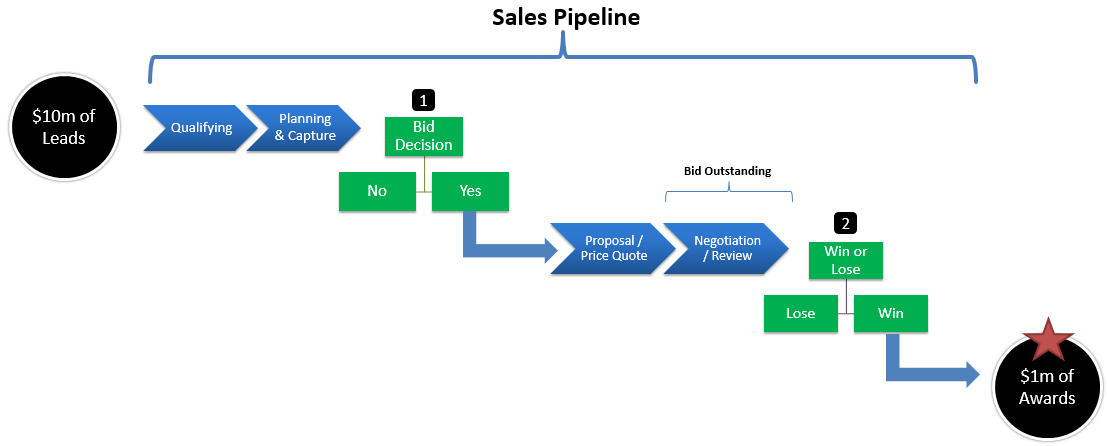
**Capstone Project: Using the Sales Pipeline to predict the $ amount of “Awards” won.**

**Project Background**:

The diagram pictured below is a visual representation of a company’s sales pipeline. Using a made up number, suppose that $10m in leads (or opportunities) enter the pipeline. The leads will progress through the “Qualifying” and the “Planning & Capture” stages, whereby proposal analysts collect data on those leads. 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. Those that are “YES” for BID will progress through the next stages of the pipeline. At the end of the Proposal / Price Quote stage, whereby a proposal is submitted to the client, the opportunity is said to be “Outstanding”, meaning we are waiting for the client to provide feedback for “Negotiation / Review”. Finally, Decision 2 – the Win or Lose Decision will be made. If we lose the bid, the $ amount falls out of the pipeline, while if we win the bid, the $ amount of the opportunity becomes an “Award”. In hypothetical example pictured in diagram, we started with $10m in opportunities or leads, and ended up with $1m in “Awards” or opportunities won.

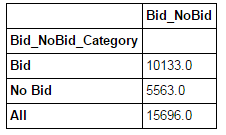


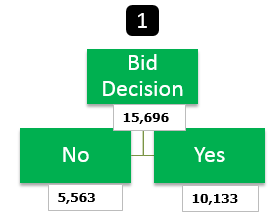
**Data Set**:

* I have 6 years of data
* 15,696 rows progressed through Decision 1 – the Bid Decision
* 9,096 rows progressed through to Decision 2 – the Win or Lose Decision

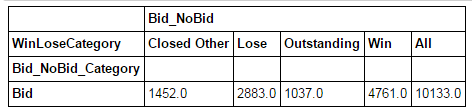
# Can you count something interesting?

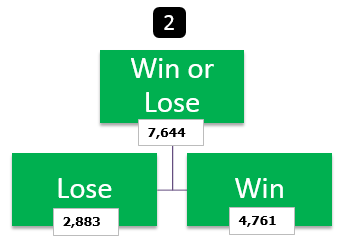
1. 65% of leads get “Bid on” (10,133 / 15,696) as pictured in below Data Frame and diagram.





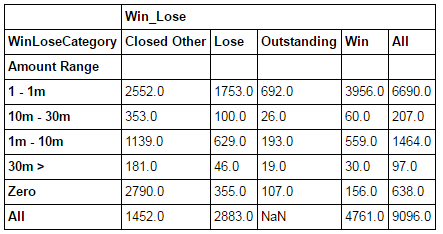
1. As shown in Data Frame pictured below, of the 10,133 total "Bid", 1,037 are "Outstanding" (i.e. not decided on yet), and 1,452 are "Closed Other" (i.e. client cancelled it before a Win / Lose decision was made). Therefore, only 7,644 are eligible for a "Win Rate" calculation. Alternatively you can simply add the Wins of 4,761 plus the Loses of 2,883 to get to the same 7,644. Therefore, the Win Rate is 62% (i.e. 4761 / 7644).

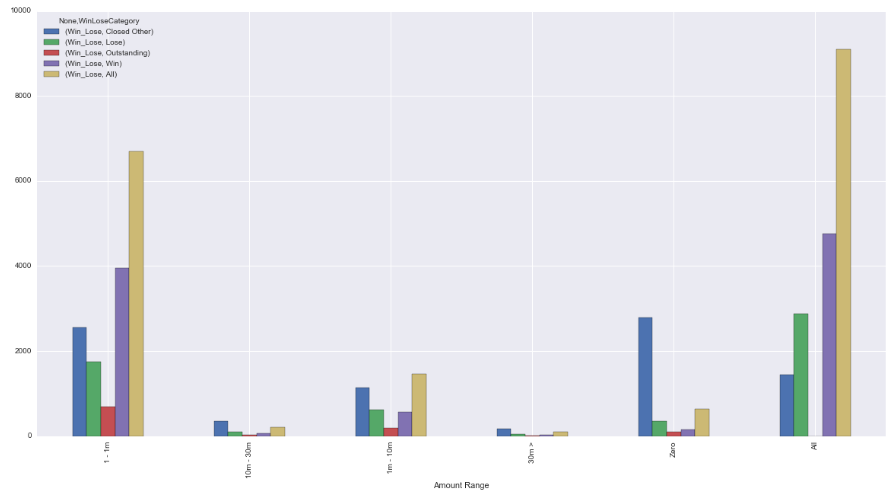




1. The below Data Frame and bar plot count the Wins and Losses by Amount Range. The Amount Range is the estimated $ contract amount of the opportunity (lead). As you can see, the opportunities whose amounts are < 1- 1m (i.e. greater than zero and less than $1m), have more wins (3,956) than losses (1,753). However, in the other higher amount ranges, there are more losses than wins. One anomaly I do notice is that there are amounts that zero (labelled Zero in below Data Frame). We actually won 156 opportunities with a zero Amount. Why is the Amount zero? Is it because the Amount was not agreed upon? Or, is the data wrong and the Amount never got recorded.

**Hypothesis to test**: do we have more wins in amounts less than $1m because amounts over $1m are more competitive?



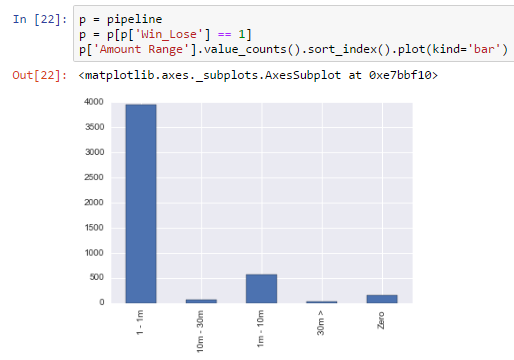


# Can you find some trends (high, low, increase, decrease, anomalies)?

1. The range in the “Amount” column is very broad, with there being amount with a $ value of zero to a max of $4 billion. This makes it complicated to plot the Amount column and have the graph look correctly.
2. One anomaly I do notice is that there are 3,408 observations in the “Amount” column with a $ value of zero. We actually won 156 opportunities with a zero Amount. Why is the Amount zero? Is it because the Amount was not agreed upon at the time we won the Award? Or, is the data wrong and the Amount never got recorded. I may consider excluding these 3,408 observations completely from the Data Set

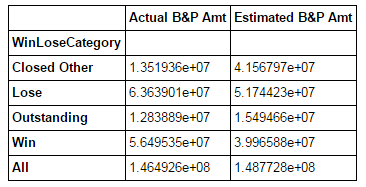
# Can you make a bar plot or a histogram?

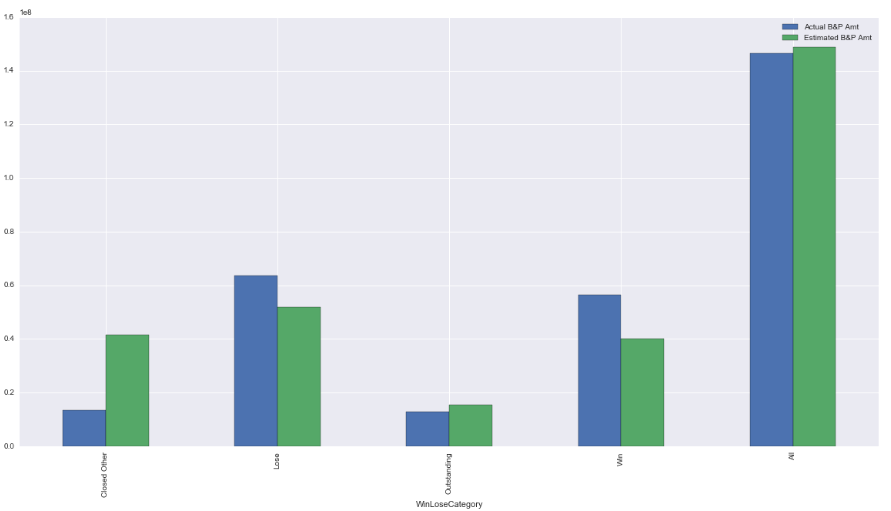
1. Yes, the below bar plot shows the “Amount Range” column by count of Wins (wins = 1). As you can see, there’s a lot more wins in the Amounts in the 1 – 1m range.



# Can you compare two related quantities?

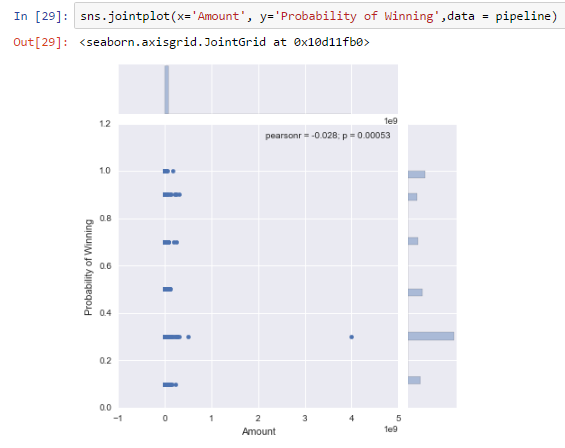
1. Yes, the below DataFrame and bar graph compare “Actual B&P Amt” ($ amount spent on Business & Proposal bids) versus “Estimated B&P Amt”. Interestingly, we spent more on B&P (both Actual and Estimated) for Losses than on Wins. Therefore, there were 2,883 Losses and we spent $63m on Actual B&P. On the other hand, we had 4,761 wins and we only spent $56m on Actual B&P. This lower B&P $ amount may be related to the fact that we have more wins in Amounts that are less than $1m (as mentioned in point 6 above), and the lower the $ amount, the less we are likely to spend on B&P.





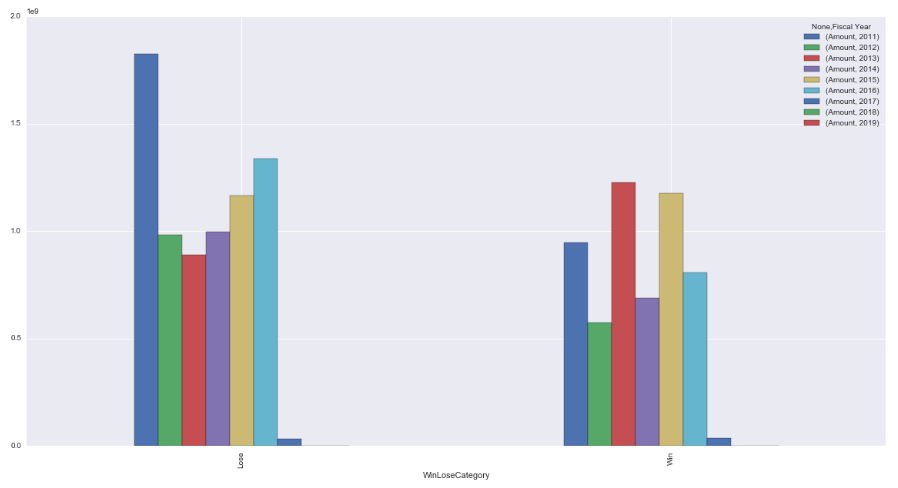
# Can you make a scatterplot?

1. Yes, I made a scatterplot between the “Amount” column and the “Probability of Winning” (this column is labelled by the project leaders and is not necessarily an actual probability) column. Unfortunately given that the “Amount” column has such a large range (with max being $4 billion), it makes the scatter look a bit weird. Although below graph includes all the $ 4 billion amount, I also did a scatter whereby I removed the $ 4 billion opportunity. My conclusion is that there is basically no correlation between the “Amount” column and the “Probability of Winning” column. In other words, all different kinds of amounts can have different probabilities of winning (i.e. it’s not like higher amounts get labelled as having a higher probability of winning).

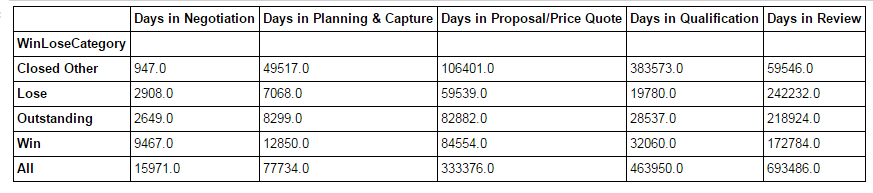


# Can you make a time-series plot?

1. Yes, the below time-series bar plot, show the $ amount of wins and losses by fiscal year. There is plenty of up-down variability in the Win category, while in the Lose category there seems to be an upward trend from 2013 to 2016, which makes sense because as the company grows you would want to make more bids and that can cause more losses, and hopefully more wins (but not necessarily).



1. The below DataFrame sums up the number of days spent in each of the pipeline stages summarized by Win / Lose Category.
   1. Comparing Wins versus losses, Wins spend more days in each of the pipeline stages, which implies that more care is being given to the Wins (by proposal analysts or for other reasons), and that in itself may explain why we won those opportunities.



# Conclusions:

Above I have summarized some interesting facts about the data, such as:

* Amounts less than $1m have a higher count of wins.
* B&P amount spent is lower for wins, which is counterintuitive at first, but makes more sense if you consider that amounts lower than $1m have a higher count of wins (and the lower amounts should have lower B&P expenditures).
* Comparing Wins versus losses, Wins spend more days in each of the pipeline stages which implies that more care is being given to the Wins (by proposal analysts or for other reasons), and that in itself may explain why we won those opportunities.

Although these are some interesting points, I don’t think that a complete story of what is going on will be complete until I start applying some inferential statistics to the data or even doing some machine learning to try to predict wins. This is to be continued.