Largest predictor of CO2 output per capita of a country

Om een beeld te kunnen krijgen van de grootste voorspellende variabelen van CO2 uitstoot zijn er datasets gezocht voor vijf interessante variabelen. Daarnaast is uiteraard een CO2 dataset gebruikt en een dataset met universele landcodes.

Om te beginnen is de CO2 dataset bekeken en gebruiksklaar gemaakt. Deze dataset is verkregen van ouwroldindata.org[[1]](#footnote-1) Enkele kolommen met te lange namen zijn hernoemd en de data is gefilterd voor landcodes die voorkomen in de landcodes dataframe[[2]](#footnote-2). In de CO2 dataset kwamen namelijk ook continenten en gebieden voor, terwijl in dit onderzoek alleen landen worden gebruikt.

Als eerste variabele is het nationale GDP gebruikt. Deze dataset is verkregen van ourworldindata.org[[3]](#footnote-3) en bevat informatie over het GDP per land en over het jaartal waarin dit GDP is vastgesteld. Om de data gebruiksklaar te maken is er één kolom gedropt. Het dataframe is vervolgens samengenomen met het CO2 dataframe met een ‘inner merge’. Uiteindelijk is de Pearsonr berekend.

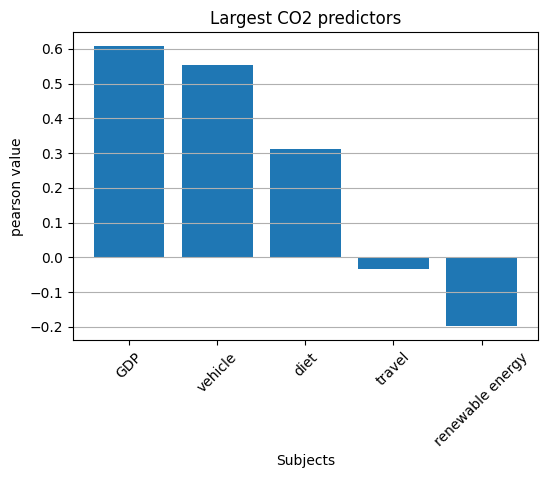
Als tweede variabele is het aantal auto’s per duizend inwoners gebruikt. Deze data is verkregen van ourworldindata.org[[4]](#footnote-4). Eén kolom in deze dataset is hernoemd wegens een te lange naam. Vervolgens is het dataframe samengenomen met het CO2 dataframe met een ‘inner merge’. Uiteindelijk is de Pearsonr berekend.

Als derde variabele is de hoeveelheid gegeten groente per inwoner gebruikt. Deze data is verkregen van ourworldindata.org[[5]](#footnote-5). In dit dataframe stond wederom een onnodig lange kolomnaam en deze is daarom hernoemd. Vervolgens is het dataframe samengenomen met het CO2 dataframe en is de Pearsonr berekend.

Als vierde variabele is de reisafstand naar werk gebruikt. Deze data is verkregen van Nationmaster.com[[6]](#footnote-6). Deze dataset beschikte over data van 77 landen. Omdat er verder geen dataset is gevonden met data van alle landen is uiteindelijk deze dataset gebruikt. De dataset maakte gebruik van landnamen en deze bleken te verschillen van de gebruikte landnamen in de CO2 dataset. Om de twee datasets aan elkaar te koppelen is gebruik gemaakt van een dataset met landcodes. Enkele landcodes zijn handmatig nog gewijzigd. Uiteindelijk is de dataset met de reisafstand samengenomen met de CO2 dataset en in de Pearsonr berekend.

Als laatste variabele is het percentage aan herbruikbare energie gebruikt. Deze data is verkregen van ourworldindata.org[[7]](#footnote-7). In dit dataframe is een lange kolomnaam ingekort en is de data gefilterd op landcodes. Vervolgens is dit dataframe samengenomen met de CO2 dataset en is de Pearsonr berekend.

Na berekenen van alle Pearsonr’s, is voor elke Pearsonr de r-waarde vastgelegd in een variabele. Deze variabelen zijn uiteindelijk in een plot m.b.v. een bardiagram gevisualiseerd per thema. Alle p-waarden zijn erg klein behalve die van diet. Hier is de p-waarde 0.77.



Uit de figuur blijkt dat het GDP de grootste bepaler is van CO2 uitstoot voor een land.

1. https://ourworldindata.org/grapher/co-emissions-per-capita [↑](#footnote-ref-1)
2. https://www.iban.com/country-codes [↑](#footnote-ref-2)
3. https://ourworldindata.org/grapher/gdp-per-capita-world-bank-data-vs-penn-world-table?tab=table&time=1990 [↑](#footnote-ref-3)
4. https://ourworldindata.org/grapher/registered-vehicles-per-1000-people [↑](#footnote-ref-4)
5. https://ourworldindata.org/grapher/vegetable-consumption-per-capita [↑](#footnote-ref-5)
6. https://www.nationmaster.com/country-info/stats/Transport/Commute/Distance [↑](#footnote-ref-6)
7. https://ourworldindata.org/grapher/share-electricity-renewables?tab=table [↑](#footnote-ref-7)