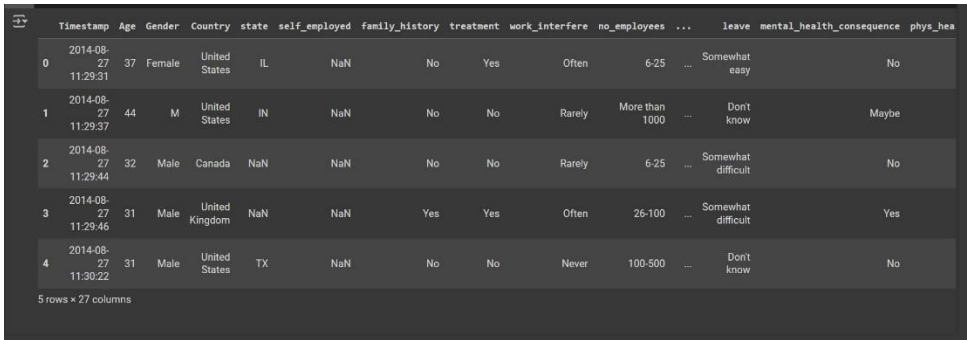


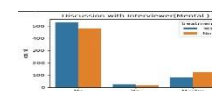
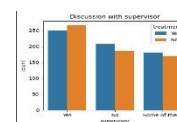
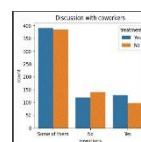
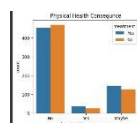
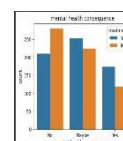
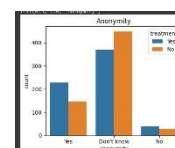
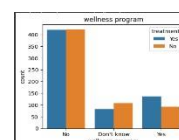
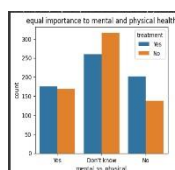
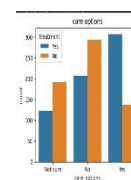
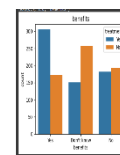
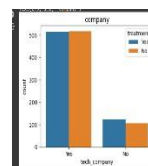
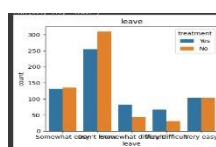
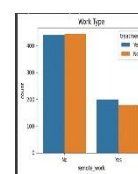
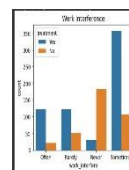
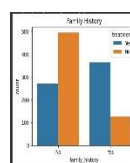
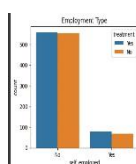
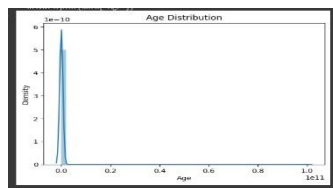
DataCollectionandPreprocessingPhase

Date	15June2024
TeamID	740035
ProjectTitle	Mentalhealthprediction
MaximumMarks	6Marks

DataExplorationandPreprocessingReport

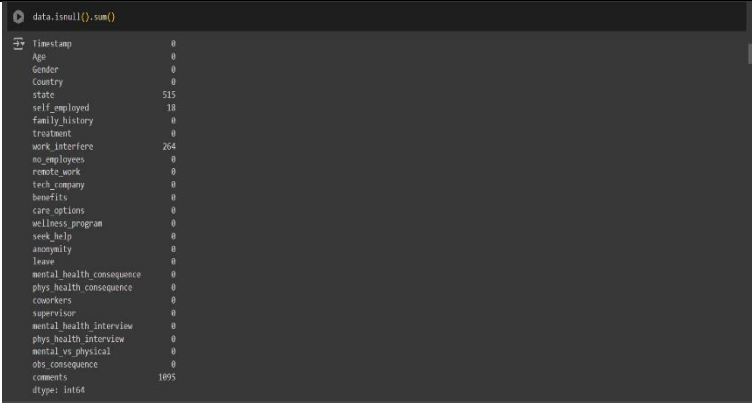
Dataexplorationformentalhealthpredictioninvolvesanalyzingandunderstanding datasetstoidentifypatternsandcorrelationsrelevanttomentalhealthfactors. Preprocessingincludescleaningdata,handlingmissingvalues,andtransforming variablestoprepareforpredictivemodelingandanalysis.

Section	Description
Data Overview	<p><u>Dimension:</u> 5rows*27columns</p> <p><u>Descriptivestatics</u></p> 
Univariate Analysis	



BivariateAnalysis

MultivariateAnalysis																																																																																												
OutliersandAnomalies	-																																																																																											
DataPreprocessingCo eScreenshots																																																																																												
LoadingData	<div><pre>[] data = pd.read_csv("../content/survey.csv")</pre><pre>data.head()</pre><table><thead><tr><th></th><th>Timestamp</th><th>Age</th><th>Gender</th><th>Country</th><th>state</th><th>self_employed</th><th>family_history</th><th>treatment</th><th>work_interfere</th><th>no_employees</th><th>...</th><th>leave</th><th>mental_health_consequence</th><th>phys_hoa</th></tr></thead><tbody><tr><td>0</td><td>2014-08-27 11:29:31</td><td>37</td><td>Female</td><td>United States</td><td>IL</td><td>NaN</td><td>No</td><td>Yes</td><td>Often</td><td>6-25</td><td>...</td><td>Somewhat easy</td><td>No</td><td></td></tr><tr><td>1</td><td>2014-08-27 11:29:37</td><td>44</td><td>M</td><td>United States</td><td>IN</td><td>NaN</td><td>No</td><td>No</td><td>Rarely</td><td>More than 1000</td><td>...</td><td>Don't know</td><td>Maybe</td><td></td></tr><tr><td>2</td><td>2014-08-27 11:29:44</td><td>32</td><td>Male</td><td>Canada</td><td>NaN</td><td>NaN</td><td>No</td><td>No</td><td>Rarely</td><td>6-25</td><td>...</td><td>Somewhat difficult</td><td>No</td><td></td></tr><tr><td>3</td><td>2014-08-27 11:29:46</td><td>31</td><td>Male</td><td>United Kingdom</td><td>NaN</td><td>NaN</td><td>Yes</td><td>Yes</td><td>Often</td><td>26-100</td><td>...</td><td>Somewhat difficult</td><td>Yes</td><td></td></tr><tr><td>4</td><td>2014-08-27 11:30:22</td><td>31</td><td>Male</td><td>United States</td><td>TX</td><td>NaN</td><td>No</td><td>No</td><td>Never</td><td>100-500</td><td>...</td><td>Don't know</td><td>No</td><td></td></tr></tbody></table><p>5 rows x 27 columns</p></div>			Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave	mental_health_consequence	phys_hoa	0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...	Somewhat easy	No		1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...	Don't know	Maybe		2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	...	Somewhat difficult	No		3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	...	Somewhat difficult	Yes		4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...	Don't know	No	
	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave	mental_health_consequence	phys_hoa																																																																														
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...	Somewhat easy	No																																																																															
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...	Don't know	Maybe																																																																															
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	...	Somewhat difficult	No																																																																															
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	...	Somewhat difficult	Yes																																																																															
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...	Don't know	No																																																																															

HandlingMissingData	 <pre>data.isnull().sum() Timestamp 0 Age 0 Gender 0 Country 0 state 515 self_employed 18 family_history 0 treatment 0 work_interfere 264 no_employees 0 remote_work 0 tech_company 0 benefits 0 care_options 0 wellness_program 0 seek_help 0 anonymity 0 leave 0 mental_health_consequence 0 phys_health_consequence 0 coworkers 0 supervisor 0 mental_health_interview 0 phys_health_interview 0 mental_vs_physical 0 obs_consequence 0 comments 1895 dtype: int64</pre>
---------------------	--

DataTransformation	
FeatureEngineering	Attachedthecodesinfinalsubmission.
SaveProcessedData	-