HR Analytics

Group:

- Aryama Ray
- Arvind Kumar
- Anand Naik
- Chetankumar Trivedi

Problem Statement

Problem Statement: Company is having annual attrition of 15% and that needs replacement of talent pool from job market.

This is bad for the company due to following issues:

- Delay meeting projects timelines – Reputation loss amongst customers & partners.
- Need recruiting good number of new talents – Sizeable staff is to be maintained in HR.
- Lead time to acclimatize new recruits – Need to spend good amount of resources (time, money & energy) in training before they start delivering.

Objective and Deliverables

Objective:

- Find out the factors causing attrition using Data Analytics Tools.
- Understand the impact of each identified factors on attrition.

And therefore facilitate arriving at…

✓ Changes to be done in workplace to make employees stay.

Deliverables:

- Find out probability of attrition using logistic regression model.
- Build a robust model which can predict the attrition well.

Modelling Approach

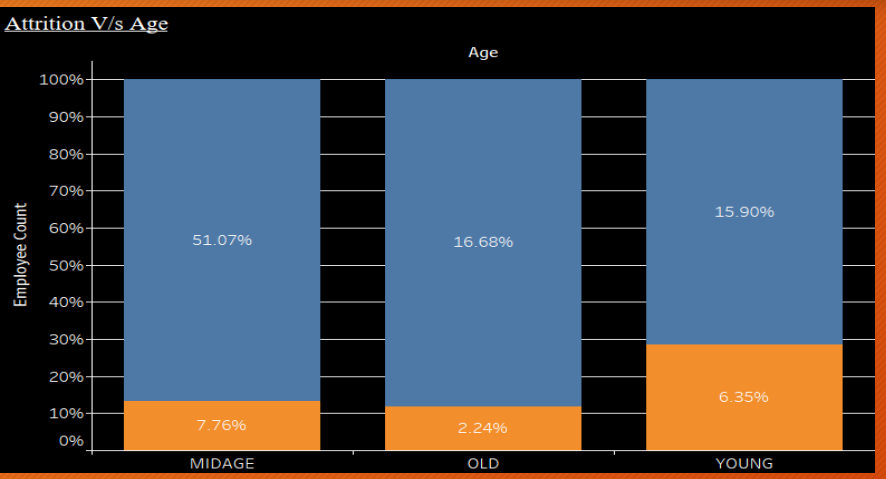
Steps Followed	Activities Carried Out	Actions Executed
Data Acquisition	1. Identify Data Sources	1. Five data sources identified and acquired - General Data, Employee Survey Ratings, Manager Survey Rating and Attendance Data. Employee ID found to be unique key across sources.
Data Preparation	1. De-duplication 2. Missing value treatment 3. Removal of outliers 4. Derive new metrics	1. Did not find any duplicates across 4410 records. 2. Attendance data ▪ Columns with all missing values – Office holiday and hence no attendance, completely removed. ▪ Missing values found in pair on some of the dates for few employees - Employee might be on leave or travel and hence not come to office. Did not removed but did not considered for derived metrics. Final dataset – Only 2% of total records were having missing values hence dropped. 3. Monthly Income has a wide range, found outliers and removed. 4. Average time in office is the only newly derived metrics based on attendance data.

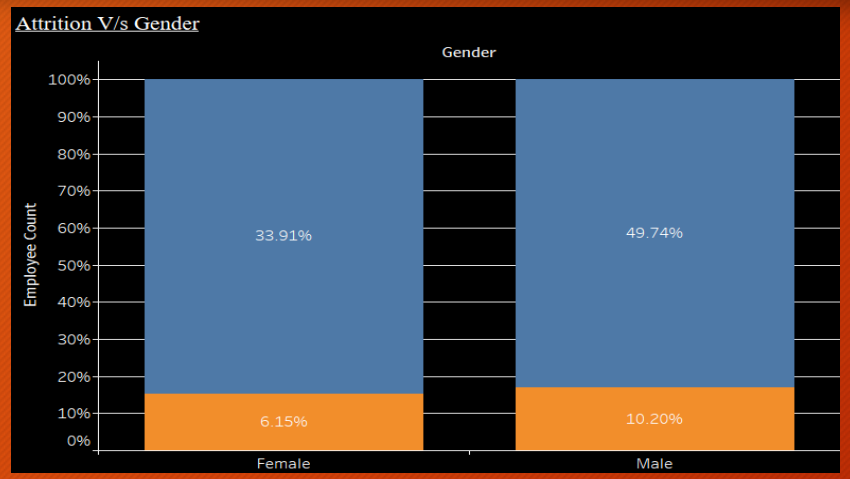
Modelling Approach (Contd…)

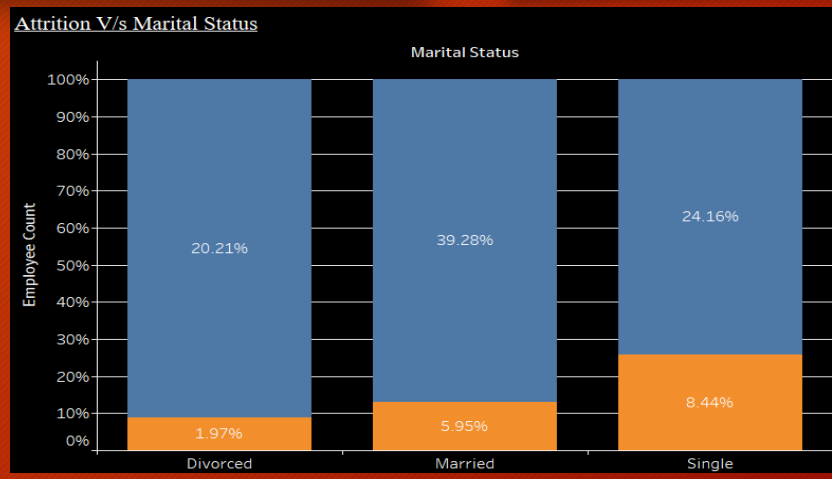
Steps Followed	Activities Carried Out	Actions Executed
Data Preparation	5. Binning of categorical variables 6. Normalization of numeric variables 7. Prepare final clean dataset	5. All categorical variables are logically binned. 6. Found different ranges for different variables. Scaled them to a common range. 7. Some variables have single value, dropped them. Converted all variables to numeric variables – Total 3969 records with 68 variables.
Exploratory Data Analysis	1. Understand relationship between Attrition and other parameters	1. Plotted attrition v/s all parameters and found relationship between them.
Build Regression Model	1. Prepare Training & Test data 2. Build model	1. Have sufficient records with attrition, hence standard 70:30 ratio taken to prepare training and test datasets. 2. Final model built with only few (# 8) but high impact parameters for stability of model.
Test the Model	1. Test robustness of model	1. Model tested with optimal probability cut-off through Sigmoid Curve, Accuracy, Sensitivity, Specificity, KS Statistics and Lift & Gain Chart of the model.

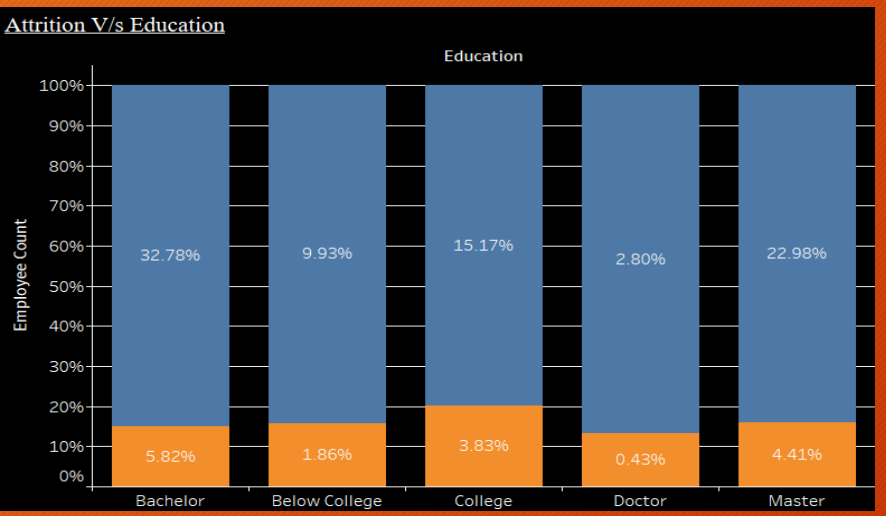
Colour Legends: *Attrition* *No Attrition*

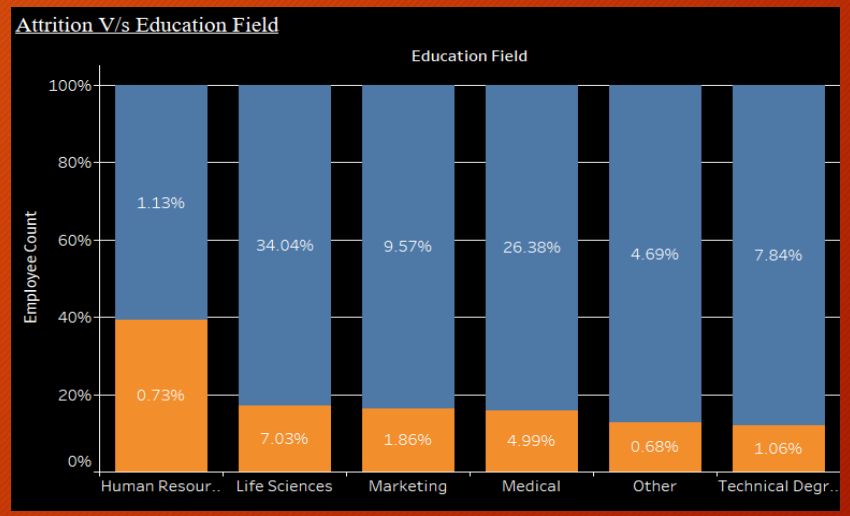
Exploratory Data Analysis *(Personal Attributes)*









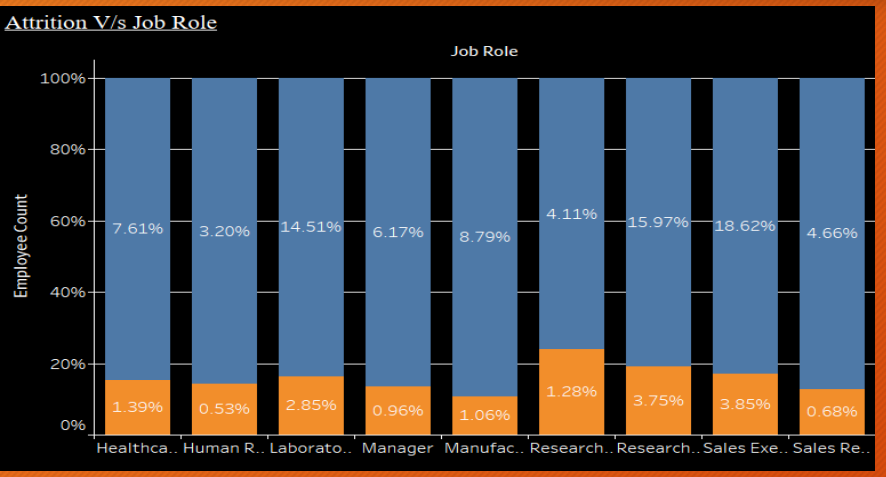


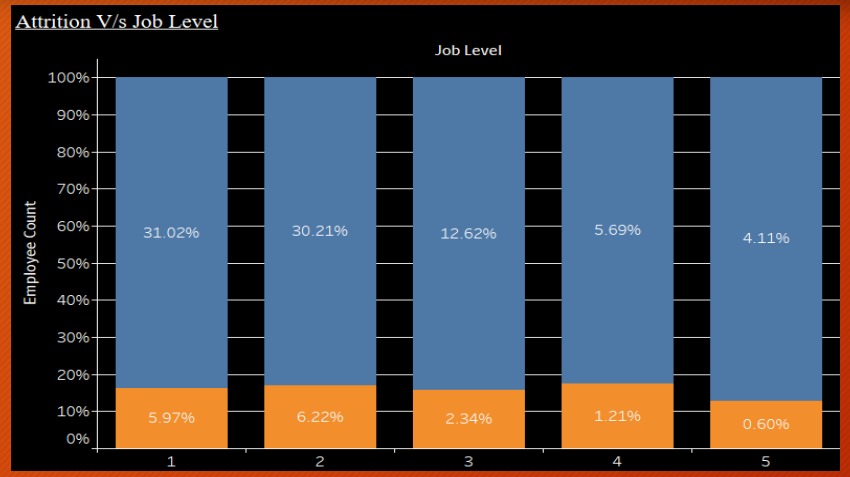
Observation:

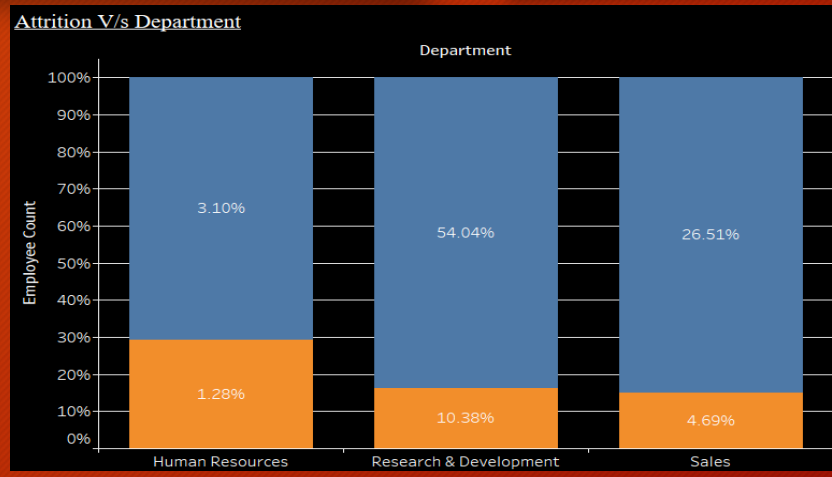
- *Young professionals contribute more to attrition rate as compared to older ones.*
- *Single employees have more attrition rate both in absolute terms and its own bin.*
- *Education and Gender do not seem to have impact on attrition.*

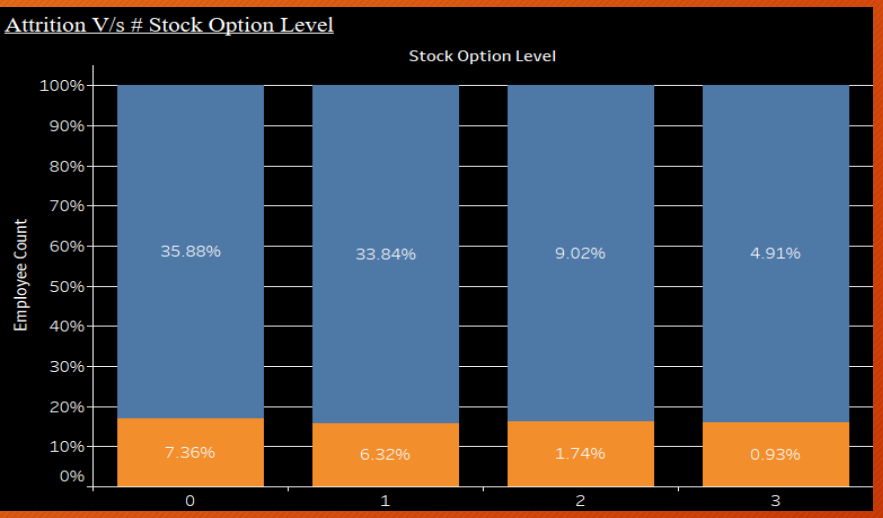
Colour Legends: Attrition | No Attrition

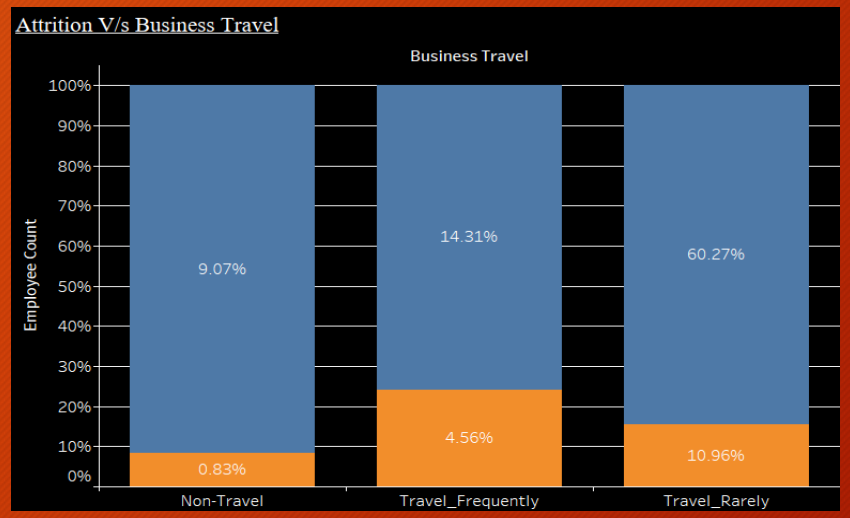
Exploratory Data Analysis *(Company Labels)*









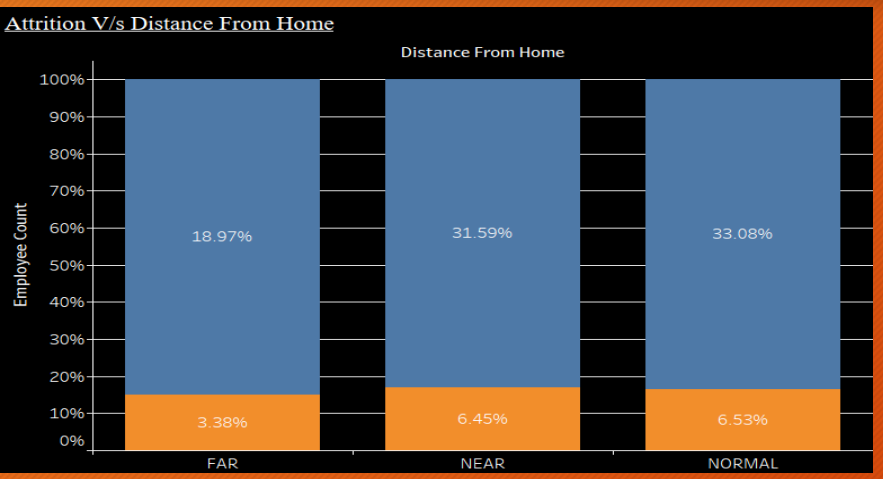


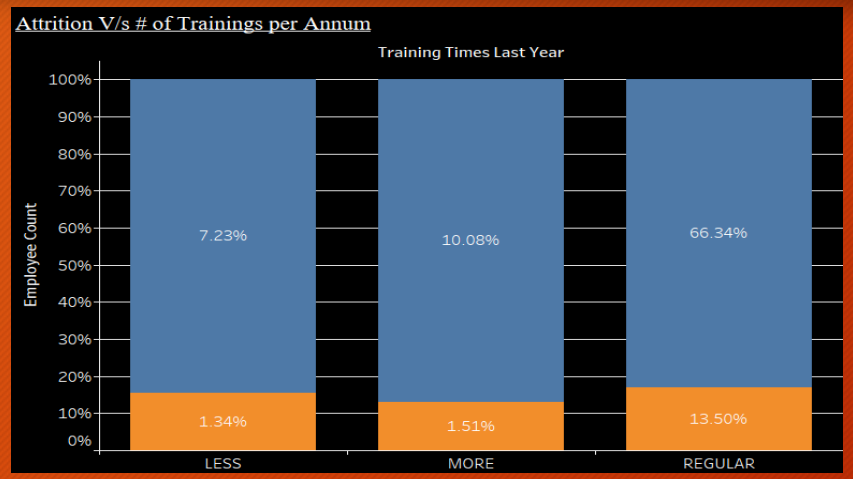
Observation:

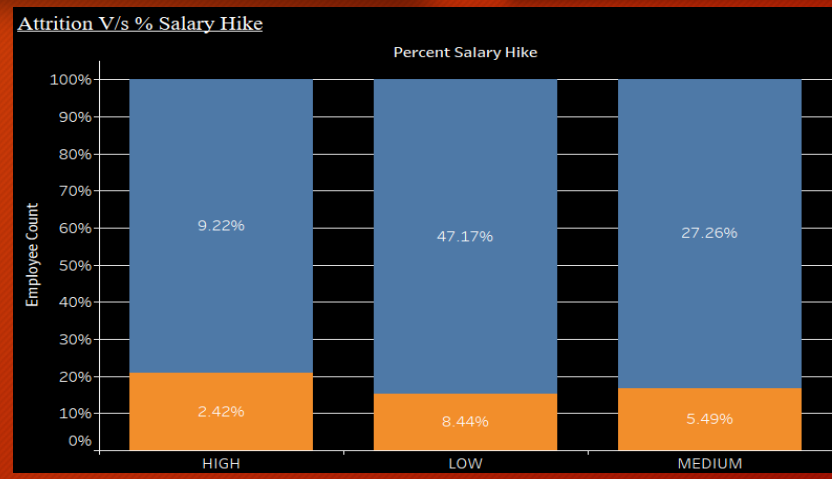
- *Attrition rate is evenly distributed across all job roles, job levels and stock option levels.*
- *HR has more % ratio of attrition v/s no attrition. Though absolute % of attrition is low in HR.*
- *Employees travelling frequently are more prone to attrition.*

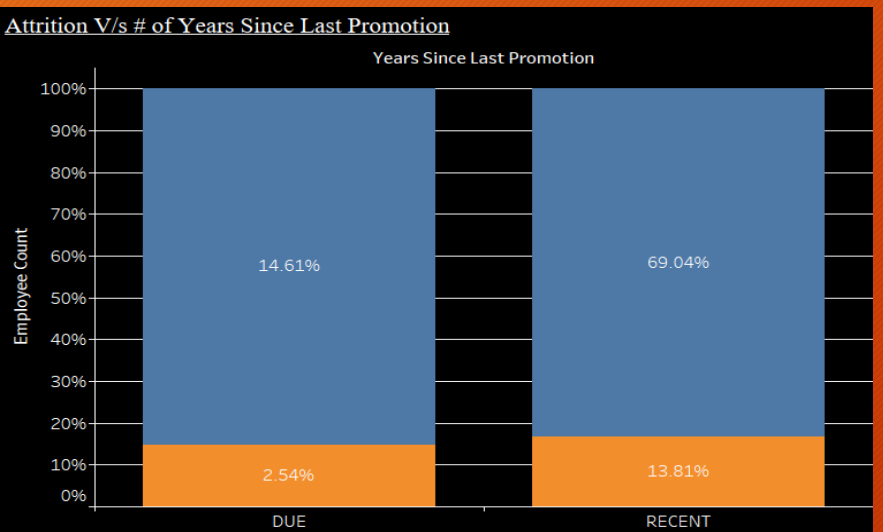
Colour Legends: *Attrition* *No Attrition*

Exploratory Data Analysis *(Company Labels)*







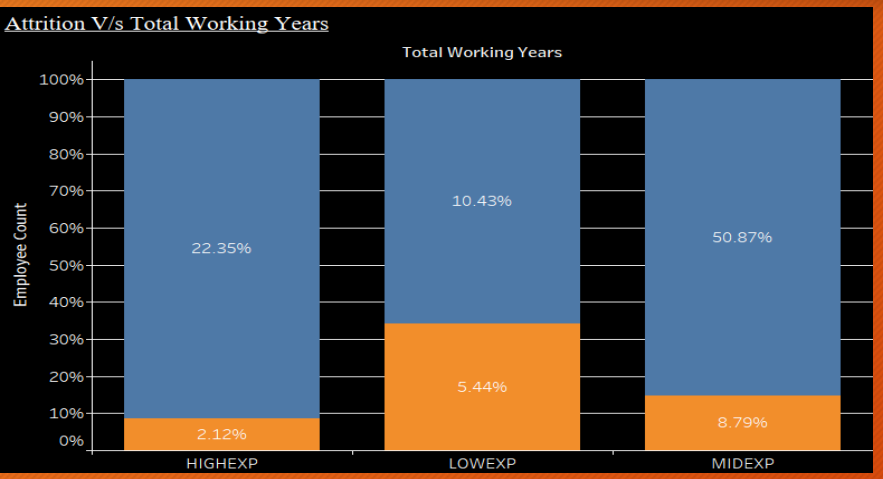


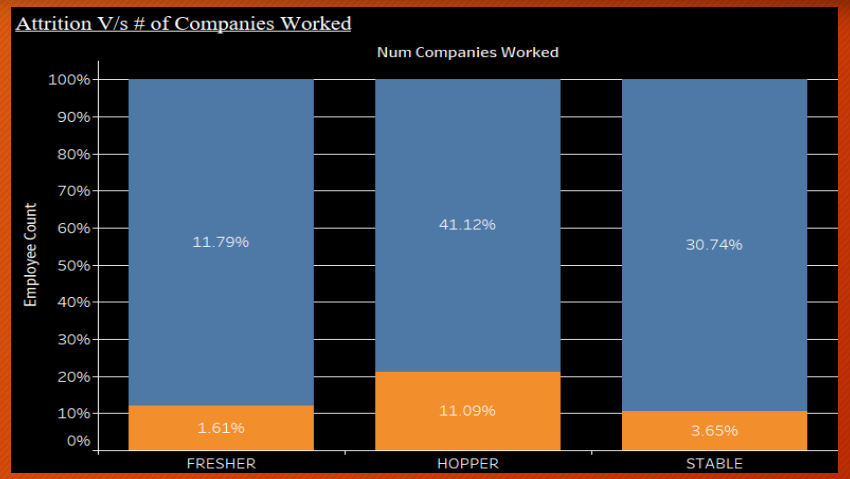
Observation:

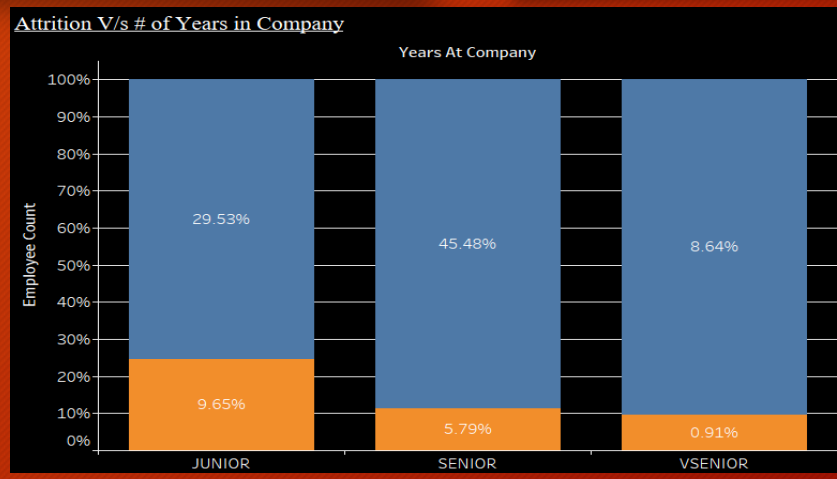
- *Travel distance from home to company does not seem to be a significant factor for an attrition.*
- *% Salary hike does not show clear linear relationship with attrition.*
- *Attrition is more amongst employees who have been imparted regular trainings.*
- *It is observed that employees recently promoted have more attrition.*

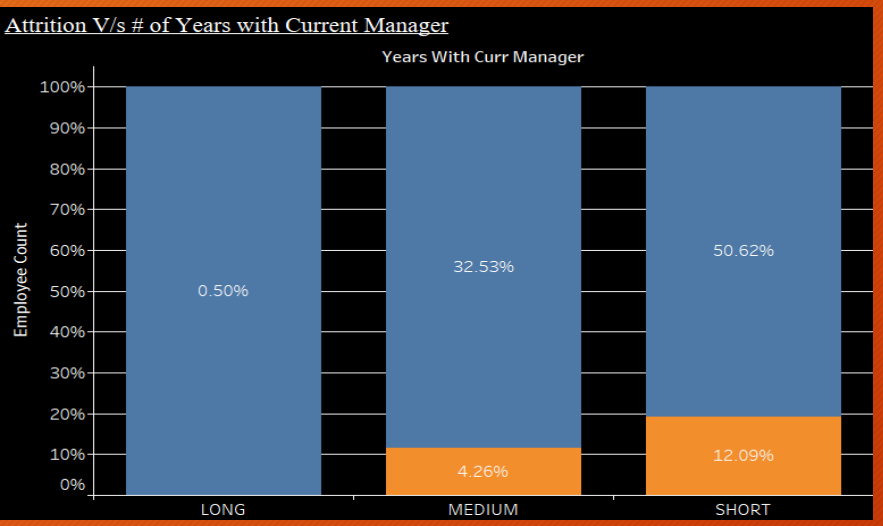
Colour Legends: *Attrition* *No Attrition*

Exploratory Data Analysis *(Working History)*







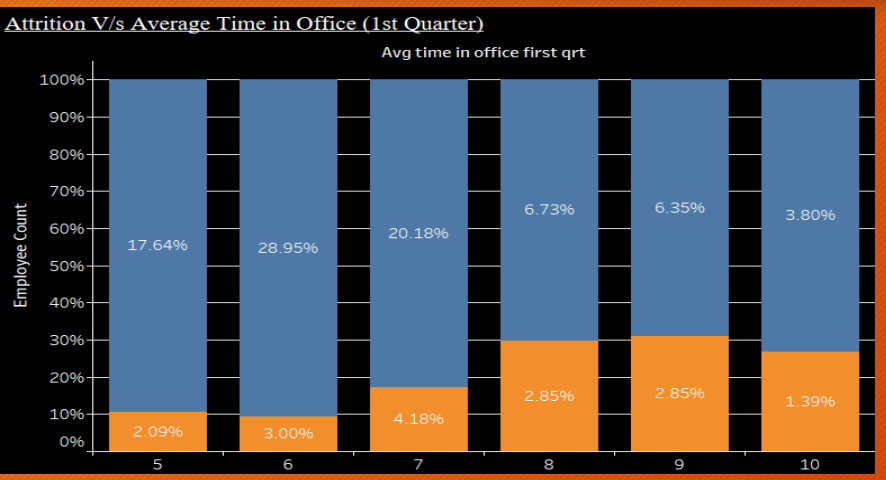


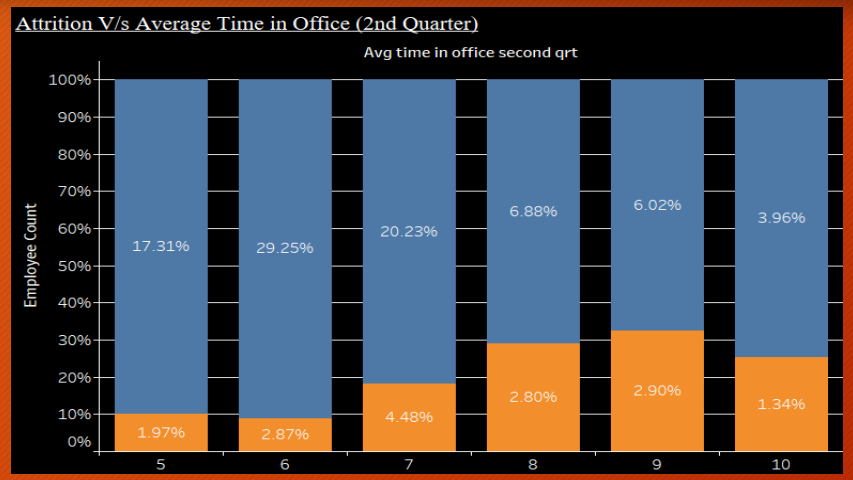
Observation:

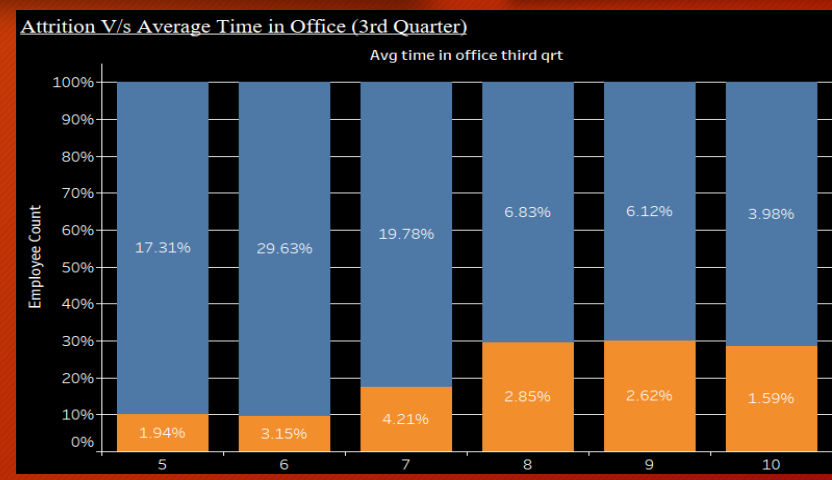
- *Ratio of % Attrition v/s No Attrition is higher for employees having:*
  - *lesser total working years*
  - *more number of companies worked*
  - *lesser working years in company*
  - *lesser working years with the same manager*
- *In absolute terms, % attrition is very high for employees who have hopped to many companies in the past.*

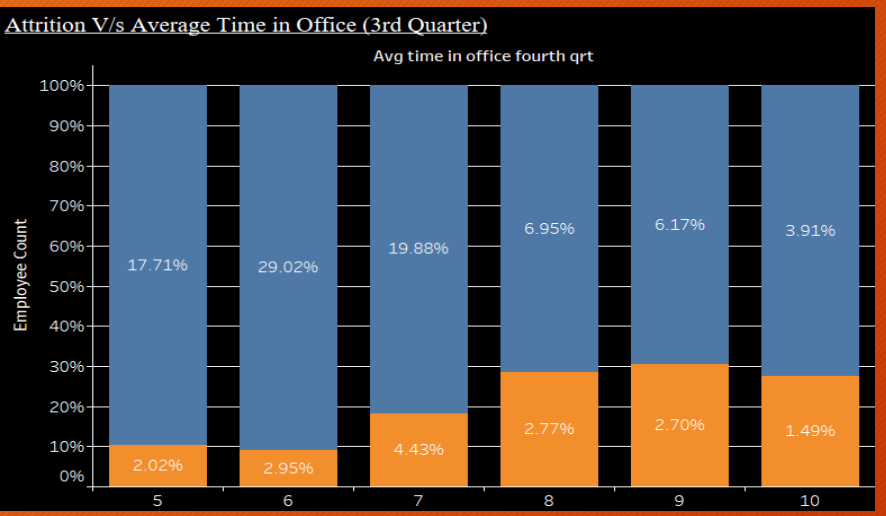
Colour Legends: Attrition | No Attrition

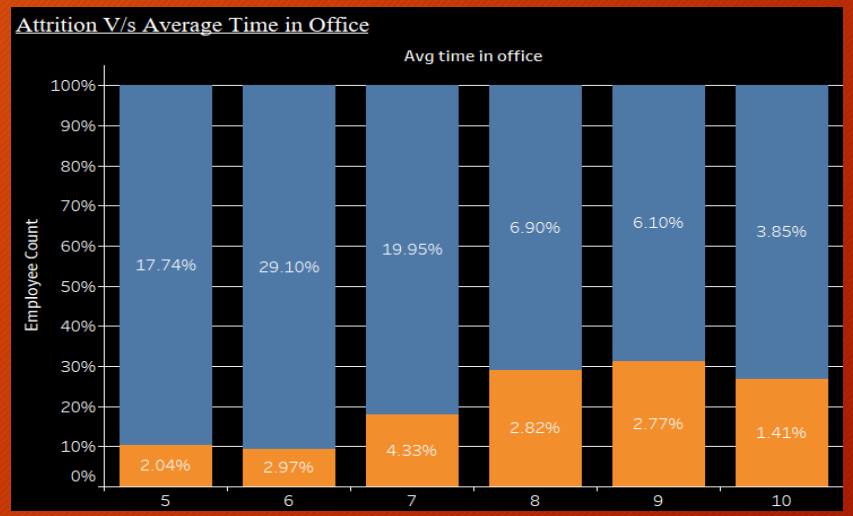
Exploratory Data Analysis *(Time in Office)*









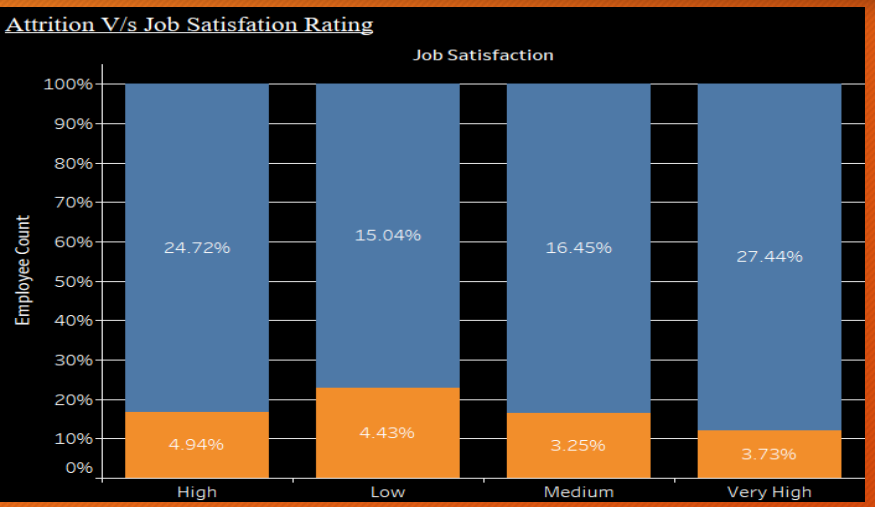


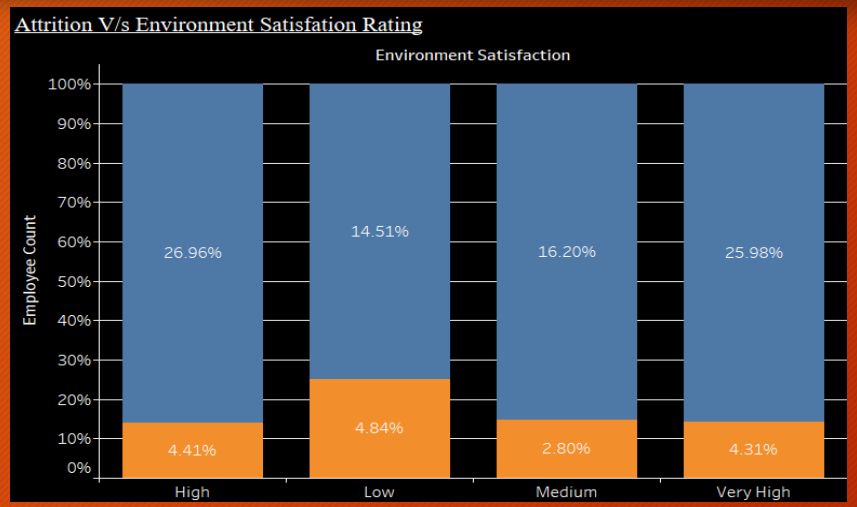
Observation:

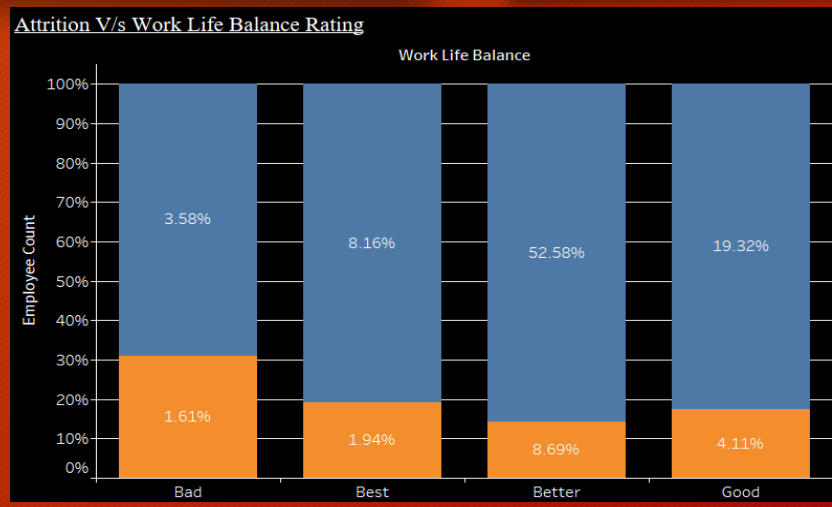
- *Attrition is more amongst the employees who stay late in office.*
- *There is a high co-relation amongst these time related variables.*

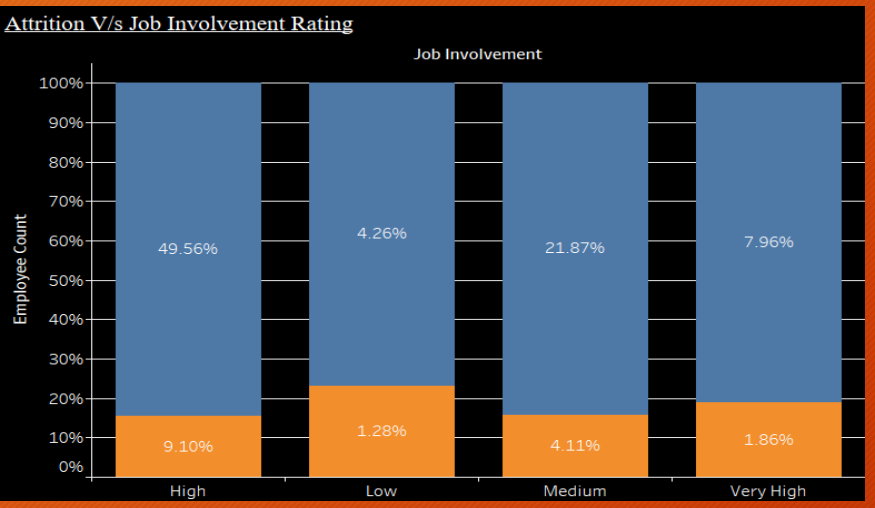
Colour Legends: Attrition | No Attrition

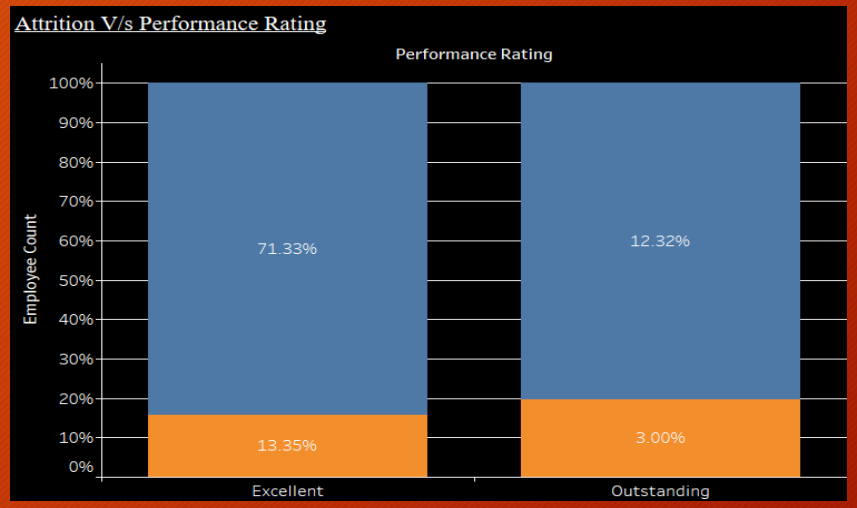
Exploratory Data Analysis *(Survey Ratings)*







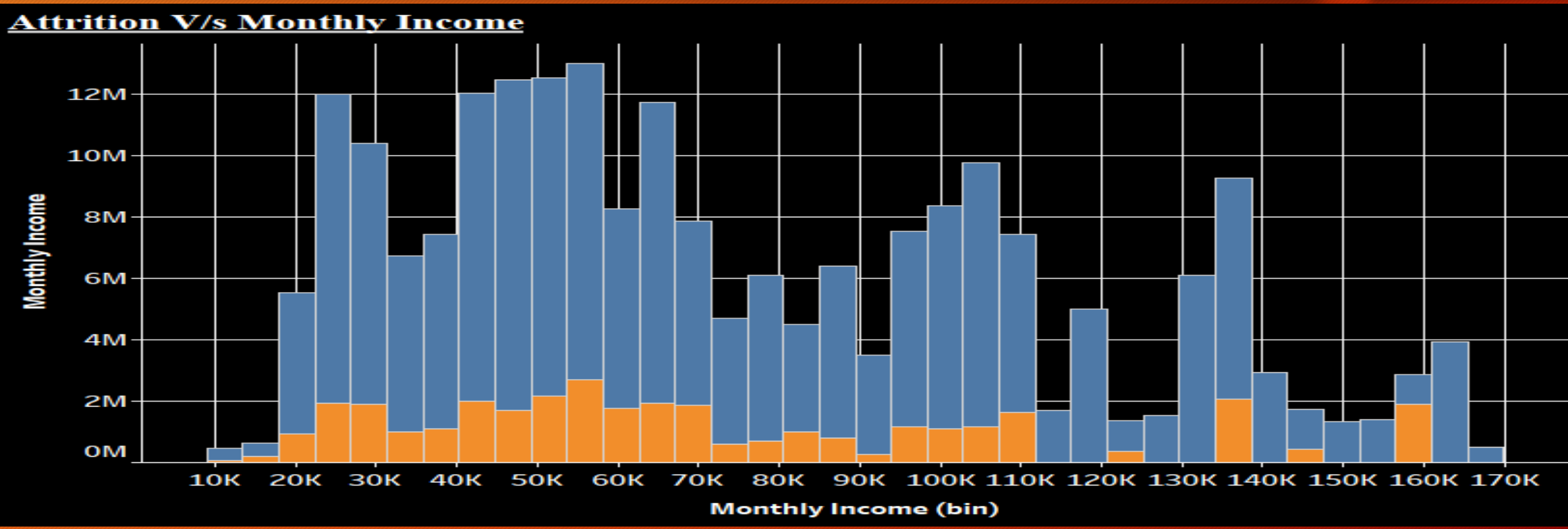




Observation:

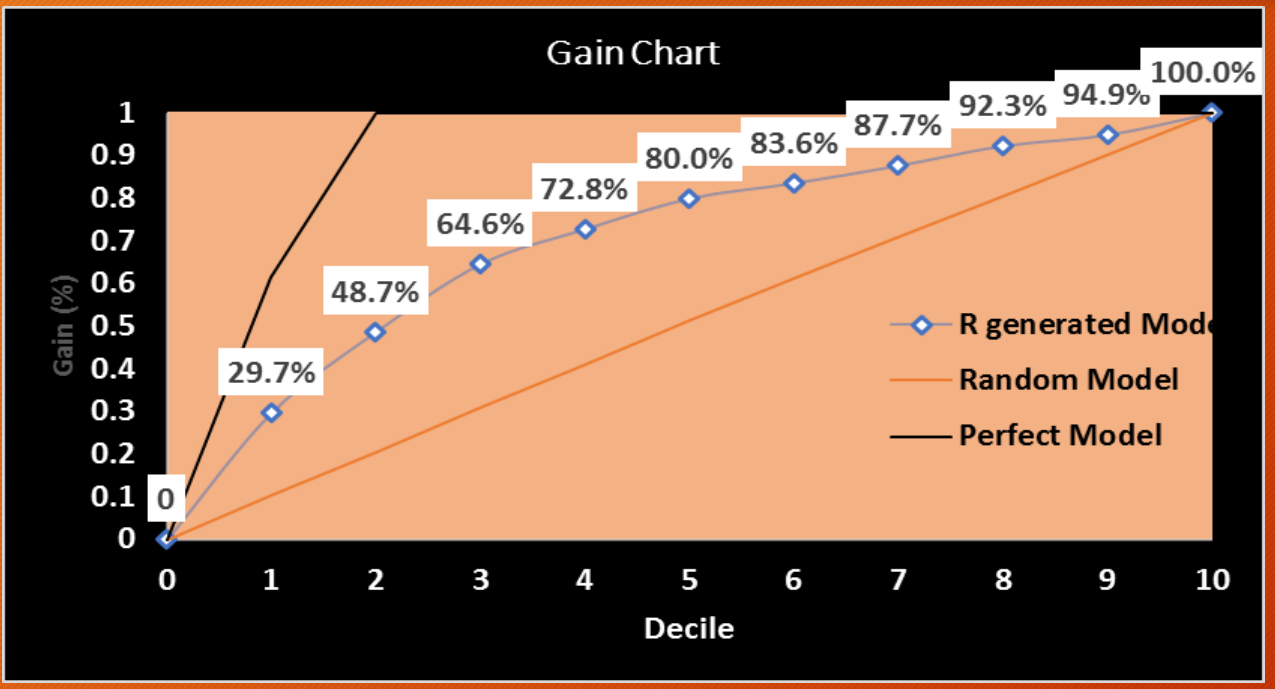
- *Employees under lower bucket of survey ratings are more likely to leave.*

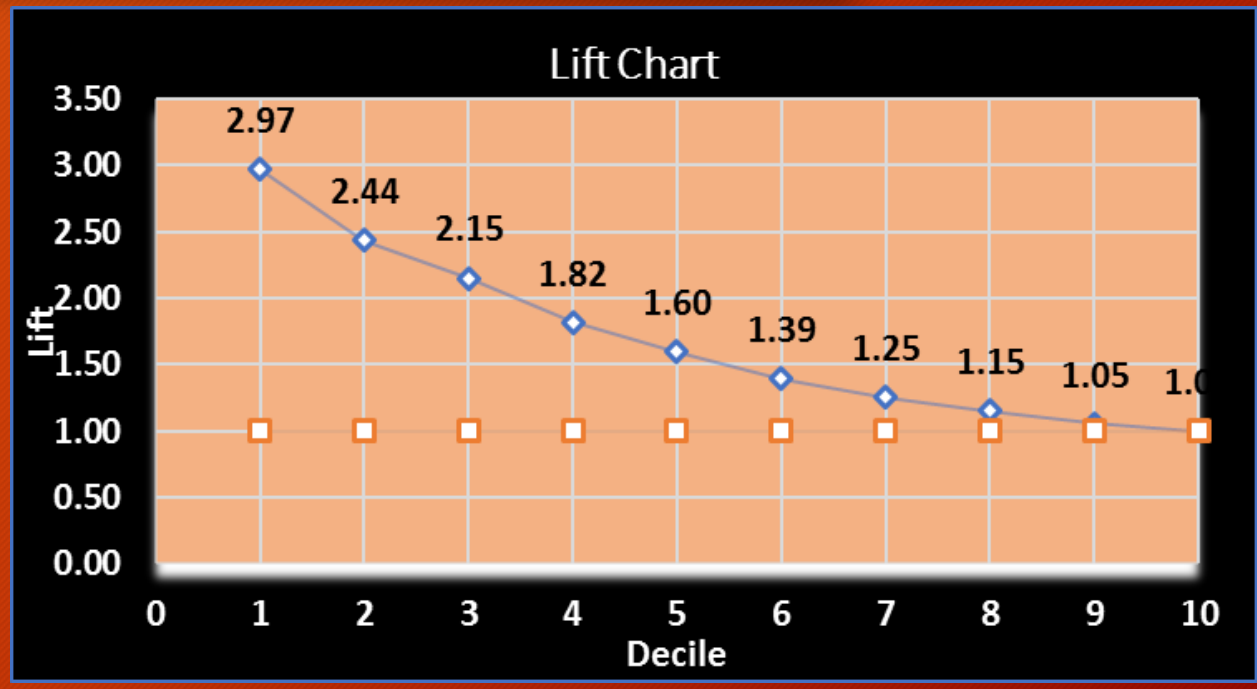
Exploratory Data Analysis *(Monthly Income)*



Observation: It seems that attrition is distributed across all income levels.

Robustness of Model *(Model's Discriminative Power Measures)*





1. *Model has good level of Accuracy = 72.71% , Sensitivity = 67.69% and Specificity = 73.70% at optimal cut-off probability of 0.1775758*
2. *It has a decent Gain and Lift as compared to random model.*
3. *It has KS Statistics of 41.39% at 3rd decile.*

Key Findings and Recommendations *(Results of a Model)*

Our model has found that following parameters impacts significantly to Attrition:

#	Factor	Finding	Recommendations
1	Age	Attrition is higher for younger employees (< 30 yrs. Age) (Age brings more Maturity !!)	1. Re-define the selection criteria and process of recruitment. Be mindful of previous career history and family background while selecting a candidate. 2. Incentivise the Job Profile where travel is more frequent. 3. High potential employees who are promoted, shall be given higher responsibilities upon promotion and may be covered with a better Stock Option Level to promote retention. 4. Review the work culture and organization structure where employees regularly stay long hours in office. Need to understand the root causes and address them. 5. Interact with employees who have given low ratings in employee survey, understand the reasons and accordingly need to address them in a time-bound manner.
2	Marital Status	Employees with a "Single" status are likely to leave more as compared married ones (More freedom !!)	
3	No. of Companies Worked	Employees who have worked more # of companies in a career earlier are prone to leave faster (Hopping Mentality !!)	
4	Business Travel	Employees in job profiles where travel is frequent, tend to leave the company (Work Life Balance !!)	
5	Years Since Last Promotion	Attrition is higher for employees who are recently promoted (> 46% are promoted current year & > 78% are promoted in less than 3 years back) (What's more for me now !!)	
6	Time Spent in Office	Attrition is more amongst the employees who have to sit long hours in office (Regular slogging wears you out !!)	
7	Environment Satisfaction	Employees who haver rated low for Environment and Job Satisfation parameters in employee survey are more prone to leave (Money alone don't buy happiness !!)	
8	Job Satisfaction		

Thank You