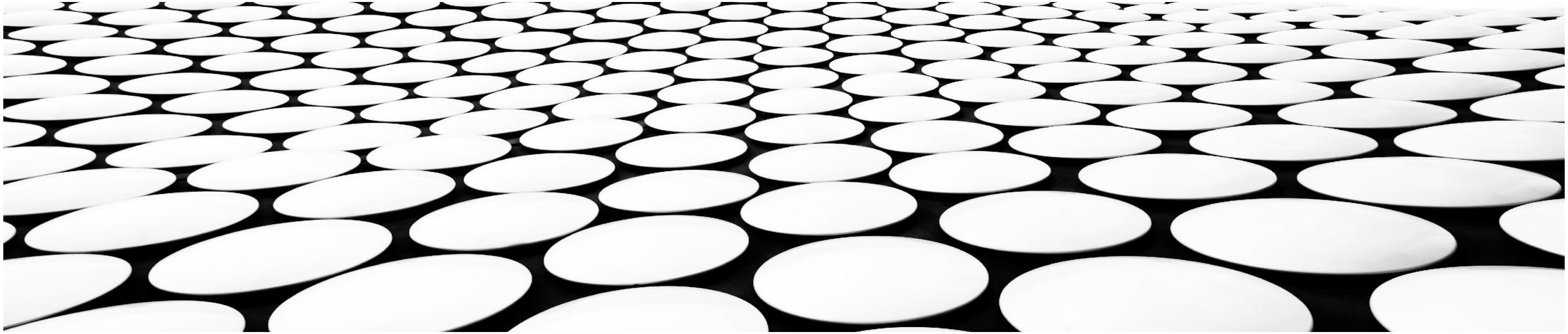
STUDENT DETAILS

NAME : GARIMA SHARMA
SKILLSBUILD EMAIL ID : 502garima@gmail.com
STUDENT ID AICTE : STU61644435acd731633961013
COLLEGE NAME : DAYALBAGH EDUCATIONAL INSTITUTE
COLLEGE STATE : UTTAR PRADESH
INTERNSHIP DOMAIN AND INTERNSHIP START AND END DATE : ARTIFICIAL INTELLIGENCE / 12.06.2023 – 24.07.2023



PROJECT TITLE/PROBLEM STATEMENT

PROJECT TITLE : AI MENTAL FITNESS TRACKER

PROBLEM STATEMENT : Mental health is a crucial aspect of overall well-being, and with the increasing prevalence of mental health disorders globally, it has become essential to develop effective tools for monitoring and improving mental fitness. Existing self-assessment methods often rely on subjective reporting, which can be prone to biases and inaccuracies. Therefore, there is a need for a reliable and objective system that can track and evaluate an individual's mental health status using advanced technologies like Artificial Intelligence (AI).

AGENDA

MENTAL FITNESS TRACKER USING REGRESSION

PROJECT OVERVIEW

The project overview above provides a brief summary of the AI Mental Fitness Tracker using regression, outlining its objectives, methodology, expected outcomes, and potential impact. The actual implementation of the project may vary depending on available resources, data availability, and specific requirements.

WHO ARE THE END USERS OF THIS PROJECT?

Regression-based AI Mental Fitness Tracker's end customers may include those who want to keep track of and advance their mental fitness. By offering precise assessments, individualized recommendations, and ongoing monitoring, the project seeks to empower people to actively manage their mental health.

YOUR SOLUTION AND ITS VALUE PROPOSITION

Solution

The solution is an AI-powered Mental Fitness Tracker that utilizes regression analysis to accurately assess and monitor an individual's mental fitness. By analyzing various factors such as mood, stress levels, sleep patterns, physical activity, and social interactions, our tracker provides a comprehensive and data-driven evaluation of mental well-being.

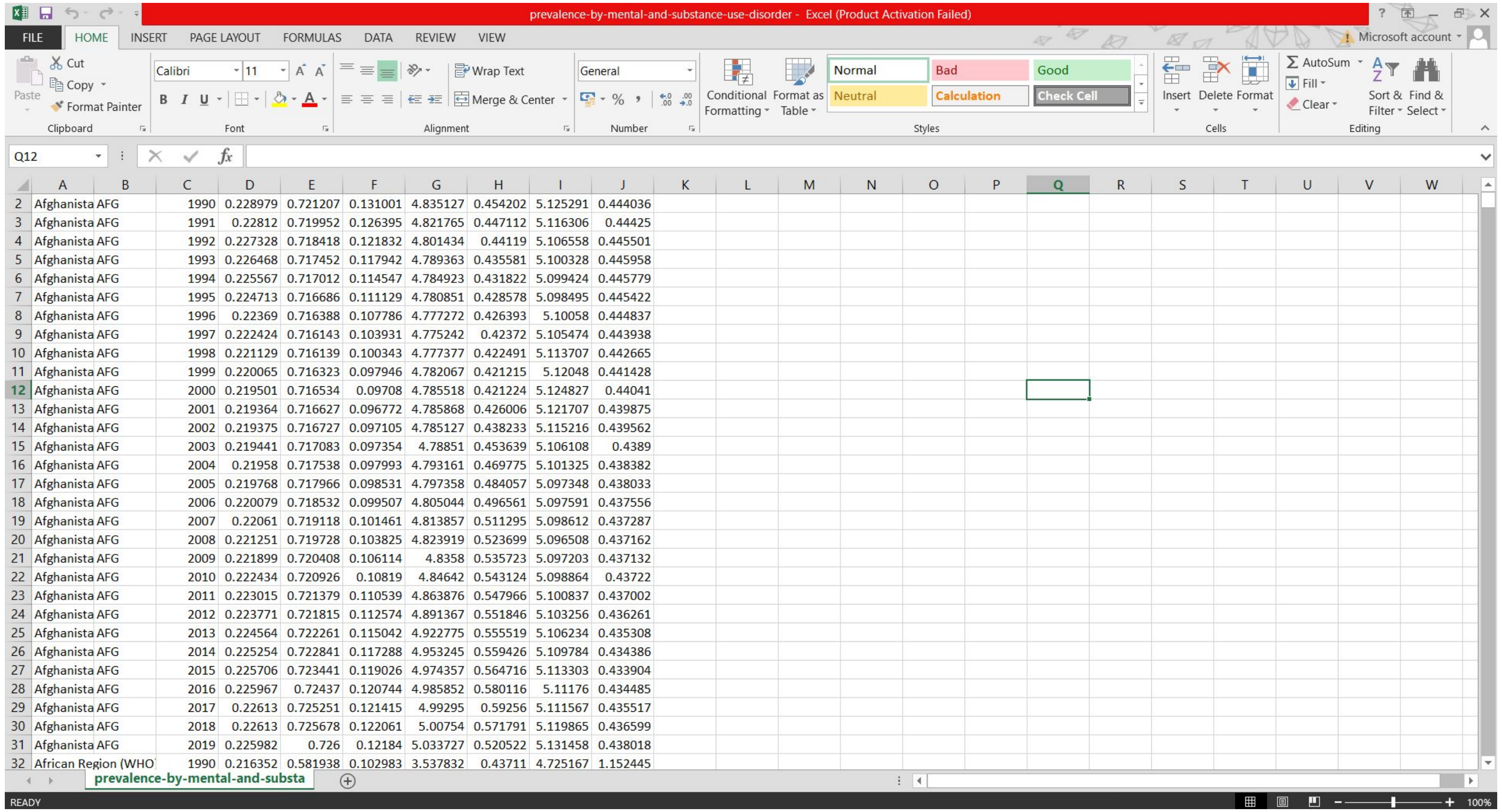
Value Proposition

Empowering Individuals: Our solution empowers individuals to take an active role in managing their mental health. By providing accurate assessments, continuous monitoring, and personalized recommendations, it enables users to make informed decisions and proactively improve their mental well-being. Early Intervention and Prevention: The AI Mental Fitness Tracker facilitates early intervention by detecting potential signs of mental health decline. By identifying patterns and changes in mental fitness levels, individuals can take proactive measures to prevent the onset or escalation of mental health issues.

DATASET OF MENTAL FITNESS TRACKER ANALYSIS AND PREDICTION

FIRST DATASET

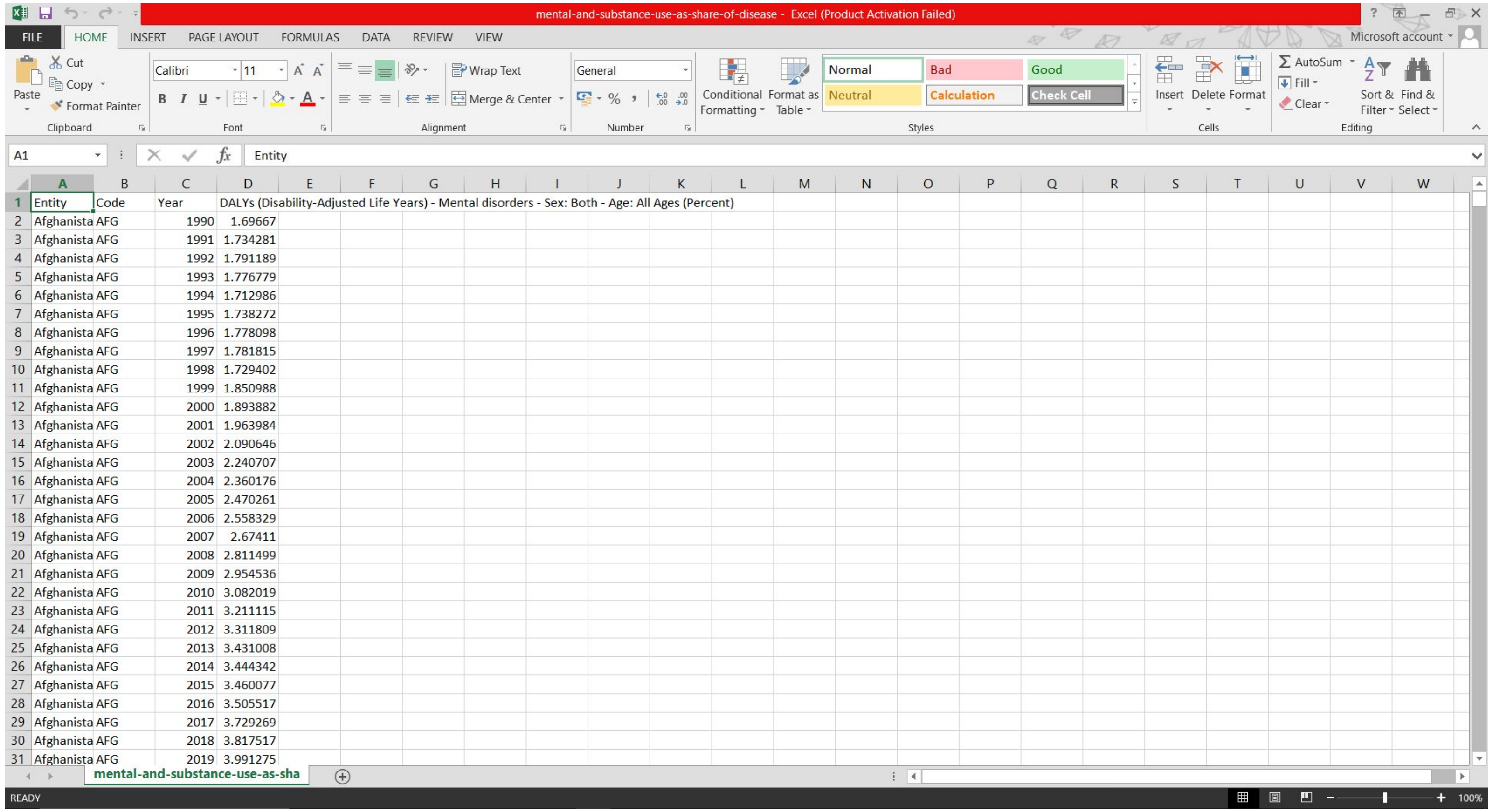
- https://drive.google.com/file/d/1B14pHI1jpVKSkaB0qJ59Bs55c2QjZYA2/view?usp=drive_link



	A	B	C	D	E	F	G	H	I	J
2	Afghanista	AFG	1990	0.228979	0.721207	0.131001	4.835127	0.454202	5.125291	0.444036
3	Afghanista	AFG	1991	0.22812	0.719952	0.126395	4.821765	0.447112	5.116306	0.44425
4	Afghanista	AFG	1992	0.227328	0.718418	0.121832	4.801434	0.44119	5.106558	0.445501
5	Afghanista	AFG	1993	0.226468	0.717452	0.117942	4.789363	0.435581	5.100328	0.445958
6	Afghanista	AFG	1994	0.225567	0.717012	0.114547	4.784923	0.431822	5.099424	0.445779
7	Afghanista	AFG	1995	0.224713	0.716686	0.111129	4.780851	0.428578	5.098495	0.445422
8	Afghanista	AFG	1996	0.22369	0.716388	0.107786	4.777272	0.426393	5.10058	0.444837
9	Afghanista	AFG	1997	0.222424	0.716143	0.103931	4.775242	0.42372	5.105474	0.443938
10	Afghanista	AFG	1998	0.221129	0.716139	0.100343	4.777377	0.422491	5.113707	0.442665
11	Afghanista	AFG	1999	0.220065	0.716323	0.097946	4.782067	0.421215	5.12048	0.441428
12	Afghanista	AFG	2000	0.219501	0.716534	0.09708	4.785518	0.421224	5.124827	0.44041
13	Afghanista	AFG	2001	0.219364	0.716627	0.096772	4.785868	0.426006	5.121707	0.439875
14	Afghanista	AFG	2002	0.219375	0.716727	0.097105	4.785127	0.438233	5.115216	0.439562
15	Afghanista	AFG	2003	0.219441	0.717083	0.097354	4.78851	0.453639	5.106108	0.4389
16	Afghanista	AFG	2004	0.21958	0.717538	0.097993	4.793161	0.469775	5.101325	0.438382
17	Afghanista	AFG	2005	0.219768	0.717966	0.098531	4.797358	0.484057	5.097348	0.438033
18	Afghanista	AFG	2006	0.220079	0.718532	0.099507	4.805044	0.496561	5.097591	0.437556
19	Afghanista	AFG	2007	0.22061	0.719118	0.101461	4.813857	0.511295	5.098612	0.437287
20	Afghanista	AFG	2008	0.221251	0.719728	0.103825	4.823919	0.523699	5.096508	0.437162
21	Afghanista	AFG	2009	0.221899	0.720408	0.106114	4.8358	0.535723	5.097203	0.437132
22	Afghanista	AFG	2010	0.222434	0.720926	0.10819	4.84642	0.543124	5.098864	0.43722
23	Afghanista	AFG	2011	0.223015	0.721379	0.110539	4.863876	0.547966	5.100837	0.437002
24	Afghanista	AFG	2012	0.223771	0.721815	0.112574	4.891367	0.551846	5.103256	0.436261
25	Afghanista	AFG	2013	0.224564	0.722261	0.115042	4.922775	0.555519	5.106234	0.435308
26	Afghanista	AFG	2014	0.225254	0.722841	0.117288	4.953245	0.559426	5.109784	0.434386
27	Afghanista	AFG	2015	0.225706	0.723441	0.119026	4.974357	0.564716	5.113303	0.433904
28	Afghanista	AFG	2016	0.225967	0.72437	0.120744	4.985852	0.580116	5.11176	0.434485
29	Afghanista	AFG	2017	0.22613	0.725251	0.121415	4.99295	0.59256	5.111567	0.435517
30	Afghanista	AFG	2018	0.22613	0.725678	0.122061	5.00754	0.571791	5.119865	0.436599
31	Afghanista	AFG	2019	0.225982	0.726	0.12184	5.033727	0.520522	5.131458	0.438018
32	African Region (WHO)		1990	0.216352	0.581938	0.102983	3.537832	0.43711	4.725167	1.152445

SECOND DATASET

- https://drive.google.com/file/d/1BHotRwLT2xVQEHTdiDY3Sj6P_3g7489L/view?usp=drive_link



Entity	Code	Year	DALYs (Disability-Adjusted Life Years) - Mental disorders - Sex: Both - Age: All Ages (Percent)
Afghanista	AFG	1990	1.69667
Afghanista	AFG	1991	1.734281
Afghanista	AFG	1992	1.791189
Afghanista	AFG	1993	1.776779
Afghanista	AFG	1994	1.712986
Afghanista	AFG	1995	1.738272
Afghanista	AFG	1996	1.778098
Afghanista	AFG	1997	1.781815
Afghanista	AFG	1998	1.729402
Afghanista	AFG	1999	1.850988
Afghanista	AFG	2000	1.893882
Afghanista	AFG	2001	1.963984
Afghanista	AFG	2002	2.090646
Afghanista	AFG	2003	2.240707
Afghanista	AFG	2004	2.360176
Afghanista	AFG	2005	2.470261
Afghanista	AFG	2006	2.558329
Afghanista	AFG	2007	2.67411
Afghanista	AFG	2008	2.811499
Afghanista	AFG	2009	2.954536
Afghanista	AFG	2010	3.082019
Afghanista	AFG	2011	3.211115
Afghanista	AFG	2012	3.311809
Afghanista	AFG	2013	3.431008
Afghanista	AFG	2014	3.444342
Afghanista	AFG	2015	3.460077
Afghanista	AFG	2016	3.505517
Afghanista	AFG	2017	3.729269
Afghanista	AFG	2018	3.817517
Afghanista	AFG	2019	3.991275

HOW DID YOU CUSTOMIZE THE PROJECT AND MAKE IT YOUR OWN

- Perform Exploratory Data Analysis and after that perform Random Forest and Linear Regression to make model more efficient.

MODELLING

1. Gather pertinent information about indicators of mental health, such as mood, stress levels, sleep habits, activity levels, and social interactions. Self-reporting, wearable technology, questionnaires, or other sources are all possible ways to get this information.
2. Data preprocessing: To assure the acquired data's quality and consistency, clean and preprocess it. Deal with missing values, outliers, and any other data quality problems that can impair the models' accuracy.
3. Identifying and extracting significant features that are a good indicator of a person's degree of mental fitness from the obtained data is known as feature engineering. In this step, pertinent variables are chosen, variables may need to be transformed, and new features that capture crucial elements of mental health are created.
4. Depending on the nature of the data and the issue at hand, select the best regression models. Regression models that are often used include those that use linear, polynomial, multivariate, or more complex methods like ridge or support vector regression.

CONTD…

5. Split the preprocessed data into training and validation sets before starting to train the models. The chosen regression models should be trained by fitting the data to them using the training set. Throughout this process, the models pick up on the connections between the input features and the goal variable (mental fitness level).
6. Model Evaluation: Use appropriate evaluation measures, such as mean squared error (MSE), R-squared value, or root mean squared error (RMSE), to gauge the effectiveness of the trained regression models. Analyze the models' accuracy in predicting a person's degree of mental fitness using the provided features.
7. Model refinement: To enhance the performance of the models, fine-tune them by modifying their hyper parameters or investigating various regression procedures. In this iterative process, many model configurations are tested, and the model with the greatest performance is chosen based on the assessment metrics.
8. Deploy the models in a system that enables ongoing monitoring of mental health indicators once they have been trained and validated. The models should be able to quickly analyze fresh data, spot patterns, and offer up-to-date evaluations of a person's level of mental fitness.

CODE

Step 1 :- Import Required Libraries

```
!pip install matplotlib-venn
```

```
!apt-get -qq install -y libfluidsynth1
```

```
E: Package 'libfluidsynth1' has no installation candidate
```

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
```

Step 2 :- Load and Prepare the Data

```
df1 = pd.read_csv("/content/drive/MyDrive/IBM SkillBuild Internship Project/prevalence-by-mental-and-substance-use-disorder.csv")


df2 = pd.read_csv("/content/drive/MyDrive/IBM SkillBuild Internship Project/mental-and-substance-use-as-share-of-disease.csv")
```

Step 3 :- Merging both the Dataset

```
data = pd.merge(df1,df2)

data.head(10)
```

	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)	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)	DALYs (Disability-Adjusted Life Years) - Mental disorders - Sex: Both - Age: All Ages (Percent)
0	Afghanistan	AFG	1990	0.228979	0.721207	0.131001	4.835127	0.454202	5.125291	0.444036	1.696670
1	Afghanistan	AFG	1991	0.228120	0.719952	0.126395	4.821765	0.447112	5.116306	0.444250	1.734281
2	Afghanistan	AFG	1992	0.227328	0.718418	0.121832	4.801434	0.441190	5.106558	0.445501	1.791189
3	Afghanistan	AFG	1993	0.226468	0.717452	0.117942	4.789363	0.435581	5.100328	0.445958	1.776779
4	Afghanistan	AFG	1994	0.225567	0.717012	0.114547	4.784923	0.431822	5.099424	0.445779	1.712986
5	Afghanistan	AFG	1995	0.224713	0.716686	0.111129	4.780851	0.428578	5.098495	0.445422	1.738272
6	Afghanistan	AFG	1996	0.223690	0.716388	0.107786	4.777272	0.426393	5.100580	0.444837	1.778098
7	Afghanistan	AFG	1997	0.222424	0.716143	0.103931	4.775242	0.423720	5.105474	0.443938	1.781815
8	Afghanistan	AFG	1998	0.221129	0.716139	0.100343	4.777377	0.422491	5.113707	0.442665	1.729402
9	Afghanistan	AFG	1999	0.220065	0.716323	0.097946	4.782067	0.421215	5.120480	0.441428	1.850988

Step 4 :- Data Cleaning

```
1 data.isnull().sum()
```

```
Entity                                                                                     0
Code                                                                                     690
Year                                                                                       0
Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)                   0
Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)                0
Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)                0
Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)               0
Prevalence - Drug use disorders - Sex: Both - Age: Age-standardized (Percent)              0
Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)            0
Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)           0
DALYs (Disability-Adjusted Life Years) - Mental disorders - Sex: Both - Age: All Ages (Percent)    0
dtype: int64
```

Step 4 :- Data Cleaning

```
data.drop('Code',axis=1,inplace=True)
```

```
data.head(10)
```

	Entity	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)	Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)	Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)	DALYs (Disability-Adjusted Life Years) - Mental disorders - Sex: Both - Age: All Ages (Percent)
0	Afghanistan	1990	0.228979	0.721207	0.131001	4.835127	0.454202	5.125291	0.444036	1.696670
1	Afghanistan	1991	0.228120	0.719952	0.126395	4.821765	0.447112	5.116306	0.444250	1.734281
2	Afghanistan	1992	0.227328	0.718418	0.121832	4.801434	0.441190	5.106558	0.445501	1.791189
3	Afghanistan	1993	0.226468	0.717452	0.117942	4.789363	0.435581	5.100328	0.445958	1.776779
4	Afghanistan	1994	0.225567	0.717012	0.114547	4.784923	0.431822	5.099424	0.445779	1.712986
5	Afghanistan	1995	0.224713	0.716686	0.111129	4.780851	0.428578	5.098495	0.445422	1.738272
6	Afghanistan	1996	0.223690	0.716388	0.107786	4.777272	0.426393	5.100580	0.444837	1.778098
7	Afghanistan	1997	0.222424	0.716143	0.103931	4.775242	0.423720	5.105474	0.443938	1.781815
8	Afghanistan	1998	0.221129	0.716139	0.100343	4.777377	0.422491	5.113707	0.442665	1.729402
9	Afghanistan	1999	0.220065	0.716323	0.097946	4.782067	0.421215	5.120480	0.441428	1.850988

```
data.size,data.shape
```

```
(68400, (6840, 10))
```

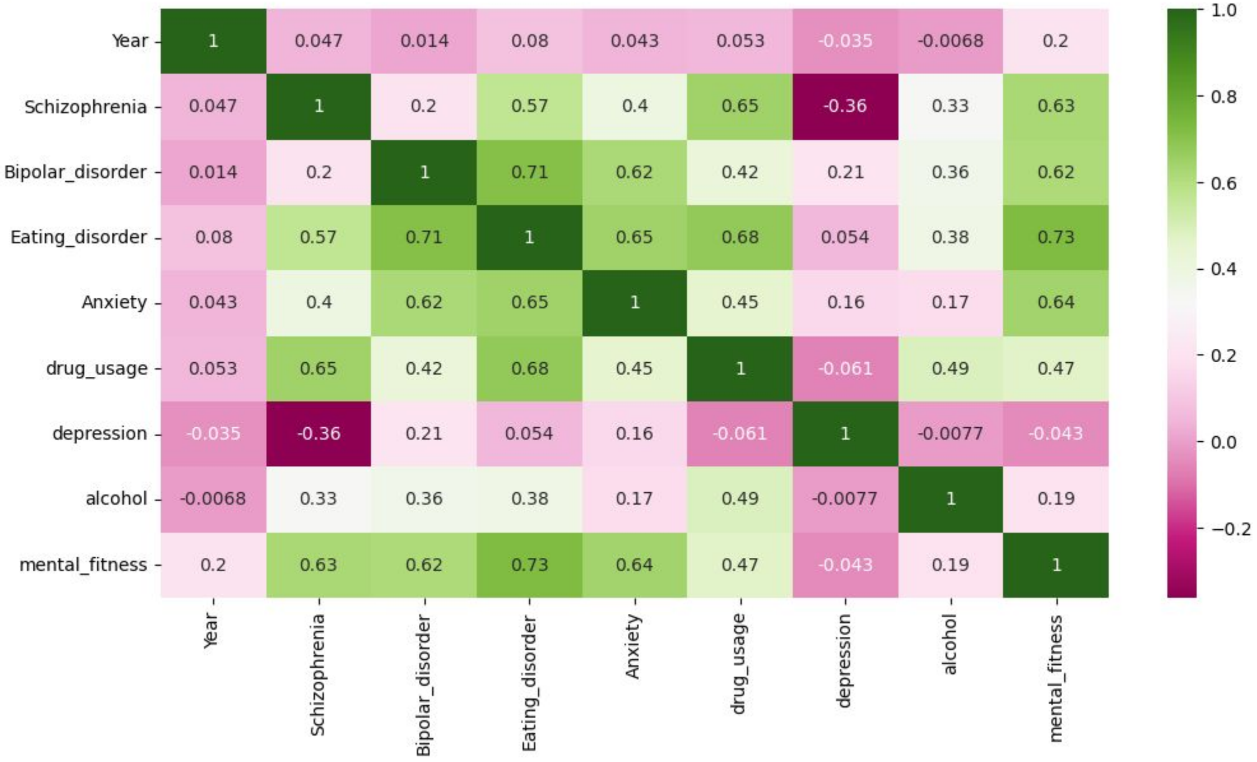
```
data.set_axis(['Country','Year','Schizophrenia', 'Bipolar_disorder', 'Eating_disorder','Anxiety','drug_usage','depression','alcohol','mental_fitness'], axis='columns', inplace=True)
```

Step 5 :- Exploratory Analysis

1. Heap Map

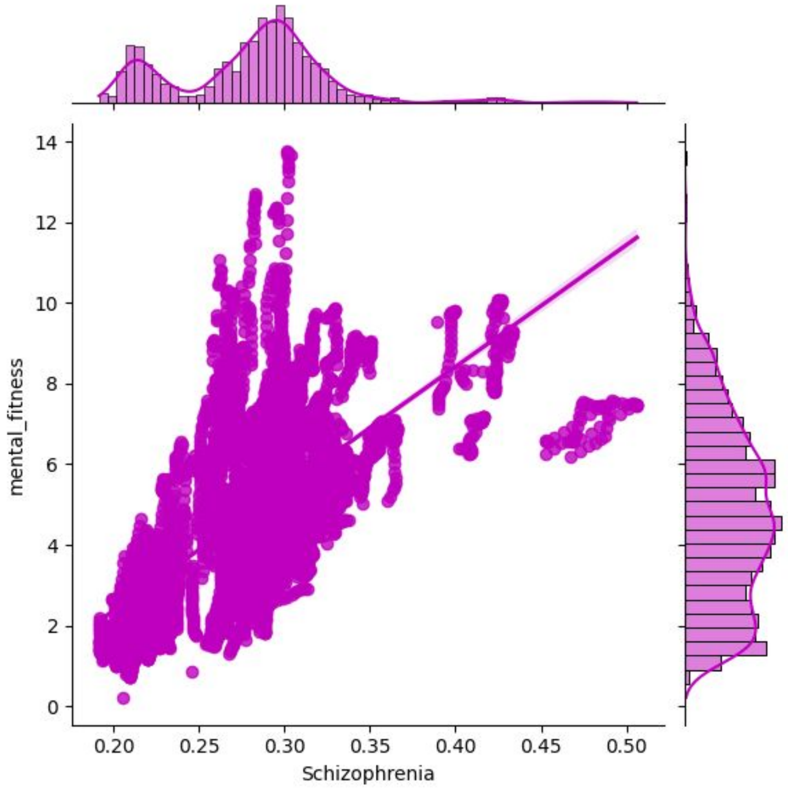
```
plt.figure(figsize=(12,6))
sns.heatmap(data.corr(),annot=True,cmap='PiYG')
plt.plot()
```

[]



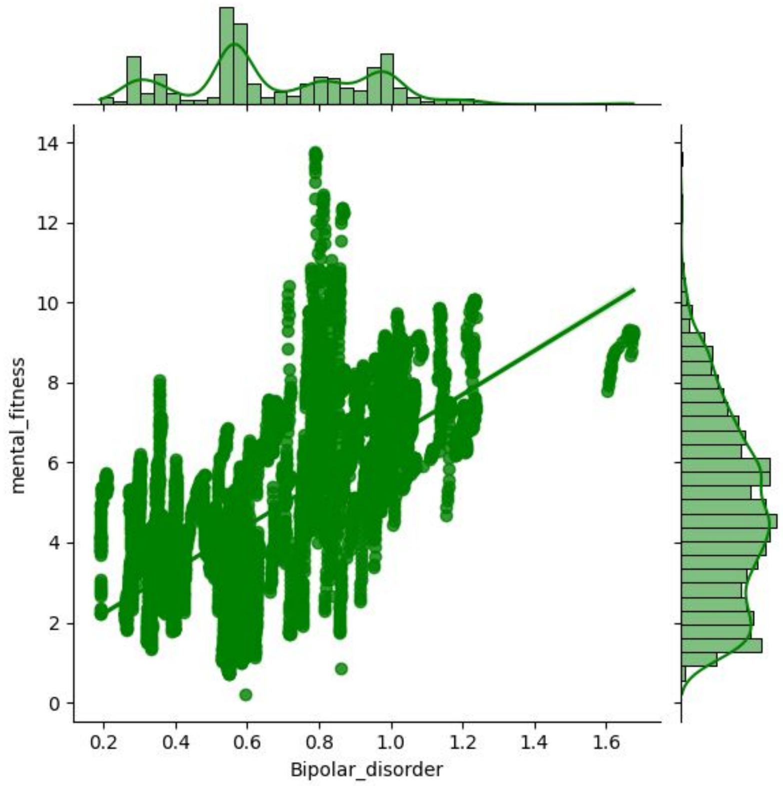
2. Joint Plot

```
sns.jointplot(data,x='Schizophrenia',y='mental_fitness',kind='reg',color='m')
plt.show()
```



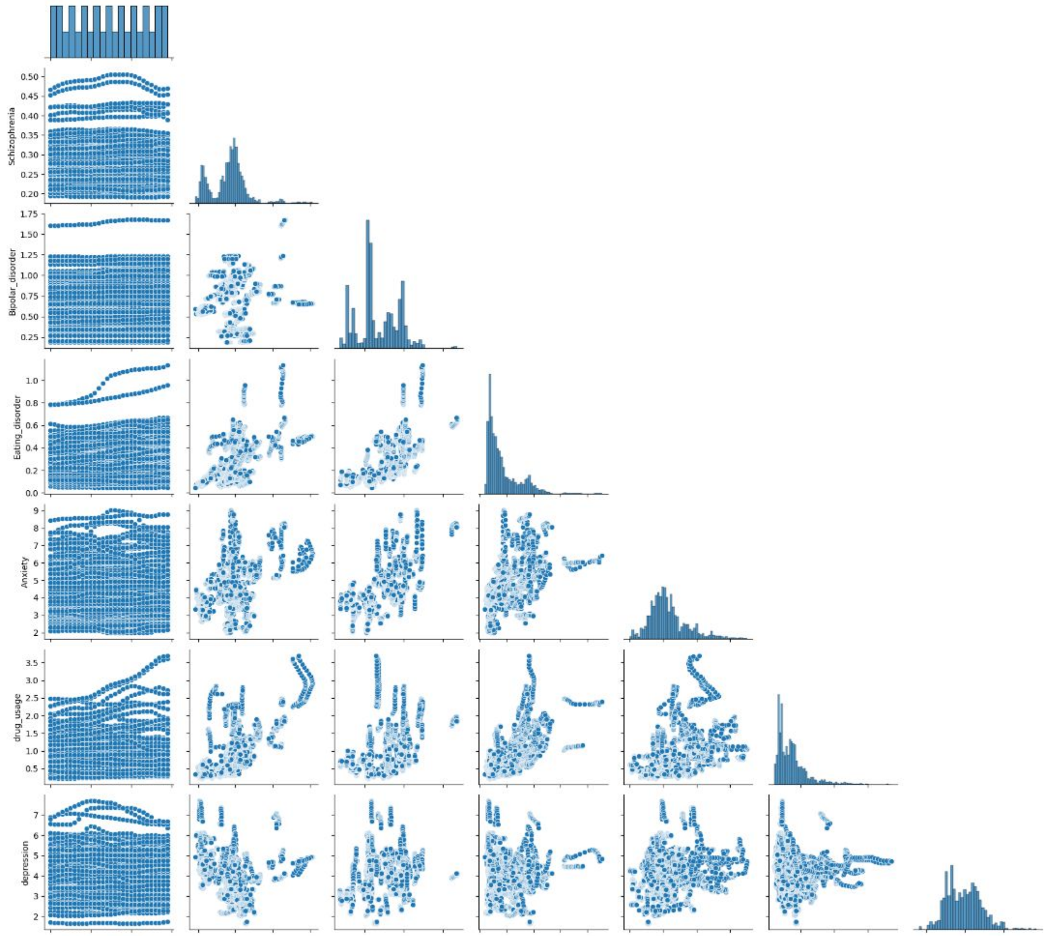
3. Joint Plot

```
sns.jointplot(data,x='Bipolar_disorder',y='mental_fitness',kind='reg',color='green')
plt.show()
```



3. Pair Plot

```
sns.pairplot(data,corner=True)
plt.show()
```

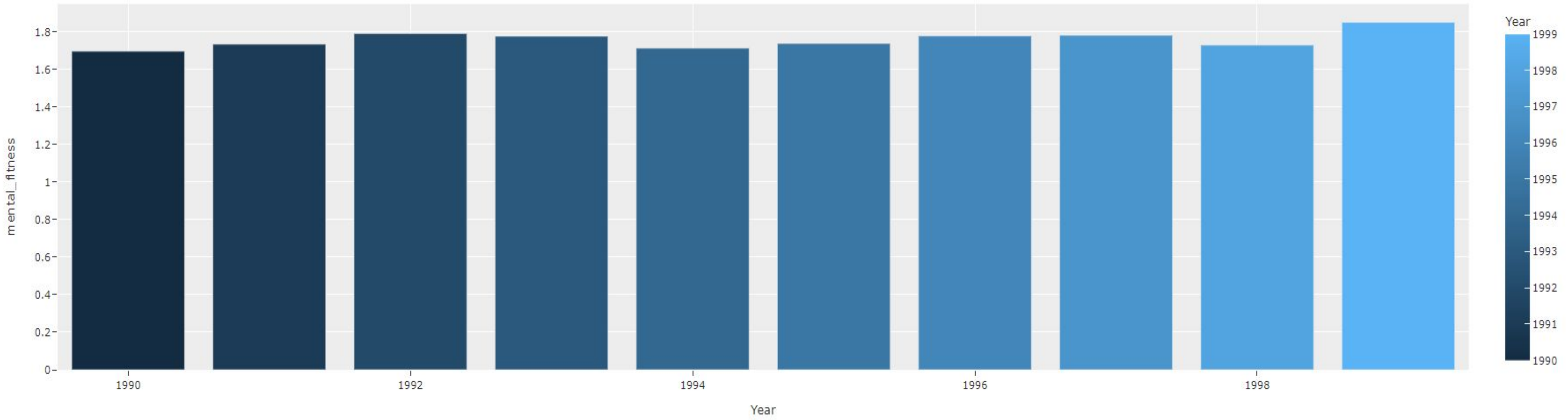


3. Pie Plot

```
fig = px.pie(data, values='mental_fitness', names='Year')
fig.show()
```

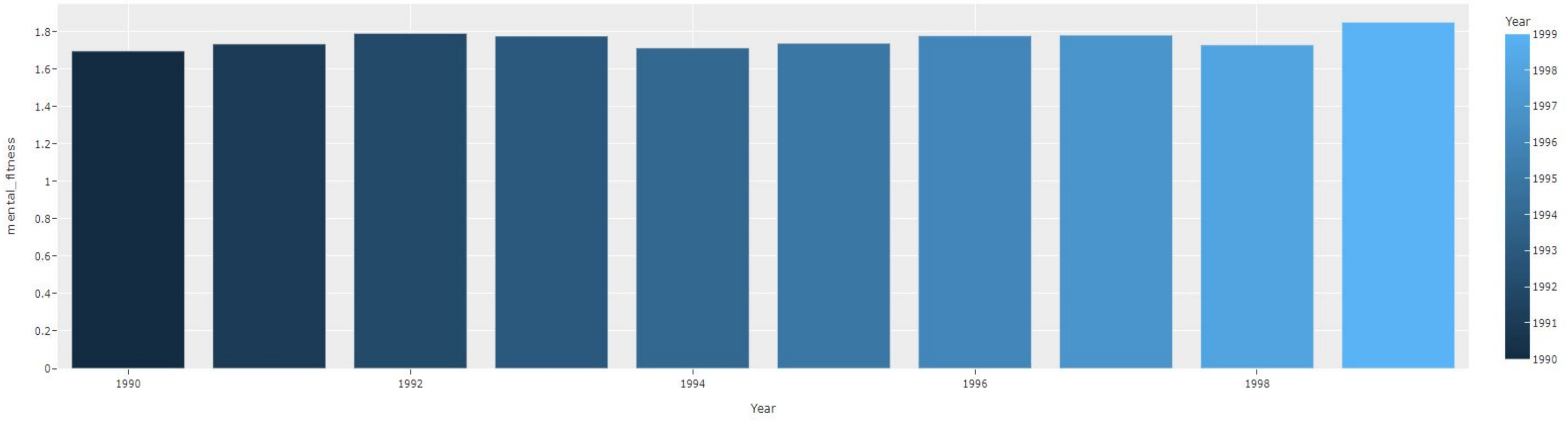
3. Bar Plot

```
fig=px.bar(data.head(10),x='Year',y='mental_fitness',color='Year',template='ggplot2')
fig.show()
```



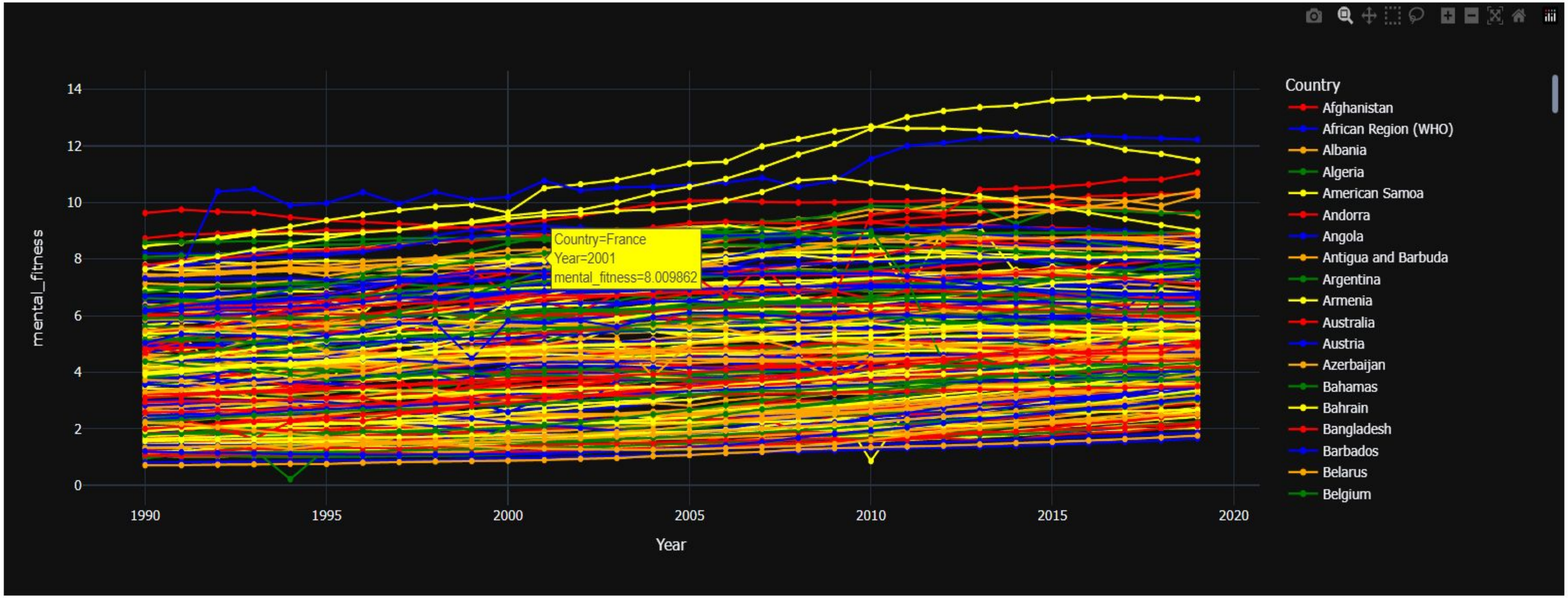
3. Bar Plot

```
fig=px.bar(data.head(10),x='Year',y='mental_fitness',color='Year',template='ggplot2')
fig.show()
```



Step 6 :- Year wise Variation in Mental Fitness of Different Countries

```
fig = px.line(data, x="Year", y="mental_fitness", color='Country',markers=True,color_discrete_sequence=['red','blue','orange','green','yellow'],template='plotly_dark')
fig.show()
```



Step 7 :- Pre-processing

```
from sklearn.preprocessing import LabelEncoder
l=LabelEncoder()
for i in df.columns:
    if df[i].dtype == 'object':
        df[i]=l.fit_transform(df[i])
```

```
X = df.drop('mental_fitness',axis=1)
y = df['mental_fitness']

from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(X, y, test_size=0.2, random_state=2)
```


```
X = df.drop('mental_fitness',axis=1)
y = df['mental_fitness']

from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(X, y, test_size=0.2, random_state=2)
```

Step 8 :- Perform Linear Regression

```
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
lr = LinearRegression()
lr.fit(xtrain,ytrain)

# model evaluation for training set
ytrain_pred = lr.predict(xtrain)
mse = mean_squared_error(ytrain, ytrain_pred)
rmse = (np.sqrt(mean_squared_error(ytrain, ytrain_pred)))
r2 = r2_score(ytrain, ytrain_pred)

print("The model performance for training set")
print("--------------------------------------")
print('MSE is {}'.format(mse))
print('RMSE is {}'.format(rmse))
print('R2 score is {}'.format(r2))
print("\n")

# model evaluation for testing set
ytest_pred = lr.predict(xtest)
mse = mean_squared_error(ytest, ytest_pred)
rmse = (np.sqrt(mean_squared_error(ytest, ytest_pred)))
r2 = r2_score(ytest, ytest_pred)

print("The model performance for testing set")
print("--------------------------------------")
print('MSE is {}'.format(mse))
print('RMSE is {}'.format(rmse))
print('R2 score is {}'.format(r2))
```

Step 9 :- Perform Random Forest

```python
from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor()
rf.fit(xtrain, ytrain)

# model evaluation for training set
ytrain_pred = rf.predict(xtrain)
mse = mean_squared_error(ytrain, ytrain_pred)
rmse = (np.sqrt(mean_squared_error(ytrain, ytrain_pred)))
r2 = r2_score(ytrain, ytrain_pred)

print("The model performance for training set")
print("--------------------------------------")
print('MSE is {}'.format(mse))
print('RMSE is {}'.format(rmse))
print('R2 score is {}'.format(r2))
print("\n")

# model evaluation for testing set
ytest_pred = rf.predict(xtest)
mse = mean_squared_error(ytest, ytest_pred)
rmse = (np.sqrt(mean_squared_error(ytest, ytest_pred)))
r2 = r2_score(ytest, ytest_pred)

print("The model performance for testing set")
print("--------------------------------------")
print('MSE is {}'.format(mse))
print('RMSE is {}'.format(rmse))
print('R2 score is {}'.format(r2))
```

RESULTS

- Linear Regression

```
The model performance for training set
--------------------------------------
MSE is 1.389959372405798
RMSE is 1.1789653821914357
R2 score is 0.7413245790025275


The model performance for testing set
--------------------------------------
MSE is 1.1357545319272384
RMSE is 1.0657178481789813
R2 score is 0.7638974087055272
```

RESULTS

- Random Forest

```
The model performance for training set
--------------------------------------
MSE is 0.004887683106083917
RMSE is 0.0699119668303211
R2 score is 0.9990903881722959


The model performance for testing set
--------------------------------------
MSE is 0.029734631044776378
RMSE is 0.172437324975703 5
R2 score is 0.9938187141292371
```

LINKS

Google Colab -- https://colab.research.google.com/drive/1dgvYCir5r2xDwM0ZOhVGR4taDWDZz8QT?usp=sharing

Drive -- https://drive.google.com/drive/folders/1lZlUGdS-gPfyxAs7SANVaFZQjAMLD4Kj?usp=sharing