****

**Group 6 - Hexagon**

**DATA CLEANSING AND ASSUMPTIONS ADOPTED:**

ShiftID on all tables means the following, first digit is LineID and second digit is TeamID, 11 is Line 1 Team 1, 23 is Line 2 Team 3 etc.

**For DBDowntime**

Created new column ShiftID (concatenate LineID + TeamID)

Downtime seconds (changed to minutes in slides) calculated from this table

**For DBMachineUtilization**

Loaded into pandas, Converted Date to DateTime, Created new column for Date only.

Created new column ShiftID (concatenate LineID + TeamID)

We noticed that each LineID has a unique 'ProductionFBETheoretical' value. First, we filled the Null values with zero. Then. replaced all the rows where ‘ProductionFBE’ existed & 'ProductionFBETheoretical' was 0 and 'ScrapFBE' was 0. There were 3 cases (1 for each LineID). So we filled the values based on the unique value per LineID.

How we handled the missing values. Enlarge font below to see.

Filled Theoretical production amounts into missing columns based on the following conditions (Python Pandas code after loading the dataframe and fixing the null to zero values):

machineutil.loc[(machineutil['ProductionFBE']>0) & (machineutil['ProductionFBETheoretical']==0) &( machineutil['ScrapFBE']==0) & (machineutil["LineID"] == 3), "ProductionFBETheoretical"] =

machineutil.loc[(machineutil['ProductionFBETheoretical']>0) & (machineutil["LineID"] == 3)]['ProductionFBETheoretical'].iloc[0]

machineutil.loc[(machineutil['ProductionFBE']>0) & (machineutil['ProductionFBETheoretical']==0) &( machineutil['ScrapFBE']==0) & (machineutil["LineID"] == 2), "ProductionFBETheoretical"] =

machineutil.loc[(machineutil['ProductionFBETheoretical']>0) & (machineutil["LineID"] == 2)]['ProductionFBETheoretical'].iloc[0]

machineutil.loc[(machineutil['ProductionFBE']>0) & (machineutil['ProductionFBETheoretical']==0) &( machineutil['ScrapFBE']==0) & (machineutil["LineID"] == 1), "ProductionFBETheoretical"] =

machineutil.loc[(machineutil['ProductionFBETheoretical']>0) & (machineutil["LineID"] == 1)]['ProductionFBETheoretical'].iloc[0]

We assume that ProductionFBE within DBMachineUtilization contains all the production pieces. We use that to calculate Machine Utilization because it contains the highest quantities of Production amongst all the tables even if we sum up the quantities from the other tables.

Total production pieces calculated by Taking a sum of the ProductionFBE column.

Machine Utilization calculated by Taking sum of ProductionFBE less the sum of ScrapFBE which is then divided by ProductionFBETheoretical

**For DBProduction**

This table show individual pieces produced with their ID’s this information has already been captured in our other tables. Only 21.98% of the pieces in this table have been recorded out of all the pieces produced in DBMachineUtlization.

**For DBRightFirstTime**

Created new column ShiftID (concatenate LineID + TeamID)

We assume that this table recorded only the pieces that were right the first time, we believe there is a lot of missing data for this table so we considered this table as a “sample size” of total production (Population) to calculate the following metrics.

ReworkRate calculated by taking the sum of ReworkFBE over sum of ProductionFBE

RightFirstTime calculated by taking the sum of ProductionFBE less the sum of ScrapFBE less the sum of ReworkFBE over the sum of ProductionFBE

ScrapRate calculated by taking the sum of ScrapFBE over the sum of ProductionFBE

**New Tables Created**

Time - Created a table of times to match time across tables

Calendar - Created a calendar table to match dates across tables

ShiftID - Created a ShiftID table to match ShiftID across tables

LineID - Created a LineID table to match LineID across tables

TeamID - Created a TeamID table to match TeamID across tables

With these tables we are able to easily merge across tables and able to create a correlation between the DBMachineUtilization[‘ProductionFBE’] and DBProduction['ScrapFBE']/[‘ReworkFBE’] columns.

**Future**

We could use more data points. 5 days’ worth of data is not enough.

We need to look at the data collection process because there is a mis-match between the data collected. Example, the total number of pieces produced in the DBMachineUtilization table don’t match the totals from the other tables.

Some analysis of ShiftID shows that there is a trend between ProductionFBE and Scrap and Rework rates due to different teams and shifts. Further investigation could prove fruitful. See PBIX Appendix “Trend Between Production and Rework/Scrap Rate Within DBRightFirstTime By ShiftID” and “Trend Between Total Production and Rework/Scrap Rate By ShiftID” scatter plots.

**ANSWERS**

1. KPI are listed in the dashboard
2. Charts are listed in the dashboard
3. Filters are at the top of the dashboard
4. (i) Weak negative correlation between (MachineUtilization) ProductionFBE and (Production) ScrapFBE for Date which is -0.3

Graph in Dashboard

(ii) Weak negative correlation between (MachineUtilization) ProductionFBE and (Production) ReworkFBE for Date which is -0.2

Graph in Dashboard

We want to add that 5 data points are not enough to fully determine the correlation, we can see what we think might be outliers but it is not easy to tell with this little data.