WORK SAMPLE

- ANALYZING ATTRITION

Ann Xiao
Yifeng Wang
02.19.2019

EXECUTIVE SUMMARY

This analysis portraits the commonalities of employees who have left the company in the past year. The determinants of attrition were explored, followed by proposals in action plan and next steps.

Who are most likely to leave:

- Millennials and Gen X who have mid-to-high level of education;
- Generally worked for more companies comparing to their peers and have stayed in the company for shorter time period;
- Ranked lower in job levels, earn less, and have less stock options although their performance may be satisfactory;
- They are likely to leave during their 1st year at the company, between 5-7 years of total working years, and less likely so after year 10;
- Have lower income, though not necessarily have worse performance;
- Prefer less overtime, and are less satisfied in work-life balance, work environment, job involvement, and training opportunities;
- Also tend to have more have more flexible skill sets from educational field

REASEARCH METHODOLOGY

01

Data Exploration Analysis

What Happened ?

02

Hypothesis Verification

Is It True and Why Did it Happen?

03

Next Steps & Action Plan

What Should We Do About It?

1.

Data Exploration

DATA OVERVIEW

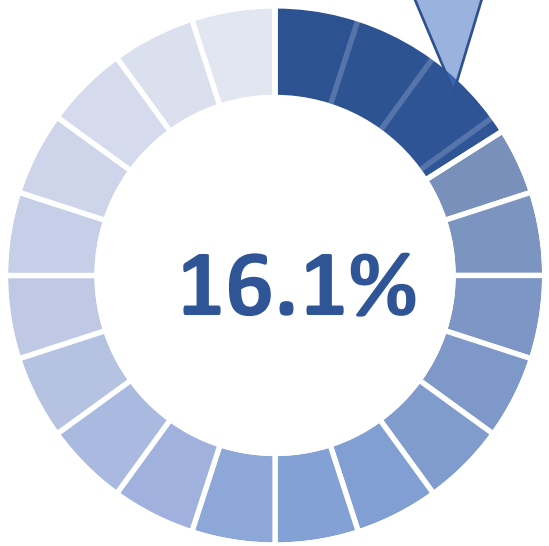
The dataset is 100% populated with 1470 records, attrition info (Yes/No), and 34 variables. After cleaning, it's ready for visualization.

The important variables possibly containing explanatory and predictive info for attrition can be categorized into **3** main buckets:

- **Personal Descriptive Info**
  - Education Field
  - Age
  - Total Working Years
  - Etc.
- **Total Compensation**
  (both monetary and intangible)
  - Monthly Income
  - Stock Options
  - Job Level
  - Training times Last Year
  - Etc.
- **Employee Satisfaction**
  - Work-life balance
  - Travel Intensity
  - Job Satisfaction
  - Environment Satisfaction
  - Relationship Satisfaction
  - Job Involvement
  - Etc.

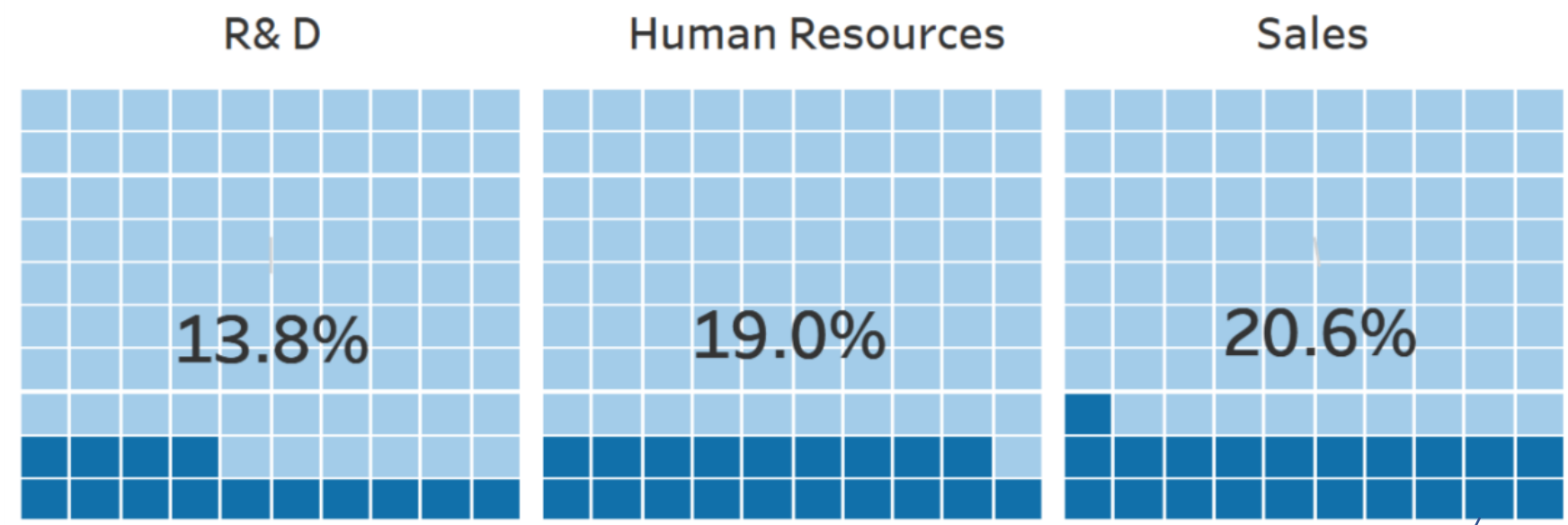
Attrition is defined as leaving the company altogether. It does not include transfers to other departments or cost centers.

For last year, the company's attrition rate is 16.1%, which is 2% higher than last year.



ATTRITION

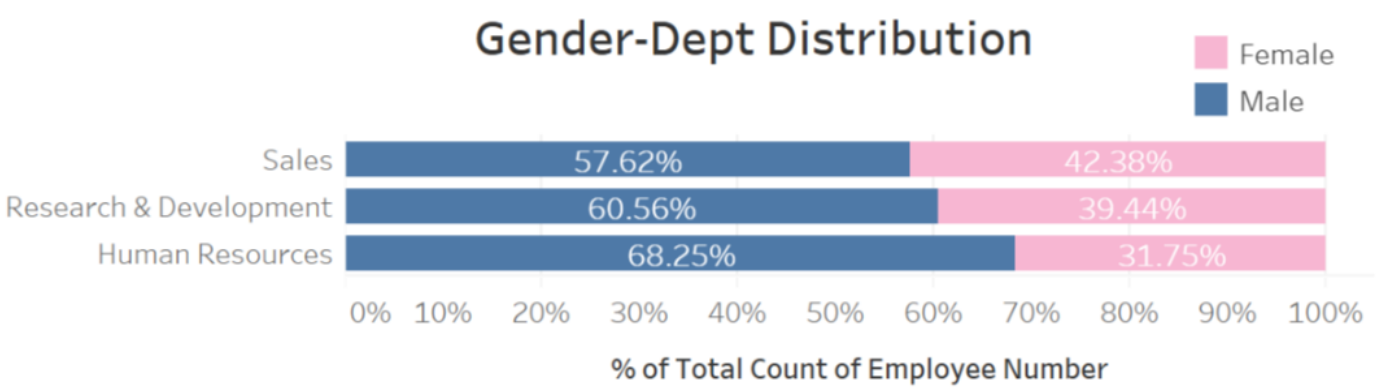
What are the attrition rate by department?

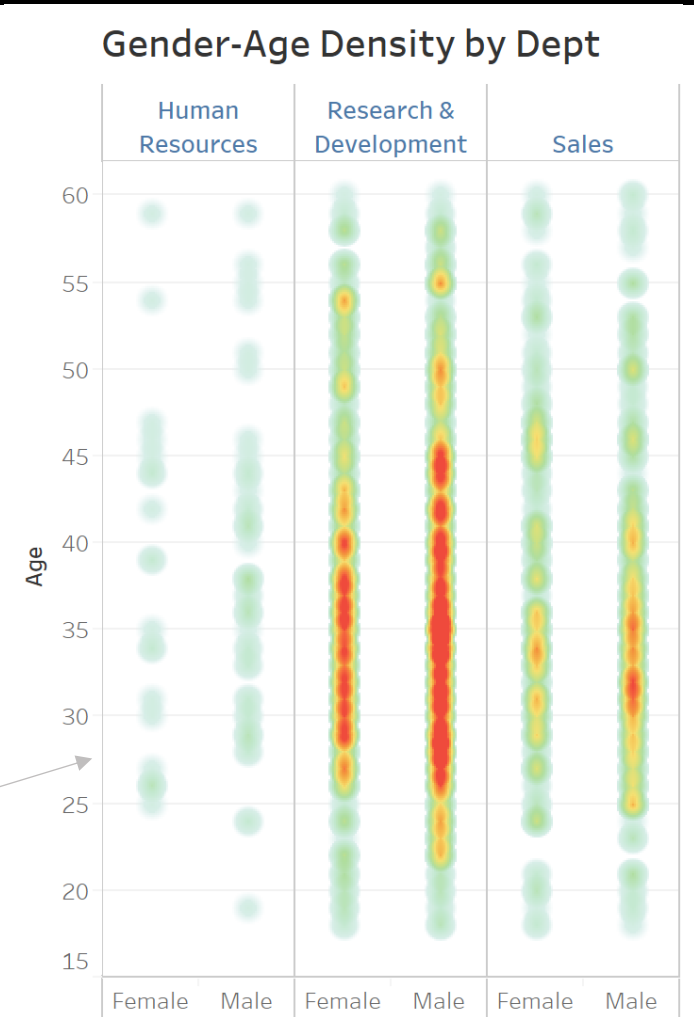


Sales is the department with the highest attrition.

Data Exploration | **Hypothesis Verification** | **Next Steps & Action Plan**

Gender distribution is relatively in reasonable scale, hypothetically attrition is not caused by gender bias.

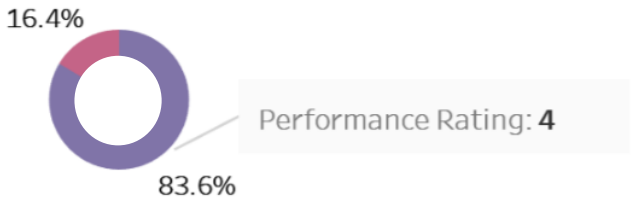


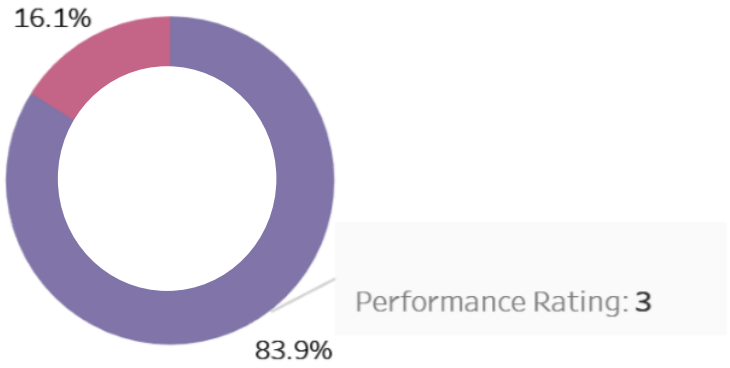


The density plot shows where the concentration of demographic distribution in both gender and age. The darker area shows it's most densely concentrated between around 27 – 46 in R&D and relatively evenly distributed among gender.

Data Exploration | **Hypothesis Verification** | **Next Steps & Action Plan**

So who left the company?

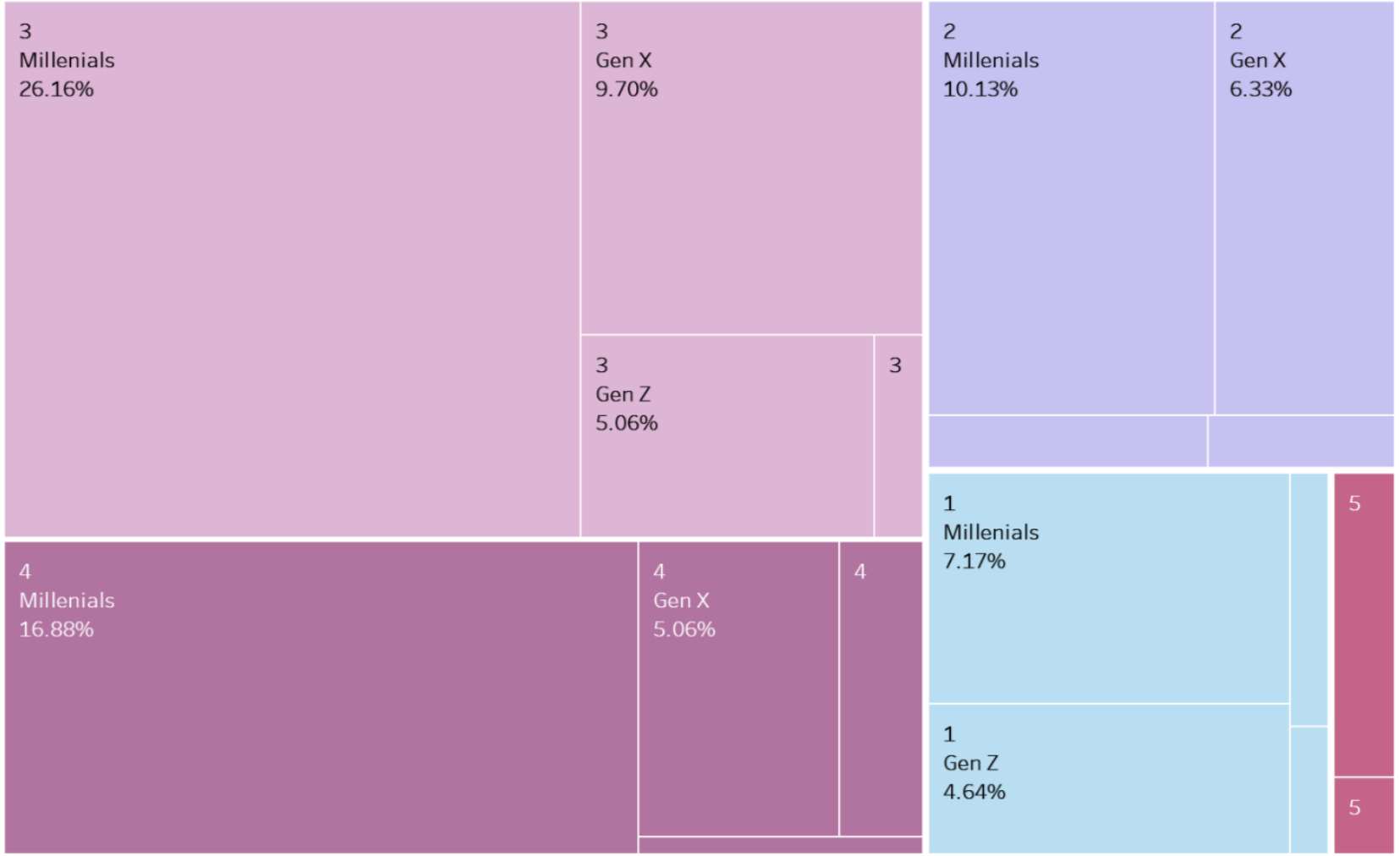




Performance in Relation to Attrition

Fortunately, we also didn't lose our high performers disproportionately.

From this chart we can see that in every education level we lose more millennials, mainly followed by Gen X, especially in Education level 3&4. Generally speaking, highly-educated middle-aged workers are leaving, possibly creating a gap in workforce succession in the future.



2.

Hypothesis Verification

Thru Chi-square Testing, age group in Gen X and Millennials is verified to have a significantly higher attrition rate, with P-value = 3.20768E-13.

Observation

Attritio	Boome	Gen X	Millenia	Gen Z	Total
No	58	443	30	702	1233
Yes	11	52	27	142	232
Total	69	495	57	844	1465

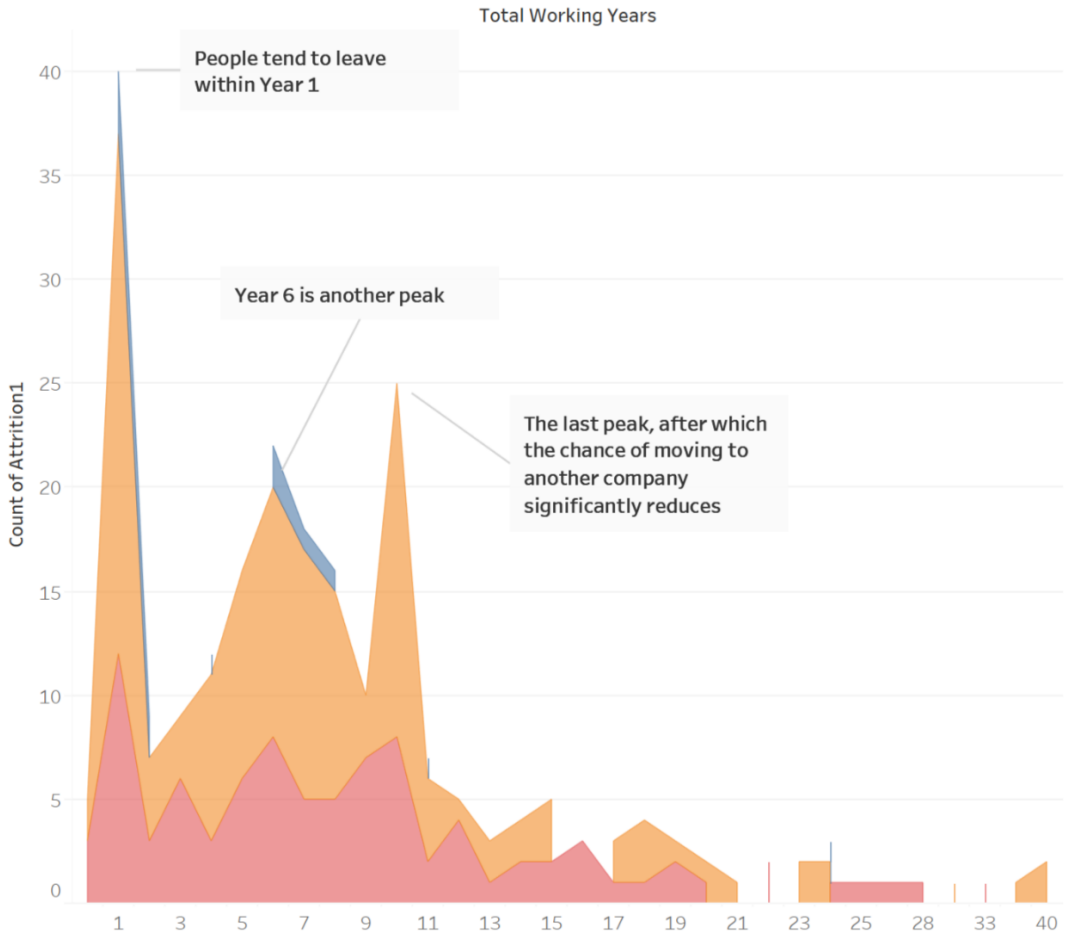
Expectation

Attritio	Boome	Gen X	Millenia	Gen Z	Total
No	58	417	48	710	1233
Yes	11	78	9	134	232
Total	69	495	57	844	1465

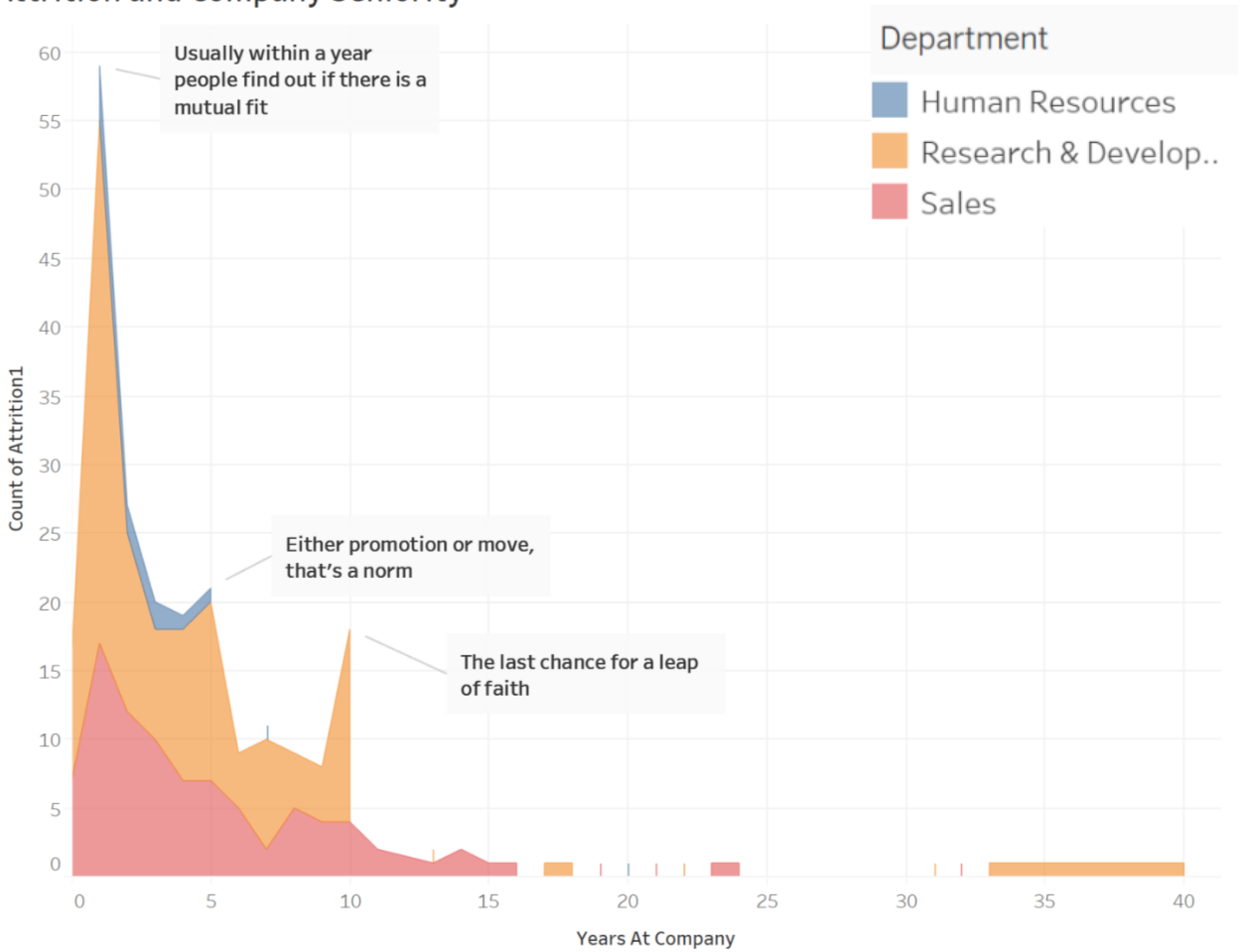
If generation is not an influencing factor, the differences wouldn't be significant.

It is observed from the chart below that there are several peaks in working years and years at the company that people tend to leave, which matches our experience – people tend to leave after they find they don't fit with the company(usually within shorter time frame), or for career development purposes after 3-5 years, and usually settle down if they stay with the company for more than 10 years.

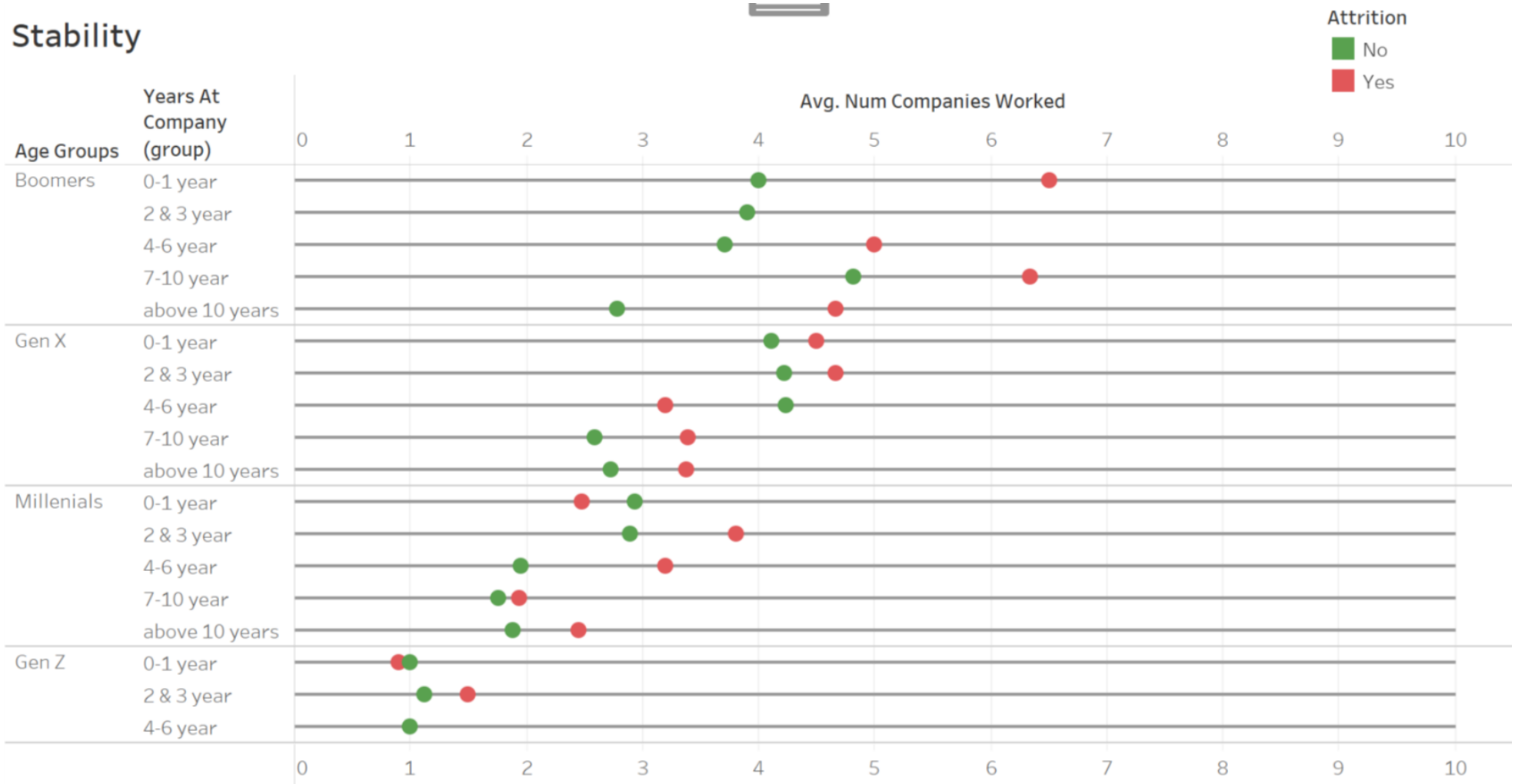
Attrition and Seniority



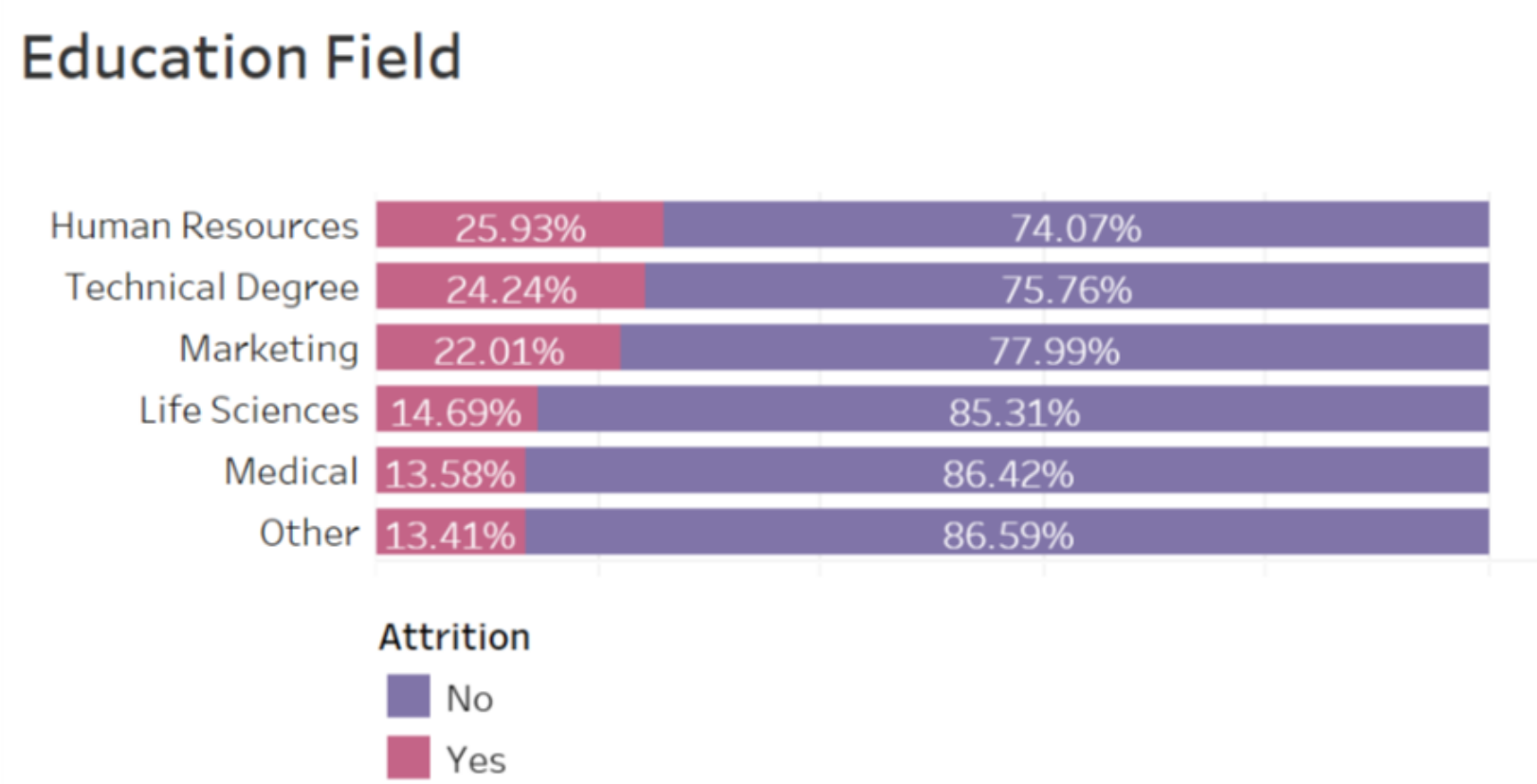
Attrition and Company Seniority



It is also observed that seniority and company seniority has a **significant** positive correlation with attrition, and number of companies worked for has a **significant** negative relation. P-values through **one-tailed t-test** are 1.2E-07, 1.2E-07, and 0.05 respectively.

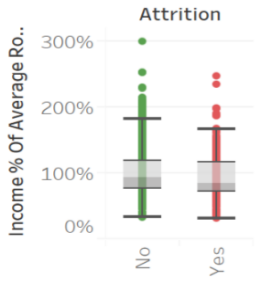


Education field also makes a difference, a possible explanation is that it could be related to external labor market demand and skill sets transferability.

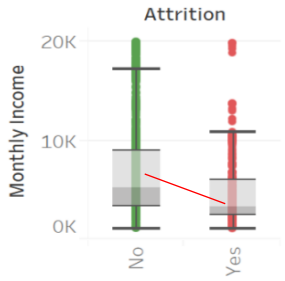


By looking at monthly income, it is observed that employees are not unhappy about their relative income level comparing to peers, but absolute income level – at an monthly income level of less than 5k, the people who left have a significant lower level of income distribution. Combining to previously examination into performance and the last graph, it shows that when people have a lower income and their performance is not incentivized by monetary rewards, they leave the company, either voluntarily or involuntarily.

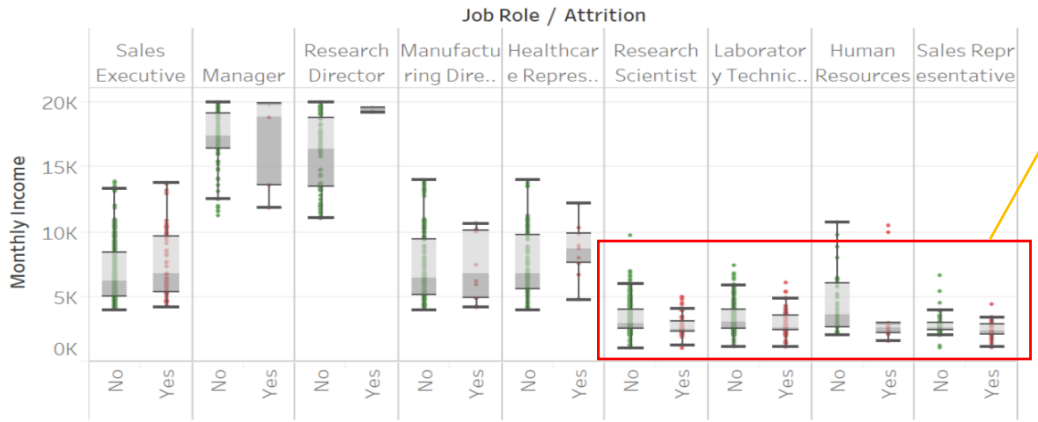
Effect of Salary Comparing to Average Salary by Postion



Effect of Salary Comparing to Overall Salary

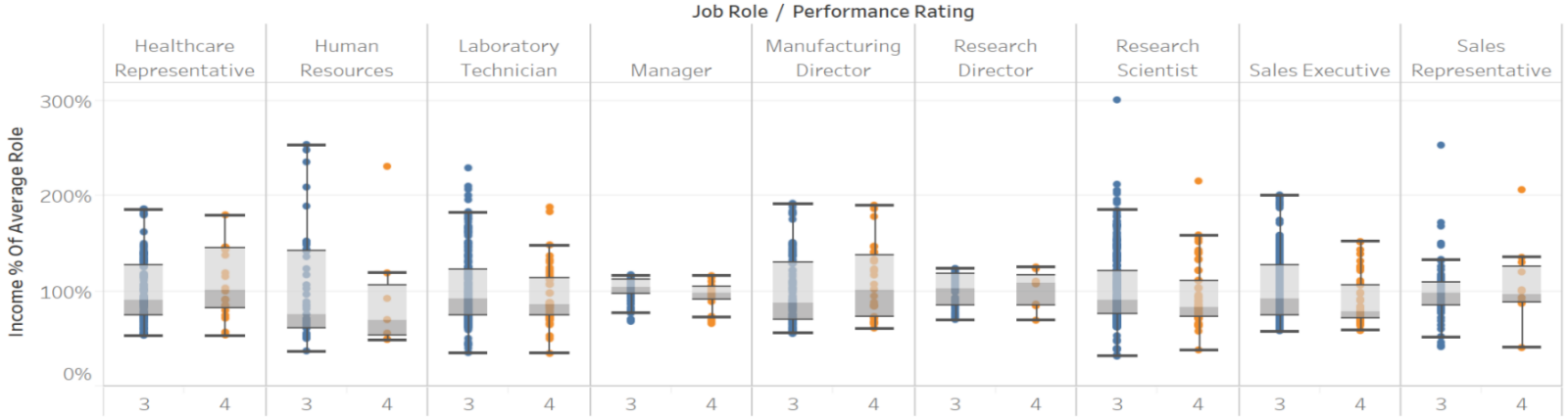


Who left because of lower salary



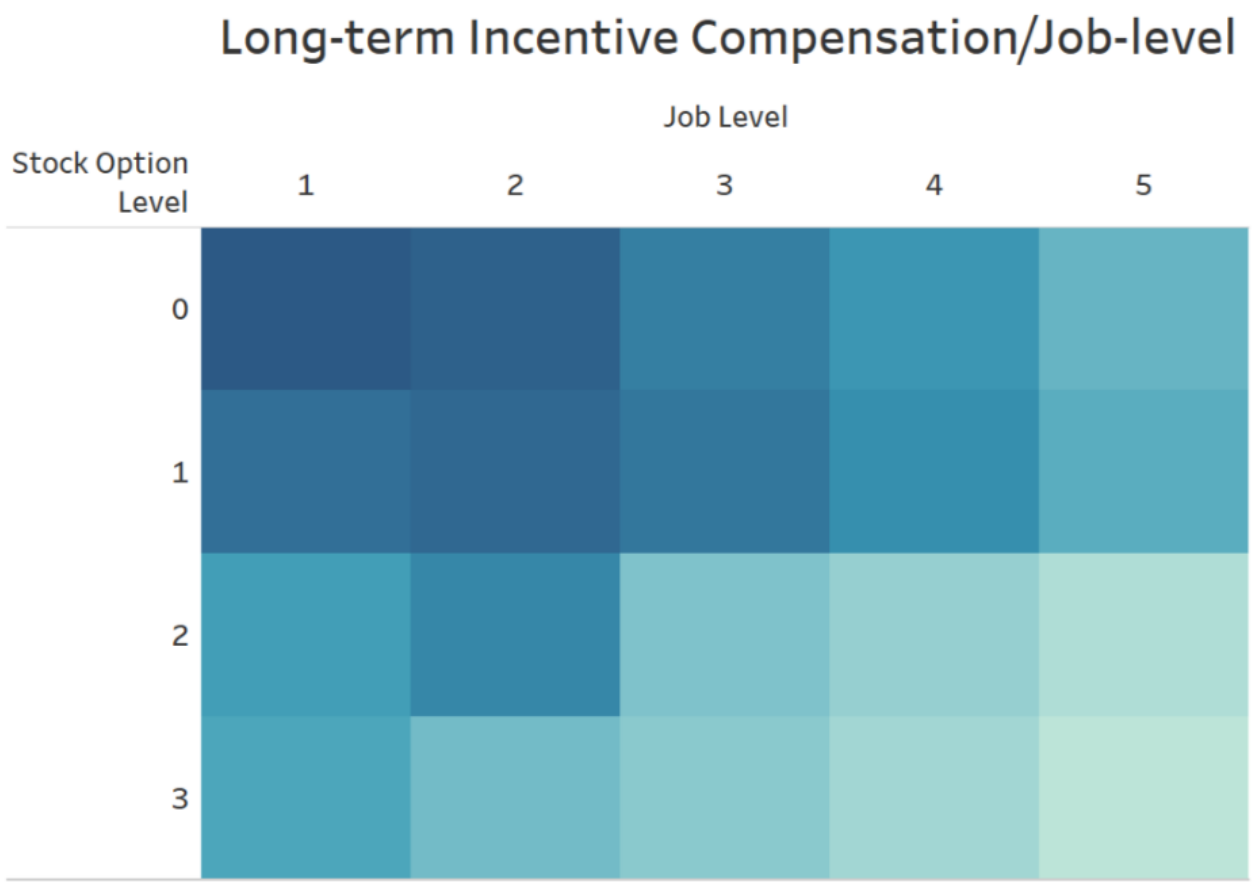
Income differences exist here

Performance is not Incentivized by Income

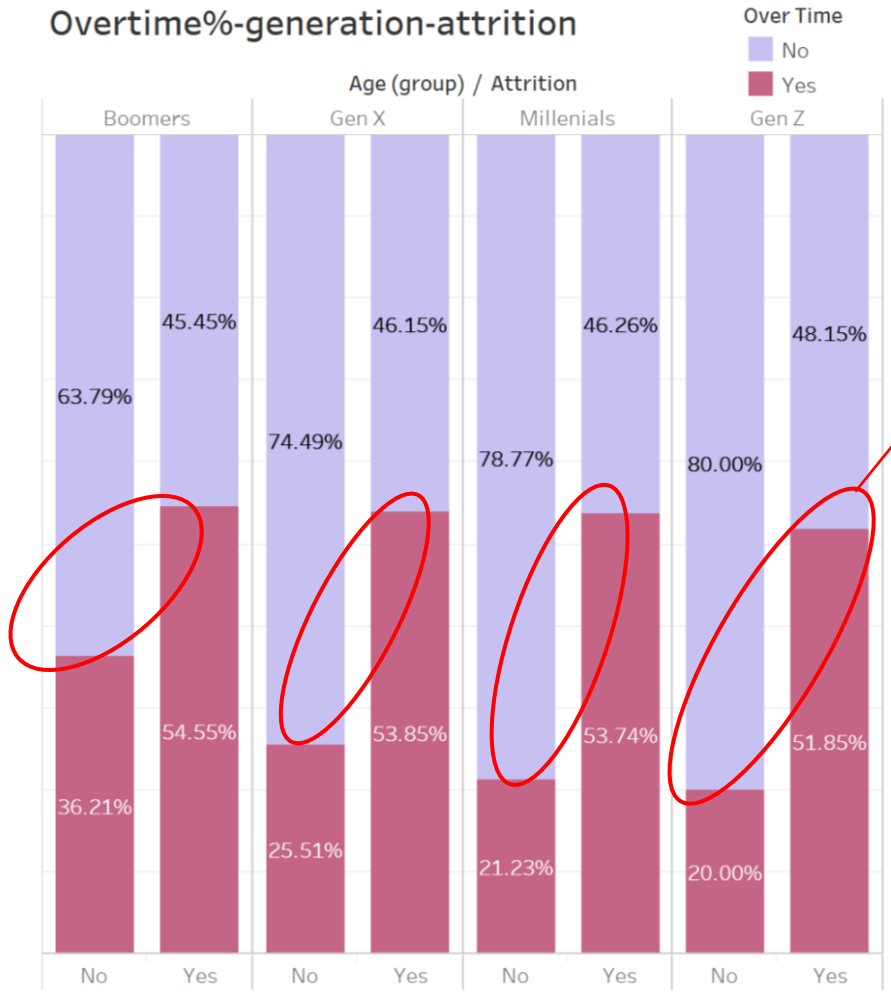


Data Exploration | **Hypothesis Verification** | **Next Steps & Action Plan**

In this graph, the darker the color, the more account of attrition there is. This shows people tend to leave when they are at lower job level and not incentivized by long-term incentive. This could prove the effectiveness of long-term talent retain of stock options.



Millennials and Gen X could also place high importance on whether they need to work overtime or not.



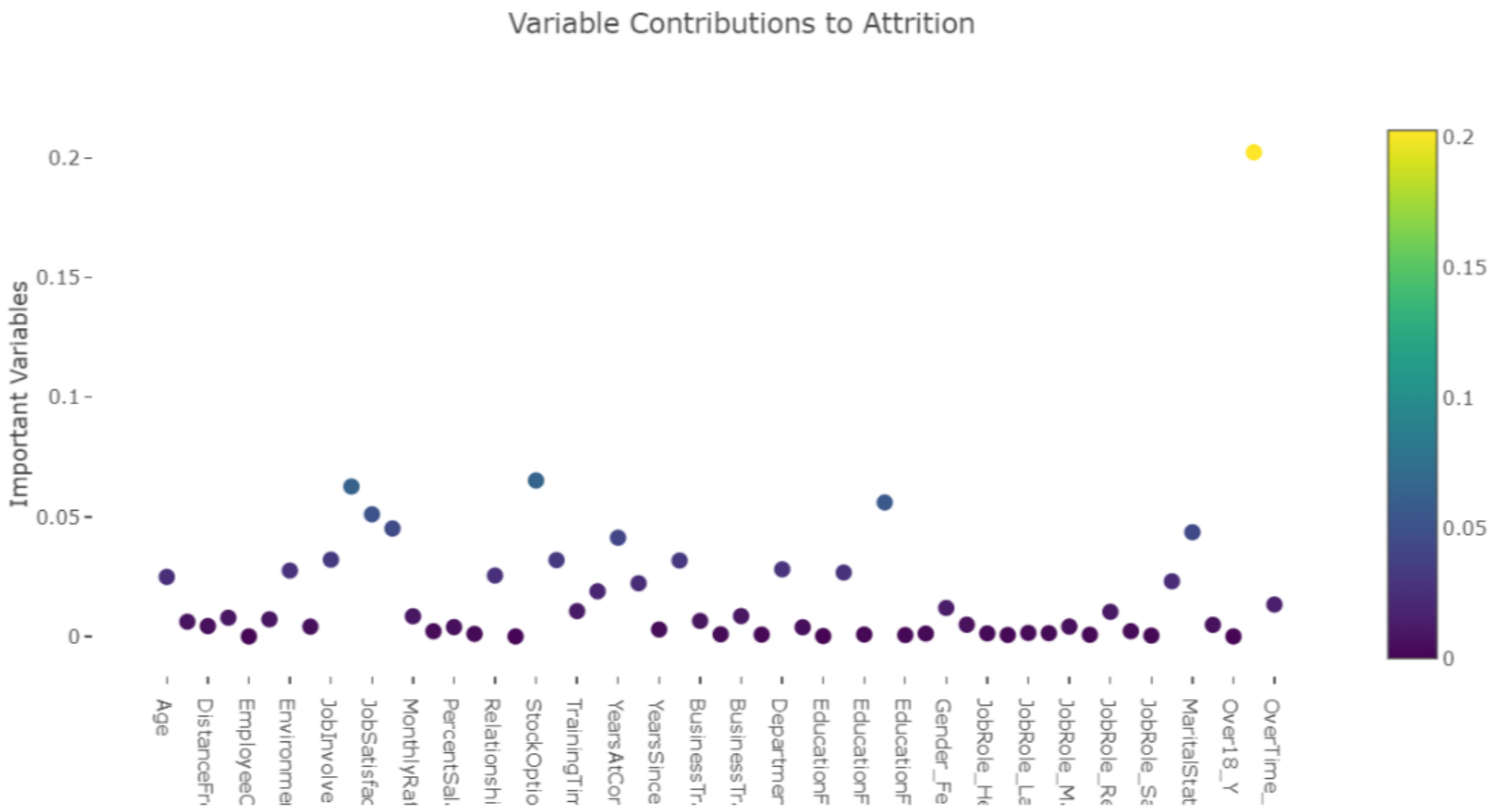
The younger the generation, the less tolerable they are to overtime

Generally the people who left have less fulfillment, happiness, and career development opportunities.

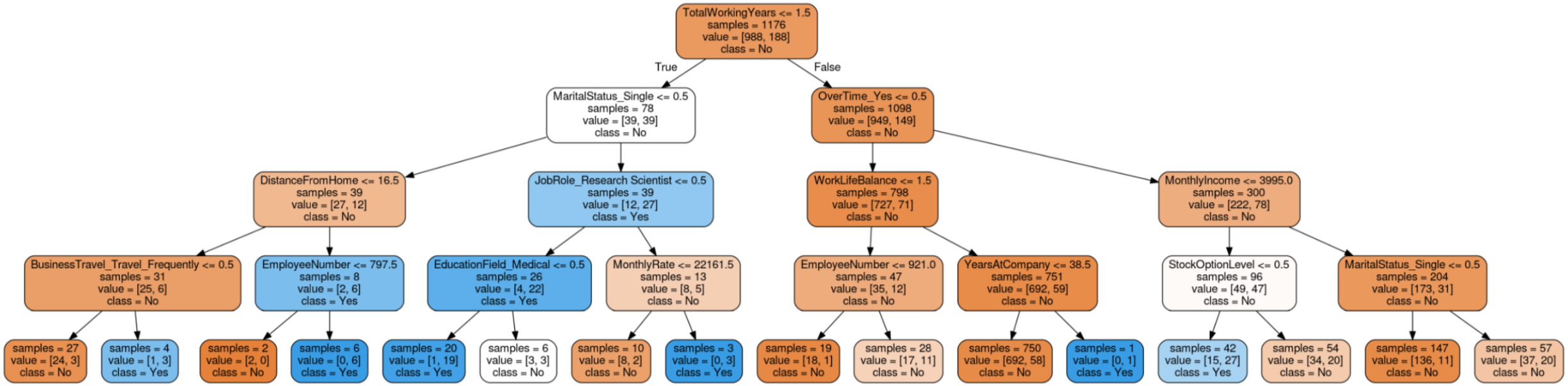
Data Exploration | **Hypothesis Verification** | Next Steps & Action Plan

To verify the above findings and hypothesis, a machine learning model was built to both examine the most important variables contributing to attrition and also predict future attritions according to those variables. A random forest model was used here.

The model evaluates and verifies that the most contributing variable is overtime, followed by Job Satisfaction, Monthly income, Age, Stock Option, Education Field. It also provides additional insights in variables such as Marital Status and Distance from Home for further analysis.



This chart give specific importance ranking and related weighting.



3.

Next Steps & Action Plan

Action Plan

From above observation, hypothesis, testing and confirmation, the following are proposals for future actions:

- For **short term**:
  - **Predictions** from the Machine Learning model can be used as input for workforce planning.
  - A **scoring and flagging system** can be built for proactive measures. Responsible HRBP can be alerted of employees above flagging score， then may conduct stay interview or take other appropriate measures accordingly.

- For **middle to long term**:
  - A more **systematic change management** approach can be taken that incorporates considerations including performance management, compensation adjustment, promotion, and intangible benefit. Because these would have a lot of intricacies in implementation, they need to be more carefully planned, managed, and closely monitored on effectiveness.
  - A department with higher attrition （e.g. Sales 20.6%） may conduct **pilot program** first, and if effective, it can be extended to other departments.

Next Steps

For the next steps, the following can be considered according to resource budget and priority:

- Use the quantitative result to drill down and get more **qualitative information** beyond numbers, e.g. stay interview, employee engagement survey with targeted questions.

- **Further analyze** the data set in combination of data resources and other measurements, e.g. the healthiness of attrition (voluntary/involuntary), time and trend, geographic locations, manager name. Possible insights may be generated to use as input to arbitrage local team size, compete for talent with competitors, and provide managers leadership and communication trainings if needed.

- Support the development of attrition **scoring and flagging system**.

- Explore the influence of **other easily overlooked forces** such as causal relationship, interdependency, and Matthew Effect etc..

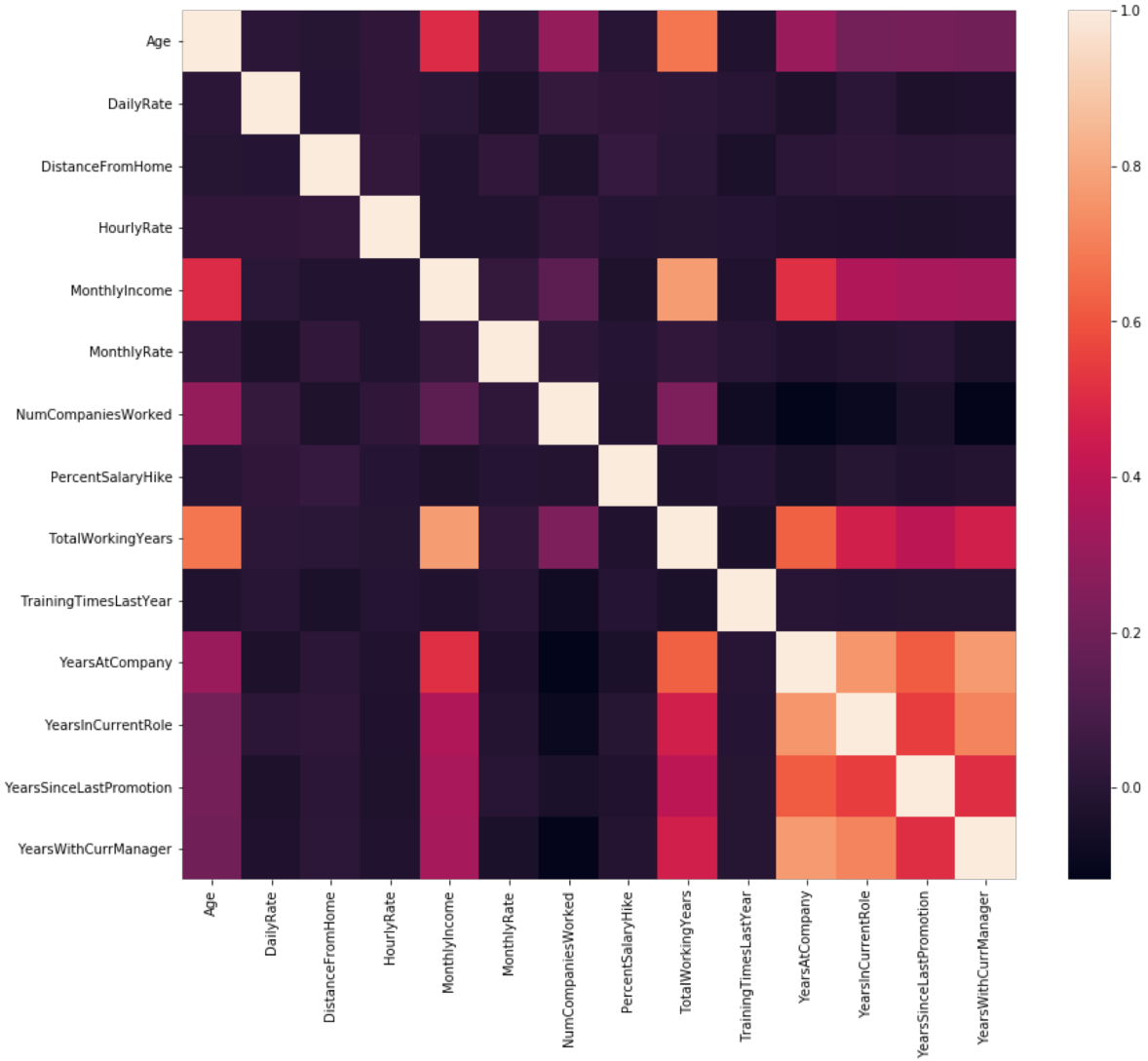
THANK YOU

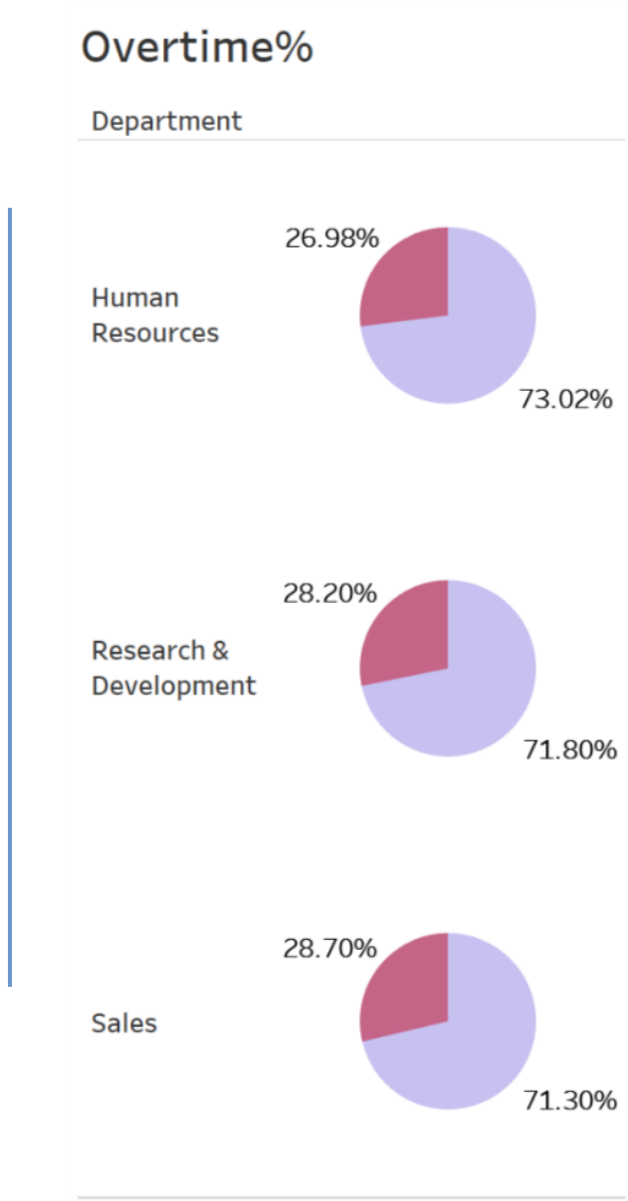
APPENDIX

- other visualizations that are not the main scope of the discussion here but may may provide further insights.

Income Difference

Correlation between variables




Overtime%
Department
Human Resources
26.98%
73.02%
Research & Development
28.20%
71.80%
Sales
28.70%
71.30%

With chi-square, the observation is also verified that attrition is not significantly different by gender or performance, which is a good thing – again, it means attrition is not caused by gender bias, and we didn't lose good performers.

Observed Attrition Distribution			
	Male	Female	Total
Yes	150	87	237
No	732	501	1233
Total	882	588	1470

Observed Attrition Distribution			
	3	4	Total
Yes	200	37	237
No	1044	189	1233
Total	1244	226	1470

Expected Attrition Distribution			
	Male	Female	Total
Yes	142	95	237
No	740	493	1233
Total	882	588	1470

Expected Attrition Distribution			
	3	4	Total
Yes	201	36	237
No	1043	190	1233
Total	1244	226	1470

P (Chi Square Test): **0.25879**

P (Chi Square Test): **0.91181**