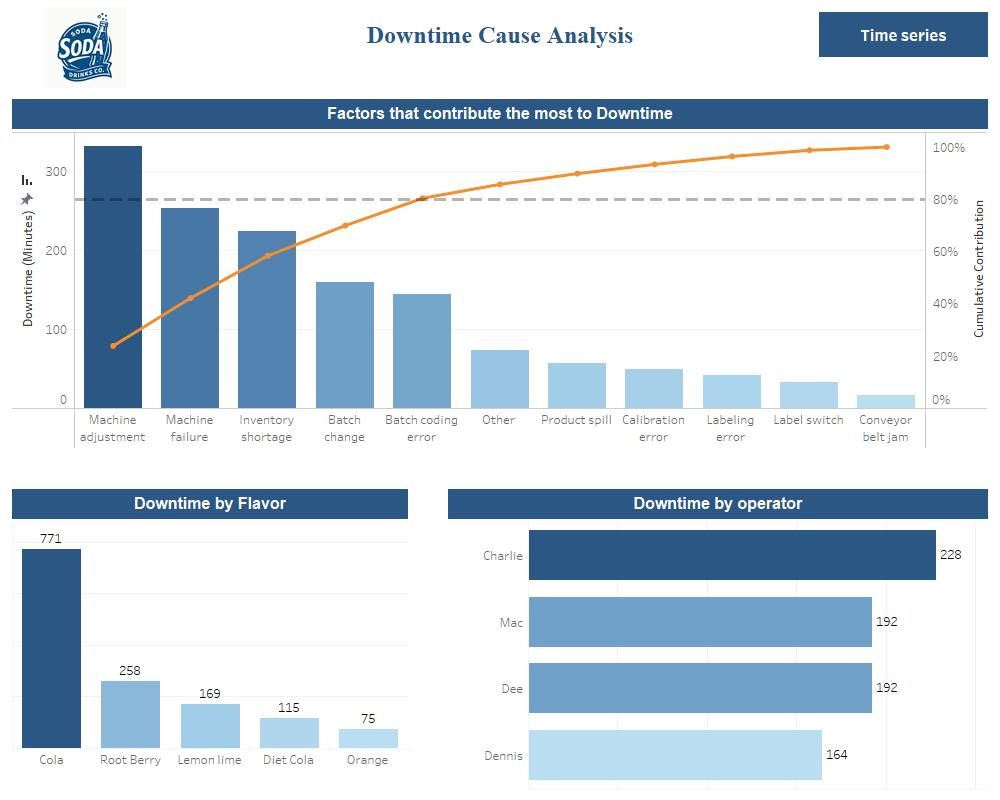


Figure : Tableau dashboard for downtime cause analysis

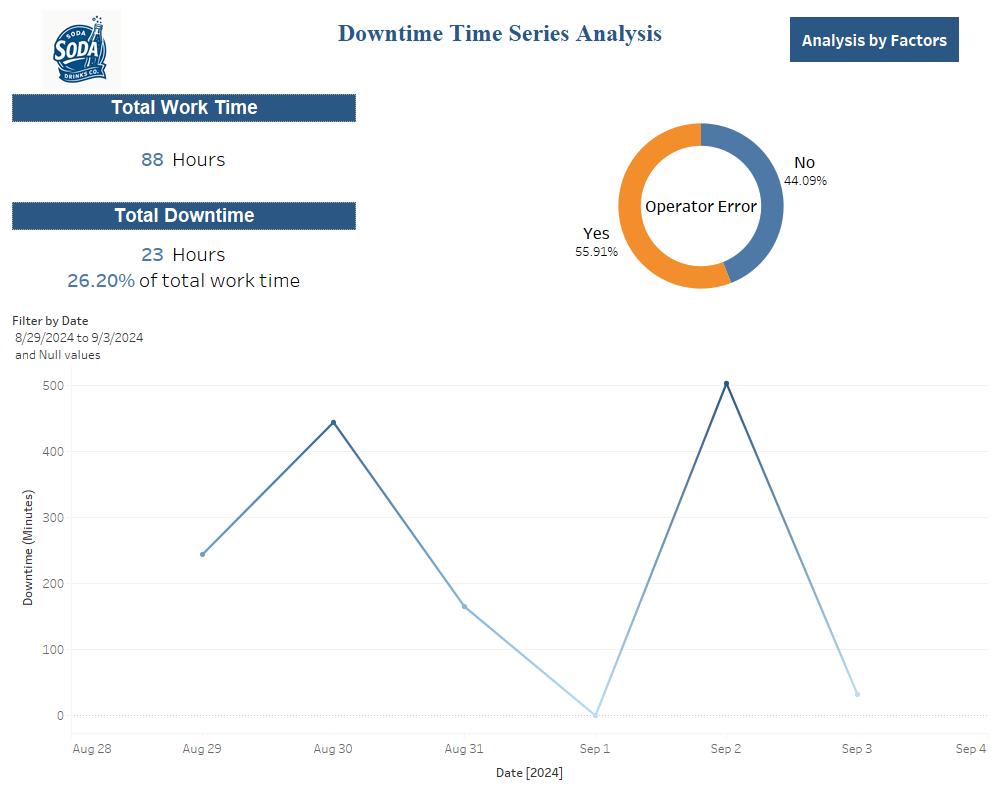


Figure : Tableau dashboard for time-series analysis of downtime