

Linear Regression Analysis 7 and 8 Variables

March 14, 2020

```
[3]: import pandas as pd
from sklearn.linear_model import LinearRegression
import scipy.stats as stats
import statsmodels.api as sm
from statsmodels.formula.api import ols
import matplotlib.pyplot as plt
import numpy as np
pd.set_option('display.max_rows', 600)
gapminder = pd.read_csv('gapminder.csv')
gapminder = gapminder.loc[gapminder['time'] <= 2020]
gapminder = gapminder.loc[gapminder['time'] >= 1970]
gapminderVariables = list()
for col in gapminder.columns:
    if(col != 'geo' and col != 'time' and col !=_
    ↪'child_mortality_0_5_year_olds_dying_per_1000_born' and col !=_
    ↪'child_mortality_0_5_year_olds_more_years_version_7' and col !=_
    ↪'under_five_mortality_from_cme_per_1000_born' and col !=_
    ↪'infant_mortality_rate_per_1000_births' and col !=_
    ↪'under_five_mortality_from_ihme_per_1000_born' and col !=_
    ↪'dead_kids_per_woman' and col != 'newborn_mortality_rate_per_1000' and col !
    ↪= 'all_causes_deaths_in_children_1_59_months_per_1000_births' and col !=_
    ↪'all_causes_deaths_in_newborn_per_1000_births' and col !=_
    ↪'maternal_mortality_ratio_who' and col !=_
    ↪'maternal_mortality_ratio_per_100000_live_births' and col !=_
    ↪'prematurity_deaths_in_newborn_per_1000_births' and col !=_
    ↪'stillbirths_per_1000_births' and col != 'life_expectancy_female' and col !=_
    ↪'life_expectancy_male' and col != 'hdi_human_development_index' and col !=_
    ↪'life_expectancy_at_birth_with_projections' and col !=_
    ↪'life_expectancy_years' and col != 'life_expectancy_at_birth_data_from_ihme'_
    ↪and col != 'literacy_rate_youth_total_percent_of_people_ages_15_24' and col !
    ↪= 'North America' and col != 'Central America' and col != 'South America'_
    ↪and col != 'Europe' and col != 'Balkans' and col != 'Middle East' and col !=_
    ↪'Coastal Africa' and col != 'Landlocked Africa' and col != 'Northern Asia'_
    ↪and col != 'Southern Asia' and col != 'Pacific Islands' and col != 'Greater_
    ↪Australia'):
        gapminderVariables.append(col)
```

```

        elif(col != 'geo' and col != 'time' and col != 'id'):
            ↪ 'child_mortality_0_5_year_olds_dying_per_1000_born'):
                gapminder = gapminder.drop(columns = col, axis = 1)
correlationStudy = gapminder.corr()
childMortalityCorrelation = correlationStudy.
    ↪ child_mortality_0_5_year_olds_dying_per_1000_born.sort_values()
print(childMortalityCorrelation)
highCorrelationVariables = gapminder[['geo', 'time',
    ↪ 'child_mortality_0_5_year_olds_dying_per_1000_born',
    ↪ 'primary_school_completion_percent_of_girls',
    ↪ 'at_least_basic_water_source_overall_access_percent',
    ↪ 'at_least_basic_sanitation_overall_access_percent',
    ↪ 'literacy_rate_youth_female_percent_of_females_ages_15_24',
    ↪ 'sepsis_deaths_in_newborn_per_1000_births',
    ↪ 'diarrhoeal_deaths_in_newborn_per_1000_births',
    ↪ 'pneumonia_deaths_in_children_1_59_months_per_1000_births',
    ↪ 'birth_asphyxia_deaths_in_newborn_per_1000_births']]
highCorrelationVariables = highCorrelationVariables.
    ↪ loc[highCorrelationVariables['time'] <= 2020]
highCorrelationVariables = highCorrelationVariables.
    ↪ loc[highCorrelationVariables['time'] >= 1970]
vifCalculation = highCorrelationVariables
currentCountry = ''
previousCountry = ''
averages = list()
currentCountry = ''
countryData = pd.DataFrame()
countryList = list()
for country in vifCalculation['geo']:
    if(country != currentCountry):
        countryData = vifCalculation[vifCalculation['geo'] == country]
        previousCountry = currentCountry
        currentCountry = country
        countryList.append(countryData)
for country in countryList:
    for variable in country.columns:
        if(variable != 'geo' and variable != 'time' and variable != 'id'):
            count = vifCalculation[variable].count()
            summation = vifCalculation[variable].sum()
            average = summation/count
            vifCalculation[variable] = vifCalculation[variable].fillna(average)
print(vifCalculation)
linregression = LinearRegression()
vif = ""
for x in vifCalculation:
    for y in vifCalculation:
        if(x != y):

```

```

if(y != 'geo' and y != 'time' and x != 'geo' and x != 'time'):
    xValues = vifCalculation[x].values.reshape(-1,1)
    yValues = vifCalculation[y].values.reshape(-1,1)
    linregression.fit(xValues, yValues)
    prediction = linregression.predict(xValues)
    score = linregression.score(xValues, yValues)
    if(score < 1):
        vif = str(1 / (1 - score))
    else:
        vif = ""
        vif = "estimably equal to 1. This will cause a divide by_
→zero error"

    print("When " + y + " is compared to " + x + " the relationship_
→has an R Square equal to " + str(score) + " and a variance inflation factor_
→" + vif + ".\n")

```

```

primary_school_completion_percent_of_girls
-0.855468
at_least_basic_water_source_overall_access_percent
-0.852226
at_least_basic_sanitation_overall_access_percent
-0.848195
literacy_rate_youth_female_percent_of_females_ages_15_24
-0.842420
primary_completion_rate_total_percent_of_relevant_age_group
-0.836248
at_least_basic_sanitation_urban_access_percent
-0.828710
literacy_rate_adult_total_percent_of_people_ages_15_and_above
-0.824839
at_least_basic_water_source_rural_access_percent
-0.821059
mean_years_in_school_women_of_reproductive_age_15_to_44
-0.819884
mean_years_in_school_women_15_to_24_years
-0.818229
literacy_rate_youth_male_percent_of_males_ages_15_24
-0.813916
literacy_rate_adult_female_percent_of_females_ages_15_above
-0.813607
births_attended_by_skilled_health_staff_percent_of_total
-0.811122
literacy_rate_adult_male_percent_of_males_ages_15_and_above
-0.811102
at_least_basic_sanitation_rural_access_percent
-0.806423
mean_years_in_school_men_15_to_24_years

```

-0.805339
at_least_basic_water_source_urban_access_percent
-0.800932
mean_years_in_school_women_25_to_34_years
-0.800178
primary_school_completion_percent_of_boys
-0.793543
mean_years_in_school_men_25_to_34_years
-0.791785
mean_years_in_school_women_percent_men_25_to_34_years
-0.784416
ratio_of_girls_to_boys_in_primary_and_secondary_education_perc
-0.783895
contraceptive_use_percent_of_women_ages_15_49
-0.775856
mean_years_in_school_men_25_years_and_older
-0.775669
mean_years_in_school_men_35_to_44_years
-0.771402
mean_years_in_school_women_25_years_and_older
-0.762752
mean_years_in_school_women_35_to_44_years
-0.761273
ratio_of_young_literate_females_to_males_percent_ages_15_24
-0.748589
mean_years_in_school_men_45_to_54_years
-0.746449
mean_years_in_school_men_55_to_64_years
-0.723205
mean_years_in_school_women_45_to_54_years
-0.723056
dtp3_immunized_percent_of_one_year_old
-0.717784
cholesterol_fat_in_blood_men_mmolperl
-0.716004
math_achievement_4th_grade
-0.715412
body_mass_index_bmi_men_kgperm2
-0.713236
female_service_workers_percent_of_female_employment
-0.690792
mean_years_in_school_men_65_plus_years
-0.689566
mean_years_in_school_women_55_to_64_years
-0.685860
cholesterol_fat_in_blood_women_mmolperl
-0.676642
male_service_workers_percent_of_male_employment

-0.673795
food_supply_kilocalories_per_person_and_day
-0.666601
math_achievement_8th_grade
-0.657882
gdpperemployee_us_inflation_adjusted
-0.655953
sugar_per_person_g_per_day
-0.654651
body_mass_index_bmi_women_kgperm2
-0.653836
male_industry_workers_percent_of_male_employment
-0.651935
median_age_years
-0.648601
urban_population_percent_of_total
-0.645050
mean_years_in_school_women_65_plus_years
-0.642192
mcv_immunized_percent_of_one_year-olds
-0.640909
female_salaried_workers_percent_of_non_agricultural_female_employment
-0.617336
age_at_1st_marriage_women
-0.614997
salaried_workers_percent_of_non_agricultural_employment
-0.604993
population_aged_40_59_years_male_percent
-0.598606
smoking_adults_percent_of_population_over_age_15
-0.598527
roads_paved_percent_of_total_roads
-0.596992
population_aged_40_59_years_both_sexes_percent
-0.595854
fixed_line_subscribers_per_100_people
-0.594202
service_workers_percent_of_employment
-0.593949
cars_trucks_and_buses_per_1000_persons
-0.588383
colonandrectum_cancer_new_cases_per_100000_women
-0.587266
municipal_water_withdrawal_cu_meters_per_person
-0.587244
hourly_labour_cost_constant_2011_usd
-0.583769
population_aged_40_59_years_female_percent

-0.579127
corruption_perception_index_cpi
-0.578147
breast_cancer_new_cases_per_100000_women
-0.576178
colonandrectum_cancer_new_cases_per_100000_men
-0.569270
corruption_perception_index_cpi_pre2012
-0.561754
medical_doctors_per_1000_people
-0.560037
population_aged_60plus_years_both_sexes_percent
-0.558455
population_aged_60plus_years_male_percent
-0.556233
population_aged_60plus_years_female_percent
-0.555761
alternative_gdppercapita_ppp_inflation_adjusted_from_pwt
-0.553490
sex_ratio_0_14_years
-0.549512
lung_cancer_new_cases_per_100000_men
-0.531706
alternative_gdp_per_capita_ppp_pwt_7_1
-0.529623
lung_cancer_deaths_per_100000_men
-0.523551
alternative_gdp_per_capita_ppp_pwt_8_0
-0.514163
smoking_women_percent_of_women_over_age_15
-0.511783
broadband_subscribers_per_100_people
-0.509712
gnipercapita_constant_2010_us
-0.507289
subsistence_incomes_per_person
-0.507058
colonandrectum_cancer_deaths_per_100000_men
-0.501062
alternative_gdp_per_capita_ppp_wb
-0.496131
gnipercapita_ppp_current_international
-0.495497
income_per_person_with_projections
-0.495413
all_forms_of_tb_detection_rate_percent
-0.494903
male_salaried_workers_percent_of_non_agricultural_male_employment

-0.492849
gdpperworking_hour_us_inflation_adjusted
-0.489458
internet_users
-0.488359
electricity_generation_per_person
-0.485954
government_share_of_total_health_spending_percent
-0.484713
total_health_spending_per_person_international_dollar
-0.483725
smoking_men_percent_of_men_over_age_15
-0.471508
electricity_use_per_person
-0.468646
government_health_spending_per_person_international_dollar
-0.465478
residential_electricity_use_per_person
-0.460529
gdppercapita_us_inflation_adjusted
-0.457950
democracy_score_use_as_color
-0.448862
personal_computers_per_100_people
-0.443531
lung_cancer_new_cases_per_100000_women
-0.442429
all_forms_of_tb_detection_rate_percent_dots_only
-0.441753
income_per_person_gdppercapita_ppp_inflation_adjusted
-0.439191
income_per_person_long_series
-0.436530
pab_immunized_percent_of_newborns
-0.423721
population_in_urban_agglomerations_m_1_million_percent_of_total
-0.421576
prostate_cancer_new_cases_per_100000_men
-0.416867
nuclear_power_generation_per_person
-0.415751
industry_workers_percent_of_employment
-0.415448
gnipercapita_atlasmethode_current_us
-0.409188
colonandrectum_cancer_deaths_per_100000_women
-0.403385
total_health_spending_per_person_us

-0.402446
energy_use_per_person
-0.402390
lung_cancer_deaths_per_100000_women
-0.400447
government_and_society_aid_given_percent_of_aid
-0.397735
population_aged_20_39_years_male_percent
-0.397158
time
-0.393888
aid_given_per_person_2007_us
-0.389990
government_health_spending_per_person_us
-0.389843
cell_phones_per_100_people
-0.389260
energy_supply_per_person_toe
-0.384580
alcohol_consumption_per_adult_15plus_litres
-0.379229
population_aged_20_39_years_both_sexes_percent
-0.377948
services_percent_of_gdp
-0.376490
co2_emissions_tonnes_per_person
-0.374198
oil_consumption_per_cap
-0.361355
privately_owned_other_wooded_land_percent
-0.356089
average_age_of_dollar_billionaires_years
-0.351576
expenditure_per_student_primary_percent_of_gdp_per_person
-0.347750
population_aged_20_39_years_female_percent
-0.347452
industrial_water_withdrawal_percent_of_total
-0.344993
income_share_of_2nd_richest_20percent
-0.340478
income_share_of_middle_20percent
-0.337932
multisector_cross_cutting_aid_given_percent_of_aid
-0.335555
hepb3_immunized_percent_of_one_year_old
-0.314841
government_health_spending_of_total_gov_spending_percent

-0.306402
income_share_of_2nd_poorest_20percent
-0.305692
privately_owned_forest_land_percent
-0.304148
exports_percent_of_gdp
-0.301728
art_coverage_percent_cd4_l_350
-0.301436
coal_consumption_per_cap
-0.284194
female_industry_workers_percent_of_female_employment
-0.277779
infectious_tb_detection_rate_percent
-0.276295
bad_teeth_per_child_12_yr
-0.269684
economic_growth_over_the_past_10_years
-0.266710
hib3_immunized_percent_of_one_year_olds
-0.262884
suicide_age_45_59_per_100000_people
-0.258984
infectious_tb_detection_rate_percent_dots_only
-0.254076
patents_granted_total
-0.252784
health_aid_given_percent_of_aid
-0.251194
tax_revenue_percent_of_gdp
-0.249765
co2_intensity_of_economic_output_kg_co2_per_2011_ppp_of_gdp
-0.248325
aid_given_percent_of_gni
-0.246285
high_technology_exports_percent_of_manufactured_exports
-0.240294
suicide_age_60plus_per_100000_people
-0.239530
suicide_age_30_44_per_100000_people
-0.238325
patent_applications_total
-0.235896
hydro_power_generation_per_person
-0.231281
suicide_per_100000_people
-0.226379
natural_gas_production_total

-0.223630
infectious_tb_treatment_dots_completed_percent
-0.222962
natural_gas_production_per_person
-0.221098
stomach_cancer_new_cases_per_100000_men
-0.221024
suicide_women_per_100000_people
-0.220864
pump_price_for_gasoline_us_per_liter
-0.220007
income_share_of_poorest_20percent
-0.213607
total_health_spending_percent_of_gdp
-0.211769
suicide_men_per_100000_people
-0.206026
patents_in_force_total
-0.204348
breast_cancer_deaths_per_100000_women
-0.203978
hydro_power_generation_total
-0.196162
oil_proved_reserves_total
-0.193310
external_debt_total_us_not_inflation_adjusted
-0.190619
gdppercapita_growth_over_next_10_years
-0.188660
merchandise_trade_percent_of_gdp
-0.187008
oil_proven_reserves_per_person
-0.182956
blood_pressure_sbp_men_mmhg
-0.180146
sex_ratio_15_24_years
-0.179944
suicide_age_15_29_per_100000_people
-0.179040
prostate_cancer_number_of_male_deaths
-0.177238
imports_percent_of_gdp
-0.176439
investments_percent_of_gdp
-0.175372
total_gdp_us_inflation_adjusted
-0.172713
oil_production_total

-0.171844
nuclear_power_generation_total
-0.168930
imports_unit_value_index_2000100
-0.167052
total_sulfur_emission_kilotonnes
-0.165059
trade_balance_percent_of_gdp
-0.164920
colonandrectum_cancer_number_of_new_female_cases
-0.164830
residential_electricity_use_total
-0.161513
market_value_of_listed_companies_percent_of_gdp
-0.159897
present_value_of_debt_percent_of_gni
-0.157912
total_gdp_ppp_inflation_adjusted
-0.157582
income_share_of_poorest_10percent
-0.156067
colonandrectum_cancer_number_of_new_male_cases
-0.154240
industry_percent_of_gdp
-0.154016
total_gni_ppp_current_international
-0.150175
armed_forces_personnel_percent_of_labor_force
-0.148876
colonandrectum_cancer_number_of_female_deaths
-0.148303
aid_received_per_person_current_us
-0.147141
suicide_total_deaths
-0.146889
cumulative_co2_emissions_tonnes
-0.145984
breast_cancer_number_of_new_female_cases
-0.145652
prostate_cancer_number_of_new_male_cases
-0.142448
personal_computers_total
-0.141525
colonandrectum_cancer_number_of_male_deaths
-0.140074
oil_consumption_total
-0.139643
yearly_co2_emissions_1000_tonnes

-0.137972
sulfur_emissions_per_person_kg
-0.131091
water_and_sanitation_aid_given_percent_of_aid
-0.127515
water_withdrawal_cu_meters_per_person
-0.123634
arms_exports_us_inflation_adjusted
-0.122930
males_aged_55_64_unemployment_rate_percent
-0.121879
broadband_subscribers
-0.118921
population_density_per_square_km
-0.118655
sex_ratio_15_49_years
-0.117509
forest_products_removal_total_dollar
-0.115905
planted_forest_area_ha
-0.114858
energy_production_total
-0.114248
energy_production_per_person
-0.109673
lung_cancer_number_of_new_female_cases
-0.109089
lung_cancer_number_of_new_male_cases
-0.108268
lung_cancer_number_of_male_deaths
-0.106514
electricity_generation_total
-0.105536
breast_cancer_number_of_female_deaths
-0.105502
aged_55_64_unemployment_rate_percent
-0.105060
municipal_water_withdrawal_percent_of_total
-0.103920
lung_cancer_number_of_female_deaths
-0.103345
total_number_of_dollar_billionaires
-0.101974
natural_gas_proved_reserves_total
-0.101242
dollar_billionaires_per_million_people
-0.098041
natural_gas_proven_reserves_per_person

-0.097763
armed_forces_personnel_total
-0.096611
traffic_mortality_15_29_per_100000_people
-0.096057
urban_population
-0.095386
male_long_term_unemployment_rate_percent
-0.094665
cell_phones_total
-0.093068
long_term_unemployment_rate_percent
-0.088980
population_aged_60plus_years_total_number
-0.088408
tsunami_deaths_annual_number
-0.087093
arms_imports_us_inflation_adjusted
-0.086081
males_aged_15_24_unemployment_rate_percent
-0.084667
foreign_direct_investment_net_inflows_percent_of_gdp
-0.084225
desalinated_water_produced_billion_cu_meters
-0.081016
oil_production_per_person
-0.080877
gdp_per_capita_yearly_growth
-0.080312
gdppercapita_growth_percent_per_year
-0.077867
stomach_cancer_deaths_per_100000_men
-0.074346
stomach_cancer_number_of_new_female_cases
-0.072068
female_long_term_unemployment_rate_percent
-0.071417
males_aged_15plus_unemployment_rate_percent
-0.070148
stomach_cancer_number_of_new_male_cases
-0.067573
aged_15plus_unemployment_rate_percent
-0.067190
total_reserves_percent_of_debt_to_foreigners
-0.063427
aged_15_24_unemployment_rate_percent
-0.059107
stomach_cancer_number_of_male_deaths

-0.057648
stomach_cancer_number_of_female_deaths
-0.056403
traffic_total_deaths
-0.052773
stomach_cancer_new_cases_per_100000_women
-0.052362
population_aged_40_59_years_total_number
-0.052046
plane_crash_affected_annual_number
-0.051826
plane_crash_deaths_annual_number
-0.051290
coal_consumption_total
-0.048594
aid_given_2007_us
-0.048467
liver_cancer_number_of_new_female_cases
-0.047691
primary_forest_area_ha
-0.046015
tsunami_affected_annual_number
-0.044222
exports_unit_value_index_2000100
-0.044222
foreign_direct_investment_net_outflows_percent_of_gdp
-0.043320
forest_coverage_percent
-0.042991
males_aged_15_64_labour_force_participation_rate_percent
-0.042939
liver_cancer_number_of_female_deaths
-0.042818
males_aged_25_54_unemployment_rate_percent
-0.042727
liver_cancer_number_of_new_male_cases
-0.041267
sex_ratio_all_age_groups
-0.039474
liver_cancer_number_of_male_deaths
-0.038847
forest_area_sq_km
-0.038616
forest_land_total_area_ha
-0.034581
aged_25_54_unemployment_rate_percent
-0.031019
females_aged_15_24_unemployment_rate_percent

-0.027858
males_aged_25_54_labour_force_participation_rate_percent
-0.027648
trade_balance_us_not_inflation_adjusted
-0.027263
wood_removal_cubic_meters
-0.027083
extreme_temperature_deaths_annual_number
-0.026342
population_aged_20_39_years_total_number
-0.025110
surface_area_sq_km
-0.023919
females_aged_25_54_unemployment_rate_percent
-0.022846
female_population_with_projections
-0.018069
male_population_with_projections
-0.014429
aged_25_54_labour_force_participation_rate_percent
-0.013612
females_aged_15plus_unemployment_rate_percent
-0.012102
extreme_temperature_affected_annual_number
-0.010459
storm_affected_annual_number
-0.010364
population_total
-0.010286
blood_pressure_sbp_women_mmhg
-0.009551
internal_renewable_water_cu_meters_per_person
-0.007709
total_population_with_projections
-0.006621
central_bank_discount_rate_annual_percent
-0.006432
earthquake_affected_annual_number
-0.005334
flood_affected_annual_number
-0.005308
females_aged_55_64_unemployment_rate_percent
-0.005197
biomass_stock_in_forest_tons
-0.001116
total_water_withdrawal_billion_cu_meters
0.001073
population_aged_15_19_years_total_number

0.002986
cervical_cancer_number_of_new_female_cases
0.004122
gdp_total_yearly_growth
0.006102
epidemic_affected_annual_number
0.006187
prostate_cancer_deaths_per_100000_men
0.008784
earthquake_deaths_annual_number
0.009812
sex_ratio_above_50_years
0.011715
drought_affected_annual_number
0.012096
renewable_water_cu_meters_per_person
0.015017
suicide_age_0_14_per_100000_people
0.016117
population_aged_10_14_years_total_number
0.016896
females_aged_25_54_labour_force_participation_rate_percent
0.017908
flood_deaths_annual_number
0.020573
military_expenditure_percent_of_gdp
0.029023
population_aged_5_9_years_total_number
0.031765
inflation_annual_percent
0.032616
cervical_cancer_number_of_female_deaths
0.034701
aged_15_64_labour_force_participation_rate_percent
0.036930
storm_deaths_annual_number
0.037195
drought_deaths_annual_number
0.041600
population_aged_0_4_years_total_number
0.049873
forest_products_removal_per_ha_dollar
0.050634
data_quality_income_per_person
0.051865
tb_programme_dots_population_coverage_percent
0.052854
other_deaths_in_newborn_total_deaths

0.062423
murder_total_deaths
0.065878
all_forms_of_tb_number_of_new_cases_reported
0.068037
new_births_total_number_estimated
0.072623
infectious_tb_number_of_new_cases_reported
0.073115
murdered_children_0_14_per_100000_people
0.078500
aid_received_total_us_inflation_adjusted
0.082589
females_aged_15_64_labour_force_participation_rate_percent
0.083600
all_forms_of_tb_number_of_existing_cases_estimated
0.087692
debt_servicing_costs_percent_of_exports_and_net_income_from_abroad
0.089003
traffic_deaths_women_per_100000_people
0.092336
population_policies_aid_given_percent_of_aid
0.096486
debt_to_foreigners_by_public_and_private_percent_of_gni
0.097567
traffic_mortality_children_0_14_per_100000_people
0.101523
infectious_tb_number_of_new_cases_estimated
0.101880
epidemic_deaths_annual_number
0.104354
agricultural_land_percent_of_land_area
0.104948
all_forms_of_tb_number_of_new_cases_estimated
0.106518
net_barter_terms_of_trade_2000_100
0.112716
other_social_services_aid_given_percent_of_aid
0.125979
data_quality_children_per_woman
0.134570
expenditure_per_student_secondary_percent_of_gdp_per_person
0.134892
ncd_deaths_in_children_1_59_months_total_deaths
0.136148
stomach_cancer_deaths_per_100000_women
0.140165
data_quality_life_expectancy

0.140259
all_forms_of_tb_number_of_deaths_estimated
0.143695
number_of_people_in_poverty
0.145529
pneumonia_deaths_in_newborn_total_deaths
0.148367
economical_infrastructure_aid_given_percent_of_aid
0.148378
injury_deaths_in_children_1_59_months_total_deaths
0.157992
traffic_deaths_per_100000_people
0.169761
newborn_deaths
0.170470
diarrhoeal_deaths_in_newborn_total_deaths
0.173948
traffic_mortality_60plus_per_100000_people
0.181495
newly_hiv_infected_number_all_ages
0.183177
people_living_with_hiv_number_all_ages
0.185598
motorcycle_deaths_per_100000_people
0.185632
pertussis_deaths_in_children_1_59_months_total_deaths
0.189762
males_aged_65plus_unemployment_rate_percent
0.190354
newly_hiv_infected_percent_age_15_49
0.191240
traffic_deaths_men_per_100000_people
0.193433
males_aged_15_24_employment_rate_percent
0.196087
all_causes_deaths_in_newborn_total_deaths
0.200243
number_of_child_deaths
0.202981
annual_hiv_deaths_number_all_ages
0.205010
tb_with_hivplus_number_of_new_cases_estimated
0.205689
tb_with_hivplus_number_of_existing_cases_estimated
0.206024
aged_65plus_unemployment_rate_percent
0.207664
males_aged_15plus_employment_rate_percent

0.211790
traffic_mortality_45_59_per_100000_people
0.213564
education_aid_given_percent_of_aid
0.217876
prematurity_deaths_in_newborn_total_deaths
0.221517
tb_with_hivplus_number_of_deaths_estimated
0.230080
aged_15_24_employment_rate_percent
0.230507
traffic_mortality_30_44_per_100000_people
0.233827
working_hours_per_week
0.236232
females_aged_15_24_employment_rate_percent
0.236366
males_aged_15plus_labour_force_participation_rate_percent
0.238025
females_aged_15plus_employment_rate_percent
0.239034
congenital_deaths_in_newborn_total_deaths
0.239652
birth_asphyxia_deaths_in_newborn_total_deaths
0.244483
females_aged_15plus_labour_force_participation_rate_percent
0.246371
aged_15plus_employment_rate_percent
0.253195
all_forms_of_tb_new_cases_per_100000_reported
0.259990
measles_deaths_in_children_1_59_months_total_deaths
0.260858
maternal_deaths_total_number
0.264607
females_aged_65plus_unemployment_rate_percent
0.268795
sepsis_deaths_in_newborn_total_deaths
0.270854
liver_cancer_new_cases_per_100000_men
0.277114
children_out_of_school_primary_female
0.278009
ncd_deaths_in_children_1_59_months_per_1000_births
0.279120
battle_deaths_per_100_000
0.279256
aged_15plus_labour_force_participation_rate_percent

0.280231
aid_received_percent_of_gni
0.283125
urban_poverty_percent_urban_people_below_national_urban
0.284251
malaria_number_of_cases_reported
0.285275
liver_cancer_new_cases_per_100000_women
0.290706
inequality_index_gini
0.291040
hiv_deaths_in_children_1_59_months_total_deaths
0.291436
children_out_of_school_primary
0.293621
adults_with_hiv_percent_age_15_49
0.303902
children_out_of_school_primary_male
0.305835
income_share_of_richest_20percent
0.311287
tb_with_hivplus_new_cases_per_100000_estimated
0.311796
tb_with_hivplus_existing_cases_per_100000_estimated
0.311996
malaria_number_of_deaths_reported
0.315354
income_share_of_richest_10percent
0.320138
murdered_women_per_100000_people
0.322565
malaria_deaths_per_100000_reported
0.327510
tetanus_deaths_in_newborn_total_deaths
0.333650
liver_cancer_deaths_per_100000_men
0.339556
meningitis_deaths_in_children_1_59_months_total_deaths
0.341433
gapminder_gini
0.342453
car_deaths_per_100000_people
0.343807
malaria_deaths_in_children_1_59_months_total_deaths
0.349156
pneumonia_deaths_in_children_1_59_months_total_deaths
0.351006
liver_cancer_deaths_per_100000_women

0.353263
murdered_15_29_per_100000_people
0.354010
tb_with_hivplus_deaths_per_100000_estimated
0.354167
agricultural_water_withdrawal_percent_of_total
0.359437
malaria_cases_per_100000_reported
0.366366
other_deaths_in_newborn_per_1000_births
0.370915
all_causes_deaths_in_children_1_59_months_total_deaths
0.376721
population_growth_annual_percent
0.377955
population_growth_annual_percent_with_projections
0.381243
diarrhoeal_deaths_in_children_1_59_months_total_deaths
0.387082
population_aged_15_19_years_male_percent
0.389015
urban_population_growth_annual_percent
0.407469
murdered_30_44_per_100000_people
0.412309
population_aged_15_19_years_both_sexes_percent
0.417350
other_infections_deaths_in_children_1_59_months_total_deaths
0.420404
data_method_maternal_mortality
0.424561
murder_per_100000_people
0.425192
murdered_men_per_100000_people
0.425942
murdered_45_59_per_100000_people
0.430553
population_aged_15_19_years_female_percent
0.431083
infectious_tb_new_cases_per_100000_reported
0.437078
expenditure_per_student_tertiary_percent_of_gdp_per_person
0.443783
out_of_pocket_share_of_total_health_spending_percent
0.446265
injury_deaths_in_children_1_59_months_per_1000_births
0.449886
hiv_deaths_in_children_1_59_months_per_1000_births

0.459324
rural_poverty_percent_rural_people_below_national_rural
0.473420
private_share_of_total_health_spending_percent
0.485206
measles_deaths_in_children_1_59_months_per_1000_births
0.495028
murdered_60plus_per_100000_people
0.508139
production_sector_aid_given_percent_of_aid
0.537736
population_aged_10_14_years_male_percent
0.548911
drownings_per_100000_people
0.557160
population_aged_10_14_years_female_percent
0.559180
population_aged_10_14_years_both_sexes_percent
0.560518
female_family_workers_percent_of_female_employment
0.562656
male_family_workers_percent_of_male_employment
0.570320
congenital_deaths_in_newborn_per_1000_births
0.582633
all_forms_of_tb_new_cases_per_100000_estimated
0.585424
cervical_cancer_new_cases_per_100000_women
0.586304
family_workers_percent_of_employment
0.589860
alternative_poverty_percent_below_nationally_defined_poverty
0.589877
infectious_tb_new_cases_per_100000_estimated
0.600328
underweight_children
0.600790
all_forms_of_tb_deaths_per_100000_estimated
0.612516
surviving_kids_per_woman
0.615561
malnutrition_weight_for_age_percent_of_children_under_5
0.618406
females_aged_65plus_labour_force_participation_rate_percent
0.628866
falls_deaths_per_100000_people
0.629278
population_aged_5_9_years_female_percent

0.652181
males_aged_65plus_labour_force_participation_rate_percent
0.657067
population_aged_5_9_years_male_percent
0.662152
population_aged_5_9_years_both_sexes_percent
0.663966
aged_65plus_labour_force_participation_rate_percent
0.666317
cervical_cancer_deaths_per_100000_women
0.669661
ifpri_underweight_children
0.670207
tetanus_deaths_in_newborn_per_1000_births
0.670606
poisonings_deaths_per_100000_people
0.672319
residential_energy_use_percent
0.674266
all_forms_of_tb_existing_cases_per_100000_estimated
0.681486
agriculture_workers_percent_of_employment
0.692921
children_and_elderly_per_100_adults
0.705636
agriculture_percent_of_gdp
0.733753
male_self_employed_percent_of_male_employment
0.734867
female_agriculture_workers_percent_of_female_employment
0.735043
burns_deaths_per_100000_people
0.739889
population_aged_0_4_years_female_percent
0.739897
teen_fertility_rate_births_per_1000_women_ages_15_19
0.750962
self_employed_percent_of_employment
0.755169
population_aged_0_4_years_both_sexes_percent
0.757348
male_agriculture_workers_percent_of_male_employment
0.760384
female_self_employed_percent_of_female_employment
0.760608
population_aged_0_4_years_male_percent
0.762051
poverty_percent_people_below_550_a_day

```

0.765574
pertussis_deaths_in_children_1_59_months_per_1000_births
0.775066
energy_from_solid_biofuels_percent
0.785773
malaria_deaths_in_children_1_59_months_per_1000_births
0.785891
crude_death_rate_deaths_per_1000_population
0.807075
pneumonia_deaths_in_newborn_per_1000_births
0.820177
children_per_woman_total_fertility_with_projections
0.825955
children_per_woman_total_fertility
0.828776
poverty_percent_people_below_320_a_day
0.831533
extreme_poverty_percent_people_below_190_a_day
0.831917
other_infections_deaths_in_children_1_59_months_per_1000_birt
0.832290
maternal_deaths_lifetime_risk_per_1000
0.837476
crude_birth_rate_births_per_1000_population
0.838020
diarrhoeal_deaths_in_children_1_59_months_per_1000_births
0.845715
meningitis_deaths_in_children_1_59_months_per_1000_births
0.848874
sepsis_deaths_in_newborn_per_1000_births
0.854768
diarrhoeal_deaths_in_newborn_per_1000_births
0.856222
pneumonia_deaths_in_children_1_59_months_per_1000_births
0.867087
birth_asphyxia_deaths_in_newborn_per_1000_births
0.922118
child_mortality_0_5_year-olds_dying_per_1000_born
1.000000
data_quality_population
NaN
income_mountains
NaN
Name: child_mortality_0_5_year-olds_dying_per_1000_born, dtype: float64
   geo  time  child_mortality_0_5_year-olds_dying_per_1000_born \
171  afg  1970                                306.000000
172  afg  1971                                300.400000
173  afg  1972                                294.700000

```


174	afg	1973	288.900000
175	afg	1974	282.700000
...
79294	world	2016	66.249344
79295	world	2017	66.249344
79296	world	2018	66.249344
79297	world	2019	66.249344
79298	world	2020	66.249344

	primary_school_completion_percent_of_girls	\
171	77.269723	
172	77.269723	
173	77.269723	
174	77.269723	
175	4.401400	
...	...	
79294	77.269723	
79295	77.269723	
79296	77.269723	
79297	77.269723	
79298	77.269723	

	at_least_basic_water_source_overall_access_percent	\
171	84.775057	
172	84.775057	
173	84.775057	
174	84.775057	
175	84.775057	
...	...	
79294	84.775057	
79295	84.775057	
79296	84.775057	
79297	84.775057	
79298	84.775057	

	at_least_basic_sanitation_overall_access_percent	\
171	72.610103	
172	72.610103	
173	72.610103	
174	72.610103	
175	72.610103	
...	...	
79294	72.610103	
79295	72.610103	
79296	72.610103	
79297	72.610103	
79298	72.610103	

	literacy_rate_youth_female_percent_of_females_ages_15_24 \
171	83.959326
172	83.959326
173	83.959326
174	83.959326
175	83.959326
...	...
79294	83.959326
79295	83.959326
79296	83.959326
79297	83.959326
79298	83.959326

	sepsis_deaths_in_newborn_per_1000_births \
171	2.43827
172	2.43827
173	2.43827
174	2.43827
175	2.43827
...	...
79294	2.43827
79295	2.43827
79296	2.43827
79297	2.43827
79298	2.43827

	diarrhoeal_deaths_in_newborn_per_1000_births \
171	0.252356
172	0.252356
173	0.252356
174	0.252356
175	0.252356
...	...
79294	0.252356
79295	0.252356
79296	0.252356
79297	0.252356
79298	0.252356

	pneumonia_deaths_in_children_1_59_months_per_1000_births \
171	6.889712
172	6.889712
173	6.889712
174	6.889712
175	6.889712
...	...
79294	6.889712
79295	6.889712

79296	6.889712
79297	6.889712
79298	6.889712

	birth_asphyxia_deaths_in_newborn_per_1000_births
171	3.945409
172	3.945409
173	3.945409
174	3.945409
175	3.945409
...	...
79294	3.945409
79295	3.945409
79296	3.945409
79297	3.945409
79298	3.945409

[13001 rows x 11 columns]

When primary_school_completion_percent_of_girls is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.2633377090089972 and a variance inflation factor 1.3574741265156105.

When at_least_basic_water_source_overall_access_percent is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.08458870130100526 and a variance inflation factor 1.0924051313559542.

When at_least_basic_sanitation_overall_access_percent is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.08513199031616525 and a variance inflation factor 1.093053849752147.

When literacy_rate_youth_female_percent_of_females_ages_15_24 is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.02830726209936496 and a variance inflation factor 1.0291319066153808.

When sepsis_deaths_in_newborn_per_1000_births is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.004745596988137368 and a variance inflation factor 1.0047682250626333.

When diarrhoeal_deaths_in_newborn_per_1000_births is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.004761761115112728 and a variance inflation factor 1.0047845439705452.

When pneumonia_deaths_in_children_1_59_months_per_1000_births is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.004883368851376524 and a variance inflation factor 1.0049073331693188.

When birth_asphyxia_deaths_in_newborn_per_1000_births is compared to child_mortality_0_5_year_olds_dying_per_1000_born the relationship has an R Square equal to 0.005522908240507052 and a variance inflation factor 1.0055535801541045.

When child_mortality_0_5_year_olds_dying_per_1000_born is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.263337709008997 and a variance inflation factor 1.35747412651561.

When at_least_basic_water_source_overall_access_percent is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.1040383934013065 and a variance inflation factor 1.1161192540339577.

When at_least_basic_sanitation_overall_access_percent is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.09624867383055347 and a variance inflation factor 1.1064990678780011.

When literacy_rate_youth_female_percent_of_females_ages_15_24 is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.03708042270836065 and a variance inflation factor 1.0385083277802443.

When sepsis_deaths_in_newborn_per_1000_births is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.005785962226538377 and a variance inflation factor 1.0058196344114152.

When diarrhoeal_deaths_in_newborn_per_1000_births is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.005048610918810703 and a variance inflation factor 1.005074228725358.

When pneumonia_deaths_in_children_1_59_months_per_1000_births is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.005843810507399461 and a variance inflation factor 1.0058781613685692.

When birth_asphyxia_deaths_in_newborn_per_1000_births is compared to primary_school_completion_percent_of_girls the relationship has an R Square equal to 0.006551909456686268 and a variance inflation factor 1.0065951200863479.

When child_mortality_0_5_year_olds_dying_per_1000_born is compared to at_least_basic_water_source_overall_access_percent the relationship has an R Square equal to 0.08458870130100504 and a variance inflation factor

1.092405131355954.

When `primary_school_completion_percent_of_girls` is compared to `at_least_basic_water_source_overall_access_percent` the relationship has an R Square equal to 0.10403839340130638 and a variance inflation factor 1.1161192540339575.

When `at_least_basic_sanitation_overall_access_percent` is compared to `at_least_basic_water_source_overall_access_percent` the relationship has an R Square equal to 0.737881392515068 and a variance inflation factor 3.8150668111476422.

When `literacy_rate_youth_female_percent_of_females_ages_15_24` is compared to `at_least_basic_water_source_overall_access_percent` the relationship has an R Square equal to 0.025018327916181637 and a variance inflation factor 1.0256603058626839.

When `sepsis_deaths_in_newborn_per_1000_births` is compared to `at_least_basic_water_source_overall_access_percent` the relationship has an R Square equal to 0.03192117740345124 and a variance inflation factor 1.0329737379419512.

When `diarrhoeal_deaths_in_newborn_per_1000_births` is compared to `at_least_basic_water_source_overall_access_percent` the relationship has an R Square equal to 0.028557156471334836 and a variance inflation factor 1.029396640946578.

When `pneumonia_deaths_in_children_1_59_months_per_1000_births` is compared to `at_least_basic_water_source_overall_access_percent` the relationship has an R Square equal to 0.03523018897426211 and a variance inflation factor 1.0365166784570152.

When `birth_asphyxia_deaths_in_newborn_per_1000_births` is compared to `at_least_basic_water_source_overall_access_percent` the relationship has an R Square equal to 0.03716178581482843 and a variance inflation factor 1.0385960852688814.

When `child_mortality_0_5_year_olds_dying_per_1000_born` is compared to `at_least_basic_sanitation_overall_access_percent` the relationship has an R Square equal to 0.08513199031616515 and a variance inflation factor 1.0930538497521467.

When `primary_school_completion_percent_of_girls` is compared to `at_least_basic_sanitation_overall_access_percent` the relationship has an R Square equal to 0.09624867383055347 and a variance inflation factor 1.1064990678780011.

When `at_least_basic_water_source_overall_access_percent` is compared to

at_least_basic_sanitation_overall_access_percent the relationship has an R Square equal to 0.737881392515068 and a variance inflation factor 3.8150668111476422.

When literacy_rate_youth_female_percent_of_females_ages_15_24 is compared to at_least_basic_sanitation_overall_access_percent the relationship has an R Square equal to 0.02971356170620887 and a variance inflation factor 1.0306234948088722.

When sepsis_deaths_in_newborn_per_1000_births is compared to at_least_basic_sanitation_overall_access_percent the relationship has an R Square equal to 0.03535760999474857 and a variance inflation factor 1.0366535934571113.

When diarrhoeal_deaths_in_newborn_per_1000_births is compared to at_least_basic_sanitation_overall_access_percent the relationship has an R Square equal to 0.029471551086662995 and a variance inflation factor 1.0303664989106307.

When pneumonia_deaths_in_children_1_59_months_per_1000_births is compared to at_least_basic_sanitation_overall_access_percent the relationship has an R Square equal to 0.0336167997211384 and a variance inflation factor 1.0347862004548898.

When birth_asphyxia_deaths_in_newborn_per_1000_births is compared to at_least_basic_sanitation_overall_access_percent the relationship has an R Square equal to 0.040469517109196884 and a variance inflation factor 1.0421763746236319.

When child_mortality_0_5_year_olds_dying_per_1000_born is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.02830726209936496 and a variance inflation factor 1.0291319066153808.

When primary_school_completion_percent_of_girls is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.03708042270836054 and a variance inflation factor 1.038508327780244.

When at_least_basic_water_source_overall_access_percent is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.02501832791618175 and a variance inflation factor 1.0256603058626839.

When at_least_basic_sanitation_overall_access_percent is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.02971356170620898 and a variance inflation factor 1.0306234948088722.

When sepsis_deaths_in_newborn_per_1000_births is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.0012524669851509127 and a variance inflation factor 1.0012540376258754.

When diarrhoeal_deaths_in_newborn_per_1000_births is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.0015207908358094715 and a variance inflation factor 1.0015231071632253.

When pneumonia_deaths_in_children_1_59_months_per_1000_births is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.000673760620244046 and a variance inflation factor 1.0006742148796794.

When birth_asphyxia_deaths_in_newborn_per_1000_births is compared to literacy_rate_youth_female_percent_of_females_ages_15_24 the relationship has an R Square equal to 0.001290996058689653 and a variance inflation factor 1.00129266488396.

When child_mortality_0_5_year_olds_dying_per_1000_born is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.004745596988137146 and a variance inflation factor 1.004768225062633.

When primary_school_completion_percent_of_girls is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.005785962226538377 and a variance inflation factor 1.0058196344114152.

When at_least_basic_water_source_overall_access_percent is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.03192117740345146 and a variance inflation factor 1.0329737379419515.

When at_least_basic_sanitation_overall_access_percent is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.03535760999474846 and a variance inflation factor 1.036653593457111.

When literacy_rate_youth_female_percent_of_females_ages_15_24 is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.001252466985150802 and a variance inflation factor 1.0012540376258754.

When diarrhoeal_deaths_in_newborn_per_1000_births is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.8248256302940663 and a variance inflation factor 5.708597677152806.

When pneumonia_deaths_in_children_1_59_months_per_1000_births is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.681017908720277 and a variance inflation factor 3.1349722361782253.

When birth_asphyxia_deaths_in_newborn_per_1000_births is compared to sepsis_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.9334311802256414 and a variance inflation factor 15.022047910562275.

When child_mortality_0_5_year_olds_dying_per_1000_born is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.004761761115112839 and a variance inflation factor 1.0047845439705454.

When primary_school_completion_percent_of_girls is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.005048610918810925 and a variance inflation factor 1.0050742287253582.

When at_least_basic_water_source_overall_access_percent is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.028557156471334836 and a variance inflation factor 1.029396640946578.

When at_least_basic_sanitation_overall_access_percent is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.029471551086663213 and a variance inflation factor 1.030366498910631.

When literacy_rate_youth_female_percent_of_females_ages_15_24 is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.0015207908358092492 and a variance inflation factor 1.0015231071632251.

When sepsis_deaths_in_newborn_per_1000_births is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.8248256302940664 and a variance inflation factor 5.7085976771528095.

When pneumonia_deaths_in_children_1_59_months_per_1000_births is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.6850245403297122 and a variance inflation factor 3.174850513899676.

When birth_asphyxia_deaths_in_newborn_per_1000_births is compared to diarrhoeal_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.8539298251973123 and a variance inflation factor 6.84602453136518.

When child_mortality_0_5_year_olds_dying_per_1000_born is compared to pneumonia_deaths_in_children_1_59_months_per_1000_births the relationship has an R Square equal to 0.004883368851376413 and a variance inflation factor 1.0049073331693188.

When primary_school_completion_percent_of_girls is compared to pneumonia_deaths_in_children_1_59_months_per_1000_births the relationship has an R Square equal to 0.0058438105073995725 and a variance inflation factor

1.0058781613685694.

When `at_least_basic_water_source_overall_access_percent` is compared to `pneumonia_deaths_in_children_1_59_months_per_1000_births` the relationship has an R Square equal to 0.03523018897426211 and a variance inflation factor 1.0365166784570152.

When `at_least_basic_sanitation_overall_access_percent` is compared to `pneumonia_deaths_in_children_1_59_months_per_1000_births` the relationship has an R Square equal to 0.0336167997211384 and a variance inflation factor 1.0347862004548898.

When `literacy_rate_youth_female_percent_of_females_ages_15_24` is compared to `pneumonia_deaths_in_children_1_59_months_per_1000_births` the relationship has an R Square equal to 0.0006737606202443791 and a variance inflation factor 1.00067421487968.

When `sepsis_deaths_in_newborn_per_1000_births` is compared to `pneumonia_deaths_in_children_1_59_months_per_1000_births` the relationship has an R Square equal to 0.681017908720277 and a variance inflation factor 3.1349722361782253.

When `diarrhoeal_deaths_in_newborn_per_1000_births` is compared to `pneumonia_deaths_in_children_1_59_months_per_1000_births` the relationship has an R Square equal to 0.6850245403297122 and a variance inflation factor 3.174850513899676.

When `birth_asphyxia_deaths_in_newborn_per_1000_births` is compared to `pneumonia_deaths_in_children_1_59_months_per_1000_births` the relationship has an R Square equal to 0.7523482403502538 and a variance inflation factor 4.037928102809767.

When `child_mortality_0_5_year_olds_dying_per_1000_born` is compared to `birth_asphyxia_deaths_in_newborn_per_1000_births` the relationship has an R Square equal to 0.005522908240506942 and a variance inflation factor 1.0055535801541045.

When `primary_school_completion_percent_of_girls` is compared to `birth_asphyxia_deaths_in_newborn_per_1000_births` the relationship has an R Square equal to 0.006551909456686045 and a variance inflation factor 1.0065951200863477.

When `at_least_basic_water_source_overall_access_percent` is compared to `birth_asphyxia_deaths_in_newborn_per_1000_births` the relationship has an R Square equal to 0.03716178581482821 and a variance inflation factor 1.0385960852688811.

When `at_least_basic_sanitation_overall_access_percent` is compared to

birth_asphyxia_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.04046951710919666 and a variance inflation factor 1.0421763746236317.

When literacy_rate_youth_female_percent_of_females_ages_15_24 is compared to birth_asphyxia_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.0012909960586893199 and a variance inflation factor 1.0012926648839597.

When sepsis_deaths_in_newborn_per_1000_births is compared to birth_asphyxia_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.9334311802256414 and a variance inflation factor 15.022047910562275.

When diarrhoeal_deaths_in_newborn_per_1000_births is compared to birth_asphyxia_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.8539298251973123 and a variance inflation factor 6.84602453136518.

When pneumonia_deaths_in_children_1_59_months_per_1000_births is compared to birth_asphyxia_deaths_in_newborn_per_1000_births the relationship has an R Square equal to 0.7523482403502538 and a variance inflation factor 4.037928102809767.

```
[62]: # seven variable models
yValues = highCorrelationVariables.
↳child_mortality_0_5_year_olds_dying_per_1000_born.values.reshape(-1,1)
exclusionOne = pd.DataFrame()
for x1 in highCorrelationVariables:
    if(x1 != 'geo' and x1 != 'time' and x1 !=_
↳'child_mortality_0_5_year_olds_dying_per_1000_born'):
        xValues = []
        exclusionOne = highCorrelationVariables.drop(columns = x1)
        for values in exclusionOne:
            if(values != 'geo' and values != 'time' and values !=_
↳'child_mortality_0_5_year_olds_dying_per_1000_born'):
                xValues.append(highCorrelationVariables[values])
        xArray = np.array(xValues, np.float64)
        xArray = xArray.transpose()
        print(xArray.shape)
        estimate = sm.OLS(yValues, xArray)
        estimation = estimate.fit()
        outputString = "\nPredicting_
↳child_mortality_0_5_year_olds_dying_per_1000_born using "
        count = 1
        for column in exclusion1.columns:
```

```

        if(column != 'geo' and column != 'time' and column != '
↳'child_mortality_0_5_year_olds_dying_per_1000_born'):
            if(count < 7):
                outputString += "x" + str(count) + " = " + column + ", "
            else:
                outputString += "x" + str(count) + " = " + column
            count += 1
        outputString += ".\n"
    print(outputString)
    print(estimation.summary())

```

(13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 = primary_school_completion_percent_of_girls, x2 = at_least_basic_water_source_overall_access_percent, x3 = at_least_basic_sanitation_overall_access_percent, x4 = literacy_rate_youth_female_percent_of_females_ages_15_24, x5 = sepsis_deaths_in_newborn_per_1000_births, x6 = diarrhoeal_deaths_in_newborn_per_1000_births, x7 = pneumonia_deaths_in_children_1_59_months_per_1000_births.

OLS Regression Results

```

=====
=====
Dep. Variable:                y      R-squared (uncentered):
0.562
Model:                      OLS      Adj. R-squared (uncentered):
0.562
Method:                     Least Squares      F-statistic:
2382.
Date:                       Sat, 14 Mar 2020      Prob (F-statistic):
0.00
Time:                       20:42:19      Log-Likelihood:
-71659.
No. Observations:           13001      AIC:
1.433e+05
Df Residuals:               12994      BIC:
1.434e+05
Df Model:                   7
Covariance Type:            nonrobust
=====
=====

```

	coef	std err	t	P> t	[0.025	0.975]
x1	0.1099	0.097	1.130	0.259	-0.081	0.301
x2	-0.9981	0.066	-15.076	0.000	-1.128	-0.868
x3	0.5935	0.077	7.683	0.000	0.442	0.745

x4	-27.1506	5.755	-4.718	0.000	-38.431	-15.870
x5	-127.8999	29.404	-4.350	0.000	-185.536	-70.264
x6	-0.8020	0.988	-0.812	0.417	-2.739	1.135
x7	46.3804	4.818	9.627	0.000	36.937	55.824

Omnibus:	4624.779	Durbin-Watson:	0.108
Prob(Omnibus):	0.000	Jarque-Bera (JB):	14928.226
Skew:	1.848	Prob(JB):	0.00
Kurtosis:	6.728	Cond. No.	7.88e+03

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 7.88e+03. This might indicate that there are strong multicollinearity or other numerical problems.
- (13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 = primary_school_completion_percent_of_girls, x2 = at_least_basic_water_source_overall_access_percent, x3 = at_least_basic_sanitation_overall_access_percent, x4 = literacy_rate_youth_female_percent_of_females_ages_15_24, x5 = sepsis_deaths_in_newborn_per_1000_births, x6 = diarrhoeal_deaths_in_newborn_per_1000_births, x7 = pneumonia_deaths_in_children_1_59_months_per_1000_births.

OLS Regression Results

Dep. Variable:	y	R-squared (uncentered):	0.644
Model:	OLS	Adj. R-squared (uncentered):	0.644
Method:	Least Squares	F-statistic:	3357.
Date:	Sat, 14 Mar 2020	Prob (F-statistic):	0.00
Time:	20:42:19	Log-Likelihood:	-70313.
No. Observations:	13001	AIC:	1.406e+05
Df Residuals:	12994	BIC:	1.407e+05
Df Model:	7		
Covariance Type:	nonrobust		

coef	std err	t	P> t	[0.025	0.975]
------	---------	---	------	--------	--------

```

-----
x1          -1.7151      0.031   -54.686      0.000      -1.777      -1.654
x2          -0.4073      0.033   -12.216      0.000      -0.473      -0.342
x3           1.7169      0.064    26.790      0.000       1.591       1.842
x4         -29.8320      5.182    -5.757      0.000     -39.990     -19.674
x5        -143.9762     26.390    -5.456      0.000    -195.704    -92.248
x6          -1.5120      0.889    -1.701      0.089     -3.254       0.230
x7          51.4620      4.292    11.989      0.000     43.048     59.876
=====
Omnibus:                4986.491   Durbin-Watson:                0.273
Prob(Omnibus):          0.000   Jarque-Bera (JB):            20783.439
Skew:                   1.880   Prob(JB):                     0.00
Kurtosis:               7.922   Cond. No.                     7.61e+03
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 7.61e+03. This might indicate that there are strong multicollinearity or other numerical problems.

(13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 =
primary_school_completion_percent_of_girls, x2 =
at_least_basic_water_source_overall_access_percent, x3 =
at_least_basic_sanitation_overall_access_percent, x4 =
literacy_rate_youth_female_percent_of_females_ages_15_24, x5 =
sepsis_deaths_in_newborn_per_1000_births, x6 =
diarrhoeal_deaths_in_newborn_per_1000_births, x7 =
pneumonia_deaths_in_children_1_59_months_per_1000_births.

OLS Regression Results

```

=====
=====
Dep. Variable:                y   R-squared (uncentered):
0.641
Model:                      OLS   Adj. R-squared (uncentered):
0.641
Method:                     Least Squares   F-statistic:
3311.
Date:                       Sat, 14 Mar 2020   Prob (F-statistic):
0.00
Time:                       20:42:19   Log-Likelihood:
-70371.
No. Observations:           13001   AIC:
1.408e+05
Df Residuals:               12994   BIC:
1.408e+05

```

Df Model: 7
Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
x1	-1.7712	0.032	-55.883	0.000	-1.833	-1.709
x2	-0.2780	0.050	-5.612	0.000	-0.375	-0.181
x3	1.6227	0.071	22.942	0.000	1.484	1.761
x4	-30.2649	5.209	-5.810	0.000	-40.475	-20.055
x5	-152.7686	26.542	-5.756	0.000	-204.796	-100.742
x6	-1.4955	0.894	-1.672	0.094	-3.248	0.257
x7	53.8425	4.325	12.449	0.000	45.364	62.321
Omnibus:	4852.987	Durbin-Watson:	0.279			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	19482.245			
Skew:	1.839	Prob(JB):	0.00			
Kurtosis:	7.737	Cond. No.	7.99e+03			

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 7.99e+03. This might indicate that there are strong multicollinearity or other numerical problems.
- (13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 = primary_school_completion_percent_of_girls, x2 = at_least_basic_water_source_overall_access_percent, x3 = at_least_basic_sanitation_overall_access_percent, x4 = literacy_rate_youth_female_percent_of_females_ages_15_24, x5 = sepsis_deaths_in_newborn_per_1000_births, x6 = diarrhoeal_deaths_in_newborn_per_1000_births, x7 = pneumonia_deaths_in_children_1_59_months_per_1000_births.

OLS Regression Results

Dep. Variable:	y	R-squared (uncentered):
0.634		
Model:	OLS	Adj. R-squared (uncentered):
0.634		
Method:	Least Squares	F-statistic:
3219.		
Date:	Sat, 14 Mar 2020	Prob (F-statistic):
0.00		
Time:	20:42:19	Log-Likelihood:
-70488.		

No. Observations: 13001 AIC:
1.410e+05
Df Residuals: 12994 BIC:
1.410e+05
Df Model: 7
Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
x1	-1.6040	0.031	-51.335	0.000	-1.665	-1.543
x2	1.5127	0.080	18.804	0.000	1.355	1.670
x3	-1.0112	0.060	-16.920	0.000	-1.128	-0.894
x4	-46.0706	5.180	-8.894	0.000	-56.224	-35.917
x5	-261.5193	25.988	-10.063	0.000	-312.460	-210.579
x6	-1.3259	0.903	-1.468	0.142	-3.096	0.444
x7	81.5629	4.037	20.202	0.000	73.649	89.477
Omnibus:	4383.900		Durbin-Watson:		0.280	
Prob(Omnibus):	0.000		Jarque-Bera (JB):		15653.442	
Skew:	1.690		Prob(JB):		0.00	
Kurtosis:	7.181		Cond. No.		7.44e+03	

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 7.44e+03. This might indicate that there are strong multicollinearity or other numerical problems.
- (13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 = primary_school_completion_percent_of_girls, x2 = at_least_basic_water_source_overall_access_percent, x3 = at_least_basic_sanitation_overall_access_percent, x4 = literacy_rate_youth_female_percent_of_females_ages_15_24, x5 = sepsis_deaths_in_newborn_per_1000_births, x6 = diarrhoeal_deaths_in_newborn_per_1000_births, x7 = pneumonia_deaths_in_children_1_59_months_per_1000_births.

OLS Regression Results

```
=====
=====
Dep. Variable: y R-squared (uncentered): 0.645
Model: OLS Adj. R-squared (uncentered): 0.645
Method: Least Squares F-statistic: 3372.
```

Date: Sat, 14 Mar 2020 Prob (F-statistic):
 0.00
 Time: 20:42:19 Log-Likelihood:
 -70294.
 No. Observations: 13001 AIC:
 1.406e+05
 Df Residuals: 12994 BIC:
 1.407e+05
 Df Model: 7
 Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
x1	-1.7463	0.032	-55.333	0.000	-1.808	-1.684
x2	0.7380	0.088	8.368	0.000	0.565	0.911
x3	-0.8197	0.060	-13.738	0.000	-0.937	-0.703
x4	1.5286	0.070	21.766	0.000	1.391	1.666
x5	-153.2077	25.881	-5.920	0.000	-203.939	-102.477
x6	-0.4722	0.883	-0.534	0.593	-2.204	1.260
x7	28.1621	2.756	10.217	0.000	22.759	33.565

Omnibus: 4948.760 Durbin-Watson: 0.261
 Prob(Omnibus): 0.000 Jarque-Bera (JB): 20222.538
 Skew: 1.873 Prob(JB): 0.00
 Kurtosis: 7.827 Cond. No. 8.81e+03

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
 [2] The condition number is large, 8.81e+03. This might indicate that there are strong multicollinearity or other numerical problems.
 (13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 =
 primary_school_completion_percent_of_girls, x2 =
 at_least_basic_water_source_overall_access_percent, x3 =
 at_least_basic_sanitation_overall_access_percent, x4 =
 literacy_rate_youth_female_percent_of_females_ages_15_24, x5 =
 sepsis_deaths_in_newborn_per_1000_births, x6 =
 diarrhoeal_deaths_in_newborn_per_1000_births, x7 =
 pneumonia_deaths_in_children_1_59_months_per_1000_births.

OLS Regression Results

Dep. Variable:	y	R-squared (uncentered):
0.645		


```

Model:                                OLS    Adj. R-squared (uncentered):
0.645
Method:                               Least Squares    F-statistic:
3375.
Date:                                 Sat, 14 Mar 2020    Prob (F-statistic):
0.00
Time:                                 20:42:19    Log-Likelihood:
-70291.
No. Observations:                     13001    AIC:
1.406e+05
Df Residuals:                         12994    BIC:
1.406e+05
Df Model:                             7
Covariance Type:                      nonrobust
=====
              coef      std err          t      P>|t|      [0.025      0.975]
-----
x1            -1.7471      0.032     -55.372      0.000      -1.809      -1.685
x2             0.7534      0.088      8.574      0.000       0.581       0.926
x3            -0.8308      0.059     -13.965      0.000      -0.947      -0.714
x4             1.5472      0.069     22.443      0.000       1.412       1.682
x5            -32.7742      5.064      -6.472      0.000     -42.701     -22.848
x6            -1.7266      0.877      -1.969      0.049      -3.445      -0.008
x7            40.2988      4.157      9.694      0.000      32.150      48.447
=====
Omnibus:                     4968.801    Durbin-Watson:                0.262
Prob(Omnibus):                0.000    Jarque-Bera (JB):            20315.964
Skew:                         1.881    Prob(JB):                    0.00
Kurtosis:                     7.832    Cond. No.                    2.17e+03
=====

```

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 2.17e+03. This might indicate that there are strong multicollinearity or other numerical problems.
(13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 = primary_school_completion_percent_of_girls, x2 = at_least_basic_water_source_overall_access_percent, x3 = at_least_basic_sanitation_overall_access_percent, x4 = literacy_rate_youth_female_percent_of_females_ages_15_24, x5 = sepsis_deaths_in_newborn_per_1000_births, x6 = diarrhoeal_deaths_in_newborn_per_1000_births, x7 = pneumonia_deaths_in_children_1_59_months_per_1000_births.

OLS Regression Results

```

=====
=====
Dep. Variable:                y    R-squared (uncentered):
0.646
Model:                        OLS    Adj. R-squared (uncentered):
0.646
Method:                        Least Squares    F-statistic:
3383.
Date:                          Sat, 14 Mar 2020    Prob (F-statistic):
0.00
Time:                          20:42:19    Log-Likelihood:
-70281.
No. Observations:              13001    AIC:
1.406e+05
Df Residuals:                  12994    BIC:
1.406e+05
Df Model:                       7
Covariance Type:                nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
x1	-1.7465	0.032	-55.396	0.000	-1.808	-1.685
x2	0.7207	0.088	8.188	0.000	0.548	0.893
x3	-0.8107	0.060	-13.601	0.000	-0.927	-0.694
x4	1.4640	0.071	20.557	0.000	1.324	1.604
x5	-26.9809	5.140	-5.249	0.000	-37.057	-16.905
x6	-128.5652	26.073	-4.931	0.000	-179.672	-77.458
x7	44.0216	3.961	11.113	0.000	36.257	51.787

```

=====
Omnibus:                      4946.752    Durbin-Watson:                0.260
Prob(Omnibus):                  0.000    Jarque-Bera (JB):              20087.057
Skew:                           1.875    Prob(JB):                      0.00
Kurtosis:                       7.798    Cond. No.                      8.88e+03
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 8.88e+03. This might indicate that there are strong multicollinearity or other numerical problems.

(13001, 7)

Predicting child_mortality_0_5_year_olds_dying_per_1000_born using x1 =
primary_school_completion_percent_of_girls, x2 =
at_least_basic_water_source_overall_access_percent, x3 =
at_least_basic_sanitation_overall_access_percent, x4 =
literacy_rate_youth_female_percent_of_females_ages_15_24, x5 =
sepsis_deaths_in_newborn_per_1000_births, x6 =

```
diarrhoeal_deaths_in_newborn_per_1000_births, x7 =
pneumonia_deaths_in_children_1_59_months_per_1000_births.
```

OLS Regression Results

```
=====
=====
Dep. Variable:          y    R-squared (uncentered):
0.643
Model:                OLS    Adj. R-squared (uncentered):
0.642
Method:              Least Squares    F-statistic:
3338.
Date:                Sat, 14 Mar 2020    Prob (F-statistic):
0.00
Time:                20:42:19    Log-Likelihood:
-70336.
No. Observations:      13001    AIC:
1.407e+05
Df Residuals:          12994    BIC:
1.407e+05
Df Model:              7
Covariance Type:      nonrobust
=====
=====
```

	coef	std err	t	P> t	[0.025	0.975]
x1	-1.7471	0.032	-55.178	0.000	-1.809	-1.685
x2	0.8583	0.088	9.805	0.000	0.687	1.030
x3	-0.8920	0.059	-15.024	0.000	-1.008	-0.776
x4	1.7647	0.066	26.904	0.000	1.636	1.893
x5	14.7963	3.303	4.479	0.000	8.322	21.271
x6	-43.3271	25.460	-1.702	0.089	-93.232	6.577
x7	2.7979	0.816	3.429	0.001	1.199	4.397

```
=====
=====
Omnibus:              5002.834    Durbin-Watson:              0.262
Prob(Omnibus):        0.000    Jarque-Bera (JB):          20763.067
Skew:                 1.889    Prob(JB):                  0.00
Kurtosis:             7.904    Cond. No.                  8.65e+03
=====
=====
```

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 8.65e+03. This might indicate that there are strong multicollinearity or other numerical problems.

```
[67]: yValues = highCorrelationVariables.
      ↪ child_mortality_0_5_year_olds_dying_per_1000_born.values.reshape(-1,1)
```

```

xValues = []
for column in highCorrelationVariables:
    if(column != 'geo' and column != 'time' and column != '
    ↳'child_mortality_0_5_year_olds_dying_per_1000_born'):
        xValues.append(highCorrelationVariables[column])
xValues = np.array(xValues , np.float64)
xValues = xValues.transpose()
estimate = sm.OLS(yValues, xValues)
estimation = estimate.fit()
print(estimation.summary())

```

OLS Regression Results

```

=====
=====
Dep. Variable:                y      R-squared (uncentered):
0.646
Model:                        OLS    Adj. R-squared (uncentered):
0.646
Method:                        Least Squares    F-statistic:
2960.
Date:                          Sat, 14 Mar 2020    Prob (F-statistic):
0.00
Time:                          20:43:30    Log-Likelihood:
-70280.
No. Observations:              13001    AIC:
1.406e+05
Df Residuals:                  12993    BIC:
1.406e+05
Df Model:                      8
Covariance Type:               nonrobust
=====
=====

```

	coef	std err	t	P> t	[0.025	0.975]
x1	-1.7466	0.032	-55.401	0.000	-1.808	-1.685
x2	0.7139	0.088	8.093	0.000	0.541	0.887
x3	-0.8080	0.060	-13.547	0.000	-0.925	-0.691
x4	1.4627	0.071	20.536	0.000	1.323	1.602
x5	-27.6879	5.176	-5.349	0.000	-37.834	-17.542
x6	-123.4209	26.447	-4.667	0.000	-175.260	-71.581
x7	-1.0316	0.889	-1.161	0.246	-2.774	0.710
x8	46.0607	4.333	10.629	0.000	37.567	54.555

```

=====
=====
Omnibus:                      4946.228    Durbin-Watson:                0.260
Prob(Omnibus):                 0.000    Jarque-Bera (JB):              20089.927
Skew:                          1.874    Prob(JB):                      0.00
Kurtosis:                     7.799    Cond. No.                      9.01e+03
=====
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.01e+03. This might indicate that there are strong multicollinearity or other numerical problems.

[]: