

Data Gedreven Werken



Uitwerkingen presentatie

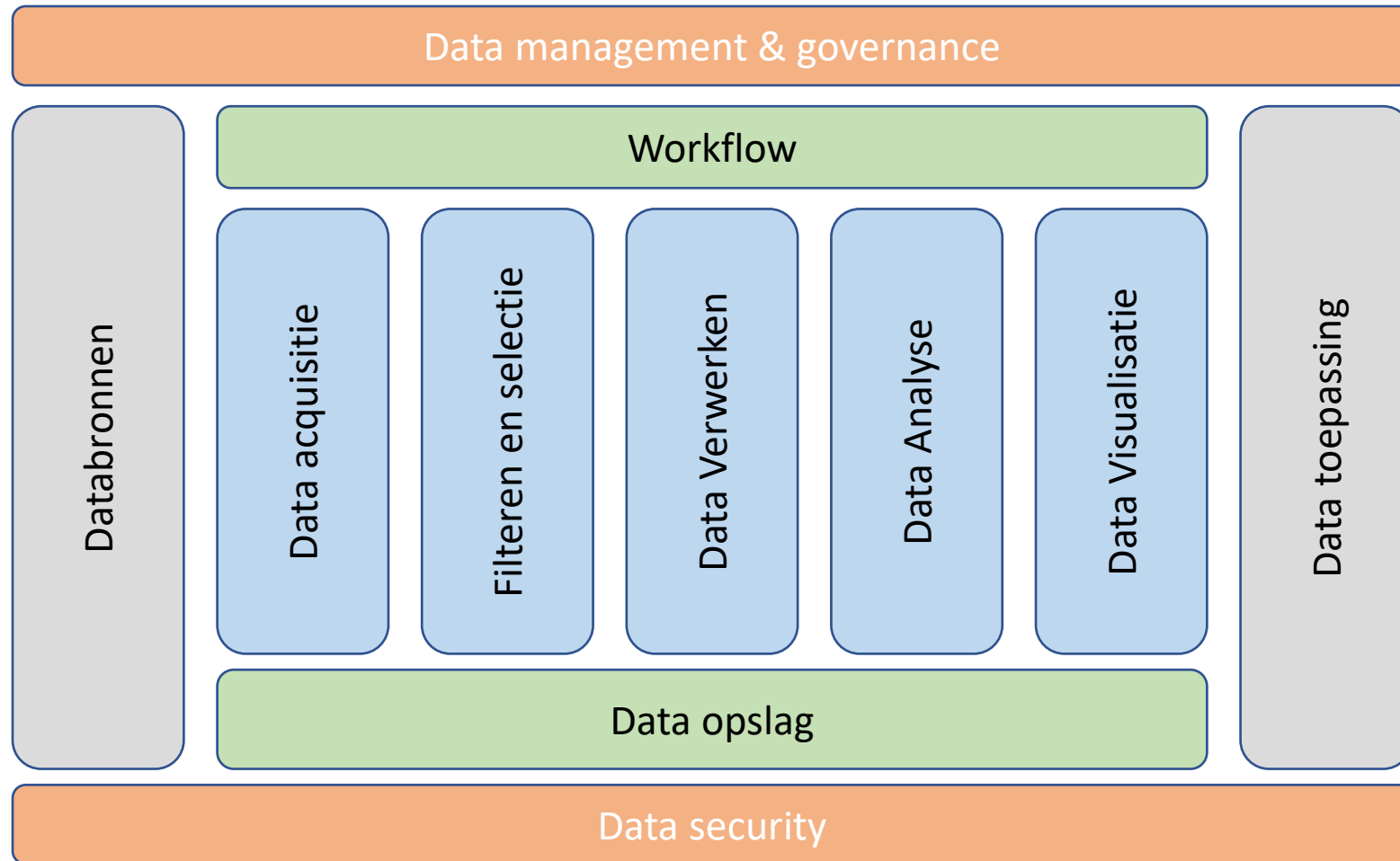
1 Opgave data aspecten

In een brandstof app worden de volgende gegevens aan de chauffeur getoond. Kan hij/zij op basis daarvan een beeld krijgen van de rijstijl. Benoem knelpunten voor de vier aspecten

Ritdatum	Status	Brandstofverbruik
10-5	1	20
21-07-2018	5	24.30
	3	30.573
17-16	3	56

Op welke van deze aspecten kunnen knelpunten optreden, geef aan waarom?

Raamwerk voor data gedreven oplossingen



2 Opgave brandstof verbruik auto App

We gaan een case uitwerken in een aantal opgaven rond data gedreven werk. Namelijk

- **Een landelijke app ontwikkelen voor het analyseren van brandstofverbruik van een auto**
- **Rekening houdend met stads- en snelwegverkeer**
- **Data moet zo veel mogelijk geautomatiseerd verzameld worden**

Welke risico's en kansen komen bij je op als je naar het raamwerk en deze app kijkt?

Template case

- Alleen landelijke app ontwikkelen of Europees
- Wat is stads- en snelwegverkeer
- Data moet zo veel mogelijk geautomatiseerd verzameld worden, dus inzet van sensoren
- Alleen brandstof of ook elektrisch

Kansen

- Nieuwe dienstverlening aanbieden aan bestuurders
- Bijdrage leveren aan milieu
- Interessante dataverwerking voor andere stakeholders

Risico's

- Kwaliteit van de data die verzameld wordt
- Security en privacy gerelateerde zaken
- Databron eigenaren zien dit als een bedreiging
- Beperkte marktpotentie

3 Opgave Project Rollen

- Maak een lijst van rollen die je verwacht binnen een data gedreven project
- Splits de rollen op in noodzakelijk en optioneel
- Wat verandert er bij een Cloud oplossing qua rollen?

Rollen uitwerking

- Data analyst/Scientist
- Data Engineer/Ontwikkelaar
- Project manager/Product owner
- Technisch architect/Data architect
- Gebruikers
- Data steward/Eigenaar
- Ambassadeur/Visionair
- Security/Privacy officer
- ICT beheer/ondersteuning
- Domein expertise
- Tester

Cloud

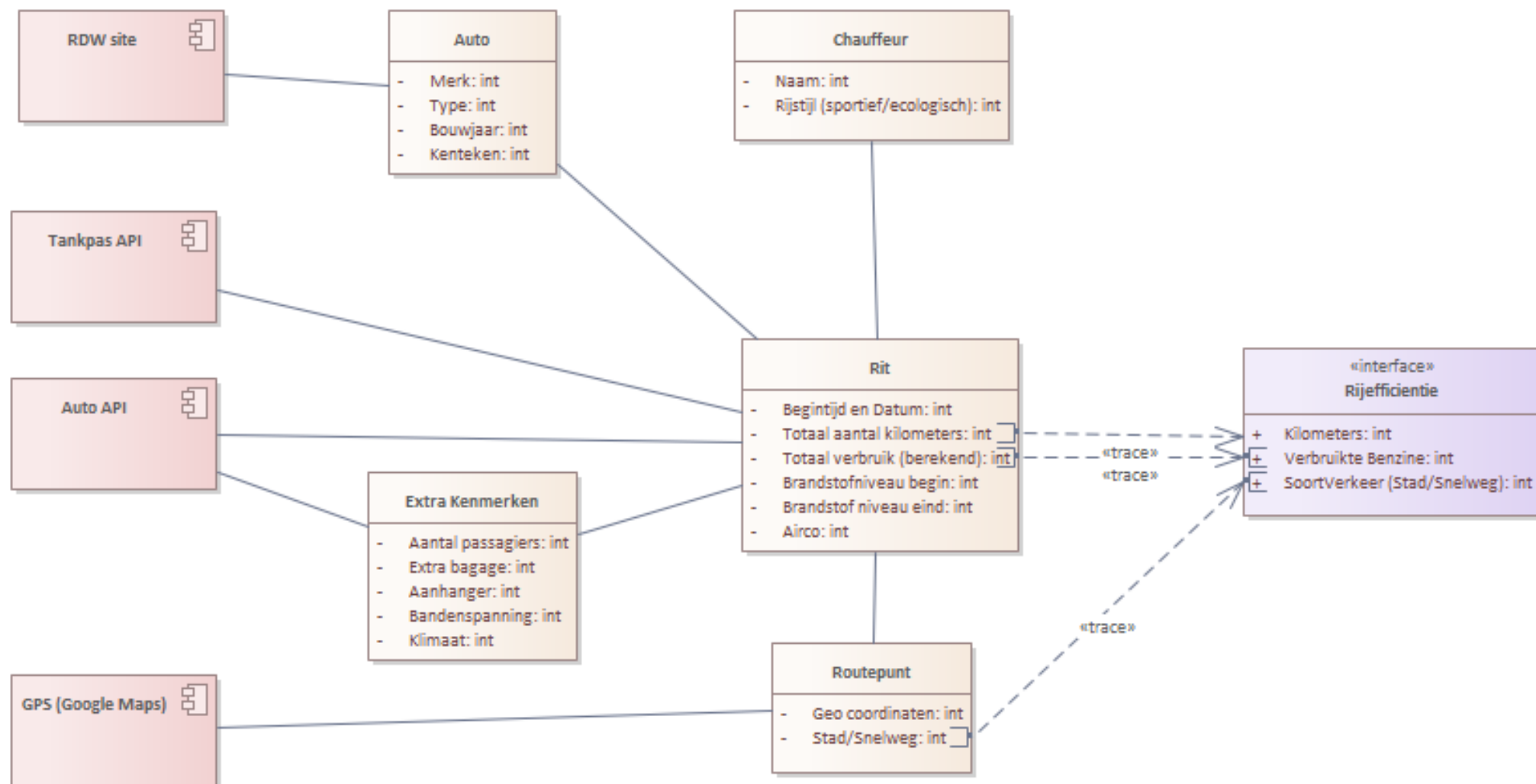
- *ICT Inkoop/Contract*
- *Cloud Architect*

4 Opgave datamodel

- Neem de eerste inventarisatie als startpunt
- Maak een eenvoudig conceptueel datamodel met daarin een definitie van stadsverkeer en snelwegverkeer
- Maak op basis daarvan logisch datamodel van een paar entiteiten en benoem een aantal attributen
- Benoem ook het datatype, kies uit:
 - Tekst
 - Getal
 - Bedrag
 - Datum
 - Datum en tijd

Template Conceptueel datamodel

Dataentiteit	Definitie
Auto	Een auto is een voertuig met wielen dat wordt aangedreven door een motor, meestal een verbrandingsmotor, en is ontworpen voor het vervoeren van passagiers en bagage over de weg. Auto's hebben meestal vier wielen en bieden plaats aan een bestuurder en meerdere passagiers. Ze worden gebruikt voor persoonlijk vervoer en zijn een belangrijk onderdeel van het dagelijks leven voor veel mensen over de hele wereld.
Stadsverkeer	In stedelijke gebieden wordt het brandstofverbruik vaak verhoogd door frequente stops en starts, verkeerslichten, en filevorming. Langzaam rijdend verkeer en stationair draaien leiden tot inefficiënt brandstofgebruik. Voertuigen verbruiken meer brandstof tijdens het optrekken en afremmen dan wanneer ze constant rijden.
Snelwegverkeer	Op snelwegen is het brandstofverbruik meestal lager per kilometer dan in stedelijke gebieden. Dit komt doordat voertuigen met een constante, hogere snelheid rijden, wat leidt tot een efficiënter gebruik van brandstof.



Template Logisch datamodel

Dataentiteit	Attribuut	Datatype
Auto	Model	Tekst
	Merk	Tekst
	Cilinderinhoud	Getal
	Bouwjaar	Getal
Rit	Begintijd	Datumtijd
	Eindtijd	Datumtijd
	Totaalverbruik	Getal met vier cijfers achter de komma
Chauffeur	Naam	Tekst
	Rijstijl	Tekst (categorie)

5 Opgave databehoefte en databronnen

- Maak een eerste lijst van databronnen voor een data gedreven toepassing (brandstof app)
- Benoem ook een aantal datamodel entiteiten
- Maak gebruik van het invulmodel op de volgende slide
- In het voorbeeld is de eerste regel een voorbeeld

Template bron inventarisatie

Dataentiteit	Bron	Bronsoort	Structuur
Autokenmerken	RDW register	Overheid	Gestructureerd
Sensoren in auto	OBD device	Zelf	Semi gestuctureerd
Klimaat	KNMI	Overheid	Gestructureerd
Positie	GPS	Zelf of GPS aanbieder	Gestructureerd
Rit	Bestuurder	Zelf	Gestructureerd

6 Opgave databronnen en APIs

- Neem de eerste inventarisatie als startpunt
- Voeg daar een inventarisatie van de API aan toe
- Voor het soort API kun je kiezen voor
 - Databasetoegang
 - JSON/XML API
 - Bestandsuitwisseling
 - Webpagina
- Maak gebruik van het invulmodel op de volgende slide
- In het voorbeeld is de eerste regel een voorbeeld

Template bron en API inventarisatie

Dataentiteit	Bron	API	API soort
Auto	RDW register	https://www.rdw.nl/particulier/voertuigen/auto/gegevens-bekijken-of-aanpassen	Webpagina
Sensoren in auto	OBD device	https://carapi.app/features/obd-code-api	Json
Klimaat	KNMI	https://www.knmi.nl/nederland-nu/klimatologie/daggegevens	Json/bestand
Positie	GPS	https://developers.google.com/maps/documentation/geolocation/overview	Json
Rit	Bestuurder	Webportaal	Database

7 Opgave databases

- Op de cursussite staat een eenvoudig model voor auto's
- Je kunt hierbij eenvoudig queries maken gebruik makend van de concepten die behandeld zijn

8 Opgave datagovernance

- Neem de eerste inventarisatie als startpunt
- Voeg daar een inventarisatie van data governance rollen aan toe
- Maak onderscheid tussen data eigenaar en data steward
- Neem je inventarisatie van de projectrollen als startpunt
- Maak gebruik van het invulmodel op de volgende slide
- In het voorbeeld is de eerste regel een voorbeeld

Template data governance

Dataentiteit	Bron	Data eigenaar	Data steward
Auto	RDW register	Productowner app en platform	Applicatiebeheerder dataplatform
Sensoren in auto	OBD device	Eigenaar auto	Applicatiebeheerder dataplatform
Klimaat	KNMI	KNMI	Contractmanager
Positie	GPS	Eigenaar auto	Applicatiebeheerder dataplatform
Rit	Bestuurder	Eigenaar auto	Applicatiebeheerder dataplatform

9 Opgave datakwaliteit

Bij de brandstof app zijn een aantal datakwaliteiten van belang. Je kunt dit doen door voor de hele toepassing datakwaliteiten te benoemen.

Dit kan ook per dataentiteit. Hier houden we het eenvoudig. Geef de drie belangrijkste datakwaliteiten (score 1,2 en 3).

Kwaliteit	Score 1,2,3
ACCURAATHEID	1
COMPLEETHEID	
CONSISTENTIE	
ACTUALITEIT	
PRECISIE	2
PRIVACY	3
REDELIJKHEID	
REFERENTIELE INTEGRITEIT	
TIJDIGHEID	3
UNIEKHEID	
VALIDITEIT	

10 Opgave data security

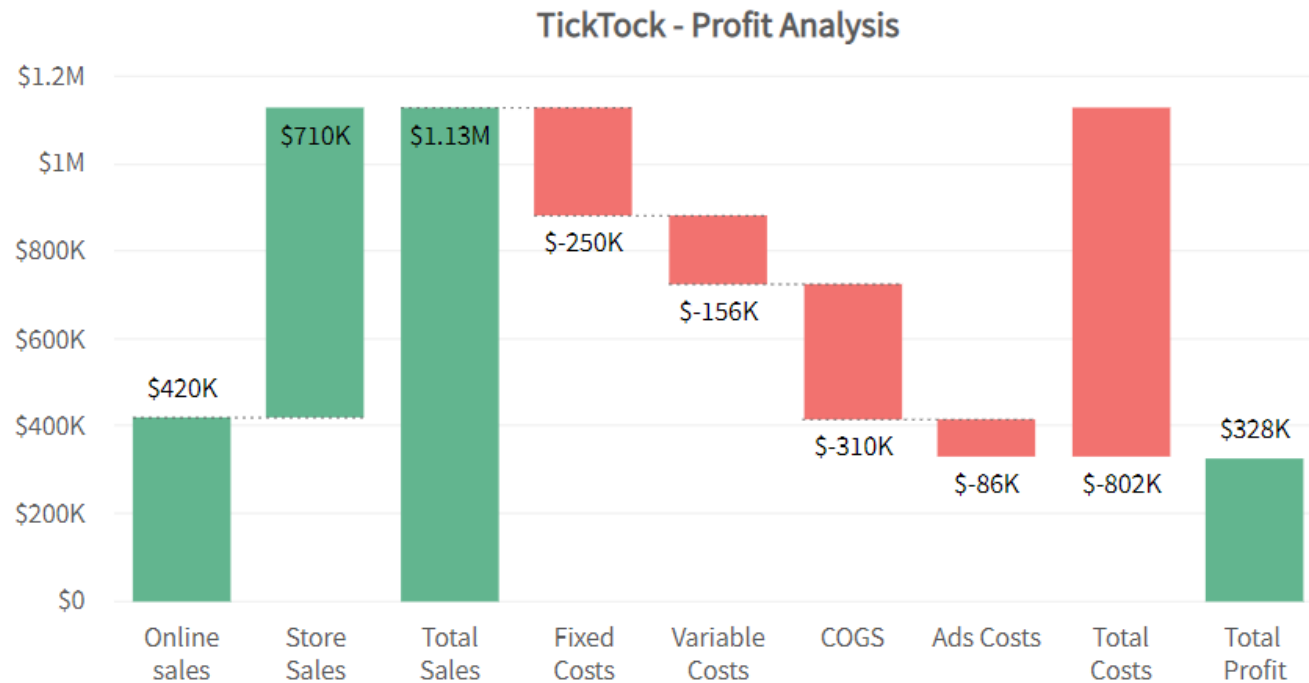
- Geef aan voor een aantal entiteiten in de BIVP classificatie hoog, middel of laag voor onze brandstof app toepassing
- Hoog betekent dat er veel maatregelen genomen moeten worden om dit kenmerk te borgen, laag is dat er weinig risico's zijn voor deze BIVP kenmerken.
- De PDF zit ook in het cursusmateriaal
- Gebruik het sjabloon op de volgende slide

<https://jarnobaselier.nl/biva-data-classificatie/>

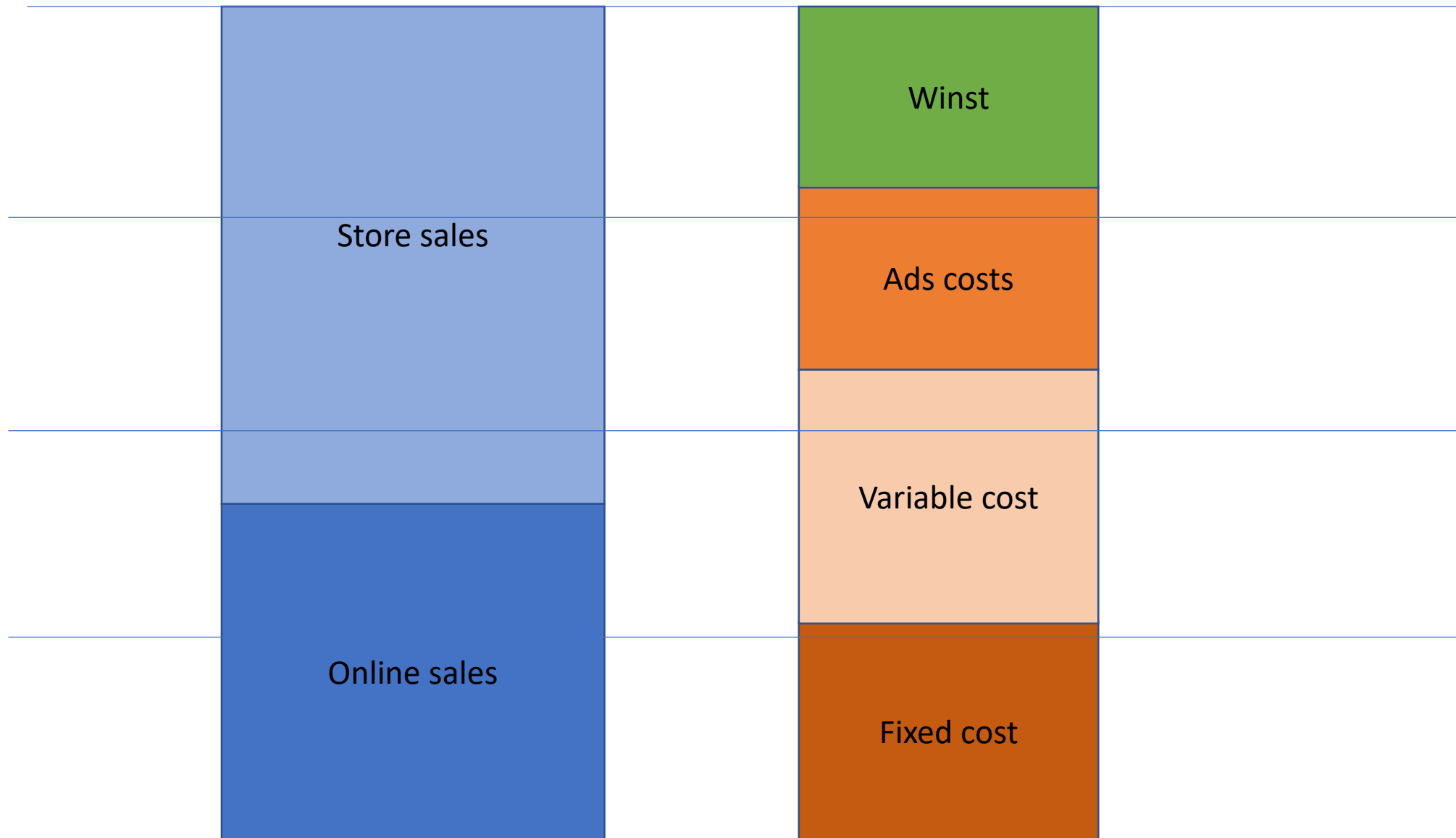
Uitwerking security

Tabel/Object	Beschikbaarheid	Integriteit	Vertrouwelijkheid	Privacy
Auto	Hoog	Hoog	Laag	Laag
Rit	Hoog	Hoog	Hoog	Middel
Chauffeur	Hoog	Hoog	Hoog	Hoog

11 Extra Opgave Data Visualisatie



- Bediscussieer in tweetallen wat hier wordt gevisualiseerd
- Kun je een alternatief verzinnen
- Wat denk je over de doelgroep?



12 Opgave data analyse en visualisatie

Je wilt graag verzamelde autodata verzamelen op basis van het type auto en bouwjaar. Je wilt daarbij graag kijken of er een verband is tussen deze dimensies

- **Welk soort analyse vorm is dit (beschrijvend, diagnostisch, voorspellend of voorschrijvend)?**

Diagnostisch

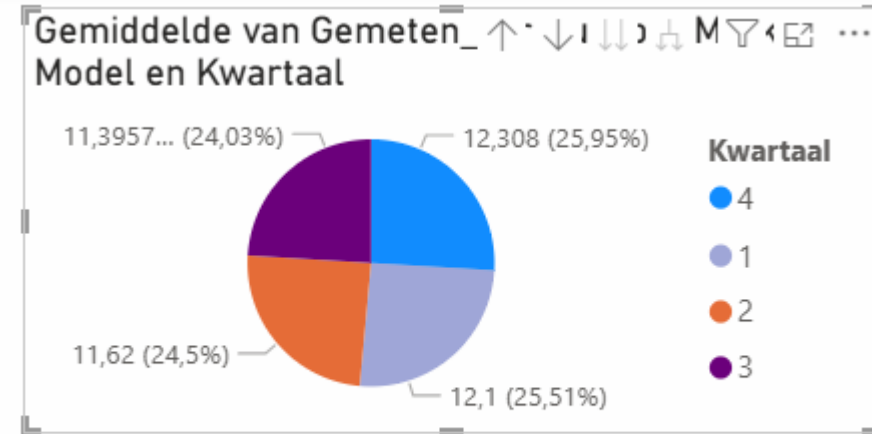
- **Kies een algoritme of een combinatie daarvan**

Drill down en Rollup

- **Wat is een goede visualisatie volgens jou**

Drill down visualisatie

Merk	Peugeot	Totaal	
Jaar	5008	Totaal	
2022	11,72	11,72	11,72
Qtr 2	11,62	11,62	11,62
Qtr 3	11,40	11,40	11,40
Qtr 4	12,31	12,31	12,31
oktober	12,02	12,02	12,02
november	12,56	12,56	12,56
1	12,56	12,56	12,56
december	12,39	12,39	12,39
Totaal	11,72	11,72	11,72



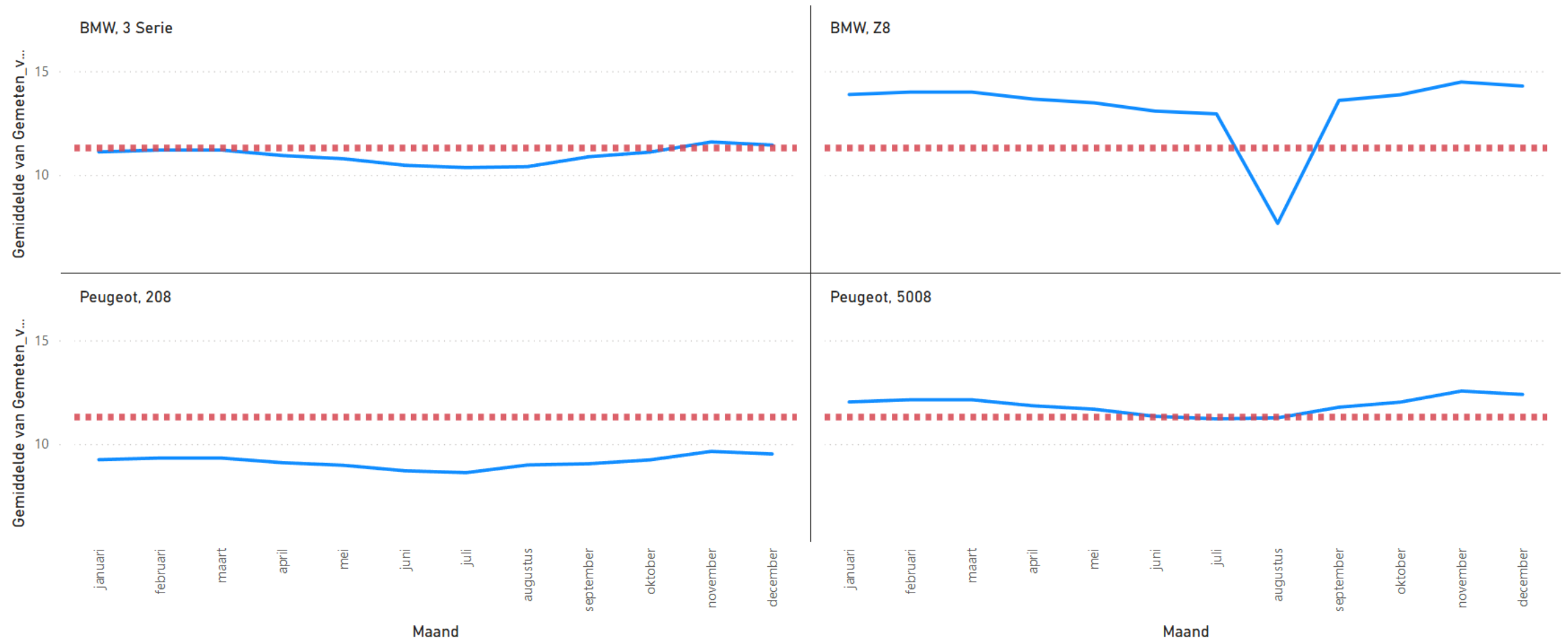
13 Extra opgave data analyse en visualisatie

- **Je wilt voorspellen wanneer het brandstofverbruik zo hoog wordt dat vervanging van de auto relevant wordt.**
- **Welk soort analyse vorm is dit (beschrijvend, diagnostisch, voorspellend of voorschrijvend)?**
- **Kies een algoritme of een combinatie daarvan**
- *Je wilt het brandstofverbruik van je auto visualiseren inclusief hoog en laag*
- **Wat is een goede visualisatie volgens jou**
- **Maak een schets**

14 Uitwerking data analyse en visualisatie

- Welk soort analyse vorm is dit (beschrijvend, diagnostisch, voorspellend of voorschrijvend)?
- **Voorspellend gebaseerd op supervised**
- Kies een algoritme of een combinatie daarvan
- **Classificatie in combinatie met timeseries**
- *Je wilt het brandstofverbruik van je auto visualiseren inclusief hoog en laag*
- *Grafiek met het brandstofverbruik van de auto in combinatie met een lijn die als bovengrens dient*
- Wat is een goede visualisatie volgens jou
- **Doelgroep is eindgebruiker zonder domeinkennis dus eenvoudig in gebruik**

Visualisatie



15 Opgave IoT

- Je wilt het brandstofverbruik van je auto analyseren
- Probeer een aantal sensoren en actuatoren te benoemen
- Zie bijvoorbeeld: <https://www.nvsemi.com/blog/how-many-sensors-are-in-a-car>

Sensor	Actuator

Uitwerking sensoren actuatoren

- Niveaumeter brandstof
- GPS sensor
- Snelheidssensor
- Bandenspanning (druk)
- Flow/gas sensor
- Aanhangsensor
- Gewichtsensensor in de stoelen
- Klimaat sensor
- Motor temp sensor

- Aan/uit actuator rijstijl actuator
- Geluidsignaal als tank leeg
- Lampje of display over verbruik etc
- Snelheidsbegrenzer

- *Actuator airbag*
- *Trillingen lane assist*

Discussie

Noem redenen waarom een organisatie Big Data in zal zetten ook als er al veel relationele databases aanwezig zijn in de organisatie?

Als het teveel data wordt voor een relationele databases bijvoorbeeld door sensor data, laag gestructureerde data dan moet je op een gegeven moment wel

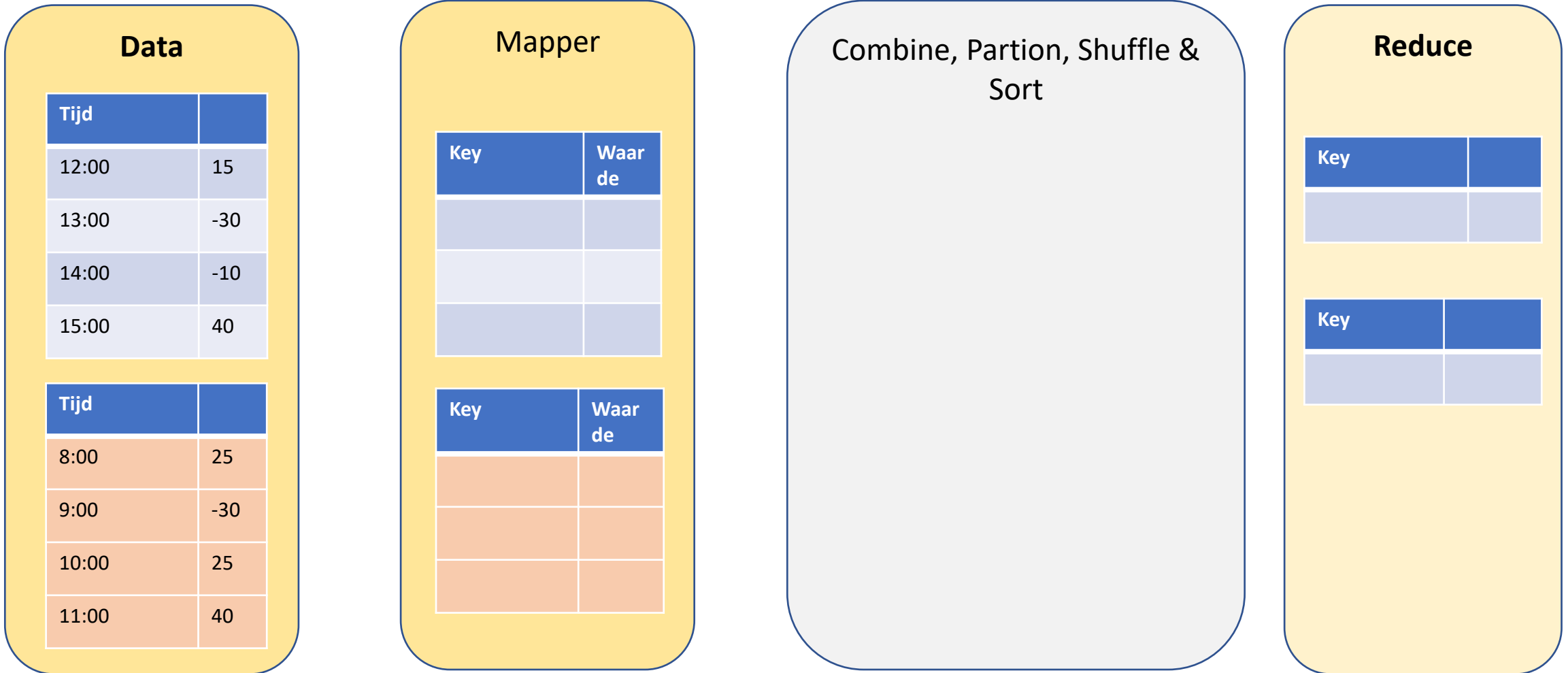
16 Extra Opgave Mapreduce

- Voor de hoeveelheid brandstof wordt een model opgesteld hoeveel er gebruikt en aangeleverd wordt. Verbruik is positief en aanvullen is negatief
- Bereken hoeveel er in de voorbeeld dataset verbruikt en aangeleverd wordt
- **Extra** wat is het maximale aanlevering inclusief de tijdsloten wanneer die aanlevering plaatsvindt
- Presenteer je model aan de groep

Tijd	
12:00	15
13:00	-30
14:00	-10
15:00	40

Tijd	
8:00	25
9:00	-30
10:00	25
11:00	40

Case brandstof opgave MapReduce



Case brandstof uitwerking MapReduce

Data

Tijd	Waarde
12:00	15
13:00	-30
14:00	-10
15:00	40

Tijd	Waarde
8:00	25
9:00	-30
10:00	25
11:00	40

Mapper

Eenheid	
Verbruik	55
Aanleveren	-40
Max	-30, 13

Eenheid	
Verbruik	90
Aanleveren	-30
Max	-30, 9

Combine, Partion, Shuffle & Sort

Tijd	
Verbruik	55,90

Tijd	
Aanleveren	-40,-30

Tijd	
Max	-30, [13:00, 9:00]

Reduce

Tijd	
Verbruik	145

Tijd	
Aanleveren	-70

Tijd	
Max	-30,[13:00, 9:00]

17 Opgave big data

- Je wilt het brandstofverbruik van je auto analyseren
- Is de brandstofanalyse een big data toepassingen
- Geef argumenten waarom wel of niet