IMPORT DATA

```
import tensorflow as tf
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
import seaborn as sns
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
import sklearn
```

READING DATA

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
df1=pd.read_csv("mental-and-substance-use-as-share-of-disease.csv")
df2=pd.read_csv("prevalence-by-mental-and-substance-use-disorder.csv")
```

```
df1.head()
```

	Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Mental disorders - Sex: Both - Age: All Ages (Percent)
0	Afghanistan	AFG	1990	1.696670
1	Afghanistan	AFG	1991	1.734281
2	Afghanistan	AFG	1992	1.791189
3	Afghanistan	AFG	1993	1.776779
4	Afghanistan	AFG	1994	1.712986

```
df2.head()
```

	Entity	Code	Year	Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)	Preval Depr disor Sex: Age standa (Pe
0	Afghanistan	AFG	1990	0.228979	0.721207	0.131001	4.835127	0.454202	5
1	Afghanistan	AFG	1991	0.228120	0.719952	0.126395	4.821765	0.447112	5
2	Afghanistan	AFG	1992	0.227328	0.718418	0.121832	4.801434	0.441190	5
3	Afghanistan	AFG	1993	0.226468	0.717452	0.117942	4.789363	0.435581	5

```
df1.describe(),df1.info()
df2.describe(),df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6840 entries, 0 to 6839
Data columns (total 4 columns):
 #   Column                                                                                Non-Null Count  Dtype
---  ------                                                                                --------------  -----
 0   Entity                                                                                6840 non-null   object
 1   Code                                                                                  6150 non-null   object
 2   Year                                                                                  6840 non-null   int64
 3   DALYs (Disability-Adjusted Life Years) - Mental disorders - Sex: Both - Age: All Ages (Percent)  6840 non-null   float64
dtypes: float64(1), int64(1), object(2)
memory usage: 213.9+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6840 entries, 0 to 6839
Data columns (total 10 columns):
```

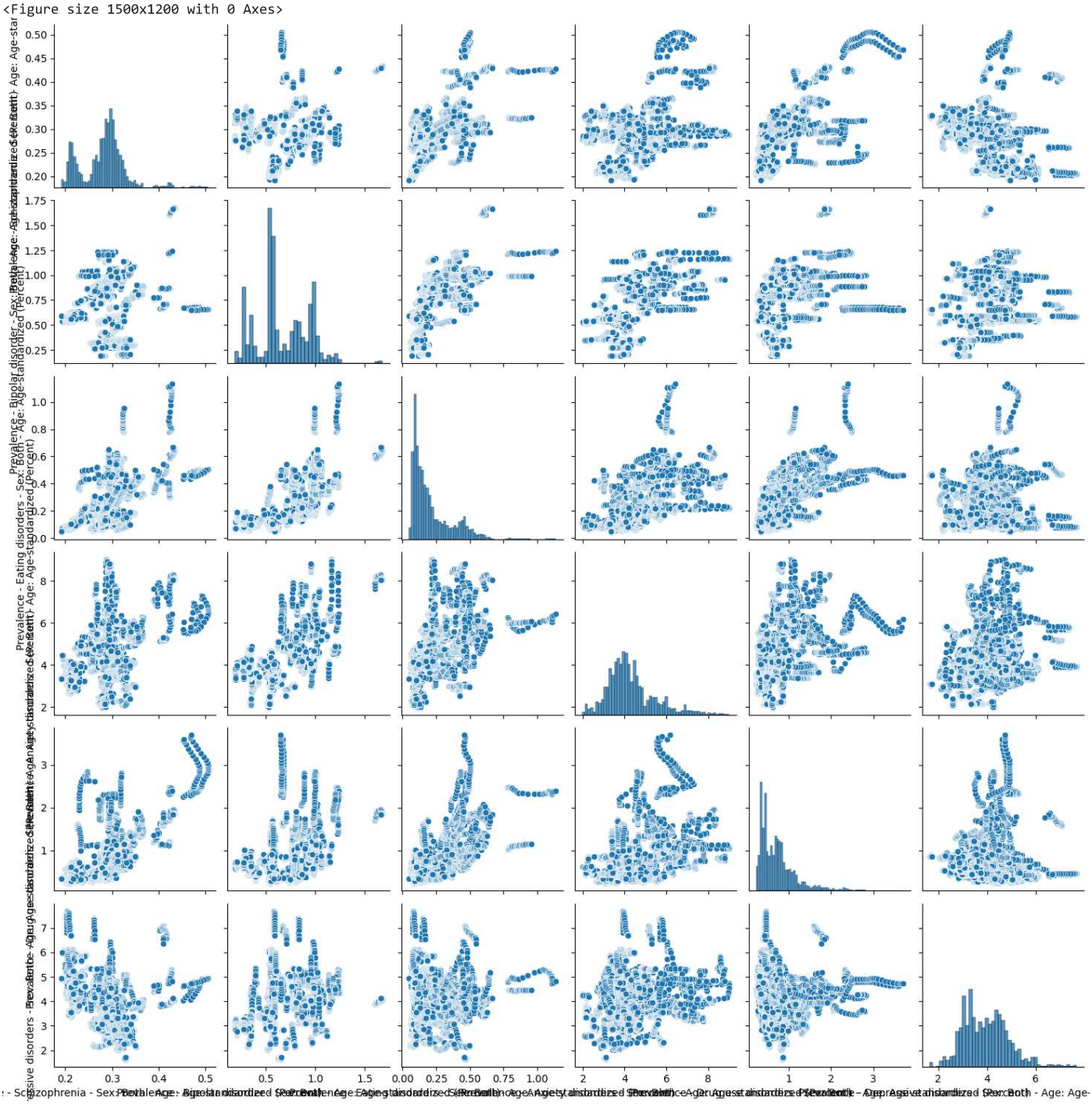
```
 #   Column                                                                        Non-Null Count  Dtype
---  ------                                                                        --------------  -----
 0   Entity                                                                        6840 non-null   object
 1   Code                                                                          6150 non-null   object
 2   Year                                                                          6840 non-null   int64
 3   Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)      6840 non-null   float64
 4   Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)   6840 non-null   float64
 5   Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)   6840 non-null   float64
 6   Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)  6840 non-null   float64
 7   Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent) 6840 non-null   float64
 8   Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)  6840 non-null  float64
 9   Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)  6840 non-null  float64
dtypes: float64(7), int64(1), object(2)
memory usage: 534.5+ KB
(              Year  \
 count  6840.000000
 mean   2004.500000
 std       8.656074
 min    1990.000000
 25%    1997.000000
 50%    2004.500000
 75%    2012.000000
 max    2019.000000

        Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)  \
 count                                       6840.000000
 mean                                          0.281167
 std                                           0.047561
 min                                           0.191621
 25%                                           0.255468
 50%                                           0.287456
 75%                                           0.304760
 max                                           0.506018

        Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)  \
 count                                       6840.000000
 mean                                          0.673891
 std                                           0.258594
 min                                           0.189344
 25%                                           0.539791
 50%                                           0.591893
 75%                                           0.897248
 max                                           1.676204
```

```
df=pd.concat(objs=[df2,df1],axis=1)
```

DATA VISUALIZATION

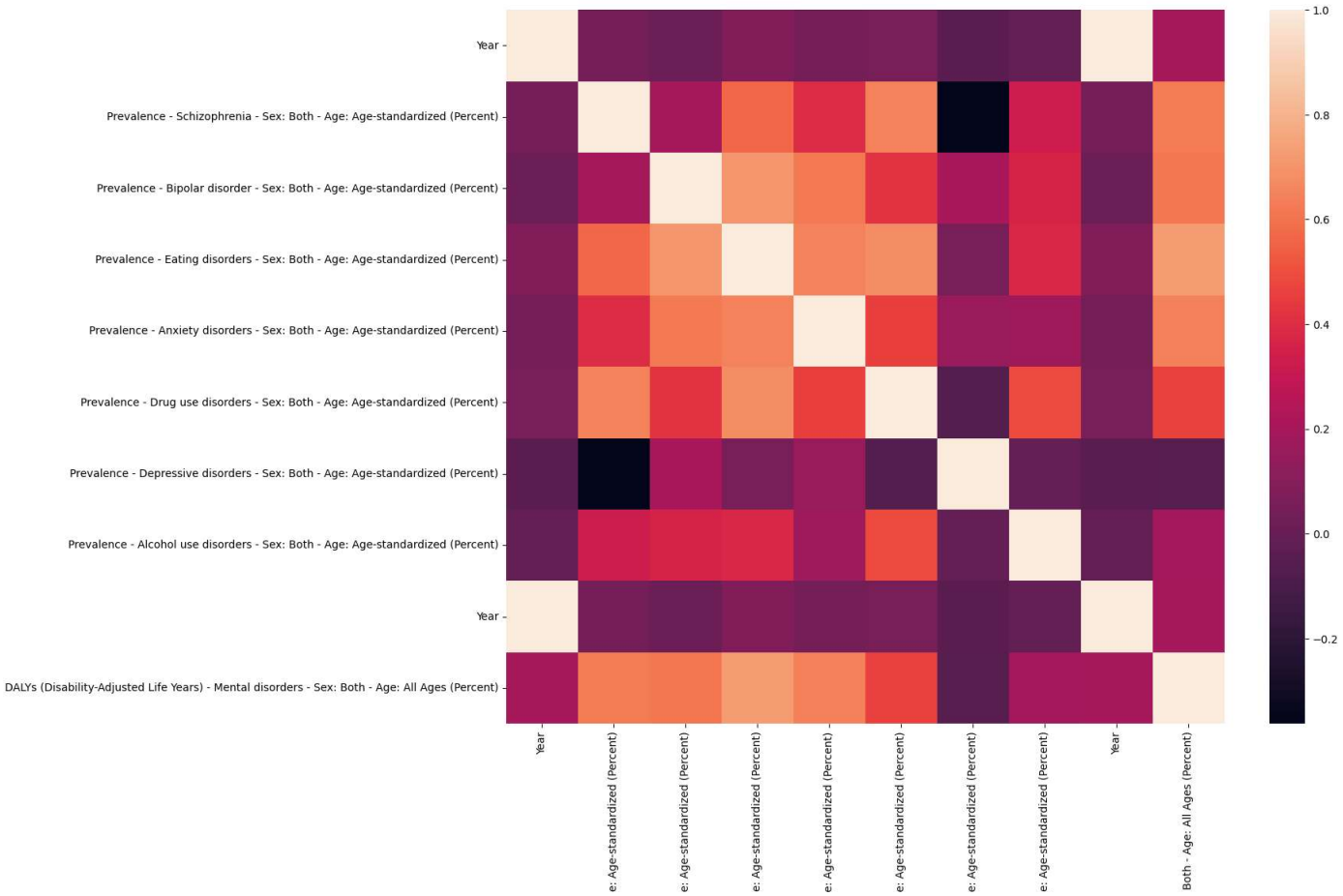
```
corr=df.corr()
plt.figure(figsize=(15,12))
sns.pairplot(df[['Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)',]])
```

```
<ipython-input-12-910495745933>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version
  corr=df.corr()
<seaborn.axisgrid.PairGrid at 0x7fb92102f040>
<Figure size 1500x1200 with 0 Axes>
```



```
plt.figure(figsize=(15,12))
sns.heatmap(corr)
```

```
<Axes: >
```



DATA PREPROCESSING

```
df.drop(['Entity','Code','Year'],axis=1,inplace=True)
df=df.fillna(df.mean())

x=df[['Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)',
       'Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)',]].to_numpy()

y=df[['DALYs (Disability-Adjusted Life Years) - Mental disorders - Sex: Both - Age: All Ages (Percent)']].to_numpy()


scaler=StandardScaler()
x=scaler.fit_transform(x)

x_train,x_test,y_train,y_test=train_test_split(x,y)
```

ML IMPLEMENTATION

```
ml=RandomForestRegressor()
ml.fit(x_train,y_train)
predicted_values=ml.predict(x_test)
```

```
<ipython-input-16-bfcd872d59c9>:2: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the s
  ml.fit(x_train,y_train)
```

MODEL EVALUATION AND METRICS

```
plt.figure(figsize=(15,12))
plt.plot(y_test[:100])
plt.plot(predicted_values[:100])
plt.legend(['true','predicted'])
plt.title('Mean Square Error '+str(sklearn.metrics.mean_squared_error(y_test,predicted_values)))
plt.show()
```