

Mental Health In Tech



Dataset Used: Mental Health in Tech Survey (By Open Sourcing Mental Illness, LTD)

Research Question

Does the organizational culture of tech companies affect the mental health of employees? If so, how do companies and employees respond to such issues?

Project Timeline



Data Cleaning and EDA

Cleaned Data

- Eliminated non-tech responses
- Removed NAN Values
- Dealt with outliers
- Standardized columns

Drew Correlations

- Related external and internal factors with mental health issues

Visualized Data

- Visualized categorical data
- Used scatter, bar and pie plots

Key Findings

Influence of External and Internal Factors

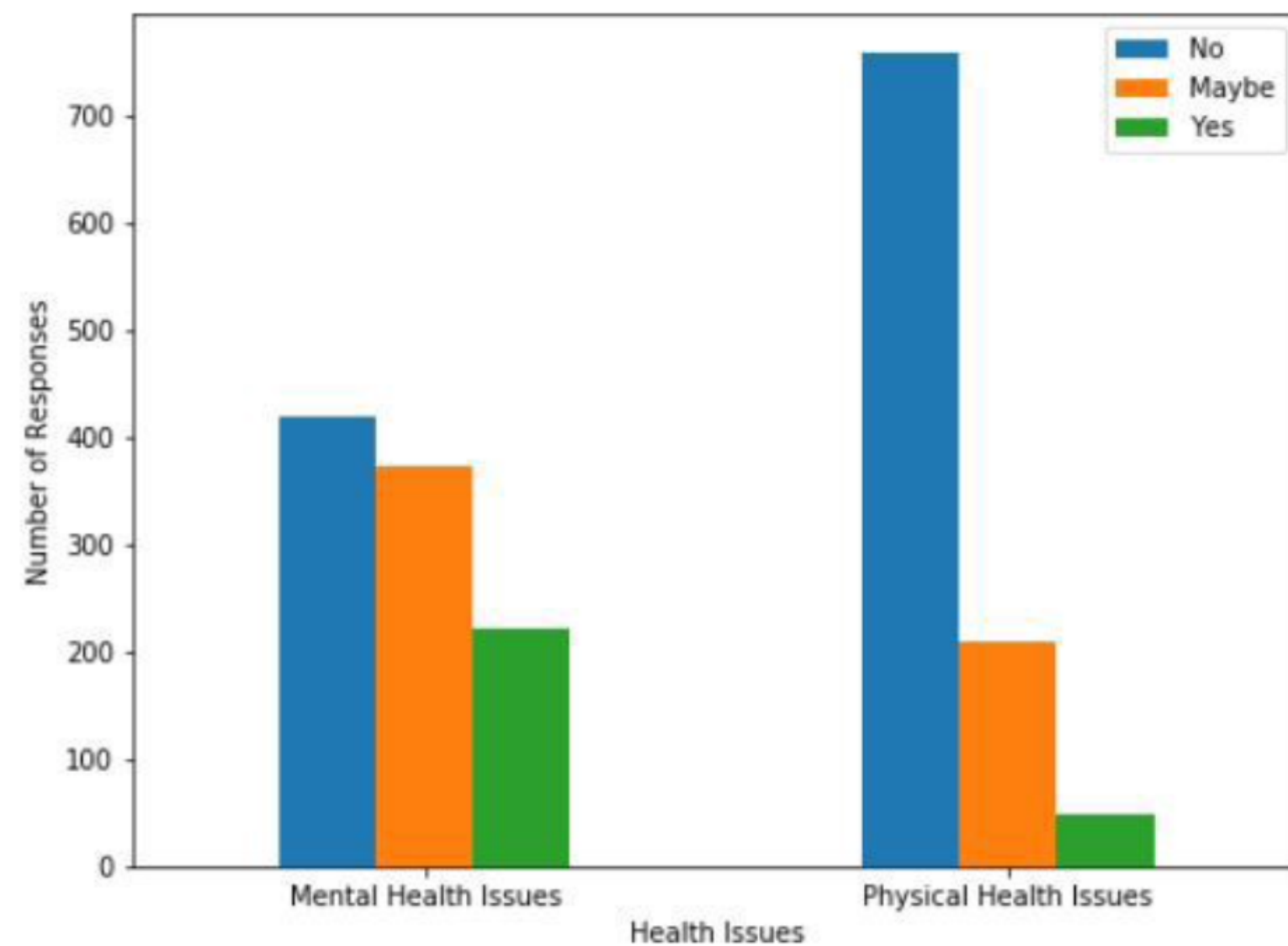
Prevalence of mental health issues does not depend on factors like self-employment, remote work, and organizational size

Mental vs Physical Health

Employers differentiated between the way they dealt with mental and physical health issues

Openness

Employees were insecure about sharing their mental health issues as they feared that these issues can lead to negative consequences, but in reality, observed consequences turn out to be quite low.



Negative consequences of discussing health issues

Machine Learning

Feature Engineering

- Chose features for the model.
- Encoded categoric features.
- Passed encoded data

Test Train Split

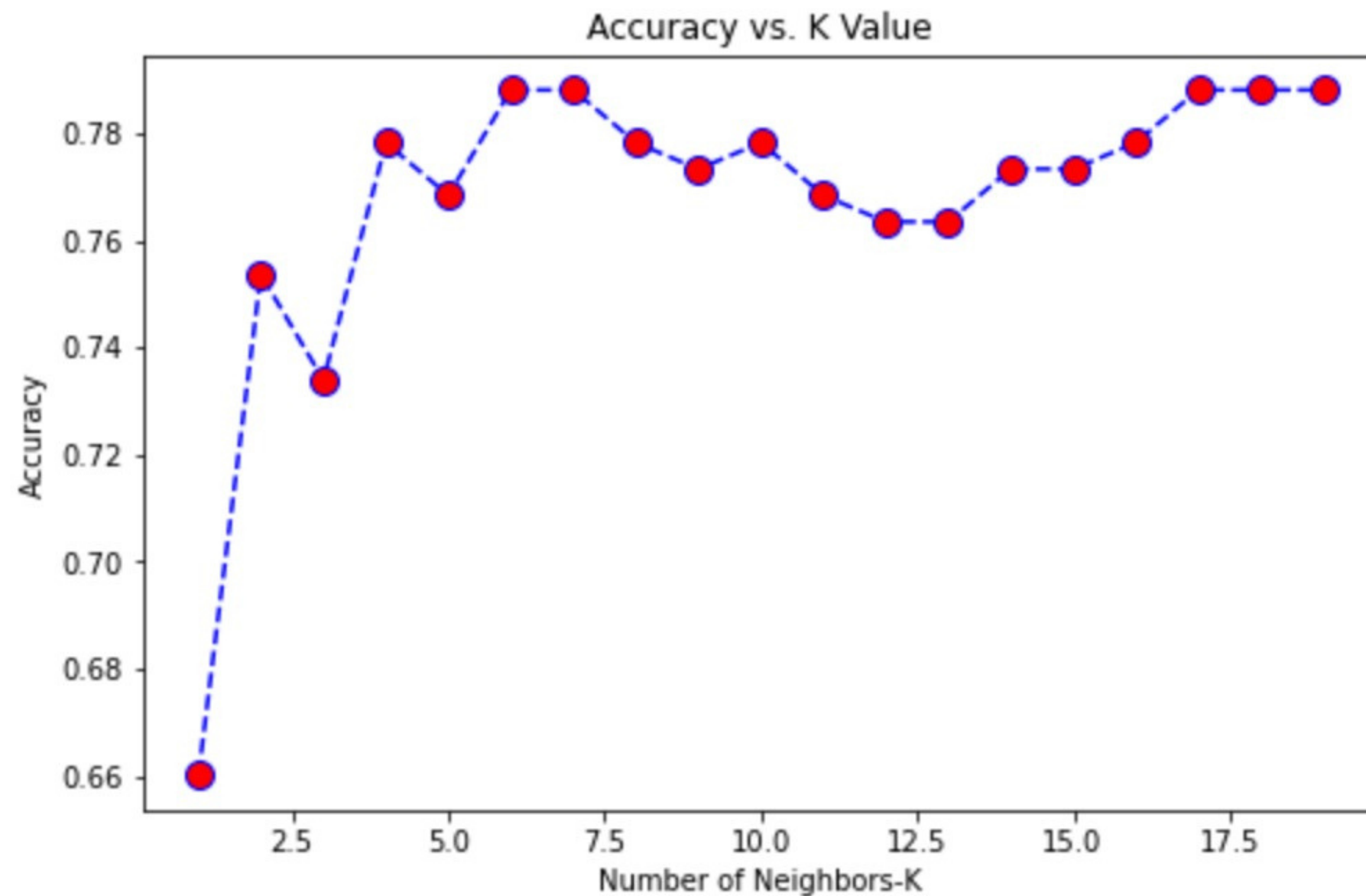
- Split data into train and test.
- Split Ratio: 80: 20.

KNN Classifier

- Trained the classifier.
- Tested the classifier for $k=5$.
- Found optimal value of k

Key Findings

Maximum Accuracy: 78.81773% at K = 6



Classification

A classifier was trained to distinguish between the presence or absence of mental health issues in tech employees based on factors including self-employment, remote-work, age, gender and family-history of mental health issues.

Optimal K Value

Achieved optimum K value of 6 neighbours

Overfitting

Observed a unique trend (overfitting) after a certain k-value (i.e k = 16)

Statistical Inference

Set Hypothesis

- Noted true population prevalence of mental health issues in tech
- Set null and alternate hypothesis

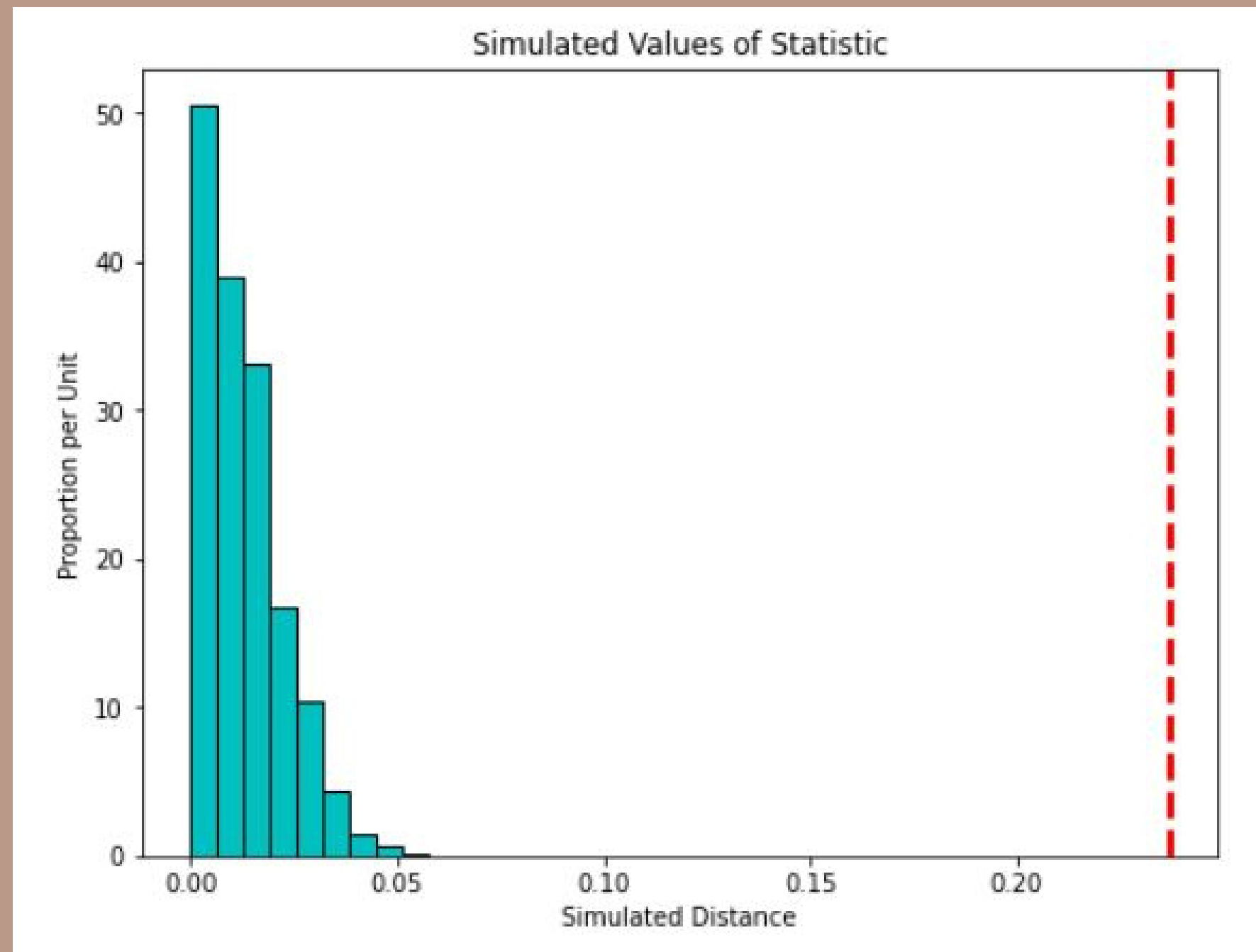
Sampling

- Set test statistic
- Repeated the experiment
- Simulated the test statistic 10000 times

P-value

- Calculated p-value
- Plotted empirical distribution
- Drew inferences

Key Findings



p-value

The p-value turned out to be 0.

Null Hypothesis Rejected

Null Hypothesis was rejected.

Sample vs Population

The prevalence of mental health issues depicted from our data set is not the representative of the entire tech-workspace population.

Observations and Analysis

Biased Dataset

- Demographically Biased
- Age
- Gender

Missing Features

- 2014 Survey
- Potential Questions
- Added in subsequent surveys

Confounding Variable

- Variable corresponding to missing questions.
- US Data
- Factors like Stress, Anxiety etc

Conclusion

- Interestingly, none of the actual work factors had any significant impact on the mental health of employees at least as much as we expected.
- Employers and employees differentiate between the way they deal with mental and physical health issues.
- Employees were insecure about opening up about their mental health issues with others within the organization.
- Our dataset is not a representation of the entire tech-workspace population.



Group Members

AbracaData

Hareem Raza 22100277

Samia Ishaque 22100095

Hassaan Ahmad Waqar 22100136

Syed Muhammad Daniyal Zaidi 22100238

Maneeha Ejaz 22110124

