Frito-Lay Analysis

DS 3605 Project 2

Christy Wachira

Agenda

- Overview

- Data Visualization

- Linear Regression model

- Naïve Bayes model
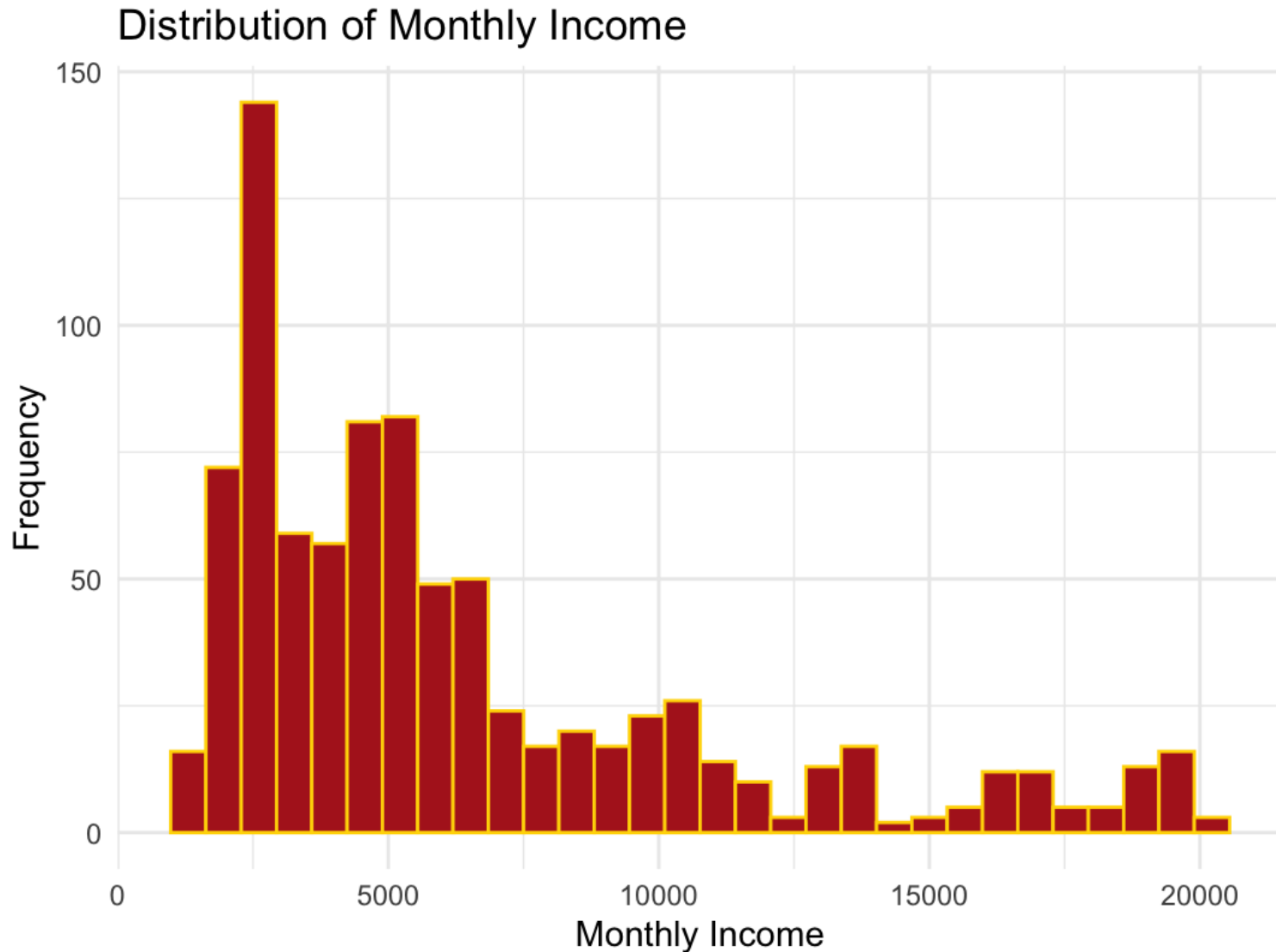
- Key findings

Overview

» Objective: Analyze existing employee data to identify the top factors contributing to attrition and provide job role-specific trends.

» Employee Dataset:
   o 870 employees
   o 37 variables
   o No missing values
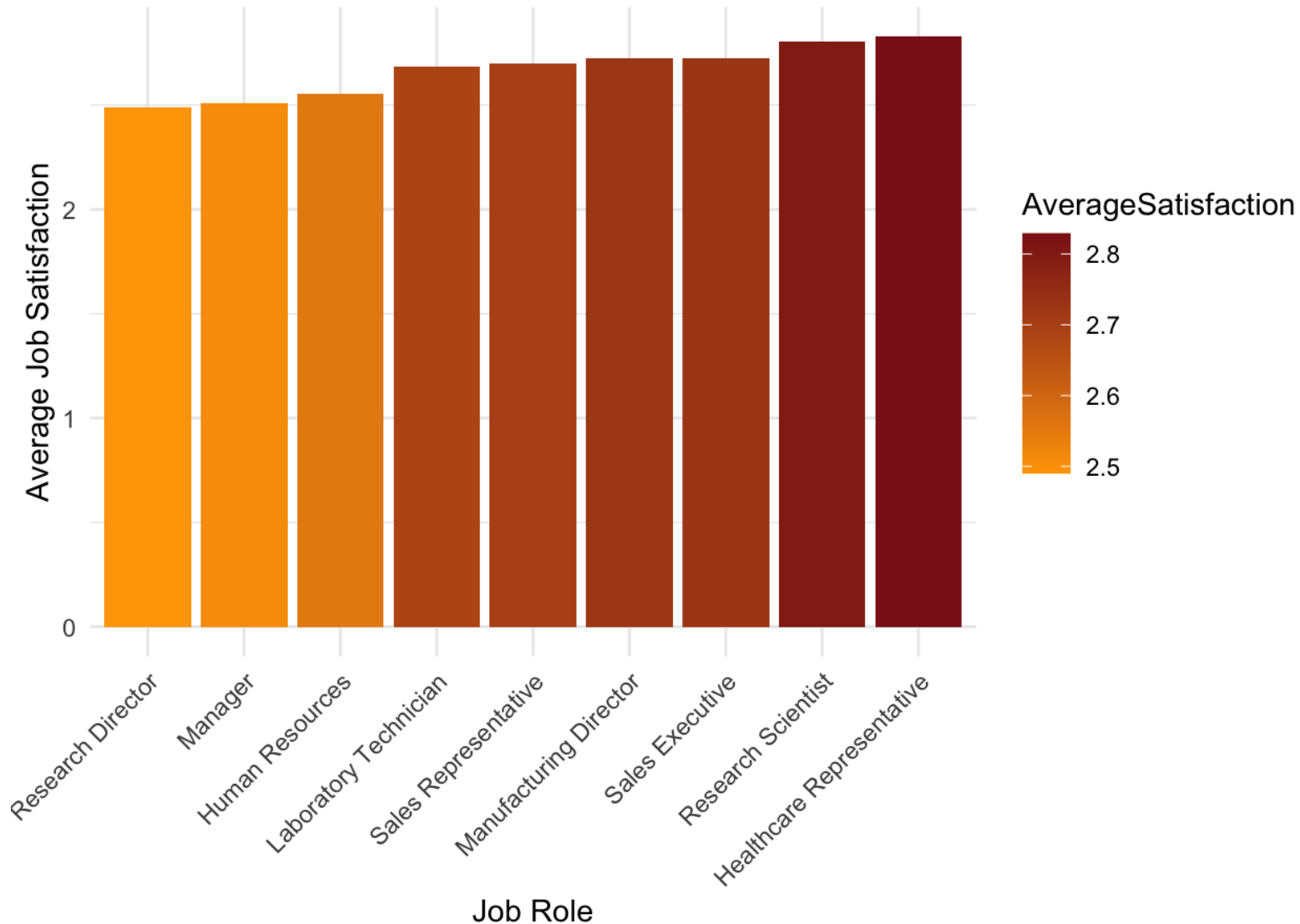
Distribution of variables for monthly income

Distribution of Monthly Income

- Majority of employees earn between $2,500 and $5,000

- Distribution shows right skewness—fewer employees with high incomes.
- Rshiny app: http://127.0.0.1:5276/

Job role specific trends

Average Job Satisfaction by Job Role

- Job satisfaction across various job roles
- Assumption of normality violated
- Assumption of equal variances met
- Kruskal-Wallis test (p-value: 0.542)

Salary Model

» Linear Regression Model (predicting salary)

  » Significant Predictors:

  - Job Level: Strong positive impact on monthly income, with each level increase boosting income by approximately $3,723.
  - Total Working Years: Each additional year of experience increases income by about $59.

  » Model Performance:

  - High Accuracy: The model explains about 91.65% of the variance in monthly incomes.
  - Residual Error $1,360.
  - RMSE on Test Data: $1502.41

Attrition Model

» Naive Bayes Model (predicting attrition)

  » 82.76% accuracy, with confidence interval ranging from 76.31% to 88.05

  » 89.04% sensitivity, indicating high accuracy in predicting 'No Attrition'

  » 50% specificity, showing room for improvement in predicting 'Yes Attrition'

  » Positive Predictive Value: 90.28% for correctly predicting 'No Attrition'

  » Negative Predictive Value: 46.67% for correctly predicting 'Yes Attrition'

Key Findings

» Top 3 factors contributing to attrition

  o Job Level

  o Overtime

  o Monthly Income

Thank You!

Christy Wachira

cwachira@smu.edu