



# MENTAL HEALTH IN TECH

by:  
Marcelina Oset  
Wiktoria Kuna

# GOALS REMAIN UNCHANGED

We'd like to be able to answer the following questions:

- As a tech industry employee, should I seek treatment?
- Is this company a good fit for my mental health?

# DATA PREPROCESSING

```
##### Genders #####
```

```
['male' 'Male' 'Female' 'female' 'M' 'm' 'Male-ish' 'maile' 'Trans-female'  
'Cis Female' 'F' 'something kinda male?' 'Cis Male' 'Woman' 'f' 'Mal'  
'Male (CIS)' 'queer/she/they' 'non-binary' 'Femake' 'woman' 'Make' 'Nah'  
'Enby' 'fluid' 'Genderqueer' 'Female ' 'Androgyne' 'Agender'  
'cis-female/femme' 'Guy (-ish) ^_^' 'male leaning androgynous' 'Male '  
'Man' 'Trans woman' 'msle' 'Neuter' 'Female (trans)' 'queer'  
'Female (cis)' 'Mail' 'cis male' 'A little about you' 'Malr' 'femail'  
'Cis Man' 'ostensibly male, unsure what that really means']
```

```
##### Missing data summary #####
```

	Missing	Percent of total records
comments	1095	0.869738
state	515	0.409055
work_interfere	264	0.209690
self_employed	18	0.014297
seek_help	0	0.000000
obs_consequence	0	0.000000

Mental Health in Tech Survey from [kaggle.com](https://www.kaggle.com)

# MAIN DATA PROCESSING OPERATIONS

## GENDER

- Grouped similar answers (male, Male, M)
- Added group queer/other for remaining answers

## COMMENTS

- Dropped as 87% of records were empty

## WORK INTERFERENCE

- Replaced NaN with 'Don't know'

## SELF EMPLOYMENT

- Dropped empty records - they were only 1% and assuming any answer could be misleading

## AGE

- Removed answers outside of [0, 100] range - some of the countries don't specify lowest legal working age (or allow any age with given circumstances)



# WORK IN PROGRESS

- Data correlation, distribution visualisation
- Study logistic regression accuracy answering both questions

# STRATEGY

In order to answer questions bothering us we'd want to use following models:

- Logistic Regression
- kNN
- Xgboost
- Random Forest

And later compare their accuracy to choose the most appropriate one.

As for tools:

- PyTorch
- pandas (for data processing)
- seaborn (data visualization)



THANK YOU FOR  
YOUR ATTENTION

by:  
Marcelina Oset  
Wiktorja Kuna