

The CYO Project

Sales Opportunities Analysis

The CYO Project

Analysis Framework



- Let's play the role of a sales executive at an automotive supply wholesaler, and investigate a sales execution issue.
 - We have not been converting enough opportunities lately.
-
- We want to better understand our sales pipeline and which deals our sales teams can expect to win or lose,
 - so that we can better anticipate performance gaps.
-
- From the data that we pull out of our CRM database, we want to find the patterns in sales wins and losses,
 - and uncover what can lead to successful sales opportunities.

The CYO Project

Data (1 of 4)

1. The dataset is a sample provided by IBM in their Watson Analytics community that can be downloaded [here](#).

The "WA_Fn UseC_Sales Win Loss.csv" file is a dataset that covers sales activities for, amongst others, carrying out a win/loss analysis, a cost effective way of generating the insights to increase revenues and grow the business.

2. This sample has a format of 78,000 rows and 19 columns/variables as follows:

Variables	Description
Client Size by Employee Count	<p>Employee sized by number of clients</p> <ul style="list-style-type: none">• 1: < 1k• 2: [1K, 5K]• 3: [5K, 10K]• 4: [10K, 30K]• 5: $\geq 30K$

The CYO Project

Data (2 of 4)

Variables	Description
Client Size by Revenue	Client size based on annual revenue <ul style="list-style-type: none">• 1: < \$1M• 2: [\$1M, \$10M]• 3: [\$10M, \$50M]• 4: [\$50M, \$100M]• 5: ≥ \$100M
Competitor Type	An indicator if a competitor has been identified: Known, Unknown, None
Deal Size by Category	Categorical grouping of the opportunity amount (OpportunityAmountUSD) <ul style="list-style-type: none">• 1: < 10K• 2: [\$10K, 25K]• 3: [\$25K, \$50K]• 4: [\$50K, \$100K]• 5: [\$100K, \$250K]• 6: [\$250K, \$500K]• 7: ≥ \$500K
Opportunity Number	A unique generated number assigned to the opportunity
Opportunity Results	A closed opportunity is won or loss. Values could be Win/Loss Good example of a Target Field for Predict

The CYO Project

Data (3 of 4)

Variables	Description
Region	Name of the Region: Mid-Atlantic, Midwest, Northeast, Northwest, Pacific, Southeast, Southwest
Route to Market	The opportunities' route to market: Fields Sales, Other, Reseller, Telecoverage, Telesales
Supplies Group	Reporting supplies group: Car Accessories, Car Electronics, Performance & Non-auto, Tires & Wheels
Supplies SubGroup	Reporting supplies subgroup. Values are: Batteries & Accessories, Car Electronics, Exterior Accessories, Garage & Car Care, Interior Accessories, Motorcycle Parts, Performance Parts, Replacement Parts, Shelters & RV, Tires & Wheels, Towing & Hitches
Opportunity Amount (USD)	Sum of line item revenue estimates by sales representative in American currency
Sales Stage Change Count	Actually a count of number of times an opportunity changes sales stages (back and forwards)
Elapsed Days In Sales Stage	The number of days between the change in sales stages. The counter is reset for each new sales stage

The CYO Project

Data (4 of 4)

Variables	Description
Ratio Days Identified To Total Days	Ratio of total days the opportunity has spent in sales stage: Identified/Validating over total days in sales process
Ratio Days Qualified To Total Days	Ratio of total days the opportunity has been spent in sales stage: Qualified/Gaining Agreement over total days in sales process
Ratio Days Validated To Total Days	Ratio of total days the Opportunity has presence in sales stage: Validated/Qualifying over total days in sales process
Revenue From Client Past Two Years	Revenue identified from this client in past two years <ul style="list-style-type: none">• 0: 0• 1: [1, 50K]• 2: [50K, 400K]• 3: [400K, 1.5M]• 4: $\geq 1.5M$
Total Days Identified Through Closing	Total days the opportunity has spent in Sales Stages from Identified/Validating to Gained Agreement/closing
Total Days Identified Through Qualified	Total days the opportunity has spent in Siebel Stages from Identified/Validating to Qualified/Gaining Agreement

The CYO Project

Methodology (1 of 3)

1. Exploratory Data Analysis

In this section, we explore the data in two main steps:

- **Initial exploration**

- Dataset structure, variable formats,
- Missing values,
- Duplicated information,
- Correlation analysis.

The CYO Project

Methodology (2 of 3)

1. Exploratory Data Analysis

- **In-depth exploration for first insights**

In this sub-section, we will intuitively explore some features of our dataset such as:

- *Opportunity.Amount.USD,*
- *Region,*
- *Deal.Size.Category,*
- *Route.To.Market.*

to possibly uncover some patterns in the data and gain some insights.

Opportunity.Result is our dependent variable (prediction target).

The CYO Project

Methodology (3 of 3)

2. Create subsets for the project

We want to create two subsets as follows:

- sales dataset, which contains 90% of our sample dataset, to analyze our sales wins and losses.
- validation dataset, which is the remaining 10%, for the purpose of validation of our decision support model.

3. Predictive Model

In this section, we will go through a couple of Machine Learning methods to build a model to support our decision.

The CYO Project

Results and Discussion (1 of 20)

1. Initial data analysis exploration

- **Dataset structure and variable formats**

- So, we have in our dataset 78,025 observations of 19 variables, which is quite a number of possible predictors (18 independent variables in total considering the *Opportunity.Result* as the dependent variable).
- Some features have a wrong type as numeric instead of categorical variables, e.g. *Deal.Size.Category* is an integer when it should be category. We convert our variables to the right class.
- The dataset has no missing value so we don't need to handle this.

Opportunity.Number	Supplies.Subgroup	Supplies.Group	Region	Route.To.Market	Elapsed.Days.In
1641984	Exterior Accessories	Car Accessories	Northwest	Fields Sales	76
1658010	Exterior Accessories	Car Accessories	Pacific	Reseller	63
1674737	Motorcycle Parts	Performance & Non-auto	Pacific	Reseller	24
1675224	Shelters & RV	Performance & Non-auto	Midwest	Reseller	16
1689785	Exterior Accessories	Car Accessories	Pacific	Reseller	69
1692390	Shelters & RV	Performance & Non-auto	Pacific	Reseller	89

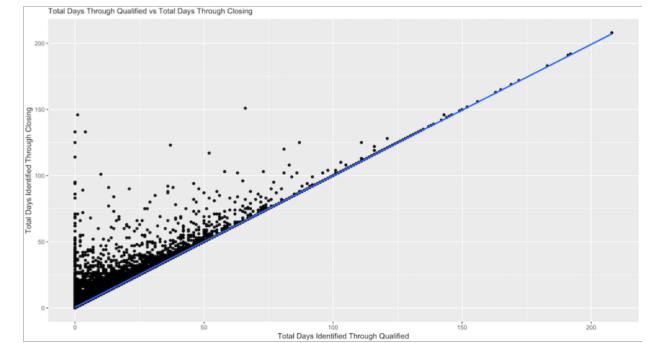
The CYO Project

Results and Discussion (2 of 20)

1. Initial data analysis exploration

- **Dataset structure and variable formats**

- Some information is duplicated. *Opportunity.Number* should be unique but is duplicated in some observations. As the duplication occurrences rate is very low and for the sake of simplicity, we just remove these duplications.
- A correlation analysis for both numeric and categorical variables show that some variables are interdependent. That is an useful information to bear in mind when we will perform our analytics later.



The CYO Project

Results and Discussion (3 of 20)



2. In-depth exploration for first insights

- **Success rate**

We have in our clean dataset a total 60,281 lost opportunities and 17,548 won opportunities so a global success rate of 23%, which confirms our low performance, that we have not been converting enough opportunities.

- **What could explain that?**

Let's explore some features, intuitively, whether they could be strong predictors of our win / loss outcomes.

The CYO Project

Results and Discussion (4 of 20)

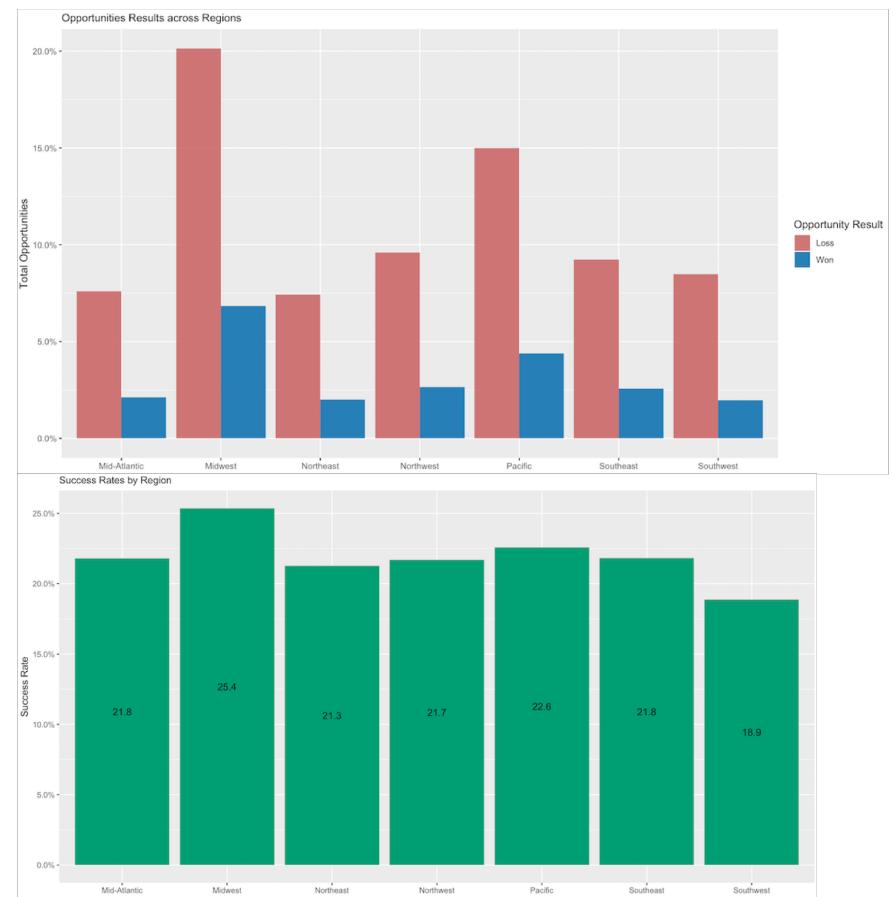
2. In-depth exploration for first insights

Deal conversions across regions?

We check our performance across regions, whether there are regions that underperformed so that they draw our figures to the bottom.

Midwest and Pacific are our biggest areas in terms of opportunity amounts. Nevertheless:

Our deal conversion rates across all regions are similarly low (between 19 and 25%), so there is surely room for improving our sales efficiency!



The CYO Project

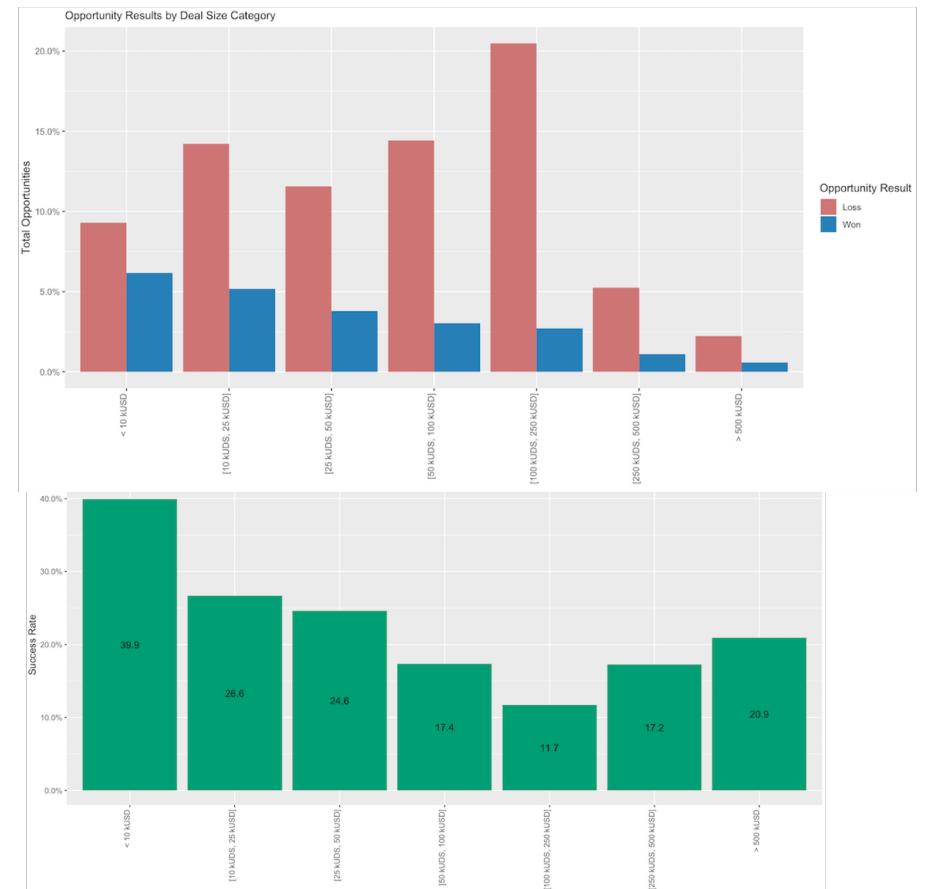
Results and Discussion (5 of 20)

2. In-depth exploration for first insights

Deal conversions by deal size category?

We check whether our performance is related to the deal size category. Could the size of the deal explain whether we successfully close the deal or not?

We won more opportunities for deal size < 10 kUSD and our success rate decreases as the deal size increases until a level of 100 to 250 kUSD.



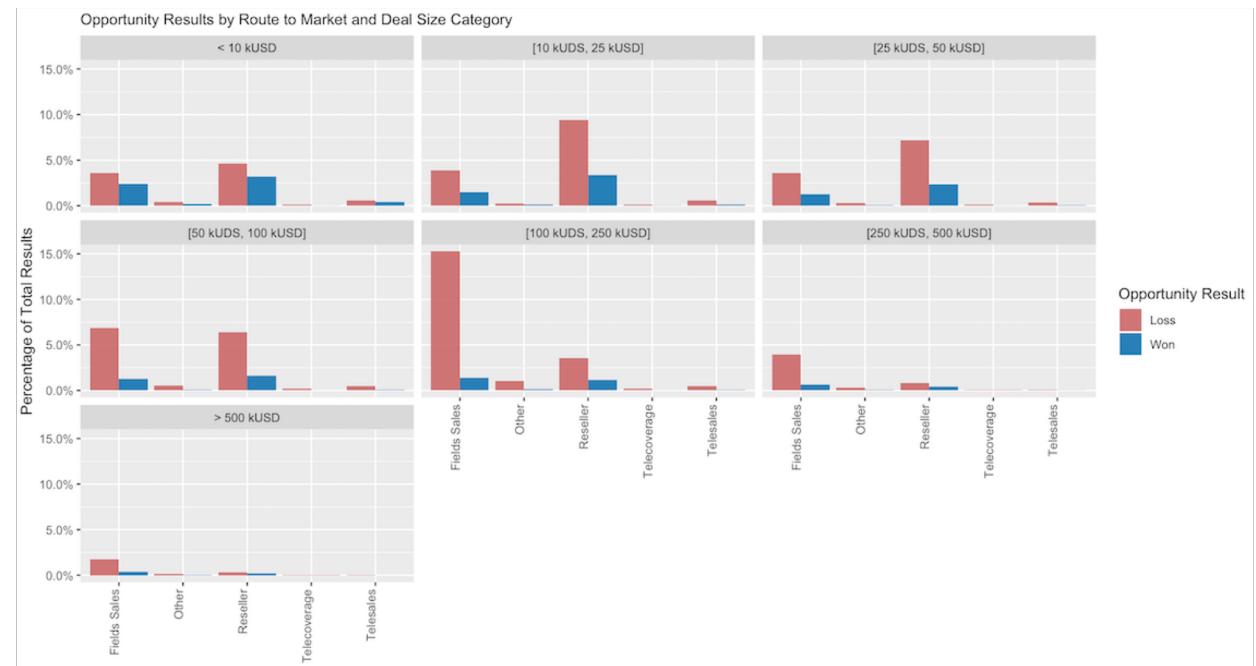
The CYO Project

Results and Discussion (6 of 20)

2. In-depth exploration for first insights

Deal conversions by sales channels?

We check our performance by route to market, whether there is a sales channel that is more relevant depending on the size of the deal.



The CYO Project

Results and Discussion (7 of 20)

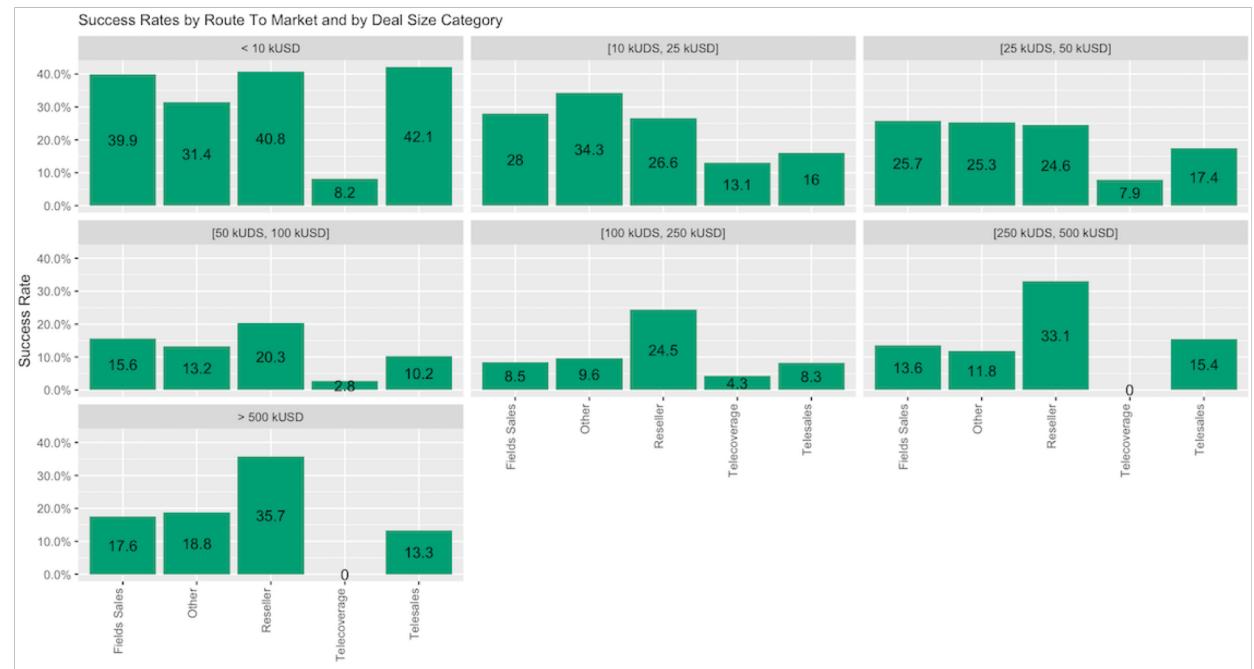
2. In-depth exploration for first insights

Deal conversions by sales channels?

There is **no clear trend of successful routes to market across deal sizes.**

However, based on our findings, we may consider shifting our resources as:

- Telesales for opportunities ≤ 10 kUSD,
- Field Sales for opportunities > 250 kUSD,
- Reseller for all opportunities,
- Telecoverage should be abandoned.



The CYO Project

Results and Discussion (8 of 20)

➤ Understanding what drives our sales!

- So, with a simple managerial approach, looking at a few variables, we have not been able to uncover the patterns that allow us to determine the successful sales profiles.



The best performance we could achieve was a modest 40% of deals conversion with Telesales route to market for opportunities < 10 kUSD.

- **We want to understand what drives our sales, which deals our sales team can expect to win or loose.** In other terms, we want to understand the **why** behind what's happening.

The CYO Project

Results and Discussion (9 of 20)



- With such a large dataset including 19 variables, we couldn't manually explore each and every variable, not even talking about possible combinations.
 - That is where we brought in **Machine Learning approaches** to help us to identify the most significant variables and predict the opportunity results.
- **Methods to perform a dimension reduction of our dataset so that we can identify the most significant variables.**

Two very common methods for identifying significant variables are **Decision Tree** and **Random Forests**.

- The Decision Tree best feature for analytics is that it is very **easy to interpret** and **results are actionable!**
- Random Forests improve the robustness of our predictions as they aggregate many Decision Trees.

The CYO Project

Results and Discussion (10 of 20)

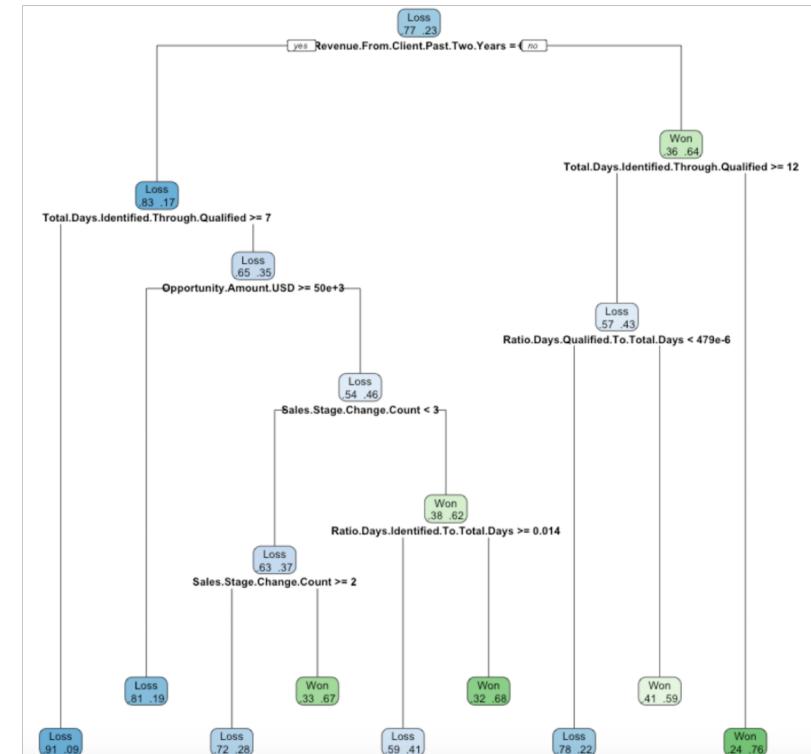
3. Predictive Models

- **Decision Tree**

We achieved a decent overall accuracy of 84% in predicting win/loss opportunities, with a negative predictive value (the proportion of predicted won opportunities which are real won deals) of 69%.

More importantly, the Decision Tree helped us to understand not only the relationships and associations between features but also the decision rules to generate that tree.

We will focus on two or three features only as we want to keep our insights interpretable and above all actionable.



The CYO Project

Results and Discussion (11 of 20)

3. Predictive Models

- **Most significant features as given by the Decision Tree model**
 - So the feature of first importance is ***Revenue.From.Client.Past.Two.Years***, i.e. the business that we've had with that Customer during the past two years,
 - then the second feature is ***Total.Days.Identified.Through.Qualified***, i.e. the number of days to qualify an opportunity from its identification.
We may note that *Total.Days.Identified.Through.Closing* can be used as a second feature too. These two variables are strongly correlated as we have seen before and basically bear the same information.
 - The third significant feature is ***Sales.Stage.Change.Count***, i.e. the number of times an opportunity changes sales stages (back and forwards) in the sales pipeline.

The CYO Project

Results and Discussion (12 of 20)

3. Predictive Models

- **Random Forest**

With a Random Forest model, we improved our overall accuracy to 88% with a negative predictive value of 76%.

Our top 3 features for prediction are:

- ***Revenue.From.Client.Past.Two.Years,***
 - ***Total.Days.Identified.Through.Qualified,***
 - ***Elapsed.Days.In.Sales.Stage***, i.e. the number of days between the change in sales stages (the counter is reset for each new sales stage).
-

The CYO Project

Results and Discussion (13 of 20)



3. Predictive Models

- **Validation**

Through validation process, we achieved a final **overall accuracy of 88% for predicting win/loss opportunities and an accuracy of 77% for predicting won opportunities that are real won deals!**

Let's see now the insights we can gain from our best predictive model!

The CYO Project

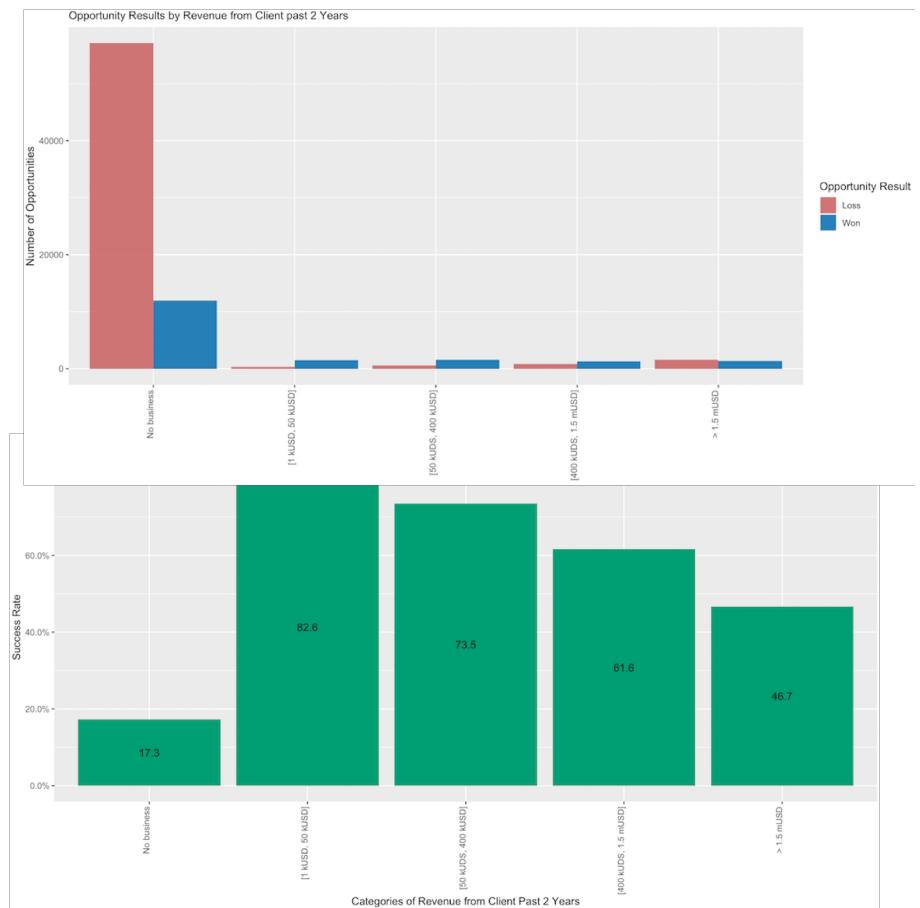
Results and Discussion (14 of 20)

3. Insights with one predictor

We used the most significant predictor identified by our models, i.e. ***Revenue.From.Client.Past.Two.Years.***

Looking at the Client purchase history, if they have bought from us less than 50,000 USD in the past 2 years, we have an **83% chance to successfully close the deal**.

When it comes to very big opportunities (≥ 1.5 mUSD), we close the deal half the time. On the other end, gaining new customers is a real challenge with a success rate of only 17%.



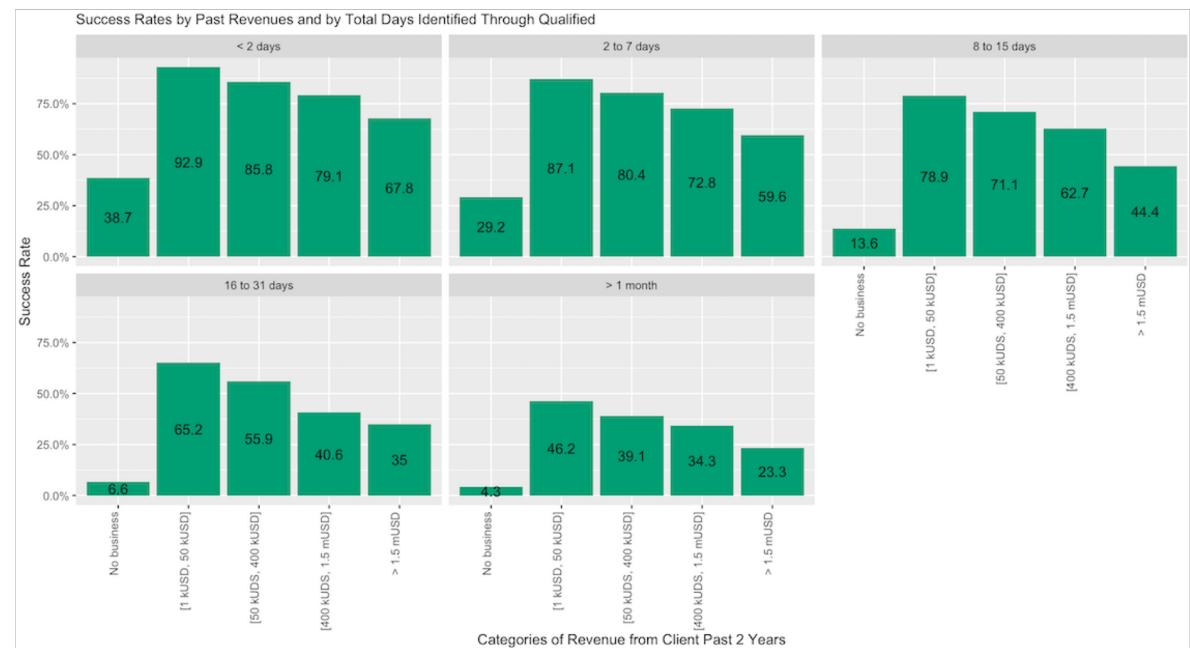
The CYO Project

Results and Discussion (15 of 20)

3. Insights with two predictors

We used the two most important predictors identified by our models, i.e.

- ***Revenue.From.Client.Past.Two.Years,***
- ***Total.Days.Identified.Through.Qualified.***



The CYO Project

Results and Discussion (16 of 20)

3. Insights with two predictors

- If we can qualify, within 2 days, an opportunity with customers having purchased for less than 50,000 USD in the last 2 years, we reach **a probability of 93% to successfully close the deal.**
- As a general rule, **the chances of winning a deal decreases as it stays longer in the pipeline.** This could help to formulate thresholds based on how many days a deal is in a pipeline and create alert mechanisms to expedite qualification.
- We also see the **same decrease trend with the increase of purchase history value, for a given qualification time frame.** For example, with an opportunity qualification of 2 to 7 days, we have an 87% chance of successful deal with customers valued at less than 50,000 USD and 60% with those at more than 1.5 mUSD.

The CYO Project

Results and Discussion (17 of 20)

3. Insights with two predictors

- We may also note that **an opportunity is more likely to result in a loss if the Client didn't buy anything from us within the last 2 years** but if we are able to qualify a deal within a week with a new Customer, we have more chance of success than our global (over the whole dataset) rate of 23%, as seen in the beginning of our analysis.



The CYO Project

Results and Discussion (18 of 20)

3. Insights with three predictors

We may want to uncover more complex relationships by adding more features, for example with three predictors as:

- ***Revenue.From.Client.Past.Two.Years,***
- ***Total.Days.Identified.Through.Qualified,***
- and ***Opportunity.Amount.USD.***

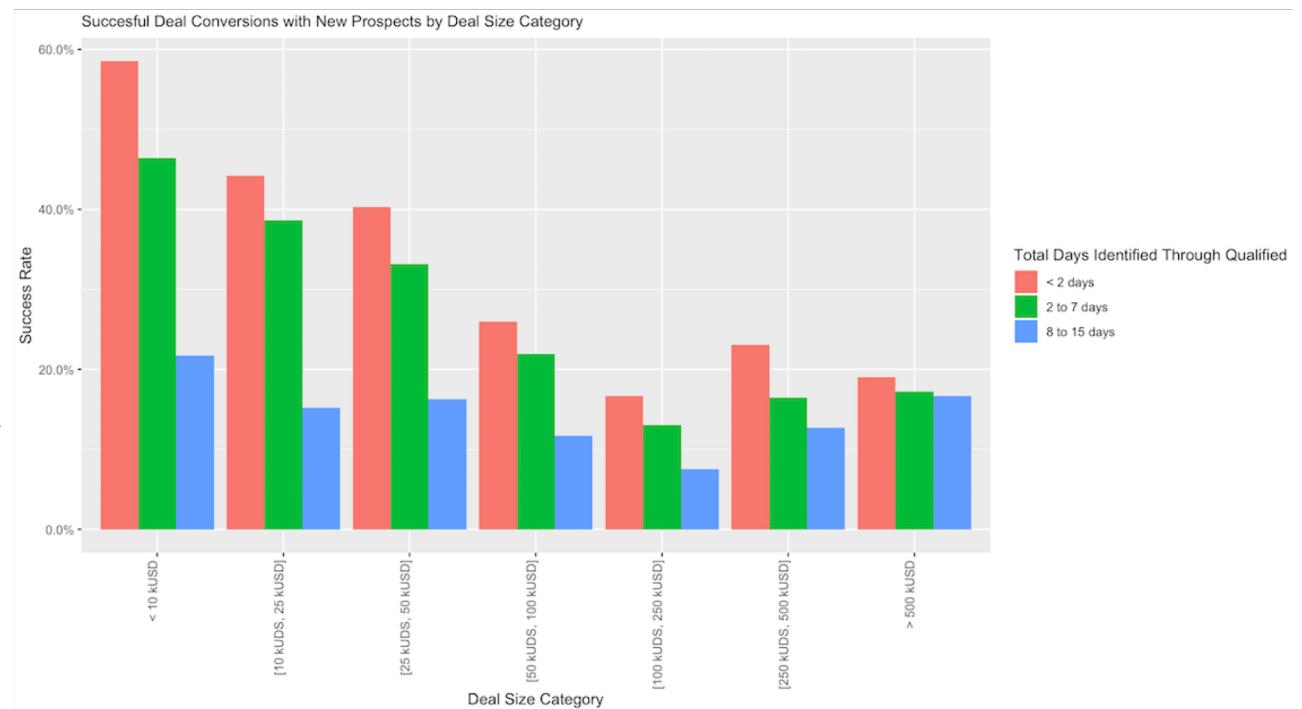
Let's say we want to know how we perform with new prospects, with whom we managed to qualify the opportunities, regardless of the USD value, within two weeks.

The CYO Project

Results and Discussion (19 of 20)

3. Insights with three predictors

- As we know by now, **the chances of successfully closing a deal are higher if we expedite the qualification of the opportunity.**
- Nevertheless, this finding is balanced by the size of the deal. For higher value deals, the impact of qualification speed show a more nuanced picture.



The CYO Project

Results and Discussion (20 of 20)

3. Insights with more predictors?

At this stage, adding more predictors won't help much but may degrade the interpretability of our findings.

Let's keep in mind that we want, above all, **insights that are relevant and actionable!**

The CYO Project

Conclusion (1 of 3)

- We started with a set of 78,000 rows and 19 variables of data extracted from our CRM and tried to intuitively interpret it with a **managerial approach**. We tried to understand what drives our sales and why we have not been converting enough deals.
- We looked at some variables that could be strong indicators of our sales performance. We looked at sales results by sales amounts, region, deal size category and route to market. We gained **some interesting insights but none of them uncovered success patterns**. The best performance we could achieve was a modest 42% deals conversion with Telesales route to market and opportunities of less than 10,000 USD.
- With such a large dataset, we couldn't realistically explore each and every variable to gain insights about what opportunities we can expect to win. We had to automatize our exploration process with **machine learning approach** to determine the most significant features, which could strongly predict the opportunities results, specifically the won deals.

The CYO Project

Conclusion (2 of 3)

- We tried Decision Tree and Random Forests models and achieved very good results. Random Forest yields an **overall prediction accuracy of 88% and 77% accuracy on won deals**. More importantly, Random Forest could drastically reduce the dimension of our dataset and provide the most significant features for predicting the opportunities results. Random Forest is a great fit for the job that we had in hands!
- So, we could **interpret our large initial dataset in terms that our sales managers can understand**.

We uncovered patterns about our opportunities, sales pipeline and what drives our win and losses. We built easy visualizations to help to **understand the profiles of the most likely successful sale opportunities**. For example, we realized that:

- the chances of winning a deal decreases as it stays longer in the pipeline,
- or that an opportunity is more likely to result in a win if the Client has purchased from us up to 50,000 USD.

The CYO Project

Conclusion (3 of 3)

- **These sales profiles are extremely valuable and more importantly, actionable in the hands of our sales teams.**

When reviewing their deal pipeline, our managers can anticipate gaps and correct their sales strategies accordingly. They can focus on the right deals and optimize their progression through the pipeline.

- Note that we could uncover more complex relationships by adding predictors according to their significance given by our predictive model, nevertheless we should always keep in mind the need of interpretability and at the end of the day, we want our **insights to be actionable by our sales managers!**

The CYO Project



For your
time reading
this report!