

POPULATION GROWTH PREDICTIONS

```
In [244]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [245]: population = pd.read_csv("C:/Users/MAHE/population_modified.csv")
```

```
In [246]:
```

Out[246]:

	LOCATION	TIME	Value	RATIO	LE
0	Andhra Pradesh	1951	31.1	965	63.4
1	Andhra Pradesh	1961	35.3	968	64.3
2	Andhra Pradesh	1971	38.3	970	65.6
3	Andhra Pradesh	1981	40.4	971	66.8
4	Andhra Pradesh	1991	44.5	972	67.3

```
In [247]:
```

```
In [248]:
```

Out[248]:

	LOCATION	TIME	Value
0	Andhra Pradesh	1951	31.1
1	Andhra Pradesh	1961	35.3
2	Andhra Pradesh	1971	38.3
3	Andhra Pradesh	1981	40.4
4	Andhra Pradesh	1991	44.5

```
In [249]: from sklearn.preprocessing import LabelEncoder
```

```
In [250]:
```

Out[250]: LabelEncoder()

```
In [251]:
```

In [252]:

```
Out[252]: array([[ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  1,  1,  1,  1,  1,  1,  1,
  1,  1,  1,  2,  2,  2,  2,  2,  2,  2,  2,  2,  2,  2,  3,  3,  3,  3,
  3,  3,  3,  3,  3,  3,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  5,
  5,  5,  5,  5,  5,  5,  5,  5,  5,  6,  6,  6,  6,  6,  6,  6,  6,  6,
  6,  6,  7,  7,  7,  7,  7,  7,  7,  7,  7,  7,  7,  8,  8,  8,  8,  8,
  8,  8,  8,  8,  8,  9,  9,  9,  9,  9,  9,  9,  9,  9,  9,  9, 10, 10,
 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11,
 11, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 13,
 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 15, 15, 15,
 15, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16,
 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18, 18,
 18, 18, 18, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 20, 20, 20, 20,
 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 21, 21, 21, 21, 21, 21])
```

In [253]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
("""Entry point for launching an IPython kernel.
```

In [254]:

```
Out[254]:
```

	LOCATION	TIME	Value	Country
0	Andhra Pradesh	1951	31.1	0
1	Andhra Pradesh	1961	35.3	0
2	Andhra Pradesh	1971	38.3	0
3	Andhra Pradesh	1981	40.4	0
4	Andhra Pradesh	1991	44.5	0

In [255]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
errors=errors,
```

In [256]:

In [257]:

In [258]: `from sklearn import preprocessing`

In [259]: `from sklearn.model_selection import train_test_split`

```
In [260]: from sklearn.linear_model import LinearRegression
lm = LinearRegression()
```

```
Out[260]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
In [261]:
```

```
In [262]: from sklearn import metrics
print('MAE:', metrics.mean_absolute_error(y_test, predictions))
print('MSE:', metrics.mean_squared_error(y_test, predictions))
```

```
MAE: 22.220072825029213
MSE: 907.2005620132772
RMSE: 30.119770284868995
```

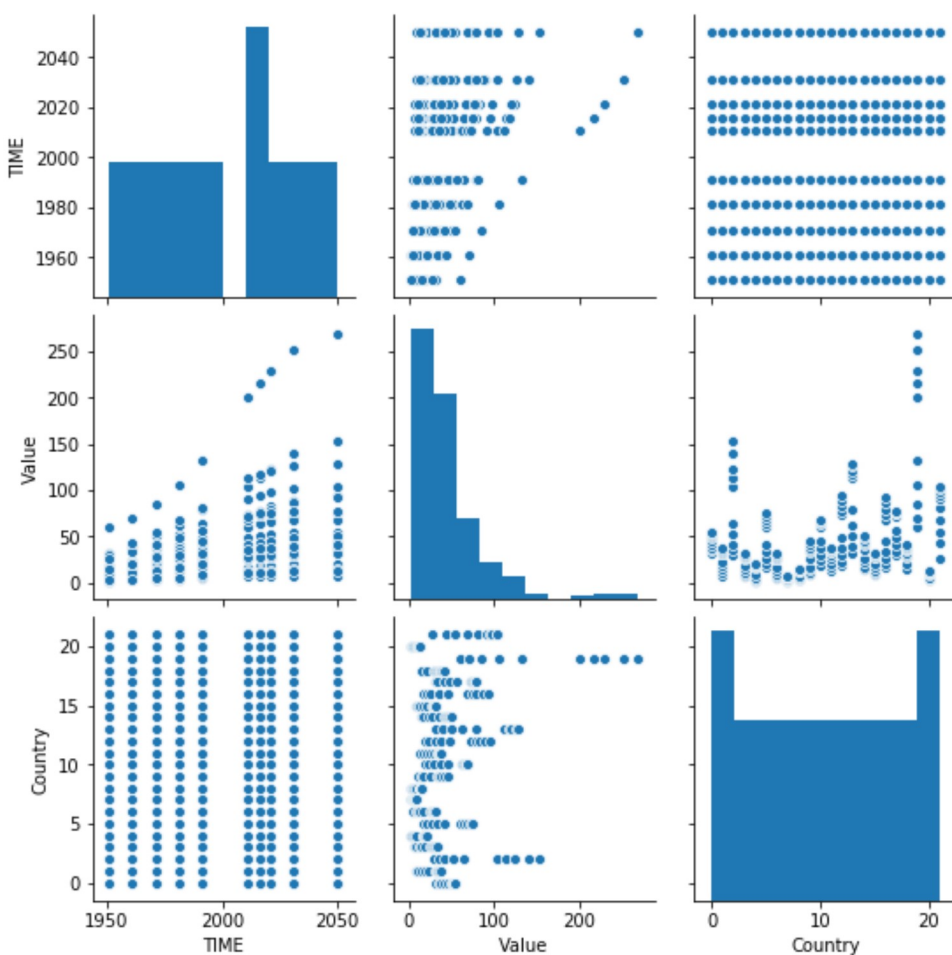
```
In [263]:
```

```
Out[263]:
```

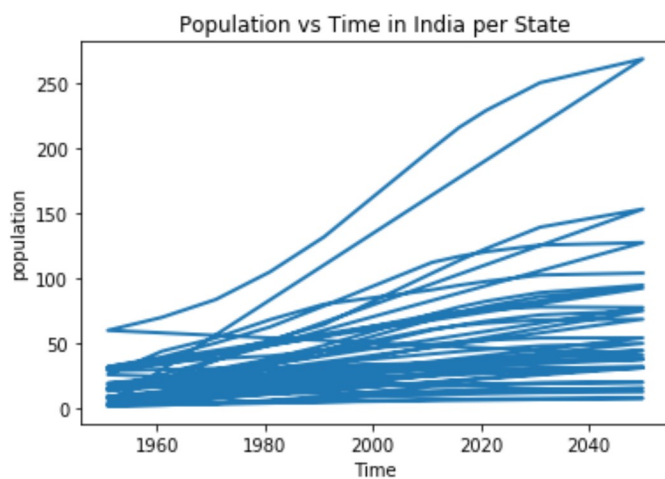
	TIME	Value	Country
0	1951	31.1	0
1	1961	35.3	0
2	1971	38.3	0
3	1981	40.4	0
4	1991	44.5	0

In [264]:

Out[264]: <seaborn.axisgrid.PairGrid at 0x19c1cc3f3c8>



```
In [265]: plt.plot(df['TIME'],df['Value'],linewidth=2.0)
plt.title('Population vs Time in India per State')
plt.xlabel('Time')
plt.ylabel('population')
```



In []:

In [266]: X = df.drop(['Value'],axis=1)

In [268]:

In [271]:

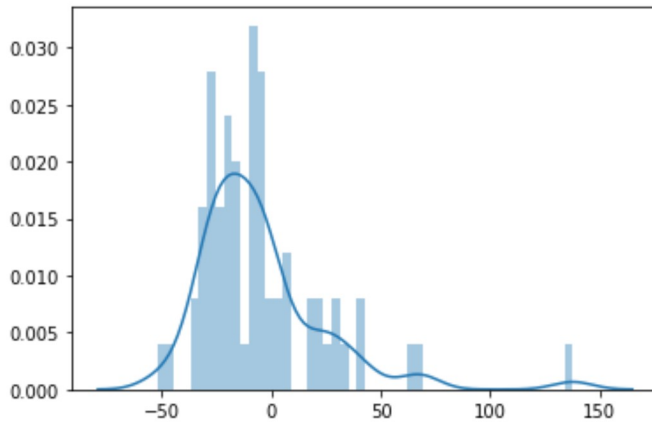
In [272]:

Predictions vs test

predictions

y test

In [274]:

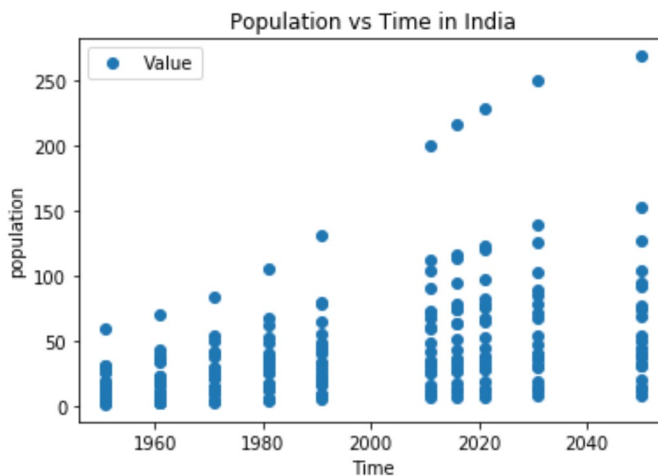


In [275]:

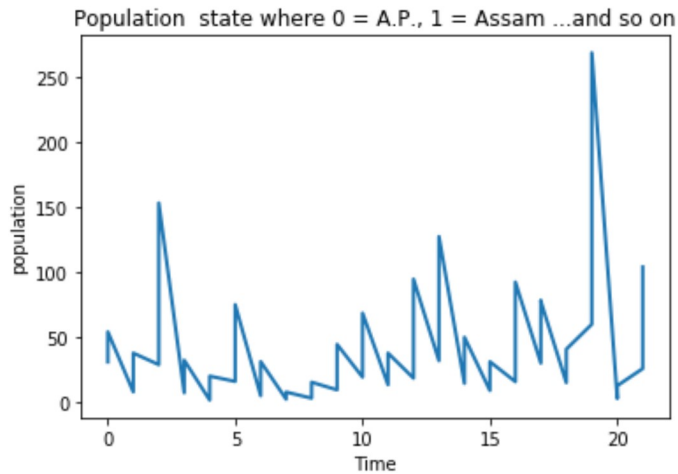
Out[275]:

	LOCATION	TIME	Value	RATIO	LE
0	Andhra Pradesh	1951	31.1	965	63.4
1	Andhra Pradesh	1961	35.3	968	64.3
2	Andhra Pradesh	1971	38.3	970	65.6
3	Andhra Pradesh	1981	40.4	971	66.8
4	Andhra Pradesh	1991	44.5	972	67.3

```
In [276]: population.plot(x='TIME', y='Value', style='o')
plt.title('Population vs Time in India')
plt.xlabel('Time')
plt.ylabel('population')
```



```
In [277]: plt.plot(df['Country'],df['Value'],linewidth=2.0)
plt.title('Population state where 0 = A.P., 1 = Assam ...and so on')
plt.xlabel('Time')
plt.ylabel('population')
```



```
In [ ]:
```

AS WE CAN CAN SEE ABOVE THE POPULATION IN INDIA IS EXPECTED TO BE HIGHEST IN UTTARPRADESH ABOVE 250 MILLION

```
In [ ]:
```

SEX RATIO IN INDIA PER STATES

```
In [278]:
```

```
In [279]:
```

```
Out[279]:
```

	LOCATION	TIME	RATIO
0	Andhra Pradesh	1951	965
1	Andhra Pradesh	1961	968
2	Andhra Pradesh	1971	970
3	Andhra Pradesh	1981	971
4	Andhra Pradesh	1991	972

```
In [280]:
```

```
Out[280]: LabelEncoder()
```

```
In [281]:
```

In [282]:

```
Out[282]: array([[ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  1,  1,  1,  1,  1,  1,  1,
  1,  1,  1,  2,  2,  2,  2,  2,  2,  2,  2,  2,  2,  3,  3,  3,  3,
  3,  3,  3,  3,  3,  3,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  5,
  5,  5,  5,  5,  5,  5,  5,  5,  5,  6,  6,  6,  6,  6,  6,  6,  6,  6,
  6,  6,  7,  7,  7,  7,  7,  7,  7,  7,  7,  7,  7,  8,  8,  8,  8,  8,
  8,  8,  8,  8,  8,  9,  9,  9,  9,  9,  9,  9,  9,  9,  9,  9, 10, 10,
 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11,
 11, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 13,
 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 15, 15, 15,
 15, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16,
 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18, 18,
 18, 18, 18, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 20, 20, 20, 20,
 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 21, 21, 21, 21, 21, 21])
```

In [283]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
"""Entry point for launching an IPython kernel.
```

In [284]:

Out[284]:

	LOCATION	TIME	RATIO	State
0	Andhra Pradesh	1951	965	0
1	Andhra Pradesh	1961	968	0
2	Andhra Pradesh	1971	970	0
3	Andhra Pradesh	1981	971	0
4	Andhra Pradesh	1991	972	0

In [285]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
errors=errors,
```

In [286]:

In [287]:

In [288]:

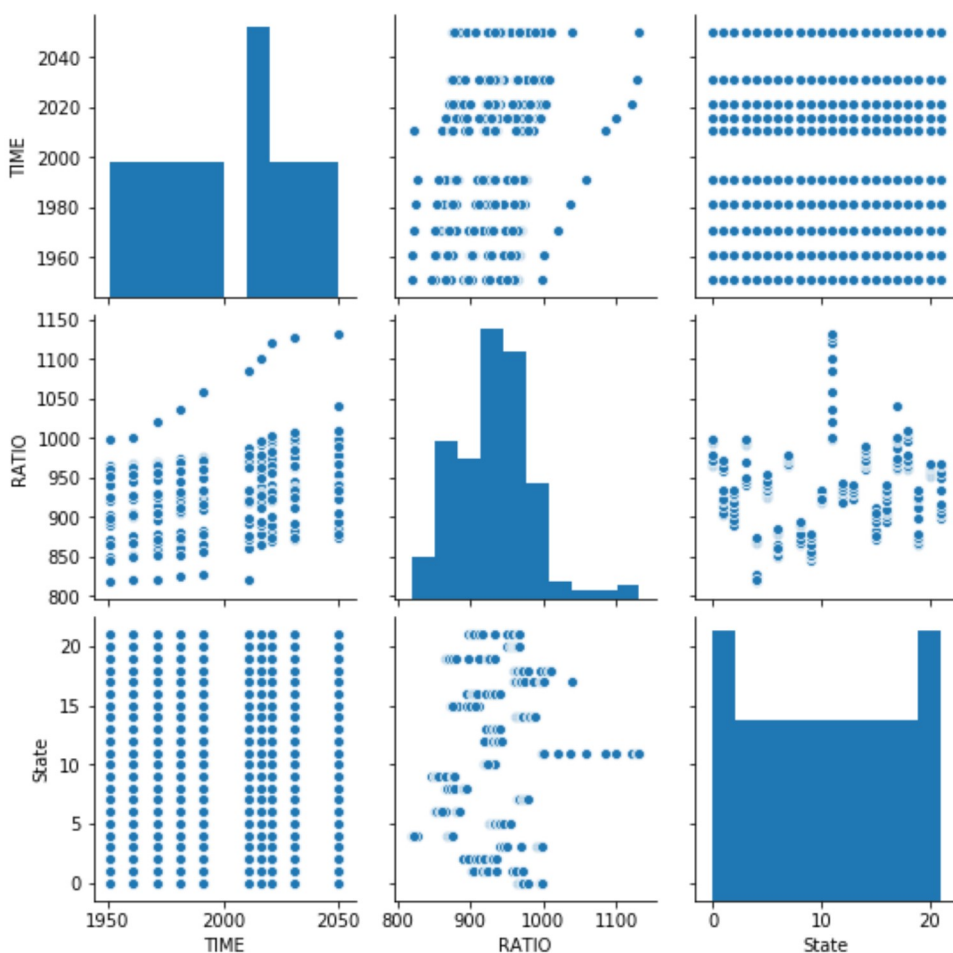
Out[288]:

	TIME	RATIO	State
0	1951	965	0
1	1961	968	0
2	1971	970	0
3	1981	971	0
4	1991	972	0

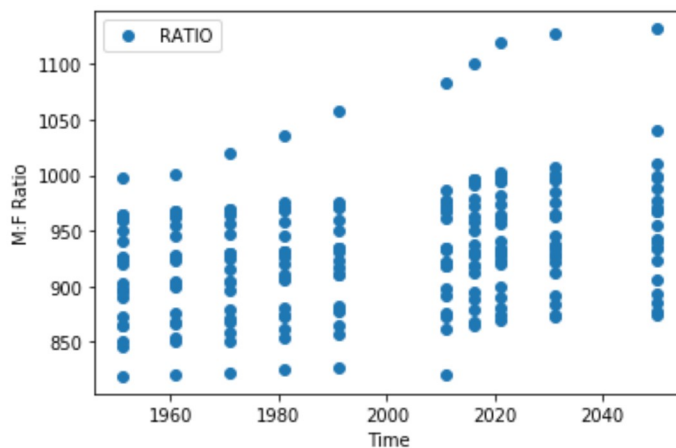
In [289]: `from sklearn import preprocessing`In [290]: `from sklearn.model_selection import train_test_split`In [291]: `from sklearn.linear_model import LinearRegression`
`lml = LinearRegression()`Out[291]: `LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)`

In [292]:

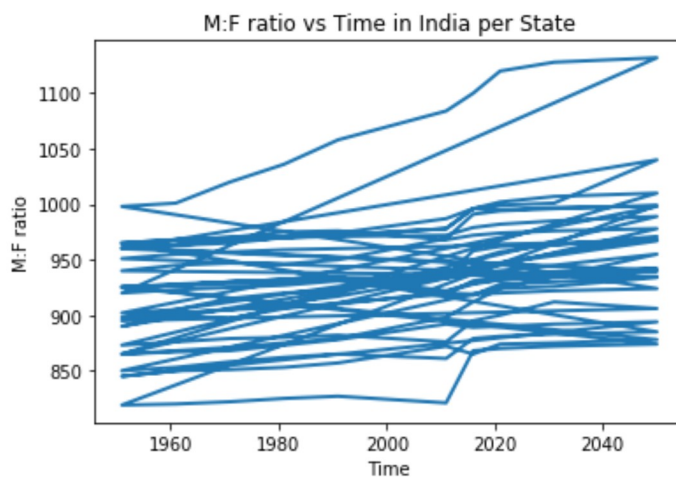
In [293]:

Out[293]: `<seaborn.axisgrid.PairGrid at 0x19c1f991208>`

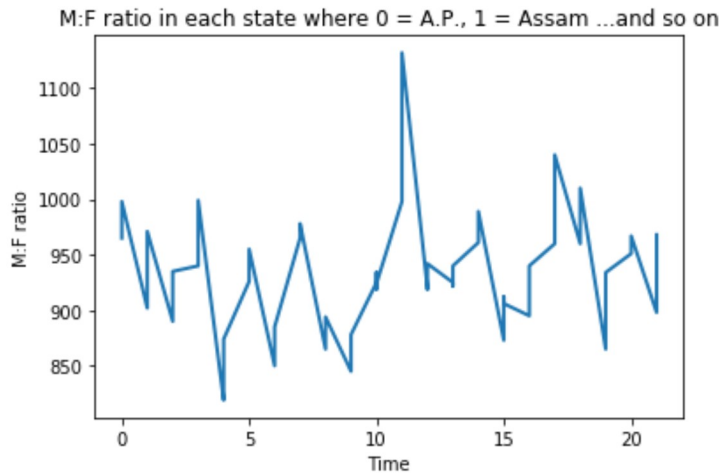
```
In [294]: population.plot(x='TIME', y='RATIO', style='o')
plt.title('')
plt.xlabel('Time')
plt.ylabel('M:F Ratio')
```



```
In [295]: plt.plot(df1['TIME'],df1['RATIO'],linewidth=2.0)
plt.title('M:F ratio vs Time in India per State')
plt.xlabel('Time')
plt.ylabel('M:F ratio')
```



```
In [296]: plt.plot(df1['State'],df1['RATIO'],linewidth=2.0)
plt.title('M:F ratio in each state where 0 = A.P., 1 = Assam ...and so on')
plt.xlabel('Time')
plt.ylabel('M:F ratio')
```



States like Kerala and Tamilnadu are predicted to have a M:F ratio > 1000 by 2050

LONGEVITY STUDY

```
In [297]:
```

```
In [298]:
```

```
Out[298]:
```

	LOCATION	TIME	LE
0	Andhra Pradesh	1951	63.4
1	Andhra Pradesh	1961	64.3
2	Andhra Pradesh	1971	65.6
3	Andhra Pradesh	1981	66.8
4	Andhra Pradesh	1991	67.3

```
In [299]:
```

```
Out[299]: LabelEncoder()
```

```
In [300]:
```

In [301]:

```
Out[301]: array([[ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  1,  1,  1,  1,  1,  1,  1,
  1,  1,  1,  2,  2,  2,  2,  2,  2,  2,  2,  2,  2,  3,  3,  3,  3,
  3,  3,  3,  3,  3,  3,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  5,
  5,  5,  5,  5,  5,  5,  5,  5,  5,  6,  6,  6,  6,  6,  6,  6,  6,  6,
  6,  6,  7,  7,  7,  7,  7,  7,  7,  7,  7,  7,  7,  8,  8,  8,  8,  8,
  8,  8,  8,  8,  8,  9,  9,  9,  9,  9,  9,  9,  9,  9,  9,  9, 10, 10,
 10, 10, 10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11,
 11, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 13, 13, 13, 13, 13, 13,
 13, 13, 13, 13, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 15, 15, 15,
 15, 15, 15, 15, 15, 15, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16, 16,
 17, 17, 17, 17, 17, 17, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18, 18, 18,
 18, 18, 19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 20, 20, 20, 20, 20,
 20, 20, 20, 20, 20, 21, 21, 21, 21, 21, 21, 21, 21, 21, 21])
```

In [302]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopy
Warning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable
/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org
/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
    """Entry point for launching an IPython kernel.
```

In [303]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWithCo
pyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable
/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org
/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
    errors=errors,
```

In [304]:

In [305]:

In [306]: `from sklearn import preprocessing`

In [307]: `from sklearn.model_selection import train_test_split`

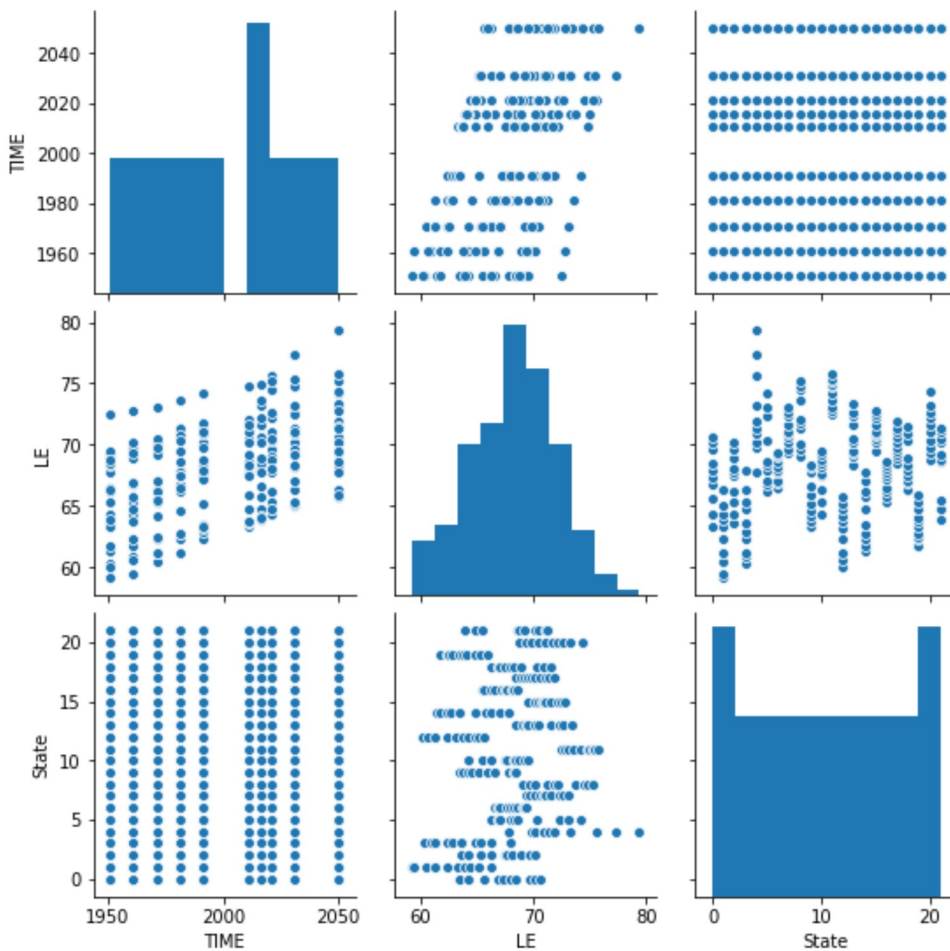
In [308]: `from sklearn.linear_model import LinearRegression`
`lm2 = LinearRegression()`

Out[308]: `LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)`

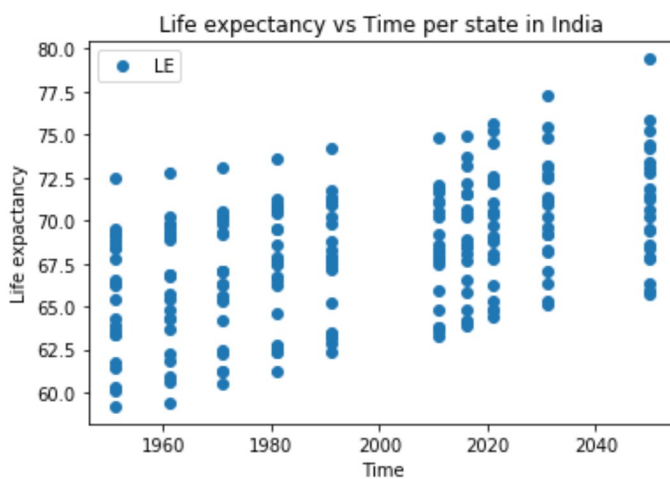
In [309]:

In [310]:

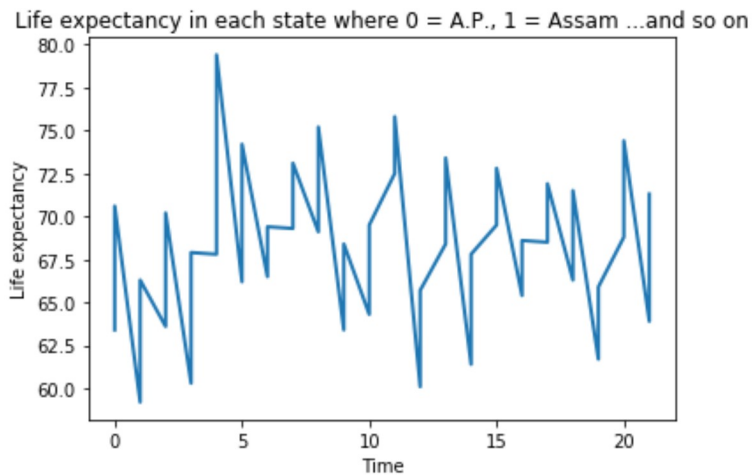
Out[310]: <seaborn.axisgrid.PairGrid at 0x19c20251bc8>



```
In [311]: population.plot(x='TIME', y='LE', style='o')
plt.title('Life expectancy vs Time per state in India')
plt.xlabel('Time')
plt.ylabel('Life expectancy')
```



```
In [312]: plt.plot(df2['State'],df2['LE'],linewidth=2.0)
plt.title('Life expectancy in each state where 0 = A.P., 1 = Assam ...and so on')
plt.xlabel('Time')
plt.ylabel('Life expectancy')
```



POPULATION GROWTH FOR INDIA

```
In [313]: populationind = pd.read_csv("C:/Users/MAHE/population_modified_india.csv")
```

```
In [314]:
```

```
Out[314]:
```

	LOCATION	TIME	Value	RATIO	LE
0	INDIA	1951	361.0	891	45
1	INDIA	1961	545.8	900	52
2	INDIA	1971	635.7	910	62
3	INDIA	1981	846.9	929	63
4	INDIA	1991	1028.7	934	70

```
In [315]:
```

```
In [316]: from sklearn.preprocessing import LabelEncoder
```

```
In [317]:
```

```
Out[317]: LabelEncoder()
```

```
In [318]:
```

In [319]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
    """Entry point for launching an IPython kernel.
```

In [320]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
    errors=errors,
```

In [321]:

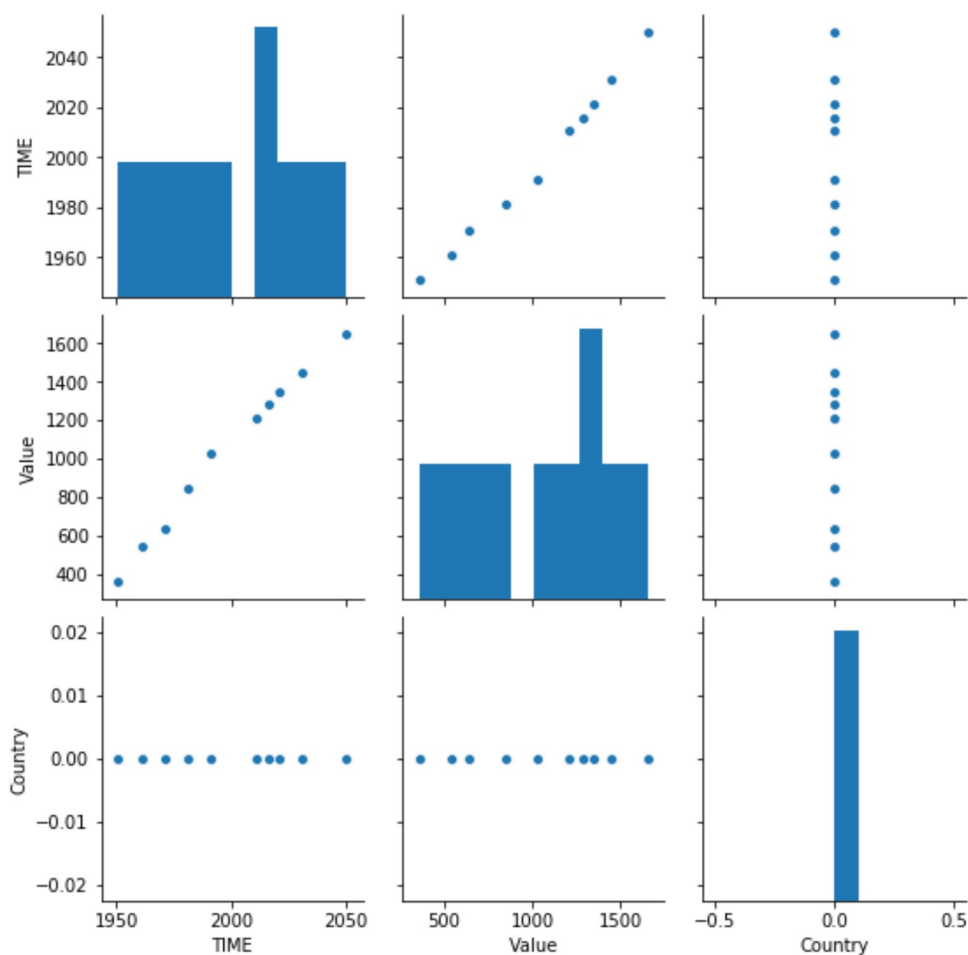
In [322]:

In [323]: `from sklearn import preprocessing`In [324]: `from sklearn.model_selection import train_test_split`In [325]: `from sklearn.linear_model import LinearRegression`
`lm = LinearRegression()`Out[325]: `LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)`

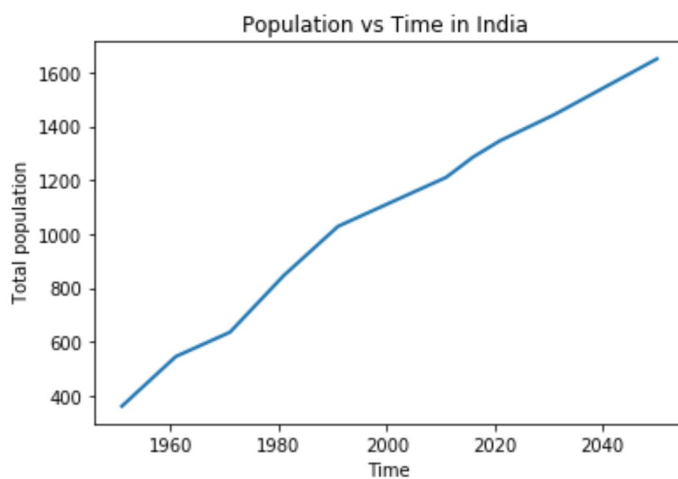
In [326]:

In [327]:

Out[327]: <seaborn.axisgrid.PairGrid at 0x19c1f931fc8>



```
In [328]: plt.plot(df3['TIME'],df3['Value'],linewidth=2.0)
plt.title('Population vs Time in India ')
plt.xlabel('Time')
plt.ylabel(' Total population')
```



GENDER RATIO IN INDIA

In [329]:

In [330]:

Out[330]: LabelEncoder()

In [331]:

In [332]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
"""Entry point for launching an IPython kernel.
```

In [333]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
errors=errors,
```

In [334]:

Out[334]:

	TIME	RATIO	Country
0	1951	891	0
1	1961	900	0
2	1971	910	0
3	1981	929	0
4	1991	934	0

In [335]:

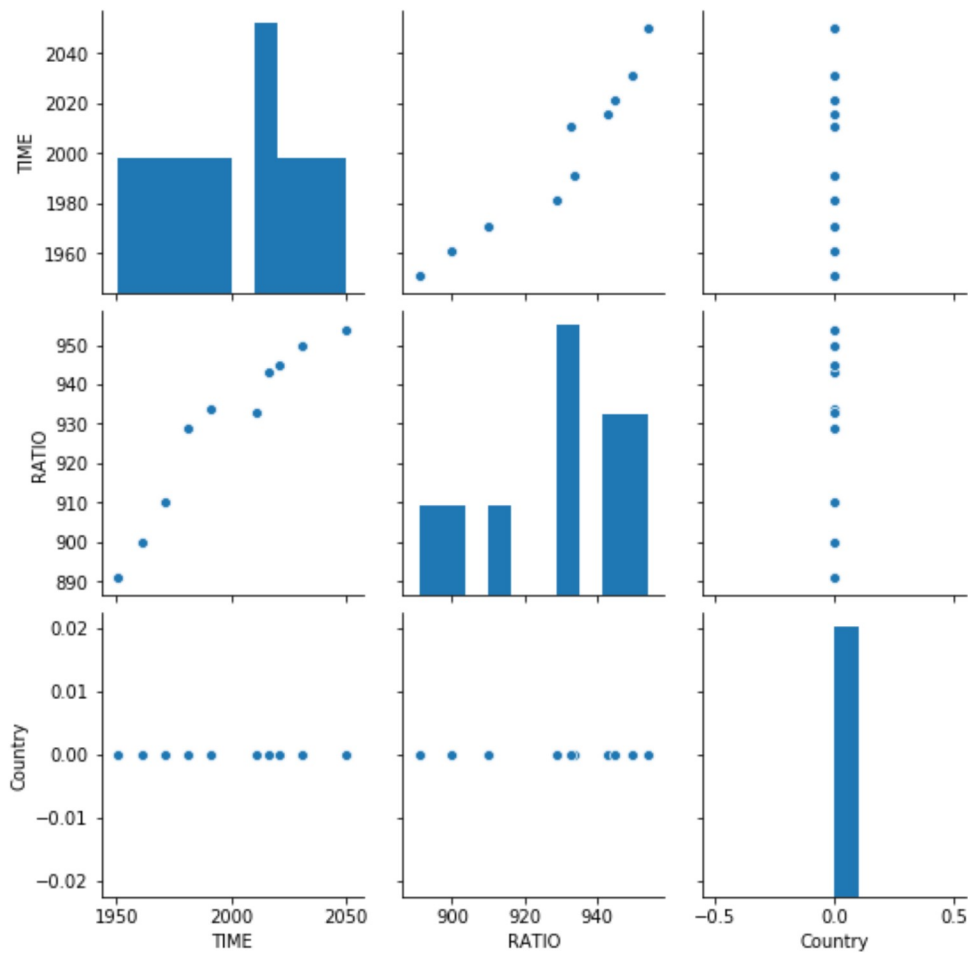
In [336]:

In [337]: `from sklearn import preprocessing`In [338]: `from sklearn.model_selection import train_test_split`In [339]: `from sklearn.linear_model import LinearRegression`
`lm = LinearRegression()`

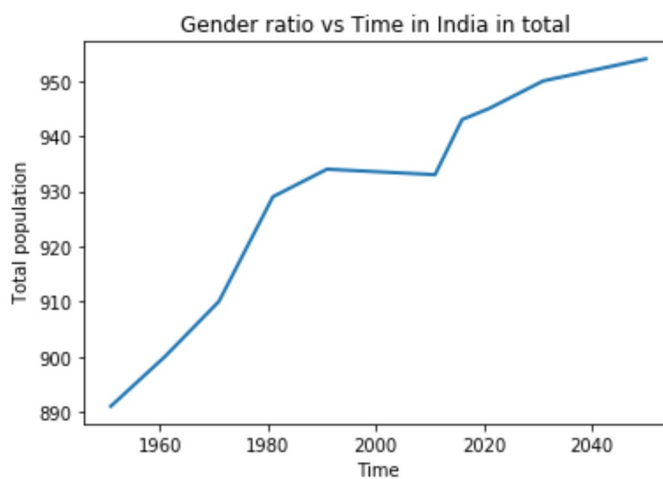
Out[339]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

In [340]:

Out[340]: <seaborn.axisgrid.PairGrid at 0x19c20c94308>



```
In [341]: plt.plot(df4['TIME'],df4['RATIO'],linewidth=2.0)
plt.title('Gender ratio vs Time in India in total ')
plt.xlabel('Time')
plt.ylabel(' Total population')
```



LIFE EXPECTANCY IN INDIA

In [342]:

In [343]:

Out[343]: LabelEncoder()

In [344]:

In [345]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
"""Entry point for launching an IPython kernel.
```

In [346]:

```
C:\Users\MAHE\anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy)
errors=errors,
```

In [347]:

Out[347]:

	TIME	LE	Country
0	1951	45	0
1	1961	52	0
2	1971	62	0
3	1981	63	0
4	1991	70	0

In [348]:

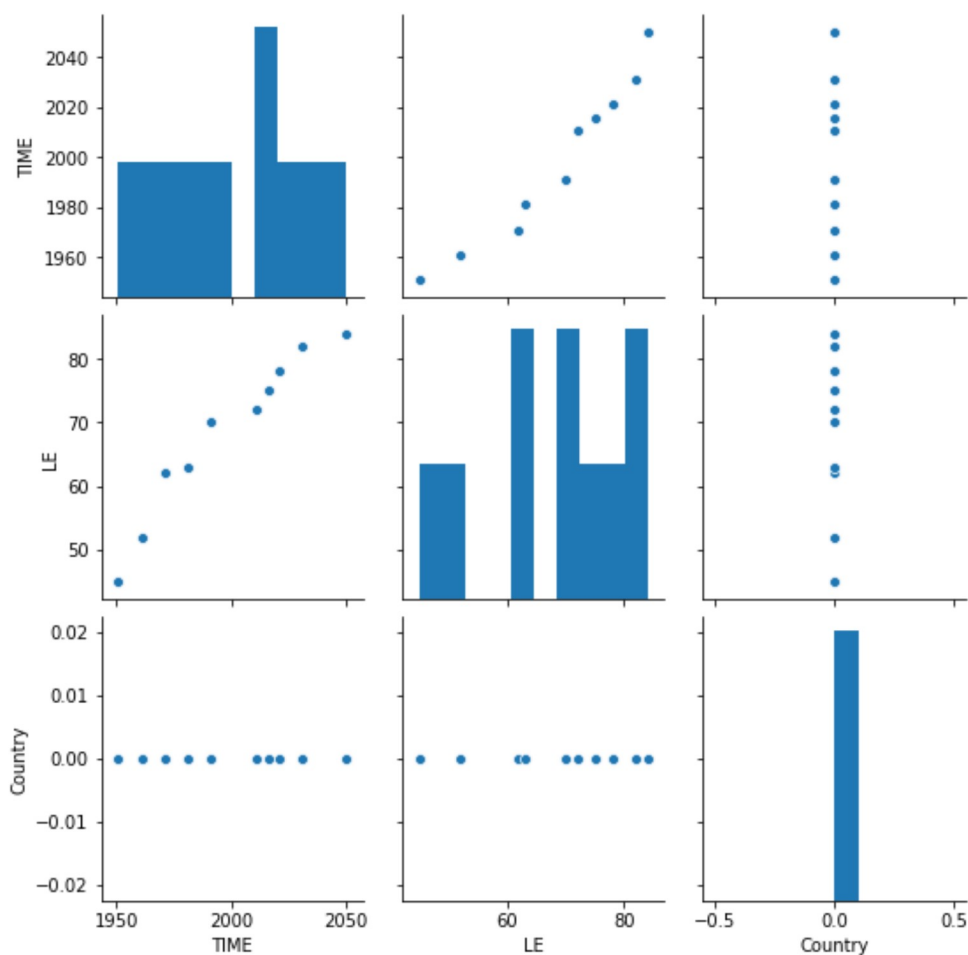
In [349]:

In [350]: `from sklearn import preprocessing`In [351]: `from sklearn.model_selection import train_test_split`In [352]: `from sklearn.linear_model import LinearRegression`
`lm = LinearRegression()`

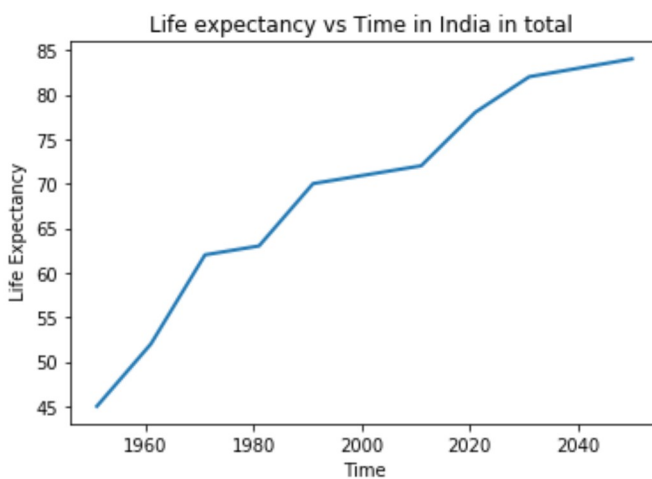
Out[352]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)

In [353]:

Out[353]: <seaborn.axisgrid.PairGrid at 0x19c22367208>



```
In [354]: plt.plot(df5['TIME'],df5['LE'],linewidth=2.0)
plt.title('Life expectancy vs Time in India in total ')
plt.xlabel('Time')
plt.ylabel(' Life Expectancy')
```



END OF PROJECT.

In []:

