Team B

CIS 450

4/29/21

Final Paper: National Industries of the Blind

Introduction:

NIB is a 501(c)(3) nonprofit organization incorporated in the state of New York, with headquarters in Alexandria, Virginia. NIB was created following the passage of the Wagner-O'Day Act in 1938, which mandated the federal government purchase products made by nonprofit agencies (NPAs) serving people who are blind. NIB is the largest employer for people who are blind. Since 1938, National Industries for the Blind (NIB) has focused on enhancing the opportunities for economic and personal independence of people who are blind, primarily through creating, sustaining and improving employment. NIB and its network of associated nonprofit agencies are the nation's largest employer of people who are blind through the sale of SKILCRAFT® and other products and services provided by the AbilityOne® Program, established by the Javits‐Wagner‐O'Day Act (41 USC 46-48c). Despite continued gains in employment, people who are blind remain one of the nation's greatest untapped labor resources. In fact, 70 percent of working‐age Americans who are blind are not employed. This statistic fuels NIB's mission to create, sustain and improve employment opportunities for people who are blind.

Situation:

The challenge NIB faces is complex due to the nature and scale of their business. NIB has a diverse product portfolio with 5 major Line of Businesses (LoBs), 22 Sub Line of Businesses (SLoBs), and over 7,000 products and services. Each year, NIB and its agencies work together on several dozen new projects with the purpose to add successful new products to the current product portfolio. Each new project could encompass one or more products that are similar in nature. Many of the new projects introduced failed to deliver solid returns, as measured by project payback time, costing NIB and agencies valuable resources. In an effort to identify opportunities to remedy this, the project's objective is to utilize past actua l projects' cost, sales, and payback time data, analyze the relationships between cost components of new projects and the projects' payback times, develop and test statistical models, at both aggregate

(across all LoBs) and individual LoB level, and use the final statistically sound models to predict the potential success or failure of new projects in the future.

The scope of the project consists of creating statistical analysis of cost attributes of new projects and their relationships to the project's payback period. The team must create statistical models that evaluate a project's chance of success or failure based on the cost components at both the aggregate level and the LoB level. There is an expectation to create Excel-based tools and macros that NIB currently doesn't use.

Analytical Techniques:

The team spent multiple weeks analyzing the data within Excel, by starting with simple questions, creating charts and graphs, and finding common statistics within the projects. Diving into the dataset, the timeline of the projects are between 2012 to 2017. Starting with the cost sheet the NPA is the name of the non-profit organization. The NSN is the specific product code within a project. The project's unique identifier is the PCN#. The dataset also includes LoB and subLob. The cost components included in the dataset are as follows: unit material cost, standard direct labor unit cost, variable overhead unit cost, fixed overhead unit cost, amortization unit cost, g&a unit cost, NIB fee unit cost, marketing unit cost, and other unit cost. The margin per unit is related to the individual non-profit organization and not NIB directly, so this attribute would not be used. The FOB origin price is how much the product costs consumers. When analyzing the data from the payback time sheet, the NSN payback time represents how long the project takes to complete. Majority of the failed projects took over 10 years. Then the most important variable is whether the project was a success or a failure.

During the preliminary meetings with NIB, the project stakeholders discussed which attributes are most important, and how they would like the data to be represented. The team needed to create cost ratios, in order to compare projects to one another. To form the ratios, the group divided each unit cost component by the FOB origin price. Now the group can use these margins to run a more accurate analysis and build a model based on the margins. Doing this in Excel, the formulas and tools that were consistently used were the VLOOKUP, IF, and simple macros.

The team decided to divide and conquer when it came to creating analysis and pivot tables. Team members presented a variety of pivot tables varying in scope, drilling down and

rolling up, *Pivot 1*, is an example of the preliminary analysis pivot tables. Once pivot tables were created, the group formed pivot charts to make the data more visually appealing. These visualizations were starting points for our analysis within Excel, R and Orange. The team used these visuals as clues on how to better understand the models they were making.

At the beginning of the project timeline, the team focused solely on the individual product level (PCN#), however, as time progressed the group formed margins of the data at the project level. Furthermore, to build a model based on the LoB or subLob the team reviewed the data to get a better understanding.

It's important to note that the team grouped the data by product successes, product failures that were less than 10 years, and products that failed within more than 10 years. There are only 121 successful products within the whole dataset, and 1379 products in total. After consulting with NIB stakeholders the group decided to disregard the products that failed within 10 years, creating a dataset that consisted of true successes and failures.

Challenges Faced:

One of the beginning challenges the team faced was fully understanding the dataset at each level. It took about two weeks of meetings for the group to analyze the cost components at just the product level, and understand what they mean. Certain cost components included in the dataset weren't relevant to NIB predicting a success or failure.

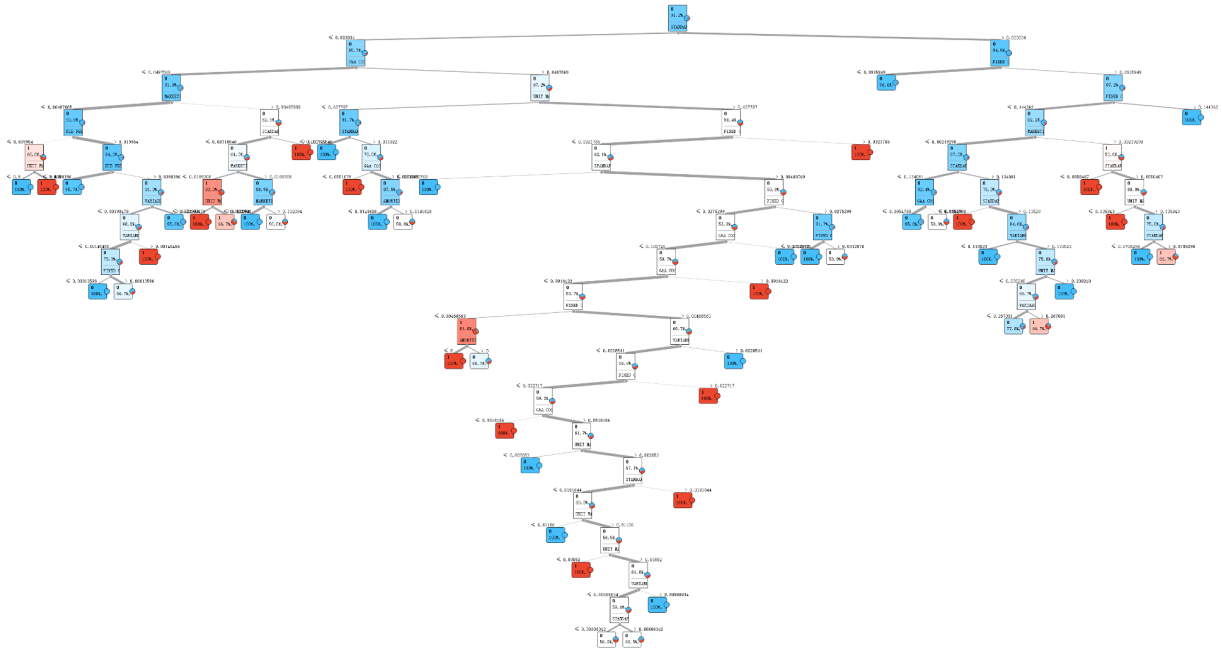
During the modeling phase the group encountered some technical difficulties with coding and improving the accuracy of each model. The group created a decision tree in R to see the relationships between the cost margins and failed products. With the limited number of successes within the dataset, it became difficult to understand the model, because of overfitting. R was producing a confusion matrix with 100% accuracy, which is shown in the table below.

Total Observations in Table: 458

	predicted success		
actual success	0	1	Row Total
0	417	0	417
	0.910	0.000	
1	0	41	41
	0.000	0.090	
Column Total	417	41	458

The group also found that calculating cost ratios at the project level was more difficult than projected. Choosing to format all the data within Excel became much more time consuming and disorganized. It was challenging to have all the data within one workbook, at the product, project, and LoB levels. Next time, the team would prefer to format the data within SQL or separate the data into multiple Excel workbooks.

We also built models through Orange. We built Decision Tree model, Logistic Regression model and Neural Network model in Orange, however, we can only see our model performance like Accuracy and ROC curve. We cannot see the detailed data in our model.



Test and Score Thu Apr 29 21, 18:47:48

Settings

Sampling type: Stratified 10-fold Cross validation
Target class: 1

Scores

Model	AUC	CA	F1	Precision	Recall
Tree	0.6342143467643467	0.9072463768115943	0.24705882352941178	0.42857142857142855	0.17355371900826447

Test and Score Thu Apr 29 21, 18:48:21

Settings

Sampling type: Stratified Shuffle split, 10 random samples with 70% data
Target class: 1

Scores

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.6200161243037232	0.9132530120481928			

Test and Score Thu Apr 29 21, 18:49:04

Settings

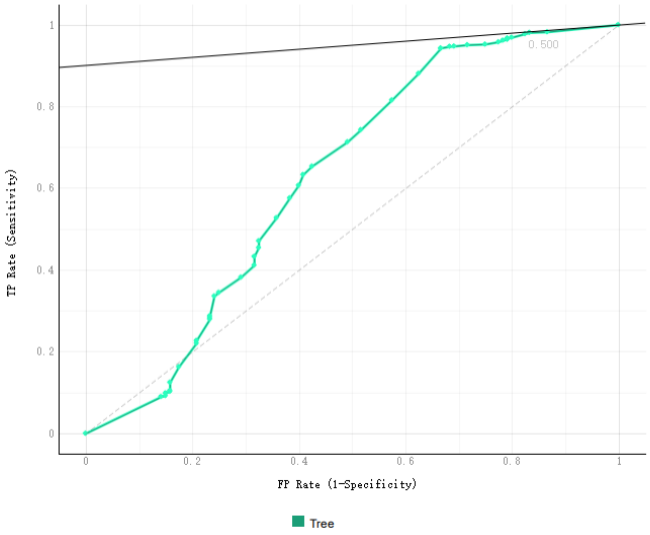
Sampling type: Stratified Shuffle split, 10 random samples with 70% data
Target class: 1

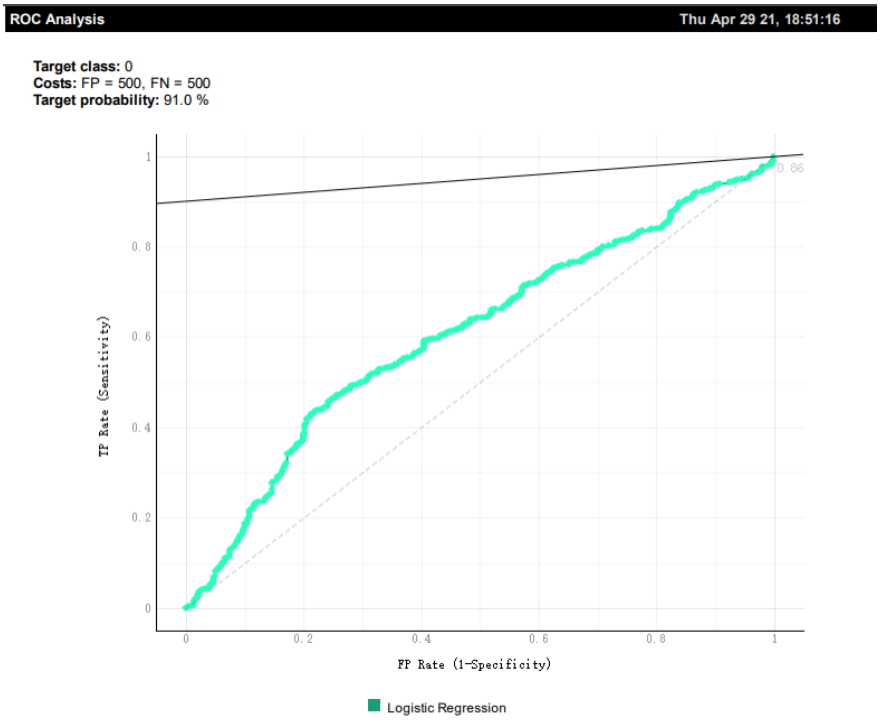
Scores

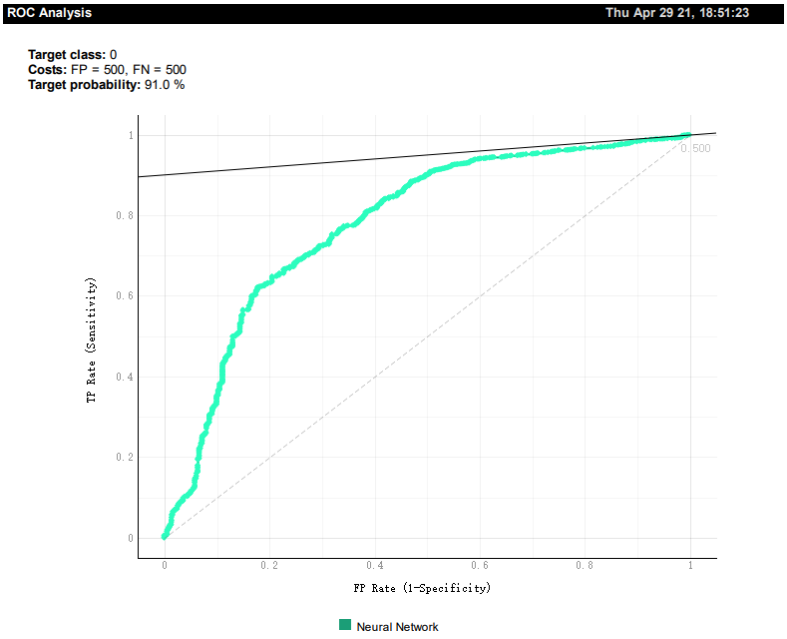
Model	AUC	CA	F1	Precision	Recall
Neural Network	0.7816915860451481	0.908433734939759	0.04040404040404041	0.2222222222222222	0.022222222222222223

ROC Analysis Thu Apr 29 21, 18:51:09

Target class: 0
Costs: FP = 500, FN = 500
Target probability: 91.0 %







The above pictures indicate our models are successfully built and had good performance. However, we cannot see which factor affects the final results in each model.

Visualizations:

The visualizations were created in Excel and Tableau. The team created multiple pivot tables, pivot charts, and various graphs to better showcase the data. These visuals were created with the intention of providing a stronger understanding about the relationships between cost margins and failed and successful projects.

LoB	Sum of Success
Commodity	38
Niche	57
Textile	26
Grand Total	121

Pivot 1

Count of Success (< 7 Years Payback Time) / Failure	Column Labels			
Row Labels	Commodity	Niche	Textile	Grand Total
Failure	162	48	13	223
Success	21	17	7	45
Grand Total	**183**	**65**	**20**	**268**

	Commodity	Niche	Textile	Grand Total
Success Rate	13.0%	35.4%	53.8%	20.2%

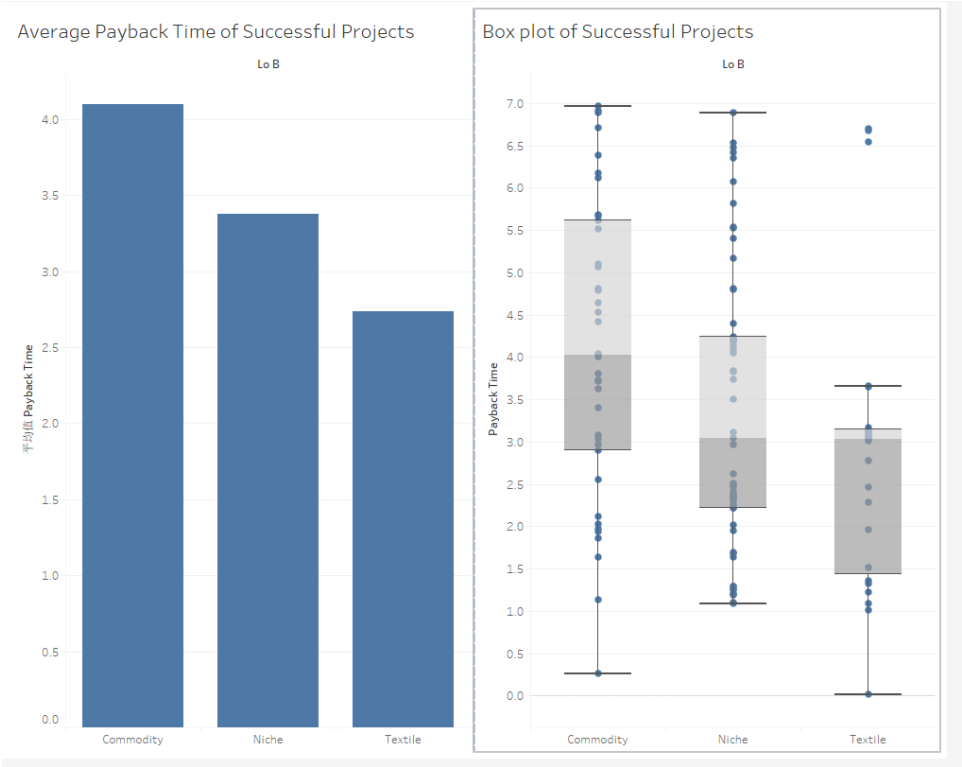
Pivot 2

LoB Success	Average of Unit_Material	Average of Standard_Direct_Labor	Average of Variable_Overhead	Average of Fixed_Overhead	Average of Amortization	Average of G_A_Cost	Average of NIB_Fee	Average of Marketing
Commodity	**0.620054703**	**0.059028263**	**0.056662587**	**0.042598271**	**0.002880431**	**0.046838685**	**0.038958501**	**0.001826918**
0	0.619988715	0.05961919	0.057626251	0.042418319	0.002898987	0.045950072	0.038989528	0.001380862
1	0.621428298	0.04672765	0.036603164	0.046344108	0.002494155	0.065335879	0.03831266	0.011111908
Niche	**0.659021309**	**0.032382836**	**0.034459695**	**0.031137478**	**0.000960654**	**0.041492159**	**0.038476104**	**0.002400196**
0	0.658300889	0.033729597	0.035622055	0.031702583	0.000570247	0.040564633	0.038633082	0.00205742
1	0.662964664	0.025011096	0.028097304	0.028044268	0.003097619	0.046569144	0.03761686	0.004276441
Textile	**0.651987025**	**0.074647999**	**0.079858281**	**0.053524425**	**0.001153516**	**0.045728607**	**0.037116714**	**0.001111934**
0	0.633816433	0.081221962	0.089854145	0.056358904	0.000506354	0.043312522	0.037322185	0.001327683
1	0.74563546	0.040766808	0.028341134	0.038915955	0.004488888	0.058180734	0.036057752	0
Grand Total	**0.634405108**	**0.053628407**	**0.053362472**	**0.040771431**	**0.002155317**	**0.045255121**	**0.038610423**	**0.001898451**

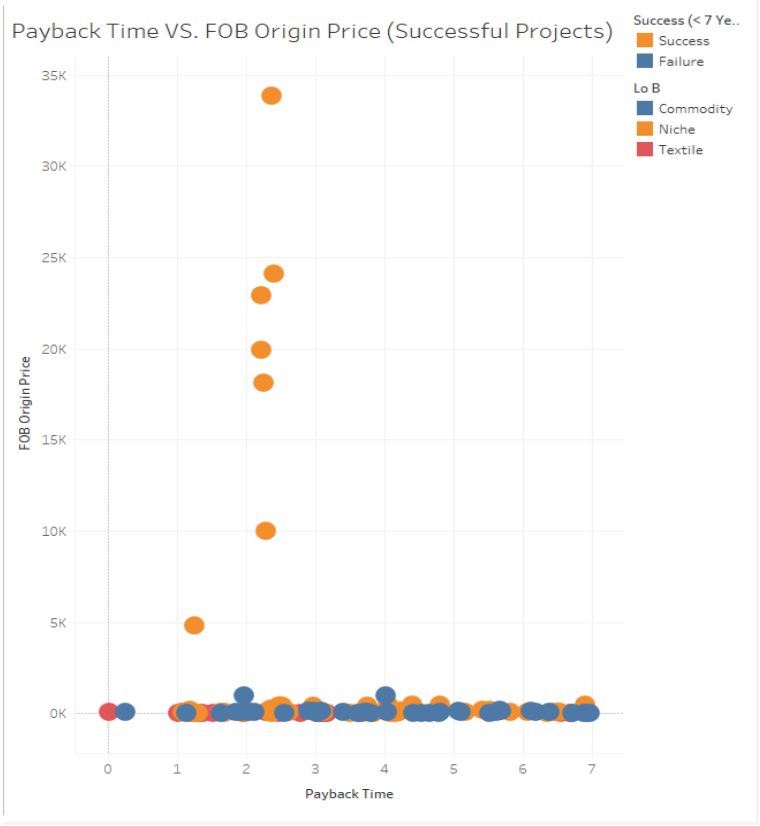
Pivot 3

Row Labels	*Sum of UNIT MATERIAL COST*	*Average of UNIT MATERIAL COST*
Commodity	***53785.471***	***63.95418668***
C-Food Service	*537.03*	*35.802*
C-Medical	*49.32*	*12.33*
Household Item	*6.41*	*3.205*
JanSan	*4298.63*	*47.76255556*
MRO	*38901.31*	*93.96451691*
Office Products	*9891.981*	*34.11027931*
Writing Instruments	*100.79*	*3.876538462*
Niche	***197486.04***	***525.2288298***
Hazardous Materials	*671.33*	*134.266*
Industrial, Organizational and Tools	*144804.41*	*2454.312034*
N-Food Service	*43.38*	*10.845*
N-Medical	*51307.92*	*172.7539394*
Personal Safety	*659*	*59.90909091*
Textile	***4041.83***	***24.9495679***
Organizational Clothing	*1229.54*	*21.19896552*
Organizational Equipment	*124.83*	*13.87*
Uniforms	*2687.46*	*28.28905263*
Grand Total	***255313.341***	***185.1438296***

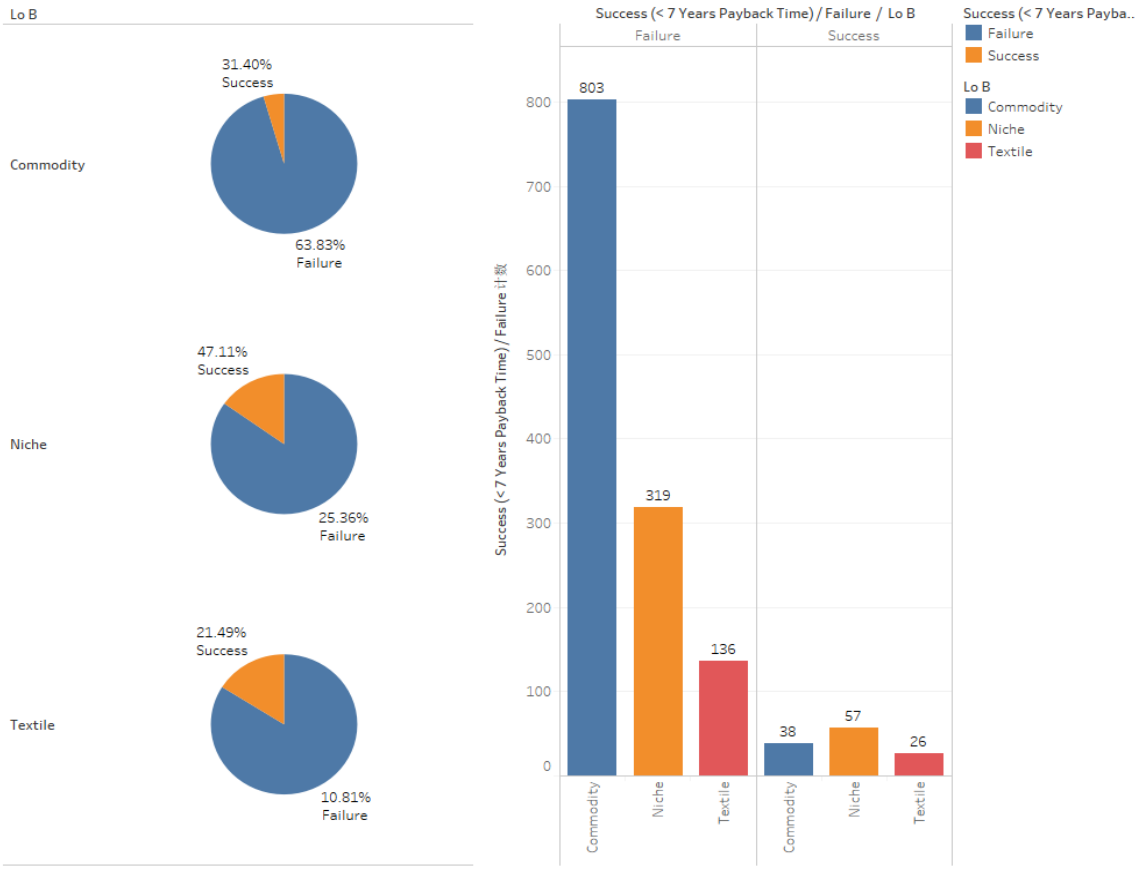
Pivot 4



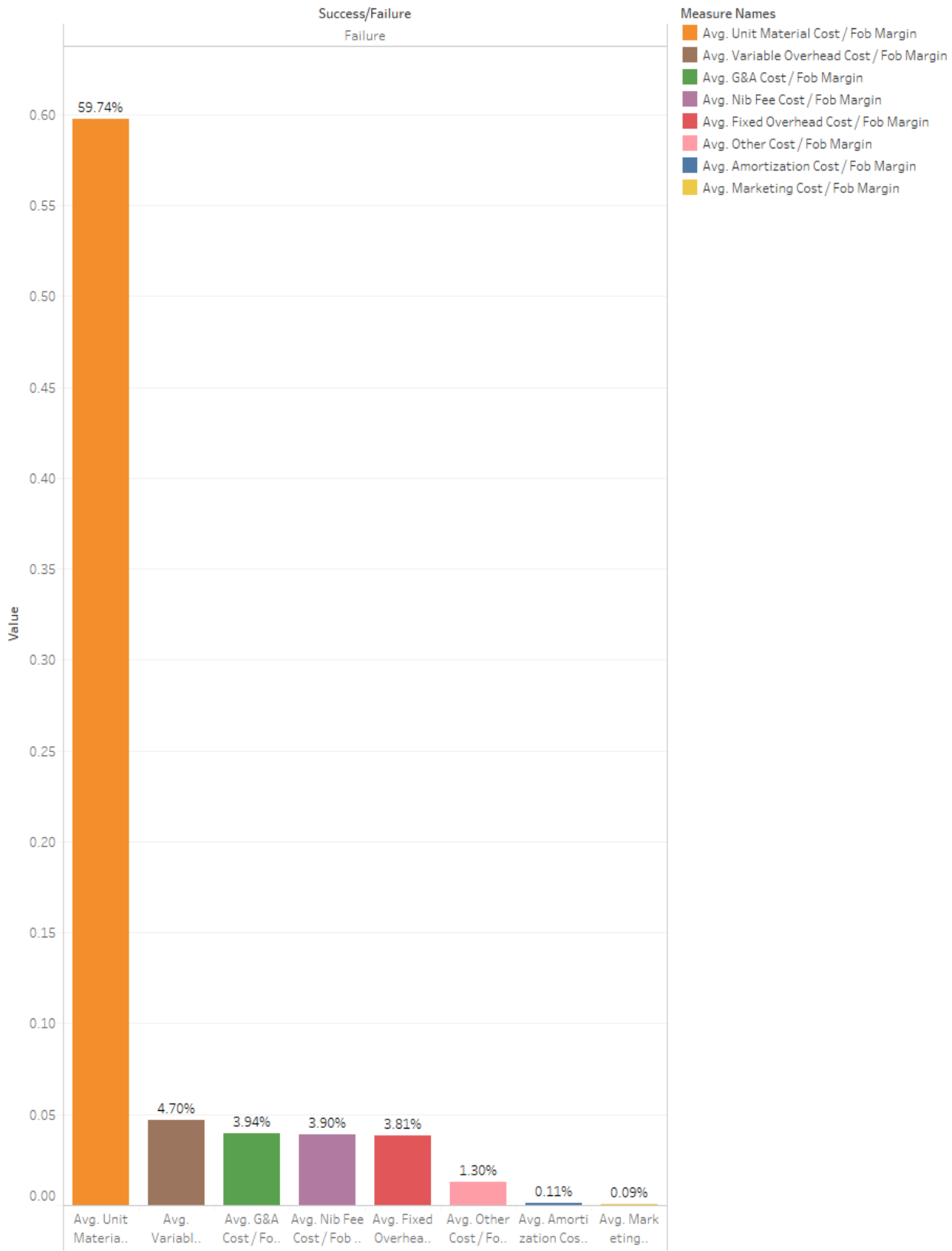
Visual 1



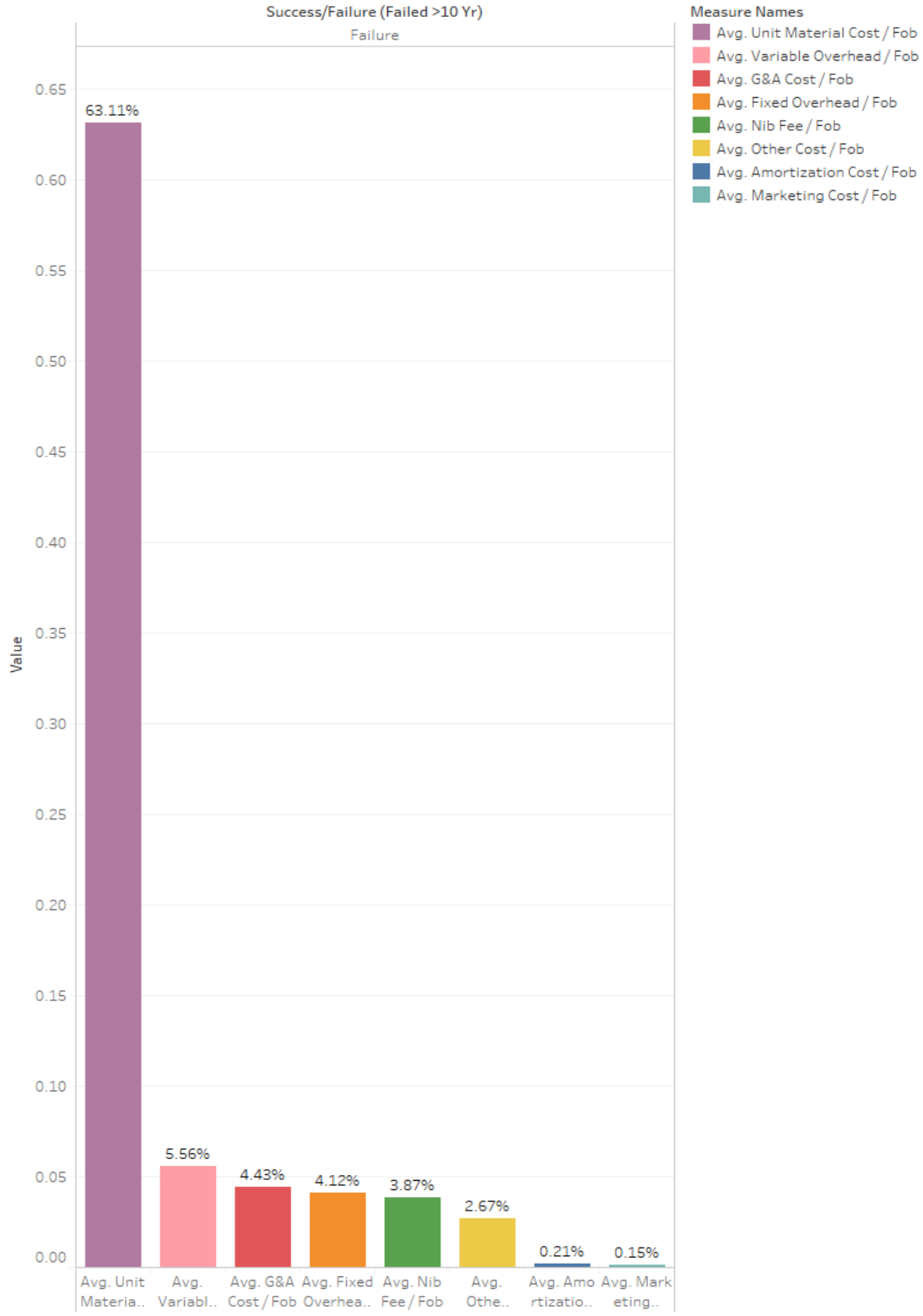
Visual 2


Lo B
Commodity
31.40%
Success
63.83%
Failure
Niche
47.11%
Success
25.36%
Failure
Textile
21.49%
Success
10.81%
Failure
Success (< 7 Years Payback Time) / Failure / Lo B
Failure
Success
Success (< 7 Years Payback Time) / Failure 计数
803
319
136
38
57
26
Commodity
Niche
Textile
Success (< 7 Years Payba..
Failure
Success
Lo B
Commodity
Niche
Textile

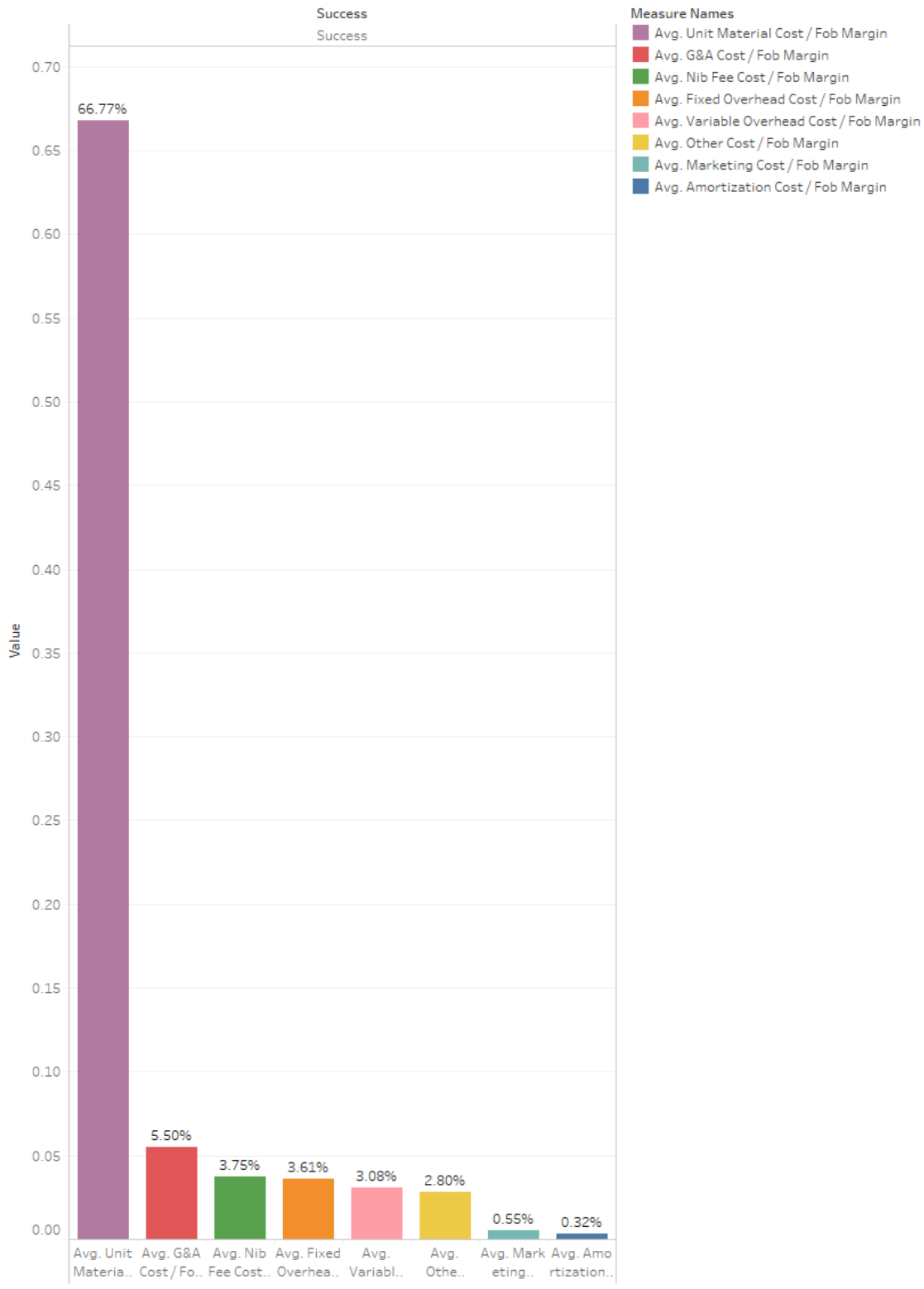
Visual 3



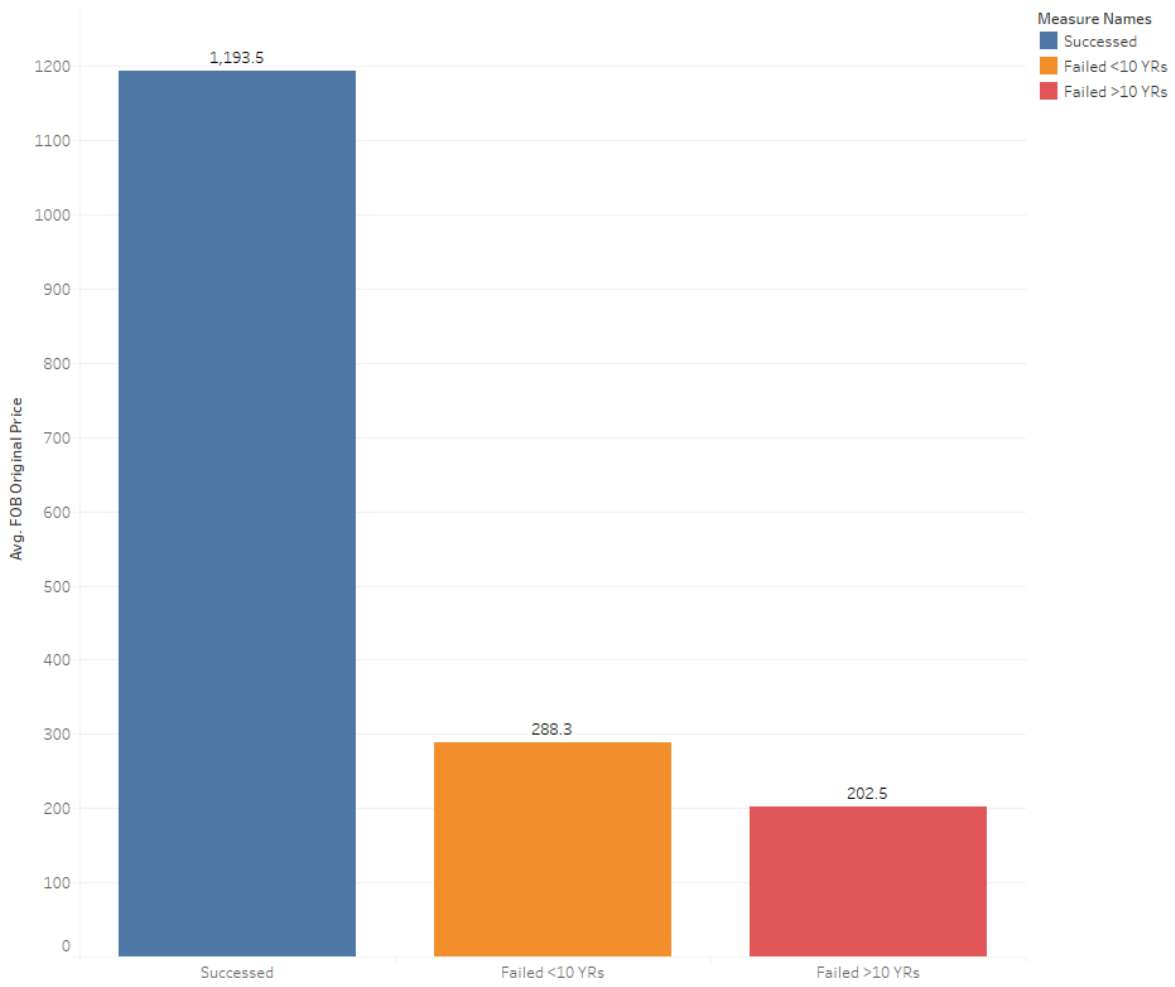
Visual 4 - Failed with payback time <10 years



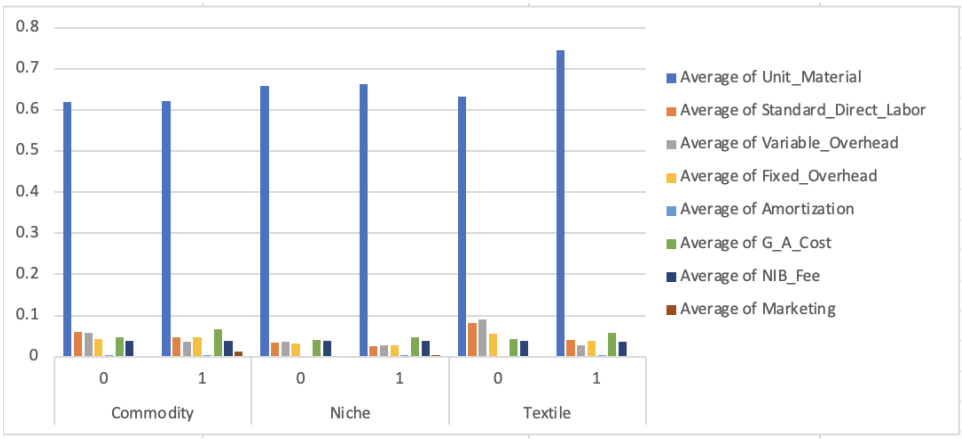
Visual 5 - Failed with payback time >10 years



Visual 6 - Successes



Visual 7



Visual 8

Results of Analysis:

Based on the visual 1 and 2, we can figure out some basic information about each category of successful projects.Visual 1 shows the average payback time of successful projects of each category. As the company's largest project category, commodities have the highest average payback time in successful projects compared with other two categories. And from the box plot, we can figure out that commodity is a very. Commodity project has very variable payback times, it has very scattered data, and the extreme range of the data is very large. Comparing different categories, textile will have the most similar payback time. From the perspective of payback time, it will be the easiest category to control.Visual 2 shows the payback time VS. the FOB origin Price of successful projects. Based on the data, the FOB Origin Prices are usually less than one thousand. But there are some extremes in the graph. And they all belong to the Niche category. We can infer that the FOB price of Niche is prone to extreme value. From the perspective of FOB price, it is also the most difficult to control among the three categories

Based on the visual analysis, we can see that the Niche type of Line of Business has the most successful project which has 57 successful projects. Among Niche types, the success rate accounts for 47.11%. That means, Niche type Line of Business may greatly influence the likelihood of a project's success. After comparing the cost ratios of the successful project, failed project with payback time less than 10 years and failed project with payback time greater than 10 years, it can be concluded with two results. First, the variable cost ratio of failed projects is high since the failed projects with payback time less than 10 years are 4.7%, failed projects with payback time greater than 10 years are 5.56% and successful projects are 3.08%. So, we can conclude that the high variable cost may cause projects to fail. Second, the G&A cost ratio and Nib cost ratio of successful projects are higher than failed projects. So, we can conclude that the G&A cost ratio and Nib cost ratio may cause projects to succeed. Moreover, after comparing the FOB original price of the successful project, failed project with payback time less than 10 years and failed project with payback time greater than 10 years, we can see that the successful projects have the highest price. That means the higher FOB original price may lead the project to be successful.

Conclusion:

Based on our results, we can see that Niche type Line of Business, G&A cost ratio and NIB cost ratio may greatly influence the likelihood of a project's success. So that the NIB can add more budget on G&A cost and NIB cost and place more products on Niche. These would increase the probability of the project's success. Since the failed projects have high variable cost, the NIB can reduce the variable cost on each project to lower the possibility of failure. The NIB can also increase the FOB original price of each product since the successful products have high FOB original price.