CEM KIRAÇ - HW2 REPORT – Advanced data analysis in python

My aim was to predict **gdp_per_capita** of countries in this study. I chose some explanatory variables and countries based on my personal judgement, whichI thought would have a correlation with gdp_per_capita. I tried to choose them from different fields to avoid multicollinearity, such as health, economics, education and law. The data source is the worldbank API. You can find the variables in the first table below. The 2nd table is the results of statsmodels library of python, just to check for if there are any differences between two linear regressions. I was surprised to see that there is no significant relationship between income_inequality and gdp_per_capita. Similarly I was expecting tertiary education of women would mean a higher gdp_per_capita but that wasnt the case. Life expectancy and rule of law, both has a very strong relationship just as expected. Employment_rate in younger people and homicide index are also significantly related to gdp. High_tech_export percentage relationship to gdp was insignificant.

Linear Regression output

Covariates	Coefficients	StandardErrors	Interval_Min	Interval_Max
Constant	-138117	46224.6	-229638	-46595
LIFE_EXPECTANCY	1607.29	430.63	754.674	2459.91
YOUTH_UNEMPLOYED	-324.887	53.735	-431.279	-218.496
HIGH_TECH_EXPORTS	-90.0541	88.3351	-264.952	84.8433
HOMICIDE_INDEX	-356.47	148.752	-650.99	-61.9508
RULE_OF_LAW_INDEX	303.544	65.0031	174.842	432.245
INCOME_INEQUALITY	152.677	189.62	-222.757	528.112
GENDER_PARITY_TERTIARY_EDUCATION	31487.3	49014	-65557	128532

Statsmodels Linear Reg. Output

OLS Regression Results

Dep. Variable:	y	R-squared:	0.814
Model:	OLS	Adj. R-squared:	0.803
Method:	Least Squares	F-statistic:	74.88
Date:	Thu, 16 Dec 2021	Prob (F-statistic):	8.86e-41
Time:	02:07:11	Log-Likelihood:	-1312.0
No. Observations:	128	AIC:	2640.
Df Residuals:	120	BIC:	2663.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1.381e+05	4.62e+04	-2.988	0.003	-2.3e+05	-4.66e+04
x1	1607.2914	430.630	3.732	0.000	754.674	2459.909
x2	-324.8873	53.735	-6.046	0.000	-431.279	-218.496
x3	-90.0541	88.335	-1.019	0.310	-264.952	84.843
x4	-356.4703	148.752	-2.396	0.018	-650.990	-61.951
x5	303.5438	65.003	4.670	0.000	174.842	432.245
x6	152.6774	189.620	0.805	0.422	-222.757	528.112
x7	3.149e+04	4.9e+04	0.642	0.522	-6.56e+04	1.29e+05

Omnibus:	5.937	Durbin-Watson:	0.591
Prob(Omnibus):	0.051	Jarque-Bera (JB):	6.074
Skew:	0.528	Prob(JB):	0.0480
Kurtosis:	2.847	Cond. No.	1.22e+04