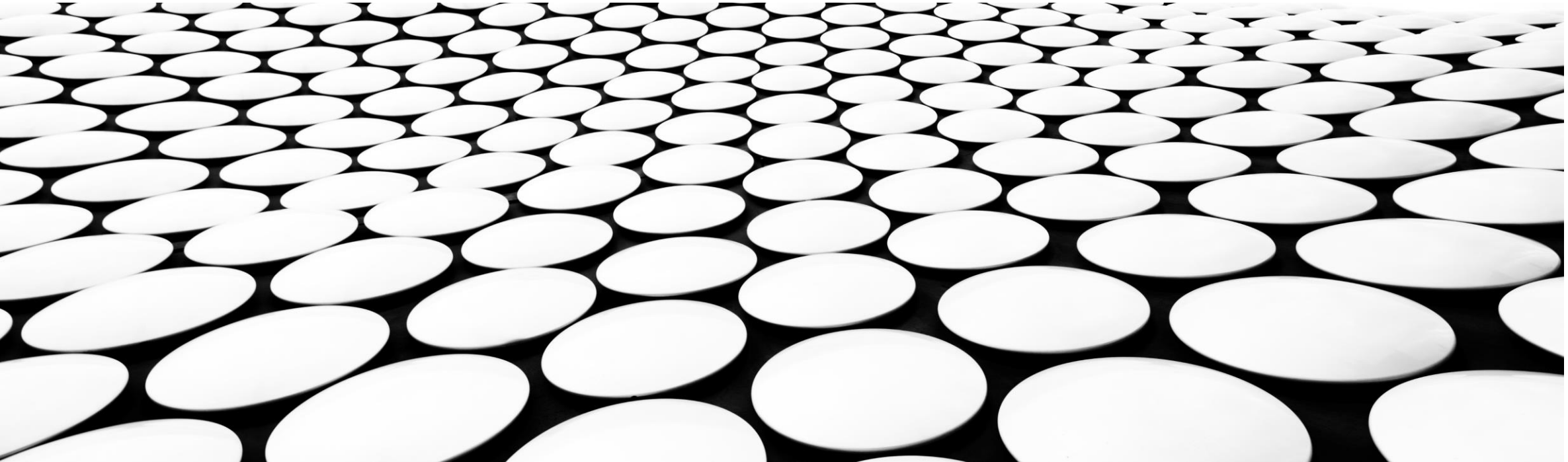# EMPLOYEE BURNOUT PREDICTION USING XGBOOST

STUDENT DETAILS

Name: Jagadeesh Ospari
Admission Number: AP22110011304
Email id: jagadeeshjagan2003@gmail.com
Degree: Btech
Year: 2nd Year
Branch: Computer Science Engineering
College/University: SRM University
Summer Internship: Intern under Edunet Foundation through SRM University
Domain: Artificial intelligence and machine learning(01-06-2024 to 25-07-2024)

PROJECT TITLE/PROBLEM STATEMENT

Project Title:

Employee Burnout Prediction Using XGBOOST

Project Statement:

In today's fast-paced work environment, employees often experience stress and depression due to heavy workloads, tight deadlines, and mental fatigue. This can result in diminished productivity and engagement, ultimately hindering organizational growth.

AGENDA

- Project Overview

- End Users

- Solution and Value Proposition

- Customization and Innovation

- Modeling Approach

- Results and Evaluation

- Conclusion and Next Steps

PROJECT OVERVIEW

• Objective: Predict employee burnout rates using machine learning

• Data source: CSV file with employee information

• Features: Gender, Company Type, WFH Setup, Resource Allocation, Mental Fatigue Score, etc.

• Target variable: Burn Rate

• Approach: Data preprocessing, exploratory data analysis, and XGBoost regression model

WHO ARE THE END USERS OF THIS PROJECT?

- Human Resources professionals

- Company management and executives

- Team leaders and managers

- Workplace wellness program coordinators

YOUR SOLUTION AND ITS VALUE PROPOSITION

• Machine learning model to predict employee burnout rates

• Comprehensive data analysis and visualization

• Value Proposition:

• Proactive approach to employee well-being

• Personalized interventions based on individual risk factors

• Cost savings through reduced turnover and increased productivity

• Improved workplace culture and employee satisfaction

HOW DID YOU CUSTOMIZE THE PROJECT AND MAKE IT YOUR OWN

•Extensive exploratory data analysis with various visualization techniques

•Feature engineering: Extracting month and day from joining date

•Handling missing values through mean imputation

•Binary encoding for categorical variables

•Standardization of features for improved model performance

•Use of XGBoost, a powerful gradient boosting algorithm

MODELLING

•Data preprocessing

•Handling missing values

•Feature engineering

•Encoding categorical variables

•Exploratory Data Analysis:

Histograms, correlation heatmap, pairplots, boxplots

•Model: XGBoost Regressor

•Train-test split: 70% training, 30% testing

•Feature scaling using StandardScaler
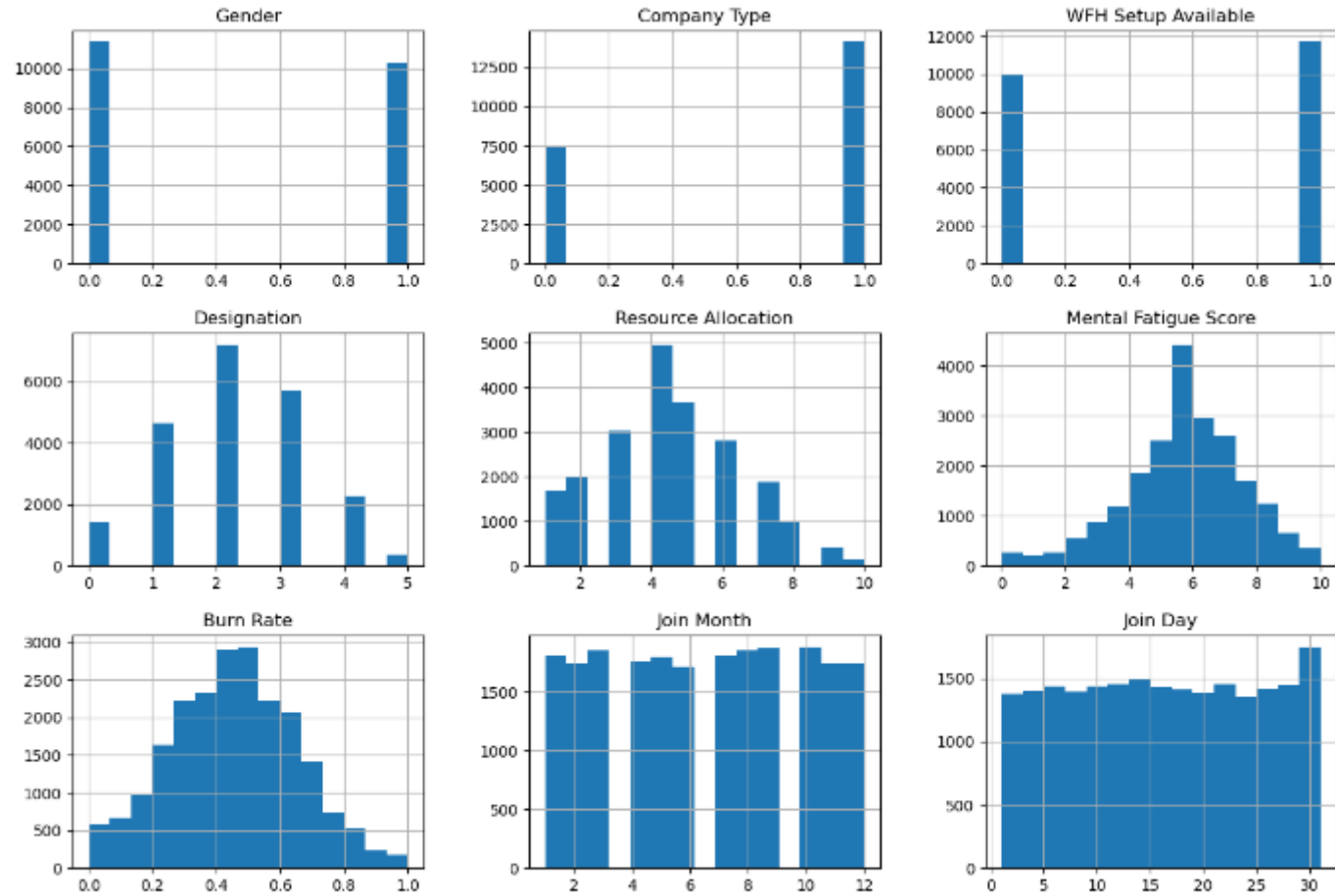
RESULTS

•**Model performance:**
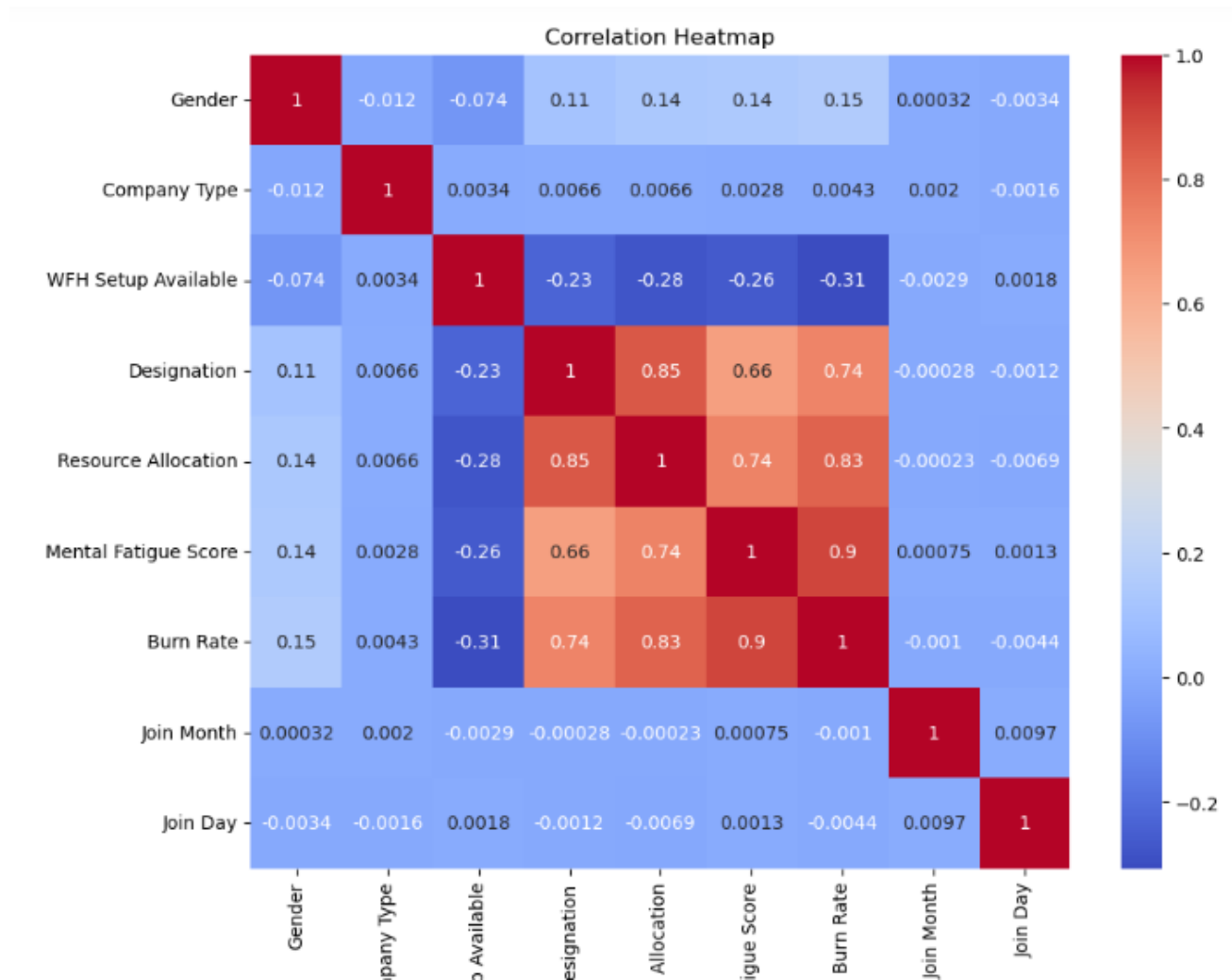
```
In [41]:  # Evaluate the model
          train_score = model.score(X_train, y_train)
          test_score = model.score(X_test, y_test)

In [42]:  |
          print(f'Training Score: {train_score}')
          print(f'Testing Score: {test_score}')

          Training Score: 0.9383634165070733
          Testing Score: 0.9035718747799601
```
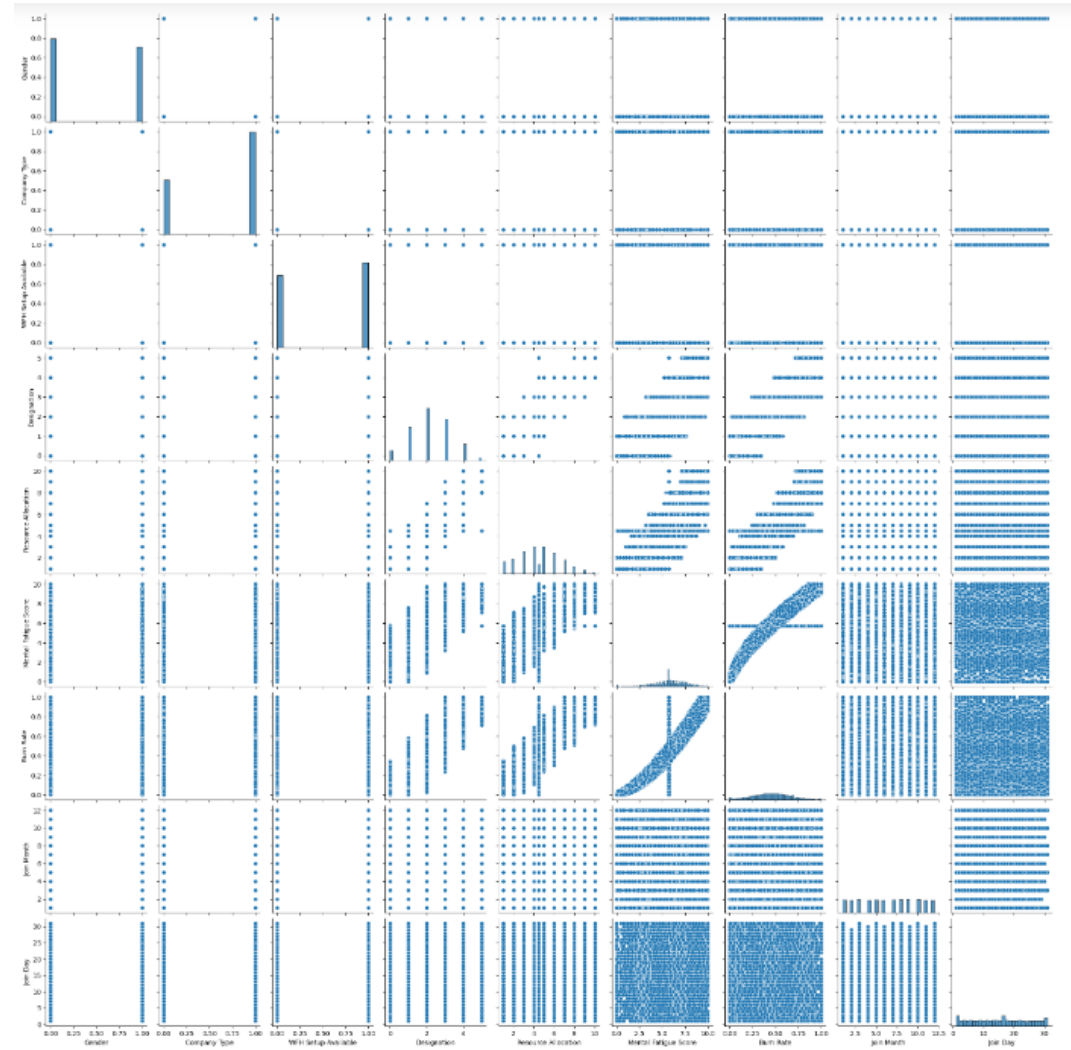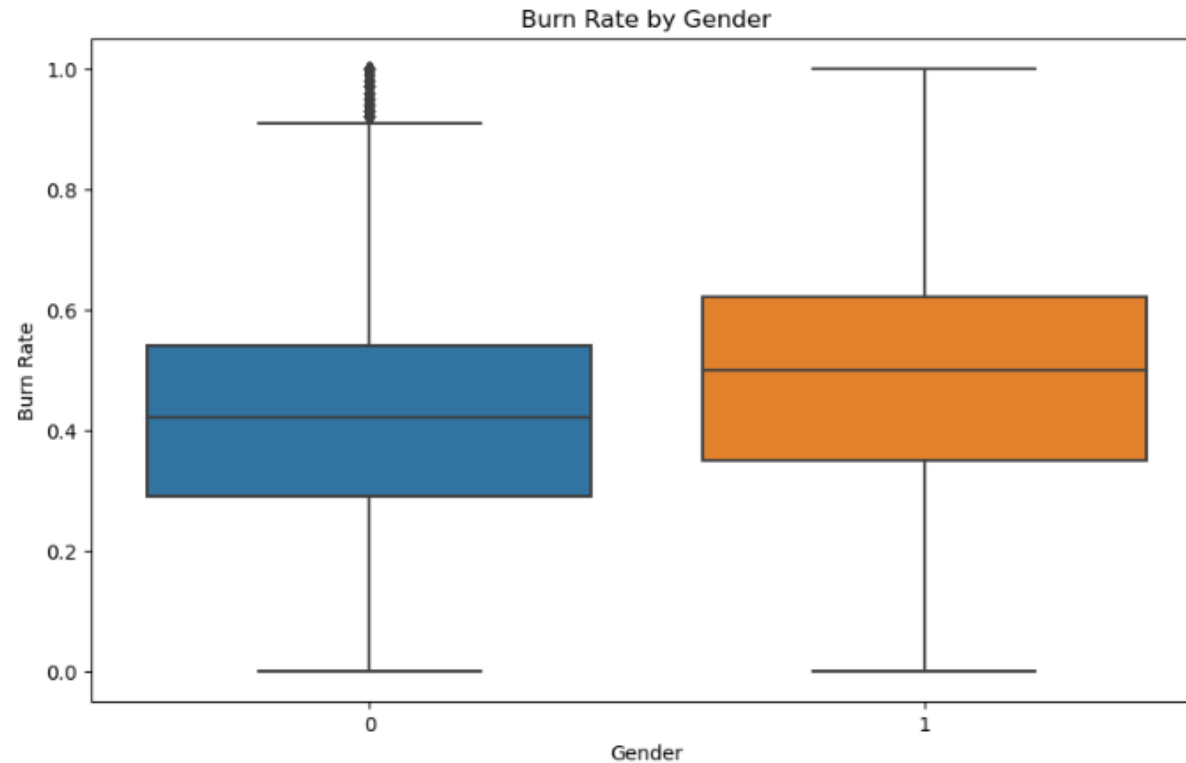
EXPLORATARY DATA ANALYSIS

Correlation Heatmap

WFH SETUP AVAILABLE HAS A MODERATE NEGATIVE CORRELATION (-0.31) WITH BURN RATE, SUGGESTING THAT EMPLOYEES WITH WORK-FROM-HOME SETUPS TEND TO HAVE SLIGHTLY LOWER BURNOUT RATES.

THE MOST IMPORTANT FEATURES IN PREDICTING BURN RATE ARE LIKELY TO BE:
MENTAL FATIGUE SCORE
RESOURCE ALLOCATION
DESIGNATION

Burn Rate by Gender

THE BOXPLOT SHOWS THAT GENDER 1 (LIKELY REPRESENTING MALES) HAS A SLIGHTLY HIGHER MEDIAN BURN RATE AND A LARGER INTERQUARTILE RANGE COMPARED TO GENDER 0 (LIKELY REPRESENTING FEMALES).

LINKS

- GitHub repository: https://github.com/123jagadeesh/Employee_BurnOut_Prediction

- Dataset source: https://www.kaggle.com/datasets/blurredmachine/are-your-employees-burning-out?select=train.csv

- XGBoost documentation: https://xgboost.readthedocs.io/

- Scikit-learn documentation: https://scikit-learn.org/