

Multiple Regression mit `california_housing.csv`



Du arbeitest mit dem California Housing Dataset, das Informationen über Häuserpreise und verschiedene Merkmale der Umgebung enthält. Ziel ist es, eine Multiple Regression zu erstellen, um die mittleren Hauspreise (`MedHouseVal`) anhand der verfügbaren Merkmale vorherzusagen.

- Erzeuge aus `california_housing.csv` ein `DataFrame`
- Mach dich mit den Daten vertraut, gibt es Datenlücken die behandelt werden müssen ?
- Erzeuge einen `seaborn heatmap` für die Korrelationsmatrix des `DataFrames`. Was ist die unabhängige Variable, was sind die abhängigen Variablen ?
- Erzeuge einen `seaborn pairplot` für die numerischen Merkmale
- Überlege dir welche Merkmale sich für eine Regression eignen könnten und probiere verschiedene Kombinationen aus
- Für jede Kombination:
 - Teile den Datensatz in Trainings- und Testdaten auf (80% Training, 20% Test).
 - Erzeuge ein Modell der linearen Regression und trainiere es mit den Testdaten
 - Bewerte die Modellleistung mit Hilfe der `score()` -Methode jeweils für Test- und Trainset
 - Ermittle die Parameter (`coef` und `intercept`) deines besten Modells

