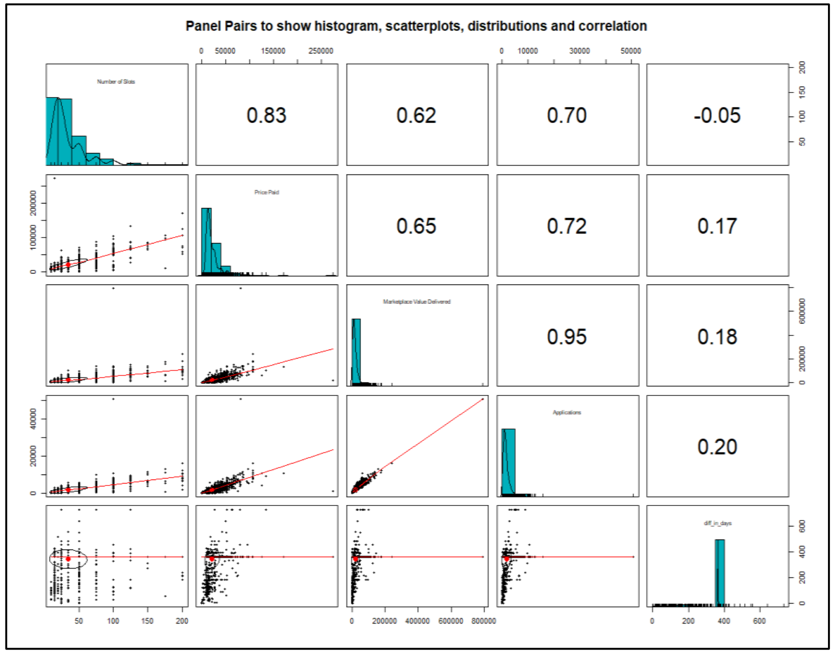
Glassdoor Data Assignment - Job Slot Retention Analysis

Vikas Srikanth

vxs180022@utdallas.edu

University of Texas - Dallas

Part 1: Job Slot Sales and Performance Distribution



Plot on the previous slide shows the Panel Pairs plot for the important variables in the Data Set (Diff_in_days is a derived column; obtained by subtracting End Date and Start Date)

Plot clearly shows that –

1. **There is <u>strong positive correlation</u> between**

a. Number of slots and Price Paid

b. Applications and Market Value Delivered

c. Number of slots and Market Value Delivered

d. Number of Slots and Application

e. Price Paid and Applications

f. Price Paid and Market Value Delivered

2. **There is a <u>weak positive correlation</u> between –**

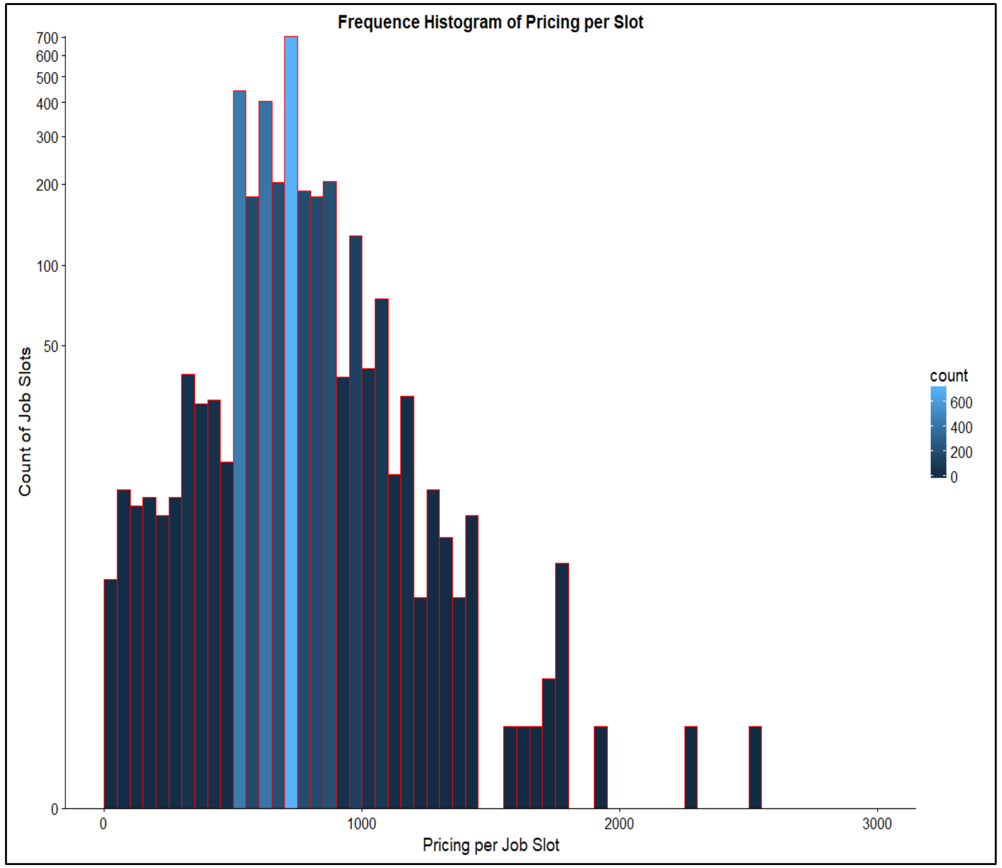
a. Number of Days (Diff_in_Days) and Applications

b. Number of Days and Market Value Delivered

c. Number of Days and Price Paid

3. **There is a <u>very weak negative/no correlation</u> between Number of Slots and Number of Days**

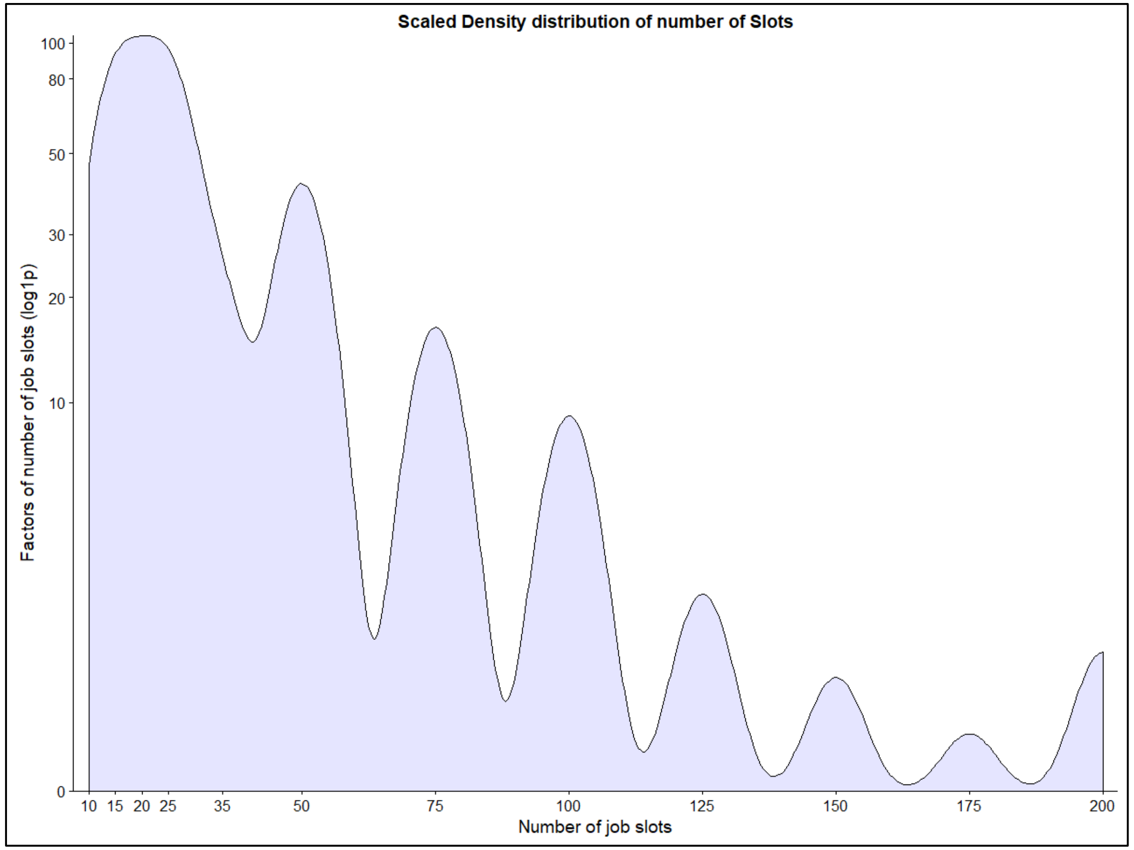
Part 1: Job Slot Sales and Performance Distribution

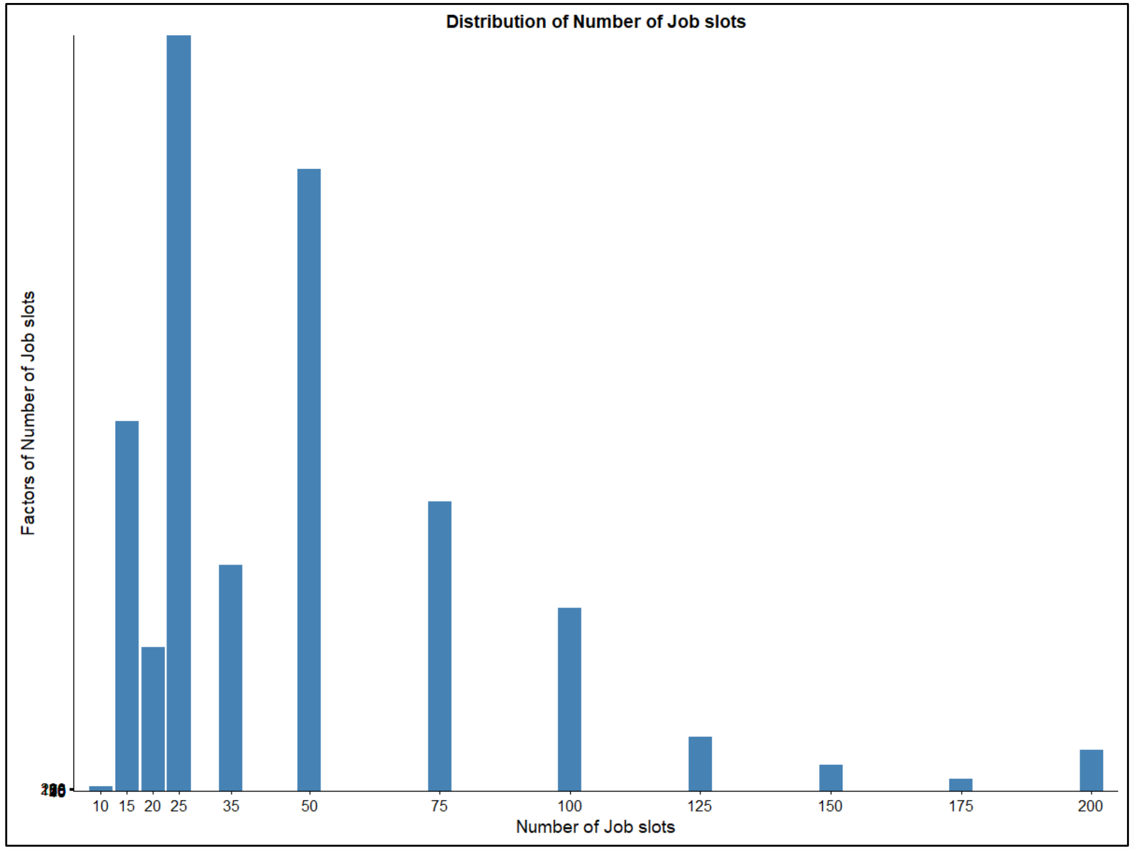


- Plot on the previous slide is a Histogram of Pricing per Job slot.
- There is one outlier in our case with pricing per slot = $18260
- Mean and median along with the histogram confirms that the job slots costing $700-750 per slot is the best selling price.
- Prices in the range of 600-800 per job slot is chosen the most by the companies.
- Histogram has more frequency around the price $600-800 per job slot.
- In this case, the mean is slightly greater than the median. Hence it slightly looks like a positively skewed distribution around its median.
- The right tail is longer.

```
> summary(pricingPerSlot)
    Min.  1st Qu.   Median     Mean  3rd Qu.      Max.
   4.108  590.700  716.000  716.300  805.500 18260.000
```

Most Popular packages based on number of slots

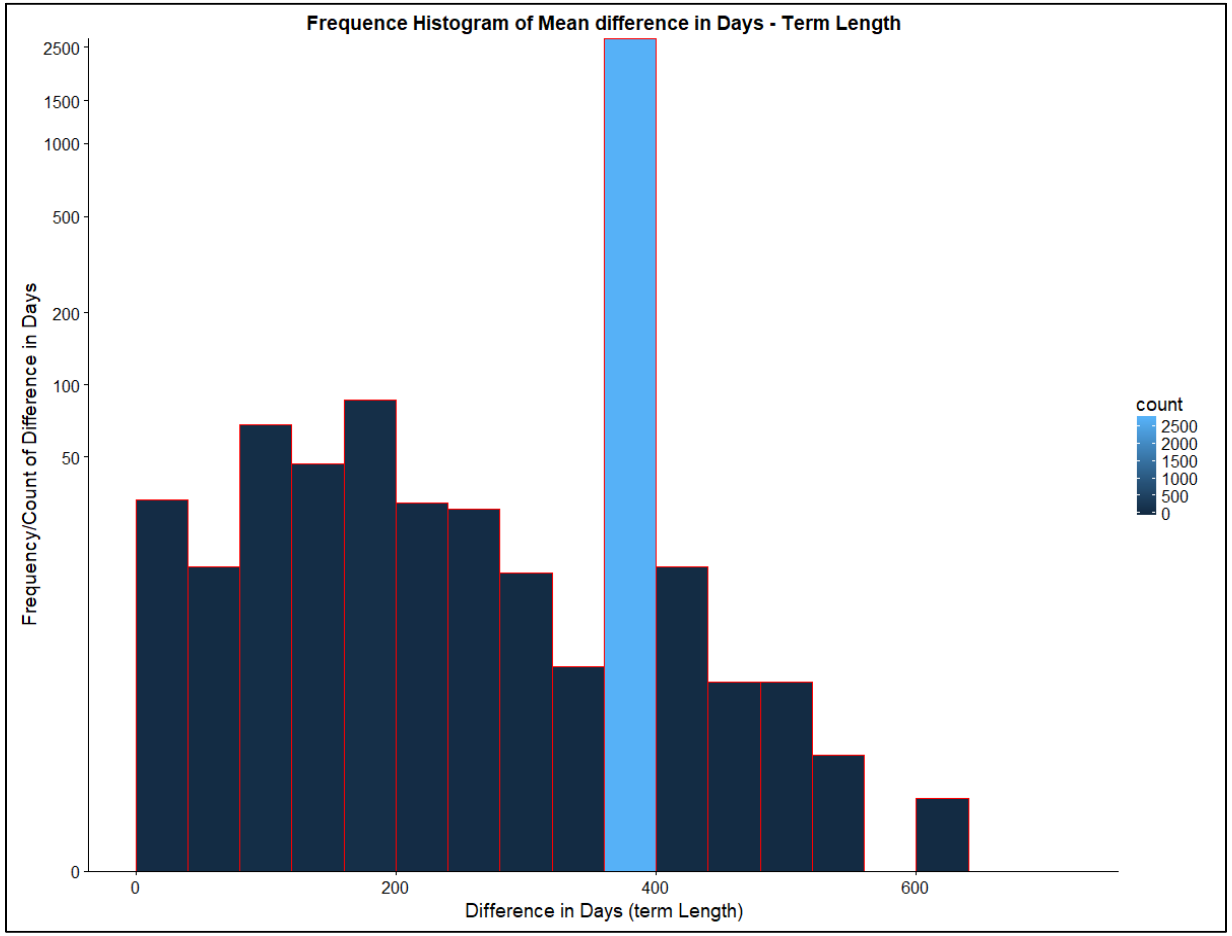


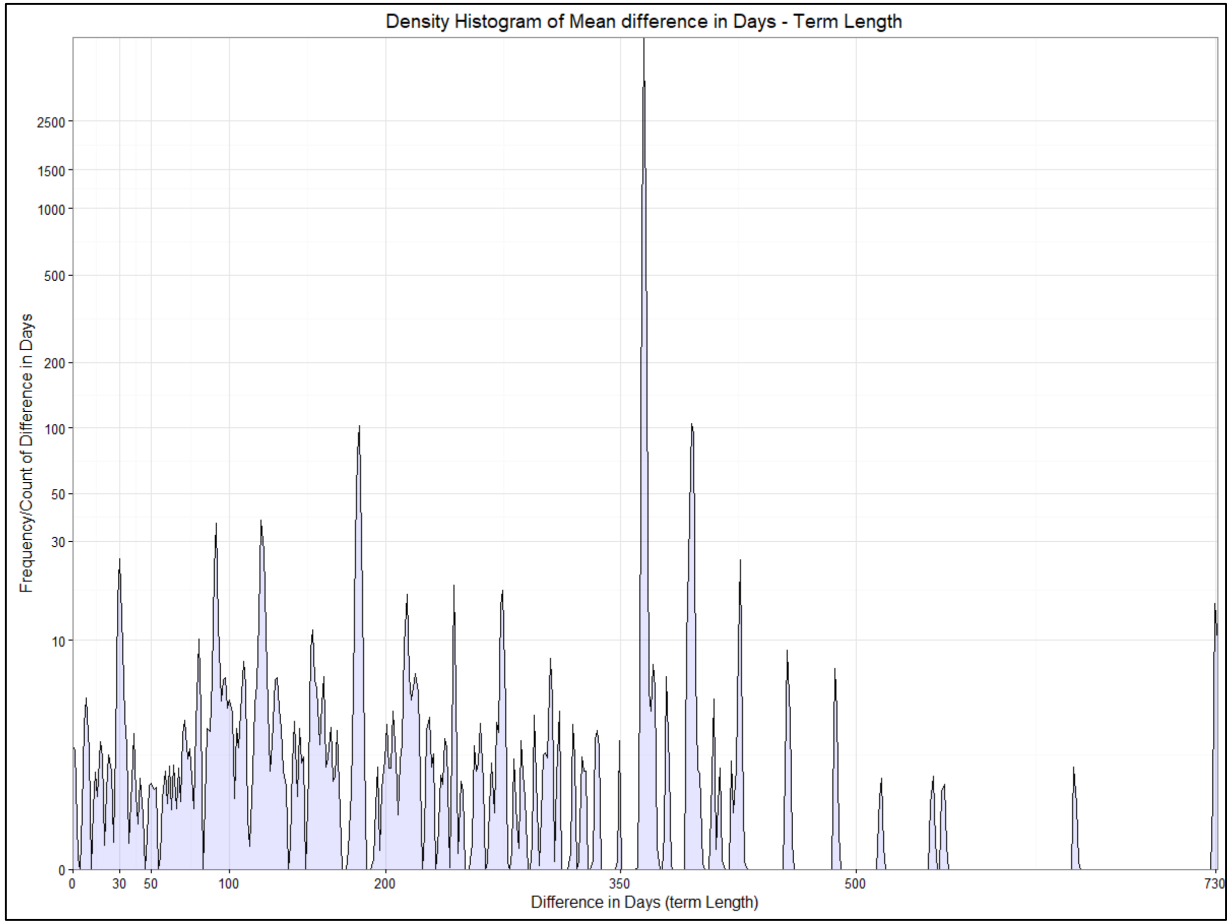


Most Popular packages based on number of slots

- Packages with 25 job slots are the most chosen packages by the companies.
- Also, from the two plots, we can see that maximum number of packages are sold when the number of slots offered is in the range 15 to 50. (specifically – 15, 25 and 50).
- There are very few companies (selected ones) which choose packages with more than 100 slots. These are the companies which are well-established and require more human resources. Hence, they have more openings (coming up).

Based on Term Length





Based on Term Length

- The most common term length irrespective of the location, number of slots and pricing, is around 1 year.
- This means that most companies buy packages with a plan for next 1 year timeline.
- Here, the mean is lesser than the median and hence it is slightly left-skewed.
- Also, when carefully reviewed, we can notice that when term length is close to 1 year, the number of slots purchased is slightly on the higher side.

```
> summary(data_glassdoor$diff_in_days)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    0.0   364.0   365.0   344.9   365.0   731.0 
```

Comparing Delivery Performance across Costumers

Based on –

Application metrics is considered as one of the important metrics. It is used to track the number of candidates/profiles applying for the particular job posted. When the number of applications increase, there are more chances of renewing the contract. This can further be improved by notifying more preferred/matching candidates whenever a job matching/suiting their profile is posted.

Cost – I am considering cost as cost per application – It will depend on how well the application matches the position you are looking for. When we focus on increasing the number of applications, we also increase the cost. Hence it is best when the right job application is submitted by the right candidate. Quantity is lesser important compared to the quality here.

Marketplace value can be higher for positions that are in demand in the market – like Data Science/Analytics compared something like sales representative. The bid Cost of marketing are more for positions that are trending. High profile companies (larger) can afford to bid more to get more relevant applications. Where as smaller companies cannot afford to spend more to get the bid.

PART 2 : Retention Analysis

Output of the Logistic Regression – Churn Analysis
There are a couple of models I built for the requirements. I am going to explain why the chosen model works the best comparing it with the next best model.

This model is the chosen model (model 1)
to predict retention/renewal.

```
Call:
glm(formula = `Renewed?` ~ `Number of Slots` + `Price Paid` +
    `Marketplace Value Delivered` + Applications + diff_in_days,
    family = "binomial", data = Data_Actual)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6759  -1.3550   0.7735   0.8710   2.0591

Coefficients:
                                  Estimate   Std. Error z value            Pr(>|z|)
(Intercept)                   -1.638984726  0.232984276  -7.035      0.000000000002 ***
`Number of Slots`              0.018362575  0.003680878   4.989      0.000000608058 ***
`Price Paid`                  -0.000042758  0.000006939  -6.162      0.000000000720 ***
`Marketplace Value Delivered`  0.000010870  0.000006974   1.559              0.1191
Applications                   0.000165946  0.000094257   1.761              0.0783 .
diff_in_days                   0.006511825  0.000669606   9.725 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3868.1  on 3077  degrees of freedom
Residual deviance: 3651.8  on 3072  degrees of freedom
AIC: 3663.8

Number of Fisher Scoring iterations: 4
```

PART 2 : Retention Analysis

Model 2 -

This is the second best model based on AIC and Chi-squared values. It is more or less the same as model 1.

```
Call:
glm(formula = `Renewed?` ~ `Number of Slots` + `Price Paid` +
    Applications + +diff_in_days, family = "binomial", data = Data_Actual)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5282  -1.3537   0.7767   0.8713   2.0460

Coefficients:
                      Estimate  Std. Error z value              Pr(>|z|)
(Intercept)        -1.63464113  0.23241197  -7.033    0.00000000000202 ***
`Number of Slots`   0.01826225  0.00366609   4.981    0.00000063128580 ***
`Price Paid`       -0.00004187  0.00000690  -6.068    0.00000000129231 ***
Applications        0.00028552  0.00005629   5.073    0.00000039243408 ***
diff_in_days        0.00651923  0.00066794   9.760 < 0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3868.1  on 3077  degrees of freedom
Residual deviance: 3654.3  on 3073  degrees of freedom
AIC: 3664.3

Number of Fisher Scoring iterations: 4
```

Analyzing the Logistic Regression Model -

- When we analyze the results of the model, we can find that there are various positive and negative reasons why a company would/wouldn’t renew their packages of Glassdoor.
- When the number of Applications offered is more, then there is better chances of the company renewing their package with Glassdoor.
- When the price of the packages offered is high, then there are lower chances of the company renewing the package. Pricing and Renewing are slightly inversely proportional with other factors considered.
- Applications : There are better chances that a company would renew their package with Glassdoor if they receive higher number of application through Glassdoor. Hence, reaching out to more relevant candidates will increase the chances of renewal/retention.
- Term Length – Term length is directly proportional to the chances of renewal. When a company is offered longer period of time for the chosen number of slots, then there are good chances of them coming back to Glassdoor for renewal.
- Now, lets look at this in a mathematical approach –

Analyzing the Logistic Regression Model -

- For every one unit change in the number of Slots, the log odds of Renewal increases by 0.0184
- For example, if Glassdoor offers 10 jobs in the first contract, then chances of renewing will increase by 0.184 solely based on Number of slots offered. Similarly, in the same scenario if 25 jobs slots were offered, then odds of Renewal increases by 0.46
- For a unit increase in the Price Paid, the log odds of Renewal decreases by 0.00004.
- For example, if the price paid is 10,000 then odds of Renewal decreases by 0.4.
- Similarly, if the price is decreases slightly, there is more chances of companies renewing.
- When the number of applications increase by 1 unit, then the odds of renewing increases by 0.000166.
- For example, when the number of applicants to a particular job is 2000, then the odds of renewing will be 0.332.
- I am very sure, if the number of relevant applications increase, then there is more chances of renewal than what the numbers suggest.
- With the term length increasing by 1 unit, then the odds of companies renewing will increase by 0.0065

Analyzing the Logistic Regression Model -

- Below the table of coefficients are fit indices, including the null and deviance residuals and the AIC. Later we show an example of how you can use these values to help assess model fit.
- We can use the confint function to obtain confidence intervals for the coefficient estimates. Note that for logistic models, confidence intervals are based on the profiled log-likelihood function. We can also get CIs based on just the standard errors by using the default method.
- The chi-squared test statistic of 155.3, with three degrees of freedom is associated with a p-value of 0.000011 indicating that the overall effect of rank is statistically significant.
- Right now the model is close to 72% accurate predicting he Renewal of contract. Can be improved by looking at more data in detail. Also, with these details one can work on increasing renewal rate focusing on companies who are not renewing due to the metrics mentioned.

```
> confint(logit.reg)
Waiting for profiling to be done...
                                       2.5 %         97.5 %
(Intercept)                    -2.104750245682 -1.19040668734
`Number of Slots`               0.011227130170  0.02566560687
`Price Paid`                   -0.000056421251 -0.00002918374
`Marketplace Value Delivered`  -0.000002800177  0.00002455983
Applications                   -0.000016986192  0.00035281975
diff_in_days                    0.005220472119  0.00784799239
```

```
Wald test:
----------

Chi-squared test:
X2 = 155.3, df = 3, P(> X2) = 0.0
```

Analyzing the Logistic Regression Model -

FINALLY -

- Based on the analysis and modeling, we can focus on what is more important with respect to the customers keeping in mind profit and popularity of Glassdoor.

- Models can be built and improved based on other factors. In our logistic regression, we can confirm that most of the customers retained are because of the number of slots, applications received, term length and sometimes the prices as well.

How well does your analysis in #3 predict retention? What other factors might you want to investigate to see if they could improve your analysis?

- I would like to investigate data relating to –

a. Company size – When the company size is higher, there are better chances that they will posting more jobs in the coming days. Providing these companies with packages that has higher number of job slots and Term length can help in retention.

b. Similarly, smaller companies will choose packages with lesser term length and slots. These companies can be charged a little higher and they'll retain if they find the right candidates through Glassdoor.

c. Relevant application – Companies will continue with Glassdoor when they see higher percentage of right candidates applying to their jobs through Glassdoor. For example, when there is an opening for Decision Science position and if they find more candidates from irrelevant background, then there is lower chances of the companies continuing with the package.

d. Records/track of conversion rate – It would be interesting to analyze data where candidates applying through Glassdoor to a particular position get selected for further rounds of interview. This assures quality candidates to a company through Glassdoor.

e. Glassdoor reviews – One of the most important reason candidates look up to Glassdoor is to find the true review about companies. Chances of companies with higher/better rating on Glassdoor retaining is more.

Based on your analysis, what modifications would you recommend we make to our ad platform algorithm to improve retention?

1. There should be more notifications to Glassdoor users when there is a job posted matching their profile. This prompting will help them apply to positions as and when they are notified.
2. Display percentage of Skill set match (embedding ATS) and recommend that job for qualified profiles only. It improves the quality of applicants for the employers.
3. Profiles of Candidates matching the job description/requirements should be made available/visible on Glassdoor to the recruiters. This increases the chances of recruiters finding the right set of candidates for further process.

4. We can work on modifying platform algorithm based on the analysis of other data mentioned in the previous slide.