

```
[ ] import pandas as pd          # Datensätze erstellen, bearbeiten,...
import numpy as np              # mathematische Berechnungen
import matplotlib.pyplot as plt # Visualisierungen
from sklearn.linear_model import LinearRegression # Lineare Regression
```

▼ Datensatz

Hier erstellen wir den Datensatz mit Alter und Größen von Mädchen und lassen ihn anzeigen.

```
[ ] daten = {'Name': ['Mia', "Tilda", "Sarah", "Amelie", "Emma", "Frida", "Clara", "Hannah", "Sophie"],
            'Alter': [9, 10, 11, 12, 13, 14, 15, 16, 17],
            'Größe': [135, 139, 144, 149, 154, 158, 162, 164, 165]}
datensatz = pd.DataFrame(daten) # hier erstellen wir den Datensatz
datensatz.set_index("Name", inplace=True) # als Index setzen wir die Namen der Mädchen
```

```
[ ] # Datensatz anzeigen lassen
datensatz
```

Name	Alter	Größe
Mia	9	135
Tilda	10	139
Sarah	11	144
Amelie	12	149
Emma	13	154
Frida	14	158
Clara	15	162
Hannah	16	164
Sophie	17	165

```
[ ] # auf einzelne Spalten zugreifen
alter = datensatz['Alter']
größe = datensatz['Größe']
```

▼ Minimum, Maximum und Durchschnitt

So berechnen wir Minima, Maxima sowie Durchschnitte unseres Datensatzes.

Das können wir entweder spaltenweise machen oder für alle Spalten zusammen.

```
[ ] # Maximum: Größe des größten Mädchens bestimmen
größe.max()
165
```

```
[ ] # Index des Maximums: Name des größten Mädchens bestimmen
größe.idxmax()
'Sophie'
```

```
[ ] # Minimum: Alter des jüngsten Mädchens bestimmen
alter.min()
9
```

```
[ ] # Index des Minimums: Name des jüngsten Mädchens bestimmen
alter.idxmin()
'Mia'
```

```
[ ] # Durchschnittsgröße der Mädchen
größe.mean()
152.22222222222223
```

```
[ ] # Durchschnittsgröße und -alter auf einmal ausgeben
datensatz.mean()

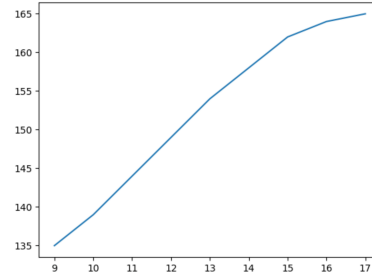
Alter    13.000000
Größe    152.222222
dtype: float64
```

▼ Grafiken (Plots) erstellen

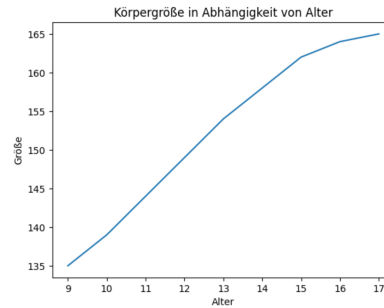
Wir erstellen eine Grafik mithilfe der Daten, die Körpergröße in Abhängigkeit von Alter anzeigt.

```
[ ] # Plot erstellen

plt.plot(alter,größe)
plt.show() # wichtig damit der Plot auch angezeigt wird
```

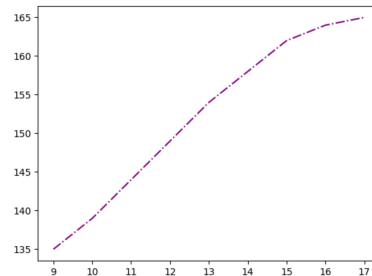


```
[ ] # Titel und Achsenbeschriftung hinzufügen
plt.plot(alter,größe)
plt.title("Körpergröße in Abhängigkeit von Alter") # Titel hinzufügen
plt.xlabel("Alter") # Beschriftung x-Achse hinzufügen
plt.ylabel("Größe") # Beschriftung y-Achse hinzufügen
plt.show()
```



```
[ ] # Linienfarbe und -art ändern
plt.plot(alter,größe, color = "purple", linestyle = "dashdot")
plt.show()

# mögliche Farben: "yellow", "orange", "red", "blue", "green", "black",...
# mögliche Linienarten: "solid", "dotted", "dashed", "dashdot",...
```



▼ Lineare Regression

Mithilfe des Modells der *linearen Regression* können wir einen linearen Zusammenhang zwischen Alter und Körpergröße finden.

```
[ ] # 1) hier treffen wir einige Vorbereitungen, um das Modell anwenden zu können
# die Spalten müssen eine bestimmte Form haben
X = alter.array.reshape(-1,1)
y = größe.array.reshape(-1,1)

# 2) Erstelle das Modell "Lineare Regression"
reg = LinearRegression()

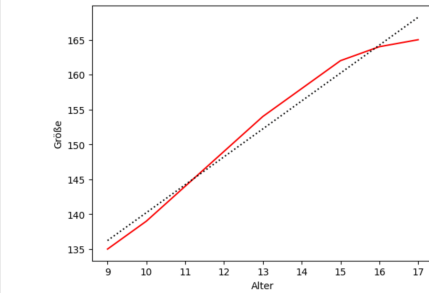
# 3) Lasse das Modell die Rechnungen ausführen, um es zu erstellen
reg.fit(X, y)

# 4) Berechne Vorhersagewerte anhand des Modells
model = reg.predict(X)
model

array([[136.22222222],
       [140.22222222],
       [144.22222222],
       [148.22222222],
       [152.22222222],
       [156.22222222],
       [160.22222222],
       [164.22222222],
       [168.22222222]])
```

```
[ ] # Grafische Veranschaulichung
```

```
plt.plot(alter,größe, color = "red")
plt.plot(alter, model, color = "black", linestyle = "dotted")
plt.xlabel('Alter')
plt.ylabel('Größe')
plt.show()
```



Vorhersage der Größe eines Mädchens mit 18 Jahren:

```
[ ] alter_vorhersage = np.array([18])
reg.predict(alter_vorhersage)

array([[172.22222222]])
```

▼ Korrelation

Die Korrelation zwischen zwei Größen (z.B. Alter und Körpergröße) gibt an, wie stark ihr linearer Zusammenhang ist. Je näher der Wert an 1 liegt, desto stärker ist der Zusammenhang.

```
[ ] # so können wir Korrelationen zwischen den einzelnen Spalten des Datensatzes anzeigen lassen
datensatz.corr()
```

	Alter	Größe
Alter	1.000000	0.987953
Größe	0.987953	1.000000