In [1]:
```
def prepare_country_stats(oecd_bli, gdp_per_capita):
    oecd_bli = oecd_bli[oecd_bli["INEQUALITY"] == "TOT"]
    oecd_bli = oecd_bli.pivot(index="Country", columns="Indicator", values="Value")
    gdp_per_capita.rename(columns={"2015":"GDP per capita"}, inplace=True)
    gdp_per_capita.set_index("Country", inplace=True)
    full_country_stats = pd.merge(left=oecd_bli, right=gdp_per_capita,
                                  left_index = True, right_index = True)
    full_country_stats.sort_values(by="GDP per capita", inplace= True)
    remove_indices = [0,1,6,8,33,34,35]
    keep_indices = list(set(range(36)) - set(remove_indices))
    return full_country_stats[["GDP per capita","Life satisfaction"]].iloc[keep_indices]
```

In [2]:
```
import matplotlib.pyplot as plt
```

In [3]:
```
import numpy as np
import pandas as pd
import sklearn.linear_model
```

In [4]:
```
# Load the data
oecd_bli = pd.read_csv("handson-ml\datasets\lifesat\oecd_bli_2015.csv", thousands=',')
gdp_per_capita = pd.read_csv("handson-ml\datasets\lifesat\gdp_per_capita.csv", thousands=",", delimiter='\t',
                             encoding='latin1', na_values="n/a")
```

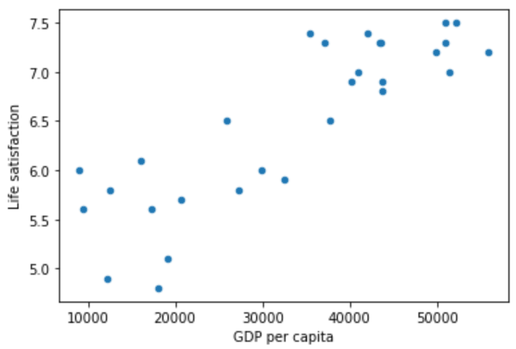
In [5]:
```
gdp_per_capita.head(5)
```

Out[5]:

	Country	Subject Descriptor	Units	Scale	Country/Series-specific Notes	2015	Estimates Start After
0	Afghanistan	Gross domestic product per capita, current prices	U.S. dollars	Units	See notes for: Gross domestic product, curren...	599.994	2013.0
1	Albania	Gross domestic product per capita, current prices	U.S. dollars	Units	See notes for: Gross domestic product, curren...	3995.383	2010.0
2	Algeria	Gross domestic product per capita, current prices	U.S. dollars	Units	See notes for: Gross domestic product, curren...	4318.135	2014.0
3	Angola	Gross domestic product per capita, current prices	U.S. dollars	Units	See notes for: Gross domestic product, curren...	4100.315	2014.0
4	Antigua and Barbuda	Gross domestic product per capita, current prices	U.S. dollars	Units	See notes for: Gross domestic product, curren...	14414.302	2011.0

In [6]:
```
   # Prepare the data
country_stats = prepare_country_stats(oecd_bli, gdp_per_capita)
x = np.c_[country_stats["GDP per capita"]]
y = np.c_[country_stats["Life satisfaction"]]
```

In [7]:
```
# Visualize the data
country_stats.plot(kind="scatter", x="GDP per capita", y="Life satisfaction")
plt.show()
```



In [8]:
```
# Select a linear model
model = sklearn.linear_model.LinearRegression()
```

In [9]:
```
# Train model
model.fit(x,y)
```

Out[9]:
```
LinearRegression()
```

In [10]:
```
# Make a prediction for Cyprus
x_new = [[22587]] # Cyprus' GDP per capita
print(model.predict(x_new))
```

```
[[5.96242338]]
```

In [ ]:
```

```