# Finns det ett samband mellan drivmedelspriser och drivkälla hos nyregistrerade bilar?

## Bakgrund:

Tanken med projektet är att undersöka i vilken grad bränsle- och elpriser påverkar vilken drivkälla vi faktiskt väljer för våra fordon. I en perfekt värld hade elbilarna varit etablerade på marknaden länge nog för att göra det möjligt att analysera huruvida det är plånboken eller miljön som styr vårt val. I skrivande stund gissar jag att underlaget är för klent för att dra några slutsatser gällande just den aspekten men jag tycker ändå att det är intressant att göra övningen för att se vad som visar sig, om inte annat rörande övriga drivmedel.

## Datakällor:

Nyregistrerade bilar efter län, kommun och drivmedel (2006M01 - 2023M03):

<https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__TK__TK1001__TK1001A/PersBilarDrivMedel>

Pris bensin, diesel och etanol (1980M12 - 2023M03):

<https://drivkraftsverige.se/fakta-statistik/priser>

Elpriser (2013M03 - 2023M02):

<https://www.scb.se/hitta-statistik/statistik-efter-amne/energi/prisutvecklingen-inom-energiomradet/elpriser-och-elavtal/#_Tabellerochdiagram>

## 

## Bearbetning av dataseten:

Samtliga datakällor är möjliga att få antingen i Excel-format eller som i fallet ”nyregistrerade bilar” direkt som csv. Jag har valt att göra grovstädningen i Excel för att efter det konvertera till csv. Datat är tämligen komplett men då det visualiseras på aningen olika sätt i respektive underlag har viss enklare bearbetning varit nödvändig för att till exempel få tidsperioderna på samma axel och göra det lättare att slå samman respektive dataset till ett. Elpriserna redovisas per typkund och elområde – här har jag (åtminstone tills vidare) valt att beräkna snittpriset för dessa per månad.

## Fält och datatyper:

Intressanta fält är priser för respektive drivmedel (datatyp float) samt antal sålda fordon (datatyp int).

## Angreppssätt och frågeställningar:

Då vi försöker förutsäga ett värde baserat på ett antal kända parametrar handlar det om ett regressionsproblem och ”supervised learning”. Då det totala antalet sålda fordon varierar över tid vore ett sannolikt bättre alternativ till att förutsäga en exakt försäljningssiffra, att jobba med marknadsandelar – d.v.s. hur stor andel nyregistrerade bilar drivs av bensin, el och så vidare. Landar jag i det får jag givetvis anpassa underlagen på lämpligt sätt. Hänsyn måste dessutom tas till den varierande ledtiden från beställd bil till leverans och registrering – är jag intresserad av hur registreringssiffrorna ser ut innevarande månad eller fyra månader efter ett givet prisläge?