Hiretual Case Competition

Hungry Terps:

Captain: Nai-jieh Wang
Group Member: Jingsi Xu
Ziyi Cheng
Mengyuan Tian
Guiran Niu
Xiaojie Xiao

CONTENTS

- Execution Summary
- Valuable Insights
- Gains and thoughts
- Q&A Time

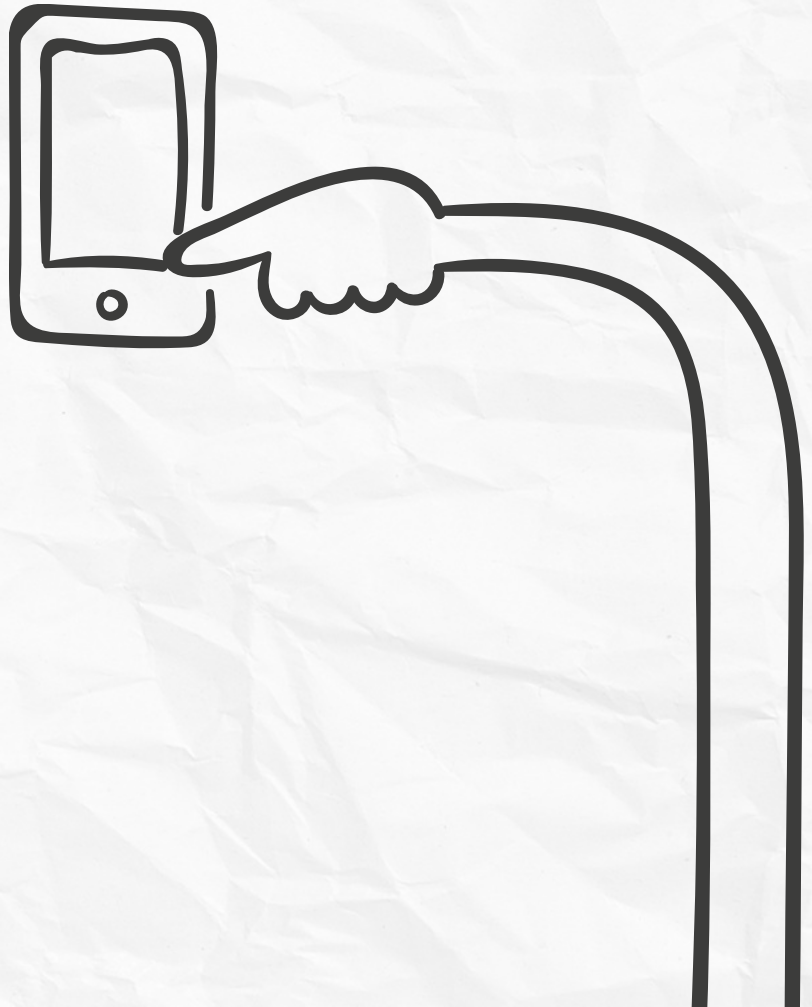
PART 01

Execution Summary

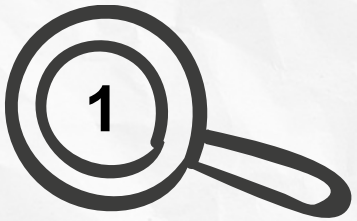
Project Goal
Data Processing

Project Goal

- Extraction
- Finding patterns
- Making analysis &predictions



Data Processing



Data Cleaning

1.Missing data
2. Data shifting
3.Missing values

Data Modeling

1. Logistic regression
2. Linear regression
3. Regression Tree
4. Boosting

PART 02

Valuable Insights

Question 1

Extract the most important features and factors that makes a person or a group of people leave a current job for new job.

Logistic Regression
Regression Tress
Boosting

Logistic Regression

```
## Call:
## glm(formula = position_is_current_new ~ ., family = "binomial",
##     data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8980  -1.1650   0.8539   1.0641   1.8241
##
## Coefficients:
##                                          Estimate Std. Error z value
## (Intercept)                            -6.182e+01  4.682e+00 -13.204
## position_company_size                   2.291e-02  3.426e-02   0.669
## Job_Day                                 1.894e-05  4.694e-06   4.036
## CoastMid                                1.948e-02  4.354e-02   0.447
## CoastWest                              -6.349e-02  4.275e-02  -1.485
## Chinese                                 1.272e-01  1.108e-01   1.148
## French                                  5.494e-03  1.010e-01   0.054
## Hindi                                   4.437e-02  9.962e-02   0.445
## Spanish                                 2.037e-02  5.779e-02   0.352
## English                                -9.717e-03  4.289e-02  -0.227
## experience                             -4.007e-02  5.210e-03  -7.690
## titledirector                          -2.207e-01  1.635e-01  -1.349
## titleexecutive                         -3.522e-01  1.614e-01  -2.182
## titlejunior                            -2.262e-01  1.616e-01  -1.400
## titlemanager                           -1.797e-01  1.579e-01  -1.138
## titlesenior                            -1.044e-01  1.542e-01  -0.677
## education_highest_DegreeBachelor        3.521e-01  9.490e-02   3.710
## education_highest_DegreeDiploma         3.817e-01  2.242e-01   1.703
## education_highest_DegreeHigh School    -1.102e+00  8.143e-01  -1.353
## education_highest_DegreeMaster          2.486e-01  9.776e-02   2.542
## education_highest_DegreeMBA             2.527e-01  1.237e-01   2.043
## education_highest_DegreePHD             2.171e-01  1.158e-01   1.875
## education_highest_DegreeUndergraduate  -7.647e-01  1.230e+00  -0.622
## education_endyear                       1.029e-02  1.076e-02   0.956
## education_startyear                     2.078e-02  1.050e-02   1.980
## careerFinancial Services               -2.499e-02  1.000e-01  -0.250
## careerHospital & Health Care           -1.628e-01  1.027e-01  -1.585
## careerInformation Technology and Services -2.268e-02  6.499e-02  -0.349
## careerMarketing and Advertising         7.296e-02  9.706e-02   0.752
## careerOthers                           -2.880e-02  5.323e-02  -0.541
## careerTelecommunications               -2.319e-01  1.110e-01  -2.090
```

```
##                                          Pr(>|z|)
## (Intercept)                               < 2e-16 ***
## position_company_size                    0.503585
## Job_Day                                  5.44e-05 ***
## CoastMid                                 0.654684
## CoastWest                                0.137522
## Chinese                                  0.250878
## French                                   0.956610
## Hindi                                    0.655985
## Spanish                                  0.724469
## English                                  0.820769
## experience                               1.47e-14 ***
## titledirector                            0.177179
## titleexecutive                           0.029074 *
## titlejunior                              0.161450
## titlemanager                             0.255139
## titlesenior                              0.498377
## education_highest_DegreeBachelor         0.000207 ***
## education_highest_DegreeDiploma          0.088603 .
## education_highest_DegreeHigh School      0.176098
## education_highest_DegreeMaster           0.011008 *
## education_highest_DegreeMBA              0.041091 *
## education_highest_DegreePHD              0.060768 .
## education_highest_DegreeUndergraduate    0.534161
## education_endyear                        0.338952
## education_startyear                      0.047757 *
## careerFinancial Services                 0.802773
## careerHospital & Health Care             0.113030
## careerInformation Technology and Services 0.727121
## careerMarketing and Advertising          0.452239
## careerOthers                             0.588458
## careerTelecommunications                 0.036623 *
```

Most Influence:

Job_Day

experience

titleexecutive

Accuracy: 59.8%

Regression Tree


```
## Regression tree:
## tree(formula = position_is_current_new ~ ., data = train)
## Variables actually used in tree construction:
## [1] "Job_Day"              "education_startyear"
## Number of terminal nodes:  3
## Residual mean deviance:  0.2113 = 3157 / 14940
## Distribution of residuals:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.7050 -0.4900  0.2950  0.0000  0.2950  0.8334

## [1] 0.6542304
```

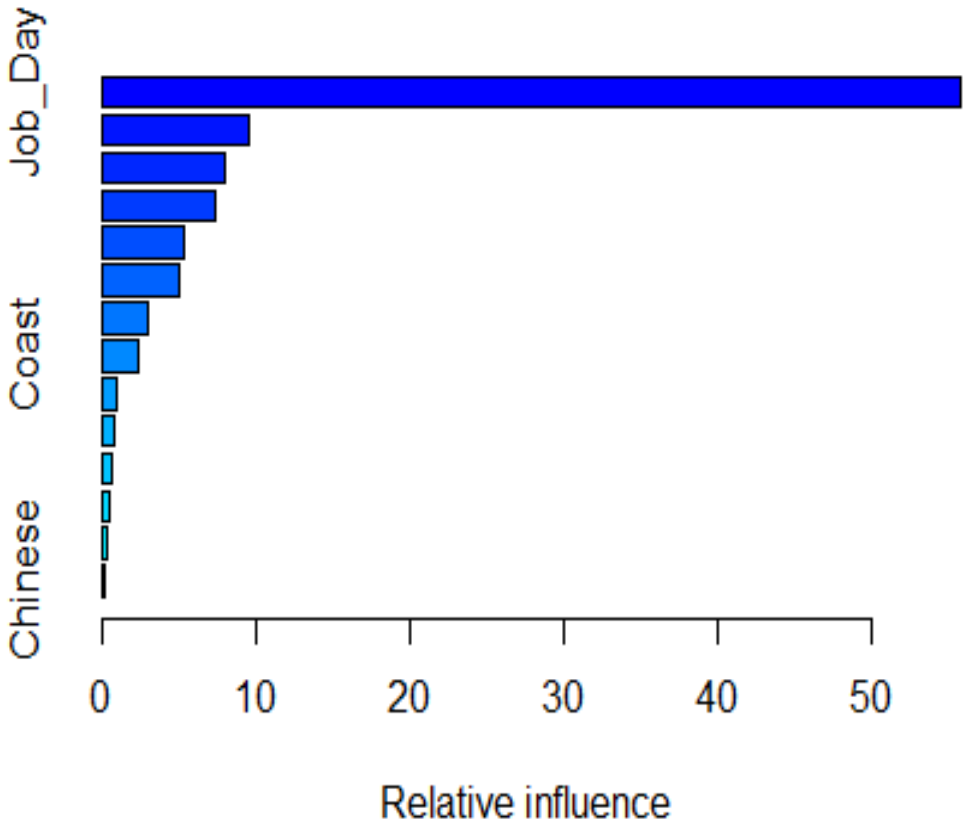
Most Influence:

Job_Day

education_startyear

Accuracy: 65.4%

Boosting




```
##                                                 var    rel.inf
## Job_Day                                     Job_Day 55.8579596
## education_endyear                 education_endyear  9.6192831
## education_startyear             education_startyear  8.0143420
## career                                       career  7.3950494
## title                                         title  5.3306501
## education_highest_Degree education_highest_Degree  5.0634195
## experience                               experience  3.0124527
## Coast                                         Coast  2.3426145
## position_company_size       position_company_size  1.0015869
## English                                     English  0.7610743
## Spanish                                     Spanish  0.5946823
## French                                       French  0.4573639
## Hindi                                         Hindi  0.3392437
## Chinese                                     Chinese  0.2102779

##                predicted
## size.test.boost    0    1
##               0 2103  911
##               1  527 2865

## [1] 0.7755229
```

Most Influence:

Job_Day

education_endyear

education_startyear

Accuracy: 77.5%

Question 2

Predict the possibility that a person leaves current job to get a new job today, in six months, in a year.

Logistic Regression

Modeling Process

1. Goal:

To predict the possibility that a person leaves current job to get a new job:

- TODAT
- IN 6 MONTHS
- IN 1 YEAR

2. Assumption:

To simplify, the situation that a person is having multiple jobs in a period is not considered here.

3. Dependent Variable:

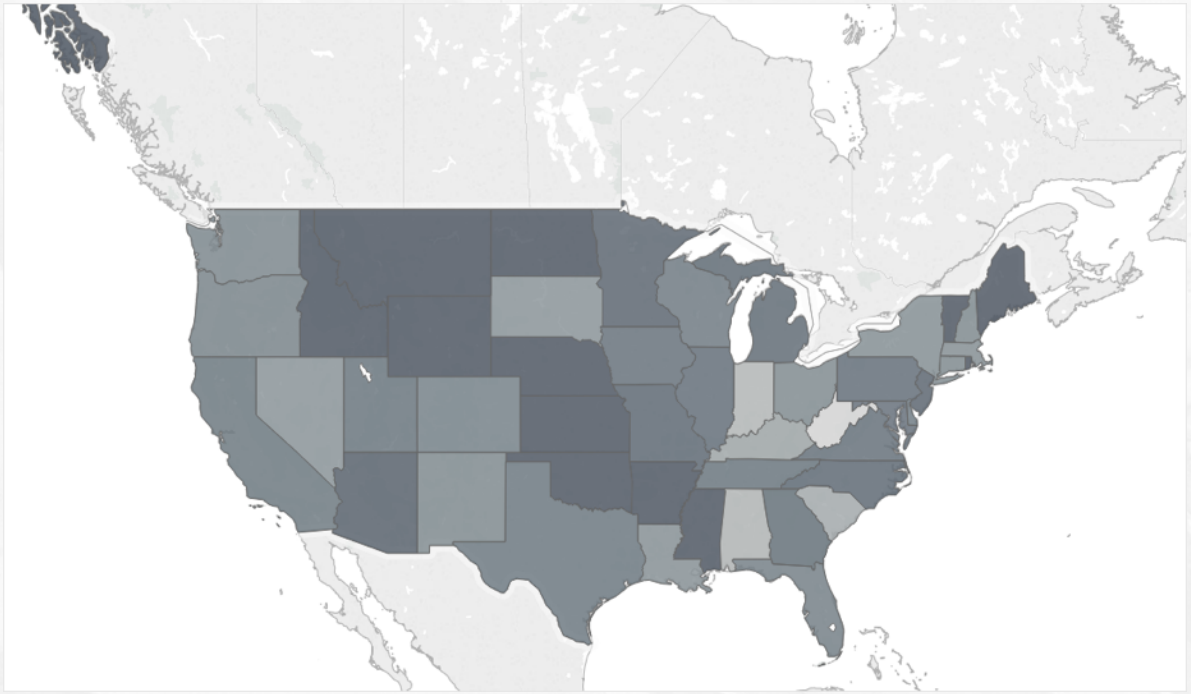
Calculate the date difference between a position's end date and this person's next new job start date.

4. Logistic Regression

Users can perform presentations on projectors or computers and print presentations

Findings

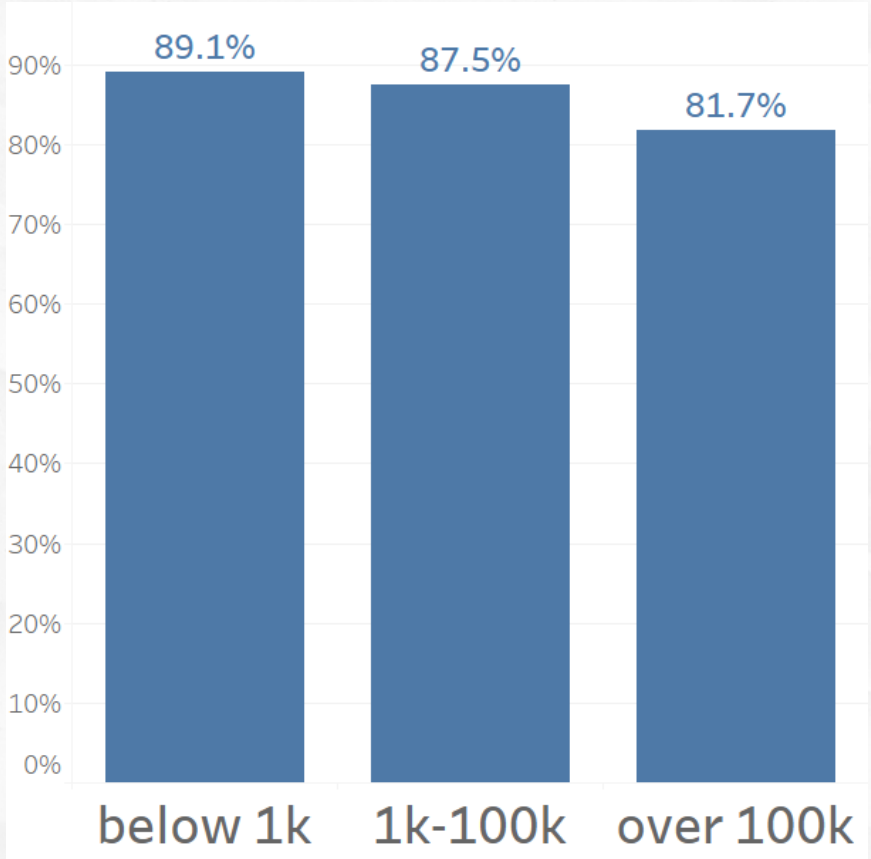
Person leaves current job to get a new job IN A YEAR



Mid Coast

Findings

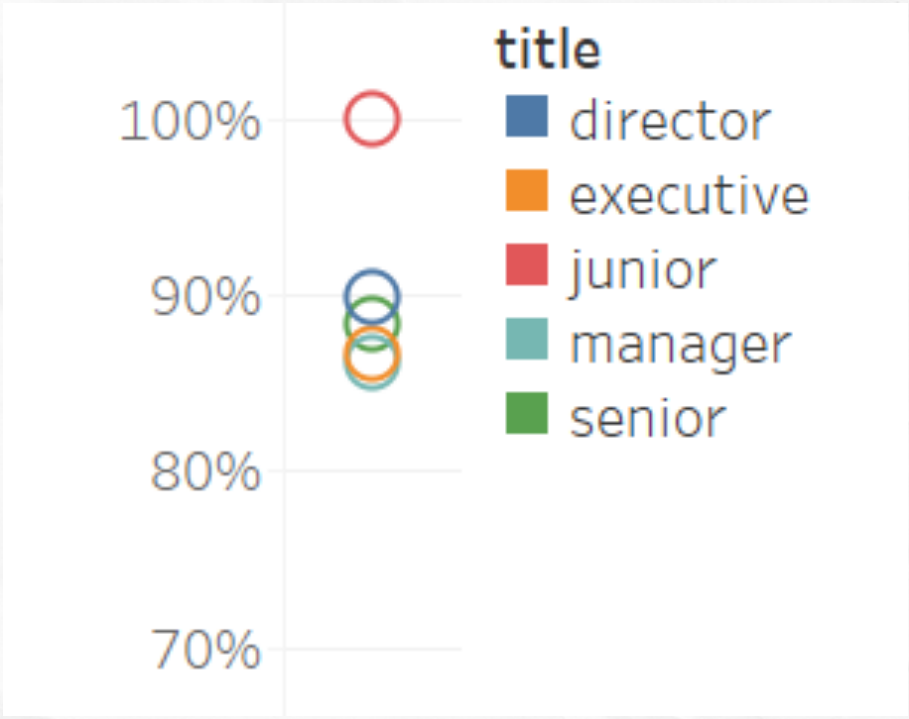
Person leaves current job to get a new job IN A YEAR



Company Size

Findings

Person leaves current job to get a new job IN A YEAR



Junior Title

Question 3

Predict the possibility that a person leaves current job before and after anniversary.

Logistic Regression

Recoding & Modeling

- Working Duration = Position_EndDate – Position _StartDate

- Job_Day(Duration)%% 365 days > 180days, “before_anniversary”=1.
- Otherwise, ”before_anniversary = 0”.

- Before_annivary = position_company_size + Chinese + Hindi + English + Spanish + experience + Coast + tit + Job_Day + career + degree

Logistic Regression

Coefficients:

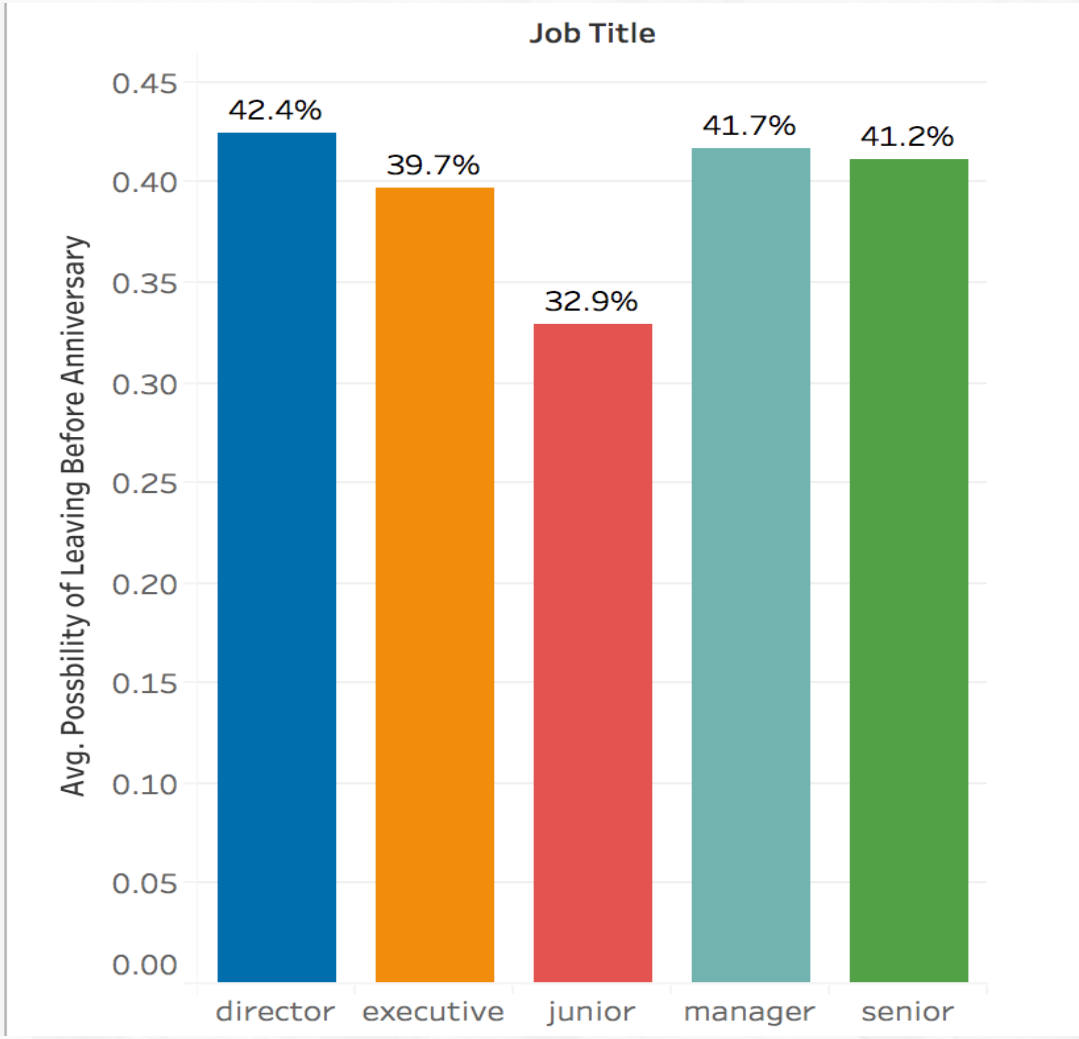
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-4.096e-01	1.213e-01	-3.376	0.000737	***
position_company_size	-2.495e-08	1.034e-07	-0.241	0.809360	
Chinese	-1.153e-01	1.010e-01	-1.142	0.253517	
Hindi	1.112e-01	8.699e-02	1.279	0.201030	
English	-3.732e-03	3.673e-02	-0.102	0.919074	
Spanish	1.652e-02	5.025e-02	0.329	0.742349	
experience	3.600e-03	3.683e-03	0.977	0.328368	
CoastMid	3.720e-02	3.729e-02	0.998	0.318501	
CoastWest	-4.346e-02	3.465e-02	-1.254	0.209821	
titexecutive	-1.544e-01	5.434e-02	-2.841	0.004497	**
titjunior	-2.838e-01	8.607e-02	-3.297	0.000977	***
titmanager	-3.311e-02	5.300e-02	-0.625	0.532138	
titsenior	-8.011e-02	4.669e-02	-1.716	0.086154	.
Job_Day	4.437e-05	3.531e-06	12.566	< 2e-16	***
careerFinancial Services	-4.350e-02	8.874e-02	-0.490	0.624024	
careerHospital & Health Care	1.484e-02	8.604e-02	0.173	0.863030	
careerInformation Technology and Services	3.642e-02	5.315e-02	0.685	0.493209	
careerMarketing and Advertising	-1.486e-01	8.248e-02	-1.802	0.071560	.
careerOthers	-2.296e-02	4.457e-02	-0.515	0.606403	
careerTelecommunications	3.502e-02	9.456e-02	0.370	0.711106	
degreeBachelor	-7.090e-02	9.235e-02	-0.768	0.442683	
degreeMaster	-4.028e-02	9.630e-02	-0.418	0.675750	
degreeMBA	-1.044e-01	1.165e-01	-0.897	0.369934	
degreePHD	-1.747e-01	1.134e-01	-1.541	0.123393	

- **Job Title: Executive**
  Junior
- **Job_day: working duration**
- **Accuracy: 60%**

Results

1. More working days → More like to leave before anniversary

2. Director > Executive > Junior
   ≈ Manager
   ≈ Senior



Predictions

Person	Job Day	Coast	Language	Career	Experience	Title	Degree	Prediction
1	3179	East	English	Information Technology and Services	15	manager	Master	43%
2	40442	East	Others	Information Technology and Services	15	senior	Bachelor	79%
3	40353	West	Others	Computer Software	15	director	Bachelor	79%
4	1990	East	Others	Others	5	manager	Bachelor	40%
5	1961	East	Others	Information Technology and Services	5	junior	Bachelor	35%

Question 4

Predict the shortest time and average time until that the person will no longer be at their current job.

Linear Regression

Predict how long a talent will leave current job

Linear Regression

Model Selection: backward

```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                               2657.431    288.432   9.213  < 2e-16 ***
CoastMid                                  -280.228     73.908  -3.792 0.000150 ***
CoastWest                                 -270.995     67.590  -4.009 6.11e-05 ***
Chinese                                    315.787    184.711   1.710 0.087347 .
Spanish                                   -308.605     97.825  -3.155 0.001609 **
English                                   -453.092     68.911  -6.575 4.97e-11 ***
experience                                 145.594      7.338  19.840  < 2e-16 ***
titledirector                             -807.455    256.391  -3.149 0.001639 **
titleexecutive                            -454.195    252.936  -1.796 0.072557 .
titlejunior                               -869.712    255.261  -3.407 0.000658 ***
titlemanager                              -950.345    247.668  -3.837 0.000125 ***
titlesenior                               -895.578    240.828  -3.719 0.000201 ***
education_highest_DegreeBachelor           141.815    163.226   0.869 0.384952
education_highest_DegreeDiploma            611.147    350.562   1.743 0.081288 .
education_highest_DegreeHigh School         22.297    578.901   0.039 0.969276
education_highest_DegreeMaster             336.273    167.822   2.004 0.045108 *
education_highest_DegreeMBA                498.054    204.999   2.430 0.015125 *
education_highest_DegreePHD                706.434    197.096   3.584 0.000339 ***
education_highest_DegreeUndergraduate     -669.387   2180.892  -0.307 0.758897
careerFinancial Services                  -573.969    170.508  -3.366 0.000763 ***
careerHospital & Health Care              -101.942    165.082  -0.618 0.536897
careerInformation Technology and Services -272.567    104.904  -2.598 0.009376 **
careerMarketing and Advertising           -277.968    158.548  -1.753 0.079579 .
careerOthers                              -108.709     85.587  -1.270 0.204039
careerTelecommunications                   253.134    184.942   1.369 0.171099
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4345 on 23975 degrees of freedom
Multiple R-squared:  0.03589,   Adjusted R-squared:  0.03492
F-statistic: 37.19 on 24 and 23975 DF,  p-value: < 2.2e-16
```

Model:

formula = Job_Day ~ Coast + Chinese + Spanish + English + experience + title + education_highest_Degree + career, data = train, na.action = na.omit)

Findings:

Who stay more time at current job?

Experienced;
Hold PhD degree

Who stay less time at current job?

Live at west coast and in the middle areas;
Speak English;
Work in financial services field

Question 5

A person is going to smaller or bigger company

Logistic Regression

Boosting

Logistic Regression

```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              0.3801227  0.0403075   9.431  < 2e-16 ***
CoastMid                                 0.0097735  0.0103691   0.943  0.34592
CoastWest                                0.0197369  0.0102376   1.928  0.05389 .
Hindi                                    0.1514299  0.0219842   6.888 5.88e-12 ***
Spanish                                 -0.0329184  0.0129291  -2.546  0.01090 *
titledirector                           -0.0006941  0.0346574  -0.020  0.98402
titleexecutive                          -0.0536238  0.0340021  -1.577  0.11480
titlejunior                              0.0047320  0.0359043   0.132  0.89515
titlemanager                             0.0564739  0.0336673   1.677  0.09348 .
titlesenior                              0.0110590  0.0327659   0.338  0.73573
education_highest_DegreeBachelor         0.0928840  0.0222945   4.166 3.11e-05 ***
education_highest_DegreeDiploma         -0.0461131  0.0532114  -0.867  0.38617
education_highest_DegreeHigh School     -0.0659343  0.1661166  -0.397  0.69144
education_highest_DegreeMaster           0.1641057  0.0230206   7.129 1.06e-12 ***
education_highest_DegreeMBA              0.1994735  0.0294271   6.779 1.26e-11 ***
education_highest_DegreePHD              0.1773463  0.0271615   6.529 6.82e-11 ***
education_highest_DegreeUndergraduate   -0.3260756  0.2859284  -1.140  0.25413
careerFinancial Services                 0.0641556  0.0235170   2.728  0.00638 **
careerHospital & Health Care            -0.0222745  0.0240012  -0.928  0.35339
careerInformation Technology and Services  0.0024654  0.0153538   0.161  0.87243
careerMarketing and Advertising         -0.1234551  0.0233563  -5.286 1.27e-07 ***
careerOthers                            -0.0631265  0.0125995  -5.010 5.50e-07 ***
careerTelecommunications                 0.1153982  0.0260997   4.421 9.87e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.2431879)

    Null deviance: 3730.3  on 14944  degrees of freedom
Residual deviance: 3628.8  on 14922  degrees of freedom
AIC: 21306
```

Independent Variables:

Coast
Hindi
Spanish
Title
Education Highest Degree
Career

Accuracy:
Full Model: 57%
Selected Model: 56.7%

Logistic Regression for Q5

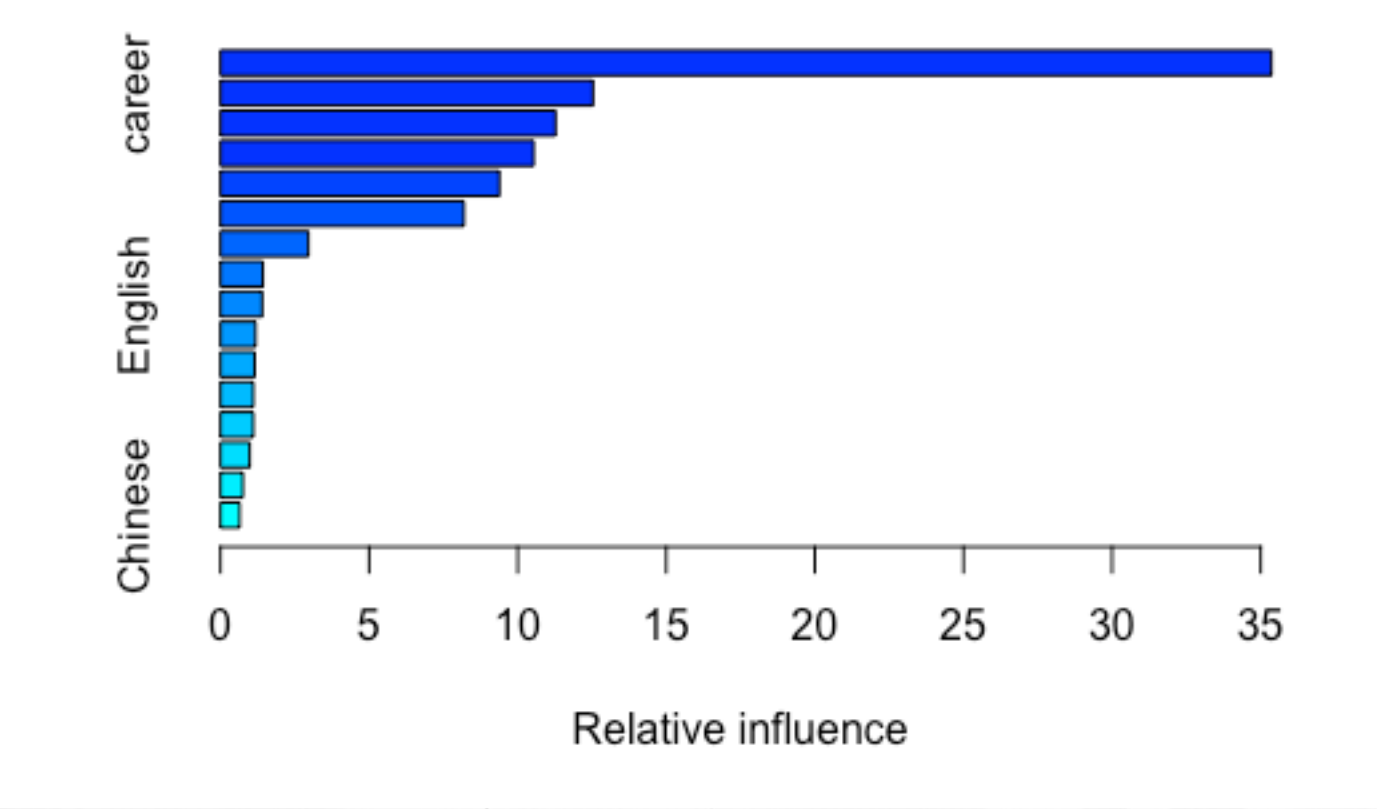
```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8587  -0.4717  -0.3275   0.5069   0.7564

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                             -6.743e-01  1.116e+00  -0.604 0.545718
position_iscurrent                       4.506e-03  1.015e-02   0.444 0.656935
Job_Day                                 -1.486e-06  1.116e-06  -1.331 0.183115
CoastMid                                 5.604e-04  1.036e-02   0.054 0.956867
CoastWest                                1.886e-02  1.022e-02   1.845 0.065005 .
Chinese                                  2.932e-02  2.605e-02   1.126 0.260271
French                                   2.491e-02  2.460e-02   1.013 0.311250
Hindi                                    1.410e-01  2.380e-02   5.923 3.23e-09 ***
Spanish                                 -2.519e-02  1.375e-02  -1.832 0.066945 .
English                                  1.384e-02  1.029e-02   1.346 0.178337
experience                              -8.318e-04  1.238e-03  -0.672 0.501483
titledirector                           -7.883e-03  3.692e-02  -0.214 0.830911
titleexecutive                          -3.971e-02  3.626e-02  -1.095 0.273528
titlejunior                              5.202e-04  3.621e-02   0.014 0.988539
titlemanager                             5.860e-02  3.544e-02   1.653 0.098264 .
titlesenior                              8.196e-03  3.446e-02   0.238 0.812032
education_highest_DegreeBachelor         8.454e-02  2.244e-02   3.768 0.000165 ***
education_highest_DegreeDiploma         -3.811e-02  5.431e-02  -0.702 0.482922
education_highest_DegreeHigh School      1.490e-01  2.218e-01   0.672 0.501750
education_highest_DegreeMaster           1.642e-01  2.313e-02   7.098 1.32e-12 ***
education_highest_DegreeMBA              2.109e-01  2.960e-02   7.124 1.09e-12 ***
education_highest_DegreePHD              1.539e-01  2.730e-02   5.637 1.76e-08 ***
education_highest_DegreeUndergraduate   -3.114e-01  2.858e-01  -1.090 0.275826
education_endyear                        4.195e-03  2.743e-03   1.529 0.126224
education_startyear                     -3.668e-03  2.677e-03  -1.370 0.170678
careerFinancial Services                 8.840e-02  2.349e-02   3.762 0.000169 ***
careerHospital & Health Care            -5.805e-03  2.431e-02  -0.239 0.811265
careerInformation Technology and Services  1.250e-02  1.547e-02   0.808 0.419079
careerMarketing and Advertising         -1.361e-01  2.279e-02  -5.972 2.40e-09 ***
careerOthers                            -6.307e-02  1.265e-02  -4.985 6.25e-07 ***
careerTelecommunications                 1.238e-01  2.697e-02   4.590 4.47e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.2427449)

    Null deviance: 3731.6  on 14944  degrees of freedom
Residual deviance: 3620.3  on 14914  degrees of freedom
AIC: 21287
```

Boosting



	var	rel.inf
Job_Day	Job_Day	35.3738922
career	career	11.9808236
education_startyear	education_startyear	11.9053516
education_endyear	education_endyear	11.1344785
title	title	8.9089788
education_highest_Degree	education_highest_Degree	7.8107674
Coast	Coast	3.4525290
experience	experience	3.2475617
English	English	1.4489430
position_iscurrent	position_iscurrent	1.2872603
Spanish	Spanish	1.1734748
Hindi	Hindi	0.8939238
French	French	0.8063561
Chinese	Chinese	0.5756591

Accuracy:
59.2%

Example

A person who is:

- Working days for current job is 3179
- Speak English
- At East
- Information Technology & Services
- Master
- Education Start Year:1996
- Education End Year:1998

53.56% works for larger company

Bonus 1

Predict the possibility that a person willing to relocate to a different location.

Logistic Regression

Results

Predict the possibility that a person is willing to relocate

```
Call:
glm(formula = relocate ~ position_company_size + Job_Day + Coast +
    Chinese + Hindi + Spanish + French + Hindi + Spanish + experience +
    career + tit + degree, family = binomial(link = "logit"),
    data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5853  -0.8893  -0.7678   1.3672   2.0403

Coefficients:
                                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                             -8.843e-01  1.389e-01  -6.365 1.95e-10 ***
position_company_size                    6.702e-07  1.150e-07   5.828 5.62e-09 ***
Job_Day                                 -4.288e-06  8.251e-06  -0.520 0.603299
CoastMid                                -2.153e-01  3.978e-02  -5.412 6.23e-08 ***
CoastWest                               -6.852e-01  3.960e-02 -17.305  < 2e-16 ***
Chinese                                 -8.053e-02  1.038e-01  -0.776 0.437955
Hindi                                    6.725e-01  8.571e-02   7.847 4.27e-15 ***
Spanish                                 -9.488e-02  5.132e-02  -1.849 0.064467 .
French                                   1.439e-01  9.180e-02   1.568 0.116952
experience                              -2.525e-02  3.976e-03  -6.352 2.12e-10 ***
careerFinancial Services                -5.134e-02  9.233e-02  -0.556 0.578166
careerHospital & Health Care             5.571e-02  9.127e-02   0.610 0.541609
careerInformation Technology and Services  2.141e-02  5.902e-02   0.363 0.716786
careerMarketing and Advertising         -2.042e-01  9.113e-02  -2.241 0.025041 *
careerOthers                             1.002e-02  4.978e-02   0.201 0.840417
careerTelecommunications                -6.717e-02  1.039e-01  -0.646 0.518110
titexecutive                             1.463e-01  5.934e-02   2.465 0.013688 *
titjunior                                2.221e-01  9.041e-02   2.457 0.014023 *
titmanager                               2.584e-02  5.915e-02   0.437 0.662246
titsenior                                7.371e-02  5.193e-02   1.419 0.155763
degreeBachelor                           4.283e-01  1.103e-01   3.883 0.000103 ***
degreeMaster                             6.253e-01  1.138e-01   5.494 3.93e-08 ***
degreeMBA                                6.211e-01  1.329e-01   4.673 2.97e-06 ***
degreePHD                                1.006e+00  1.288e-01   7.815 5.50e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

AIC: 24406

Findings

- larger size company
- less work experience
- Marketing and Advertising career path
- Executive or Junior

Predictions

Person	Job Day	Coast	Company size	Career	Experience	Title	Degree	Possibility to relocate
1	3179	East	597758	Information Technology and Services	15	manager	Master	**45%**
2	40442	East	0	Information Technology and Services	15	senior	Bachelor	**29%**
3	40353	West	4235	Computer Software	15	director	Bachelor	**15%**
4	1990	East	1040	Others	5	manager	Bachelor	**36%**
5	1961	East	85	Information Technology and Services	5	junior	Bachelor	**41%**

PART 03

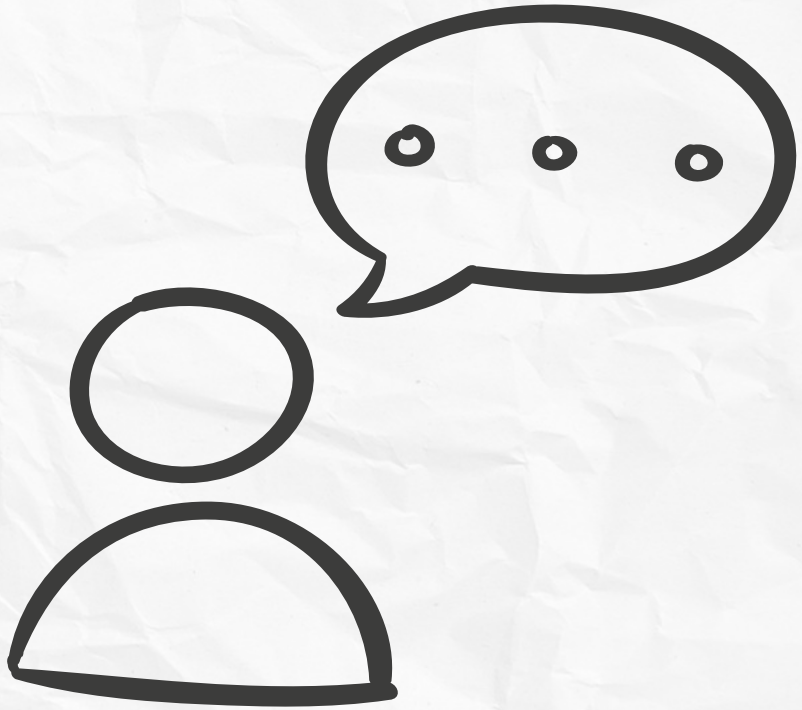
Gains and thoughts

Gains and Thoughts

PART 04

Q&A Time

Q&A



THANK YOU

Thank your to watch