



The HR department from ABC IT company that Recruit and Retain jobs ensures that its employees don't quit their jobs.



The HR team has limited funds, they cannot compensate every employee and is concerned about who they should incentivize to increase retention.



**Employee Retention Program –
Recruit and Retain Talent Team**

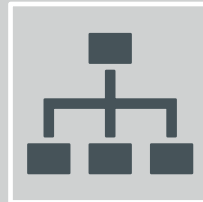
I want to Incentivize Employees to Stay in the company. Who should I be incentivizing?

HOW TO TACKLE THE BUSINESS PROBLEM?

WHAT RESOURCES DO YOU HAVE?



The strategy is to find out the probabilities of employees leaving the company, staying with the company and Uncertain (whether to leave or stay with the company) and help retention team with insights on whom they should incentivize.



A data set with 4653 observations containing employee details and their decisions to leave or stay with the company is available to make predictions.

SAS and R studio for making Predictions.

Tableau for Data Visualization

Variables Summary:

Independent Variables

- Education
- Payment Tier
 - 1 – Low
 - 2 – Medium
 - 3 - High
- Experience
- Age
- City
- Gender
- Ever Benched

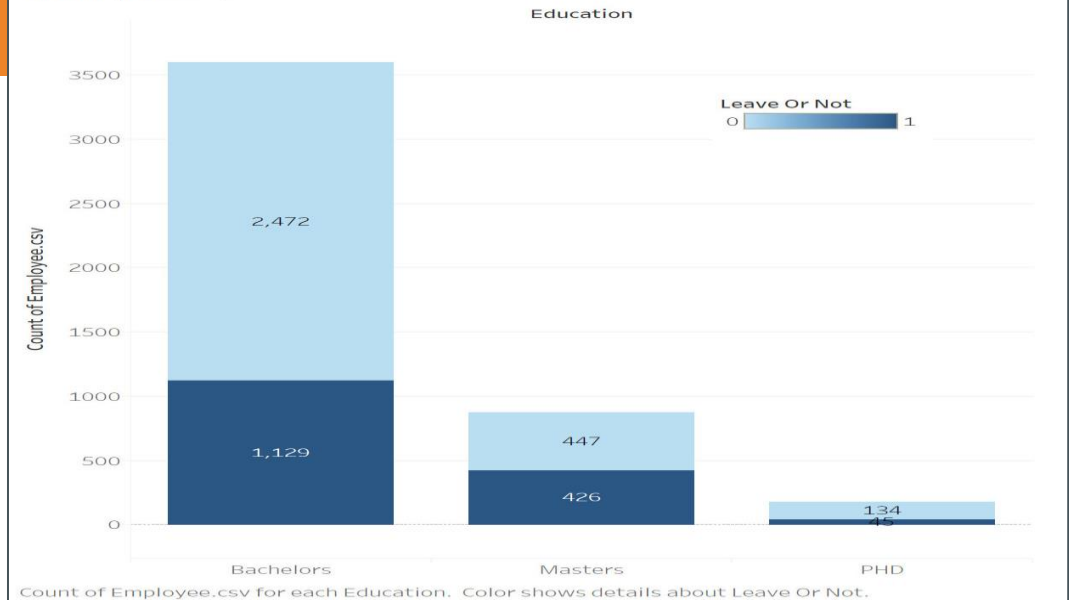
Dependent (Target) Variable

- Leave or Not
 - 0 = Stay
 - 1 = Leave

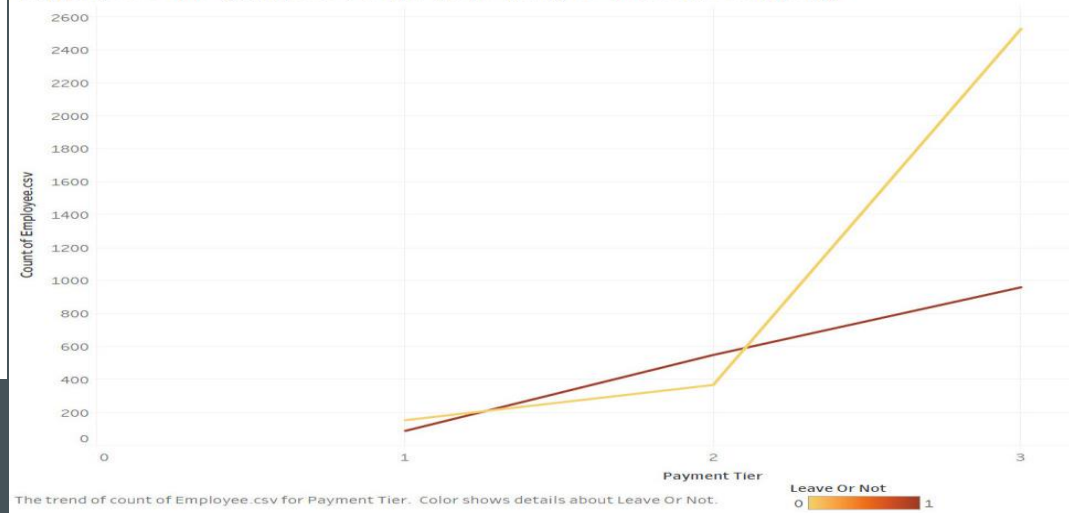
Data Visualization:

- ✓ For Data exploration we used Tableau to understand the data.
- ✓ For Data Visualization we used Bar charts, Line charts, Pie diagrams, Side by side Bar and Pie charts.

Employees with a Master's degree are more likely to leave their jobs. |



Employees of Payment tier 2 are more likely to leave the company



HOW DO YOU USE THE DATA TO ANSWER THE BUSINESS QUESTION?

WHAT ARE THE DATA MINING MODELS USED?

The goal is to find out the probability of employees who are leaving, Staying and Uncertain.

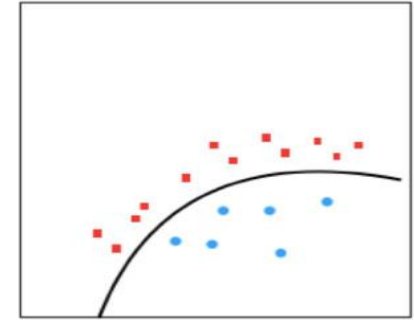
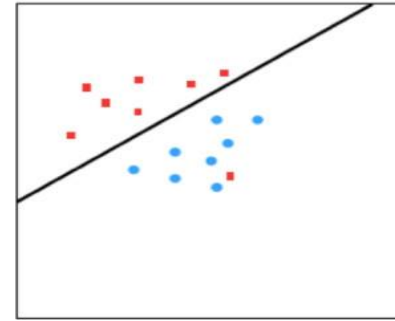
1. We take the data and divide it into three sections: 70%, 20%, and 10%.
2. We set aside 10% of the data. At the end, this will be used as new data to make predictions.
3. Use 70% of the data to train the model and 20% to validate the model.
4. Calculate the probability of leaving, which are classified as follows:
 - 0 % to 20% - indicates that the probability of employees leaving the company is very low. There is no need to Incentivize because they won't leave the company anyway.
 - 70% to 100% indicates that the probability of employees leaving the organization is very high. There is no need to incentivize because they will leave the company anyway, even if they are incentivized.
 - 20% to 70% - the probability of employees who are uncertain (Undecided) whether to leave or stay with the company. These employees need to be incentivized to obtain maximum retention.

To train, validate and test the data, we used multiple data mining models like Linear, Logistic and Decision trees.

ABOUT DATA MINING MODELS USED AND WHY DID YOU CHOOSE THEM?.

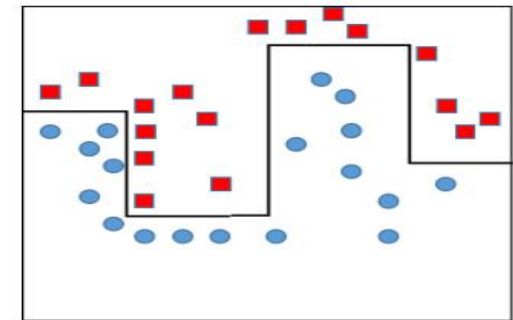
At first, we apply both Regression and Decision Tree to see which model produces better results.

Logistic Regression assumes that the data is linearly (or curvy linearly) separable in space on to exactly two planes.



Separable in space

Decision Trees are non-linear classifiers; they do not require data to be linearly separable. They Bisect the sample space in to smaller and smaller Regions



Non-linearly separable data

We are sure that our data set divides in to exactly two separable parts, so we have chosen to go with Regression model as it is performing best on our data.

We have chosen Logistic regression over Linear regression because of the following reasons,

1. The dependent variable (Target Variable) in our dataset is binary and Logistic regression is highly used in this case.
2. Also, the Misclassification rates and Average squared errors for Logistic regression are slightly less when compared to Linear regression. Please see the image on the left side of this page.
3. The event classification table for both models are given below on the right-side of the page, as you can see the False Negative values are better for Logistic regression.

Fit Statistics

Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected	Model		Valid:	Train:	Train:	Valid:
Model	Node	Model Description	Misclassification	Average	Misclassification	Average
			Rate	Squared	Rate	Squared
				Error		Error
Y	Reg	Logistic Regression	0.26882	0.19218	0.27611	0.18945
	Reg2	Linear Regression	0.27527	0.19374	0.28409	0.19142

Event Classification Table

Model Selection based on Valid: Misclassification Rate (_VMISC_)

Model		Data	Target	Target	False	True	False	True
Node	Model Description	Role	Target	Label	Negative	Negative	Positive	Positive
Reg2	Linear Regression	TRAIN	LeaveOrNot		752	1963	173	368
Reg2	Linear Regression	VALIDATE	LeaveOrNot		220	575	36	99
Reg	Logistic Regression	TRAIN	LeaveOrNot		715	1952	184	405
Reg	Logistic Regression	VALIDATE	LeaveOrNot		214	575	36	105

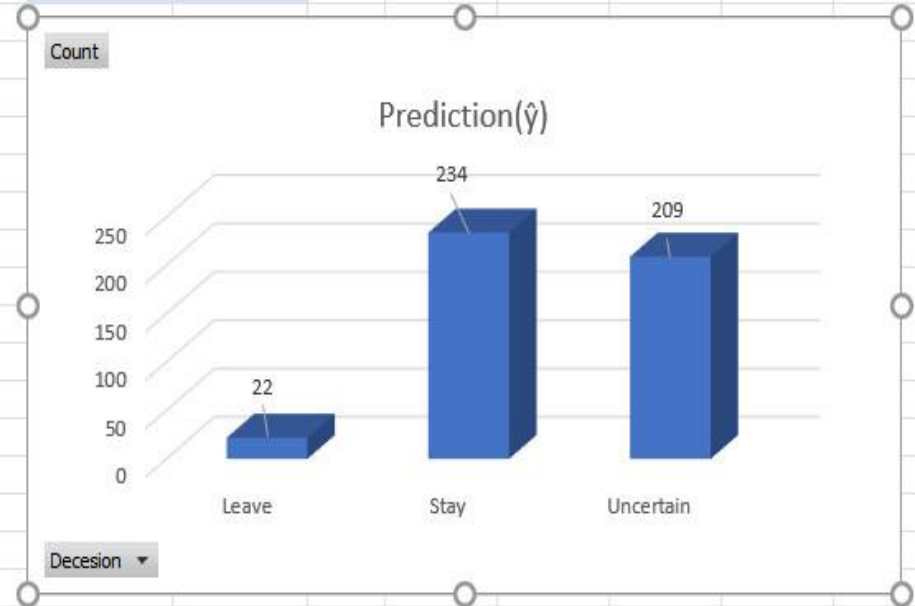
Prediction (\hat{y}):

The following are the predictions based on the 10% of data (465 observations) kept aside at the beginning of data partition.

Prob(%)	Decision	Count
0 to ≤20%	Stay	234
20% to ≤70%	Uncertain	209
70% to 100%	Leave	22
Total		465

Test_Observations	Probability	Decesion
4221	0.581429898	Uncertain
4222	0.221155617	Stay
4223	0.117732334	Stay
4224	0.768634445	Leave
4225	0.358915988	Uncertain
4226	0.231398209	Stay
4227	0.806980715	Leave
4228	0.148342024	Stay
4229	0.340450617	Uncertain
4230	0.194121188	Stay
4231	0.278679987	Stay
4232	0.110227337	Stay
4233	0.222727673	Stay
4234	0.46239642	Uncertain
4235	0.367370611	Uncertain
4236	0.318720945	Uncertain
4237	0.156794143	Stay
4238	0.304791511	Uncertain
4239	0.519720779	Uncertain
4240	0.2667256	Stay
4241	0.457075374	Uncertain
4242	0.77613803	Leave
4243	0.376601682	Uncertain
4244	0.47542477	Uncertain
4245	0.192401635	Stay
4246	0.244024448	Stay
4247	0.378437997	Uncertain
4248	0.369279011	Uncertain
4249	0.394992897	Uncertain
4250	0.463522271	Uncertain
4251	0.163742572	Stay
4252	0.322457408	Uncertain
4253	0.577132209	Uncertain

Decesion	Count
Leave	22
Stay	234
Uncertain	209
Grand Total	465

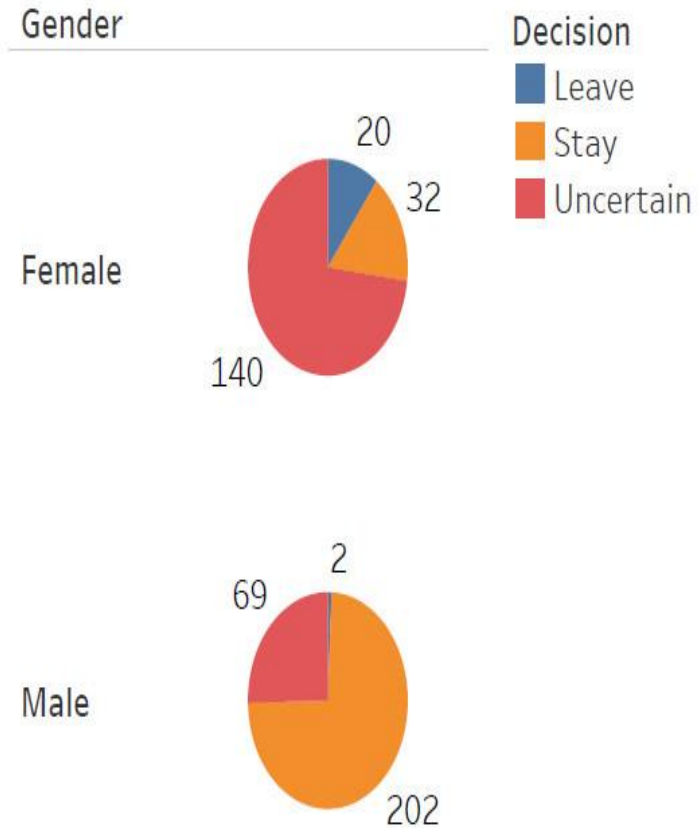


Data Prediction Insights for Management:

The Pie Chart depicts that the Female employees are more likely to leave the company compared to Male employees.

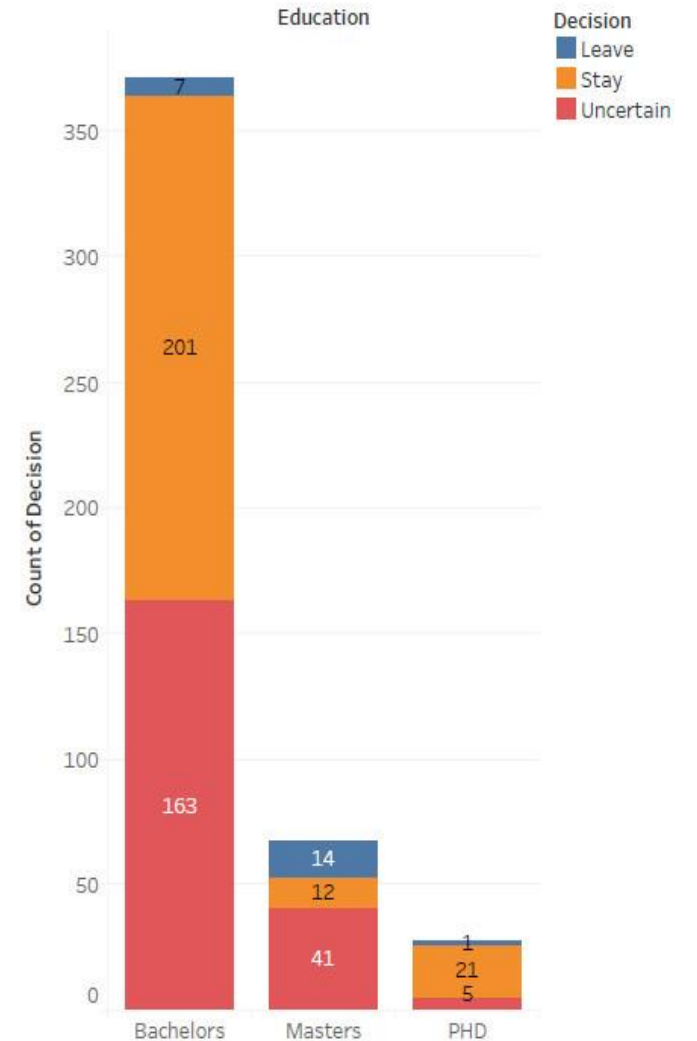
The Bar graph shows that the employees with master's degree are more likely to leave company compared to others.

Gender Vs Decision



Decision (color) broken down by Gender.

Education Vs Decision



Count of Decision for each Education. Color shows details about Decision.