Employee Retention Analytics

Content

Business Case: Analysis of a Software company's Employee retention data

Data Acquisition: Dataset was obtained in a CSV format

Data Preparation: Using Azure ML, cleaned all missing data

Data Visualization: Descriptive Statistics using Tableau

Data Visualization - Tableau

Tableau is an interactive data visualization tool used for Exploratory Data Analysis (EDA), where charts/graphs are plotted for dimensions (qualitative values) against measures (quantitative values) and dependent variables (readmit30) to get insights and understand their data. Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their statistical characteristics, often with visual methods.

Tableau is quick, simple, user-friendly, intuitive, can handle lot of data, provide statistical calculations on datasets

EDA:

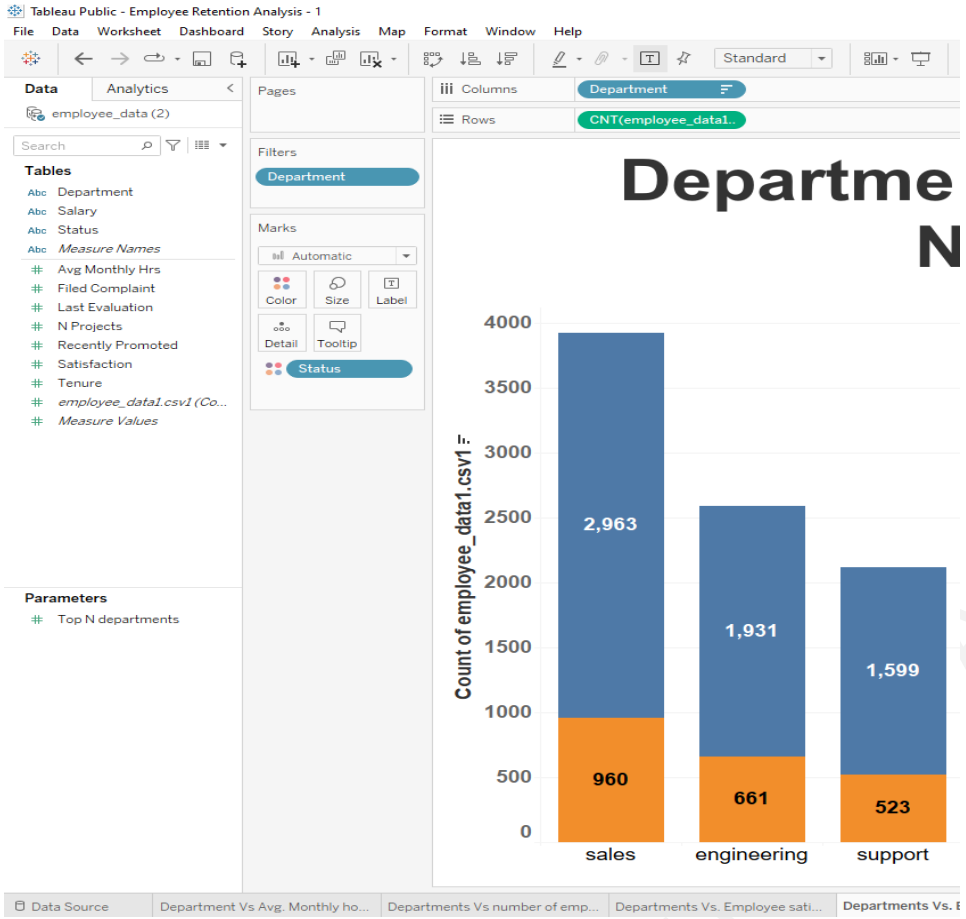
- Get a better understanding of data that may not be analyzed by standard data science algorithms.
- Understanding data patterns that may be skipped by typical machine learning algorithms.
- Drawing charts and graphs for better understanding from different angles and projects the results.
- To get a better understanding of the problem statement, visually.
- To find the hidden trends and relationship between variables.
- Assess and validate your assumptions on the variables, whether the variables help answer business problem or not.
- Screen for noise variables, missing data, outliers, etc. Find which variables need imputation, preprocessing

Tableau

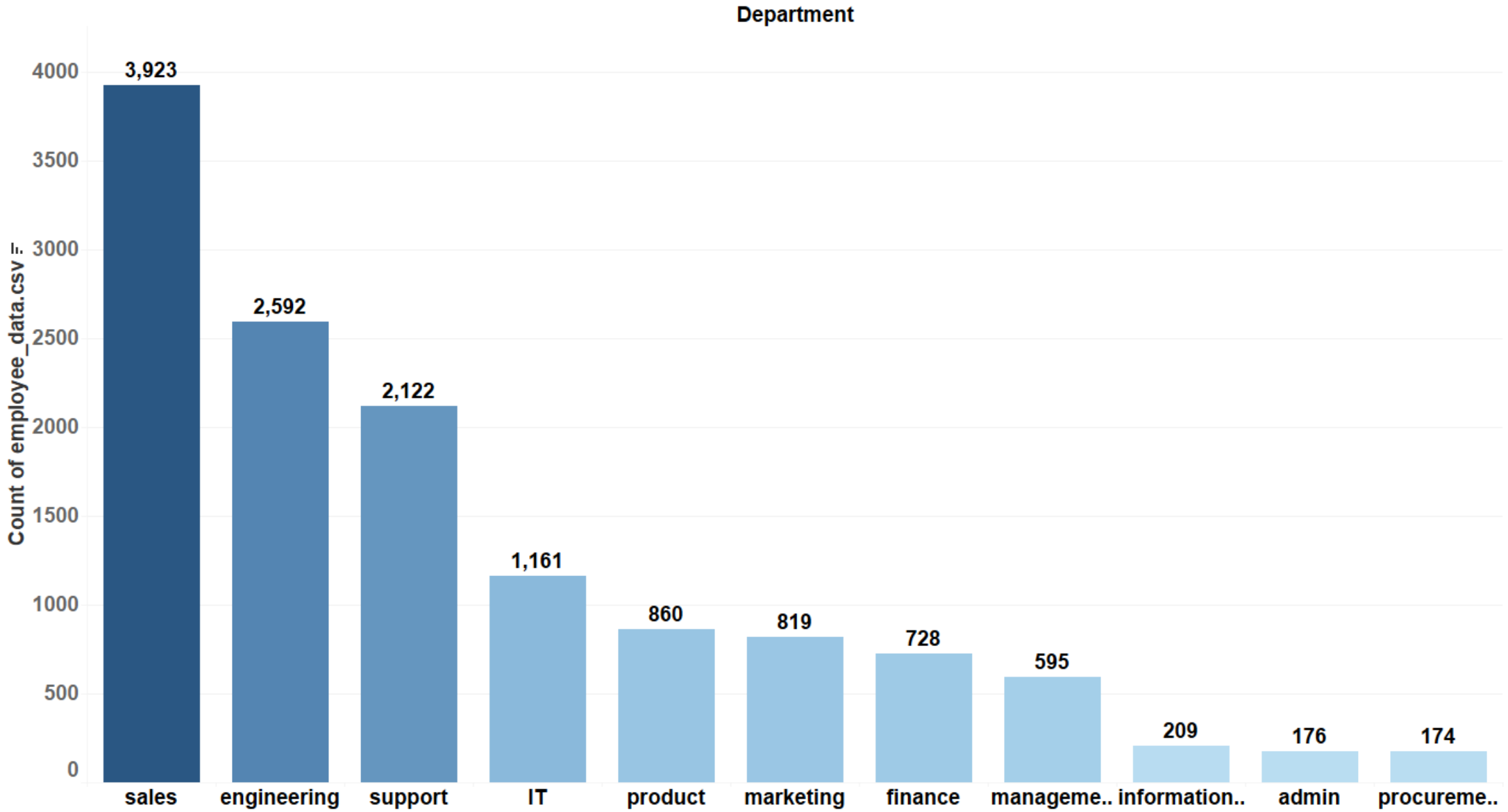
Chart Views

1. Text tables
2. Heat maps
3. Highlight tables
4. Symbol maps
5. Maps
6. Pie charts
7. Horizontal bars
8. Stacked bars
9. Side-by-side bars
10. Tree maps
11. Circle views
12. Side-by-side circles
13. Lines (continuous)
14. Lines (discrete)
15. Dual lines
16. Area charts (continuous)
17. Area charts (discrete)
18. Dual combination
19. Scatter plots
20. Histogram
21. Box and whisker plots
22. Gantt
23. Bullet graphs
24. Packed bubbles

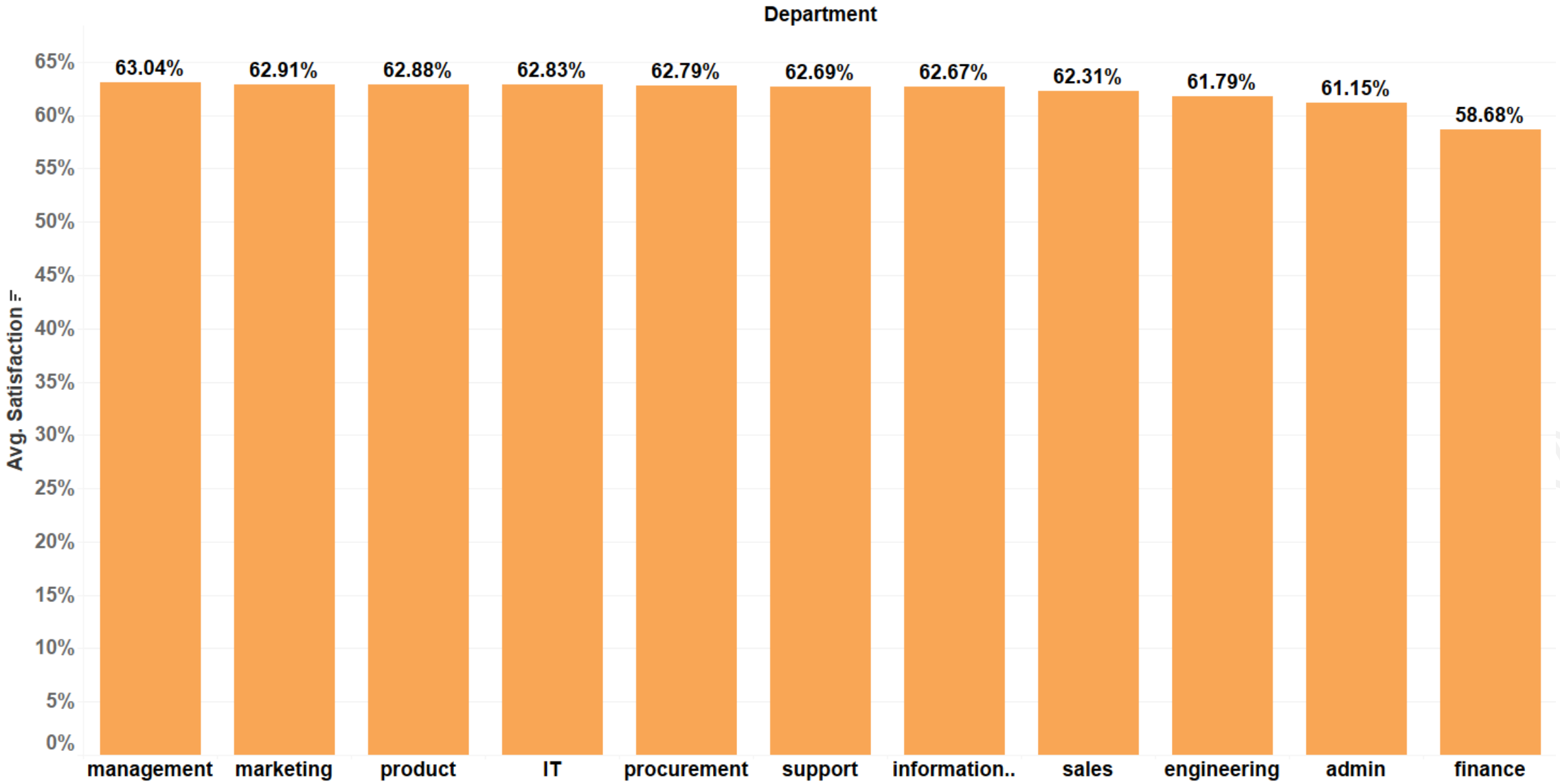
Data Pane, Marks card and Worksheet



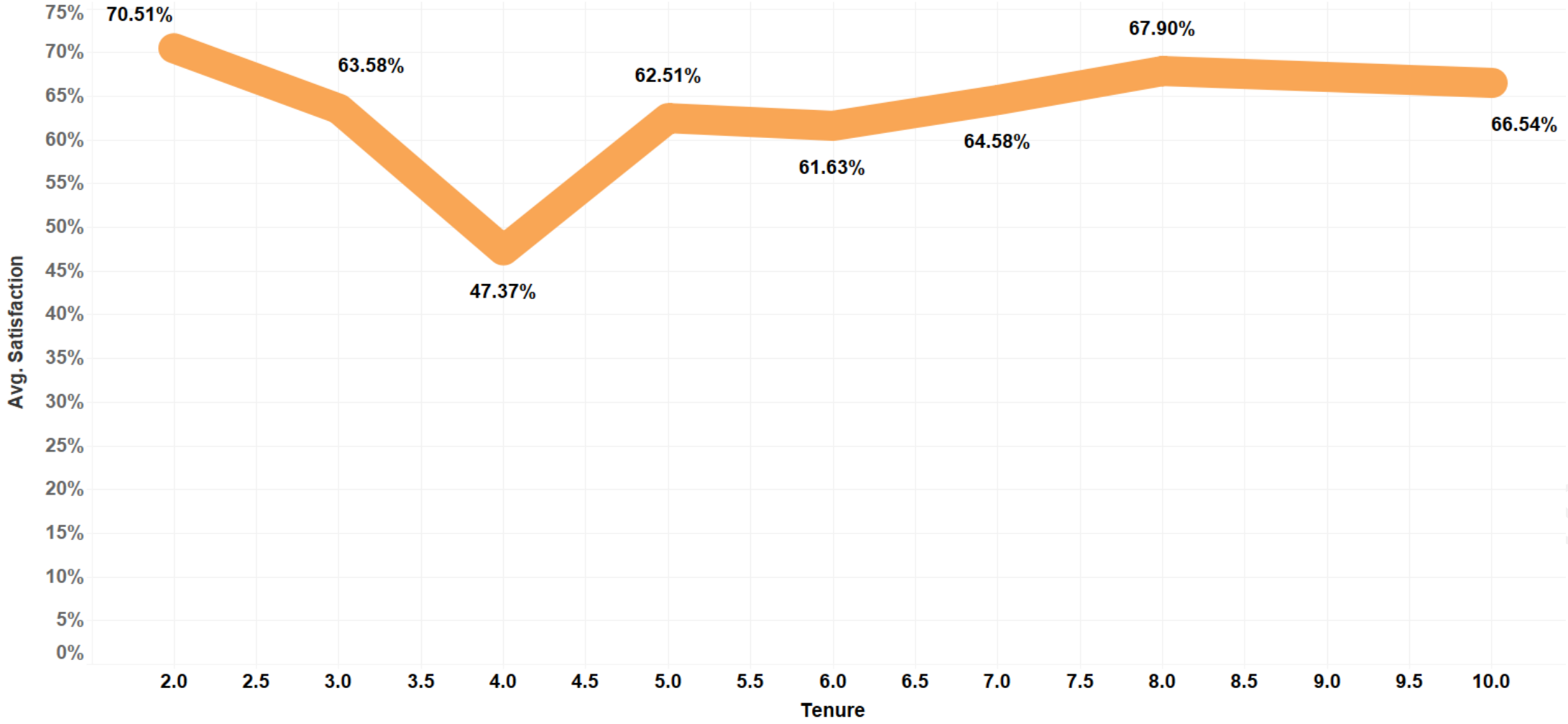
Departments Vs number of employees



Departments Vs Employee satisfaction

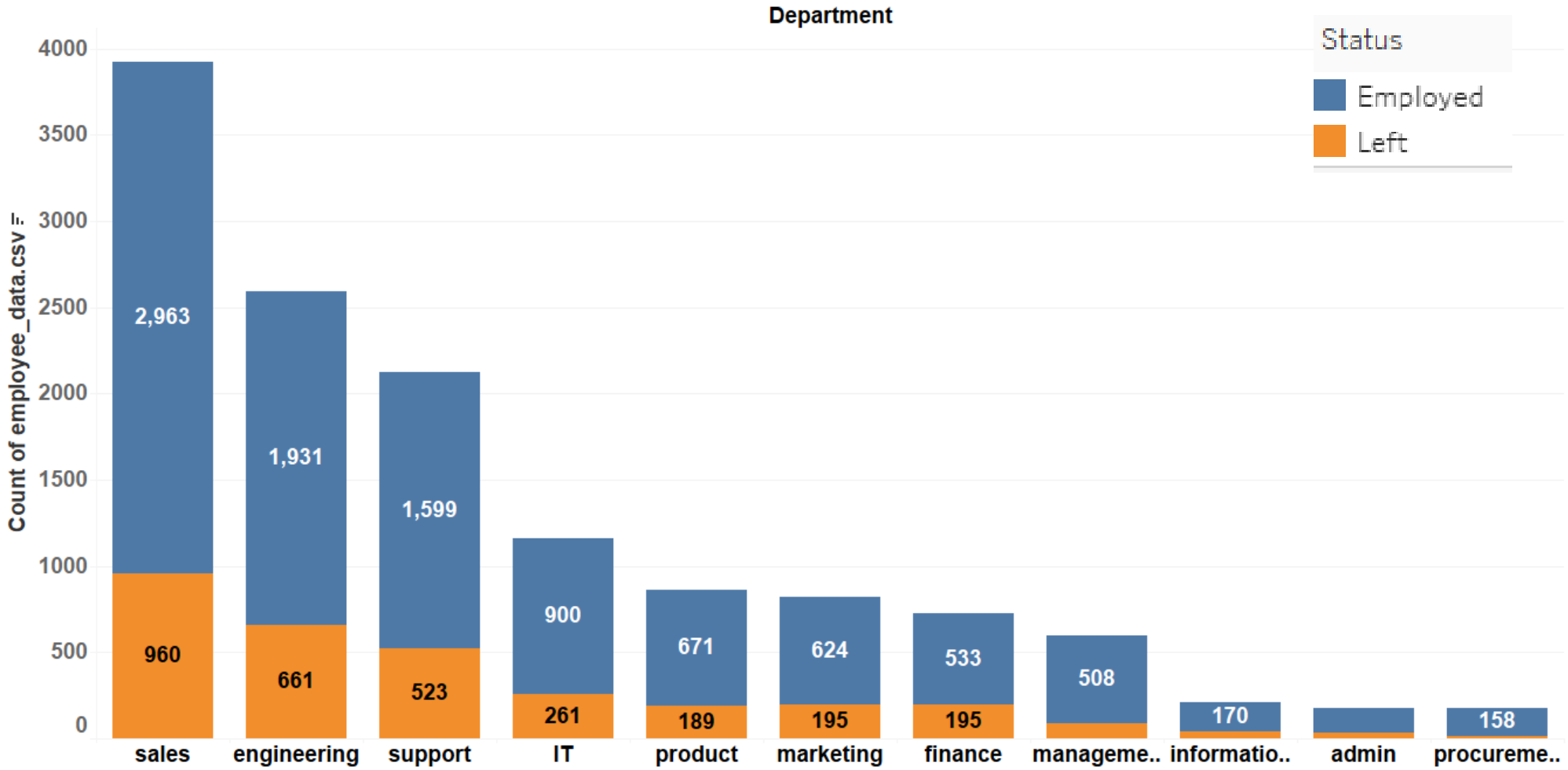


Tenure Vs Employee satisfaction



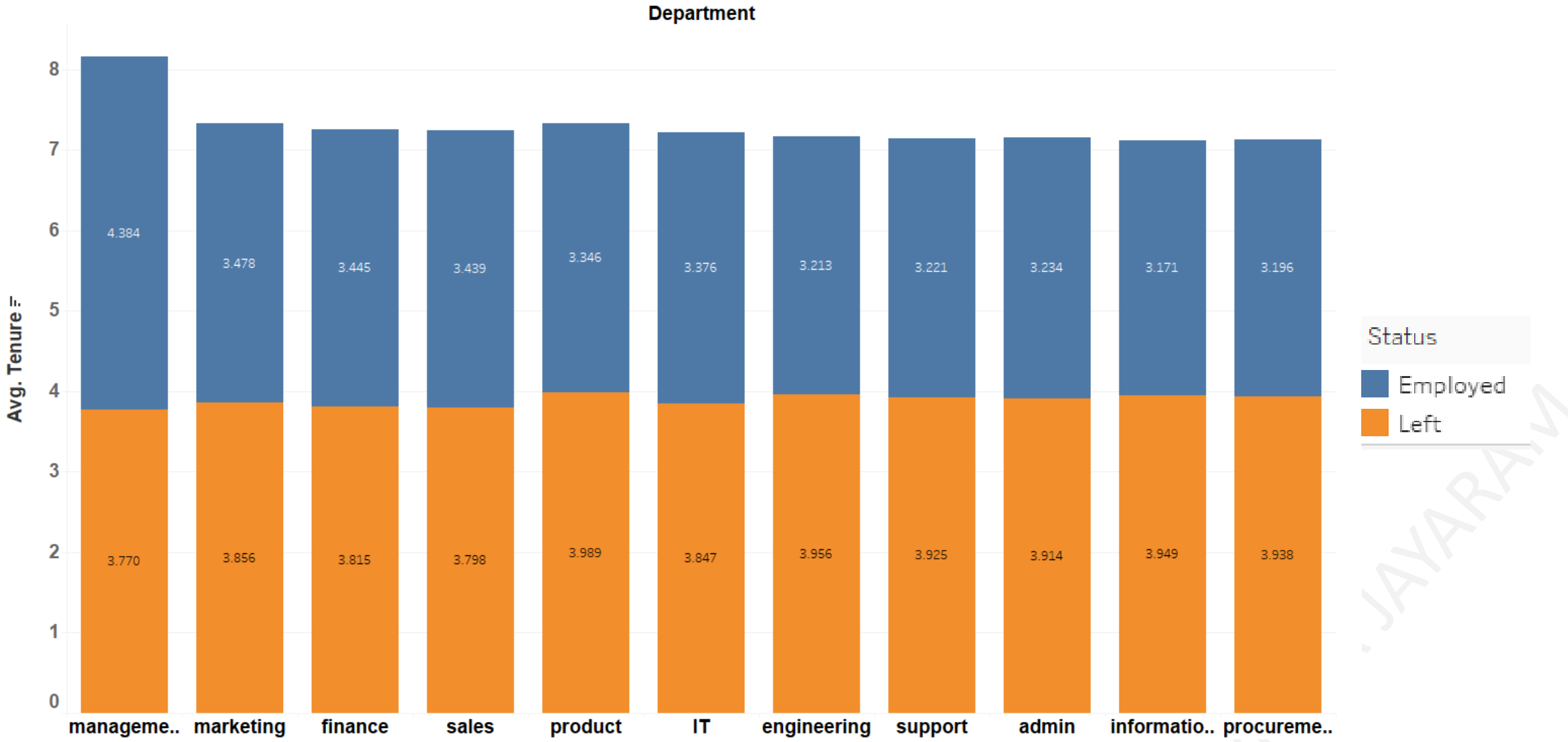
Employee satisfaction is high (70%) in employees with 2 years of tenure, then drops sharply to 47% at 4 years, then rises to 62% at 5 years. After 5 years, it plateaus out.

Departments Vs Employee status, No. of employees



The chart shows, employees employment status in each department, and the number of employees employed/left

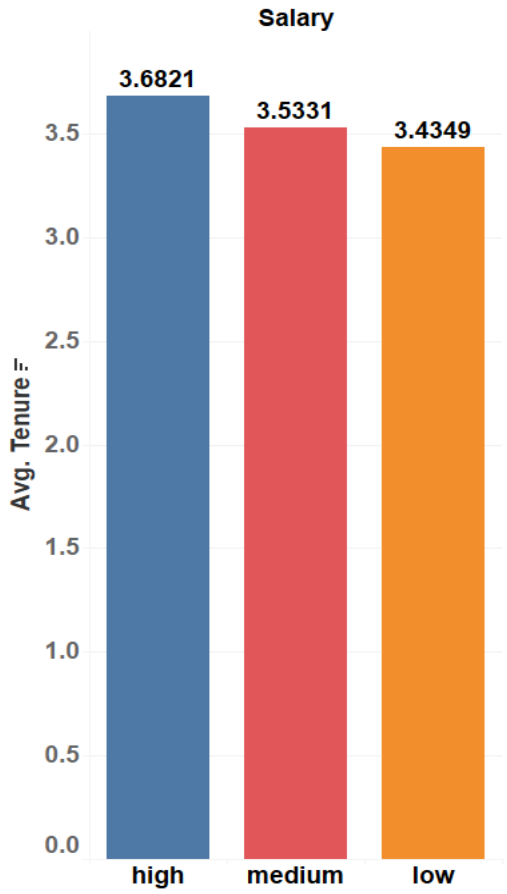
Departments Vs Employee status, tenure



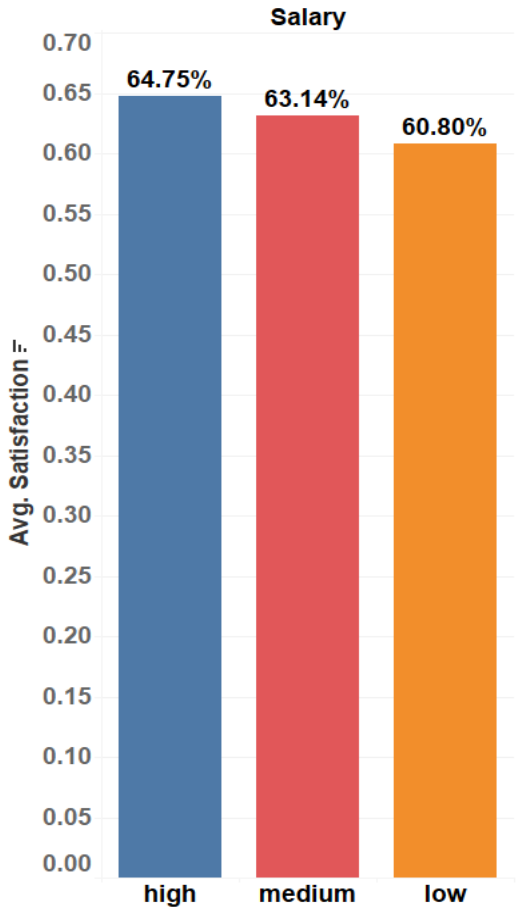
The chart shows, employees employment status in each department, and the average tenure of employees

Does Salary affect other features?

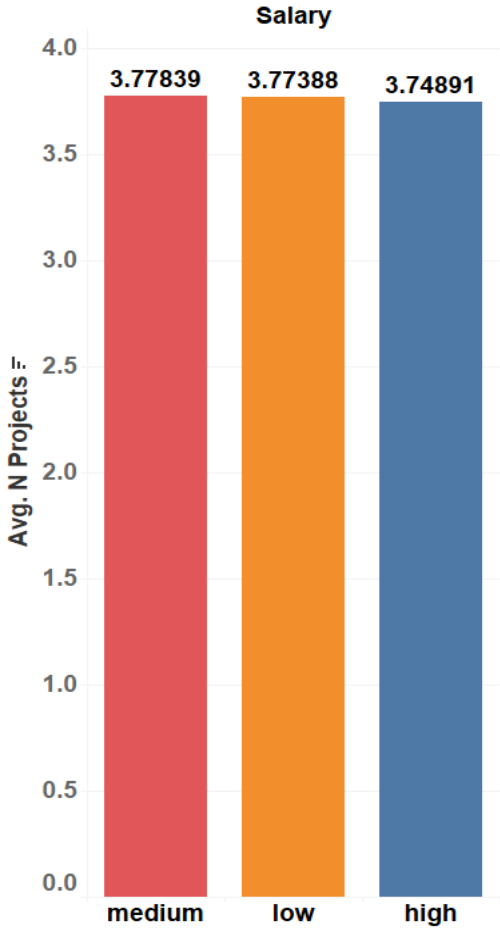
Salary Vs Avg. Tenure



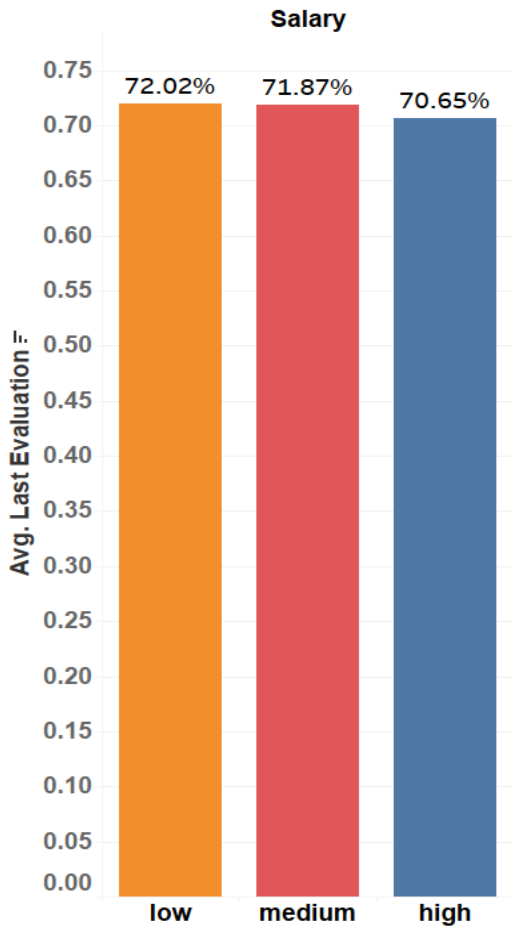
Salary Vs Satisfaction



Salary Vs No. of projects

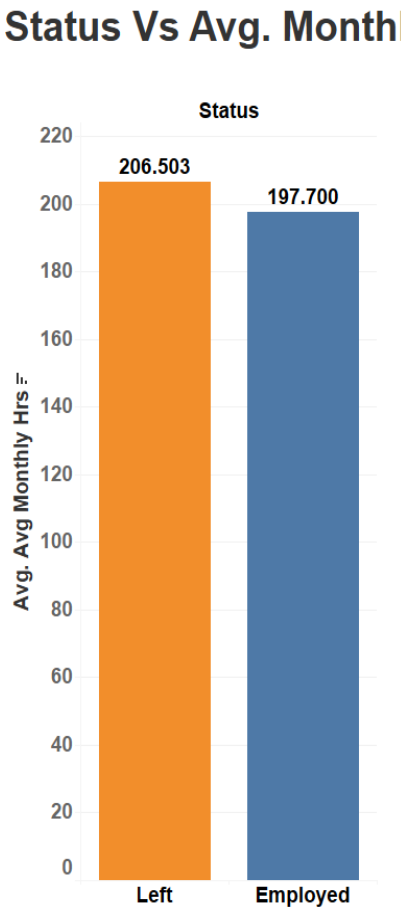


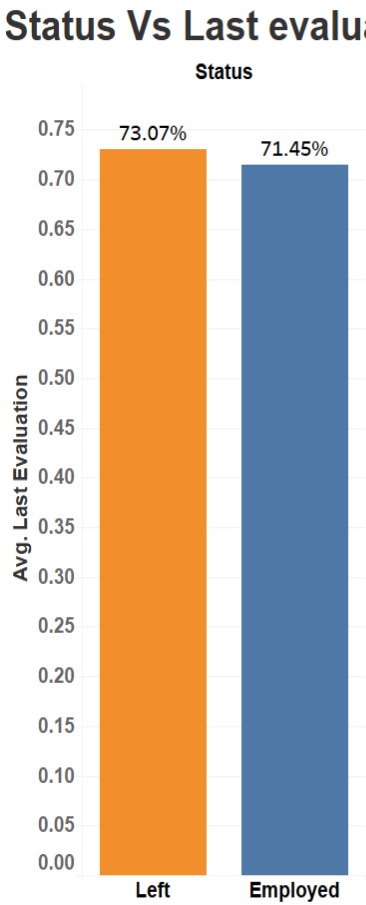
Salary Vs Last evaluation

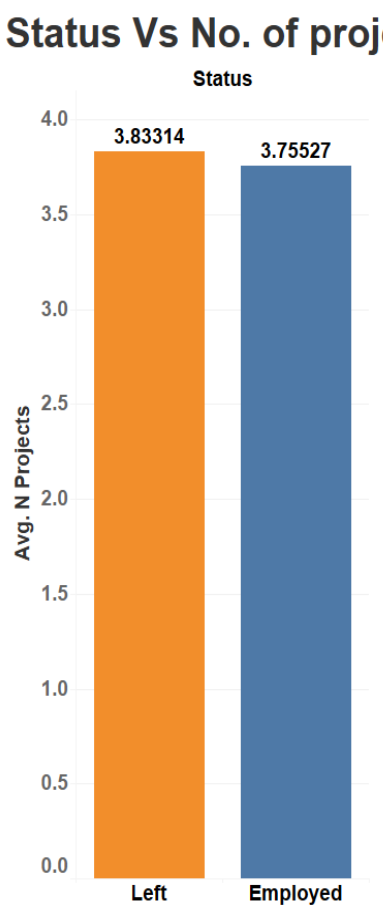


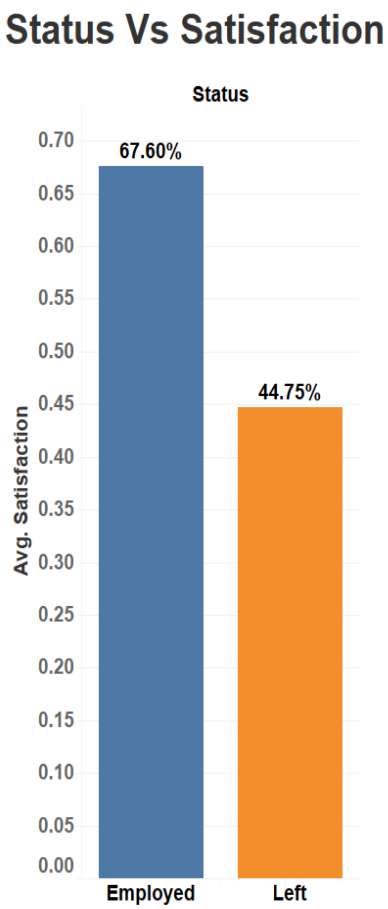
Employees with high Salary have high Employee satisfaction and high tenure, but the difference is not significant

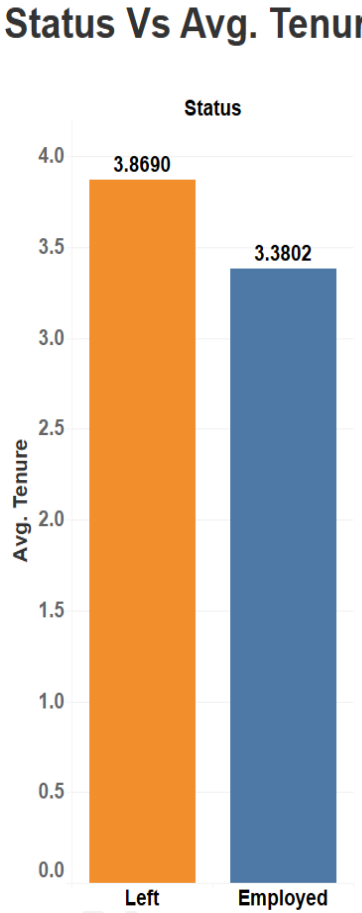
Which feature affects employee retention?











Employee satisfaction is the most impacting feature for Employee retention

Data Preparation

Total number of features = 10
Total number of records = 14249

No duplicate values found

Missing values

Department (709 records)	5%
Filed complaint (12191 records)	**86%**
Last evaluation (1532 records)	11%
Recently promoted (13949 records)	**98%**
Satisfaction (181 records)	1%
Tenure (181 records)	1%

Exploratory Data Analysis

The Features (i.e., Variables) are segregated into three categories:

Dependent Variable (Y): Variable that is being measured in the experiment. It changes as a result of the changes to the independent variables. Y values to predict:

Y : Employee Status (Employed/Left)

Noise: Variable that does not affect the dependent variable.

Independent Variable or Predictor Variable (X): Variable whose change isn't affected by any other variable in the experiment.

Independent variable is the cause, and dependent variable is the effect.

Exploratory Data Analysis - continued

Noise Variables

- Filed Complaint
- Recently promoted

Independent Variables

- Average monthly hours
- Department
- Last Evaluation
- Number of projects
- Salary
- Satisfaction
- Tenure

Data Analysis – Azure ML

- Qualitative and Quantitative analysis
- Find features impacting the predictor variable
- Find linear relationship between variables

Chi-Squared Test Analysis

Chi-squared test is a statistical method that measures how close expected values are to actual results.

Top 5 impacting features on employee retention: **Number of projects, Satisfaction, Avg. Monthly hours, Tenure, Last evaluation**

Bottom 2 impacting features on employee retention: **Salary, Department**

Independent variable	Chi-squared test value
No. of projects	4664.748952
Satisfaction	3549.141523
Avg. Monthly hours	1944.677611
Tenure	1858.047258
Last evaluation	1085.765778
Salary	350.817666
Department	78.635058

Linear Correlation Tests Analysis

The correlation coefficient **r** measures the strength and direction of a linear relationship between two variables.

r is always between +1 (Strong positive) and –1 (Strong negative).

Strong correlation: r > 0.7, Moderate correlation: 0.6 to 0.4 , Weak correlation: r < 0.4

Top 3 features that have strong linear relationship with **Employee retention status**: Satisfaction, Salary, Tenure

All other correlations are weak.

Independent Variable	R (Independent variable, Average Standard Score)
Avg. Monthly hours	-0.073514
Last evaluation	-0.037926
No. of projects	-0.026464
Salary	0.155516
Satisfaction	0.3878037
Tenure	-0.1430274

Questions?