Mental Health Accommodation in Tech Companies

Presented by Xinbei Yu
Institution: DSI at Brown University
Date Presented: 10/18/2022
Link to the project:
https://github.com/XinbeiYu00/project-XinbeiYu.git

Introduction

- **Research Question:** Which factors affect the easiness of taking a leave from work due to mental health conditions?

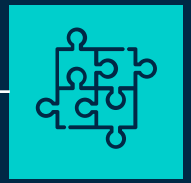
- **Why Important:** With growing attention to mental health issues in every industry, both employer and employee should start exploring their options and responsibilities.
    - **Employer Perspective:**
        - Healthy working environment
        - Appropriate benefits
        - Employee efficiency
    - **Employee Perspective:**
        - Learn their rights
        - Healthy working environment

- **Type of Problem:** Categorical

- **Target Variable:** Easiness of Taking A Leave ([leave] in the data set)

- **Data Source:** Mental Health in Tech Survey, Kaggle, https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey

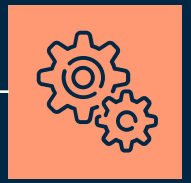
Exploratory Data Analysis



01

TARGET VARIABLE VISUALIZATION

[leave] distribution in bar plot



02

CATEGORICAL FEATURE

[obs_consequence] vs. [leave] [anonymity] vs. [leave]

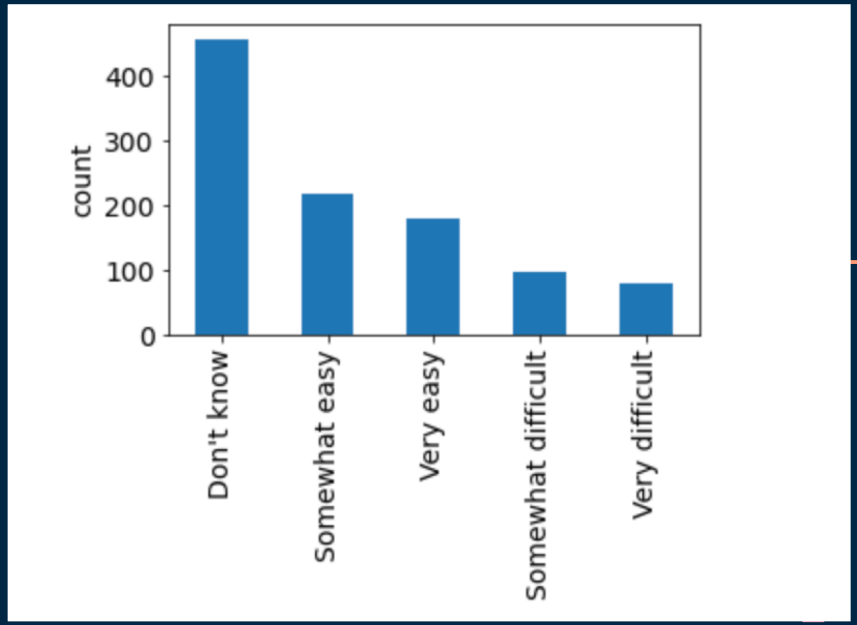
03

CONTINUOUS FEATURE

[Age] vs. [leave]

Target Variable Visualization

- Number of Data Points: 1031
- Number of Values: 5
- Most Frequent Value: Don’t know
- Interpretation: [Don’t know] needs further exploration



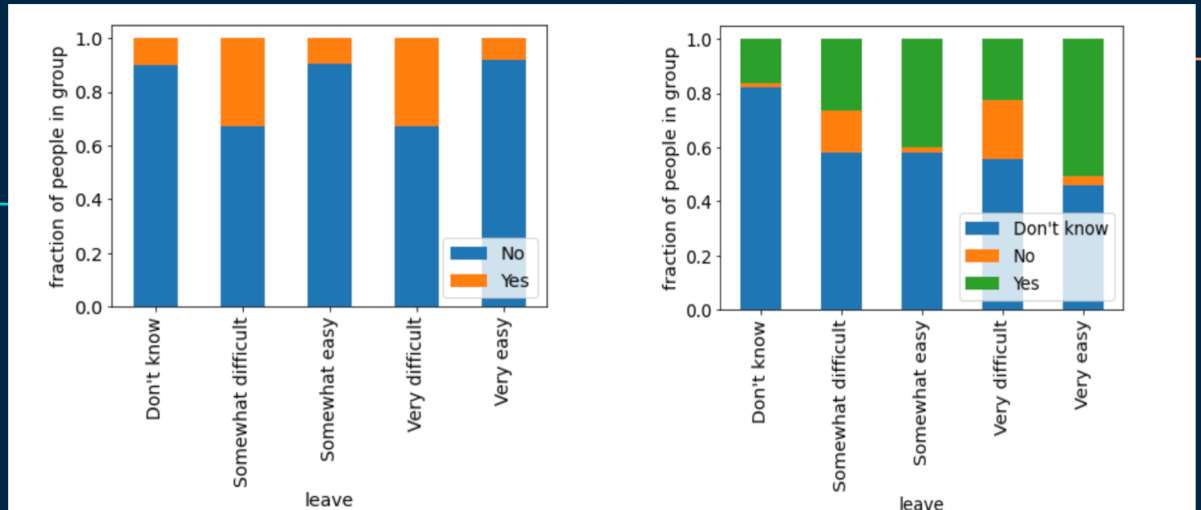
Categorical Feature

OBS_CONSEQUENCE

- Very difficult vs. Very easy
- Don’t know

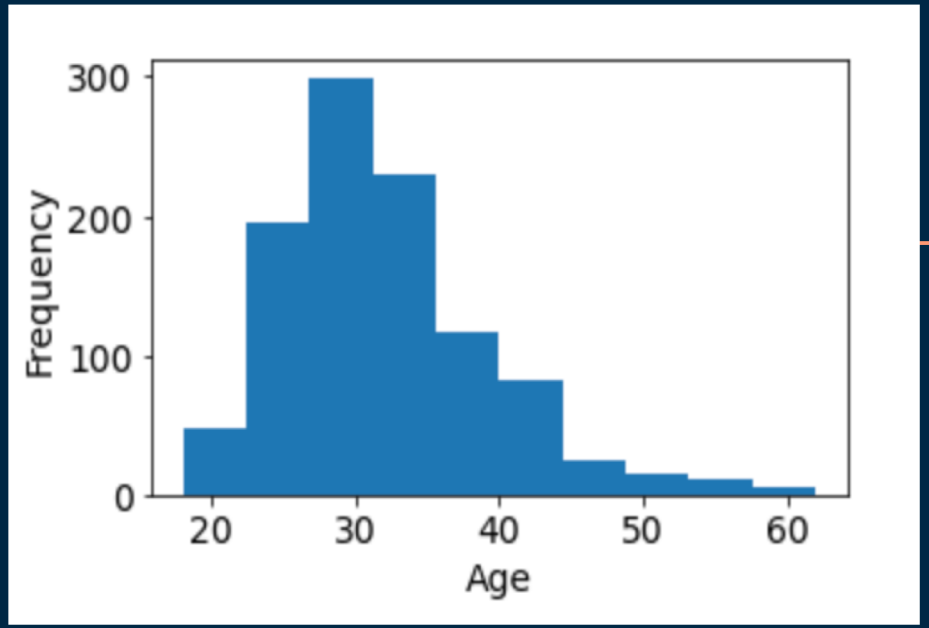
ABNOMYNITY

- Very difficult vs. Very easy
- Don’t know



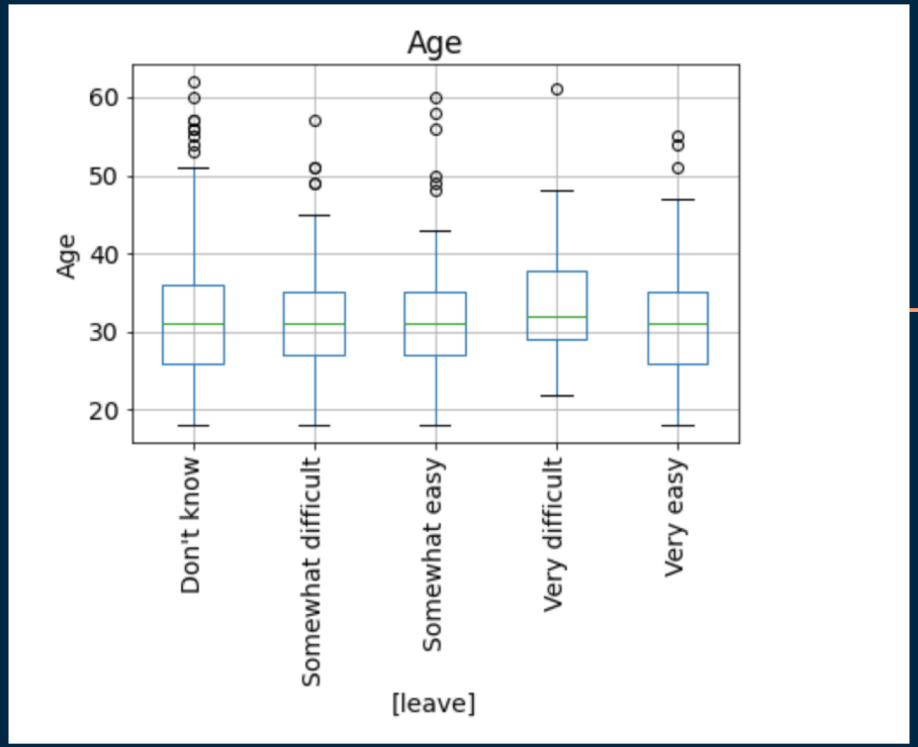
Continuous Feature

- Number of Data Points: 1025 (excludes wrong obvious wrong age)
- Skewness: right-skewed
- Mean: 32
- Range: (18, 62)



Continuous Feature

- Assumption: older people find taking a leave for mental health issue is harder
- Very difficult vs. Very easy
- Outliers



Splitting

Train_Test_Split

- Implicit Group: Company
- Lack further information on distribution method
- Assume i.i.d

StratifiedKFold

- Imbalanced Data: 455: 218: 176: 98: 78

Preprocessing

StandardScaler

- Age (tailed)

OneHotEncoder

- All other features except the dropped ones (timestamp, comments, etc)

After Splitting and Preprocessing

Features

- Number of Features:26 -> 23 -> 170
- Number of Data Points in Training Set: 615

Values

- Missing Values: state, comments, self_employed, work_interfere
- Wrong Values: age