Employee Loyalty Insights: Statistical Analysis with Python

Problem Statement

Should Employees Invest Their Entire Careers in the Company? Evaluating the company's benefits and rewards of loyalty

The main goal of this project is to evaluate how effective our current compensation and benefits system is in rewarding employees who commit to the company for the long period of time. Firstly, we want to find out if loyal employees receive extra perks like better salaries. Secondly, we aim to explore how loyalty affects career progression and lastly, whether our compensation remains fair throughout different stages of an employee's career. By investigating these areas, we hope to provide insights from the perspective of someone considering a long-term career with the company. These insights will help our BOD makes informed decisions to enhance our retention strategies, ensuring that our organization not only attracts but also retains dedicated and high-performing employees. We believe that when promoting long-term commitment, our company will benefit from stability, deep expertise, strong cultural values, and a lasting competitive edge.

SUB-PROBLEM 1

Do loyal employees receive additional benefits?

Determine whether employees who stay with us longer receive extra benefits. If long standing employees are rewarded with additional benefits, it can encourage others to remain with the company.

SUB-PROBLEM 2

Is loyalty asssociated with better career progression within the company?

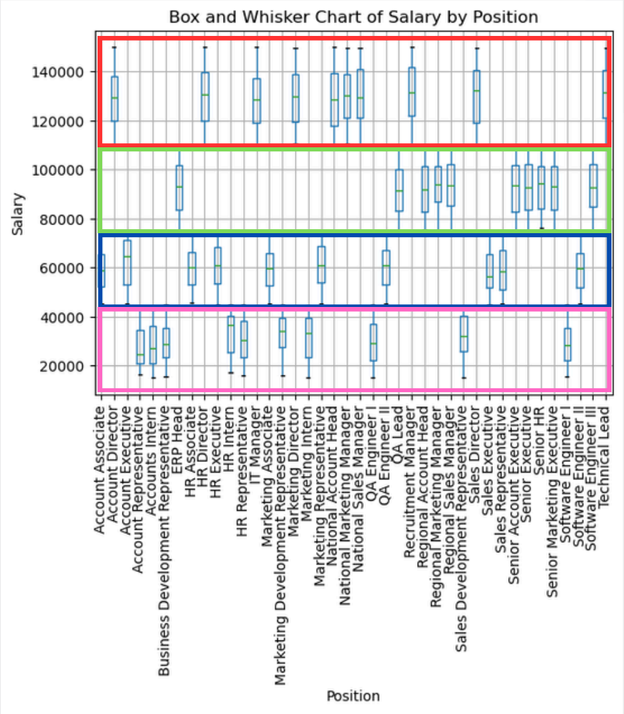
Investigate whether employees who demonstrate long-term commitment experience more favorable career trajectories, including promotions/ leadership roles compared to employees hired externally.

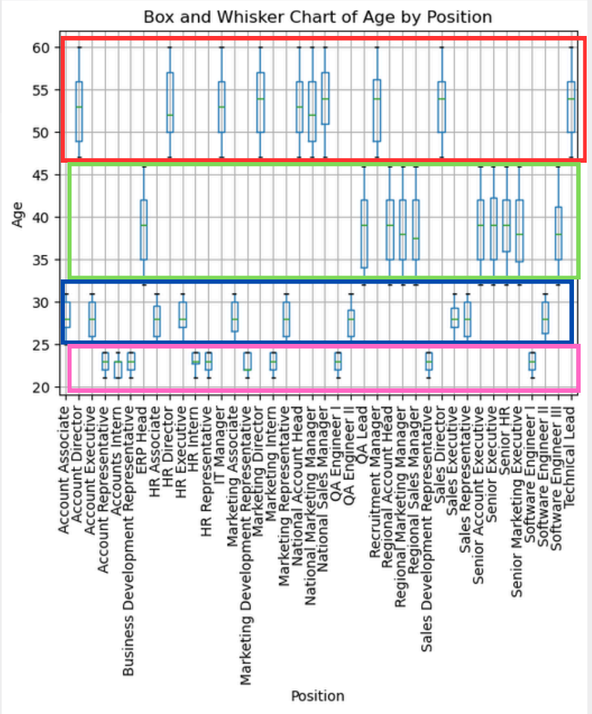
SUB-PROBLEM 3

Is compensation fair throughout all stages of employees' careers?

Assess whether compensation are equitable across department at every stage of an employee's career. Ensuring fairness helps prevent turnover, saving time and resources on recruiting/ training new hires.

Summary statistics





From our EDA, we have discovered that the Positions can be split into 4 distinct Position Levels, they are:

Position Level	Positions	Age Range
1 (most junior)	Account Rep/Intern, HR Rep/Intern, Marketing Dev Rep/Intern, QA/Software Engineer I, Business/Sales Dev Rep	21 - 24
2	Account Associate/Executive, HR Associate/Executive, Marketing Associate/Executive, QA/Software Engineer II, Sales Rep/Executive	25 - 31
3	Senior Account Executive/Regional Account Head, Senior HR/ERP Head, QA Lead/Software Engineer III, Regional Marketing Manager/Senior Marketing Exec, Regional Sales Manager/Senior Exec	32 - 46
4 (most senior)	Account Director/National Account Head, HR Director/Recruitment Manager, IT Manager/Tech Lead, Marketing Director/National Marketing Manager, National Sales Manager/Sales Director	46 and above

Other Findings from EDA

- Within each Position Level, Age Range is the same regardless of Department or Position.

Newly Created variables

- Position Level
  - Since the Age and Positions are binned into 4 categories, we have binned the Position Levels based on the Positions as shown in Table.
  - Position Level will go from most junior at Level 1 to Level 4.
- Insurance
  - NaN was transformed to str 'No' to enables analysis of no insurance holders.
- Medical Insurance
  - Performed One-Hot Encoding for Employees with Medical Insurance or both Insurance
- Life Insurance
  - Performed One-Hot Encoding for Employees with Life Insurance or both Insurance
- Age of Joining
  - Measures the Age of employee prior to joining the company
  - Age of Joining = Age - In Company Years
- Long Term Employee
  - Checks if Year of Experience = In Company Years
  - Value of 1 denotes that Employee is a Long Term Employee (LTE)
- First Role?
  - There is no data on whether the employee has been promoted, however
  - Based on the variable analysis from the box and whisker chart that showcases that Age is a discrete variable with a fixed range in relation to the Position Level, we can deduce that if the employee has been promoted from their Age of Joining
  - If the employee's Age of Joining is not within (or in fact lower than) the age range of his current Position Level, it can be deduced that he could have been promoted to the next Position Level. As such, their current position would not be their First Role.
  - As such, the only way that he could derive at that certain Position Level is **through an internal promotion in the company**. As such, his First Role? value will be 0.
  - It is also taken that First Role? for Position Level 1 (the most junior positions) will have the value of 1.

Descriptive Analysis on Compensation for Loyal Employees

Refer to notebook 01.ipynb

What are the distribution of compensation between long term employee and non-long term employees? What parts of the compensation are noticable?

Salary only

Gross salary

Insight

- From the 2 graphs, long-term employees appear to receive similar base salaries compared to non-long-term employees across all position levels, suggesting that there may be no additional pay specifically for loyalty.
- However, at lower levels, such as Position Levels 1 and 2, gross salaries for long-term employees are slightly higher than those of their non-long-term counterparts. This discrepancy may indicate the presence of supplementary benefits, such as Housing allowances, Cost of Living Adjustments (COLA), or contributions to the Central Provident Fund (CPF) provided to loyal employees.
- To determine whether the compensation for long-term employees is different from that of non-long-term employees, we investigated the gross salaries variable for each position level.

Inferential analysis on Compensation for long term employees

Refer to notebook 01.ipynb

Is long term employment a factor that can significantly influence individual salaries and housing allowances benefits?

Mean Gross Salary for Long-Term Employees (LTE) and Non-Long-Term Employees (NLTE) at each Position Level

Position level	Mean Gross Salary (LTE)	Mean Gross Salary (NLTE)	Sample Size (LTE)	Sample size (NLTE)
1	40,761.4	39,747.41	196	226
2	72,890.66	72,673.99	124	730
3	110,210.39	108,877.15	77	1604
4	152,166.70	149,901.83	45	1598

90% CI for Gross Salary for LTE and NLTE at each Position Level

Position level	Gross Salary 90% CI for LTE	Gross Salary 90% CI for NLTE
1	(39561.66, 41961.16)	(38669.58,40825.25)
2	(71273.77, 74507.56)	(72040.99, 73307.01)
3	(107727.39, 112693.40)	(108372.61, 109381.7)
4	(149390.71, 154942.7)	(149426.57,150377.10)

Insight

- Position Level 1: LTEs seem to have a slight salary advantage at this level, as the lower and upper bounds of their CI are slightly higher than those of NLTEs.
- Position Level 2: The salary ranges for LTEs and NLTEs are quite close, with some overlap. This suggests that the impact of long-term employment on salary starts to diminish at this level.
- Position Levels 3 and 4: The salary ranges for LTEs and NLTEs largely overlap at these higher levels, indicating that factors other than employment status are more influential in determining salary.
- To further investigate whether the gross salaries for LTE and NLTE are significantly different, we carried out a statistical test in the following part.

Hypothesis Testing on Compensation for long term employees

Refer to notebook 01.ipynb

Are compensation between long term employee and non-long term employees significantly different?

Hypothesis

- **Null Hypothesis (H_0):** There is no differences in average gross salaries between LTE and NLTE in the same position level.
- **Alternative Hypothesis (H_1):** There is a significant difference in average gross salaries between LTE and NLTE in the same position level.

Position level	Mean Gross Salary (LTE)	Mean Gross Salary (NLTE)	Sample Size (LTE)	Sample size (NLTE)
1	40,764.1	39,747.41	196	226
2	72,890.66	72,673.99	124	730
3	110,210.39	108,877.15	77	1604
4	152,166.70	149,901.83	45	1598

- At 90% Confidence Level, reject Null Hypothesis if p-value ≤ 0.10
- Since σ is unknown and sample sizes n1>30 and n2>30 for all Position Levels, two-tailed t-test will be conducted.

Result

- Since p-value > 0.10 for all position level, we fail to reject the Null Hypothesis at 90% Confidence level.
- From that, we can conclude that there are no significant difference in salaries between LTE & NLT for each position level.

Position level	T-statistic	P-value	Mean Difference	Result
1	NaN	NaN	NaN	Not enough NLTE to perform t-test
2	0.7181	0.4729	$-517.49	Fail to reject null hypothesis
3	-0.1558	0.8763	$96.46	Fail to reject the null hypothesis
4	0.0212	0.9831	$-15.31	Fail to reject the null hypothesis.

- Across Position Levels 2, 3, and 4, there is insufficient evidence to say that Long-Term Employees (LTE) receive significantly better Gross Salaries compared to Non-Long-Term Employees (NLTE).

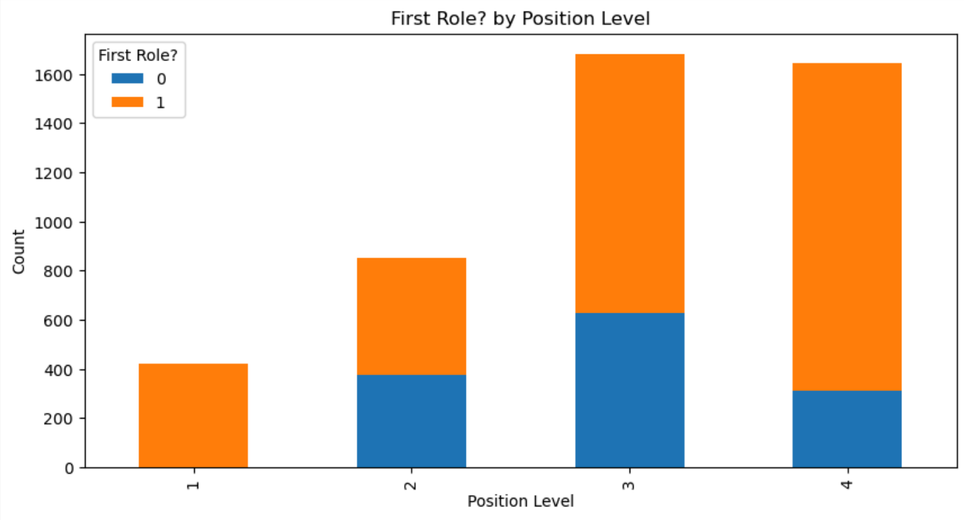
Insight

- The t-test results show no significant difference in gross salaries between long-term employees (LTE) and non-long-term employees (NLTE) at position levels 2, 3, and 4. From this, we can observe that employee loyalty does not translate into better compensations, including base salaries and other benefits.
- This finding could be a demotivating factor for loyal employees. They might feel less inclined to put in additional effort if they believe it won't be rewarded financially. Furthermore, demotivated LTEs might become more open to exploring opportunities outside the company, potentially leading to increased turnover. It is suggested that the company should create a more rewarding environment for long-term employees, increasing retention and fostering a stronger sense of loyalty within the workforce.

Descriptive Analysis on Career Progression

Refer to notebook 02.ipynb

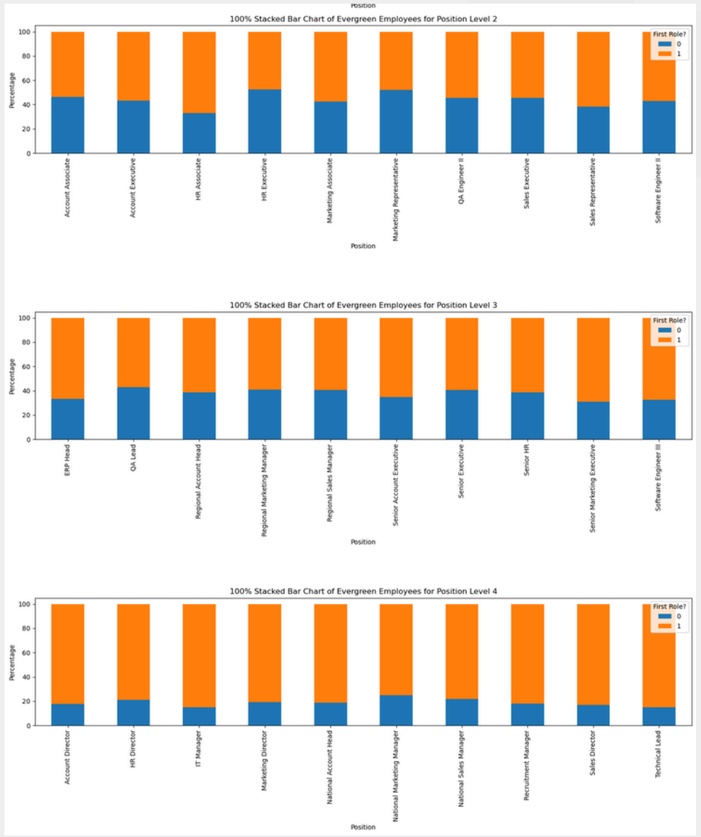
What is the likelihood that an employee would get promoted in the company? Does the company favour candidates externally or within their own organisation for senior positions?



	Position Level	Count of people not in First Posting	Count of people in First Posting	First Role?=0 (%)	Promotees from Previous Position Level (%)
0	1	0	422	0.000000	0.000000
1	2	376	478	44.028103	89.099526
2	3	628	1053	37.358715	73.536300
3	4	311	1332	18.928789	18.500892

We have identified that the number of promotions (First Role? = 0) towards Senior Positions in the company took a big dip, despite Senior Positions taking a huge majority of the Headcount.

Since we do not have promotion data, and with different Position Levels taking a fixed range of ages, we take that the employee was previously promoted if their Age of Joining < the minimum Age of their Position Level.



It is also identified that the dip is consistent for all roles, and no anomalies are identified.

Conclusion

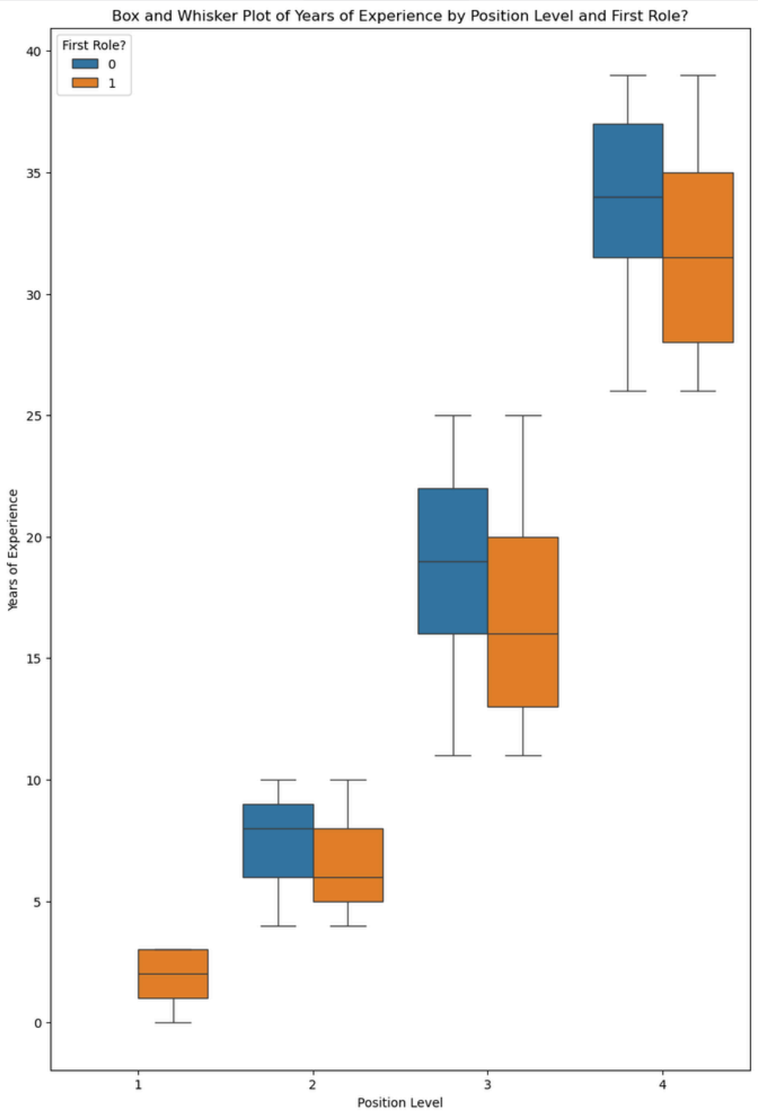
It seems that a higher proportion of Senior Positions (Position Level 4) are externally sourced, rather than internally promoted in. As such, it may seem that internal employees are less favoured for promotions to a higher Job Grade.

Inferential Analysis on Career Progression

Refer to notebook 02.ipynb

What could be some factors that make external candidates more favourable than internal employees? Are employees here truly being valued in the company?

Investigating Year of Experience



Internally Promoted Employees: Employees whose First Role?=0
Externally Sourced Employees: Employees whose First Role?=1, excluding those in Position Level 1

	Position Level	Mean Years of Experience (First Role?=0)	Mean Years of Experience (First Role?=1)	Sample Size (First Role?=0)	Sample Size (First Role?=1)
0	1	NaN	1.798578	0.0	422.0
1	2	7.672872	6.550209	376.0	478.0
2	3	19.114650	16.628680	628.0	1053.0
3	4	33.996785	31.769520	311.0	1332.0

By looking at the means of Year of Experience on internal employees promoted (First Role?=0) and employees sourced externally (First Role?=1), we can see that generally, the means of those internally promoted have greater Years of Experience than employees externally sourced.

	Position Level	Year of Experience Confidence Interval
0	1	(nan, nan)
1	2	(7.51894136462406, 7.8268033162270045)
2	3	(18.869625812727406, 19.359673550329916)
3	4	(33.68583061205451, 34.30773851977828)

This table shows the confidence interval of Year of Experience for Employees who are not currently in their First Posting in the company, which simulates as internal employees promoted, at 90% confidence level.

	Position Level	Year of Experience Confidence Interval
0	1	(1.711838903014541, 1.8853174950897247)
1	2	(6.404581374784053, 6.695837035257789)
2	3	(16.416308866076413, 16.841051057950175)
3	4	(31.59496466471264, 31.944074374326398)

This table shows the confidence interval of Year of Experience for Employees who are currently in their First Posting in the company, which simulates as external candidates, at 90% confidence level.

We can observe that the Year of Experience for internal employees promoted may appear to have slightly similar professional experience than externally sourced candidates. **We should conduct hypothesis testing to observe if it is statistically true.**

However, this tells us that Externally Sourced Employees do not have a hidden advantage due to greater professional experience than Internally Promoted Employees.

Hypothesis Testing on Career Progression

Refer to notebook 02.ipynb

Hypothesis:

Internally Promoted Employees have longer professional experience than Externally Sourced Employees

For every Position Level (except Position Level 1):

Null Hypothesis (H_0):
μ(Years Of Experience of Internally Promoted Employee) ≤ μ(Years of Experience of Externally Sourced Employee)

Alternative Hypothesis (H_1):
μ(Years Of Experience of Internally Promoted Employee) > μ(Years of Experience of Externally Sourced Employee)

At 90% Confidence Level, reject Null Hypothesis if p ≤ 0.10

	Position Level	Mean Years of Experience (First Role?=0)	Mean Years of Experience (First Role?=1)	Sample Size (First Role?=0)	Sample Size (First Role?=1)
0	1	NaN	1.798578	0.0	422.0
1	2	7.672872	6.550209	376.0	478.0
2	3	19.114650	16.628680	628.0	1053.0
3	4	33.996785	31.769520	311.0	1332.0

Since σ is unknown and sample sizes n1>30 and n2>30 for all Position Levels, we will use the **Independent t-test**

Code:

```python
from scipy.stats import ttest_ind

# Initialize a list to store the results
t_test_results = []

for level in employee_df['Position Level'].unique():
    # Separate the data into two groups based on 'First Role?' for the current position level
    group_0 = employee_df[(employee_df['Position Level'] == level) & (employee_df['First Role?'] == 0)]['Year of Experience']
    group_1 = employee_df[(employee_df['Position Level'] == level) & (employee_df['First Role?'] == 1)]['Year of Experience']

    # Perform the one-tailed t-test
    t_stat, p_value = ttest_ind(group_0, group_1, alternative='greater', equal_var=False)

    # Store the results
    t_test_results.append({
        'Position Level': level,
        'T-statistic': t_stat,
        'P-value': p_value,
        'Mean Difference': group_0.mean() - group_1.mean()
    })

t_test_results_df = pd.DataFrame(t_test_results)
t_test_results_df.sort_values('Position Level')
```

✓ 0.0s Python

Results:

	Position Level	T-statistic	P-value	Mean Difference
2	1	NaN	NaN	NaN
3	2	8.733904	6.752186e-18	1.122663
0	3	12.626228	4.874234e-35	2.485970
1	4	10.298673	4.429835e-23	2.227265

Position Level 2:

Since p-value < 0.10, we reject the Null Hypothesis at 90% Confidence Level. As such, we can conclude that for Position Level 2, the Year of Experience of Internally Promoted Employees is statistically greater than Year of Experience of Externally Sourced Employees

Position Level 3:

Since p-value < 0.10, we reject the Null Hypothesis at 90% Confidence Level. As such, we can conclude that for Position Level 3, the Year of Experience of Internally Promoted Employees is statistically greater than Year of Experience of Externally Sourced Employees

Position Level 4:

Since p-value < 0.10, we reject the Null Hypothesis at 90% Confidence Level. As such, we can conclude that for Position Level 4, the Year of Experience of Internally Promoted Employees is statistically greater than Year of Experience of Externally Sourced Employees

Conclusion

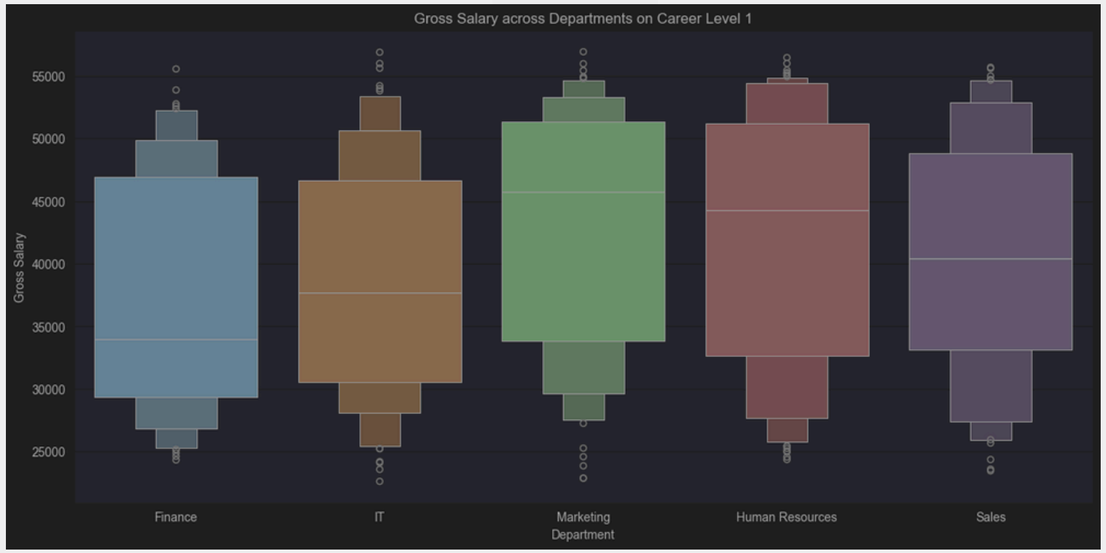
With the 3 results, we can observe Internally Promoted Employees are in fact more experienced as Externally Sourced Employees. As such, there is no added advantage of hiring Externally Sourced Employees in term of Professional Experience.

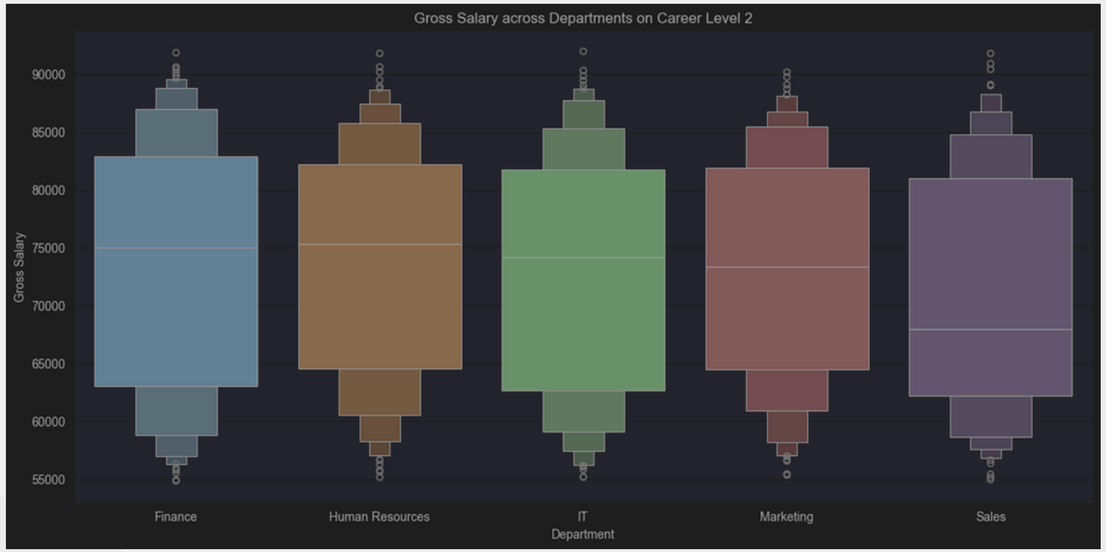
Additionally, it may deduce that staying in the company in the long run does not bring about greater career prospects, as externally sourced employees are able to assume the same roles at a lower age. This means it takes longer time for internal employees to receive the same prospects as employees sourced externally. Employees would not have an incentive to staying in the company in the long run.

Analysis on employee equality throughout the career stages

Refer to notebook 03.ipynb

Are employees treated equally across departments, career levels and personal factors?





Descriptive Analysis

The analysis has shown that there is a difference in the earned gross salary between the different departments on career level 1 & 2. Specifically the finance department on level 1 and the sales department on level 2 seem to earn less than the other departments. On the other hand are Marketing and HR on level 1 apparently higher paid.

	Finance	IT	Marketing	HR	Sales
mean level 1	37,196	38,972	42,267	42,760	40,233
mean level 2	73,095	72,457	73,136	73,482	71,112

Confidence Intervals for Gross Salary across Departments

	Department	Gross Salary Confidence Interval Career Level 1
0	Finance	(35392.75796386149, 39000.53034417858)
1	Human Resources	(40056.40679903136, 43464.231433028646)
2	IT	(37282.69118791908, 40659.91177388414)
3	Marketing	(40531.841683841754, 44002.68375079723)
4	Sales	(38211.8849354414, 42255.49323280825)

	Department	Gross Salary Confidence Interval Career Level 2
0	Finance	(71710.90057587781, 74480.76837381034)
1	Human Resources	(72275.57346817366, 74688.73407062037)
2	IT	(71105.00763613202, 73810.60198413902)
3	Marketing	(71857.89552342764, 74414.15282500081)
4	Sales	(69715.34043423017, 72510.45553069963)

Hypothesis testing on employee equality throughout the career stages

Refer to notebook 03.ipynb

Employees seem to receive different compensation for similar position in different departments

Hypothesis Test on Finance Department Career Level 1

Null Hypothesis (H_0): the finance department has the same mean gross salary as all other departments (μf = μd).
Alternative Hypothesis (H_1): the finance department has a different mean gross salary (μf != μd).
Procedure: Since σ is unknown and sample sizes n > 30 for all groups, we will use the independent t-test
T-Test-Result:
T-statistic: -3.04893, P-Value: 0.00285
Since p-value <= 0.10 we reject the Null-Hypothesis. There is sufficient evidence to conclude that μf != μd.

Hypothesis Test on Marketing & HR Department Career Level 1

Null Hypothesis (H_0): the marketing and HR department have the same mean gross salary as all other departments (μmh = μd).
Alternative Hypothesis (H_1): the marketing and HR department have a different mean gross salary (μmh != μd).
Procedure: Since σ is unknown and sample sizes n > 30 for all groups, we will use the independent t-test
T-Test-Result:
T-statistic: 3.35515, P-Value: 0.00087
Since p-value <= 0.10 we reject the Null-Hypothesis. There is sufficient evidence to conclude that μmh != μd.

Hypothesis Test on Sales Department Career Level 2

Null Hypothesis (H_0): the sales department has the same mean gross salary as all other departments (μs = μd).
Alternative Hypothesis (H_1): the sales department has a different mean gross salary (μs != μd).
Procedure: Since σ is unknown and sample sizes n > 30 for all groups, we will use the independent t-test
T-Test-Result:
T-statistic: -2.09104, P-Value: 0.03763
Since p-value <= 0.10 we reject the Null-Hypothesis. There is sufficient evidence to conclude that μs != μd.

Conclusion

The statistically significant inequality in the gross salary received in different departments of the company is an interesting discovery. Especially since there is not one department that is significantly different through all stages but different ones on each career level.

Should this phenomenon prevail, it could lead to rising employee dissatisfaction due to an unjust compensation and with that rising jealousy between departments.

Executive Summary

How can we support employee loyalty?

Increase benefits for loyal employees

The analysis has shown that there is no monetary incentive for employees to stay loyal to the company. Developing a compensation scheme that rewards long term employees can increase loyalty and overall satisfaction.

Ensure internal career opportunities

Positions throughout the company which are accounted to higher career levels are mainly taken by externally hired personnel. We propose to review our hiring procedures and develop an internal career program to support motivated employees.

Review fairness within the company

The analysis has shown that similar ranked positions are paid differently across different departments, especially in the lower career segments. This could increase jealousy and exasperation within our employees. Compensation should follow clear and transparent guidelines to prevent these emotions.