Sales Pipeline Overview

Mike Crowder

3/30/2019

Business to business (B2B) sales data has thus far been neglected in the machine learning revlolution over the past 5 years. B2B sales teams hold one of the greatest responsibilities in modern global business. The B2B sales force holds the key to the relationships to their largest customers. Our goal here in this section is to look at the pipeline and data and understand it.

Is Sales an Art or a Science?

Sales is neither an art or a science, but a process that can be managed just like any other business process. Their are many moving parts to not only an indvidual deal, but take that deal and multiple it by a few thousand over global territories and cultures you have the modern large global sales force.

With a proper customer relationship managment (CRM) system companies can collect millions of data points. That's just what is in the CRM system. Now companies have the ability to monitor sales calls with video through meeting software and voice, now giving us billions of data points. Sales needs the attention of your analytics or data science department.

About the data

This dataset is sample data from IBM's Watson Analytics. The data is from a CRM system that has been run through a process to make it somewhat model ready for analysis. The purpose of the data is to discover patterns with sales wins and losses. This could allow a Sales Operations team to communicate to executive leadership what factors contribute to winning and losing. That allows the orgainization to be proactive and possibly change the outcome.

This is a flat file that you could get out of a relational database through a data warehouse. Let's take a look at the data.

```r
pipeline <- read.csv("pipelineAnalytics.csv")
#head(pipeline)
```

```r
dim(pipeline)
```

```
## [1] 78025    19
```

The pipeline data is 78K rows with 19 variables. Let's explore some of the variables. Opp is short for opportunity.

Variable	Data Type	Description
Opportunity Number	Integer	Unique indentifier some opps may have mutliple products
Supplies Subgroup	Factor	Product subgroup
Supplies Group	Factor	Product grouping

Variable	Data Type	Description
Region	Factor	Business region in which opp was sold
Route to Market	Factor	The channel in which owns the opp
Elapsed Days In Sales Stage	Integer	The number of days an opp has been in the sales stage
Opportunity Result	Factor	Tells if the opp was won or lost
Sales Stage Change Count	Factor	How many times has the stage name changed
Total Days Identified Through Closing	Integer	The age of an opp
Total.Days.Identified.Through.Qualified	Integer	How long it took in days to qualify an opp
Opportunity Amount USD	Integer	The revenue over a 12 month period the opp repersents
Client Size By Revenue	Integer	Client size by the clients yearly revenue
Client Size by Employee Count	Integer	Client size by the number of employees
Revenue From Client Past Two Years	Integer	Client spend past two years
Competitor Type	Factor	Indicator if a competitor has been identified
Ratio Days Identified To Total Days	Integer	Ratio of total days the opportunity has spent in sales stage: Identified/Validating over total days in sales process
Ratio Days Validated To Total.Days	Integer	Ratio of total days the Opportunity has presence in sales stage: Validated/Qualifying over total days in sales process

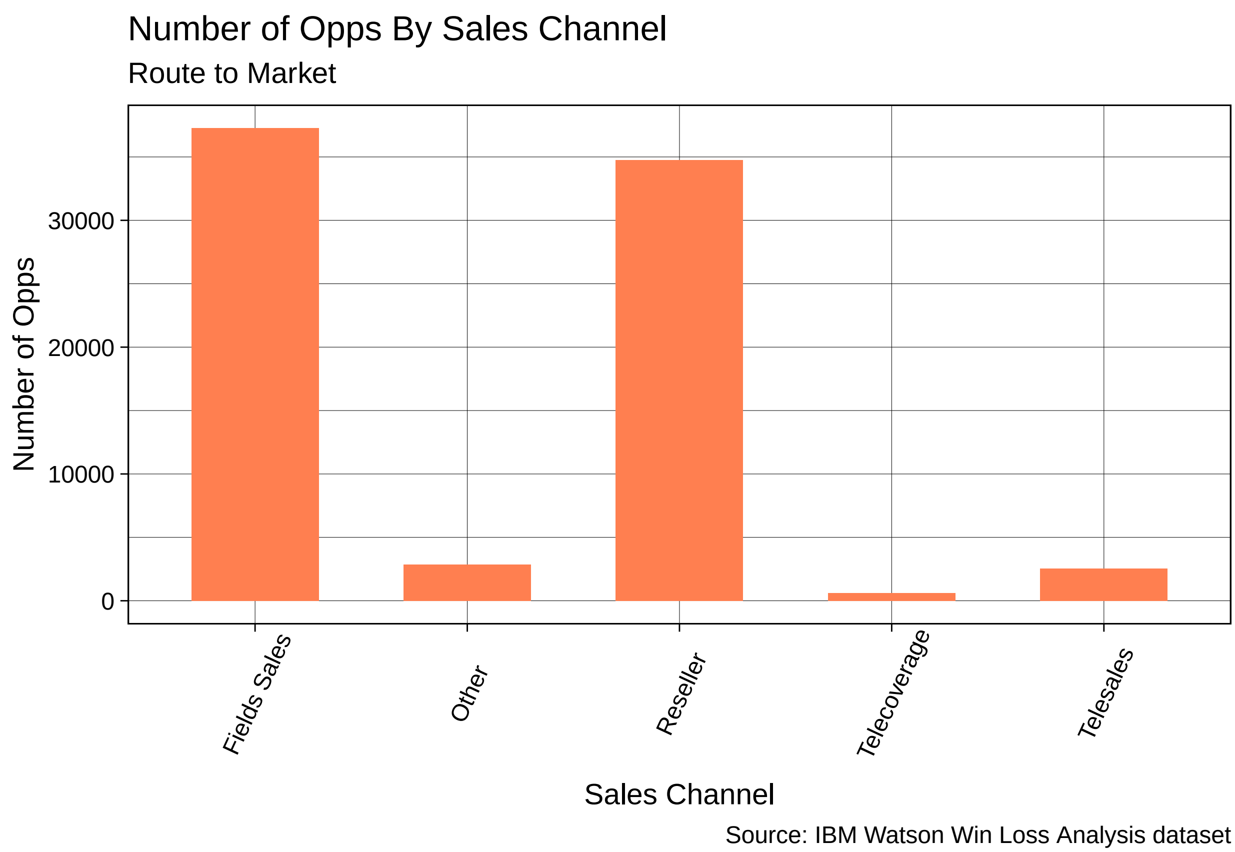
Variable	Data Type	Description
Ratio Days Qualified To Total Days	Integer	Qualified/Gaining Agreement over total days in sales process
Deal Size Category	Integer	Categorical size of the opportunity size

Analysis Overview

Lets visualize some of this information to do this we are going to use ggplot.

```
library(ggplot2)
theme_set(theme_linedraw())
options(scipen = 999)
```

```
w <- ggplot(pipeline, aes(Route.To.Market))
w + geom_bar(width = 0.6, fill = "coral") +
  labs(title = "Number of Opps By Sales Channel",
       subtitle = "Route to Market",
       caption = "Source: IBM Watson Win Loss Analysis dataset",
       y = "Number of Opps",
       x = "Sales Channel") +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6))
```

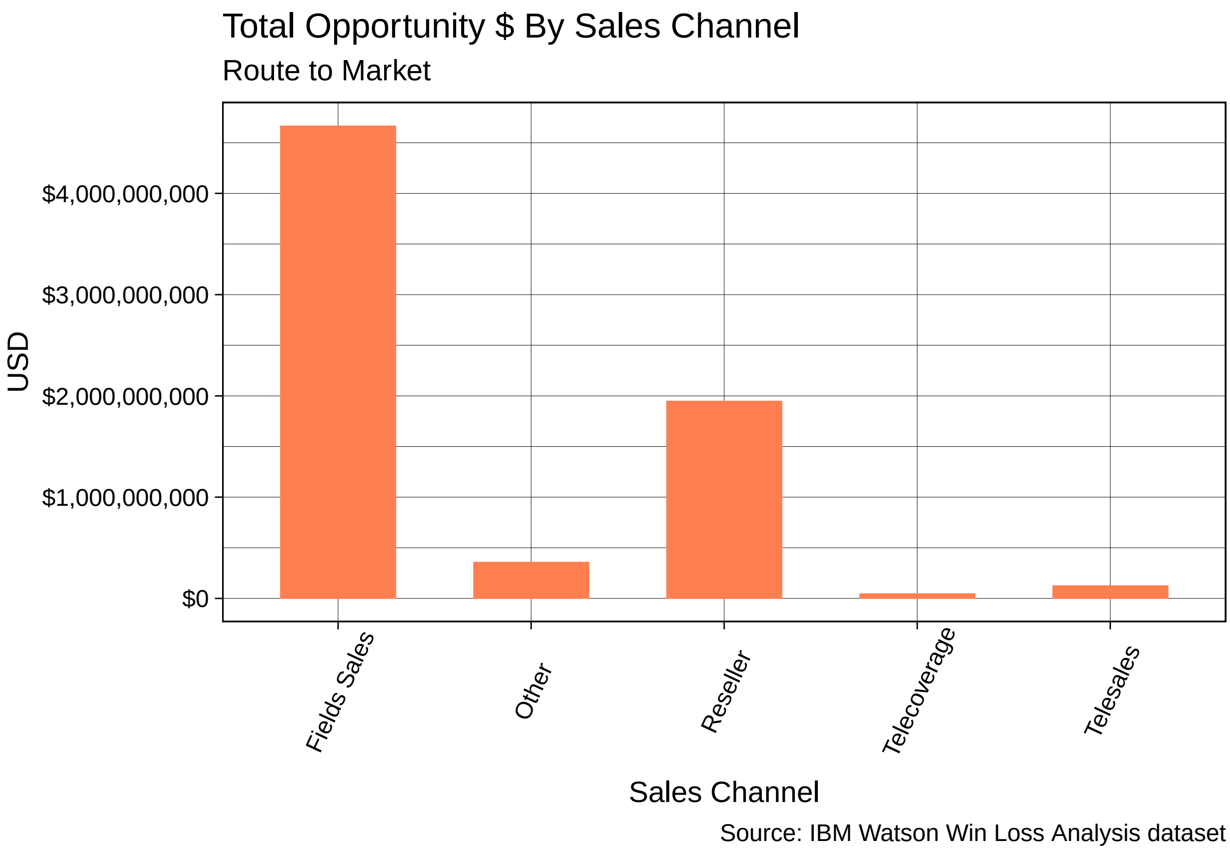


```r
table(pipeline$Route.To.Market)
```

```
##
## Fields Sales        Other     Reseller Telecoverage    Telesales
##        37262         2856        34758          619         2530
```

Most of the opportunties come from field sales and the reseller channel.

```r
w1 <- ggplot(pipeline, aes(x = Route.To.Market, y = Opportunity.Amount.USD))
w1 + geom_bar(stat = "identity", width = 0.6, fill = "coral") +
  labs(title = "Total Opportunity $ By Sales Channel",
       subtitle = "Route to Market",
       caption = "Source: IBM Watson Win Loss Analysis dataset",
       y = "USD",
       x = "Sales Channel") +
  scale_y_continuous(labels = scales::dollar) +
  theme(axis.text.x = element_text(angle = 65, vjust = 0.6))
```




```r
tapply(pipeline$Opportunity.Amount.USD, pipeline$Route.To.Market, FUN = sum)
```

```
## Fields Sales        Other     Reseller Telecoverage    Telesales
##   4669514062    357806122   1948586216     46515153    127575717
```

This is a pretty large dollar value of pipeline. The total pipeline for this sample for field sales for example is 4.7 billion dollars. The reseller channel is worth 1.95B billion dollars. This looks like a typically B2B pipeline in that our field sales generate the most opportunity while we have other channels that might and this is a guess that smaller opportunties are managed by Resellers. Let's check that guess.

```r
rtmTBL <- table(pipeline$Route.To.Market, pipeline$Client.Size.By.Revenue)
rtmTBL
```

```
##
##                    1     2     3     4     5
##   Fields Sales 27112  1837  2075  2678  3560
##   Other         2160    81   146   178   291
##   Reseller     27417  1880  2469  1690  1302
##   Telecoverage   558    10    16    13    22
##   Telesales     2257    33    50    77   113
```

```r
prop.table(rtmTBL, 1)
```

```
##
##                         1          2          3          4          5
##   Fields Sales 0.72760453 0.04929955 0.05568676 0.07186946 0.09553969
##   Other        0.75630252 0.02836134 0.05112045 0.06232493 0.10189076
##   Reseller     0.78879682 0.05408827 0.07103401 0.04862190 0.03745900
##   Telecoverage 0.90145396 0.01615509 0.02584814 0.02100162 0.03554120
##   Telesales    0.89209486 0.01304348 0.01976285 0.03043478 0.04466403
```

The relationship is not with Client Size By Revenue.

```r
cltTBL <- table(pipeline$Route.To.Market, pipeline$Revenue.From.Client.Past.Two.Years)
cltTBL
```

```
##
##                    0     1     2     3     4
##   Fields Sales 32446   590   861  1233  2132
##   Other         2564    48    54    61   129
##   Reseller     31332  1082  1091   731   522
##   Telecoverage   591    10    12     1     5
##   Telesales     2275    52    65    66    72
```

```r
prop.table(cltTBL, 1)
```

```
##
##                          0           1           2           3           4
##   Fields Sales 0.870753046 0.015833825 0.023106650 0.033090011 0.057216467
##   Other        0.897759104 0.016806723 0.018907563 0.021358543 0.045168067
##   Reseller     0.901432764 0.031129524 0.031388457 0.021031130 0.015018125
##   Telecoverage 0.954765751 0.016155089 0.019386107 0.001615509 0.008077544
##   Telesales    0.899209486 0.020553360 0.025691700 0.026086957 0.028458498
```

Nor is it by spend...

```r
oppTBL <- table(pipeline$Route.To.Market, pipeline$Deal.Size.Category)
oppTBL
```

```
##
##                   1     2     3     4     5     6     7
##   Fields Sales 4697  4235  3794  6316 12986  3592  1642
##   Other         449   304   293   486   907   288   129
##   Reseller     6081  9965  7442  6251  3640   970   409
##   Telecoverage   97   107    89   144   142    32     8
##   Telesales     771   512   350   431   399    52    15
```

```r
prop.table(oppTBL, 1)
```

```
##
##                         1           2           3           4           5
##   Fields Sales 0.126053352 0.113654662 0.101819548 0.169502442 0.348505180
##   Other        0.157212885 0.106442577 0.102591036 0.170168067 0.317577031
##   Reseller     0.174952529 0.286696588 0.214108982 0.179843489 0.104724092
##   Telecoverage 0.156704362 0.172859451 0.143780291 0.232633279 0.229402262
##   Telesales    0.304743083 0.202371542 0.138339921 0.170355731 0.157707510
##
##                         6           7
##   Fields Sales 0.096398476 0.044066341
##   Other        0.100840336 0.045168067
##   Reseller     0.027907244 0.011767075
##   Telecoverage 0.051696284 0.012924071
##   Telesales    0.020553360 0.005928854
```

It looks like there is some relationship between the channel and how large the opportunity is. Not enough information to go off of since we could have a global business and they use resellers even with large opportunities. It would appear that telesales is used for smaller opportunties.

Basic Sales Metrics

Here we are going to take a look at some basic sales metrics like Close Rate and time to close. Lets start with Close Rates.

Close Rates

Close Rates in sales is very important. Its often how we start a basic forecast in B2B sales. Close rate is simple the number of opportuntiies sold or won divided by the total number of opportunties.

We can also evaluate the preformance of some types of sales resources with Close Rates. Take Close Rates with a grain of salt when you are evaluating sales resources that work large complex opportunties. They work very well when we look at a segment of a sales team or business unit.

Let's take a look at close rates by opportunity size. This is a breakdown of opportuntiy size based on the source data.

Value	Bin Size
1	< \$10K
2	\$10K - \$25K
3	\$25K - \$50K
4	\$50K - \$100K
5	\$100K - \$250K
6	\$250K - \$500K
7	> \$500K

```r
DScloseRateTBL <- table(pipeline$Opportunity.Result, pipeline$Deal.Size.Category)
DScloseRateTBL
```

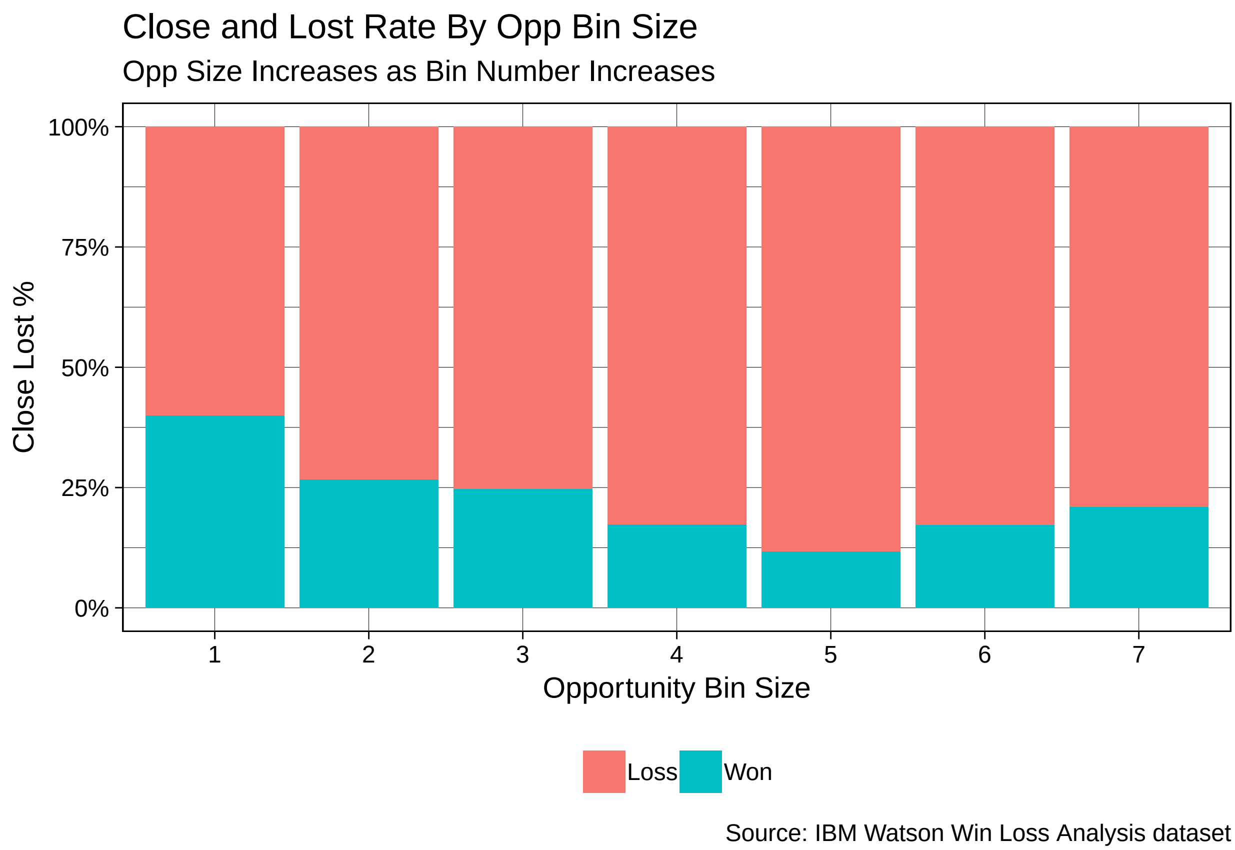
```
##
##            1     2     3     4     5     6     7
##   Loss  7264 11079  9016 11262 15954  4082  1741
##   Won   4831  4044  2952  2366  2120   852   462
```

```r
prop.table(DScloseRateTBL, 2)
```

```
##
##              1         2         3         4         5         6
##   Loss 0.6005788 0.7325927 0.7533422 0.8263869 0.8827044 0.8273206
##   Won  0.3994212 0.2674073 0.2466578 0.1736131 0.1172956 0.1726794
##
##              7
##   Loss 0.7902860
##   Won  0.2097140
```

```r
DScloseRateDF <- as.data.frame(prop.table(DScloseRateTBL, 2))
colnames(DScloseRateDF) <- c("Outcome", "OppBinSize", "CloseLostRate")
```

```r
dscrBar <- ggplot() + geom_bar(aes(y = DScloseRateDF$CloseLostRate,
  x = DScloseRateDF$OppBinSize, fill = DScloseRateDF$Outcome),
  data = DScloseRateDF, stat = "identity") +
  labs(title = "Close and Lost Rate By Opp Bin Size",
      subtitle = "Opp Size Increases as Bin Number Increases",
      caption = "Source: IBM Watson Win Loss Analysis dataset",
      y = "Close Lost %",
      x = "Opportunity Bin Size") +
  scale_y_continuous(labels = scales::percent) +
  theme(legend.position = "bottom", legend.title = element_blank())
dscrBar
```



From the Close and Lost Rate chart we can see as opportuntiy size increases the close rate decreases. Which is what happens at most sales orgs. We can see that most of the opportunties are in the bin size groupings of

4 and 5. This also happens to be the lowest close rates.

Now this is a combined sales force with resellers, telesales, etc. We could break continue to break this down, but that is what we have advanced analytics for. Let's take a look at this data from a data science prospective in the next section.