#imports necessary libraries to do basic things on the dataset

Import pandas as pd

Import numpy as np

Import seaborn as sns

Import matplotlib.pyplot as plt

Print('Successfully imported')

Successfully imported

#Reading data

Data = pd.read_csv('/kaggle/input/mental-health-in-tech-survey/survey.csv')

Data.head()

	Timestamp	Age	Gender	Countrystate	self_employed	family_history	treatment	work_interfere	no_employees	…	leave	mental_health_consequence	phys_health_consequence	coworkers	supervisor	mental_health_interview	phys_health_interview	mental_vs_physical	obs_consequence	comments
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	…	Somewhat easy	No	No	Some of them	Yes	No	Maybe	Yes	No	NaN
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	…	Don't know	Maybe	No	No	No	No	No	Don't know	No	NaN
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	…	Somewhat difficult	No	No	Yes	Yes	Yes	Yes	No	No	NaN
3	2014-08-27 11:29:46	31	Male	United Kingdom		NaN	NaN	Yes	Yes	Often	26-100	…	Somewhat difficult	Yes	Yes	Some of them	No	Maybe	Maybe	No	Yes	NaN
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	…	Don't know	No	No	Some of them	Yes	Yes	Yes	Don't know	No	NaN

5 rows × 27 columns

Preprocessing and Cleaning dataset

#Check the dataset for missing data

```
If data.isnull().sum().sum() == 0 :
   Print ('There is no missing data in our dataset')
Else:
   Print('There is {} missing data in our dataset '.format(data.isnull().sum().sum()))
```

There is 1892 missing data in our dataset

```
#Check our missing data from which columns and how many unique features they have.
Frame = pd.concat([data.isnull().sum(), data.nunique(), data.dtypes], axis = 1, sort= False)
Frame
```

	0	1	2
Timestamp	0	1246	object
Age	0	53	int64
Gender	0	49	object
Country	0	48	object
State	515	45	object
Self_employed	18	2	object
Family_history	0	2	object
Treatment	0	2	object
Work_interfere	264	4	object
No_employees	0	6	object
Remote_work	0	2	object
Tech_company	0	2	object
Benefits	0	3	object
Care_options	0	3	object
Wellness_program	0	3	object
Seek_help	0	3	object
Anonymity	0	3	object
Leave	0	5	object
Mental_health_consequence	0	3	object
Phys_health_consequence	0	3	object

Coworkers	0	3	object	
Supervisor	0	3	object	
Mental_health_interview	0	3	object	
Phys_health_interview	0	3	object	
Mental_vs_physical	0	3	object	
Obs_consequence	0	2	object	
Comments	1095	160	object	

#Look at what is in the 'Work_interfere' column to choose a suitable method to fill nan values.

Data['work_interfere'].unique()

Array(['Often', 'Rarely', 'Never', 'Sometimes', nan], dtype=object)

#Plot **work_interfere**

Ax = sns.countplot(data = data , x = 'work_interfere');

#Add the value of each parametr on the Plot

Ax.bar_label(ax.containers[0]);