Determining Cost Breakdown from Shop Floor Data
DIRECT MATERIALS

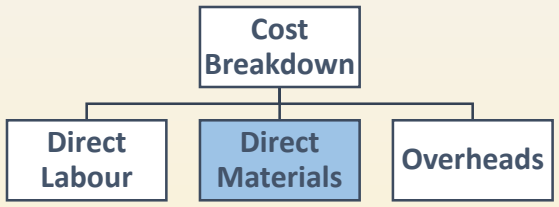
INTRODUCTION

Star Asia Trading Pte Ltd (Star Asia) aims to build Singapore's first urban smart shoe factory. This future factory space seeds big dreams, and hopes to bring designers, manufacturers, as well as people like us, closer to where our shoes are crafted. To that end they intend to fully utilize data analytics to fully understand how different components of their factory relate to their end product - the humble shoe.

Star Asia partnered with SMU-X to engage teams of undergraduates to help them implement data analytics on certain areas of their operations; these areas were divided based on how Star Asia accounted for their cost to manufacture per pair of shoes.

ANALYSIS

Cost Breakdown



Our group will be focusing on the analysis of **Direct Materials**

Shop Floor Data

End of Line Rate	•Number of Shoes made a day
Plant Summary Data	•Manpower hours, Machine Breakdown Rate
Defect Count	•Based on different defect types
KPIs	•Targets set by the company which affect shop floor data

Exploratory Data Analysis

Daily DM

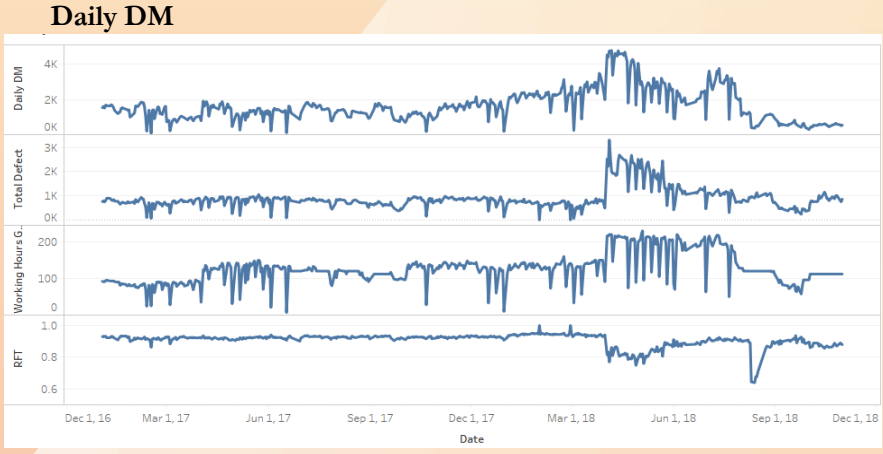


Figure 1: We chose a few variables that, with our domain knowledge, might be correlated with direct materials cost. It can be observed from the visualization that direct materials, total defect and working hours gross seem to share ***similar pattern*** while RFT (% of non-defect products) has an ***inverse relationship***.

Defect Count vs Output

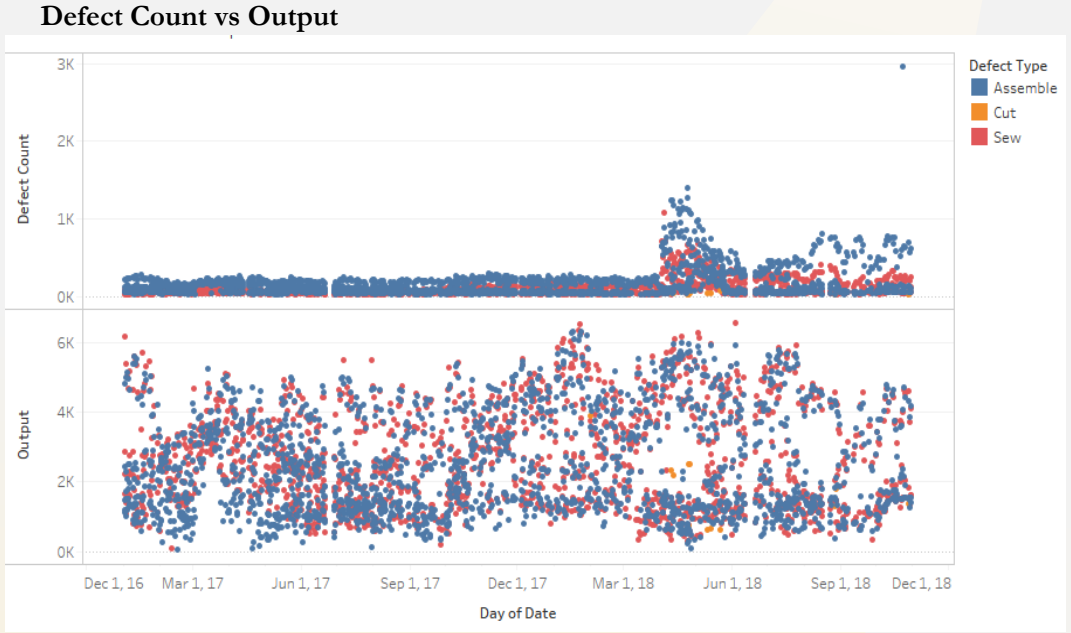


Figure 2: We plotted output against defect count with defect type filtered by different colors and plant and brand included in detail. It can be observed that while output seems to remain constant during the period, there seems to be a surge in defect count in April 2018. These are some ***outliers*** that the management might wish to further investigate.

Model Performance

We developed different models using algorithms to predict Daily and Monthly DM cost and get the following results:

			ME	RMSE	MAE	MPE	MAPE
Daily DM Data	PCA	Regression Trees	81.71	482.62	352.36	-2.26%	23.98%
		Multiple Linear Regression	24.57	505.78	420.59	-5.25%	38.62%
		Random Forest	41.83	343.43	256.55	-2.00%	17.43%
	Non-PCA	Multiple Linear Regression	53.55	400.25	316.42	-1.46%	26.74%
		MLR with Variable Selection	46.72	391.69	313.28	-1.94%	26.66%
		Random Forest	54.38	248.76	189.04	-0.59%	13.78%
Monthly DM Data	PCA	Regression Trees	-5240.57	12721.20	10450.73	-31.07%	40.35%
		Muletiple Linear Regression	-5183.11	10280.65	8863.46	-24.94%	32.48%
		Random Forest	-5727.46	9245.45	7134.35	-26.75%	29.29%
	Non-PCA	Random Forest	-3619.19	10479.74	8253.22	-23.42%	31.83%

Model Evaluation

Overall, the random forest algorithm with the original daily master sheet returns the most accurate prediction. We believe that accuracy should be prioritized as the most important factor. With strong predicting power the algorithm can provide the management with an approximation of direct materials cost given the shop floor data available. This information will help Star Asia in negotiating a more favorable contract with suppliers, ensuring a healthier profit margin for each pair of shoes.

RECOMMENDATION

1. Algorithm model:
   Random Forest

2. Proposal for future data collection
   - Data Schema
   - Using sensors on factory lines to automate data collection

3. Recommendation to change the data collection processes
   - Automatic collection of data
   - Automatic analysis through machine learning methods
   - Greater granularity of data into shoe model, shoe size, and shoe colour, leading to more data collected
   - Database Management System