

# Matson Social Skills

- Kamyab Abedi
- Kian Jalilian
- Ehsan Sajadi

DataMining Course - Spring 1400



# TABLE OF CONTENTS

## **Introduction**

Briefly elaborate on what we want to discuss.

---

## **PROBLEM & SOLUTION**

Review on our datasets

---

## **OUR PROCESS**

What we did?

## **TARGET**

What is our target?

# Our Data Set

## Life Expectancy

The data was collected from WHO and United Nations website

## Matson social skills

The data was collected by a master student of SBU

## Fish

Review on Hematology factors collected by a PHD student of TU

## Books Recommendation

The data was collected by Dr. Amirreza Asrafi



# Final Data Set

## Matson social skills

The development of social skills is an important process in young childhood and adolescence. Deficits present in childhood that are left undetected and/or untreated can lead to increased problems into adulthood.

## Life Expectancy

This study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well.



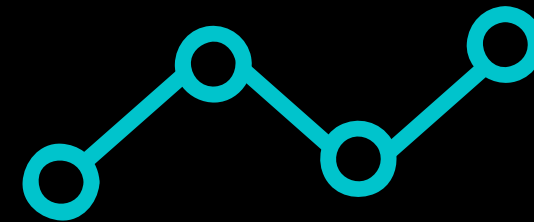


# Matson social Skills



**160 row**

Contain 80 men and women



**135 column**

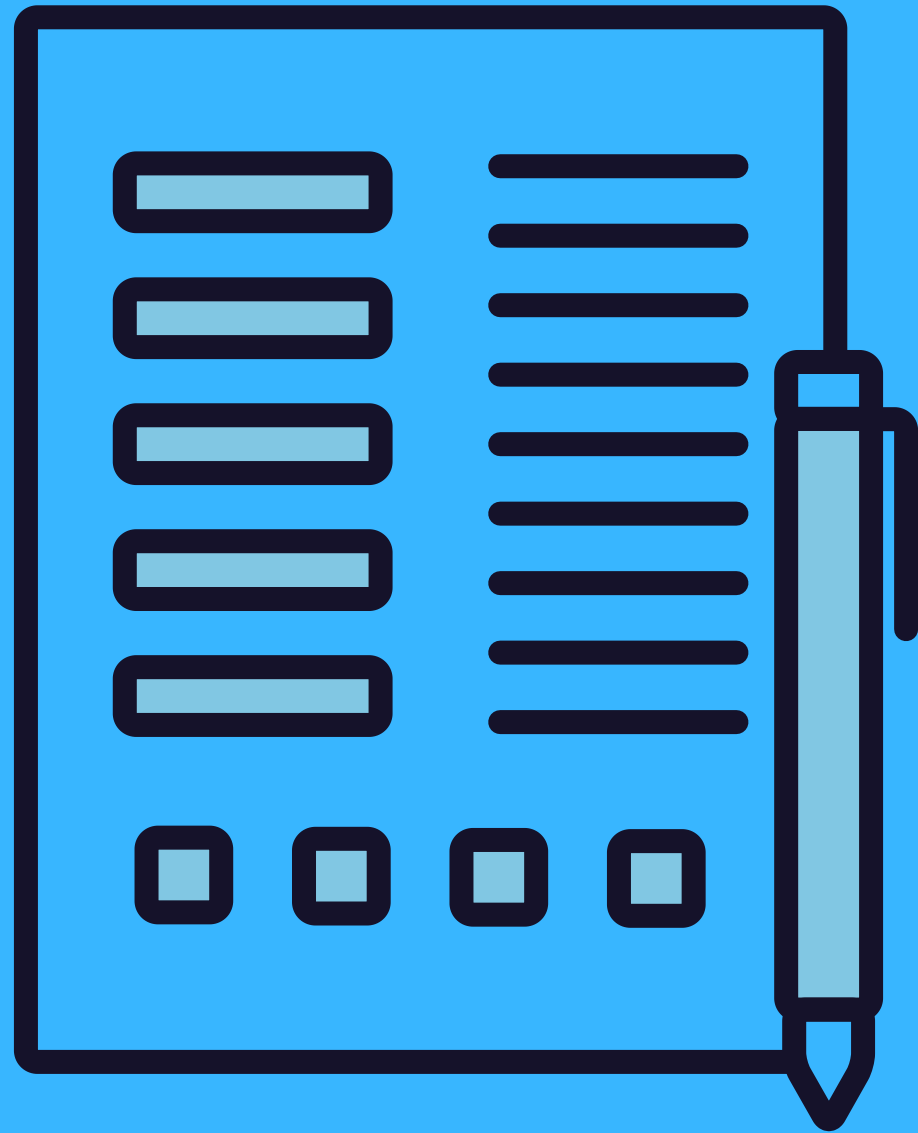
Collected by surveys



**Final data set**

Briefly elaborate on what you  
want to discuss.

# OUR PROCESS



# 1. Data Collection

The process of gathering information from different sources.

## 2. Data Profiling

The process of reviewing source data for content and quality.

مربوطه	بعد
1-18	مهارت های اجتماعی مناسب
19-29	رفتارهای غیر اجتماعی
30-41	پرخاشگری و رفتارهای تکانشی
42-47	برتری طلبی، اطمینان زیاد به خود داشتن

14	من به هنگام صحبت با دیگران، از آنها سوال هم می پرسم.	هرگز	به ندرت	برخی اوقات	اغلب اوقات	همیشه
15	من وقتی کسی را اذیت می کنم، بعدا متاسف می شوم.	هرگز	به ندرت	برخی اوقات	اغلب اوقات	همیشه
16	من به کسانی که با من خوب هستند، خوبی می کنم.	هرگز	به ندرت	برخی اوقات	اغلب اوقات	همیشه
17	من از اطرافیان احوالپرسی کرده و جویای کار آنها و مسایلی از این قبیل می شوم.	هرگز	به ندرت	برخی اوقات	اغلب اوقات	همیشه



# Matson Social Skills

13-25-28-29-48-59-65-73	رشد و بالندگی فردی
1-11-20-24-37-47-58-69-71-74	خود پیروی
5-7-9-10-21-40-41-64	ارتباط مثبت با دیگران
4-12-19-22-23-30-31-33-34-38- 45-49-52-55-57-60-61-62- 68	رضایت از زندگی
3-8-14-16-27-32-39-42-46-50- 56-63-76	معنویت
2-6-15-17-18-26-35-36-43-44- 51-53-54-66-67-70-72-75-77	شادی و خوشبینی

کاملاً موافقم	موافقم	نظری ندارم	مخالقم	کاملاً مخالفم	15	من جزء افرادی هستم که دوست دارند چیزهای جدید را تجربه کنند
موافقم	کاملاً موافقم	نظری	مخالقم	کاملاً مخالفم	16	معنی وجودی خودم را درک می کنم
کاملاً موافقم	موافقم	نظری ندارم	مخالقم	کاملاً مخالفم	17	در کل، نگرش مثبت و خوشبینانه ای نسبت به انسان ها و زندگی دارم





# 3. Data Exploration

A visual exploration to help users familiarize with information.

# Data set

```
[155]: # Discard the metadata (age and variable and row).  
le = le.drop(['age', 'variable', 'row'], axis=1)
```

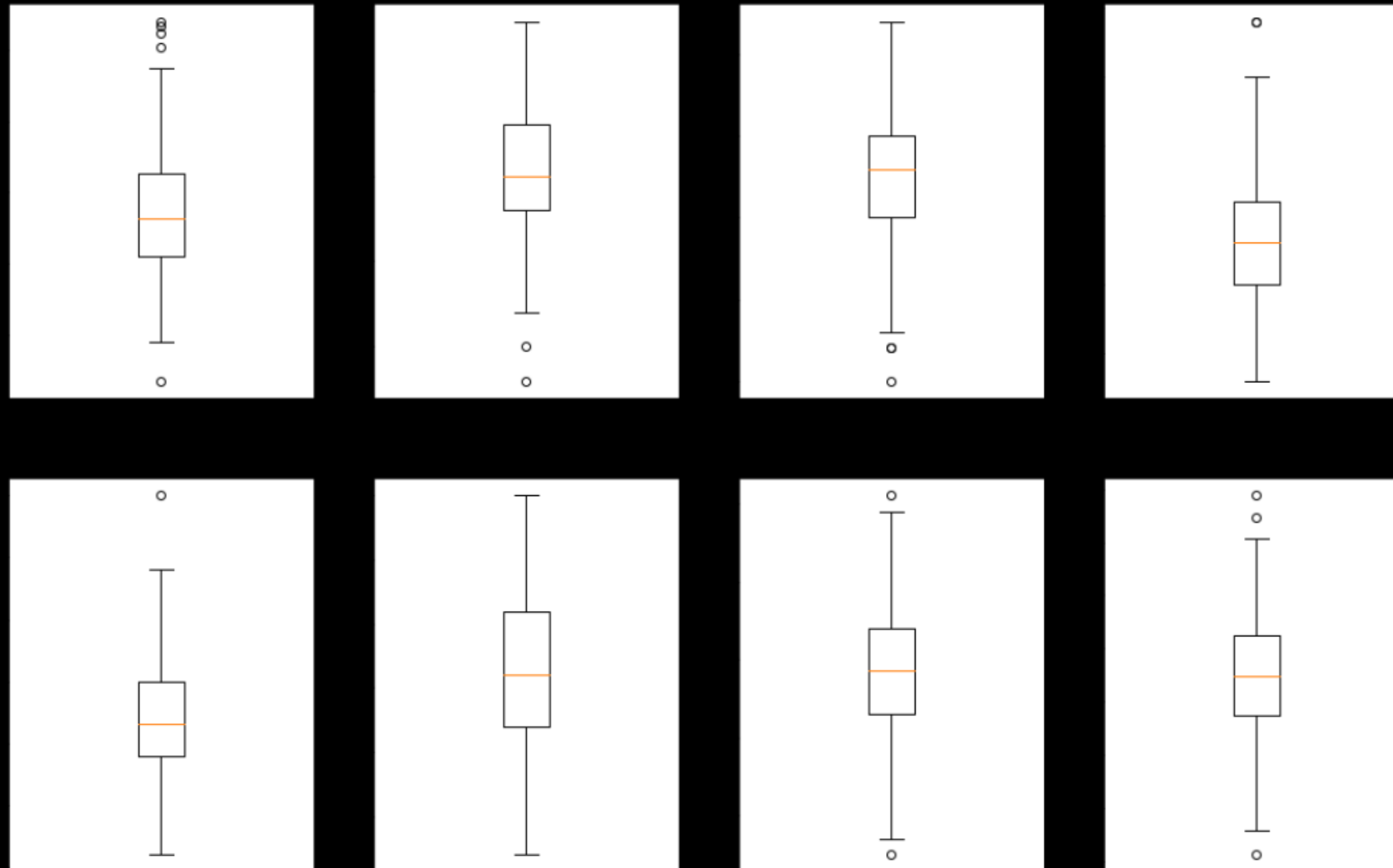
```
[156]: le.head()
```

```
[156]:
```

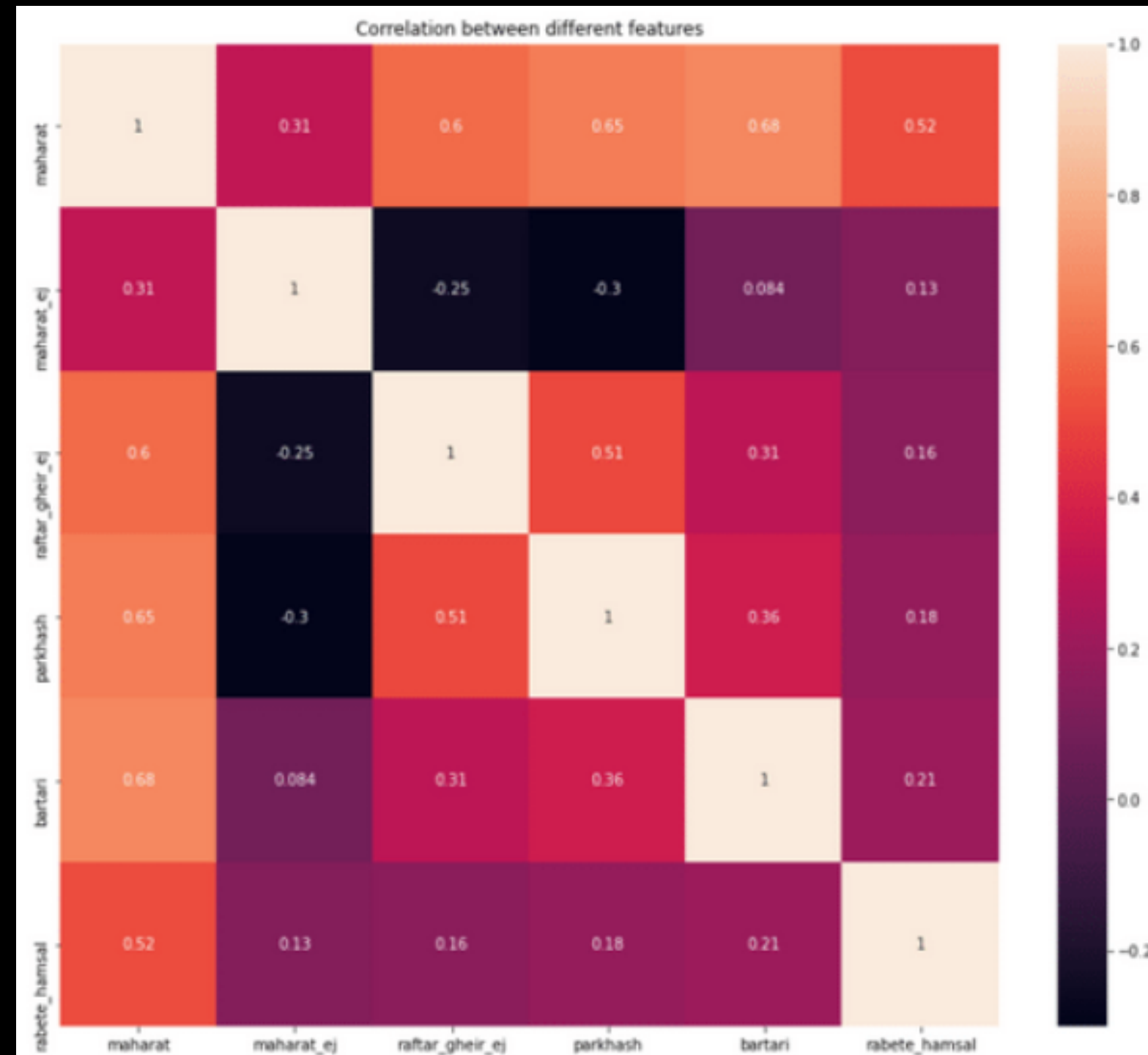
	sex	b1	b2	b3	b4	b5	b6	b7	b8	b9	...	m55	m56	maharat	roshd	khodpeiravi	ertebatm	rezayat	manaviat	shadi	behzisti
0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	5.0	168.0	15.0	28.0	16.0	42.0	22.0	48.0	170.0
1	1.0	3.0	1.0	2.0	1.0	2.0	1.0	1.0	1.0	1.0	...	5.0	5.0	170.0	18.0	26.0	17.0	43.0	21.0	47.0	171.0
2	1.0	3.0	4.0	4.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	135.0	15.0	31.0	18.0	34.0	27.0	42.0	171.0
3	1.0	2.0	5.0	2.0	2.0	2.0	2.0	1.0	2.0	2.0	...	1.0	1.0	155.0	24.0	29.0	20.0	45.0	25.0	53.0	193.0
4	1.0	1.0	1.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	152.0	24.0	25.0	12.0	40.0	22.0	38.0	159.0

5 rows × 142 columns

```
[163]: # Visualize the cleaned data for each feature using box plots.  
plt.figure(figsize=(18,30))  
  
for variable,i in col_dict.items():  
    plt.subplot(5,4,i)  
    plt.boxplot(le[variable],whis=1.5)  
    plt.title(variable)  
  
plt.show()
```



```
[164]: # Plot heatmap to visualize the correlations.  
plt.figure(figsize = (14, 12))  
sns.heatmap(le.corr(), annot = True)  
plt.title('Correlation between different features');
```





# Clustering

```
le = pd.read_csv('DataSet.csv', sep=',')
le.dataframeName = 'DataSet.csv'
le.head()
```

Python

	Variable	row	age	sex	b1	b2	b3	b4	b5	b6	...	m55	m56	maharat	roshd	khodpeiravi	ertebatm	rezayat	manaviat	shadi	behzisti
0	NaN	1	NaN	1	1	1	1	1	1	1	...	1	5	168	15	28	16	42	22	48	170
1	NaN	2	NaN	1	3	1	2	1	2	1	...	5	5	170	18	26	17	43	21	47	171
2	NaN	3	NaN	1	3	4	4	1	1	1	...	1	1	135	15	31	18	34	27	42	171
3	NaN	4	NaN	1	2	5	2	2	2	2	...	1	1	155	24	29	20	45	25	53	193
4	NaN	5	NaN	1	1	1	2	1	1	1	...	1	1	152	24	25	12	40	22	38	159

5 rows × 145 columns

...

```
# Modify the original names of the features using a standard format for all the features.
orig_cols = list(le.columns)
new_cols = []
for col in orig_cols:
    new_cols.append(col.strip().replace(' ', ' ').replace(' ', '_').lower())

le.columns = new_cols
# Discard the metadata (age and variable and row).
le.drop(['age', 'variable', 'row'], axis=1, inplace=True)
```

Python

```
le.drop(le.iloc[:,1:134], axis = 1, inplace = True)
le.head()
```

Python

	sex	maharat	roshd	khodpeiravi	ertebatm	rezayat	manaviat	shadi	behzisti
0	1	168	15	28	16	42	22	48	170
1	1	170	18	26	17	43	21	47	171
2	1	135	15	31	18	34	27	42	171
3	1	155	24	29	20	45	25	53	193
4	1	152	24	25	12	40	22	38	159



DecisionTree



```
#split dataset in features and target variable
feature_cols = ['sex', 'maharat', 'roshd', 'khodpeiravi', 'ertebatm', 'rezayat', 'manaviat', 'shadi', 'behzisti']
X = le[feature_cols] # Features
y = le.Label # Target variable
```

```
# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1) # 70% training and 30% test
```

```
# Create Decision Tree classifier object
le = DecisionTreeClassifier()
```

```
# Train Decision Tree Classifier
le = le.fit(X_train, y_train)
```

```
#Predict the response for test dataset
y_pred = le.predict(X_test)
```

```
print("Accuracy:", metrics.accuracy_score(y_test, y_pred))
```

Accuracy: 0.8541666666666666



OUR TARGET

# دختران

```
[194] for variable,i in col_dict2.items():  
      print(variable,end=' = ')  
      print(girls[variable].mean())
```

```
maharat_ej = 75.4625  
raftar_gheir_ej = 17.5375  
parkhash = 19.8875  
bartari = 15.95  
rabete_hamsal = 24.6375  
maharat = 153.475
```

# پسران

```
▶ for variable,i in col_dict2.items():  
   print(variable,end=' = ')  
   print(boys[variable].mean())
```

```
↪ maharat_ej = 69.8875  
   raftar_gheir_ej = 19.75  
   parkhash = 23.325  
   bartari = 16.525  
   rabete_hamsal = 24.575  
   maharat = 154.0625
```

```
▶ meandict = {}  
  for col in le.columns:  
    # print(le[col].mean())  
    meandict[col] = le[col].mean()
```

meandict

```
↳ {'b1': 2.4375,  
   'b10': 2.3125,  
   'b11': 2.575,  
   'b12': 2.3375,  
   'b13': 2.875,  
   'b14': 2.53125,  
   'b15': 1.6625,  
   'b16': 2.00625,  
   'b17': 2.1875,  
   'b18': 3.0125,  
   'b19': 3.10625,  
   'b2': 2.4875,  
   'b20': 3.525,  
   'b21': 2.14375,  
   'b22': 2.1625,  
   'b23': 2.1,  
   'b24': 3.80625,  
   'b25': 2.70625,  
   'b26': 3.09375,  
   'b27': 1.5875,  
   'b28': 1.76875,  
   'b29': 3.375,  
   'b3': 1.80625,  
   'b30': 3.70625,  
   'b31': 2.21875,  
   'b32': 1.66875,  
   'b33': 2.10625,  
   'b34': 2.0,  
   'b35': 3.43125,
```

```
[175] for i in range(0,160):
      print('\n###   ', i, '   ###\n')
      x = le.iloc[i]
      # print(le.iloc[i])
      # print(x['b13'])
      for variable,j in meandict.items():
          diff = x[variable] - meandict[variable]
          if abs(diff) > 2:
              print(variable, x[variable], diff, sep='   ')
```



```
###   0   ###

b33  5.0  2.89375
b40  5.0  2.1
b41  5.0  2.14375
b77  5.0  2.0625
m19  5.0  3.26875
m39  5.0  2.35625
maharat 168.0 14.231249999999989
roshd 15.0 -5.21875
ertebatm 16.0 -2.5500000000000007
manaviat 22.0 -4.399999999999999
behzisti 170.0 -16.037499999999994
maharat ej 89.0 16.325000000000003
parkhash 17.0 -4.606249999999999

###   1   ###

m15  2.0 -2.1125
m20  4.0  2.40625
m31  5.0  2.48125
m39  5.0  2.35625
m52  4.0  2.04375
m55  5.0  2.89375
maharat 170.0 16.231249999999999
roshd 18.0 -2.21875
manaviat 21.0 -5.399999999999999
behzisti 171.0 -15.037499999999994
maharat ej 80.0 7.325000000000003
parkhash 26.0 4.393750000000001
rabete hamsal 28.0 3.3937500000000007

###   2   ###
```

Do you have any questions?

**THANKS**