From Pipeline to Revenue

Predicting a company's proposal wins and losses from its sales pipeline

Andrés Forero

1. INTRODUCTION

The goal of every business is to sell products or services to potential and existing customers. Over time, a selling organization can develop a database or list of potential and existing clients to whom they want to sell their products. In a services organization, that means bidding on opportunities, contracts or work. More formally, the term "sales pipeline" has come to mean a visual representation of the stage where those potential customers are in the selling process.

My capstone project will investigate a service company's sales pipeline, and using a multi-class classification method will predict when a sales opportunity will be either **1)** a Win, **2)** a Loss, or **3)** Other.

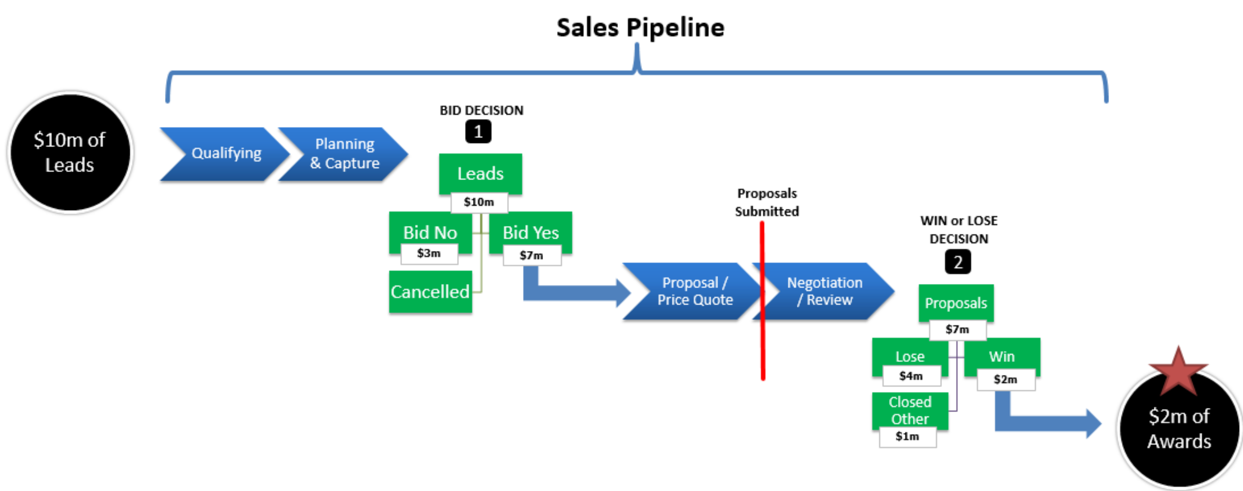
2. BACKGROUND

Let me first explain a sales pipeline through a hypothetical example. Figure 1 below is a visual representation of a company's sales pipeline. Suppose that Company B (a services company) is considering bidding on four different client's opportunities, all valued at a total of $10m:

- Client W = $3m
- Client X = $4m
- Client Y = $1m
- Client Z = $2m.

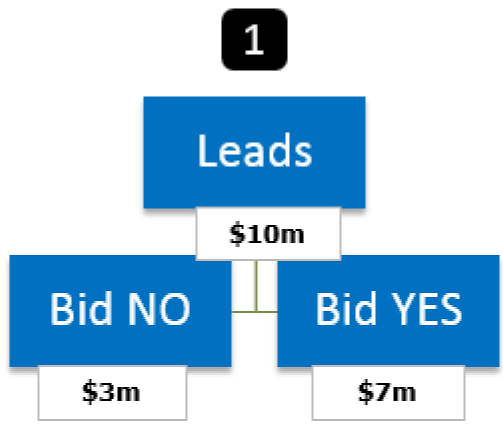
The total of $10m in leads (sales opportunities) are represented by the black circle that enters the sales pipeline.

Figure 1: Sales Pipeline Visual Representation



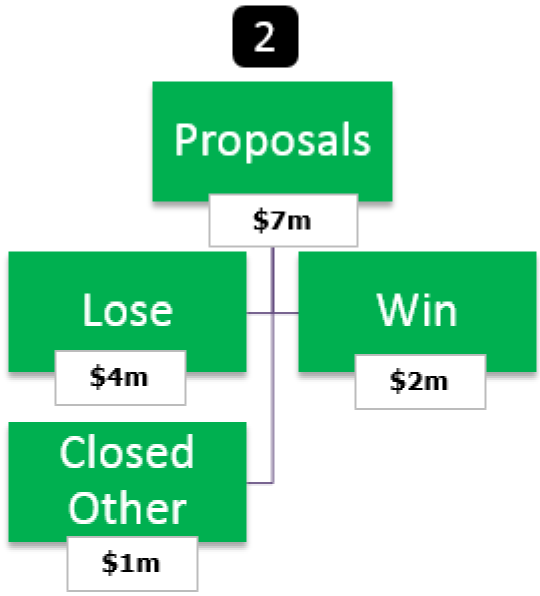
- The leads will progress through the "**Qualifying**" stage (Company B to confirm interest and fit with Clients XYZ, understand timeframe of work, develop strategic positioning)
- The leads then enter the "**Planning & Capture**" stage (begin budget discussions, analyze competition, submit for internal management approval)
- Then **Decision 1 – the Bid Decision** will be made, whereby the opportunity will either be Bid on or not. The $ amounts that are "NO" for BID will fall off the pipeline. Suppose Company B decided to not bid on Client W's $3m opportunity. Those $3m fall off the pipeline. The remaining $7m in opportunities that are "YES" for BID (Client X = $4m, Client Y = $1m, Client Z = $2m) will progress through to the next stages of the pipeline.

Figure 2: The Bid Decision



- In the "**Proposal / Price Quote**" stage, Company B iteratively develops price quote proposals for Clients X, Y and Z. Internal management presentations occur to decide on what proposal to submit.
- In the "**Review / Negotiation**" stage, Company B has submitted proposal price quotes to Clients X, Y and Z. All three clients are reviewing the proposals, and Company B is waiting to hear back. There may also be negotiation conversations occurring, should clients reply with counter offers.
- Finally, **Decision 2 – the Win or Lose Decision** will be made. Suppose that Client X's opportunity of $4m comes in as a "Loss", Client Y is in economic hardship and decides to cancel their $1m opportunity (thus categorized as "Closed Other"), and Client Z opportunity of $2m comes in as a "Win". The lost bid of $4m falls off the pipeline, the cancelled bid of $1m also falls off the pipeline, while the $2m opportunity that is won becomes an "**Award**".

Figure 3: The Win / Lose Decision for Proposals Submitted



- In this hypothetical example, we started with $10m in leads, and ended up with $2m in “Awards” or opportunities won (Client Z), while the $4m proposal (Client X) was lost, and another $4m are classified as “Bid NO” or “Closed Other” (Client W = $3m was not bid on, Client Z $1m contract cancelled).
- The $ 2m in “Awards” that were won will become revenue in future months or years, as Company B completes the services contracted with Client Z.

3. PROJECT GOAL

To summarize, the goal of my Capstone project is to classify each opportunity in the sales pipeline as either **1)** a Win, **2)** a Loss, or **3)** Bid No / Cancel / Closed Other (“BID NO” = company decided not to bid on the work; “CANCEL” = Bid cancelled; “CLOSED OTHER” = either the bidding company decided to take their proposal out of the running, OR, the client decided to cancel their contract opportunity before a win / lose response was provided to the company that submitted the proposal)

How can Company B use a model that is able to predict each of these outcomes?

There are two main ways that the model could be utilized:

1. **PREDICT WASTED BIDS**: The model can be used to predict which specific opportunities will be “Bid NO”, “Cancel or “Closed Other” before the Decision 1 (Bid vs. No Bid) point. This will allow Company B (specifically the Proposals department) to not waste time and resources (Business & Proposal Budget) submitting proposals on opportunities that will not lead to revenue. Company B can remove these opportunities from its sales pipeline (most beneficial if removed before Decision 1), and thus concentrate on either acquiring new leads, or concentrate on the opportunities currently in the pipeline that are most likely to become “Wins”.

2. **PREDICT AWARDS (PROPOSALS WON) IN $ DOLLARS**: The model can be used by the CFO of Company B to:
    a. Forecast Awards won in $ dollars for the fiscal year end.
    b. Once you know $ Awards Won, you can use that number in combination with other separate forecasts used by the CFO (i.e. the Backlog forecast and the Base Revenue Forecast) to forecast Total revenue for the current fiscal year-end, and also into future fiscal years.
    c. Understand whether the sales pipeline is large enough to sustain the CFO's future revenue and earnings growth targets.
    d. Make strategic spending decisions that will affect the future trajectory of the company. For example should the CFO commit a higher budget to Business & Proposal activities, or should he commit budget to other separate projects not related to the sales pipeline?

4. DATA SPOTLIGHT

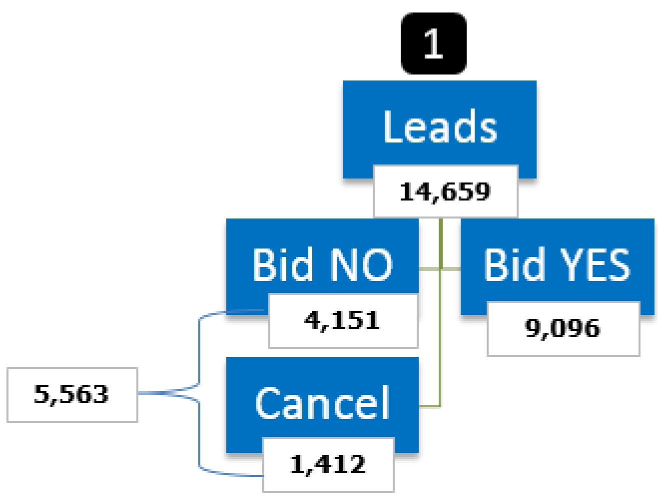
The data is historical data spanning several years for Company B, which was extracted from a Business Intelligence system as an Excel file. The source of the data is the Salesforce system. To maintain confidentiality, certain opportunity features (for example $ Amount, # of Leads / Proposals) have been manipulated (multiplied / divided) by a constant.

4.1 DECISION 1: BID vs. NO BID

As shown in Figure 4, of 14,659 Leads:

- 5,563 were either "Bid NO" (4,151) or "Cancel" (1,412)
- 9,096 were "Bid YES"

Figure 4: Number of leads / proposals available for machine learning

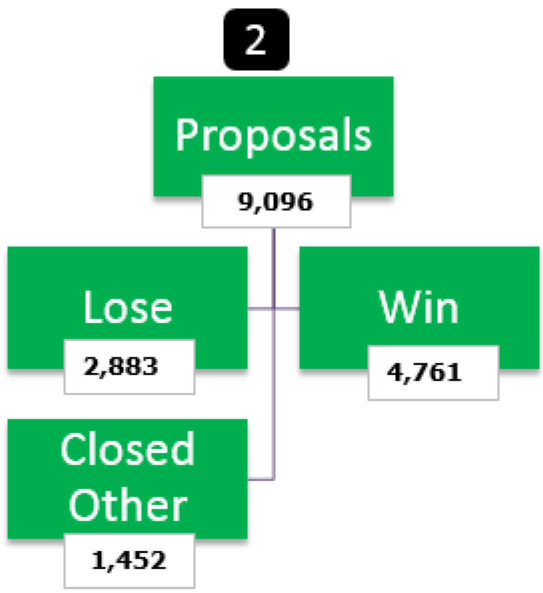


4.2 DECISION 2: WIN vs. LOSE

As shown in Figure 5, of 9,096 Total Proposals submitted:

- 2,883 were “Lose”
- 1,452 were “Closed Other”
- 4,761 were “Win”

Figure 5: Number of Proposals Submitted



4.3 SUMMARY OF TARGET CLASSES

Figure 6 shows the Target Classes that were provided to the machine learning algorithms:

Figure 6: The Win / Lose Decision for Proposals Submitted

Class Summary

Class 0 (Losses)	2,883
Class 1 (Wins)	4,761
Class 2 (Bid NO / Cancel / Closed Other)	7,015
Total	14,659

- The “Bid NO”, “Cancel” and “Closed Other” classes were combined into Class 2, because as mentioned above, one of the goals of this project is to predict the opportunities that are “Wasted Bids”, which use up valuable Company B time and resources.

4.4 DATA FEATURES

Figure 7 shows the 50 features that were utilized by the model. At this time, I decided not exclude any features, but for future analyses I can definitely only include the most essential ones.

Figure 7: Data Features

```
In [173]: X.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14659 entries, 0 to 14658
Data columns (total 50 columns):
BidWin%                              14659 non-null float64
Fiscal_Year                          14659 non-null int64
Proposal #                           14659 non-null float64
Business Unit                        14659 non-null int64
Bid_NoBid                            14659 non-null int64
BidProbability%                      14659 non-null float64
BlankBidProb                         14659 non-null int64
UndeterminedBidProb                  14659 non-null int64
LowBidProb                           14659 non-null int64
MediumBidProb                        14659 non-null int64
HighBidProb                          14659 non-null int64
AdjEstimatedBase+OptionDuration      14659 non-null int64
Amount                               14659 non-null float64
WeightedAmount                       14659 non-null float64
Zero                                 14659 non-null int64
Less1m                               14659 non-null int64
1m_10m                               14659 non-null int64
10m_30m                              14659 non-null int64
Over30m                              14659 non-null int64
EstimatedFee%                        14659 non-null float64
ProbabilityWin                       14659 non-null float64
ProbabilityWin10%                    14659 non-null int64
ProbWin15%Blank&QualStage            14659 non-null int64
ProbWin30%Blank&AllOtherStage        14659 non-null int64
ProbabilityWin30%                    14659 non-null int64
ProbabilityWin50%                    14659 non-null int64
ProbabilityWin70%                    14659 non-null int64
ProbabilityWin90%                    14659 non-null int64
ProbabilityWin100%                   14659 non-null int64
EstimatedB&PAmt                      14659 non-null float64
ActualB&PAmt                         14659 non-null float64
Commercial                           14659 non-null int64
Federal                              14659 non-null int64
Other                                14659 non-null int64
CommercialCT                         14659 non-null int64
Cost Reimbursable                    14659 non-null int64
FP                                   14659 non-null int64
Non-Fee Bearing                      14659 non-null int64
Other.1                              14659 non-null int64
T&M                                  14659 non-null int64
Days_in_Qualification                14659 non-null int64
Days_in_PlanningCapture              14659 non-null int64
Days_in_ProposalPriceQuote           14659 non-null int64
Days_in_Negotiation                  14659 non-null int64
Days_in_Review                       14659 non-null int64
AdditionalMarketsMultiAreaBlank      14659 non-null int64
EducationWorkforceDev                14659 non-null int64
EnergyEnvironment                    14659 non-null int64
FoodAgriculture                      14659 non-null int64
Health                               14659 non-null int64
dtypes: float64(9), int64(41)
```

5. DATA CLEANING & FEATURE ENGINEERING

5.1 DATA CLEANING

- I converted several categorical columns to binary (possible values are either 1 or 0). These new columns can be useful in providing insight as to whether a specific categorical feature has more or less influence on the model prediction outcomes.

- **"BidProbability %"** → several rows contain zero or "undetermined" values for this column. To clean up, any rows where the opportunity has made it through the "Bid / No Bid" Decision (in other words are in the "Proposal / Price Quote" or the "Negotiation / Review" stages), automatically get populated with 100% bid probability.

- "**AdjEstimatedBase+OptionDuration**" (*the number of months that the work on a contract that is won is estimated to last*) → several rows contain zero values for this column. To clean up, the field will be populated based on the following 4 conditions: **1)** If "Amount Range" = 1 – 1m, then 15 (months), **2)** If "Amount Range" = 1m – 10m, then 30, **3)** If "Amount Range" = 10m – 30m, then 48, **4)** If "Amount Range" = 30m >, then 60. The number of months picked for these 4 conditions is based on observation of the duration of proposals that are in the same $ dollar amount size range. Usually, the higher the $ amount of the proposal, the higher the duration.

5.2 FEATURE ENGINEERING

I created a new feature called BidWin%, which is the Bid Probability % multiplied by the Probability of Winning %. So, if the Bid Probability % is 50%, and the Probability of Winning % is 50%, then the BidWin% is 25%. With a higher BidWin%, we should expect that the the likelihood of winning the opportunity should also be higher.

- For opportunities in stages <u>after</u> Decision 1 (i.e. in the "Proposal / Price Quote", "Review / Negotiation" stages) I could have used the Probability of Winning % on its own. However, I wanted to see whether their Bid Probabilities also had any influence on their class predictions.

- On the same token, for opportunities in stages <u>before</u> Decision 1 (i.e. in the "Qualification", "Planning / Capture" stages) I could have used just their Bid Probabilities. However, I also wanted to see whether their Win Probabilities also had any influence on their class predictions. (This is possible because the Probability of Winning data is available for all opportunities, not matter what stage they are in).

6. EXPLORATORY DATA ANALYSIS

Figure 8 shows the count of proposals Won and Lost, pivoted by their $ Amount Range.

As you can see when the proposal $ amount is in the $1 – 1m range, there are more Wins than Losses. However in the other ranges there are more Losses than Wins. This implies that proposals of smaller $ amounts are easier to win than proposals of higher $ amounts because they may have less competition.

Figure 8: Count of Proposals by Amount Range

	Lose	Win
Amount Range		
1 - 1m	1,753	3,956
10m - 30m	100	60
1m - 10m	629	559
30m >	46	30
Zero	355	156
All	2,883	4,761

Figure 9 shows the count of number of days that an opportunity spends in each pipeline stage pivoted by class. I won't comment on all of them, but here are a couple of interesting points.

- For "Days in Negotiation", it makes sense that Wins have a higher count of days than the other classes because proposals that end up as Losses may not require as much negotiation time.
- Remember that the "Qualification" stage is where Company B is analyzing whether the opportunity is a good fit for them to bid on. For "Days in Qualification", Wins have a higher count of days. This implies that opportunities that spend more time in "Qualification" have a greater chance of being a good fit for Company B, and thus are more likely to be won.
- The "Review" stage is where the client is analyzing whether to award the proposal contract to Company B or the other competitors. For "Days in Review", Losses have a higher count of days than Wins. This implies that the higher the number of days spent in Review, the more likely you are to lose the proposal. This could be because there are more competitors and so the client needs more days to decide who will be the winner, or simply because Company B was deemed a to have a bad proposal (or not as good) by the client.

Figure 9: Count of number of days in Pipeline Stage

	Days in Negotiation	Days in Planning & Capture	Days in Proposal / Price Quote	Days in Qualification	Days in Review
Closed Other	947	49,517	106,401	383,573	59,546
Lose	2,908	7,068	59,539	19,780	242,232
Win	9,467	12,850	84,554	32,060	172,784

Figure 10 shows the "Actual Bid & Proposal (B&P) Amount" (*the actual $ amount that the Proposal department spends on Bid & Proposal activities)* pivoted by class.

- It is known that the lower the proposal $ Amount, the lower the Actual B&P will be. This rule does not follow a linear relationship, but you can be sure that a $500k proposal should have lower B&P expenses than a $10 million proposal. In Figure 10 it is interesting that the **4,761** proposals that were Won had B&P costs of $56 million, while the **2,883** proposals that were Lost had B&P costs of $63 million. In other words, a lower number of proposals (the Losses) had higher B&P expenses. It is possible that this could be related to the $ Amount of the proposals. The Wins had an average $ amount of $1.1 million, while the Losses had an average $ amount of $2.5 million. In the next section below (Section 7 Inferential Statistics) the hypothesis test I performed showed statistical significance in that Amounts less than $1m had a higher the Probability of Winning than Amounts greater than $1m. It then follows that since B&P expenses are less for smaller $ amount proposals, then that is why B&P expenses were also lower for Wins (average of $1.1m) as opposed to Losses (average of $2.5m).

Figure 10: Actual Business & Proposal Amount ($)

	Actual B&P Amt
Closed Other	$13,519,360
Lose	$63,639,010
Win	$56,495,350

Figure 11 shows "Adjusted Estimated Base + Option Duration" (*the number of months that the work on a contract that was won is estimated to last*) pivoted by class.

- As you can see the number of months for Wins (4,761 proposals) is lower than for Losses (2,883 proposals). This implies that the proposals we are winning are of shorter duration. Proposals of shorter duration are also usually proposals of lower $ amounts. Therefore, this coincides with the point made above that we win proposals of lower $ amounts.

Figure 11: Adjusted Estimated Base + Option Duration

	Adj Estimated Base + Option Duration
Closed Other	224,690
Lose	93,086
Win	84,936

Figure 12 shows the number of proposals by class, pivoted by Probability of Winning.

- For "Wins", as expected, the higher the Probability of Winning, the higher the number of proposals that were won.

- For “Losses” you can see that the highest concentration of number of proposals is in the 10% to 70% Probability of Winning range. In other words, the lower the Probability of Winning, the higher the number of “Losing” proposals.
- Finally, for “Closed Other” (including “BID NOs” “Cancels” and “Closed Other”), the highest concentration of number of proposals is when the Probability of Winning is 0%. That makes sense because Company B would not Bid or they would cancel proposals with a 0% Probability of winning.

Figure 12: Number of Proposals and their Probability of Winning %

Probability of Winning %:	0%	10%	30%	50%	70%	90%	100%	All
Closed Other	4,018	664	814	647	341	226	305	1,452
Lose	12	711	1,003	718	300	87	52	2,883
Win	119	169	406	583	660	864	1,960	4,761

7. INFERENTIAL STATISTICS & HYPOTHESIS TESTS

I was most interested in performing a hypothesis test that could tell me whether there is a difference in the mean “Probability of Winning” for proposals $ amounts that are less than $1 million (79%), versus $ amounts that are higher than $1 million (66%).

Figure 13, shows that I was able to reject the null hypothesis with a 1% confidence level. In other words, there is evidence that Company B is more likely to win proposals that have a $ dollar amount less than $1 million.

Figure 13: Hypothesis Testing

Hypothesis Test 1:

Ho: There is no difference in the means of Probability of Winning of the following two groups whose WinLose Category is "WIN":

- Amount <= 1m (which have mean Probability of Winning of 78.9%), versus
- Amount > 1m (which have mean Probability of Winning of 66.4%)

Ha: There is a difference in the means

Reject Ho if $p < 0.01$

```
In [8]: ProbLess1m = AmtLess1m['ProbabilityofWinning']

In [9]: ProbLess1m.mean()
Out[9]: 0.7891134124779047

In [10]: ProbOver1m = AmtOver1m['ProbabilityofWinning']

In [11]: ProbOver1m.mean()
Out[11]: 0.6643962848297207
```

Hypothesis 1: Reject Ho, given p-value is 5.792e-23 as shown below:

```
In [12]: # Perform t-test on the means of the the two series:

         stats.ttest_ind(a= ProbLess1m, b= ProbOver1m, equal_var=False) # Assume samples have equal variance input as false

Out[12]: Ttest_indResult(statistic=10.166514179400044, pvalue=5.7927076036784238e-23)
```

Even though this test showed statistical significance, does it still hold up to a test of practical significance? Practical significance asks whether the differences between the samples are big enough to have real meaning. What can happen sometimes is that even if a test shows statistical significance, effect size may show that the difference between the two groups is trivial. To test for practical significance I calculated the Effect Size using Cohen's D.

As you can see in Figure 14, Cohen's D was 46%, which according to Cohen is close to a medium effect size. Therefore in terms of Practical Significance, this medium effect size is consistent with the presence of additional competition on deals larger than $1M.

Finally, I would also clarify that this test does not imply that Company B does not win proposals of amounts over $1m. Company B is doing quite well winning the majority of proposals they bid on, having a healthy probability of winning of 66%.

Figure 14: Cohen's D and Practical Significance

```
In [13]: def CohenEffectSize(group1, group2):
             """Compute Cohen's d.

             group1: Series or NumPy array
             group2: Series or NumPy array

             returns: float
             """
             diff = group1.mean() - group2.mean()

             n1, n2 = len(group1), len(group2)
             var1 = group1.var()
             var2 = group2.var()

             pooled_var = (n1 * var1 + n2 * var2) / (n1 + n2)
             d = diff / numpy.sqrt(pooled_var)
             return d
```

Hypothesis 1: Calculate Effect Size

- Effect size is 0.46, which is close to a medium effect size, therefore, even though there was enough statistical significance to reject the Ho, in terms of Practical Significance, there is a medium effect size, which provides further evidence that there is a difference between the means of the two groups.

```
In [14]: CohenEffectSize(ProbLess1m, ProbOver1m)

Out[14]: 0.46015424887665285
```

8. ALGORITHMS TESTED

I used the algorithms listed below to perform the model fitting.

Please refer to the Jupyter Notebook, which shows the code and results for each Model.

1. **Logistic Regression**: Even though my project is a multi-class categorization project, I performed a one-versus-all logistic regression with Wins as the first class, and Losses & Other combined as the second class. Seeing the coefficients for each feature could be informative. Since the coefficients will be either positive or negative, they can give an indication of the magnitude of how a particular feature is strengthening or weakening prediction outcomes.

2. **Decision Trees:** are a flow-chart-like (or tree-like) structure, where each internal node denotes a test on a feature, each branch represents the outcome of a test, and each leaf (or terminal) node predicts a class label.1[1] They are popular because they provide fast training and prediction, and are easy to understand and interpret.

3. **Random Forests**: are an ensemble learning method (based on Decision Trees) that operates by constructing a multitude of random decision trees at training time and then outputting the class that is the most common of the classes.

4. **Support Vector Machines**: categories are divided by a clear gap (or separating hyperplane) that is made as wide as possible. The vector points that the gap lines touch are known as Support Vectors.

5. **K Nearest Neighbors**: is known as a lazy algorithm and is a non-parametric method (i.e. it doesn't have to assume a linear relationship). The algorithm calculates the distance from X (the new point to be predicted) to all points in the data. The points are sorted by increasing distance from X. Then it predicts the majority label of the "k" closes points.

6. **XGBoost** (extreme gradient boosting): is an ensemble method like Bagging (Bootstrap aggregating), but weak learners evolve over time. It uses regression trees, so the variance is reduced in each bin. Each tree learns the current error function (i.e. the gradient of the loss output at each point), so when it is added to the ensemble the overall error is hopefully reduced.

9. MODEL EVALUATION

I used scikit-learn's Classification Reports and Confusion Matrices to analyze the average **Precision** and **Recall** for each model.

As you recall from Figure 1 above, in the sales pipeline, **Decision 1** is the Bid vs. No Bid decision, while **Decision 2** is the Win vs. Lose decision.

The interesting thing about this project is that Recall is more important for Decision 1, while Precision is more important for Decision 2. I explain this in the next sections.

Therefore, the model with both the highest Precision and Recall scores for the appropriate Decision point was chosen the winner.

[1] https://en.wikipedia.org/wiki/Decision_tree

9.1 DECISION 1: BID vs. NO BID

- **False Negatives More Costly →** over-predicting “BID NO” (under-predicting “BID YES”):

Any true “”BID YES” that are misclassified as “Bid NO” (Type 2 error – False Negatives) will cause the company to not submit proposals for opportunities that in actuality turned out to be “Wins”. Therefore, Awards & Revenue would be lower than they should have been.

- **False Positives Less Costly →** Over-predicting “BID YES” (under-predicting “BID NO”):

Any true “BID NOs” that are erroneously classified as “BID YES” (Type 1 error – False Positives) will cause the company to submit proposals on too many opportunities. It is a waste of time and resources (miss-utilize B&P budget), to go after proposals that may not be a good fit with Company B, and have a greater chance of being lost. However, the higher B&P costs of over-submitting proposals are not as costly as not submitting proposals for opportunities that in actuality turned out to be “Wins”.

<u>Recall (reduce False Negatives) is More Important</u>:

To summarize the two bullets above in a different way, for Decision 1 **Recall** is more important than **Precision** because we would like to have less False Negatives in trade off to having more False Positives. In other words, getting a False Negative is more costly, while getting a False Positive is not as costly.

9.2 DECISION 2: WIN vs. LOSE

- **False Negatives Less Costly →** over-predicting Losses / Closed Other (under-predicting Wins):

Any true “”Wins” that are misclassified as “Losses” or “Closed Other” (Type 2 error – **False Negatives**) will cause Award projections to be lower than actual results (or Targets). If the CFO has aggressive revenue and earnings growth targets, and he thinks that the Awards (and also Revenue) prediction is lower than his growth targets (even though they are not; i.e. in the future actual Wins will be higher than the prediction), he will push to reverse that. For example, he may ask the Proposals department to expand Business & Proposal (B&P) efforts, so that the Pipeline size increases, and consequently Company B has a greater chance of winning more Awards. This is actually not as costly, because an organization that increases its pipeline will also grow its Awards and Revenue, and that is not a bad thing. One negative is that maybe B&P costs will go up unnecessarily during the time period, and so resources that would have gone somewhere else will now be spent on B&P activities. However, given that growing revenue and earnings is important to any business, this temporary shift in resources may not be all that bad if the CFO can shift resources around without impacting any specific area negatively (but maybe easier said than done).

- **False Positives More Costly →** over-predicting Wins (under-predicting Losses / Closed Other):

On the other hand, any true “Losses” or “Closed Other” that are erroneously classified as “Wins” (Type 1 error – **False Positives**) may cause the Awards projection to appear closer to or higher than the CFO’s present growth targets . This can be very costly for the opposite points mentioned above. If the CFO has aggressive revenue and earnings growth targets, and he thinks that Awards (and consequently also Revenue) are on track or better than target expectations, the organization may relax and not push B&P activity as hard as it should. This is very costly because in the future once Actual results come in (i.e. the number of wins achieved), and they are lower than the original prediction, the organization actually does have lower Awards and Revenue numbers. Having lower revenue is not good for any organization.

Precision (reduce False Positives) is More Important:

To summarize the two bullets above, **Precision** is more important than **Recall** when we would like to have less False Positives in trade off to having more False Negatives. In other words, getting a False Positive is more costly, while getting a False Negative is not as costly.

10. MODEL COMPARISONS

10.1 DECISION 1: BID vs. NO BID

In the model, "Bid NO", "Cancels" or "Closed Other" are represented by Class 2.

If you remember from Section 9.1 above, for Decision 1 it is more important for the model to have good **Recall** because we want to have less False Negatives, otherwise the company would not submit proposals for all actual "Wins".

As shown below in Figure 15, in terms of Precision for Class 2, Random Forest had a score of **0.97** and XGBoost had a score of **0.96**.

Figure 15: Decision 1 - Precision and Class 2

	Precision					
	Logistic	**Trees**	**Rforest**	**SVM**	**KNN**	**XGBoost**
0	0.89	0.63	0.67	0.63	0.56	0.70
1	0.80	0.75	0.81	0.80	0.73	0.81
2		0.89	**0.97**	0.97	0.87	**0.96**
avg.		0.79	0.86	0.85	0.76	0.86

As shown below in Figure 16, in terms of Recall, for Class 2 Trees had a score of **0.90** and XGBoost also had a score of **0.90**.

Figure 16: Decision 1 - Recall and Class 2

	Recall					
	Logistic	**Trees**	**Rforest**	**SVM**	**KNN**	**XGBoost**
0	0.91	0.63	0.79	0.82	0.58	0.78
1	0.77	0.75	0.82	0.80	0.72	0.83
2		**0.90**	0.89	0.85	0.86	**0.90**
avg.		0.79	0.85	0.83	0.76	0.85

Given that Recall is more important for Decision 1, I decided that XGBoost is the best choice (even though its Precision score was second best). In addition, XGBoost performed consistently well on both Precision and Recall (while Random Forest was only consistent on Precision).

10.2 DECISION 2: WIN vs. LOSE

Remember from section 7.2 above that for Decision 2, it is more important for the model to have good Precision.

As shown below in Figure 17, in terms of Precision, for "Wins" (Class 1) Random Forest and XGBoost had scores of **0.81**.

Figure 17: Decision 2 - Precision and Class 2

	Precision					
	Logistic	**Trees**	**Rforest**	**SVM**	**KNN**	**XGBoost**
0	0.89	0.63	0.67	0.63	0.56	0.70
1	0.80	0.75	**0.81**	0.80	0.73	**0.81**
2		0.89	0.97	0.97	0.87	0.96
avg.		0.79	0.86	0.85	0.76	0.86

As shown below in Figure 18, in terms of Recall, for "Wins" (Class 1) Random Forest had a score of 0.82 and XGBoost had a score of **0.83**.

Figure 18: Decision 2 - Recall and Class 1

	Recall					
	Logistic	**Trees**	**Rforest**	**SVM**	**KNN**	**XGBoost**
0	0.91	0.63	0.79	0.82	0.58	0.78
1	0.77	0.75	**0.82**	0.80	0.72	**0.83**
2		0.90	0.89	0.85	0.86	0.90
avg.		0.79	0.85	0.83	0.76	0.85

In summary, XGBoost tied for highest score for Precision, and had the best Recall score. Given that Precision is more important for Decision 2, I decided that XGBoost is the best choice (even though Random Forest performed equally well on Precision).

10.3 THE WINNER IS XGBOOST

- For Decision 1, Recall is more important, and XGBoost performed the best
- For Decision 2, Precision is more important, and XGBoost performed the best.
- Therefore, overall XGBoost is the winner.

11. XGBOOST MODEL – FEATURES IMPORTANCES

Figure 19 below shows the 11 most important features with XGBoost. Let me comment on some of the more interesting ones.

Figure 19: XGBoost Feature Importance

	Features
1	F30 = ActualB&PAmt
2	F2 = Proposal #
3	F12 = Amount
4	F29 = EstimatedB&PAmt
5	F13 = WeightedAmount
6	F44 = Days_in_Review
7	F42 = Days_in_ProposalPriceQuote
8	F19 = EstimatedFee%
9	F11 = AdjEstimatedBase+OptionDuration
10	F40 = Days_in_Qualification
11	F0 = BidWin%

1. **ActualB&PAmt** (*the actual $ amount that the Proposal department spends on Bid & Proposal activities)* was the most important feature. It makes sense that the algorithm would be able to split effectively on this feature. Opportunities that were "BID NO", "CANCELS" and some "CLOSED OTHER" had low or no B&P expenses. While for opportunities that were "BID YES" ("Wins", "Losses" and some "Closed Other"), definitely had higher B&P expenses because we bid on them.

2. **Proposal #** (*just a Company B internal numeric identifier for the proposal*) was the second most important feature. I was initially very surprised that this feature was so important, until I realized that proposals that we do not Bid on (3,975 rows), and some that are "Cancelled" (1,412 rows) do not get a Proposal # ID. This certainly makes it easy for the algorithm to split on that feature. Please note that I also tested all algorithms without the Proposal # as a feature. The decrease in Recall and Precision scores was minimal (1 to 4% depending on the algorithm), and so for this report I have not removed it because I wanted to document its effect on the results. However, if I was to deploy an XGBoost model into production, I would not include the Proposal # as a feature.

3. **Amount** (*the $ amount of the opportunity. Amount may update as opportunity moves through pipeline stages*) was the third most important feature. As you recall, in the inferential statistics section I mentioned that opportunities with amounts of less than $1 million were more likely to be won. In other words, the higher the amount, the less likely the opportunity will be won. For this reason, it makes sense that the algorithm found it easier to split on this feature.

4. **Days_in_Review** (*the number of days that a proposal that was submitted to a client has been reviewed by the client before providing a win or lose decision*) is the sixth most important feature. It makes

sense that the algorithm was able to split relatively easy on this feature. It is known that the longer a client takes in providing an answer to Company B about their proposal, the less likely they are to win the proposal.

5. **BidWin%** (*recall the section 5.2 example that if the Bid Probability % is 50%, and the Probability of Winning % is 50%, then the BidWin% is 25%)* was the eleventh most important feature. Initially I was expecting this feature to be more important, but I now I understand why it did not do as well. At inspecting the data I noticed that opportunities labelled "Wins" or "Losses" can have quite varying Bid Probabilities and Win Probabilities. For example, there are opportunities that are labelled as "Wins" that had very high or very low Bid Probabilities, as well as very high or very low Win Probabilities. The same goes for "Losses", which had very high (or low) Bid or Win Probabilities, but they still ended up labelled as "Losses". Therefore, the high variance in the probabilities makes it difficult for the algorithm to split on this feature consistently.

12. XGBOOST MODEL PERFORMANCE IN $ DOLLARS

- For Decision 1 – Bid vs. No Bid, XGBoost had Precision of 96% and Recall of 90%.
- For Decision 2 – Win vs. Lose, XGBoost had Precision of 81% and Recall of 83%.

These are great results if the goal of the model is simply trying to predict the number of outcomes for opportunities at Decision points 1 and 2. However, the metric changes if we want to predict outcomes in $ dollar amounts, which is the way the CFO of Company B would want to receive the information.

12.1 DECISION 1: BID vs. NO BID

Figure 20 below shows the confusion matrix that XGBoost produced. Please just focus on Class 2 ("Bid NO", "Cancel" or "Closed Other").

- The Total Actual "No Bids / Cancels" were 2,291 opportunities (red star).
- The Total Predicted "No Bids / Cancels" were 2,146 (blue star), of which 2,063 (black star) were predicted correctly.
- Precision was **96%** (2,063 / 2,146), while Recall was **90%** (2,063 / 2,291).

Figure 20: Decision 1 – Class 2 Confusion Matrix

	Predicted: 0 - Losses	Predicted: 1 - Wins	Predicted: 2 - No Bid / Cancel	Total Actuals
Actual: 0 - Losses	753	169	42	964
Actual: 1 - Wins	234	1,308	41	1,583
Actual: 2 - No Bid / Cancel	95	133	**2,063**	**2,291**
Total Predictions	1,082	1,610	**2,146**	4,838

Class 2 Precision:	96%
Class 2 Recall:	90%

In Figure 21 below I converted the Confusion Matrix to $ dollars.

- The Total Actual "No Bids / Cancels" were $11.5 billion (red star).
- The Total Predicted "No Bids / Cancels" were $11.1 billion (blue star), of which $10.9 billion (black star) were predicted correctly.
- Precision was **98%** ($10.9 b / $11.1 b), while Recall was **95%** ($10.9 b / $11.5 b).

Figure 21: Decision 1 – Class 2 Confusion Matrix in $ dollars

	Predicted: 0 - Losses	Predicted: 1 - Wins	Predicted: 2 - No Bid / Cancel	Total Actuals
Actual: 0 - Losses	$2,118,457,383	$283,668,187	$136,015,588	$2,538,141,158
Actual: 1 - Wins	$478,730,957	$1,362,309,483	$46,405,063	$1,887,445,504
Actual: 2 - No Bid / Cancel	$474,813,260	$121,274,875	**$10,920,912,724**	**$11,517,000,859**
Total Predictions	$3,072,001,600	$1,767,252,545	**$11,103,333,375**	$15,942,587,521

Class 2 Precision:	98%
Class 2 Recall:	95%

12.2 DECISION 1: SUMMARY OF CONFUSION MATRICES

- To summarize, Precision was **96%** by counts, while in $ dollar terms it increased by 2% to **98%**.
- On the other hand, Recall was **90%** by counts, while in $ dollar terms it decreased by 5% to **95%**.

- The lesson here is that just because Precision and Recall are certain percentages when computed based on count of proposals won, that does not mean that in $ dollar terms, the percentages will be the same, or will increase or decrease by the same magnitude.
- The reason this happens is because the opportunities vary greatly in their $ dollar amounts.
- In this case since Recall is more important for Decision 1, in $ dollar terms the Recall of 95% is still very good.

12.3 DECISION 2: WIN vs. LOSE

Figure 22 below shows the confusion matrix that XGBoost produced. Please just focus on Class 1 (the number of proposals Won).

- The Total Actual Wins were 1,583 proposals (red star).
- The Total Wins were predicted as 1,610 (blue star), of which 1,308 (black star) were predicted correctly.
- Precision was **81%** (1,308 / 1,610), while Recall was **83%** (1,308 / 1,583).

Figure 22: Decision 2 – Class 1 Confusion Matrix

	Predicted: 0 - Losses	Predicted: 1 - Wins	Predicted: 2 - No Bid / Cancel	Total Actuals
Actual: 0 - Losses	753	169	42	964
Actual: 1 - Wins	234	**1,308**	41	**1,583**
Actual: 2 - No Bid / Cancel	95	133	2,063	2,291
Total Predictions	1,082	**1,610**	2,146	4,838

Class 1 Precision:	81%
Class 1 Recall:	83%

In Figure 23 below I converted the Confusion Matrix to $ dollars.

- The Total Actual Wins were $1.89 billion (red star).
- The Total Wins were predicted as $1.76 billion (blue star), of which $1.36 billion was predicted correctly (black star).
- In $ dollar terms, Precision was **77%** ($1.36 b / $1.77 b), while Recall was **72%** ($1.36 b / $1.89 b).

Figure 23: Decision 2 – Class 1 Confusion Matrix in $ dollars

	Predicted: 0 - Losses	Predicted: 1 - Wins	Predicted: 2 - No Bid / Cancel	Total Actuals
Actual: 0 - Losses	$2,118,457,383	$283,668,187	$136,015,588	$2,538,141,158
Actual: 1 - Wins	$478,730,957	**$1,362,309,483**	$46,405,063	**$1,887,445,504**
Actual: 2 - No Bid / Cancel	$474,813,260	$121,274,875	$10,920,912,724	$11,517,000,859
Total Predictions	$3,072,001,600	**$1,767,252,545**	$11,103,333,375	$15,942,587,521

Class 1 Precision:	77%

Class 1 Recall:	72%

12.4 DECISION 2: SUMMARY OF CONFUSION MATRICES

- To summarize, Precision was **81%** by counts, while in $ dollar terms it decreased by 4% to **77%**.
- On the other hand, Recall was **83%** by counts, while in $ dollar terms it decreased by 11% to **72%**.
- The lesson here is that just because Precision and Recall are certain percentages when computed based on count of proposals won, that does not mean that in $ dollar terms, the percentages will be the same, or will increase or decrease by the same magnitude.
- The reason this happens is because the opportunities vary greatly in their $ dollar amounts.
- The good news in this case is that since Precision is more important for Decision 2, in $ dollar terms the Precision is 77%, which is still pretty good.
- However, the warning is that in $ dollar terms, Precision or Recall can decrease or increase, making it unpredictable (or risky) in production.

12.5 XGBOOST MODEL IN PRODUCTION:

Let us assume that Company B used XGBoost in production, and so the predicted number that would be presented to the CFO would be $1.76 billion (the total predictions amount). Company B would not know until much later in the future that only $1.36 billion of that prediction was correct (77% Precision).

While Precision is an important metric, to be honest, the CFO would really only care about how close the total prediction of $1.77 billion was to the total Actual value of $1.89 billion. In this case, the variance was $(120) million or 4% off.

Figure 24: XGBoost Prediction vs. Actual Wins

XG Boost Predicted Wins	$1,767,252,545
Actual Wins	$1,887,445,504
Variance	($120,192,958)

Therefore, the question becomes, is using XGBoost any better than the current method that Company B uses? I will answer that question in the next section.

12.6 PREDICTION METHOD CURRENTLY USED AT COMPANY B

Company B currently uses a "Weighted Pipeline Amount" method. In other words, suppose there is a Lead of $100,000 that is in the "Qualification" stage and has a "Bid Probability %" of 50%, and a "Win Probability %" of 75%. That means that the "Weighted Pipeline Amount" is $40,000 as shown in Figure 25.

Figure 25: Example "Weighted Pipeline Amount"

Lead Amount	$100,000
x Bid Probability	50%
x Win Probability	80%
Weighted Amount	**$40,000**

The Total Weighted Pipeline Amount for Company B for the time period analyzed is $3.08 billion. This is the number that the CFO would currently use for Award Projections, Revenue Projections and other analyses.

If you compare that to the Actual Wins of $1.88 billion, you can see that the variance is quite large at $1.2 billion.

Figure 26: Weighted Amount Prediction vs. Actual Wins

Weighted Amount	$3,088,275,476
Actual Wins	$1,887,445,504
Variance	$1,200,829,972

On the other hand, the XGBoost variance was only $(120) million.

Figure 27: XGBoost Prediction vs. Actual Wins

XG Boost Predicted Wins	$1,767,252,545
Actual Wins	$1,887,445,504
Variance	($120,192,958)

In addition, the Weighted Amount is higher than the Actual Wins. In other words, the prediction is predicting Awards to be higher, than they actually were. As mentioned above in Section 9.2, this can be costly because if the CFO has aggressive revenue and earnings growth targets, and he thinks that Awards (and also Revenue) are better or on track with target expectations, the organization may relax and not push B&P activity as hard as it should.

On the other hand, the XGBoost total prediction ($1.77 billion) is lower than Actual Wins ($1.89 billion). It is less risky that the prediction is lower than Actuals, because if the CFO has aggressive revenue and earnings growth targets, and he thinks that the Awards (and also Revenue) prediction is lower than his growth targets (even though they are not; in the future Actual Wins will come in higher than predicted), he can push the organization to reverse that in the present. View explanation in section 9.2.

12.7 XGBOOST MODEL RISKS (OR ANY OTHER MODEL RISKS)

As with any prediction method, there are inherent risks. Let me use an extreme example. Figure 28 below is the same Confusion Matrix from Figure 23, but I have made a manual adjustment to it whereby Total Predicted Wins were increased by $ 2 billion to be $3.77 billion (blue arrow), and Total Predicted "No Bid / Cancel" were reduced by $2 billion to $9.1 billion (red arrow). In other words, suppose that a $ 2 billion proposal that was "Cancelled" was misclassified as a "Win".

Figure 28: Extreme Example Manual Adjustment to Confusion Matrix

	Predicted: 0 - Losses	Predicted: 1 - Wins	Predicted: 2 - No Bid / Cancel	Total Actuals
Actual: 0 - Losses	$2,118,457,383	$283,668,187	$136,015,588	$2,538,141,158
Actual: 1 - Wins	$478,730,957	$1,362,309,483	$46,405,063	**$1,887,445,504**
Actual: 2 - No Bid / Cancel	$474,813,260	$2,121,274,875	$8,920,912,724	$11,517,000,859
Total Predictions	$3,072,001,600	**$3,767,252,545**	$9,103,333,375	$15,942,587,521

The prediction that would be provided to the CFO would be $3.77 billion, which means the variance would be $1.879 billion. This is significantly worse than using the "Weighted Pipeline Amount" method (which was off by $1.2 billion)

Figure 29: Extreme Example XGBoost Prediction vs. Actual Wins

XG Boost Predicted Wins	$3,767,252,545
Actual Wins	$1,887,445,504
Variance	$1,879,807,042

While, unlikely that such an extreme case (caused by this very large $2 billion opportunity) would be allowed without verification to be submitted as a final prediction to the CFO, this example does demonstrate a risk of forecasting in $ dollar terms with XGBoost, or any other method for that matter.

13. RECOMMENDATIONS

To summarize, the goal of my Capstone project was to classify each opportunity in the sales pipeline as either **1)** a Win, **2)** a Loss, or **3)** Bid No / Cancel / Closed Other

Using those predictions, I recommended that Company B could utilize the model in two distinct ways:

- **Predict Wasted Bids**: predict which specific opportunities will be "Bid NO", "Cancel or "Closed Other" before the Decision 1 (Bid vs. No Bid) point. This allows Company B to not waste time and resources submitting proposals on opportunities that will not lead to revenue. The Proposal department would probably get the most use from this functionality.
- **Predict Awards (Proposals Won) in $ dollars:** The CFO could use the model to forecast Awards, Revenue and make strategic spending decisions.

Given the positive results that XGBoost provided, Company B could put the model into production to provide better analytics. Here are my recommendations of how to do that:

1. **Temporary Testing Phase:**

   The Proposal department can use the model to predict what opportunities not to bid on (Wasted Bids). The Proposal department already has a good process of "Qualifying" opportunities for bids, and so the model can be a great supplement to their existing process with which to improve their accuracy even more.

   The CFO support staff can start using the XGBoost model in parallel with the existing "Weighted Pipeline Amount" method to forecast $ Awards and Revenue. For the CFO group, continuing to use the "Weighted Pipeline Amount" method will allow the CFO and other business unit leaders to stay within their normal comfort zone, while at the same time testing XGBoost. This testing time period will also allow everyone to make sure the XGBoost continues to generalize well to new data. If XGBoost performs substantially better, then the "Weight Pipeline Amount" method can be dropped for good. On the other hand, if the results were varied, but XGBoost still provides value, then both methods can be used in parallel to supplement other indefinitely.

2. **Institute Procedures to Control Outliers:**

   To protect against the risks of the XGBoost model making very bad $ predictions: before finalizing any Award or Revenue forecasts, one can verify very large deal $ amounts with the Proposal Leaders before committing to the numbers. Proposal leaders are in the best position to cross-check predictions. A threshold can be set for very large amounts that should go through this process.

3. **Improve Salesforce Data that Feeds XGBoost:**

   At the same time, improve the Salesforce Data that feeds the XGBoost prediction model. The data can be greatly enriched and cleansed. For example, it would be good to have information about competitors bidding on the same proposals. It would also be good to have information about the win rates for specific Proposal leaders (for example rate a Proposal leader with a 90% win rate as a 10, and one with a 40% win rate as a 4). This new information could be used as features that would greatly improve the Precision and Recall for the "Win" predictions.

14. FURTHER RESEARCH

Additional research I would try in the future to improve the model is to split out Class 2 into three new classes. As you recall, Class 2 currently represents "Bid NO", "Cancelled" and "Closed Other".

Class 0 would still be "Losses", Class 1 would still be "Wins", but Classes 2, 3 and 4 would be the following:

- Class 0 = Losses
- Class 1 = Wins
- **Class 2** = Bid NO
- **Class 3** = Cancelled before Decision 1
- **Class 4** = Closed Other after Decision 1

The goal of having these extra classes is to see whether Precision and Recall improve for the model, as well as to get more information gain.

I would also continue working on better Feature Engineering, and trying other techniques (maybe Principal Components Analysis) for feature selection.

15. ACKNOWLEDGMENTS

Special thanks to Andrew Vaughan for mentoring me during this project. His suggestions and input were essential in guiding me towards the proper analysis and completion of this project.