1. Situational size-up. What factors are important to understand? What are the key issues facing the firm?

* Basics
  + Who – Sanjay Shelvankar (Cofounder and CEO) and top management
  + What – use analytics to predict if someone will accept an offer or not
  + When – April 30, 2015
  + Where – India for the talent acquisition process
  + Why – talent acquisition is tough, it takes time and effort for many organizations. Analytics might help eliminate wasted time and money
* Situational sizeup
  + Talent acquisition for IT companies in India that provides viable recommendations to its customers
  + Business process re-engineering is at company’s core
    - People capability
    - Process maturity
    - Technology adoption
  + Number of people not joining the company varied between 15% and 35% of all people who accepted the offer
  + For one client 12,000 offers were rolled out but 30% did not join
  + 15 hours spent on candidate leads to a loss of 54,000 hours for one client
  + Renege increased cost of hire by 10 to 15%
* Key issues
  + key drivers that influence the candidate joining a company
  + factors to predict renege
  + devise a predictive algorithm to calculate probability of acceptance and joining after acceptance

1. Logically, based on your own understanding of business and hiring, what factors could help explain why a job candidate could be a no-show after accepting a job? Provide a list and justify your response.

Based on my understanding of business and hiring, I believe that the factors that explain why a job candidate could renege ties mainly with the interaction with the employee immediately after the offer is made.

Factors that could help explain renege:

* Candidate source
  + This may help show patterns for different types of employees
  + Direct employees should be less likely to renege because there may be a personal level of accountability when reneging
  + Agency candidates should be more likely to renege because they do not have a personal connection to the company
* Relocation
  + A candidate forced to relocate may have a harder time justifying moving locations
  + This should be dependent on the age of the employee or if the candidate has a family or supports any dependents
  + Older people tend to have a family, know where they want to live and usually don’t make big changes to their job
* Line of business
  + Certain lines of business have a higher degree of stability and value

I believe that the capability of sentiment analysis can be leveraged with the six-stage questionnaire to help identify a potential candidate that might renege. However, since that data is not available, we don’t have to worry about analyzing it.

1. Examine your data. Do missing data issues exist? How should they be dealt with? What relationships exist between the measures you have available the reneging? Based on your answer to Q2, what measures should be included and why? Are any key factors missing? Are there any obvious interactions among the measures that might be useful?

Missing data issues exist. For example, Duration to Accept is a feature that is missing over 2700 values. It also contains negative values, which shouldn’t exist for duration. My understanding is that these were errors and since their size was insignificant, I decided to compute the absolute of this column. From my understanding and tests, the best and most simple way to solve missing values is by replacing any missing value with the mean from the data set.

Missing values also exist for Percent hike offered, hike expected and naturally the hike difference. My understanding is that this issue may exist if either party decided to not answer. Since the values missing is less than 10% for each column, I decided to fill these columns with 0.

Measures to include for sure are the source of the employee, age, type of job, location, relocation and line of business.

It would also be invaluable to have and analyze questionnaire data. It is also likely that the factors that effect one company may not affect another company. Therefore, data from different clients would help explain and build a better model with effective generalization.

Employee salary, total time interviewing, number of interview rounds, position title and more details about the employee or job could help.

Columns that would have a relationship:

* CTC columns are obviously all related, the difference should be offered minus expected
* Candidate source could also be an important factor, from my understanding, employees from certain sources could be more likely to join than others
* There should also be a relationship between Notice Period and Duration to Accept, longer Duration to accept should be tied with longer duration periods

1. Build your model. What measures/factors were significant? Is your model valid? What is your model’s Sensitivity, Specificity, and Accuracy?

While building the model, I learned that there is a strong statistical significance between DOJ Extended, Notice Period, Offered Band (1 and 2), Price Hike Expected, Candidates that were Employee Referrals, across multiple Lines of Business and Age. This is based on Logit1 which included all values.

I attempted to build a better model which tried to specify the key factors. Based on my Logit1 model and a lot of testing. The critical factors are Notice Period, LOB, Offered Band, if the employee extended their acceptance date and the price hike that the candidate expected to receive.

My specificity is 78%, sensitivity is 65% and accuracy is 75%.

1. If costs for reneging are 3 times the costs of a normal non-renege hire, what thresholds for classifications might be better? Justify.

The candidates that are in band E1, E2 and E3 have some of the best indicators of whether they join or not. We can also look at factors such as Line of Business where the categorical values of SERV, CORP, AS and INTRA are excellent indicators. Candidates from direct and referred sources were also good indicators. Notice period would be one of the thresholds that we could use to correctly predict when to predict a renege hire.

Therefore, anyone with those specific bands and lines of business with notice periods of greater than 20 should not be selected.

**Recommendation**

Based on my model and research, the ideal predictors for finding renege candidates are Notice Period, Offer Band, Percentage Hike Expected, Candidate source and LOB. Therefore, the ideal LOGIT model to uses these variables.

Anyone that is from offer band E1, E2 and E3 should be closely monitored for predicting renege. As well as candidates sourced directly or referrals. We also should look at Lines of Businesses such as SERV, AS, CORP and INFRA. Younger candidates also tend to renege more often. Candidates that extend their duration of acceptance are more likely to renege. Notice periods of within 20 days is typically when people accept and join.

The Logit2 model can predict with 75% accuracy with an error rate of 16.5%. The candidates that were predicted to join and ended up joining was 65%. The candidates that we predicted wouldn’t join and did not was 78%. Therefore, this model is a good predictor to save the company time and money since it can predict reneges 78% of the time.