**Analyzing Employee Satisfaction and Job Attrition**

Joshua Wei

Colorado State University Global

MIS581: Business Intelligence and Data Analytics

Dr. Justin Bateh

10/06/2024

**Link to video presentation:** <https://youtu.be/ynP6s2xm0Tw>

**Link to GitHub account:** <https://github.com/jweicsu/MIS581-Portfolio-Project>

**Programming code:**

# Importing necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

from sklearn.linear\_model import LogisticRegression

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import confusion\_matrix, accuracy\_score, roc\_curve, auc

import statsmodels.api as sm

from scipy import stats

# Load the dataset (assuming it’s in CSV format)

df = pd.read\_csv('HR\_Employee\_Attrition.csv')

# Convert categorical variables to numerical (if not already done)

df['Attrition'] = df['Attrition'].map({'Yes': 1, 'No': 0})

# For simplicity, focus on a few variables, including Job Satisfaction, Monthly Income, Work-Life Balance

df['JobSatisfaction'] = pd.to\_numeric(df['JobSatisfaction'])

df['WorkLifeBalance'] = pd.to\_numeric(df['WorkLifeBalance'])

df['MonthlyIncome'] = pd.to\_numeric(df['MonthlyIncome'])

# Drop any missing values

df = df.dropna()

# Feature selection and scaling

X = df[['JobSatisfaction']] # You can include other factors like 'MonthlyIncome' or 'WorkLifeBalance'

y = df['Attrition']

# Split the dataset into train and test sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

# Scaling the data

scaler = StandardScaler()

X\_train\_scaled = scaler.fit\_transform(X\_train)

X\_test\_scaled = scaler.transform(X\_test)

# Logistic Regression Model

log\_model = LogisticRegression()

log\_model.fit(X\_train\_scaled, y\_train)

# Predict probabilities for test data

y\_prob = log\_model.predict\_proba(X\_test\_scaled)[:,1]

# Generate plot for logistic regression

plt.figure(figsize=(10, 6))

plt.scatter(X\_test, y\_test, color='blue', label='Actual Data')

plt.plot(X\_test, y\_prob, color='red', label='Predicted Probability', linewidth=2)

plt.xlabel('Job Satisfaction')

plt.ylabel('Attrition Probability')

plt.title('Logistic Regression of Job Satisfaction and Attrition')

plt.legend()

plt.grid(True)

plt.show()

# Select relevant features for correlation matrix

correlation\_data = df[['Attrition', 'JobSatisfaction', 'WorkLifeBalance', 'MonthlyIncome']]

# Calculate the correlation matrix

corr\_matrix = correlation\_data.corr()

# Plot the correlation matrix using seaborn

plt.figure(figsize=(8, 6))

sns.heatmap(corr\_matrix, annot=True, cmap='coolwarm', linewidths=0.5)

plt.title('Correlation Matrix of Key Variables')

plt.show()

# Perform a Chi-Square test between JobSatisfaction and Attrition

contingency\_table = pd.crosstab(df['JobSatisfaction'], df['Attrition'])

chi2, p, dof, expected = stats.chi2\_contingency(contingency\_table)

# Print Chi-Square results

print(f"Chi-Square value: {chi2}")

print(f"P-value: {p}")

# Visualize the Chi-Square test results with a bar plot

contingency\_table.plot(kind='bar', stacked=True, figsize=(10, 6))

plt.title('Chi-Square Test of Job Satisfaction and Attrition')

plt.xlabel('Job Satisfaction')

plt.ylabel('Employee Count')

plt.legend(title='Attrition', loc='upper right')

plt.show()

# Get predictions and calculate accuracy

y\_pred = log\_model.predict(X\_test\_scaled)

accuracy = accuracy\_score(y\_test, y\_pred)

print(f"Accuracy of Logistic Regression Model: {accuracy}")

# Plot ROC Curve

fpr, tpr, thresholds = roc\_curve(y\_test, y\_prob)

roc\_auc = auc(fpr, tpr)

plt.figure(figsize=(8, 6))

plt.plot(fpr, tpr, color='orange', label=f'ROC curve (area = {roc\_auc:.2f})')

plt.plot([0, 1], [0, 1], color='darkblue', linestyle='--')

plt.xlabel('False Positive Rate')

plt.ylabel('True Positive Rate')

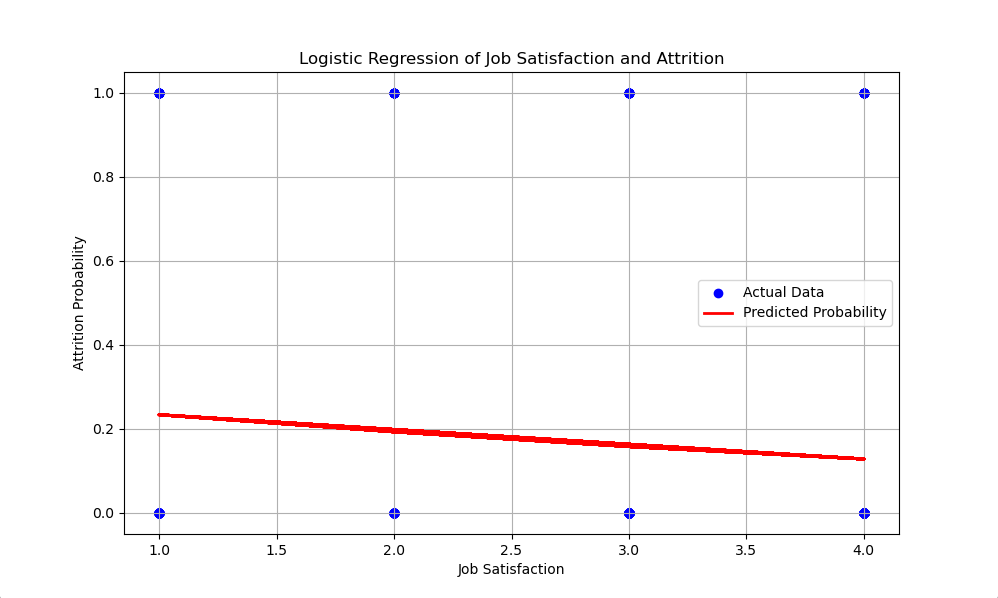
plt.title('Receiver Operating Characteristic (ROC) Curve')

plt.legend()

plt.show()

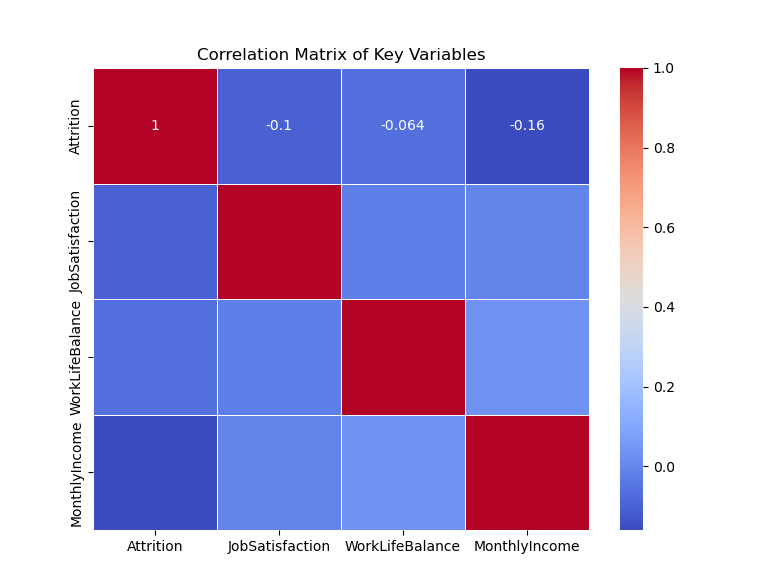
**Figure 1**

*Logistic Regression of Job Satisfaction and Attrition*



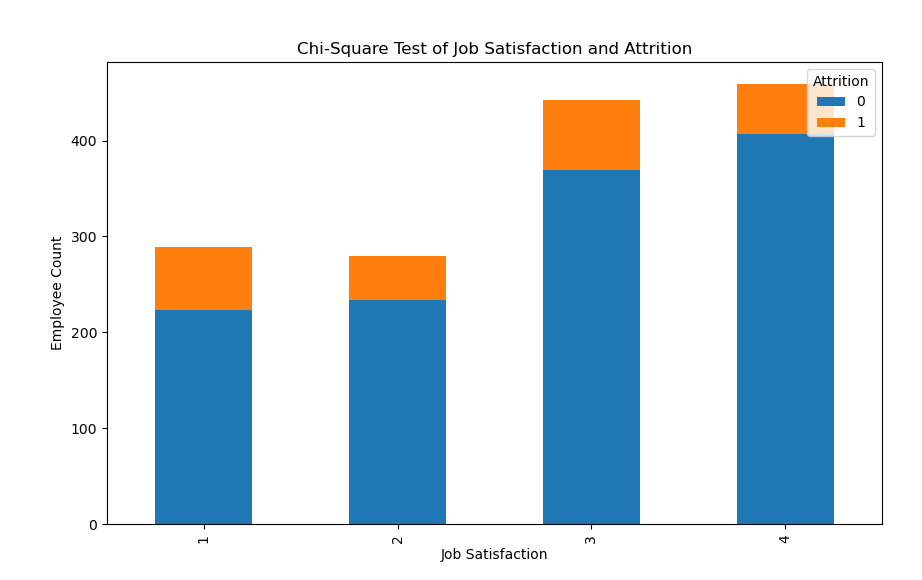
**Table 1**

*Correlation Matrix of Key Variables*



**Figure 2**

*Chi-Square Test of Job Satisfaction and Attrition*



**Figure 3**

*Receiver Operating Characteristic (ROC) Curve for Logistic Regression Model*

