Data Source : https://www.kaggle.com/abhilash04/fathersandsonheight

Import Libraries

Let's import some libraries to get started

In [31]:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import seaborn as sns
```

In [32]:

```
train = pd.read_csv("Pearson_data.csv")
```

In [33]:

```
train.head()
```

Out[33]:

	Father	Son
0	65.0	59.8
1	63.3	63.2
2	65.0	63.3
3	65.8	62.8
4	61.1	64.3

Exploratory Data Analysis

Missing Data

In [34]:

```
train.isnull()
```

Out[34]:

	Father	Son
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
...	...	...
1073	False	False
1074	False	False
1075	False	False
1076	False	False

1077	False	False

1078 rows × 2 columns

In [35]:

```
train.isnull().sum()
```

Out[35]:

```
Father    0
Son       0
dtype: int64
```

In [36]:

```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1078 entries, 0 to 1077
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Father  1078 non-null   float64
 1   Son     1078 non-null   float64
dtypes: float64(2)
memory usage: 17.0 KB
```

In [37]:

```
train.describe()
```

Out[37]:

	Father	Son
count	1078.000000	1078.000000
mean	67.686827	68.684230
std	2.745827	2.816194
min	59.000000	58.500000
25%	65.800000	66.900000
50%	67.800000	68.600000
75%	69.600000	70.500000
max	75.400000	78.400000

Data Visualization

In [38]:

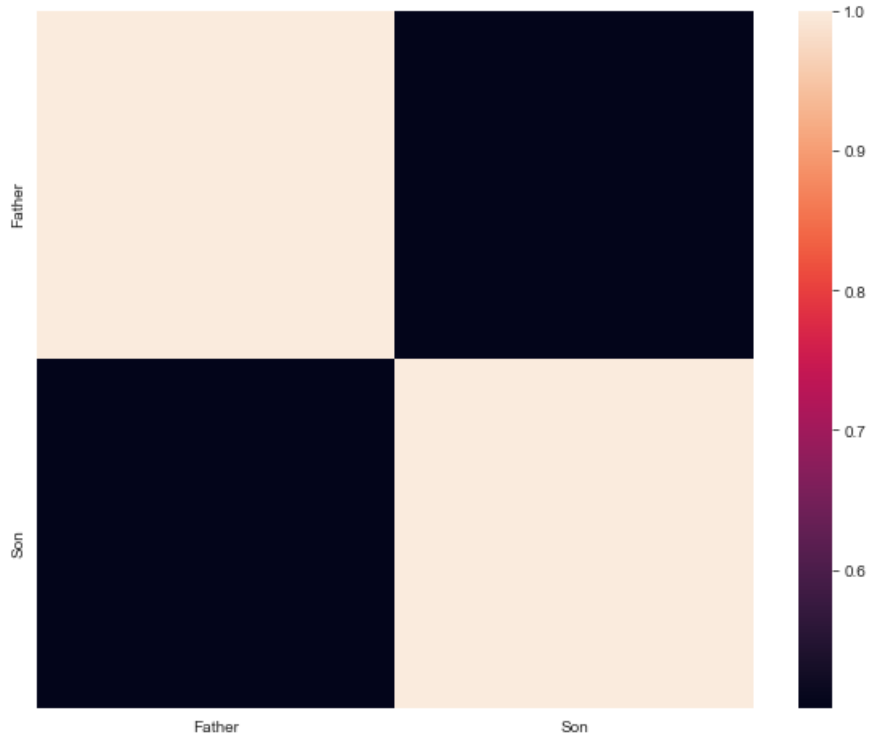
```
import plotly.express as px
fig = px.bar(train,x='Father',y = 'Son',height = 500)
fig.update_layout(title = "Father's age corresponding son's age",titlefont_size=20)
fig.show()
```

In [39]:

```
import seaborn as sns
plt.figure(figsize=(10,8))
sns.heatmap(train.corr())
```

Out[39]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3cc293240>
```



In [ ]: