Case Study in People Analytics

Identifying Key Factors for Attrition at Frito Lay®

Presented by: Celia Banks, Ph.D.

Lead Researcher

Executive Summary

A study was conducted to determine the key factors that contribute to attrition at Frito Lay. Key findings identified in the analysis of Company-supplied data are highlighted below:

- Exploratory data analysis provided plotting of various variables' associations.

- Linear regression was performed with stepwise approach to introduce variables individually to model. Goal was to identify the best variables to include in an analytic model.

- Top variables associated with attrition were identified to be Years at Company, Years in Current Role, Years Since Last Promotion. A fourth variable, Age, was also found to have a significant association with attrition on its own, notwithstanding its potential as a confounder in associating other variables to attrition.

- Given the data, a k-Nearest Neighbor analysis was the chosen approach for its ability to do multiclass classification. The analysis provided prediction accuracy at 85%.

Company Background

Formed in 1961 when C.E. Doolin and Herman W. Lay merged. The Company was acquired by PepsiCo in 1965. Currently, Frito-Lay has 29 snack brands and 55K employees. Their motto is "we go the extra mile to bring you your favorite snacks." Their future embraces *Winning with Purpose*, a theme encompassing winning in the marketplace, accelerating top-line growth, and keeping a commitment to the planet and communities. The Company's mission is to "Create more smiles with every sip and every bite."

Online human resource boards report that Frito Lay employees give it an average rating of 3.5 out of 5 stars. The boards consistently report that Sales Representatives are the happiest employees and have average rating of 4.8. [*Source: CareerBliss*] The Company operates 82 percent non-union since the last fully unionized warehouse located upstate New York voted to decertify its 15-year relationship with the Teamsters Union (there remain sparse union membership in some states). A recent 20-day walkout in Kansas saw hundreds of workers express concern over extreme overtime hours amid the COVID-19 pandemic.

Data Assumptions

- The data supplied contained a large number of categorical and discrete variables about employee data.

- Some data columns were dropped from the analysis once they were identified as having no meaningful association with attrition. Data columns were also dropped that could not be used for purposeful analysis such as Employee Number, Over 18, or Marital Status.

- The scope of inference for the analysis is delimited to observational study. As such, the absence of randomness in sample and assignment means that any associations are neither causal nor generalizable beyond the data and the study.

Report Content

Part I is an exploratory analysis of the data to seek patterns and trends in the data. Graphs are introduced that aid in visualizing data associations.
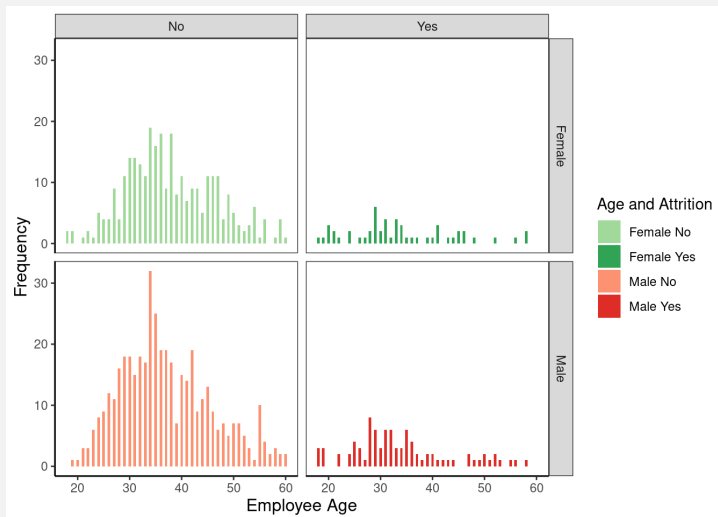
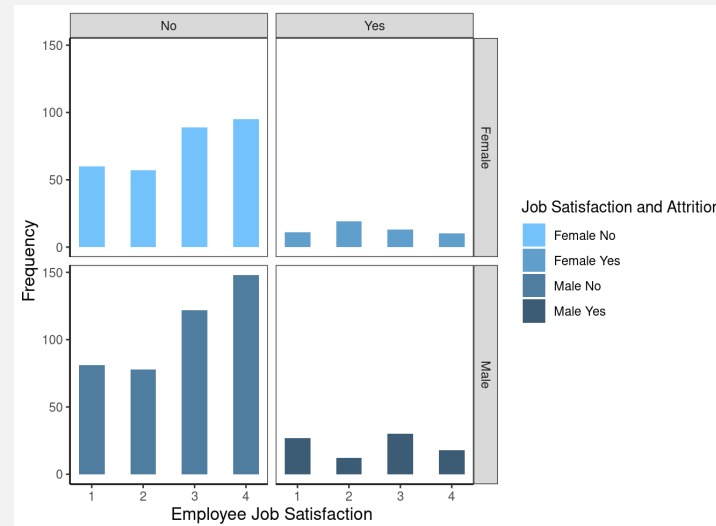Part II provides the core analysis, identifying top variables associated with attrition.

Part I – Exploratory Analysis

Visual Dashboard of Variable Associations

Age, Gender vs Attrition
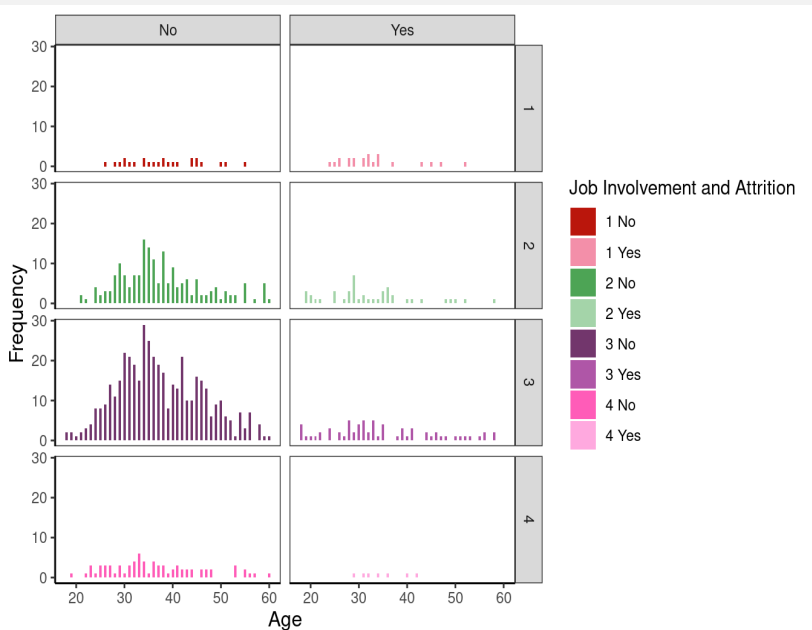
Job Satisfaction, Gender vs Attrition

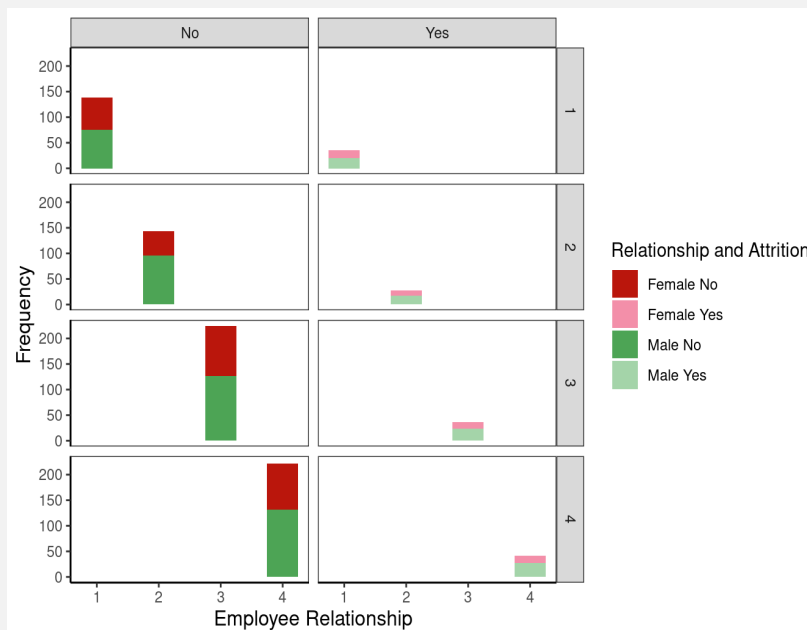Performance Rating, Age vs Attrition

Part I – Exploratory Analysis

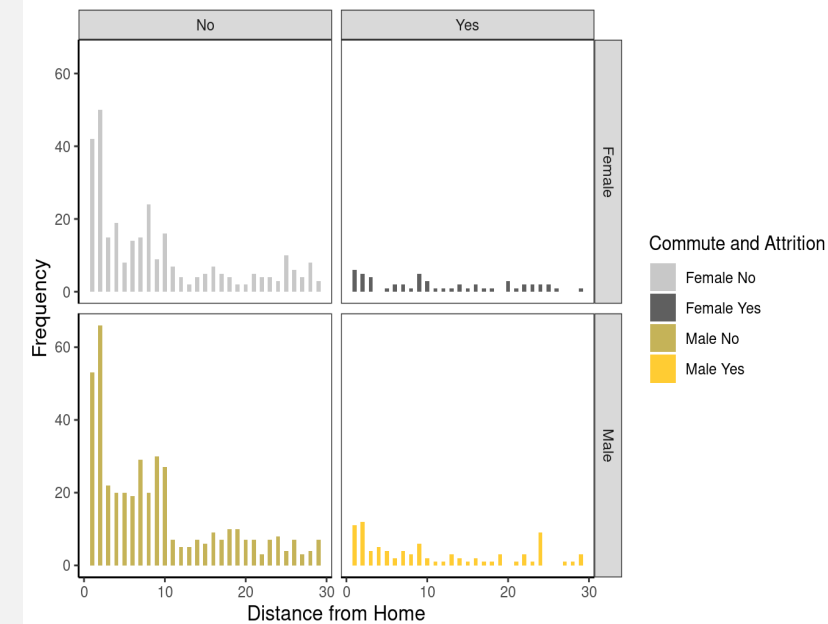## Visual Dashboard of Variable Associations

Job Involvement, Age vs Attrition
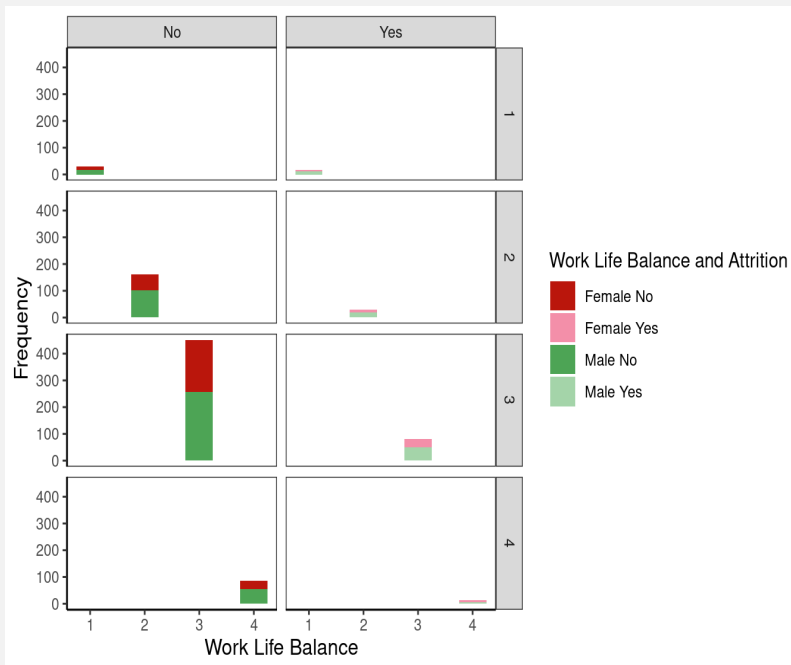
Relationship, Gender vs Attrition
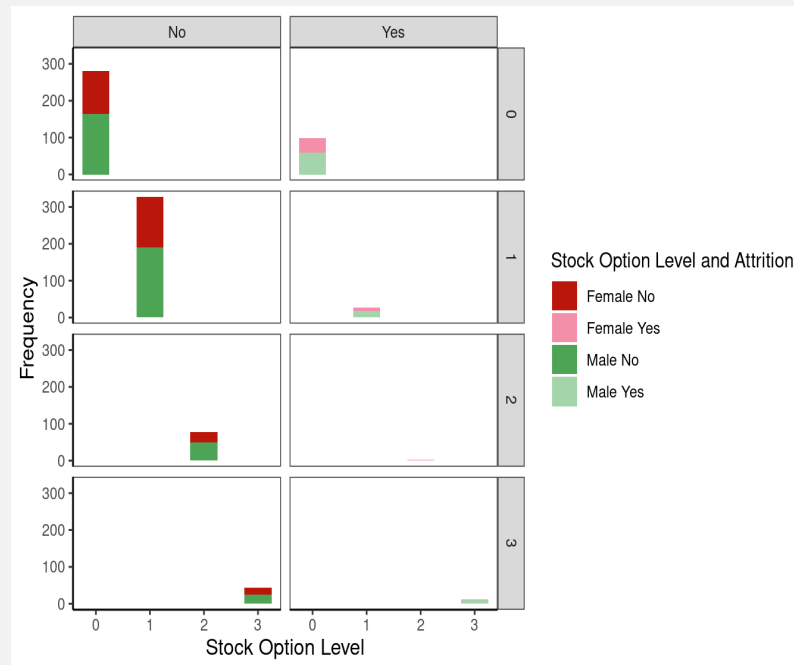
Commute, Gender vs Attrition

Part I – Exploratory Analysis

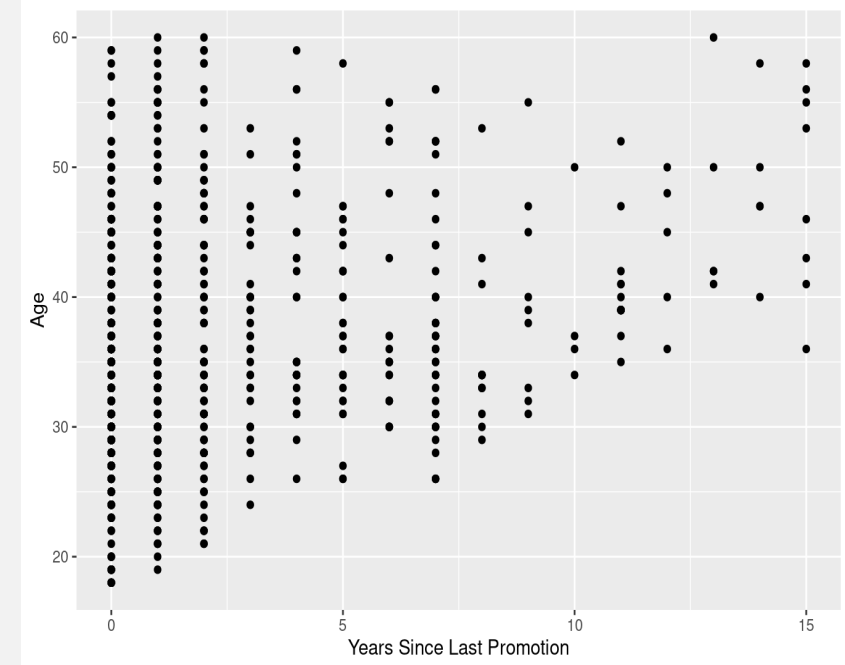## Visual Dashboard of Variable Associations

Part II – Core Analysis

Stepwise Regression

- AIC stepwise modelling was used to add each variable to the regression analysis. This approach helped in identifying the key variables associated with attrition. Specifically, four variables stood out as the top factors each of which had R^2 indicators showing strong association with attrition.

- The top variables identified in the AIC stepwise regression were:
    - (1) Years Since Last Promotion
    - (2) Years in Current Role
    - (3) Years with Company
    - (4) Age

- While Age was shown as an individual contributor, it also had potential for confounding other variables.

Part II – Core Analysis

k-Nearest Neighbor Classification Analysis

- A k-NN analysis provided multiclassification of the top four variables in association with attrition.

- The model demonstrated prediction accuracy of 85%. It showed specificity of 86% and sensitivity of 82%. Further, a 95% confidence interval of 81%, 90% suggests that the model is fairly accurate in predicting attrition.

- A No Information Rate of 96% gives assurance that we can pick the majority class of "No" and expect to be accurate.

- To validate model accuracy, a binomial test was performed and found to have a probability of 84% in predicting success.

Conclusion: Analysis Interpretation

Overall analysis suggests that the Company will experience less attrition in new employees. Given the key indicators associated with attrition, it is recommended that positive succession planning approaches be adopted to ensure that more experienced exiting employees will share knowledge with new hires.

Full Analysis Report

See the full report at:

https://i-meta-inc.shinyapps.io/People-Metrics/