Predicting Airplane Manufacturing Quality and Performance Time

IST 718 | Sept. 14, 2019

# See the source imageSee the source imageImage result for tx trainer pictures

Rebecca Boyer, Justin Pate, Sulav Rupakheti and Rachel Shoop

# Project Overview

In large manufacturing environments, the need to predict performance trends and quality rework is imperative to maintain a schedule that meets the needs of the final assembly line of the customer. This need is abundantly present in Aerospace, a local Dallas company that builds large aerostructures for delivery to final line airplane manufacturing companies. Aerospace struggles with predicting both performance to budget trends and the effects that poor quality will have on manpower.

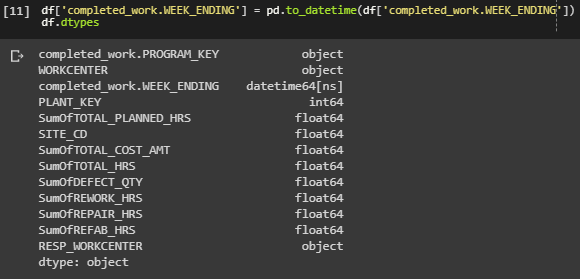
The company would greatly benefit from a time series prediction model that will improve the company’s ability to anticipate poor performance or poor quality instances. This will help inform decisions about manpower as a site or by individual programs or sections.

# Data

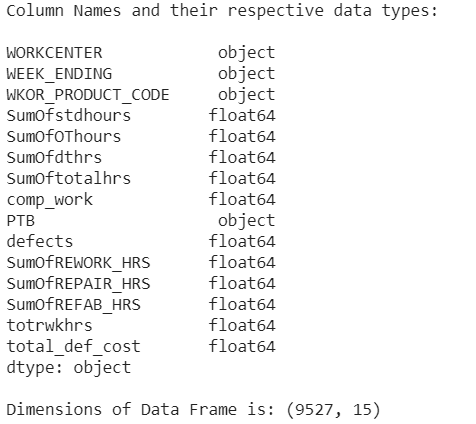
The data source is from the back end of the manufacturing system in an Oracle environment that houses time, standards, and quality data. However, the data needed to be cleaned up and transformed for ease of use with Python and in order to be run through Prophet (Figure 1, Figure 2, Figure 3).

# 

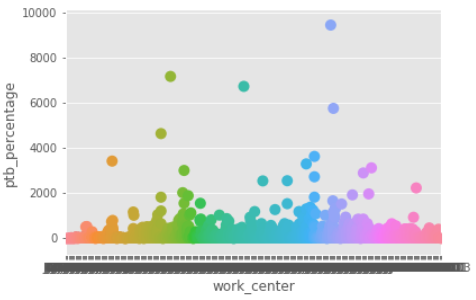
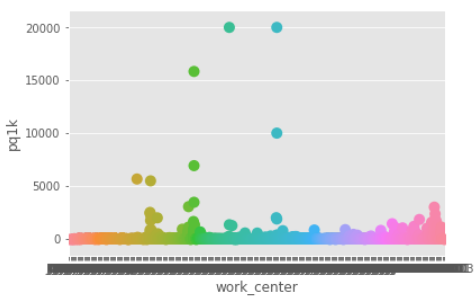
**Figure 1. Sample of raw data imported into Colab.** Missing data (“NaN”) can be observed in several rows and columns. Fifteen variables were imported with the raw dataset. Two of these variables (number of defects and completed work hours) were used to derive a sixteenth variable representing the number of defects per 1000 completed work hours.



**Figure 2. Python code and output showing data types for the dataset.** It was necessary to convert integers to float64 for downstream calculations.



**Figure 3.** **Time series plot for nine variables.**



# Objectives (Specifications)

It is pertinent in any manufacturing industry that high quality products are produced and that working hours are optimized in order to reduce cost and increase performance to budget (PTB). This goal is challenged, however, by variables that seem, at first glance, to be unpredictable. Unchecked, these variations can impact production and lead to delayed delivery schedules. In order to counter this threat, we identified the following objectives for our investigation.

1. Identify performance trends based on the number of people working and the planned hours completed creating a performance to budget variable, where **PTB** = **completed work** divided by the **sum of hours worked**. PTB values were presented as percentages.
2. Identify poor quality trends, where the number of **defects** divided by **planned hours completed** per **1000 completed work hours**.
3. Create a model that predicts product quality and performance seasonal variations that will allow the company to coordinate employees in order to avoid low-performance indicators and overcome the cost and burden of rework due to poor product quality.

## 

## Methods

## There were two main methods that we used to learn more about the trends in Aerospace’s productivity. First, we used various Python packages to observe and get to know the data better. As discussed above, this involved creating two new columns. Each project at Aerospace has a certain number of expected hours of labor. When a project is completed, these hours are totaled for the given week in the “Completed Work” column. The “Hours Worked” column, on the other hand, is a record of the actual number of hours spent by employees. Since the difference between these two is one of the primary focuses of this project, we created the PTB column using the completed work divided by the actual hours worked.

## The second derived column illustrates the relationship between the labor hours and the defects that needed to be fixed before the final product could be released to the customer. Since defects are not part of the business strategy, these require extra hours of labor to correct. However, by dividing the number of defects by the planned hours of labor, we hoped to normalize this data to make it more meaningful. We called this variable PQ1K. For both of these columns, we needed to make sure to account for the fact that some weeks did not have defects and/or completed projects. Many of the “NaN” columns in our data were really representative of a zero value.

# Once we had created this data, we could use packages in Python to get a basic idea of the relationship between our variables. This is what inspired us to create some time series predictions. We used Facebook’s Prophet to make this easier to make this simpler. We used this package to make predictions on total hours and on PQ1K.

# Results

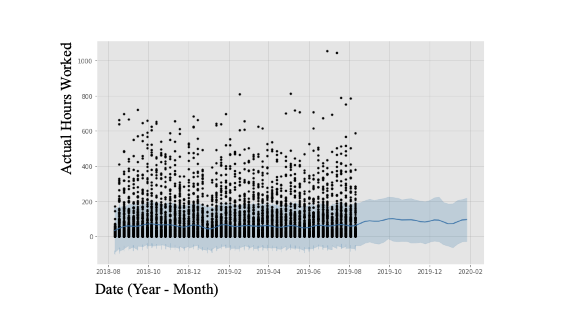
## Observations

1. Some Units and Programs are less active than others and should be considered eliminating from the data set. Descriptions of the variables included in this dataset are presented in the Glossary at the end of this document.
2. After reviewing data fields, some redundancy was found. The new data-frame will only include:

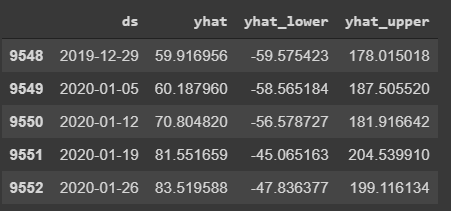
work center, week ending, product code, sum of total hours, completed work, and number of defects with a derivative for PTB = (completed work / sum of total hours) and (completed work/1000)/defects.

**Analyses**

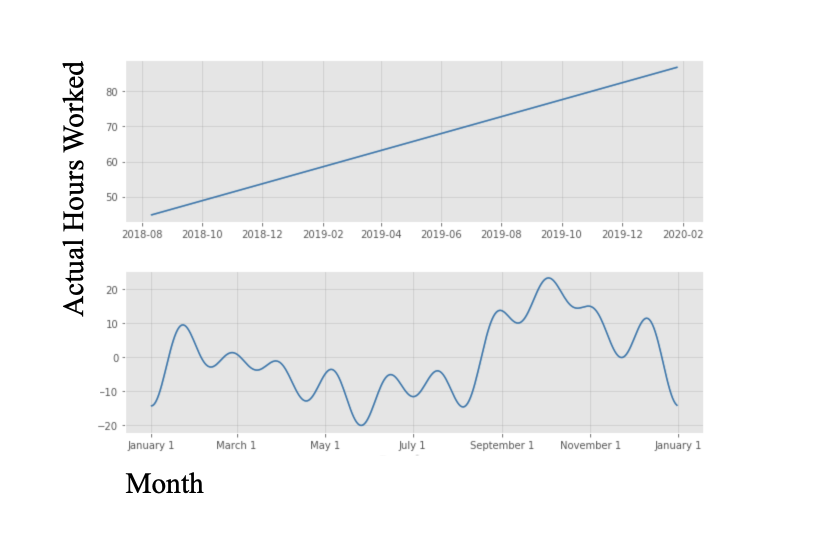
Based on initial Prophet model analysis of Actual Hours worked, there is a non-stationary and slightly upward trend (Figure 4). An initial observation indicates that sum of total hours will increase in January after a decrease in December. This decrease in December is likely due to employees taking reduced hours during the holiday season.



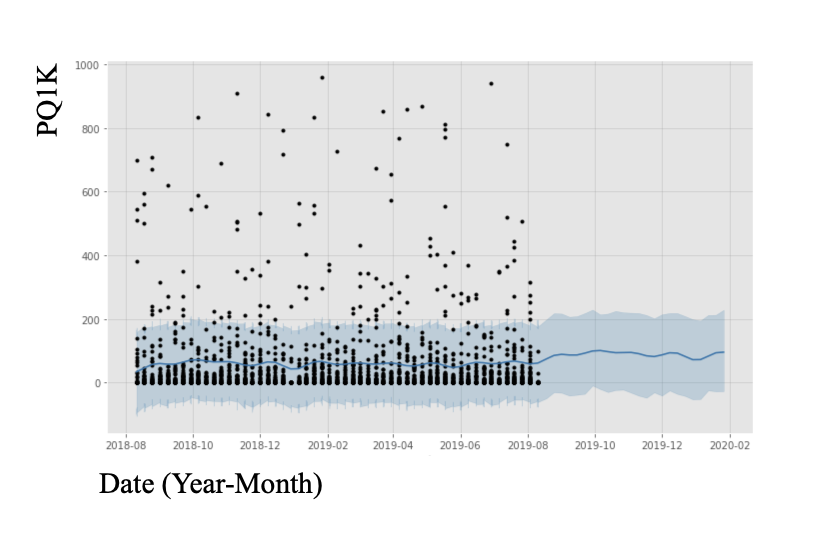
**Figure 4. Prophet forecast for actual hours worked**. Black points represent actual samples from the dataset. The dark blue line represents the mean predicted value for hours worked between September 2019 to February 2020. The light blue areas represent the confidence interval for this prediction.



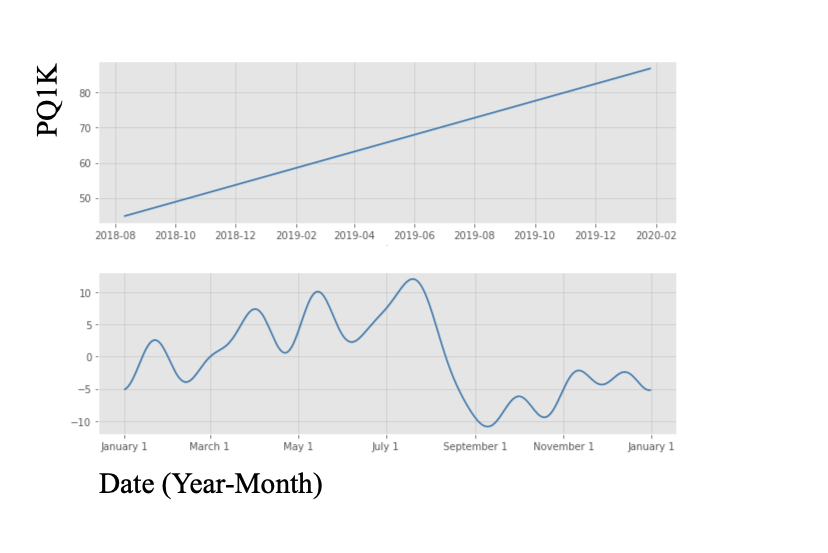
**Figure 5. Confidence interval estimates for the hours worked forecasted using Prophet.**



**Figure 6. Seasonality of hours worked illustrated using Prophet.** The top plot represents the overall trend predicted on a yearly scale. This shows that the actual hours worked has increased since 2018 and is predicted to increase into 2020. The bottom plot represents the seasonal trends in hours worked on a monthly scale.



**Figure 7. Prophet forecast for PQ1K**. Black points represent actual samples from the dataset. The dark blue line represents the mean predicted value for PQ1K between September 2019 to February 2020. The light blue areas represent the confidence interval for this prediction.



**Figure 8. Seasonality of PQ1K illustrated using Prophet.** The top plot represents the overall trend predicted on a yearly scale. This shows that PQ1K has increased since 2018 and is predicted to increase into 2020. The bottom plot represents the seasonal trends in PQ1K on a monthly scale.

# 

# Recommended Action

After considering the results of the analyses presented above, we recommend the following four steps to address the trends forecasted here.

1. Improve man-power planning. Managing schedules to increase the number of people working will counter the predicted seasonal decline in work hours and reduce the predicted number of defects.
2. Manage vacation schedules. Efficiently coordinating vacation schedules will allow the company to ensure that essential positions in the production line are covered throughout the year, reducing the number of defects that occur seasonally.
3. Continue to apply the above two steps periodically in order to predict and respond to future dips in performance and product quality. This will help the company to refine these performance measures over time.
4. Survey employees. Gathering data about the time of year employees tend to take vacation time will allow the company to better coordinate schedules and plan production accordingly.

**Conclusion**

The goal in any manufacturing company is to predict the profit and loss for forward planning. A predictable performance and quality time trend could be the missing link that this aerostructures company needs to succeed. The time series predictions will be used to predict fluctuations in efficiency and quality of production. Additionally, use of this real-world data set will have implications not only for the process of exploring, cleaning, manipulating, and analyzing the data– it will also provide a valuable scenario for interpreting the data in terms of production management and planning.

# 

# Appendix 1. Glossary of data with data types and descriptions of variables.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Data Type** | **Description** |
| WORKCENTER | object | Plan function/Business Unit |
| WEEK\_ENDING | datetime64 | Time stamp indicating was aggregated within |
| WKOR\_PRODUCT\_CODE | object | Product Identification. This defines the end item deliverable to the customer. |
| SumOfstdhours | float64 | Standard hours worked |
| SumOfOThours | float64 | Overtime hours worked |
| SumOfdthrs float64 | float64 | Double time hours worked |
| SumOftotalhrs float64 | float64 | Total hours worked |
| comp\_work float64 | float64 | Budgeted hours completed |
| PTB\_percentage | float64 | “Performance To Budget”, (Completed hours/Worked Hours) |
| defects | float64 | Number of Quality issues that |
| SumOfREWORK\_HRS | float64 | Touch labor Hours spent correcting quality issues back to engineering. |
| SumOfREPAIR\_HRS | float64 | Touch Labor hours spent performing non-engineering repair |
| SumOfREFAB\_HRS | float64 | Touch Labor hours re-building assembly for replacement |
| totrwkhrs | float64 | Total hours for defect correction |
| total\_def\_cost | float64 | Total $ for defect repair or replacement |
| pq1k | float64 | Derived value calculated as the number of defects per 1000 hours of completed work. |

# 