

## Data preparation and exploration:

The dataset was prepared with 10,000 observations over the period from January 1, 2022, to December 31, 2022. For the features, we have considered the values in the range:

- shifts = ['Morning', 'Afternoon', 'Night']
- lines = ['A', 'B', 'C']
- products = ['A', 'B', 'C', 'D']
- Batch = range(0,10)
- cycle\_time = range(100, 150)
- downtime = range(50, 100)
- maintenance\_cost = range(50, 150)

For the Features Yield, Scrap, and Rework, have used Uniform Distribution.

- Yield: have values between (0.85, 0.98) distributed uniformly, mean(0.914777)
- Scrap: have values between (0.01, 1 - yield\_data) distributed uniformly, mean(0.047387)
- Rework: (1 - yield\_data - scrap\_data, 2), mean(0.037836)

As the data was prepared considering all the conditions, the data is clean and there is no need to preprocess.

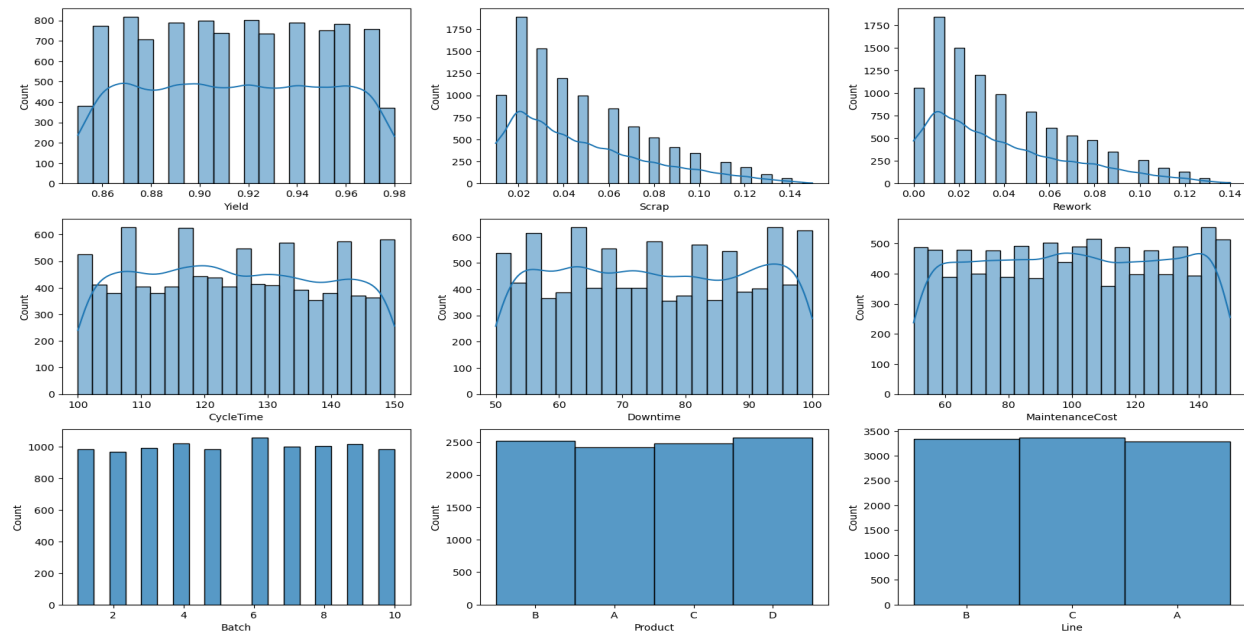


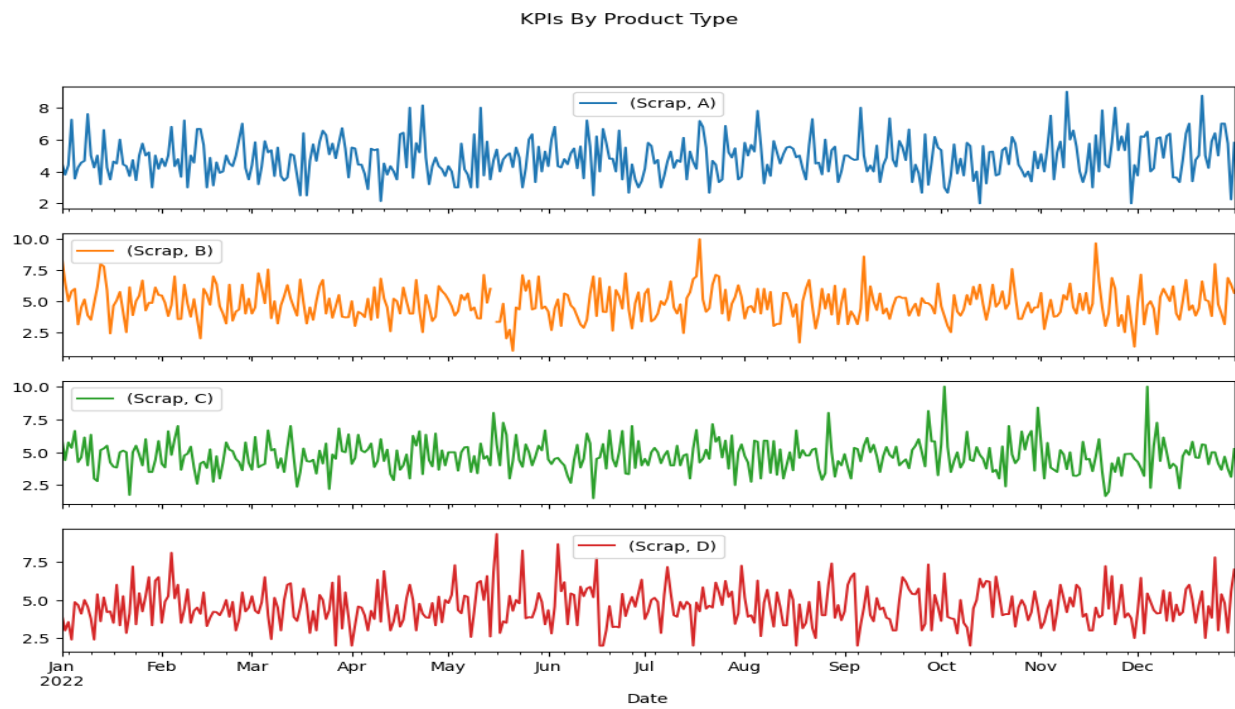
Fig 1: Visualize the distributions of the variables

## Data Analysis and Results:

### KPI 1: Average Scrap per Product by Date:

This KPI calculates the mean scrap value for each product on each date. The KPI provides insight into the overall quality of the products being produced over time. This KPI can

be used to identify trends and patterns in scrap rates, to monitor the effectiveness of quality control measures, and to identify areas where improvements can be made to reduce scrap rates and improve overall product quality. It can also be used to compare scrap rates across different products.



We have used Auto regression for forecasting the Scrap KPI for the next 30 days. The R2 score of the model is 0.70. And based on the results of Dickethe y-Fuller test, it can be observed that the chosen variable are stationary.

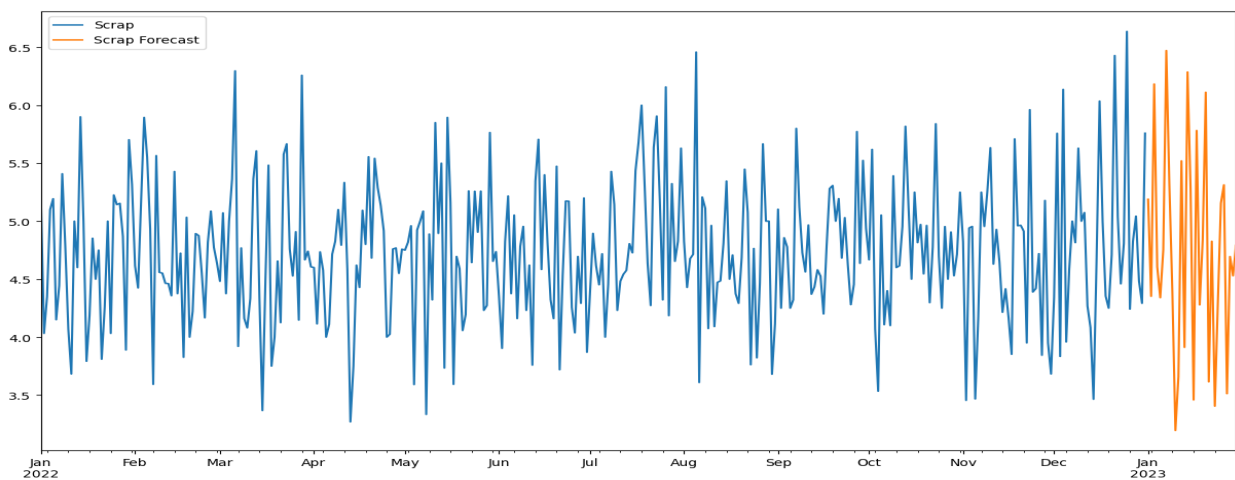
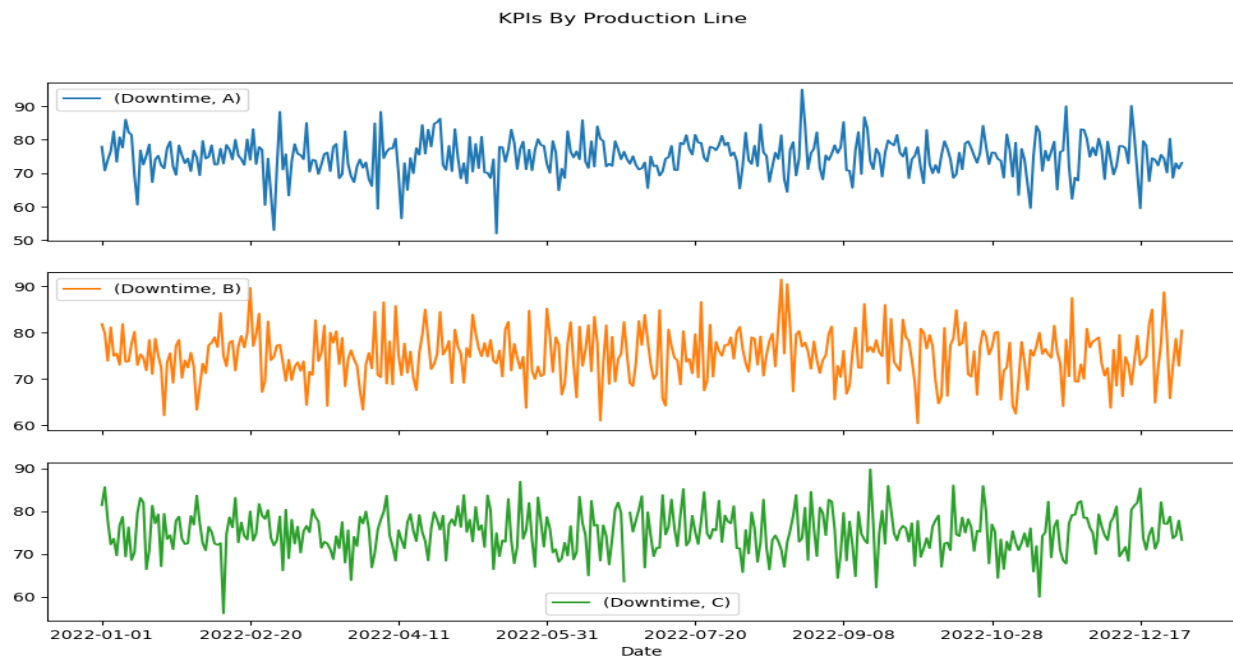


Fig 3: Forecasting the values of Scrap KPI for the next 30 days.

## KPI 2: Average Downtime by Line and Date:

The KPI calculates the average downtime for each line on each day. The KPI Identifies which lines have the highest downtime on certain days or over a certain period of time.



We have used Auto regression for forecasting the Scrap KPI for the next 30 days. The R2 score of the model is 0.70. And based on the results of the Dickethe y-Fuller test, it can be observed that the chosen variable are stationary.

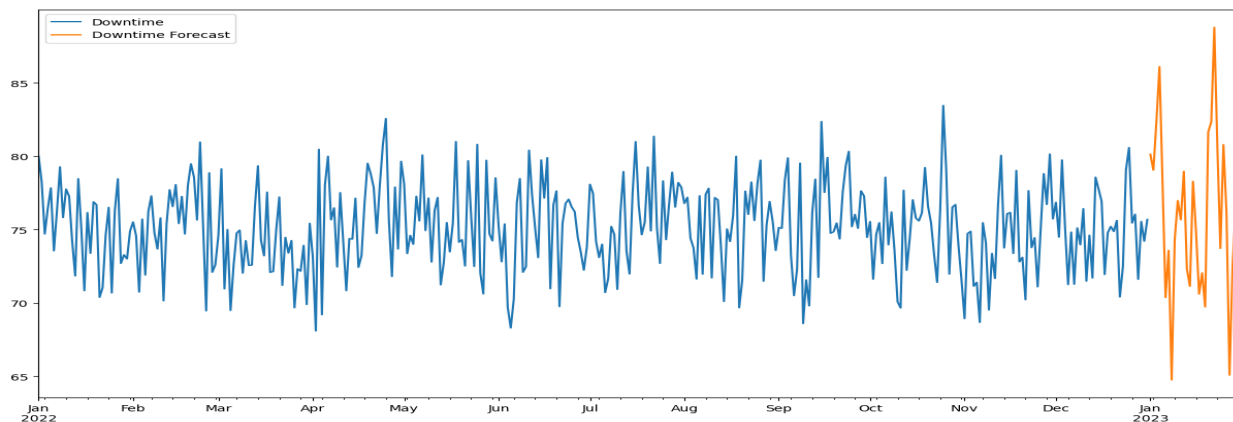


Fig 3: Forecasting the values of Downtime KPI for the next 30 days.