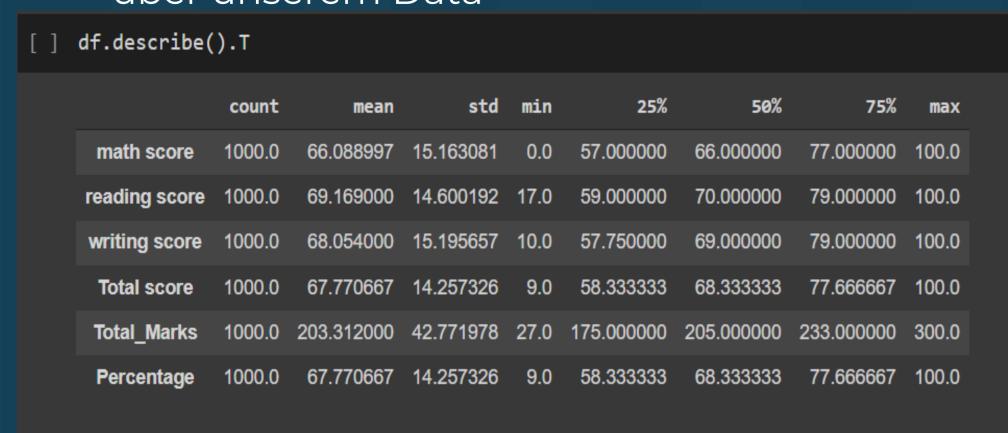
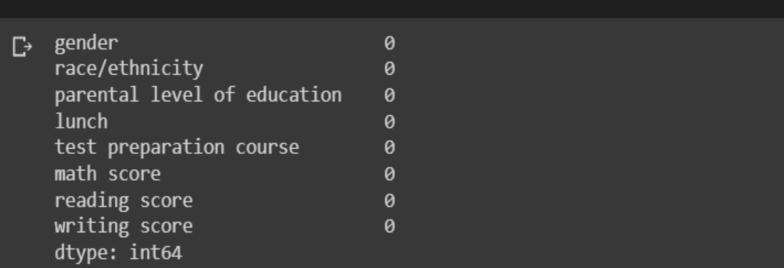
EDA und einfaches Daten über unserem Data



Gucken wir nun data info: df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 1000 entries, 0 to 999 Data columns (total 15 columns): Non-Null Count Dtype Column gender 1000 non-null object race/ethnicity 1000 non-null object parental level of education 1000 non-null object 1unch 1000 non-null object test preparation course object 1000 non-null float32 math score 1000 non-null 1000 non-null reading score writing score 1000 non-null Writing_PassStatus 1000 non-null object 1000 non-null Math_PassStatus object Reading_PassStatus 1000 non-null object OverAll_PassStatus 1000 non-null object 12 Total score 1000 non-null float64 13 Total Marks 1000 non-null float64 14 Percentage 1000 non-null float64 dtypes: float32(1), float64(3), int64(2), object(9) memory usage: 113.4+ KB

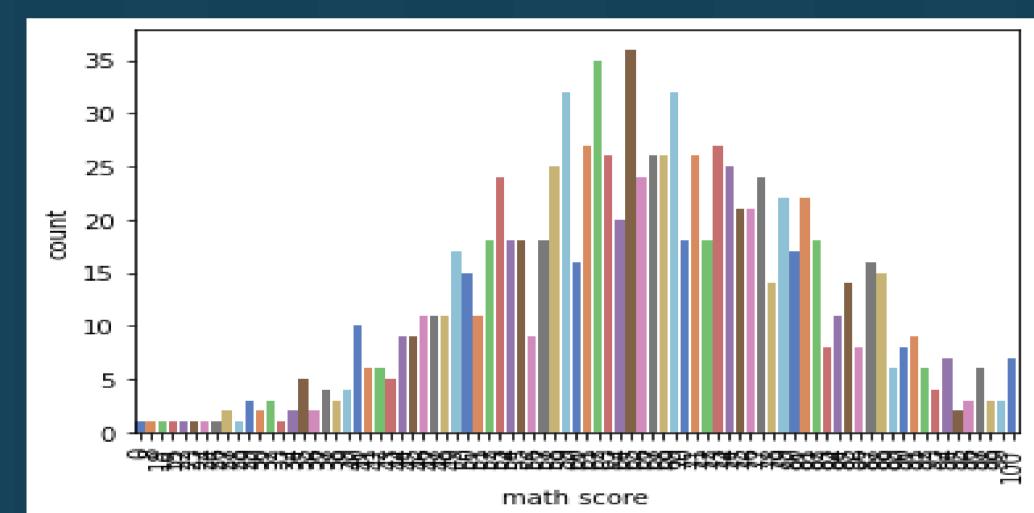
Lass uns gucken, ob es sogenannte missing values(fehlschlagene Werte) gibt df.isnull().sum()

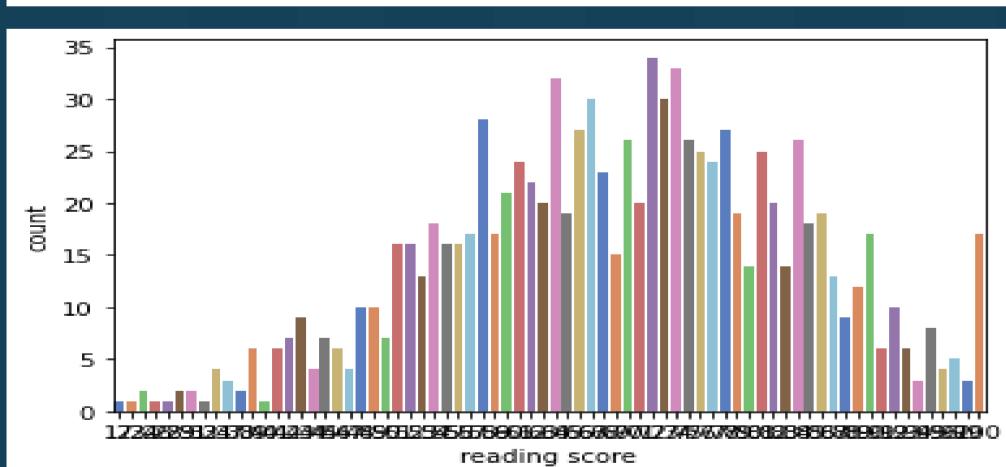


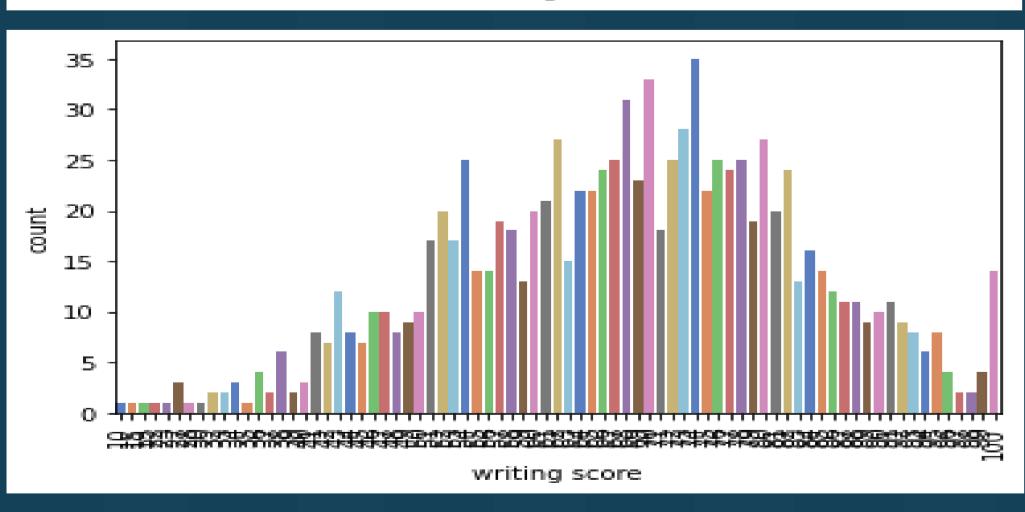
wir sehen hier keine missing Values (null Werte). Aber in echtes Leben ist es nicht so.

etwas einfaches Informationen über unserem Data print(df.describe())

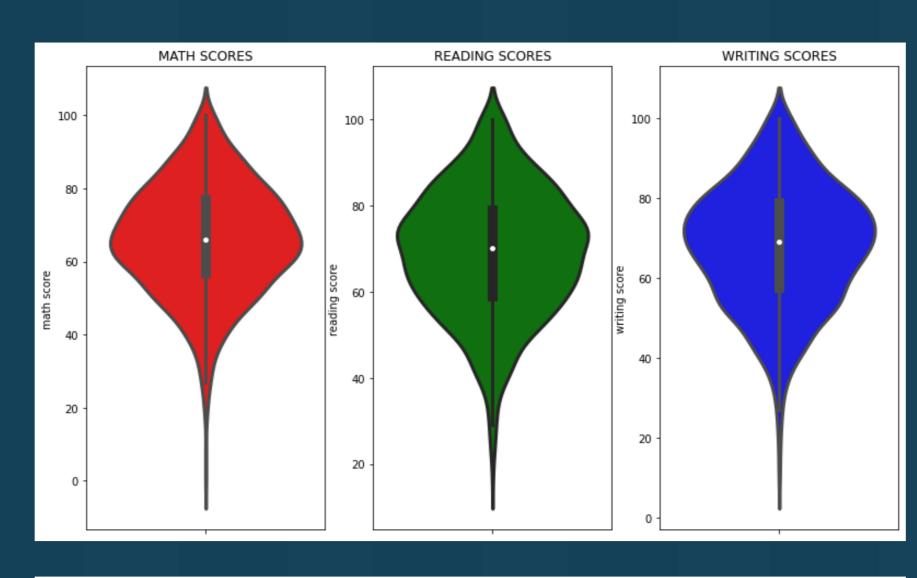
	math score	reading score	writing score
count	1000.00000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000
std	15.16308	14.600192	15.195657
min	0.00000	17.000000	10.000000
25%	57.00000	59.000000	57.750000
50%	66.00000	70.000000	69.000000
75%	77.00000	79.000000	79.000000
max	100.00000	100.000000	100.000000

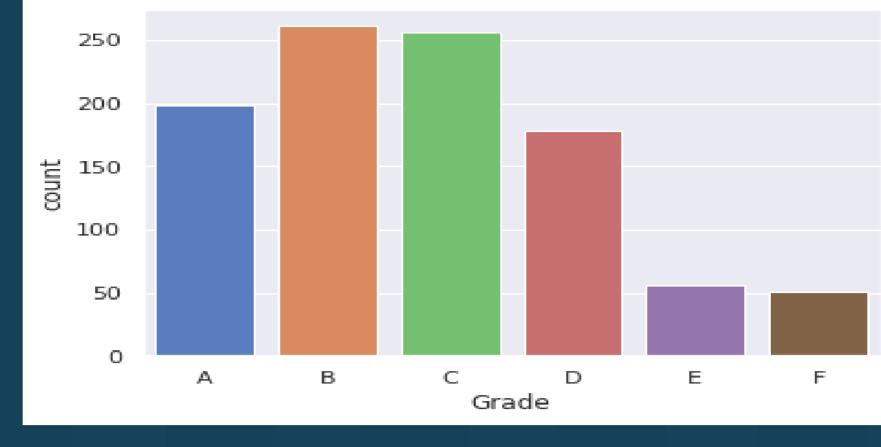




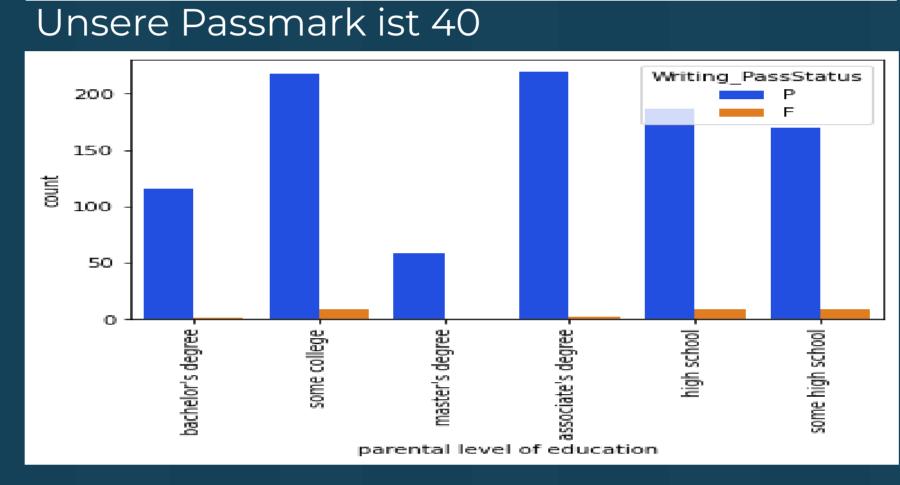


Mean Werte von Scores





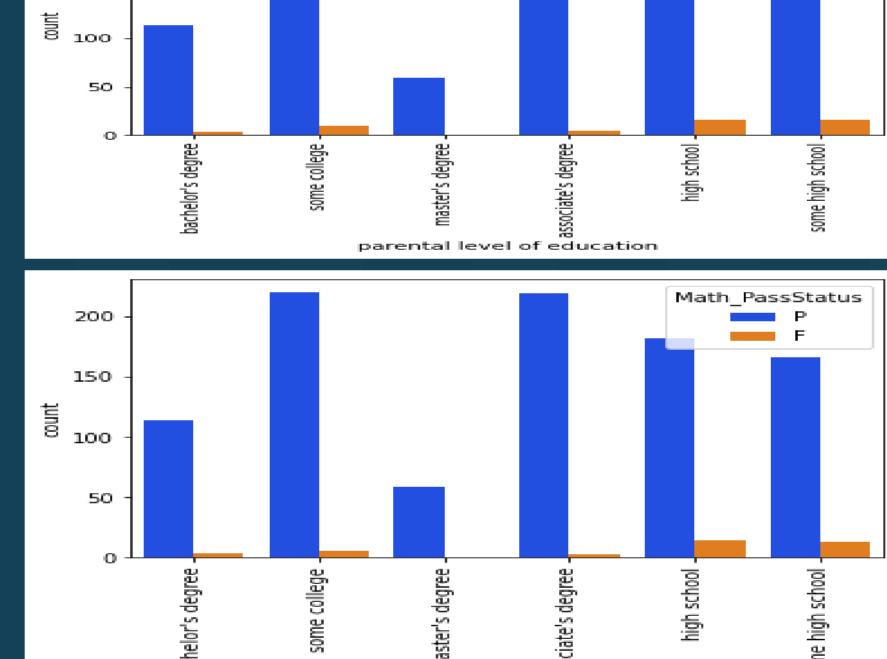


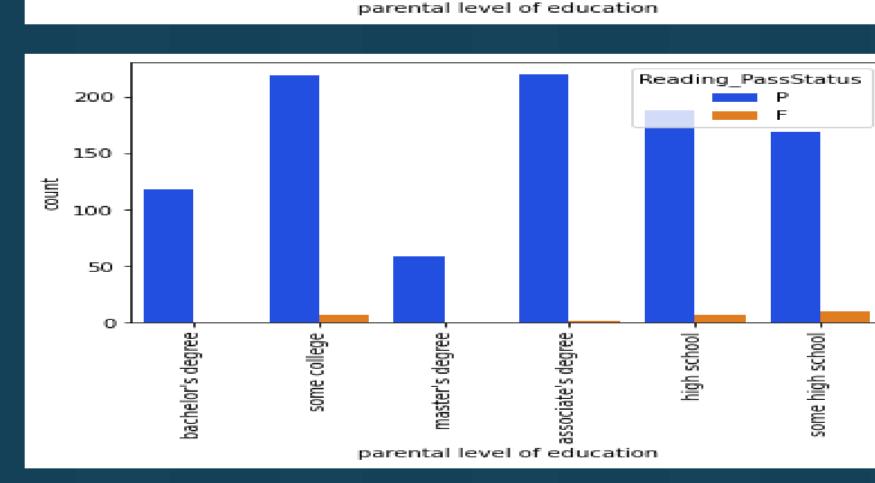


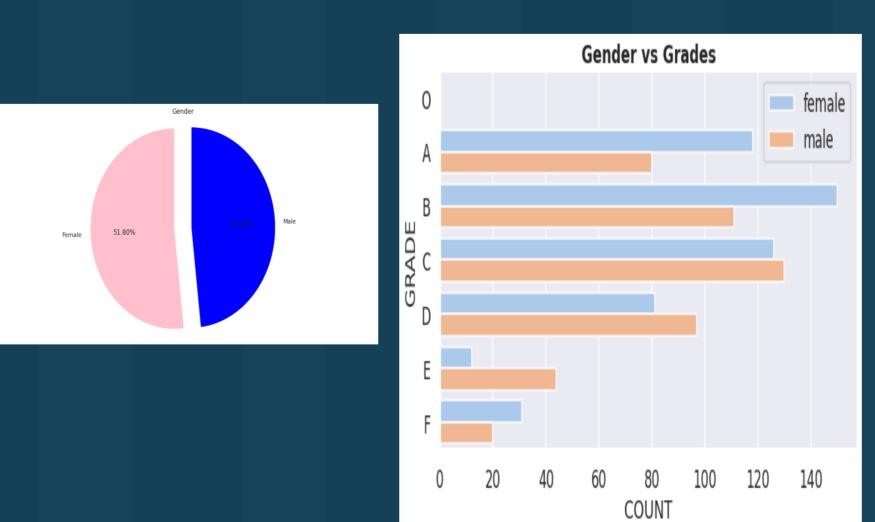
200

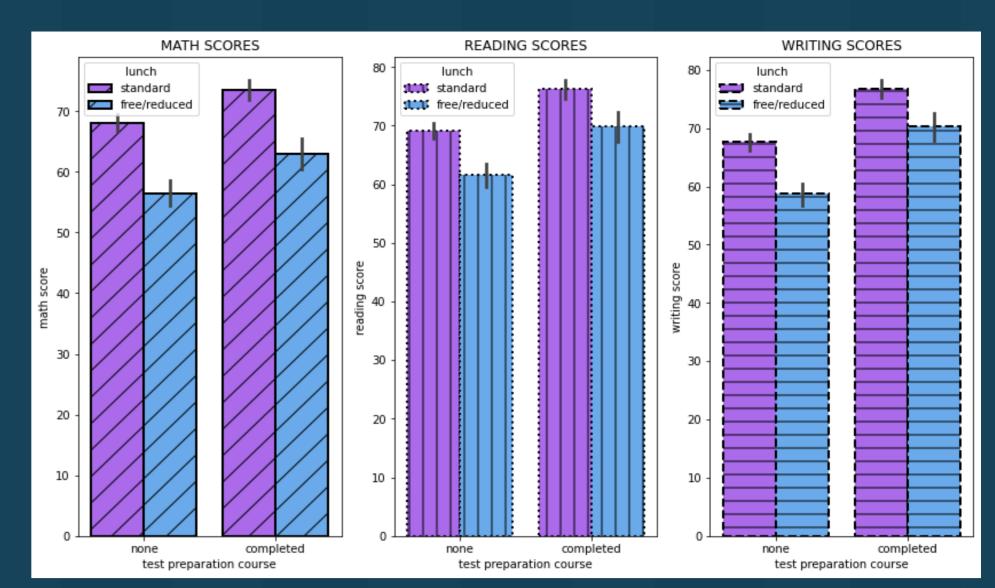
150

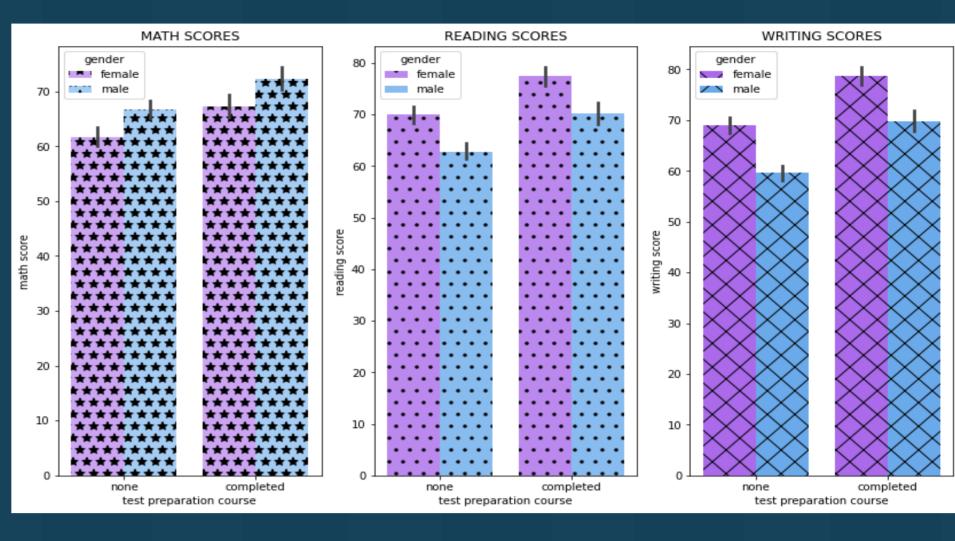
OverAll_PassStatus

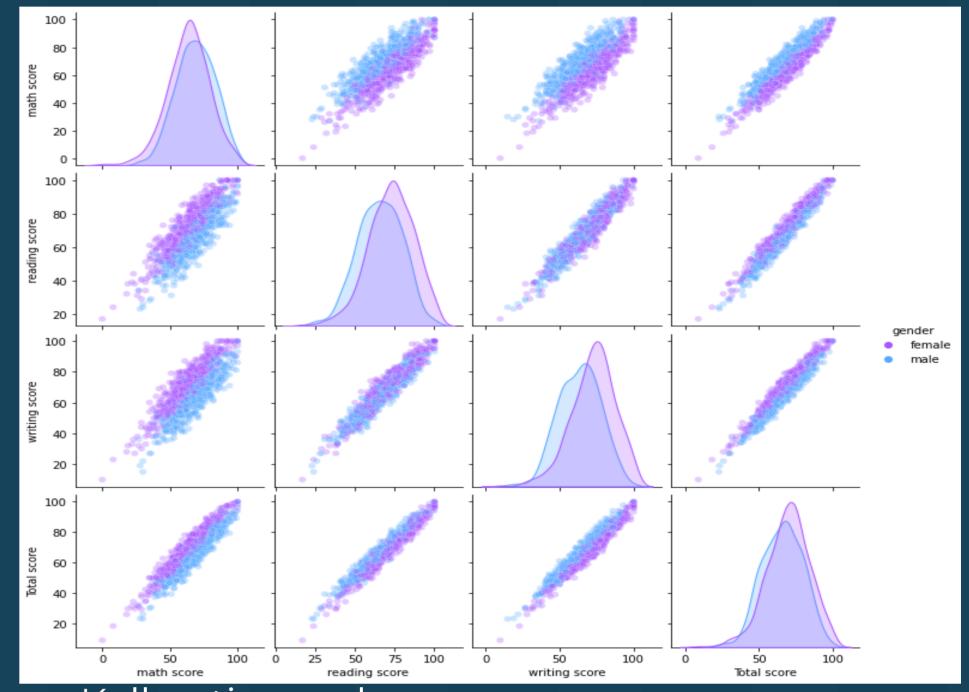


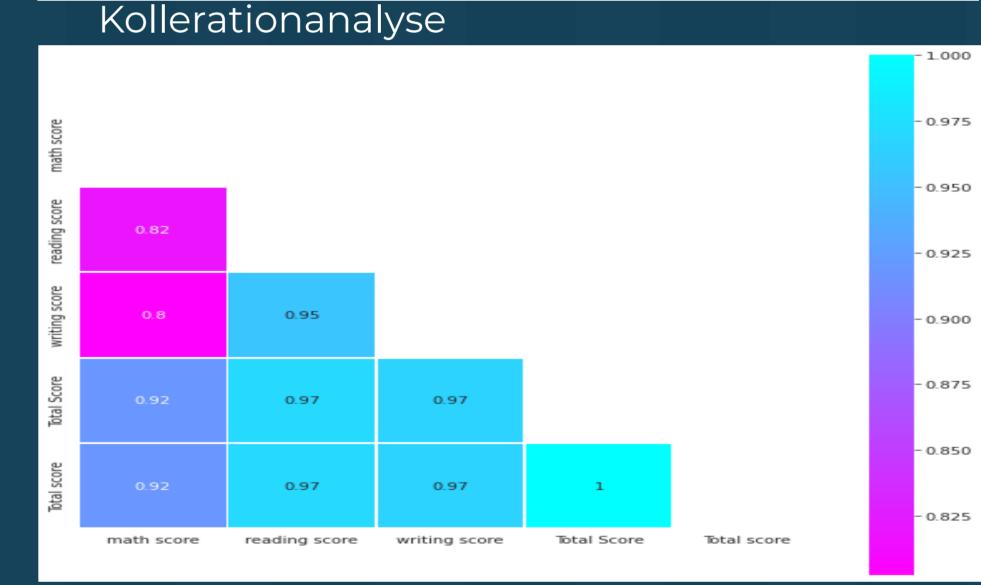


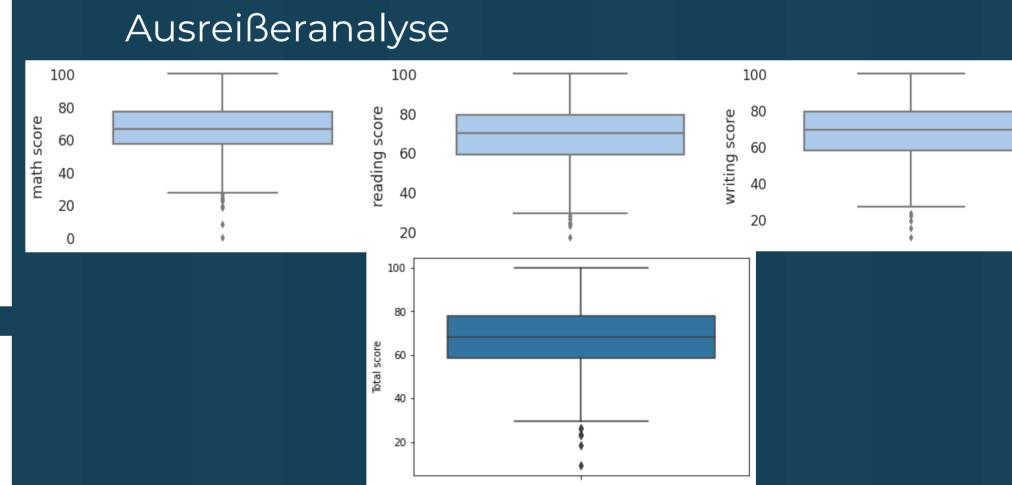




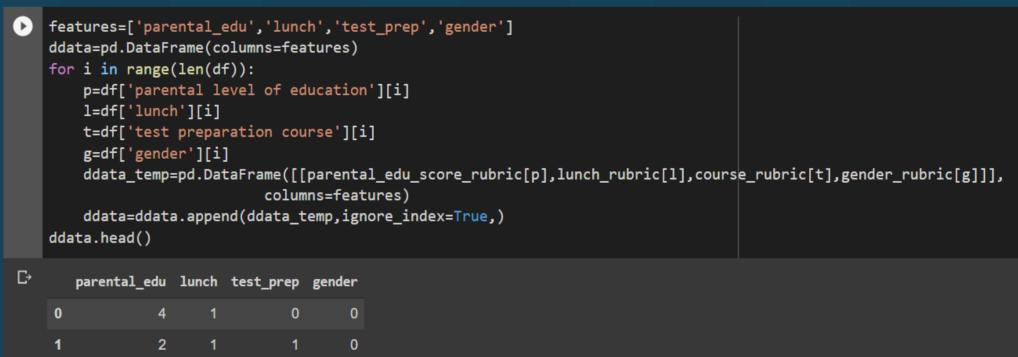








Linear Regression

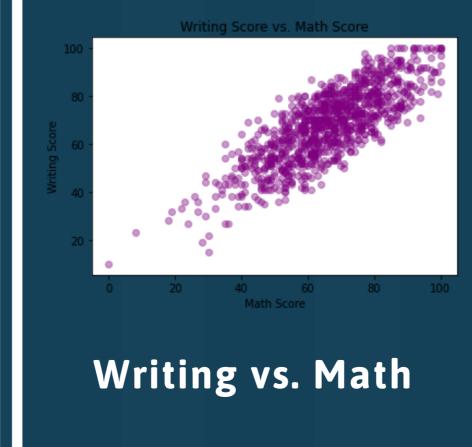


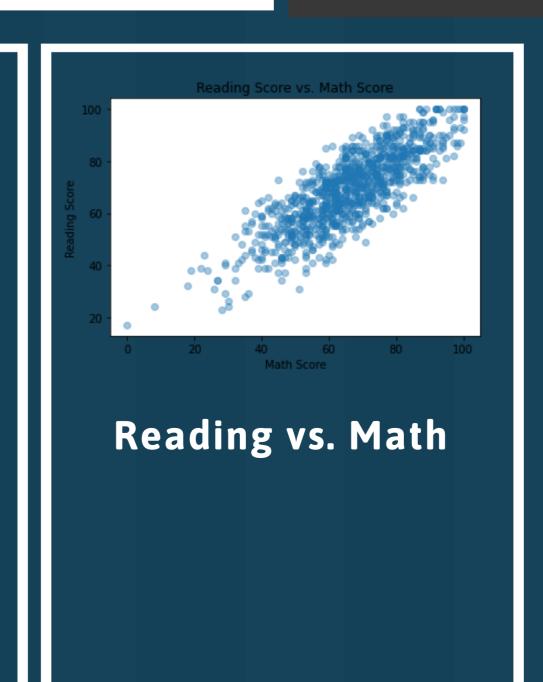
4	2	1	0	1			
regr:	model	fit(d	data,t	earRegression(otal_score) ef [0],regr.co) ef [1],regr.coe	f [2],regr.coe	f [3])

print('R^2: ',regr.score(ddata,total_score))

coef: 6.121461593336114 26.998647429712427 23.55371257956297 -11.099243727699289 R^2: 0.2154490486266485



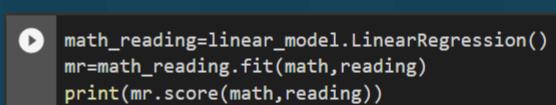




Regression der Gesamtpunktzahl auf einer Liste von Merkmalen, einschließlich Geschlecht, Bildungsstand der Eltern, Prüfungsvorbereitungskurs und Mittagessen Der niedrige R^2-Wert (nur 0,21) zeigt an, dass die Korrelation nicht signifikant ist.

Wie die folgende Regression zeigt, besteht ein gewisses Maß an Korrelation zwischen dem Mathematikergebnis und dem Leseergebnis; Es besteht eine extrem starke Korrelation zwischen Lese- und Schreibpunktzahl, die auch intuitiv Sinn macht.

Regression Werte



math_writing=linear_model.LinearRegression() mwr=math_writing.fit(math,writing) print(mwr.score(math,writing))

rw=reading_writing.fit(reading,writing)

reading_writing=linear_model.LinearRegression()

print(rw.score(reading,writing)) € 0.6684365064501092

0.6442342539264934 0.9112574888913137