

Data Story Group Project Proposal

Team information

All student names and student numbers of your team:

Obbe Reinders

Group number of your team that is registered on Canvas:

A7

Project information

Temporary title for your project:

De invloed van elektrische voertuigen op de luchtkwaliteit van landen.

Topic description in your own words:

De toename van elektrische voertuigen heeft potentieel grote gevolgen voor de luchtkwaliteit in landen. Dit project onderzoekt hoe het gebruik van elektrische voertuigen zich verspreidt in verschillende steden en welke invloed dit heeft op de luchtkwaliteit. We vergelijken landen met hoog en laag gebruik van elektrische voertuigen om de impact te analyseren.

The first perspective of your topic with arguments:

Landen waar er meer gebruik wordt gemaakt van elektrische voertuigen zullen een betere luchtkwaliteit hebben dan steden die minder gebruik maken van elektrische voertuigen. Dit wordt ondersteund door data dat laat zien dat elektrische voertuigen minder CO2 uitstoten dan niet elektrische voertuigen.

The second perspective of your topic with arguments (suggestion: 50-100 words):

De adoptie van elektrische voertuigen heeft nog steeds een beperkte invloed op de luchtkwaliteit vanwege andere vervuilingsbronnen en het huidige marktaandeel van elektrische voertuigen. Dit perspectief kijkt naar data die laat zien dat ondanks de toename van elektrische voertuigen, de luchtkwaliteit niet is verbeterd in sommige gebieden.

Dataset information

Find at least 2 datasets that you may use in the project and describe these datasets by following the instructions below. You also need to explain how the variables in the datasets can help you form perspectives/arguments, as well as providing a rough plan of visualizing the variables. If you have more than 2 datasets, duplicate the form to create another "Dataset 3" section and fill out the form, and so on.

The purpose of asking you to do this is because we want students to explore possible dataset options. If you cannot find at least 2 relevant datasets for your project, you should seriously consider changing the topic, perspectives, or arguments.

Dataset 1

Name of the dataset that you may use:

Historic sales of electric vehicles

Dataset source URL(s):

<https://www.kaggle.com/datasets/edsonmarin/historic-sales-of-electric-vehicles>

Number of records (i.e., data points) in the dataset:

834 records

Number of variables in the dataset:

8 variables

Dataset description in your own words:

Deze dataset bevat historische verkoopgegevens van elektrische voertuigen en informatie over merk, model, type voertuig, verkoopcijfers, regio, land, cumulatieve verkoop, elektrisch bereik en prijs. Deze gegevens helpen bij het analyseren van het aantal elektrische voertuigen per land.

Screenshot of the result of running `pandas.DataFrame.head(n=5)` on your dataset:

region	category	parameter	mode	powertrain	year	unit	value
Australia	Historical	EV sales	Cars	BEV	2011	Vehicles	49
Australia	Historical	EV sales	Cars	BEV	2012	Vehicles	170
Australia	Historical	EV sales	Cars	PHEV	2012	Vehicles	80
Australia	Historical	EV sales	Cars	PHEV	2013	Vehicles	100
Australia	Historical	EV sales	Cars	BEV	2013	Vehicles	190

Choose at least one variable in the dataset that your team wants to explore, explain what the variables mean, and explain how the variable(s) can help you form perspectives/arguments.

We zullen de variabele "value" en "year" gebruiken om de groei van elektrische voertuigen in verschillende regio's te analyseren. Deze variabelen helpen ons te begrijpen hoe snel elektrische voertuigen worden opgenomen in een land en hoe de verkoop over de jaren is geëvolueerd.

Report at least one relevant descriptive statistic for each variable you choose.

We zullen beide variabele gebruiken voor een relevante statistic, namelijk het gemiddelde verkochte aantal auto's per jaar. Het gemiddelde verkochte aantal auto's per jaar is: 5.223.042.

Describe at least one visualization component that you plan to use based on the variable(s) that you choose.

We willen een lijnplot gebruiken om de groei van elektrische voertuigen verkopen door de jaren heen te visualiseren. Deze lijnplot zal ons in staat stellen om de jaarlijkse toename in elektrische voertuigen verkopen te traceren, waardoor we trends in populariteit en marktaandeel kunnen identificeren. Door deze trends in kaart te brengen, kunnen we beter begrijpen hoe de markt voor elektrische voertuigen zich ontwikkelt en welke factoren de verkoop stimuleren.

Describe how your visualization component(s) can support the reason why you want to choose the variables.

De lijnplot maakt het visueel eenvoudig om de trends in elektrische voertuigen verkopen door de jaren heen te zien. Door de gegevens op deze manier te presenteren, kunnen we duidelijk de groeiende populariteit van elektrische voertuigen in verschillende regio's volgen.

Dataset 2

Name of the dataset that you may use:

CO2 Emissions

Dataset source URL(s):

<https://www.kaggle.com/datasets/ulrikthgepedersen/co2-emissions-by-country>

Number of records (i.e., data points) in the dataset:

13.953 records

Number of variables in the dataset:

4 variabels

Dataset description in your own words:

Deze dataset bevat gegevens over de CO2-uitstoot per land per jaar, uitgedrukt in kiloton. De dataset biedt een uitgebreide weergave van de jaarlijkse CO2 emissies per land, wat nuttig is voor het analyseren van trends in uitstoot van broeikasgassen.

Screenshot of the result of running `pandas.DataFrame.head(n=5)` on your dataset:

country_code	country_name	year	value
ABW	Aruba	1960	11092.675
ABW	Aruba	1961	11576.719
ABW	Aruba	1962	12713.489
ABW	Aruba	1963	12178.107
ABW	Aruba	1964	11840.743

Choose at least one variable in the dataset that your team wants to explore, explain what the variables mean, and explain how the variable(s) can help you form perspectives/arguments.

Hier zullen we weer kiezen voor de variabelen "year" en "value" om de veranderingen in CO2 uitstoot door de jaren heen te analyseren. Deze variabelen helpen ons te begrijpen hoe de CO2 uitstoot zich ontwikkelt in verschillende landen en hoe dit mogelijk gerelateerd is aan het gebruik van elektrische voertuigen of juist helemaal niet.

Report at least one relevant descriptive statistic for each variable you choose.

De gemiddelde CO2 uitstoot per jaar per land, omdat we niet van elk land het gemiddelde hier kunnen geven, geven we nu als voorbeeld het gemiddelde van Nederland. Het gemiddelde CO2 uitstoot per jaar in Nederland is ongeveer 148.624 kiloton.

Describe at least one visualization component that you plan to use based on the variable(s) that you choose.

We plannen het gebruik van een heatmap om de CO2 uitstoot per land te visualiseren, wat helpt om verschillen en trends in uitstoot te identificeren. De heatmap maakt het mogelijk om eenvoudig gebieden met hoge en lage uitstoot te vergelijken, en biedt inzicht in de effectiviteit van elektrische voertuigen.

Describe how your visualization component(s) can support the reason why you want to choose the variables.

De heatmap maakt het mogelijk om snel visuele patronen te identificeren in de CO2 uitstoot data, waardoor het gemakkelijker wordt om de impact van elektrische voertuigen gebruik op de luchtkwaliteit te analyseren.