

Big Data Discovery - Verkeersveiligheid



AGENDA / INHOUD

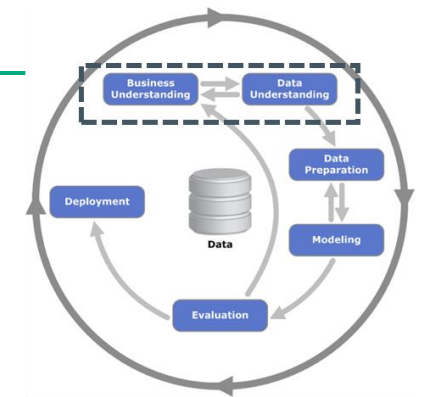
- Project overzicht & Aanloop
- Werk & Resultaten
- Samenvatting & Feedback

Project Team

HPE
Dirk Struyf
Peter Vandermeersch
Bruno Josso
Lukasz Cizak
Rafal Wawrzynczyk
Anna Mag Grzelak
MOW
Bieke Moelans
Pascal Lammar
Nathalie Gosseye
Bart Van Herbruggen

Key Dates	
21 Oktober	Business Werksessie
	Data Intake
21 November	Week 1
28 November	Week 2
5 December	Week 3 + Tussentijdse Evaluatie
12 December	Week 4
19 December	Week 5 + Tussentijdse Evaluatie
2 Januari	Week 6
9 Januari	Week 7
16 Januari	Week 8
23 Januari	Week 9 + Technische Presentatie
30 Januari	Finale Presentatie

Business Werksessie



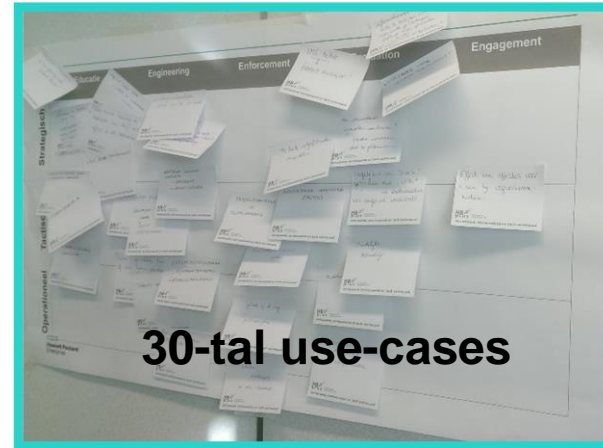
Cross Industry Standard Process for Data Mining

Context & Inspiratie

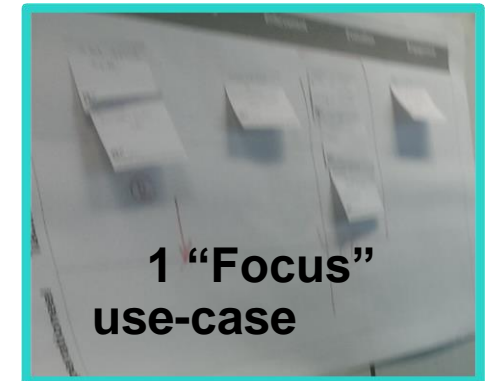
Tabel 2: Belangrijke problemen/opdringen op vlak van verkeersveiligheid, met prioritering (1, 2, 3, 4, 5) naar vervoerswijze, waarbij een hoger aantal knipjes overeenkomt met een hogere prioriteit. Deze 'ordinaire' inschatting is gebaseerd op reële cijfers en onderzoeksresultaten (zowel Vlaamse, nationale als internationale).

Probleemgebieden / weggebruikers	Persoonswagen	Lichte vrachtwagen	Vrachtwagen	Motorfiets	Bromfiets	Fiets	Voetganger	Autocars/openbaar vervoer
Snelfeild	xxx	xxx	x	xxx	xx	x(1)	-	x
ROI alcohol	xxx	xxx	x	xx	xx	x	-	x
ROI drugs	xxx			xx				
ROI medicijnen	xxx			xx	xx	xxx		
Afleiding	xxx	xxx	xxx	xx	xx	xx	xx	xx
Vermoedheid	xxx	xxx	xxx	xx	xx	xx	xx	xx
Respect veiligheidsafstand	xxx	xxx	xxx	x	x	x	-	xx
Zichtbaarheid in het verkeer	xx	xx	xx	xxx	xx	xxx	xxx	x
Gebruik persoonlijke beveiligingsmiddelen	xxx							
Roodlichtregatie	xxx							
Dode hoek	xx							
Verlenen voorrang	xxx							

Werksessie



Prioriteit & Keuze



Onderzoeksvragen & Data

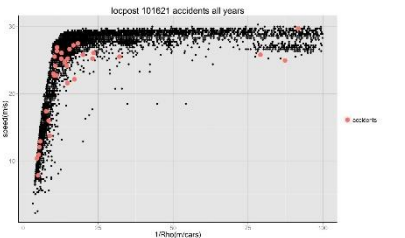
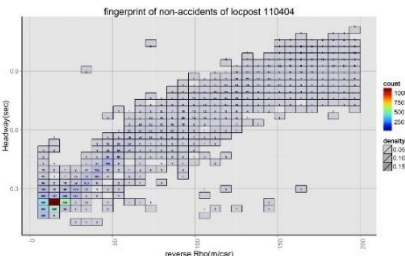
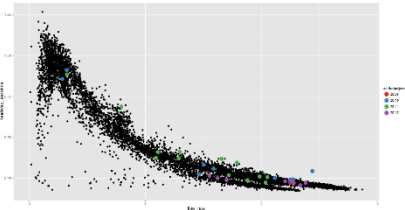


- 9-tal stappen/vragen
- 10-tal data bronnen

- Verkeersveiligheidsplan
- Vorige onderzoeken
- Ervaringen

De geselecteerde Use Case : Ongevulsfactoren op snelwegen

Vertrekkende van het vorige ongevallen onderzoek...



1. Onderzoeken van nieuwe ongevulsfactoren door gebruik te maken van meer data en ruwere informatie
2. Vergelijkingen maken tussen veilige en onveilige plaatsen.

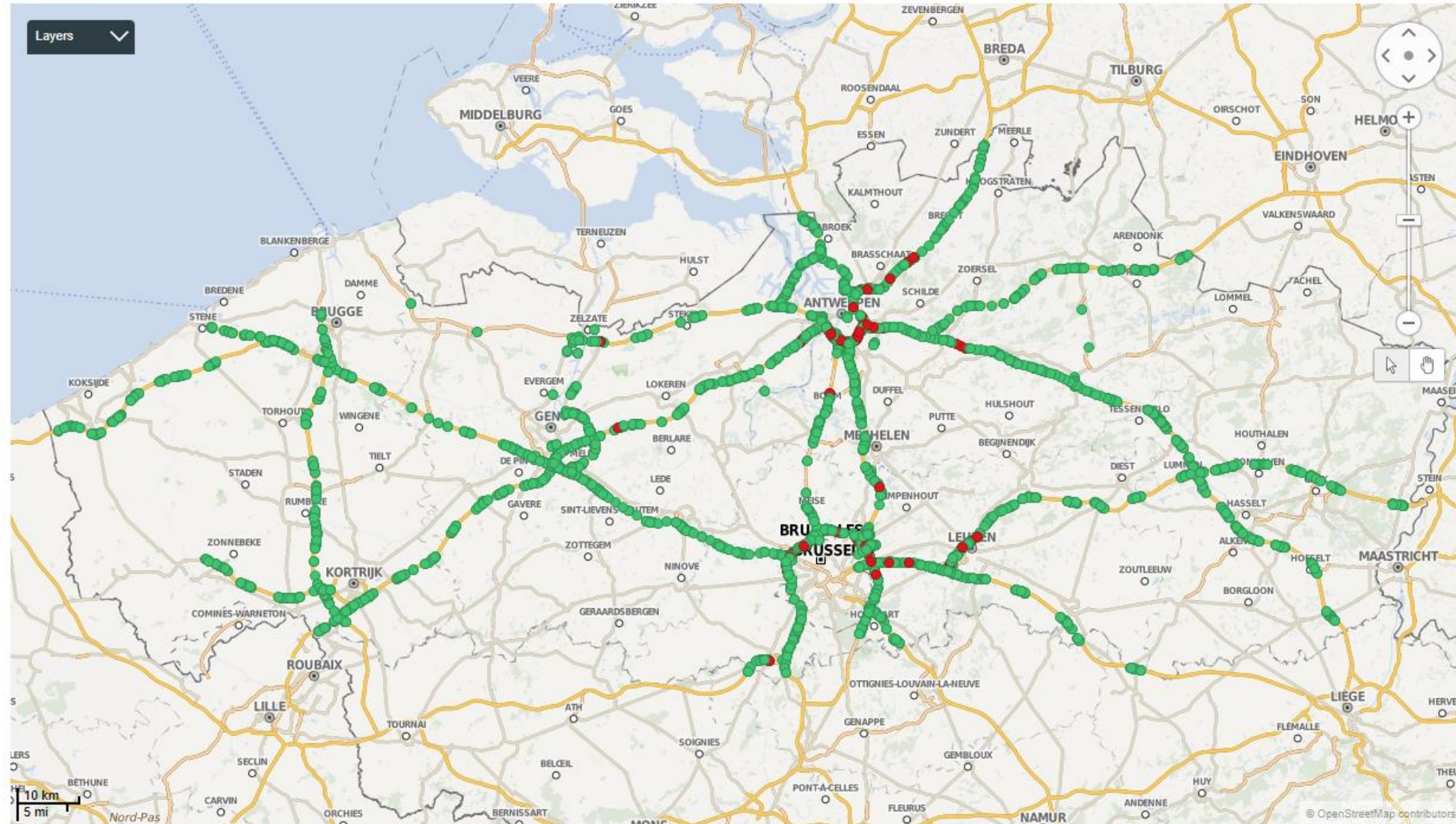
UCID	Naam	Beschrijving(en) eerder onderzoek	Doel
1	Analyse van gevaarlijke punten/segmenten	Vervolgzetting van de oorspronkelijke vingervrintanalyse. Wat zijn verzwarde factoren op de gevaarlijke punten (verkeerssamenstelling, rijgedrag, ...) Wanneer, en onder welke omstandigheden gebeuren er op deze plaatsen minder ongevallen?	(Nieuwe) factoren identificeren die tot ongevallen leiden, zodat deze kunnen aangepakt worden, en de gevolgen van ongevallen op de weg beter begrijpen
2	Gevaarlijke punten vergelijken met veilige punten	Identificeren van veilige punten, met gelijkaardige kenmerken dan gevaarlijke punten, en op basis van een vergelijk tussen beiden factoren vaststellen die de veiligheid bevorderen	"bright spots" identificeren, zodat we hieruit kunnen leren, en deze kunnen repliceren.
3	Wegwerkzaamheden, of andere, tijdelijk versturende situaties	Kijken hoe afhankelijk verkeersgedrag (snelheid, onrustigheid, ...) is van tijdelijk versturende verkeerssituaties zoals wegenwerken, metingen, afsluttingen, ... Heeft dit een positief of negatief effect op ongevallen en veiligheid? Is een situatie voor, na, of tijdens beter of slechter in termen van veiligheid?	
4	Volgafstanden/rijgedrag aanpakken	Over heel Vlaanderen op alle snelwegen de afstanden tussen voertuigen analyseren en in kaart brengen, wat is het effect van deze afstand naar veiligheid op alle wegen in Vlaanderen? Waar zijn extra controles zeker nodig? Wat is de ideale afstand in termen van veiligheid?	
5	Trajectcontrole/ flitscamera's en mobiele snelheids controles	Waar zouden trajectcontroles effectiever zijn, waar flitscamera's. Stollen we punten vast waar te snel rijden lokaal sterk voorkomt, kunnen we gehele trajecten identificeren waar extra trajectcontrole interessant zou zijn, op basis van de huidige verkeersdata?	
6	Rijgedrag analyse	Rijgedrag over heel Vlaanderen in kaart brengen op basis van historische gegevens, verschillen per regio vaststellen, gerelateerde factoren opsporen.	
7	Zichtbaarheid op de wegen	Identificeren van slechte zichtbaarheid op wegen aan de hand van windrichtingen van wegen en uren, zonsopgang en ondergang, weerdata, data over verlichting op de wegen. Vergelijk met ongevallen, rijgedrag. Model bouwen om de verlichting life aan te passen op basis van huidige/voorspelde (weers)omstandigheden	
8	Onderzoek naar inhalen vrachtwagens/middenvakrijders/	Vaststellen op basis van de verkeersdata waar en wanneer middenvakrijders voorkomen en waar en wanneer zwaar transport voertuigen inhaalbewegingen maken, en of dit tot problemen (ongevallen) of gevaarlijk rijgedrag leidt	
9	Maximale vs optimale snelheden bekijken in bepaalde omstandigheden	Wanneer wordt het gevaarlijk, zodat we de snelheid preventief kunnen aanpassen. Locaties identificeren om extra RSS borden te installeren.	
10	sensibilisatie gericht naar doelgroepen/momenten	Analyse van gedane campagnes op rijgedrag	
11	trajectsturing	Automatisch alternatieve wegen aanduiden of afsluiten om toekomstige verkeerhinder op kerfwegen te vermijden.	
12	Seizoenseffect op ongevallen	analyse van ongevallen en oorzaken van ongevallen doorheen het jaar, een seizoensgebonden trends vaststellen om campagnes/verplichte winterbanden/ABS systemen op te baseren	
13	Secundaire ongevallen-analyse	Analyse van ongevallen ten opzichte van elkaar en de wegdata, om secundaire ongevallen te identificeren en mogelijk te voorspellen, factoren bepalen om deze beter te voorkomen in de toekomst	
14	Alcoholgebruik	Alcoholgebruik bij ongevallen of controles in kaart brengen, gelinkt aan regio's, events, externe factoren, ...	

Wat bepaalt of een Locpost veilig is of onveilig?
Zijn bepaalde Locposts veiliger dan andere?
Wat zijn de bepalende factoren?
Kunnen we ongevallen voorspellen?
Welke informatie hebben we daarvoor nodig?
Hoe doen we die voorspelling het best?
Kunnen we modellen veralgemenen?
Zijn er patronen te herkennen die ongevallen veroorzaken?
Zijn er andere dimensies bepalend?



De voorbereiding

Map Chart



Totaal: 4014 LocPosts – 5703 Meetpunten
Geselecteerd : 156 LocPosts – 552 Meetpunten

De beschikbare data

meetpunt_id	meettijd	volgtijd	snelheid_kmph	elek_lengte_dm
460	3/17/2015 12:00:46 AM	3001	91	88
460	3/17/2015 12:00:55 AM	902	98	47
460	3/17/2015 12:01:23 AM	2805	134	45
460	3/17/2015 12:03:24 AM	3001	87	166
460	3/17/2015 12:03:50 AM	2590	108	47
460	3/17/2015 12:04:43 AM	3001	120	44
460	3/17/2015 12:05:33 AM	3001	123	53
460	3/17/2015 12:05:50 AM	1766	99	42
460	3/17/2015 12:05:59 AM	840	95	43
460	3/17/2015 12:06:04 AM	542	125	47
460	3/17/2015 12:06:14 AM	989	148	45
460	3/17/2015 12:06:19 AM	527	110	50
460	3/17/2015 12:07:42 AM	3001	105	47

Individuele Doortochten - 6.989.879.746

TIME_MEASURED	I123	I45	It	Ipwe	Vh123	Vh45	Vht	Vhpwe	Bt
2015-03-27 09:01:00	30	4	34	38	106	84	103	103	5
2015-03-27 09:02:00	46	0	46	46	99		99	99	6
2015-03-27 09:03:00	27	3	30	33	109	86	106	106	4
2015-03-27 09:04:00	24	6	30	36	100	87	97	97	6
2015-03-27 09:05:00	37	6	43	49	100	84	97	97	7
2015-03-27 09:06:00	22	4	26	30	106	84	102	102	5
2015-03-27 09:07:00	40	6	46	52	96	84	94	94	8
2015-03-27 09:08:00	22	6	28	34	109	88	104	104	5
2015-03-27 09:09:00	24	7	31	38	103	86	99	99	7
2015-03-27 09:10:00	37	6	43	49	102	83	98	98	7
2015-03-27 09:11:00	26	6	32	38	106	88	102	102	5
2015-03-27 09:12:00	27	5	32	37	100	86	98	98	6
2015-03-27 09:13:00	30	5	35	40	99	82	96	96	6
2015-03-27 09:14:00	32	3	35	38	102	85	101	101	5
2015-03-27 09:15:00	27	3	30	33	107	90	105	105	4
2015-03-27 09:16:00	22	1	23	24	105	84	104	104	3

1 Minuut Aggregaties - 189.937.836

Ongevallen

63.535

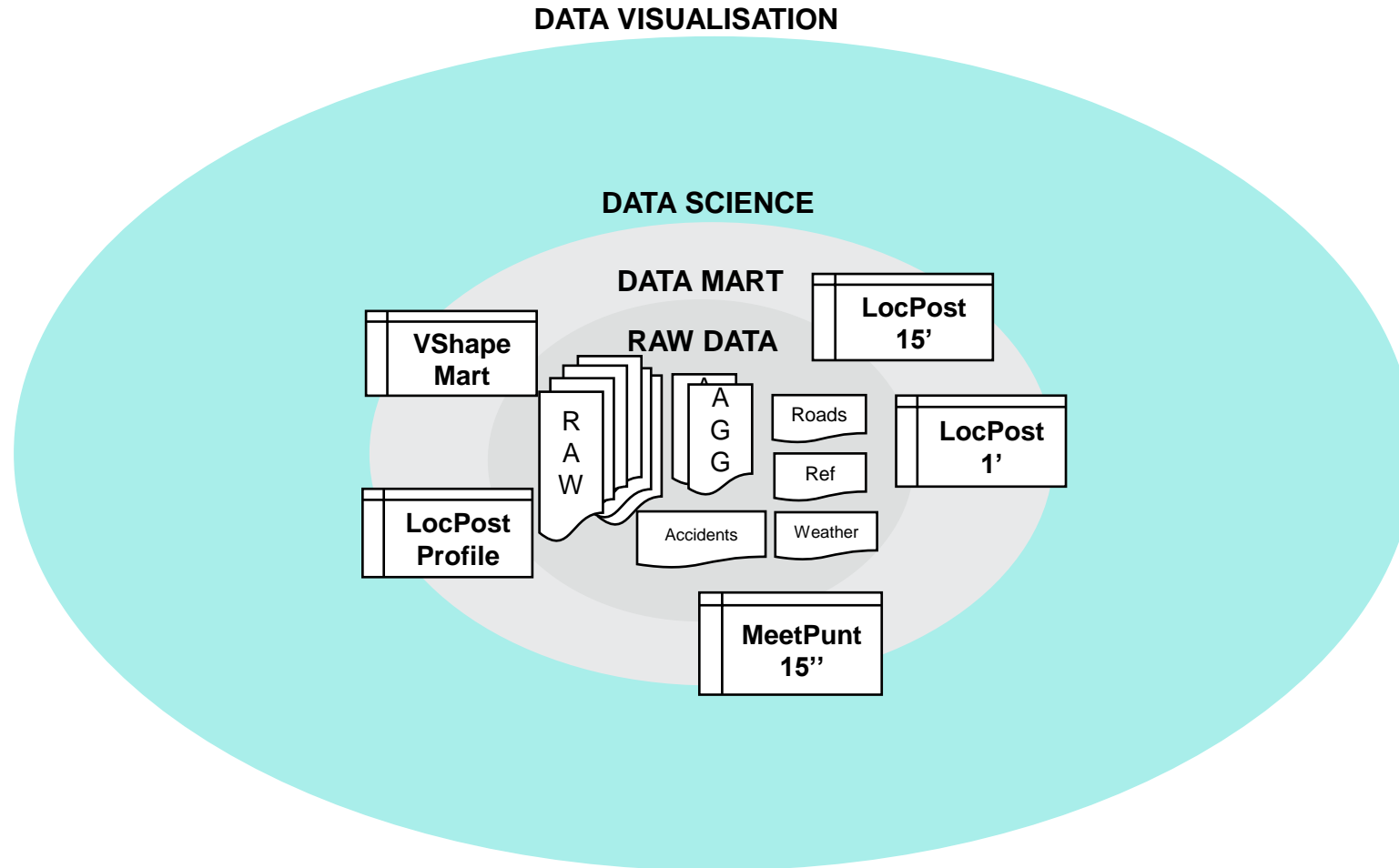
Referentie Data
(Kalender, Infrastructuur)

Meto
(Weer – Zonlicht)

Wegstatus

Events

Data Marts



Opbouw Data Marts

Aggregaties per tijdseenheid

meetpunt_id	meettijd	volatiid	snelheid kmph	elek lenate dm
460	3/17/2015 12:00:46 AM	3001	91	88
460	3/17/2015 12:00:55 AM	902	98	47
460	3/17/2015 12:01:23 AM	2805	134	45
460	3/17/2015 12:03:24 AM	3001	87	166
460	3/17/2015 12:03:50 AM	2590	108	47
460	3/17/2015 12:04:43 AM	3001	120	44
460	3/17/2015 12:05:33 AM	3001	123	53
460	3/17/2015 12:05:50 AM	1766	99	42
460	3/17/2015 12:05:59 AM	840	95	43
460	3/17/2015 12:06:04 AM	542	125	47
460	3/17/2015 12:06:14 AM	989	148	45
460	3/17/2015 12:06:19 AM	527	110	50
460	3/17/2015 12:07:42 AM	3001	105	47

Individuele Doortochten

MAX

MIN

AVG

StDev

Aantal

MAX

MIN

AVG

StDev

Aantal

MAX

MIN

AVG

StDev

Aantal

Groepering per minuut of per 15 minuten

Opbouw Data Marts

Combinaties over tijdseenheden

meetpunt id	meettijd	volatiid	snelheid kmph	elek lenate dm
460	3/17/2015 12:00:46 AM	3001	91	88
460	3/17/2015 12:00:55 AM	902	98	47
460	3/17/2015 12:01:23 AM	2805	134	45
460	3/17/2015 12:03:24 AM	3001	87	166
460	3/17/2015 12:03:50 AM	2590	108	47
460	3/17/2015 12:04:43 AM	3001	120	44
460	3/17/2015 12:05:33 AM	3001	123	53
460	3/17/2015 12:05:50 AM	1766	99	42
460	3/17/2015 12:05:59 AM	840	95	43
460	3/17/2015 12:06:04 AM	542	125	47
460	3/17/2015 12:06:14 AM	989	148	45
460	3/17/2015 12:06:19 AM	527	110	50
460	3/17/2015 12:07:42 AM	3001	105	47

Individuele Doortochten

MAX MIN AVG
StDev Aantal

MAX MIN AVG
StDev Aantal

Vorige – Verschil - Ratio

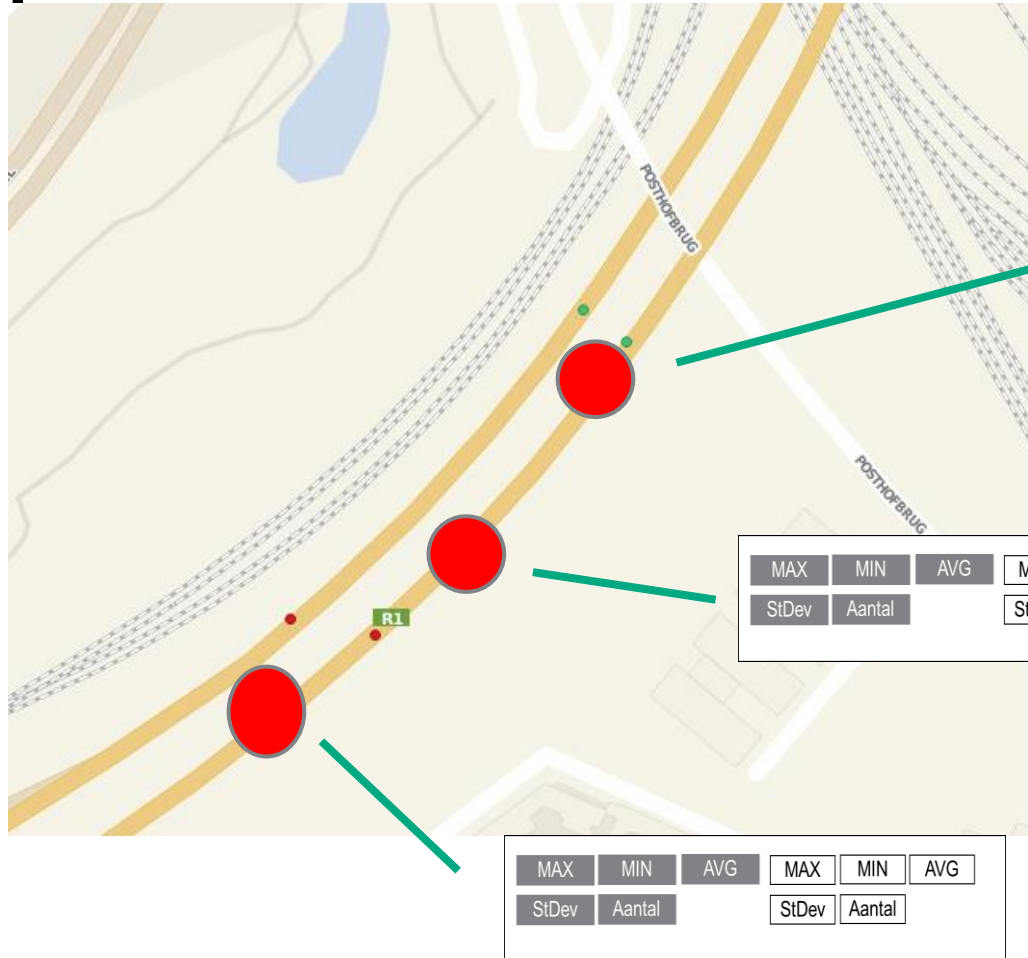
MAX MIN AVG
StDev Aantal

Vorige – Verschil - Ratio

Groepering per minuut of per 15 minuten

Opbouw Data Marts

Combinaties over locposten

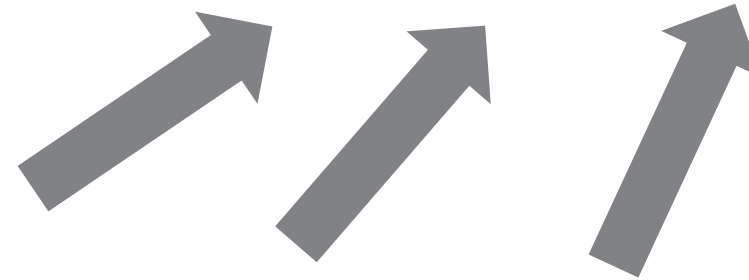


Noot: Soms ook tegenovergestelde richting genomen

Opbouw Data Marts

Toevoegen andere variabelen

MAX	MIN	AVG	MAX	MIN	AVG	MAX	MIN	AVG	MAX	MIN	AVG
StDev	Aantal		StDev	Aantal		StDev	Aantal		StDev	Aantal	



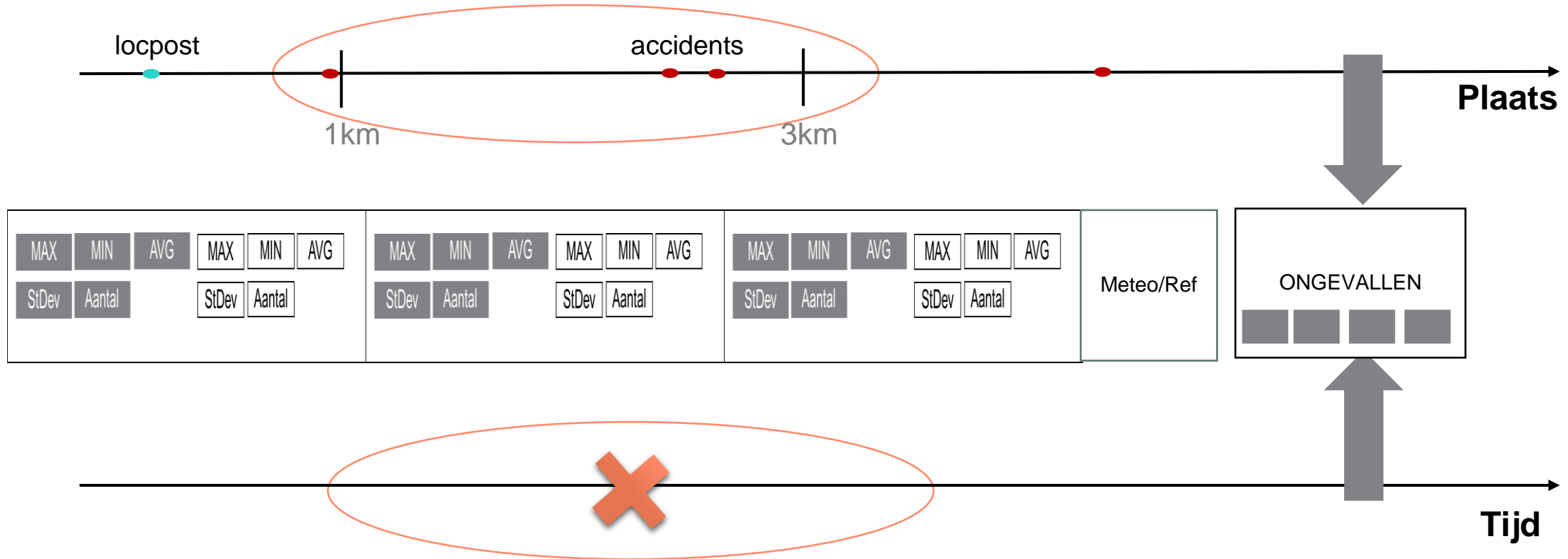
Meto
(Weer – Zonlicht)

Referentie Data
(Kalender, Infrastructuur)

Wegstatus

Opbouw Data Marts

Linken Ongevallen



Opbouw Data Marts

- Columns
- locpost
- time_chunk
- speed_min
- speed_avg
- speed_harmonic_avg
- speed_max
- speed_stddev
- speed_V
- vehicle_count
- kinetic_energy_min
- kinetic_energy_avg
- kinetic_energy_max
- kinetic_energy_sum
- kinetic_energy_stddev
- kinetic_energy_V
- momentum_min
- momentum_avg
- momentum_max
- momentum_sum
- momentum_stddev
- momentum_V
- timespan_min
- timespan_avg
- timespan_max
- timespan_stddev
- timespan_V
- vehicle_0_20_count
- vehicle_21_40_count
- vehicle_41_60_count
- vehicle_61_80_count
- vehicle_81_100_count
- vehicle_101_120_count
- vehicle_above_120_cou

- car_vehicle_count
- van_vehicle_count
- small_truck_vehicle_count
- large_truck_vehicle_count
- car_vehicle_ratio
- van_vehicle_ratio
- small_truck_vehicle_ratio
- large_truck_vehicle_ratio
- vehicle_0_20_ratio
- vehicle_21_40_ratio
- vehicle_41_60_ratio
- vehicle_61_80_ratio
- vehicle_81_100_ratio
- vehicle_101_120_ratio
- vehicle_above_120_ratio
- avg_5min_speed_min
- avg_5min_speed_avg
- avg_5min_speed_harmonic
- avg_5min_speed_max
- avg_5min_speed_stddev
- avg_5min_vehicle_count
- avg_5min_kinetic_energy_m
- avg_5min_kinetic_energy_av
- avg_5min_kinetic_energy_m
- avg_5min_kinetic_energy_st
- avg_5min_timespan_min
- avg_5min_timespan_avg
- avg_5min_timespan_max
- avg_5min_timespan_stddev
- avg_5min_vehicle_0_20_ratic
- avg_5min_vehicle_above_12

- stddev_5min_speed_harmonic_av
- stddev_5min_speed_max
- stddev_5min_speed_stddev
- stddev_5min_vehicle_count
- stddev_5min_kinetic_energy_min
- stddev_5min_kinetic_energy_avg
- stddev_5min_kinetic_energy_max
- stddev_5min_kinetic_energy_stddev
- stddev_5min_timespan_min
- stddev_5min_timespan_avg
- stddev_5min_timespan_max
- stddev_5min_timespan_stddev
- stddev_5min_vehicle_0_20_ratio
- stddev_5min_vehicle_above_120
- change_speed_min
- change_speed_avg
- change_speed_harmonic_avg
- change_speed_max
- change_speed_stddev
- change_vehicle_count
- change_kinetic_energy_min
- change_kinetic_energy_avg
- change_kinetic_energy_max
- change_kinetic_energy_stddev
- change_timespan_min
- change_timespan_avg
- change_timespan_max
- change_timespan_stddev
- change_5min_speed_min
- change_5min_speed_avg
- change_5min_speed_harmonic_a
- change_5min_speed_max
- change_5min_speed_stddev
- change_5min_vehicle_count

- stddev_5min_vehicle_count
- stddev_5min_kinetic_energy_min
- stddev_5min_kinetic_energy_avg
- stddev_5min_kinetic_energy_max
- stddev_5min_kinetic_energy_stddev
- stddev_5min_timespan_min
- stddev_5min_timespan_avg
- stddev_5min_timespan_max
- stddev_5min_timespan_stddev
- stddev_5min_vehicle_0_20_ratio
- stddev_5min_vehicle_above_120_ratic
- change_speed_min
- change_speed_avg
- change_speed_harmonic_avg
- change_speed_max
- change_speed_stddev
- change_vehicle_count
- change_timespan_max
- change_timespan_stddev
- change_5min_speed_min
- change_5min_speed_avg
- change_5min_speed_harmonic_avg
- change_5min_speed_max
- change_5min_speed_stddev
- change_5min_vehicle_count
- change_5min_kinetic_energy_min
- change_5min_kinetic_energy_avg
- change_5min_kinetic_energy_max
- change_5min_kinetic_energy_stddev
- change_5min_timespan_min
- change_5min_timespan_avg
- change_5min_timespan_max
- change_5min_timespan_stddev

- bridge_day_flag
- calendar_month
- calendar_weekday_name
- calendar_national_holiday_flag
- calendar_regional_holiday_flag
- calendar_core_construction_holiday_flag
- calendar_edge_construction_holiday_flag
- calendar_school_holiday_flag
- calendar_student_vacation_holiday_flag
- calendar_day_type
- calendar_weekday_flag
- ci prev_kinetic_energy_min
- ci prev_kinetic_energy_avg
- ci prev_kinetic_energy_max
- ci prev_kinetic_energy_sum
- prev_momentum_min
- prev_momentum_avg
- prev_momentum_max
- prev_momentum_sum
- prev_timespan_min
- prev_timespan_avg
- prev_timespan_max
- prev_vehicle_0_20_count
- prev_vehicle_21_40_count
- prev_vehicle_41_60_count
- prev_vehicle_61_80_count
- prev_vehicle_81_100_count
- prev_vehicle_101_120_count
- prev_vehicle_above_120_count
- prev_car_vehicle_count
- prev_van_vehicle_count
- prev_small_truck_vehicle_count
- prev_large_truck_vehicle_count
- prev_car_ratio
- prev_van_ratio

- sunlight_flag
- sunlight_hour_to_sunrise
- sunlight_hour_to_sunset
- weather_temperature
- weather_pressure
- weather_visibility_km
- weather_wind_direction
- weather_wind_speed
- weather_conditions
- weather_snow
- weather_rain
- weather_fog
- weather_thunderstorm
- weather_hail

- ACCIDENT_LAST_24
- ACCIDENT_LAST_12
- ACCIDENT_LAST_6
- ACCIDENT_LAST_2
- ACCIDENT_ALL_DIRS_LAST_24
- ACCIDENT_ALL_DIRS_LAST_12
- ACCIDENT_ALL_DIRS_LAST_6
- ACCIDENT_ALL_DIRS_LAST_2
- trgt_minute_1
- trgt_minute_5
- trgt_minute_10
- trgt_minute_15
- trgt_minute_20

Voorbeeld Data Mart



1. Visuele Analyse

Map Chart

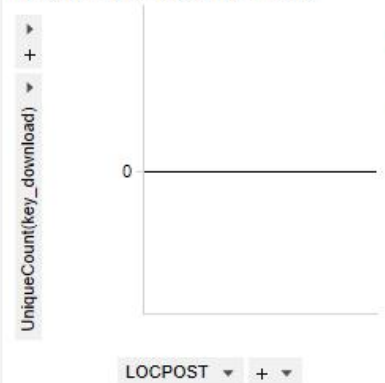


ACCIDENT_MAPPING_MEETPUNT

key_download	ACCIDENT_T...	LOC
130653/2015	10/3/2015 1:0...	1258
130653/2015	10/3/2015 1:0...	1258
130653/2015	10/3/2015 1:0...	1258
002698/2014	1/9/2014 4:38:...	1267
002698/2014	1/9/2014 4:38:...	1267
117741/2014	10/24/2014 5:...	1267
117741/2014	10/24/2014 5:...	1267
002698/2014	1/9/2014 4:38:...	1267
117741/2014	10/24/2014 5:...	1267
000707/2015	1/22/2015 6:5...	1281
000707/2015	1/22/2015 6:5...	1281
002428/2014	1/21/2014 9:5...	1281
002428/2014	1/21/2014 9:5...	1281

Data table:
ACCIDENT_M...

key_download per LOCPOST



Data limiting:
● LocPostSelection
Data table:
ACCIDENT_M...
Color by:
(None)

ACCIDENT_TIMESTAMP per LOCPOST

137101	115504	106191	102494	104691	150702	151201	109201	129902	111201	102291	102324	119302
					109991	109992	119602	109091	150401	122301	125803	109301
					118901	119301	119303	145901	12210	12514	14615	15110
					119303	125106	136601	150901	1510	1111	1111	1111
					101621	113392	158301	150001	1513	1111	1111	1111
109704	118123	104491	128101	150801	150301	130801	142120	1013	1111	1111	1111	1111

Selectie Locpost en Meetpunten

Dimension Selection

Select X-Axis: One_Over_Densit

Select Y-Axis: AVG_volgtijd

Select Z-Axis: min_s_30_cars

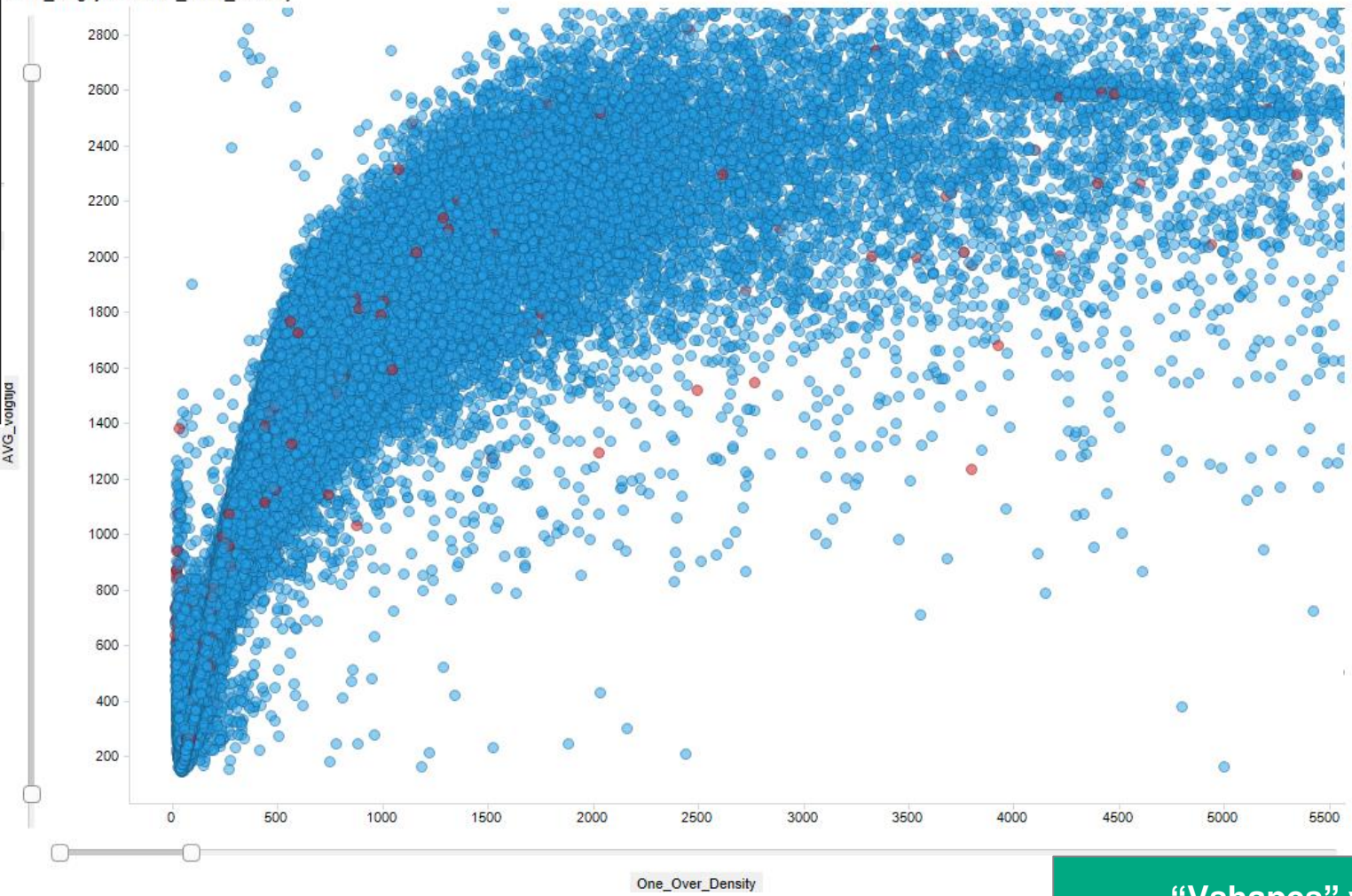
Distribution – meetpunt_id

Distribution – ACCIDENT_KEY

Dimension List:

- meetpunt_id
- meetijd_sice
- AVG_snelheid_kmph
- MAX_snelheid_kmph
- MIN_snelheid_kmph
- AVG_volgtijd
- MAX_volgtijd
- MIN_volgtijd
- avg_s_15_cars
- max_s_15_cars
- min_s_15_cars
- avg_s_30_cars
- max_s_30_cars
- min_s_30_cars
- max_speed_difference_previous_car
- min_speed_difference_previous_car
- avg_speed_difference_previous_car
- max_volgtijd_previous_car
- min_volgtijd_previous_car
- avg_volgtijd_previous_car
- avg_vt_15_cars
- max_vt_15_cars
- min_vt_15_cars
- avg_vt_30_cars
- max_vt_30_cars
- min_vt_30_cars
- number_of_cars
- ACCIDENT_KEY
- ACCIDENT_TS
- ACCIDENT_WINDOW

AVG_volgtijd vs. One_Over_Density



Data limiting:

- Meetpunt

Current page filtering

Filtering scheme

Data table: MEETPUNTE...

Marker by: meetijd_s...

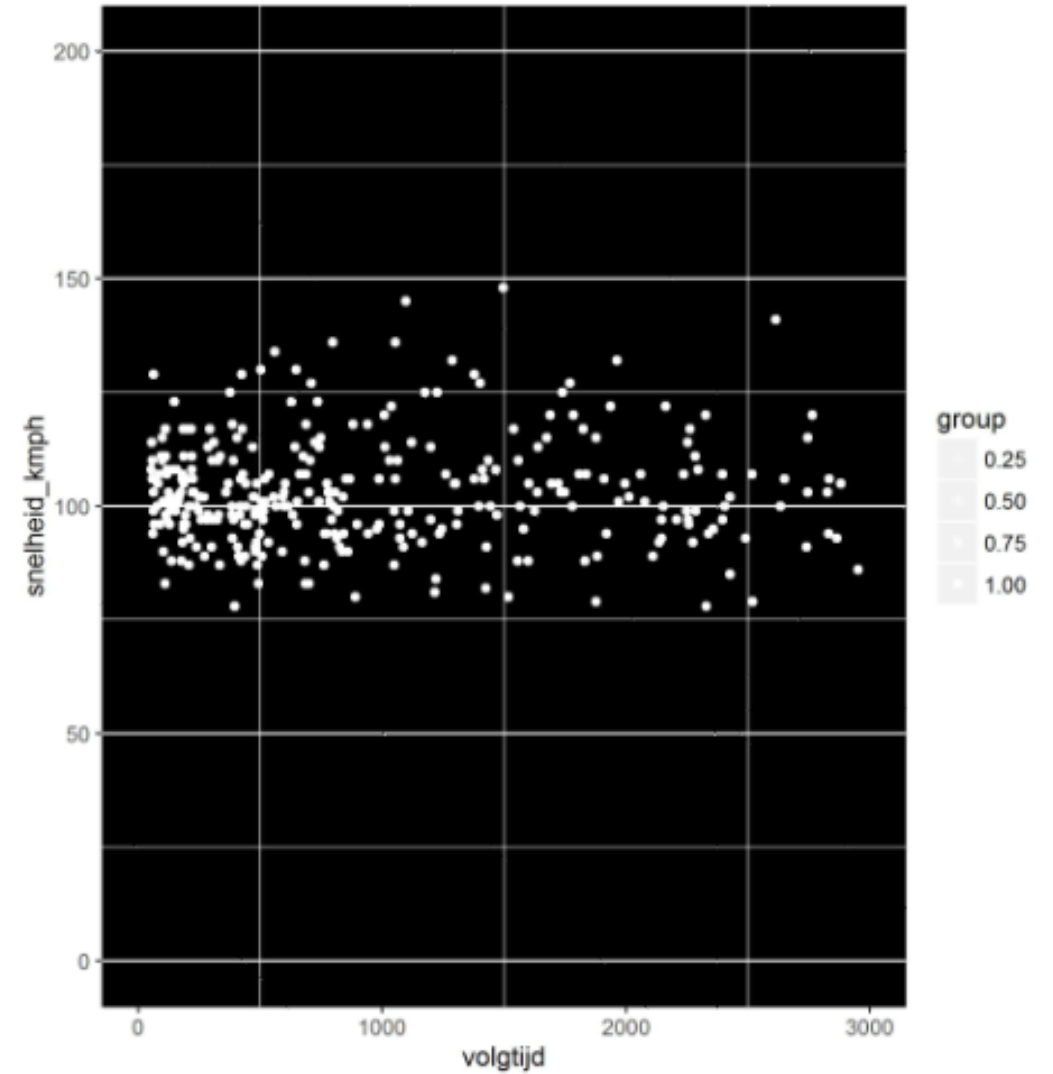
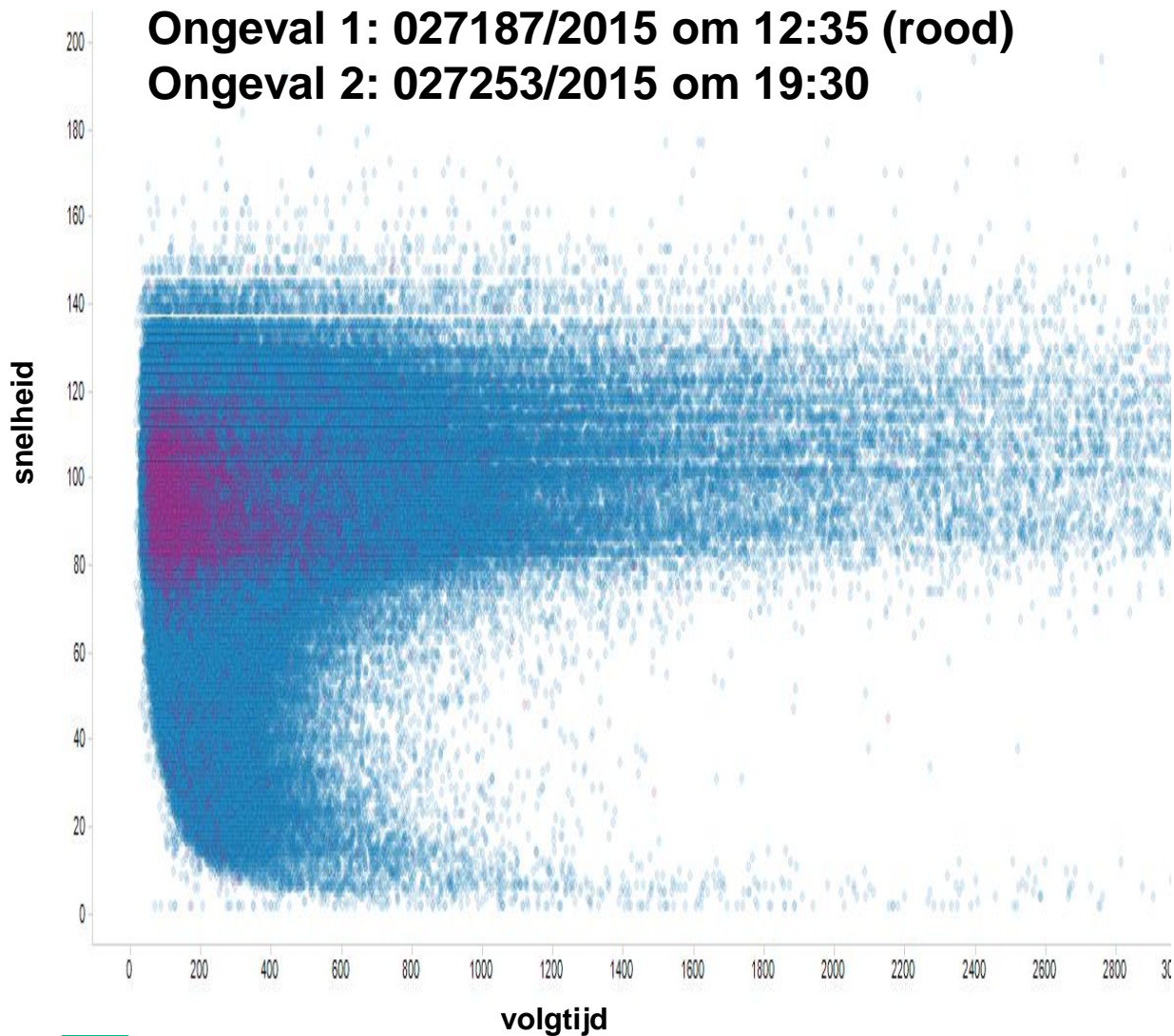
Color by: AccidentFL...

- False
- True

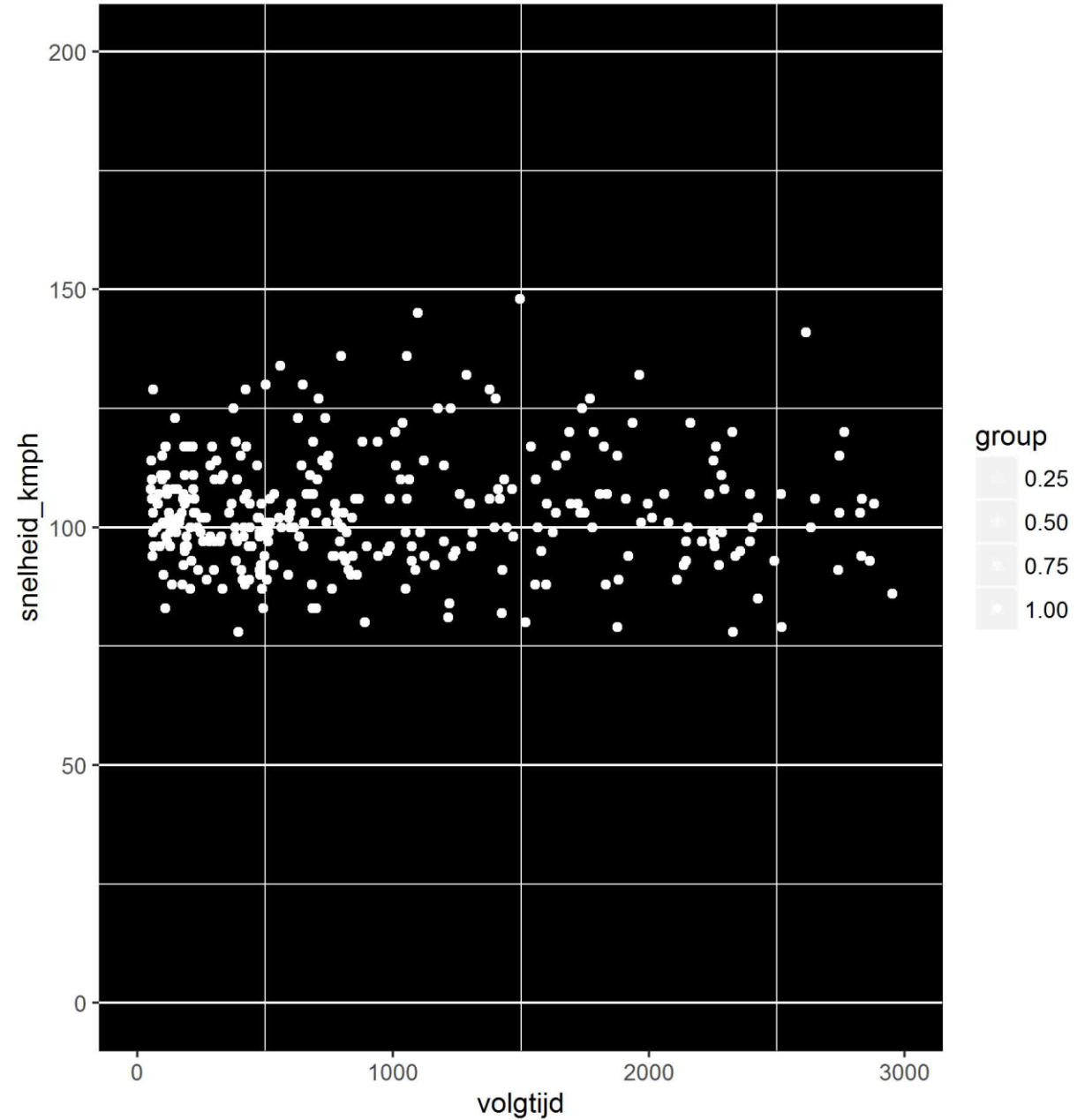
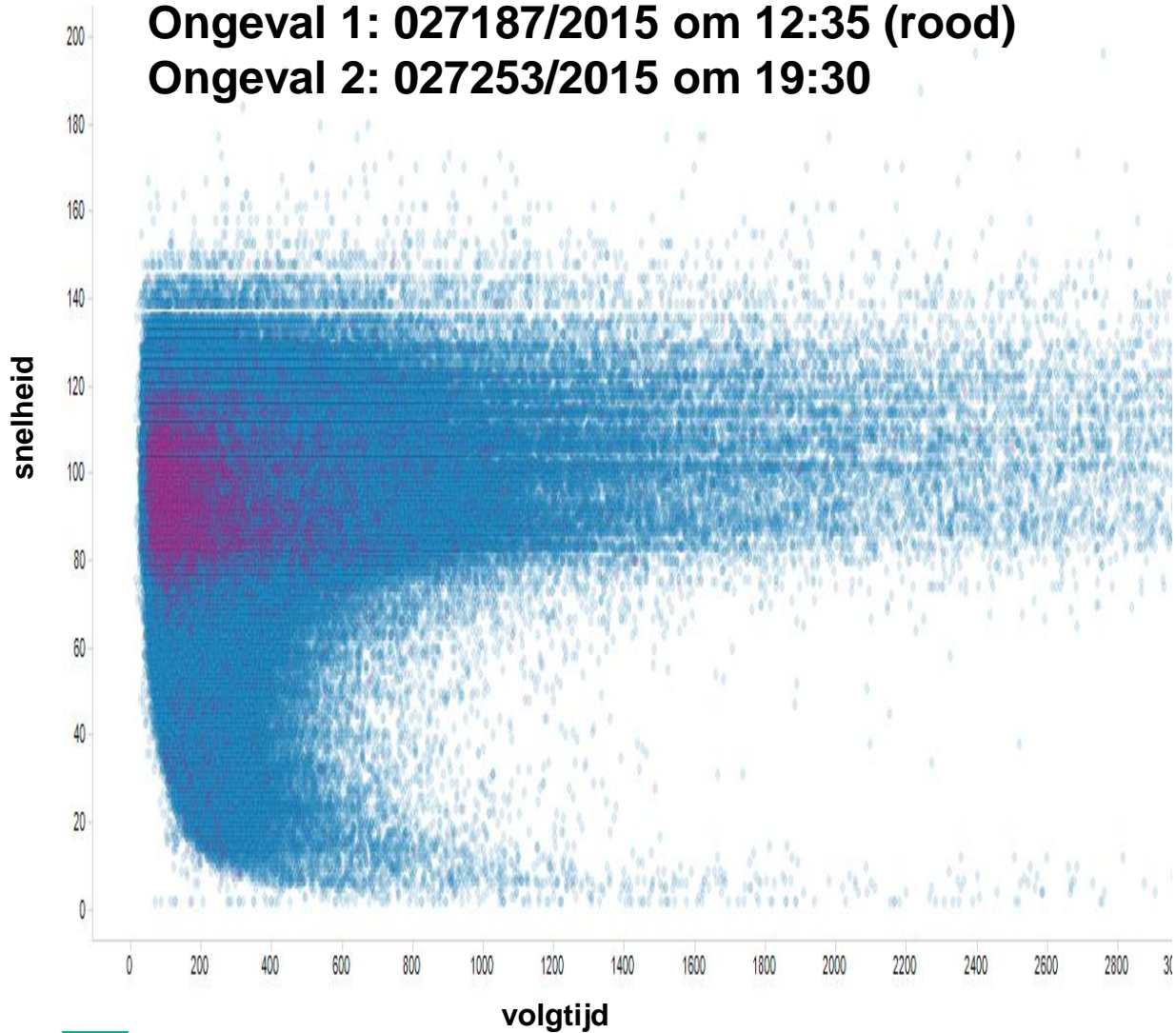
Shape by: (None)

Size by: (None)

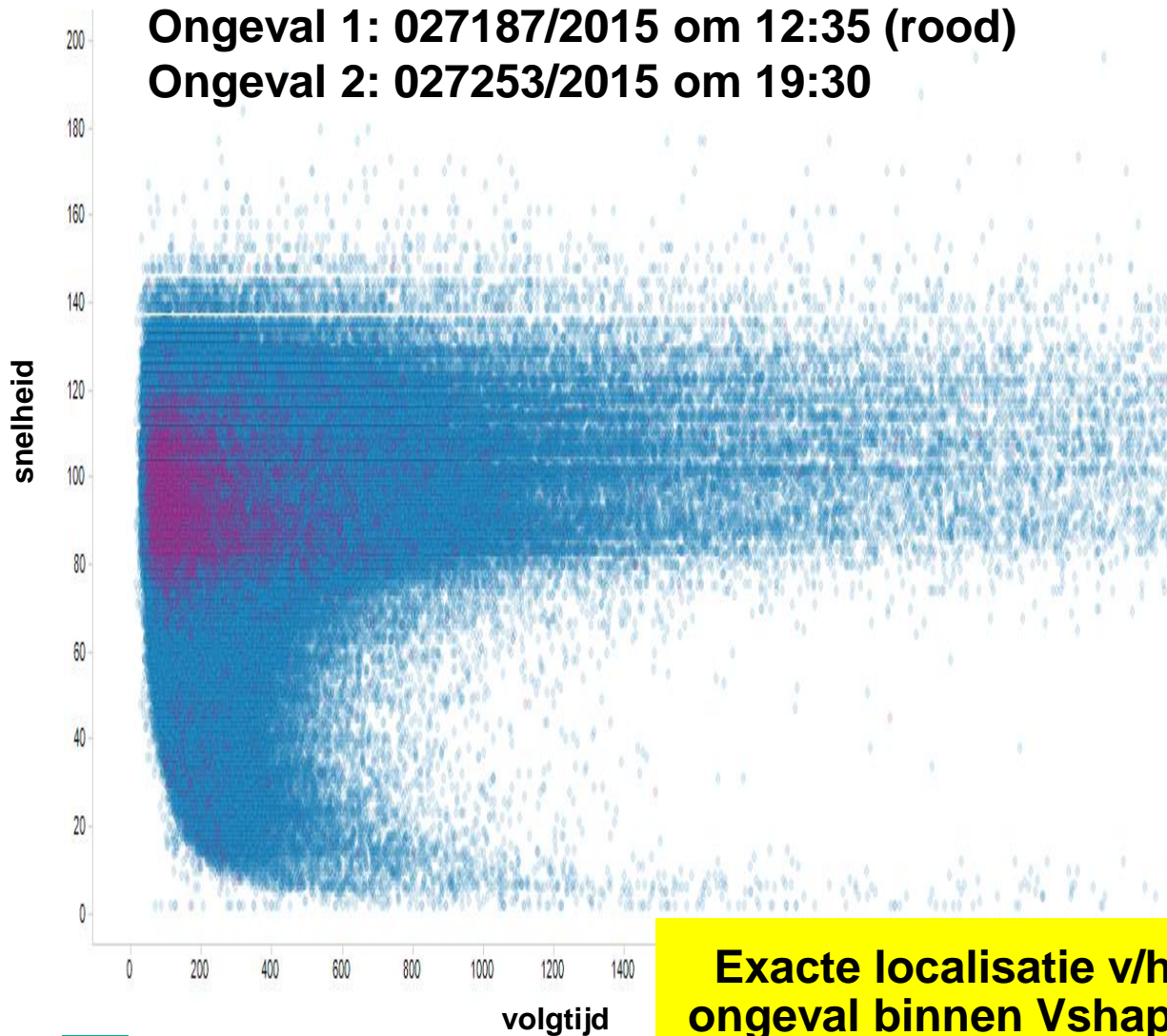
“Vshapes” voor geselecteerde Punten



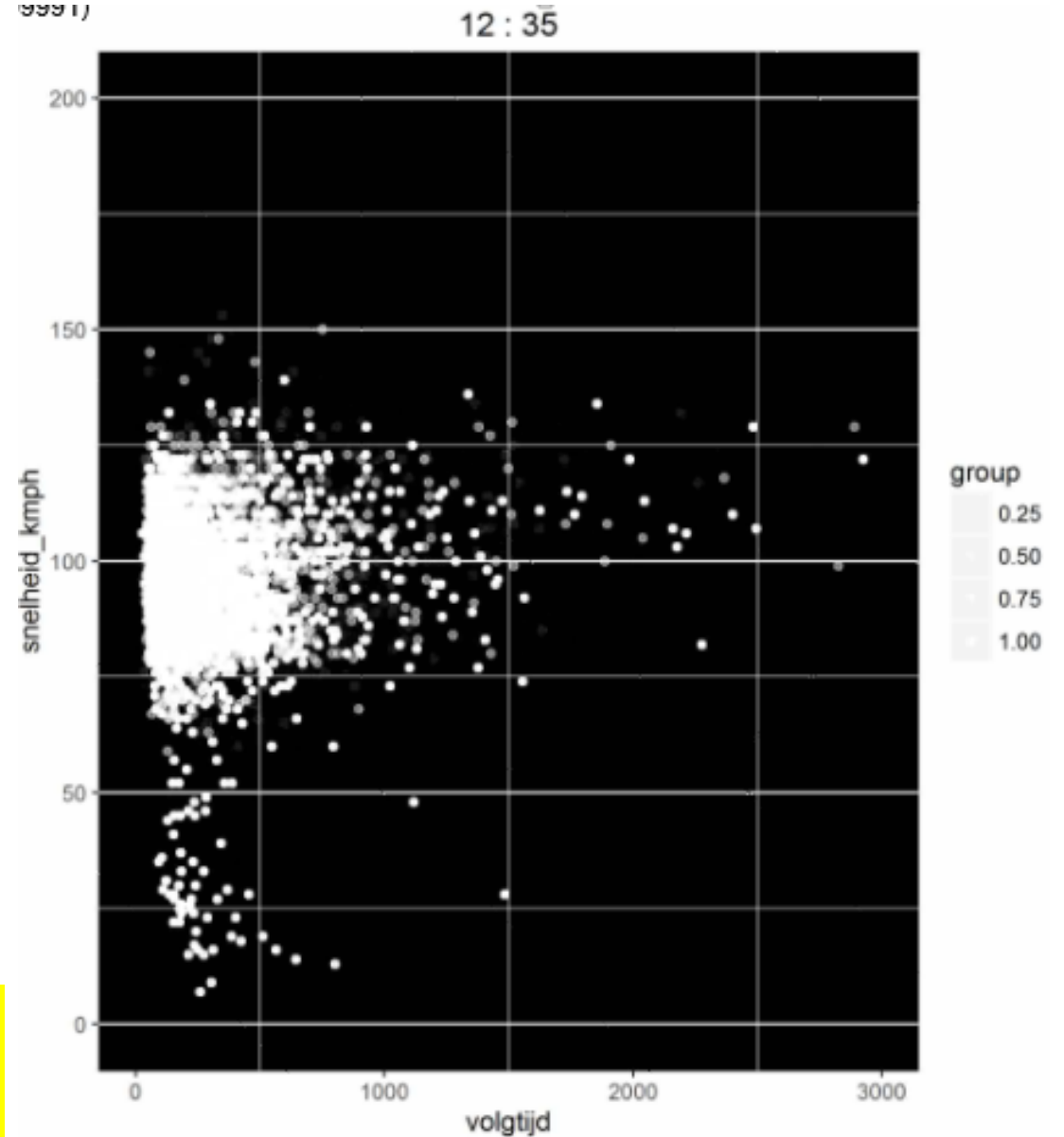
Ongeval 1: 027187/2015 om 12:35 (rood)
Ongeval 2: 027253/2015 om 19:30



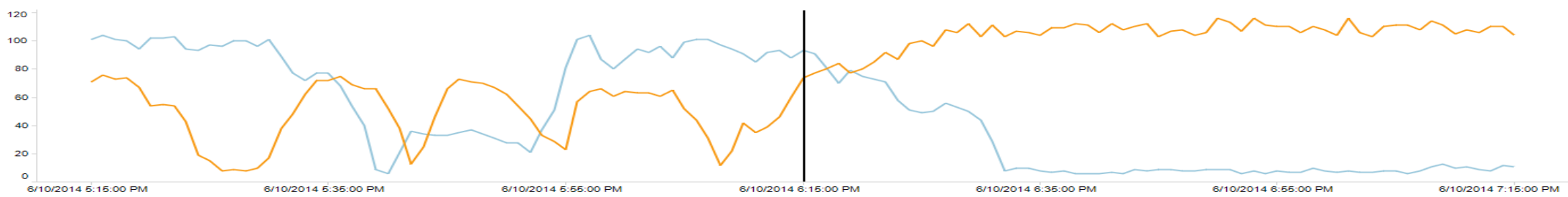
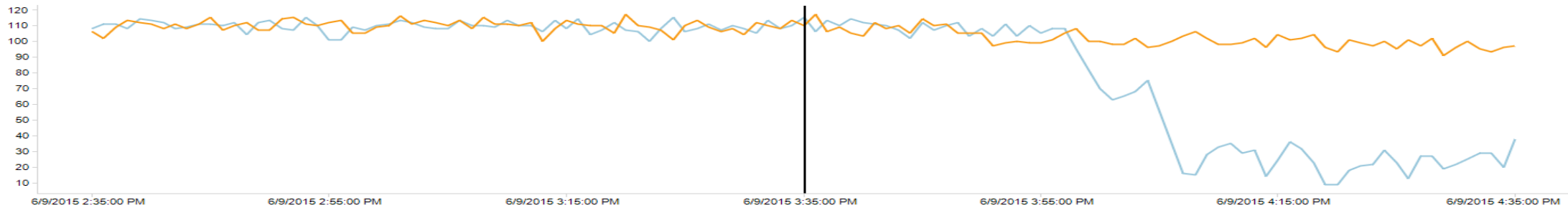
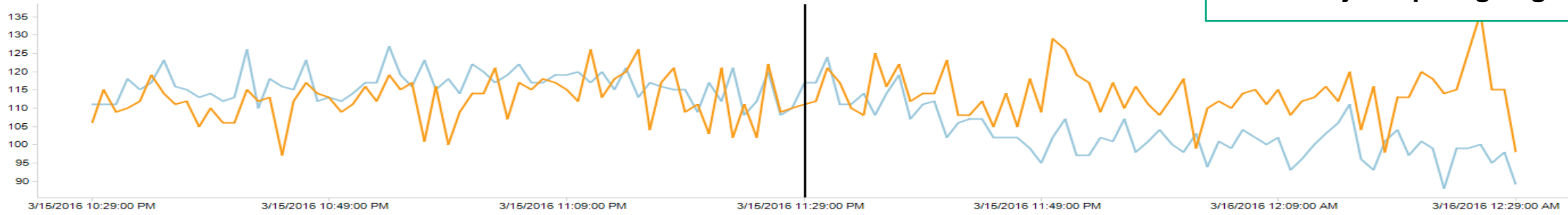
Ongeval 1: 027187/2015 om 12:35 (rood)
Ongeval 2: 027253/2015 om 19:30



**Exacte localisatie v/h
ongeval binnen Vshape
niet eenduidig.**



Betere tijdsbepaling ongeval ?



A0030001
3 Ongevallen
Hewlett Packard
Enterprise

Geen eenduidig effect van ongeval zichtbaar in snelheidsverloop.

Avg(snelheid) **Voor ongeval** **Na ongeval**

Deel 1 – Visuele Analyse

Besluiten & Relevantie

- Visuele Analyse
 - Datamart laten toe om Vshapes op te bouwen.
 - Grote variatie binnen vShape (video)
=> onderzoeken van bepaalde delen (hotspots) van de Vshape als “fingerprint” is afgezwakt.
 - Ongeval niet duidelijk herkenbaar in primaire verkeersvariabelen (cfr. snelheidsplot)
=> Piste om tijdstip van het ongeval eenvoudig te corrigeren door duidelijke snelheidsdrop is niet verder gevolgd.

Wat bepaalt of een Locpost veilig is of onveilig?

Zijn bepaalde Locposts veiliger dan andere?

Wat zijn de bepalende factoren?

Kunnen we ongevallen voorspellen?

Welke informatie hebben we daarvoor nodig?

Hoe doen we die voorspelling het best?

Kunnen we modellen veralgemenen?

Zijn er patronen te herkennen die ongevallen veroorzaken?

Zijn er andere dimensies bepalend?



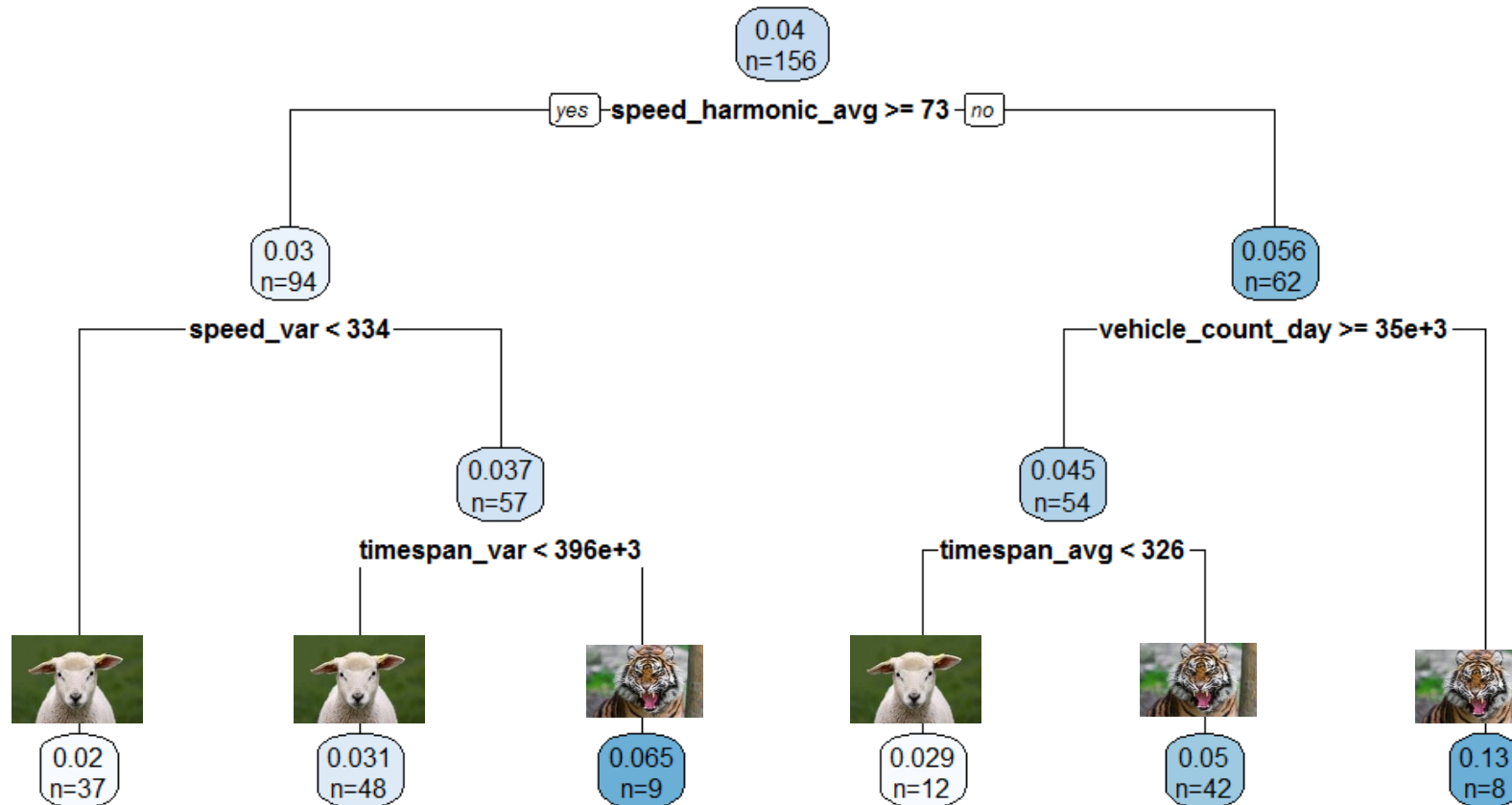
2. Classificatie & Clustering

Welke kenmerken onderscheiden veilige van onveilige LocPosts ?

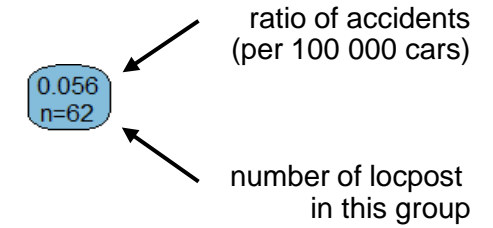


Regression tree – all data

Regression tree for all locposts movement characteristics,
entire dataset



Key:



Legende

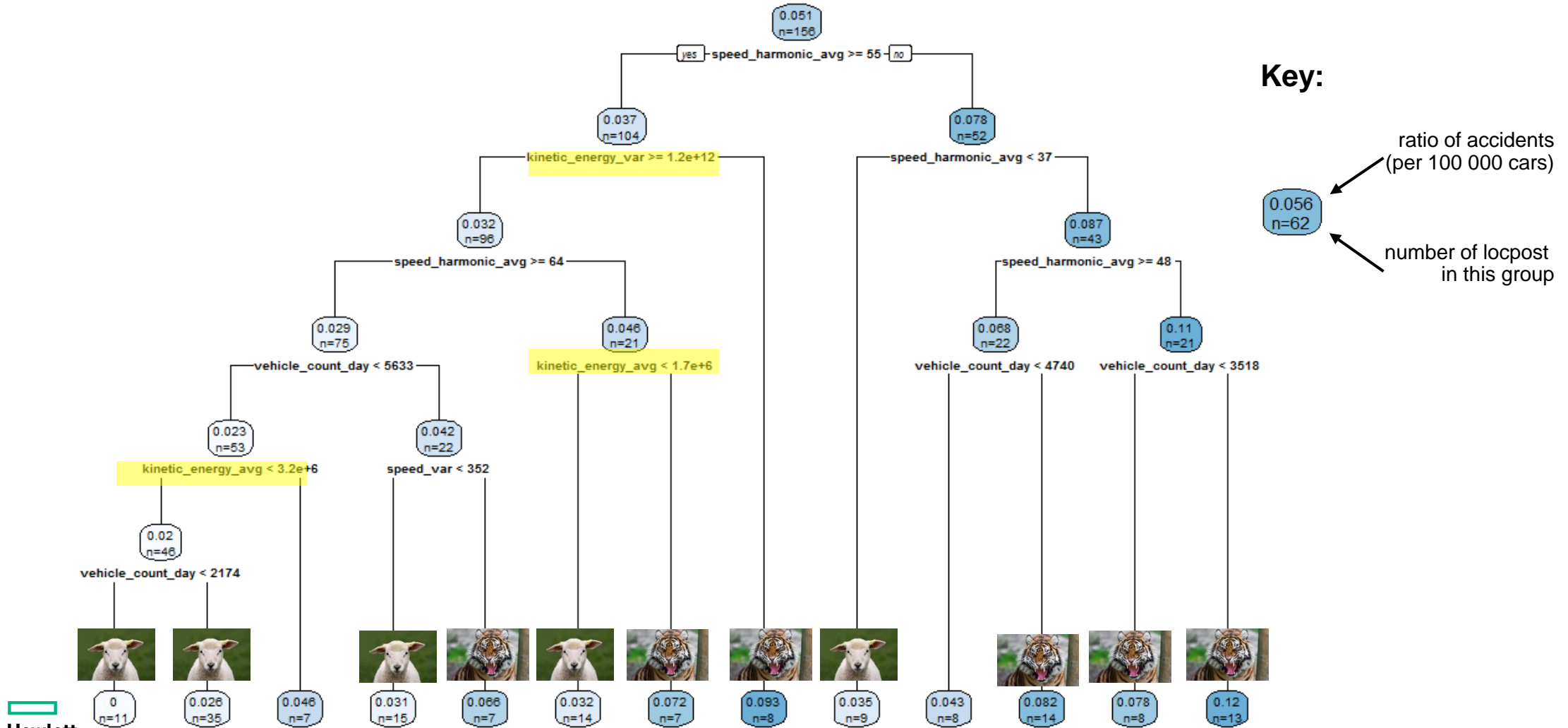
- Datamart op basis van Individuele doortochten
- Alle **156** locposten
- Geen tijdsdimensie
- Ongevallen binnen 3 KM na Locpost

Variaties

- Alle dagen / Werkdagen
- Werkuren
- Goed Weer / Slecht Weer

Regression tree

Variation: Workdays & 6am – 7pm & Bad Weather

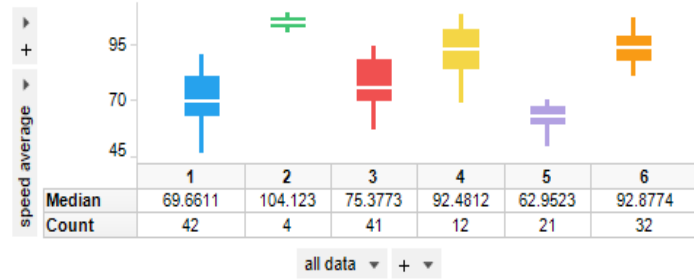


Kunnen we LocPosts groeperen op basis van verkeersvariabelen?

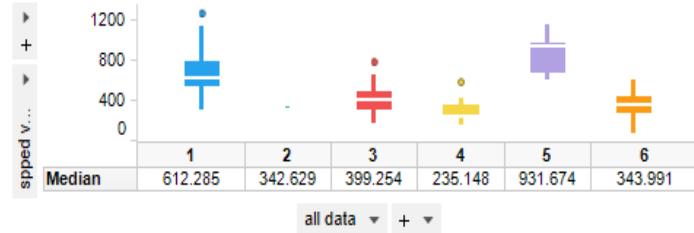


En zijn bepaalde groepen (clusters) veiliger of onveiliger dan anderen ?

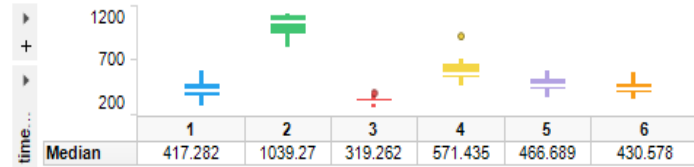
Speed average



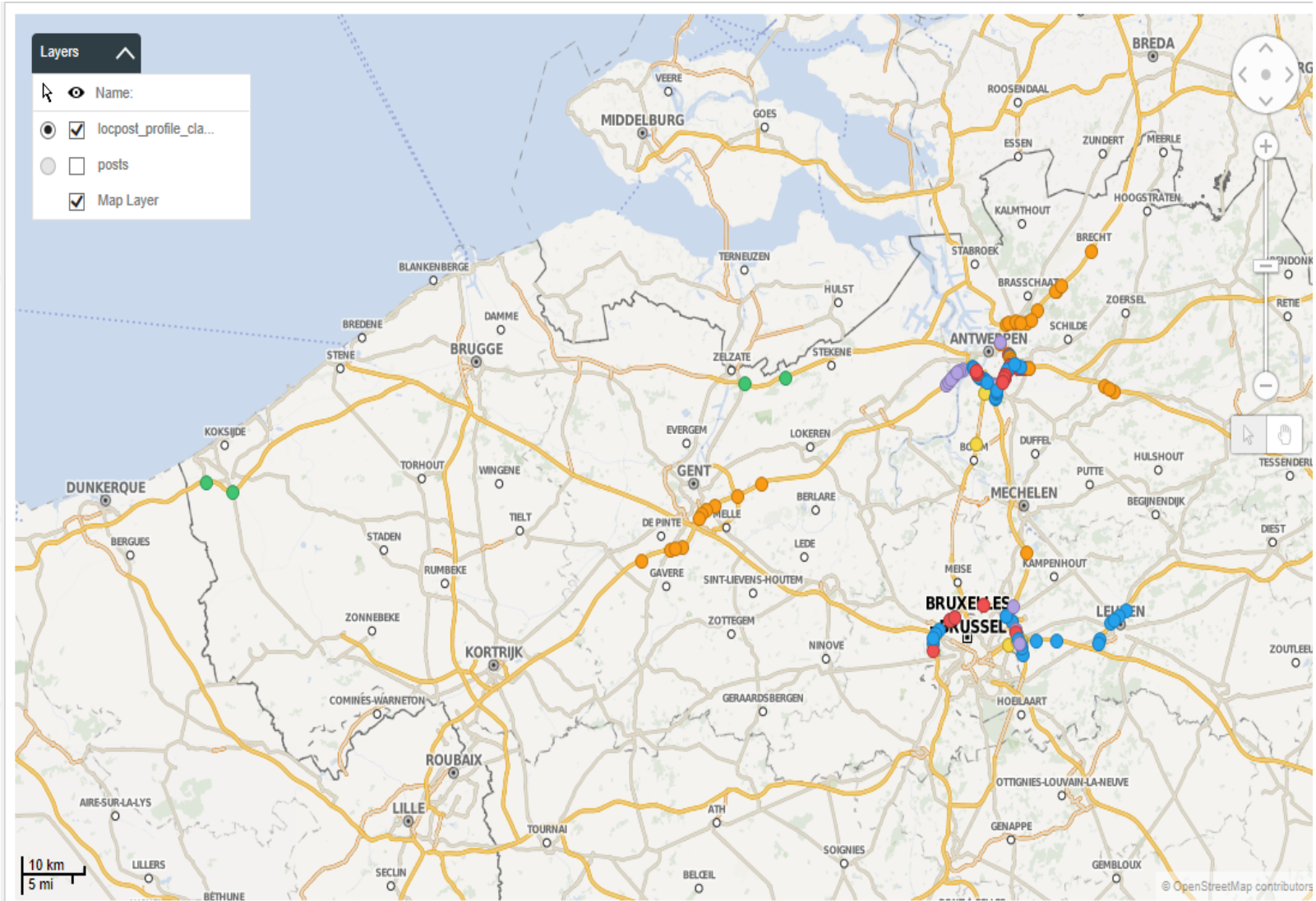
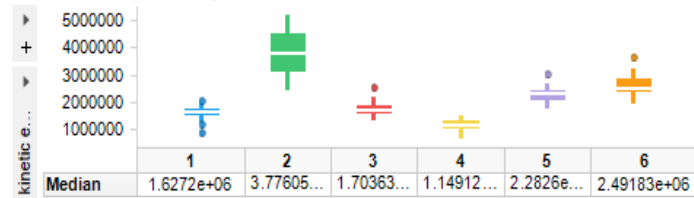
Speed variance



Timespan

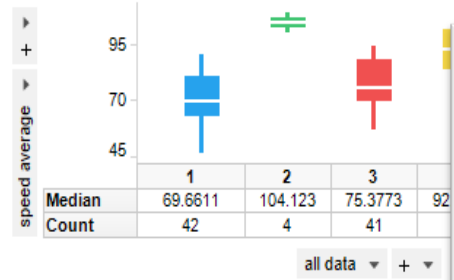


Kinetic energy average

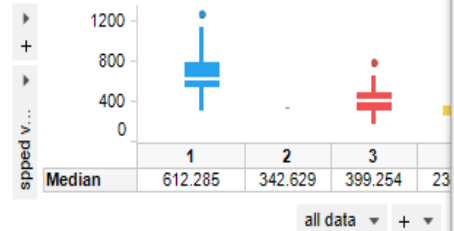


Groepering op basis van gelijke gemiddelde verkeersvariabelen

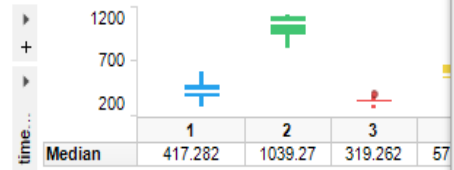
Speed average



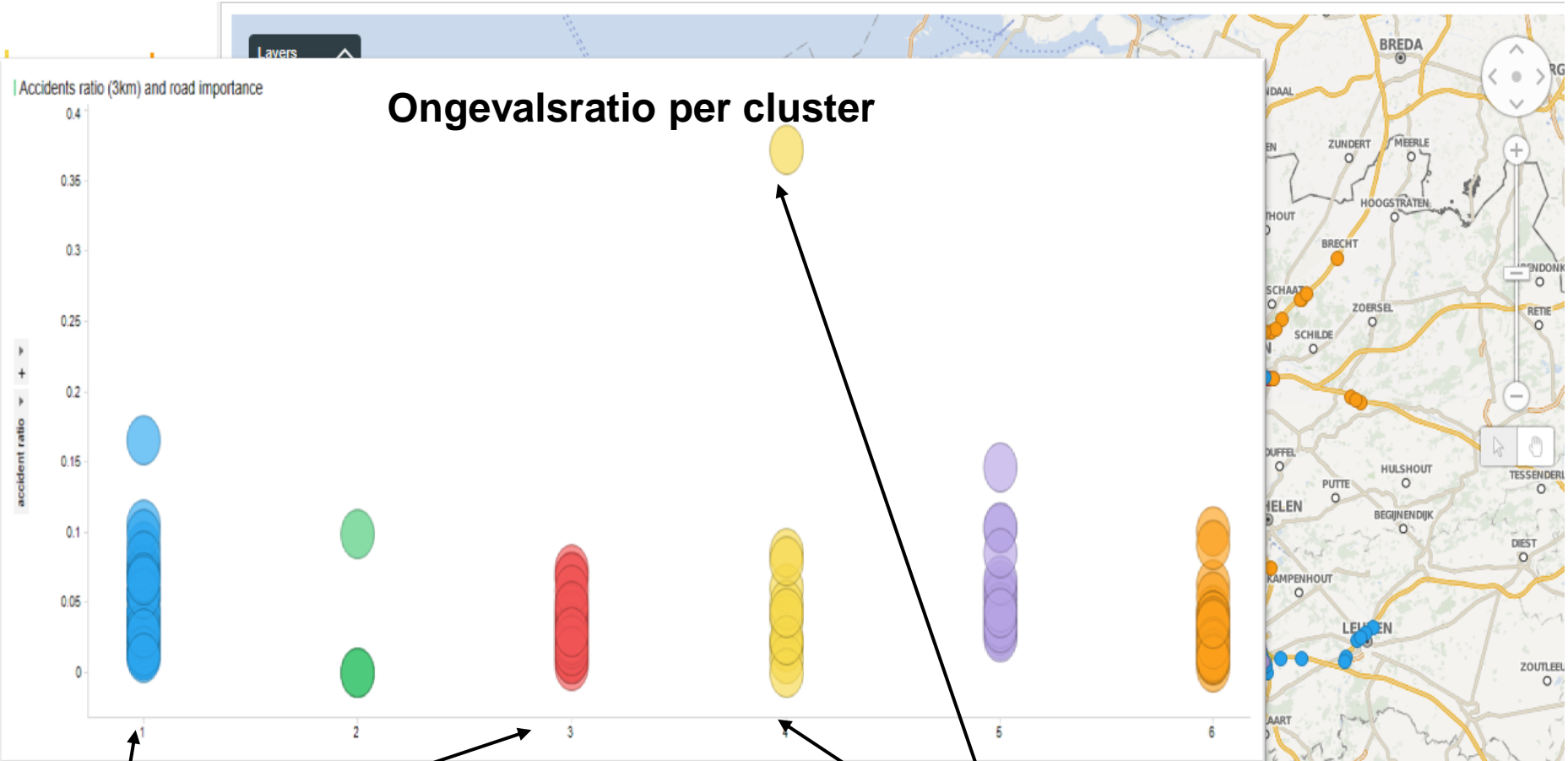
Speed variance



Timespan



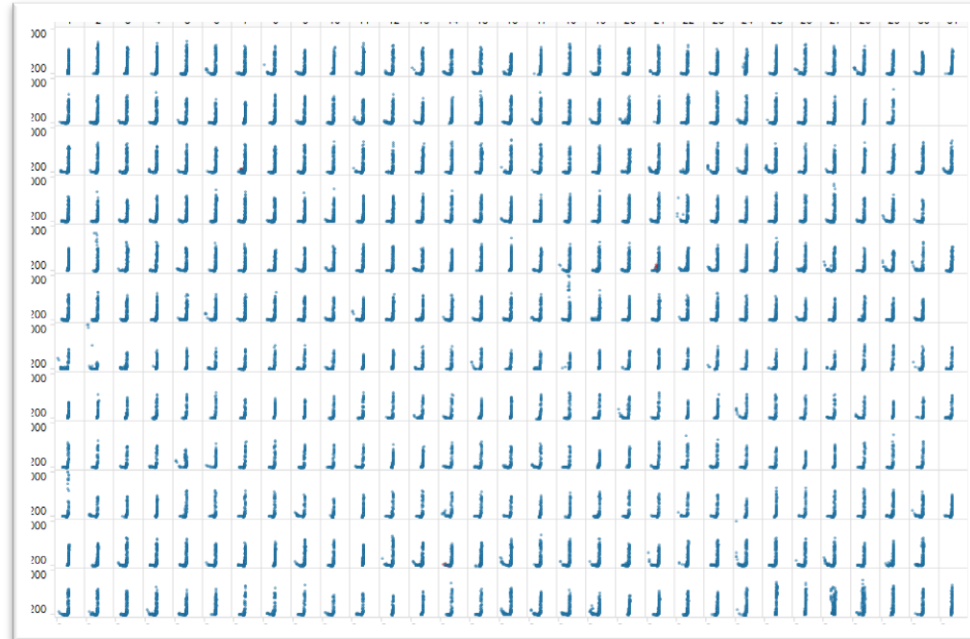
Kinetic energy average



Geen significant verschil tussen clusters

Wel duidelijke spreiding binnen de clusters

Kunnen we LocPosts groeperen op basis van de V-Shape Diagramma's



En zijn bepaalde groepen dan veiliger of onveiliger dan anderen ?

V-Shape Clustering

Aanpak

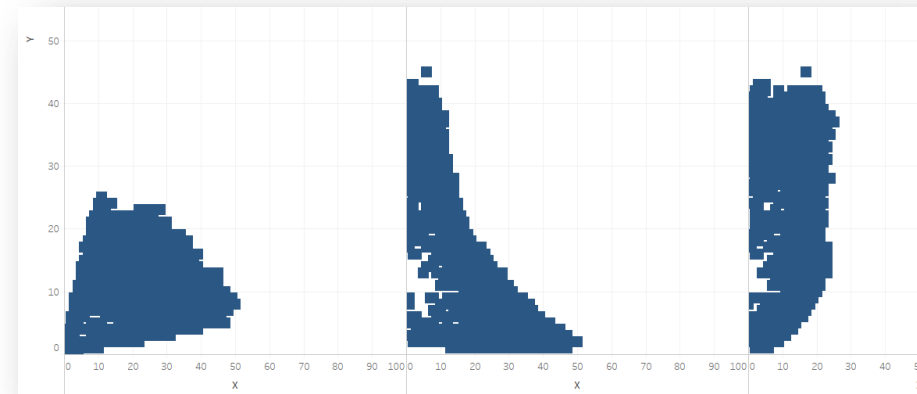
1. Berekenen van alle V-Shapes voor alle locposts

meetpunt_id	meettijd	volgtijd	snelheid_kmph	elek_lengte_dm
460	3/17/2015 12:00:46 AM	3001	91	88
460	3/17/2015 12:00:55 AM	902	98	47
460	3/17/2015 12:01:23 AM	2805	134	45
460	3/17/2015 12:03:24 AM	3001	87	166
460	3/17/2015 12:03:50 AM	2590	108	47
460	3/17/2015 12:04:43 AM	3001	120	44
460	3/17/2015 12:05:33 AM	3001	123	53
460	3/17/2015 12:05:50 AM	1766	99	42
460	3/17/2015 12:05:59 AM	840	95	43
460	3/17/2015 12:06:04 AM	542	125	47
460	3/17/2015 12:06:14 AM	989	148	45
460	3/17/2015 12:06:19 AM	527	110	50
460	3/17/2015 12:07:42 AM	3001	105	47

Individuele Doortochten



V-Shapes



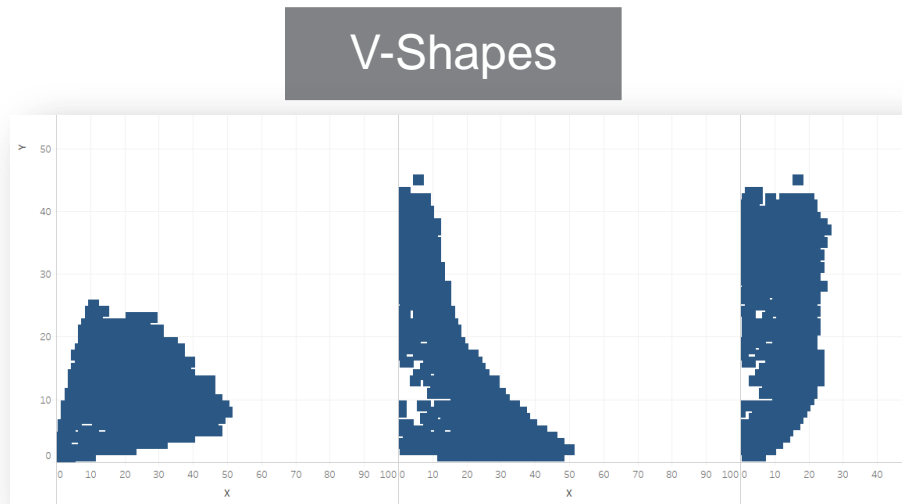
Intensiteit (Auto's per Uur)
Snelheid (Km/Uur)
Dichtheid (Auto's per KM)

(intensiteit = dichtheid × snelheid)

V-Shape Clustering

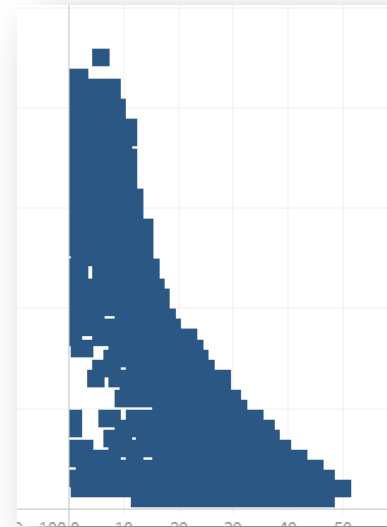
Aanpak

1. Berekenen van alle V-Shapes voor alle locposts
2. Bepalen “dimensies” van all V-Shapes



Intensiteit – Snelheid – Densiteit
(3 X 2)

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y)$$

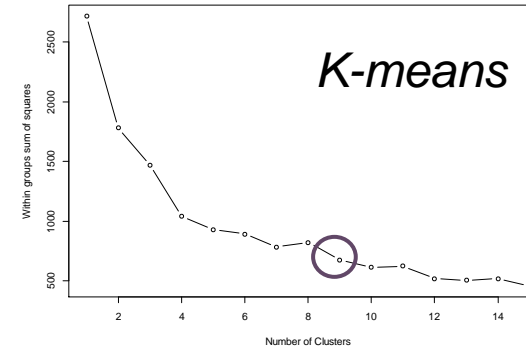


A vector of central moments for the
three V-shapes for one locpost
Central moments of order up to 2

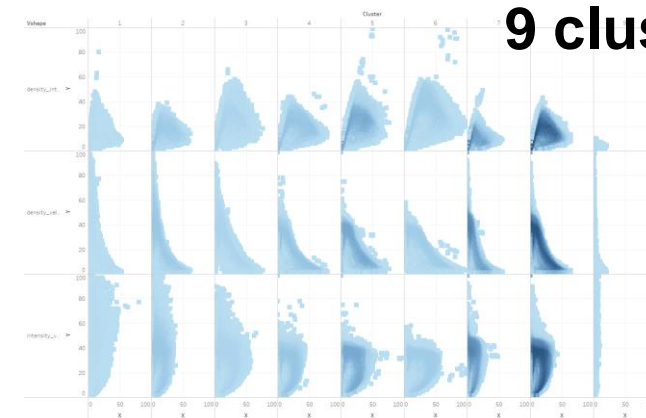
V-Shape Clustering

Aanpak

1. Berekenen van alle V-Shapes voor alle locposts
2. Bepalen dimensies van all V-Shapes
3. Groeperen (Clustering) van V-Shapes

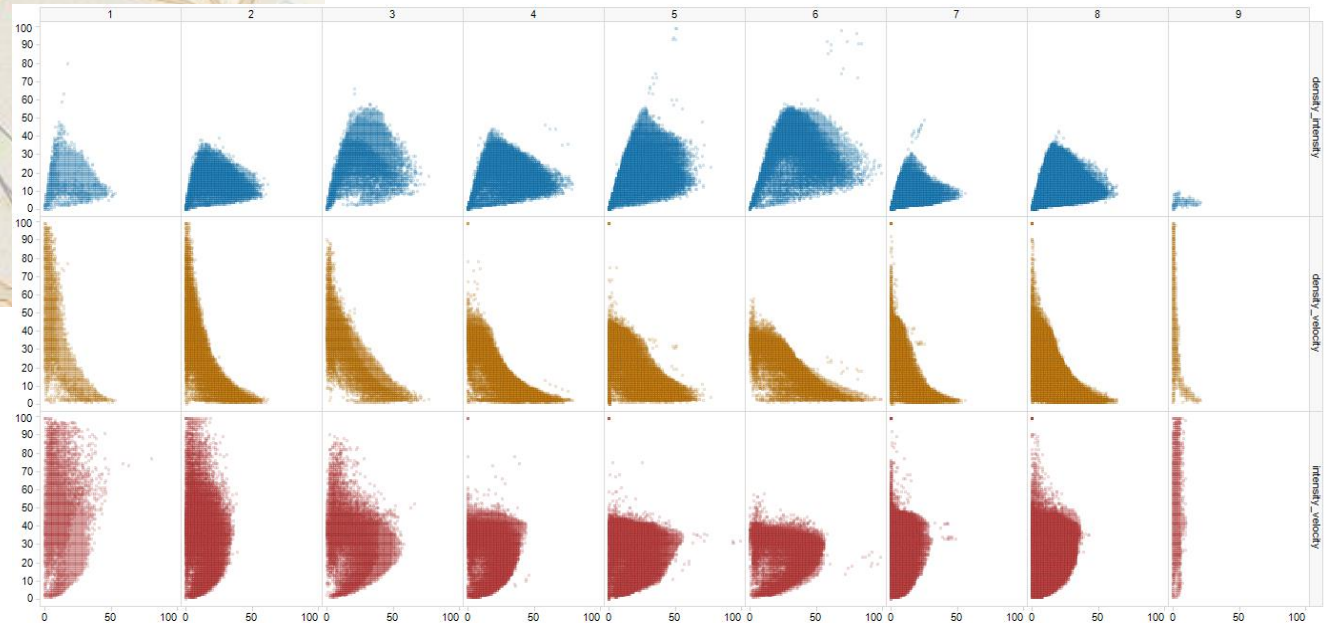
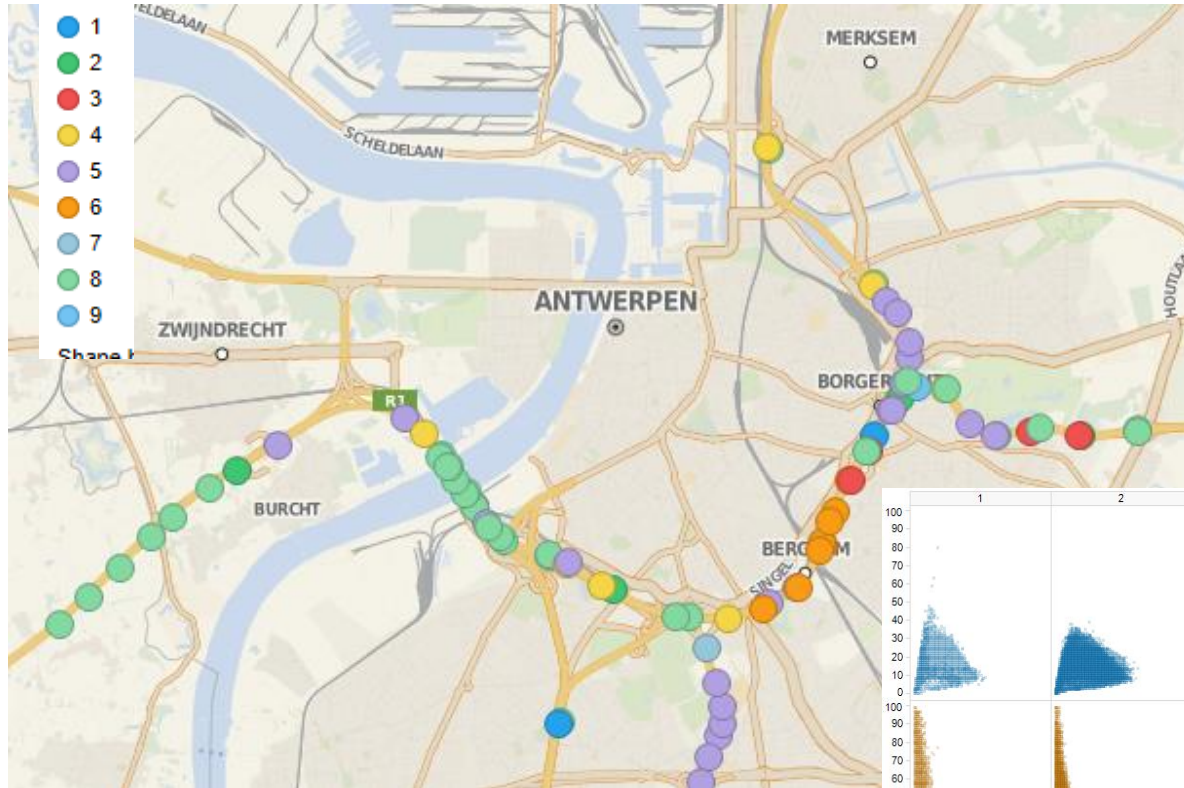


Number of clusters selection based on the within-group sum of squares



9 clusters

De gevonden clusters



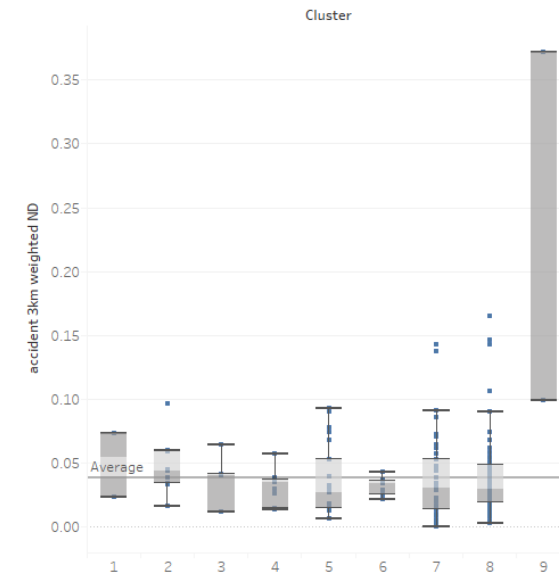
V-Shape Clustering

Aanpak

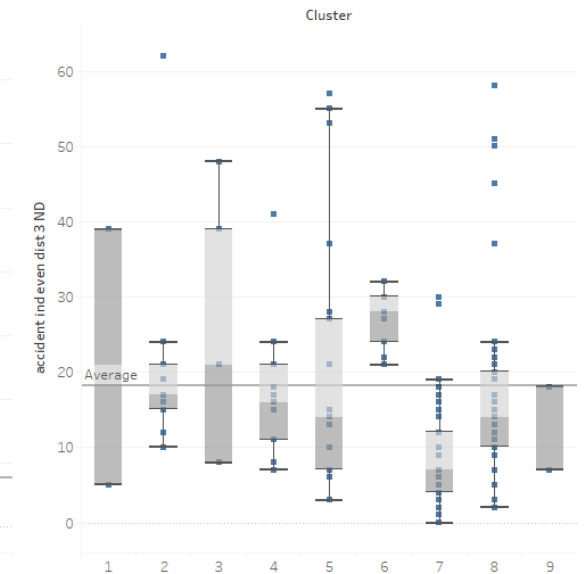
1. Berekenen van alle V-Shapes voor alle locposts
2. Bepalen dimensies van all V-Shapes
3. Groeperen van V-Shapes
4. Ongevallsstatistieken toepassen

Geen significante verschillen tussen de gevonden clusters.

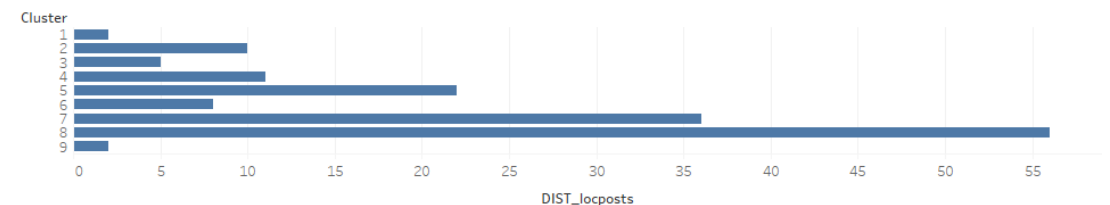
Weighted accident ratio



Absolute number of accidents



Locposts per cluster



Deel 2 – Classificatie en Clustering

Besluiten & Relevantie

- **Ongevalsclustering** gaf aan dat

- ‘primaire’ verkeersvariabelen zijn bepalende factoren en komen steeds terug – maar met wisselend belang.

- Hogere snelheid vaak indicatie van veilige situatie.

- Kinetische Energie is belangrijke nieuwe variabele. $E_k = \frac{1}{2}mv^2$

- Opdeling op basis van “**gemiddelde**” verkeerssituatie en/of vorm van de V-Shape

- Geen significante verschillen tussen gevonden clusters

- Wel spreiding binnen cluster – mogelijk relevant om te onderzoeken

Opmerking: Analyses hebben enkel gelopen op de geselecteerde ‘high-risk’ locposts waarvoor we data hadden. Mogelijk andere resultaten indien oefening ook voor de andere locposts gebruikt.

Wat bepaalt of een Locpost veilig is of onveilig?

Zijn bepaalde Locposts veiliger dan andere?

Wat zijn de bepalende factoren?

Kunnen we ongevallen voorspellen?

Welke informatie hebben we daarvoor nodig?

Hoe doen we die voorspelling het best?

Kunnen we modellen veralgemenen?

Zijn er patronen te herkennen die ongevallen veroorzaken?

Zijn er andere dimensies bepalend?



3. Via modelleringen naar een Veiligheidsindicator

Kunnen we een veiligheidsindicatie bepalen ?

$$-3,89*A+0,14*B-0,02*C-23*.13*D-2*E-0,07*F-0,04*G+1,22*H+0,06*I$$

met

A = Standaard Deviatie van de gemiddelde gemeten snelheid over de laatste 5 minuten gemeten in locpost voor het ongeval

B = Gemiddelde Minimum gemeten tijdspanne tussen 2 voertuigen voor de laatste 5 minuten in de locpost na het ongeval

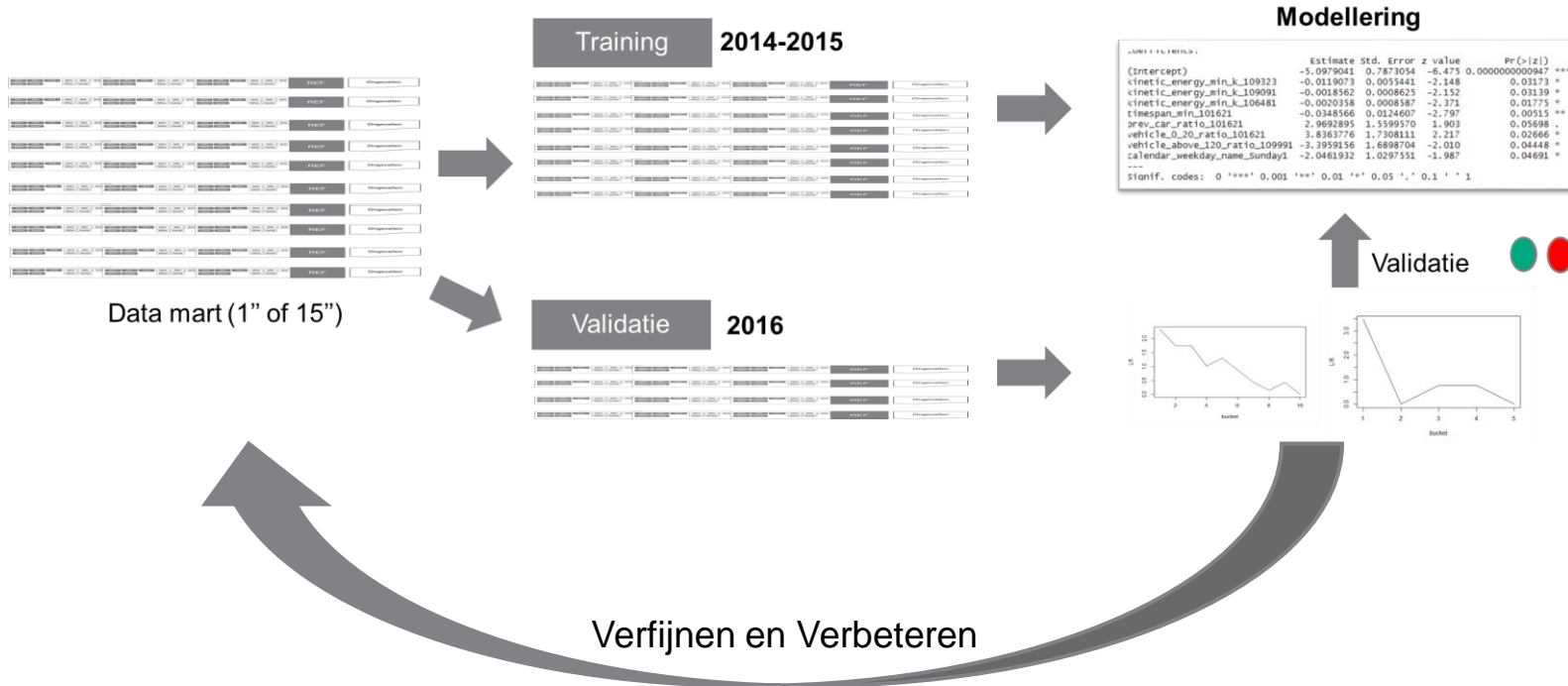
C = ...



Veiligheids Modelleringen

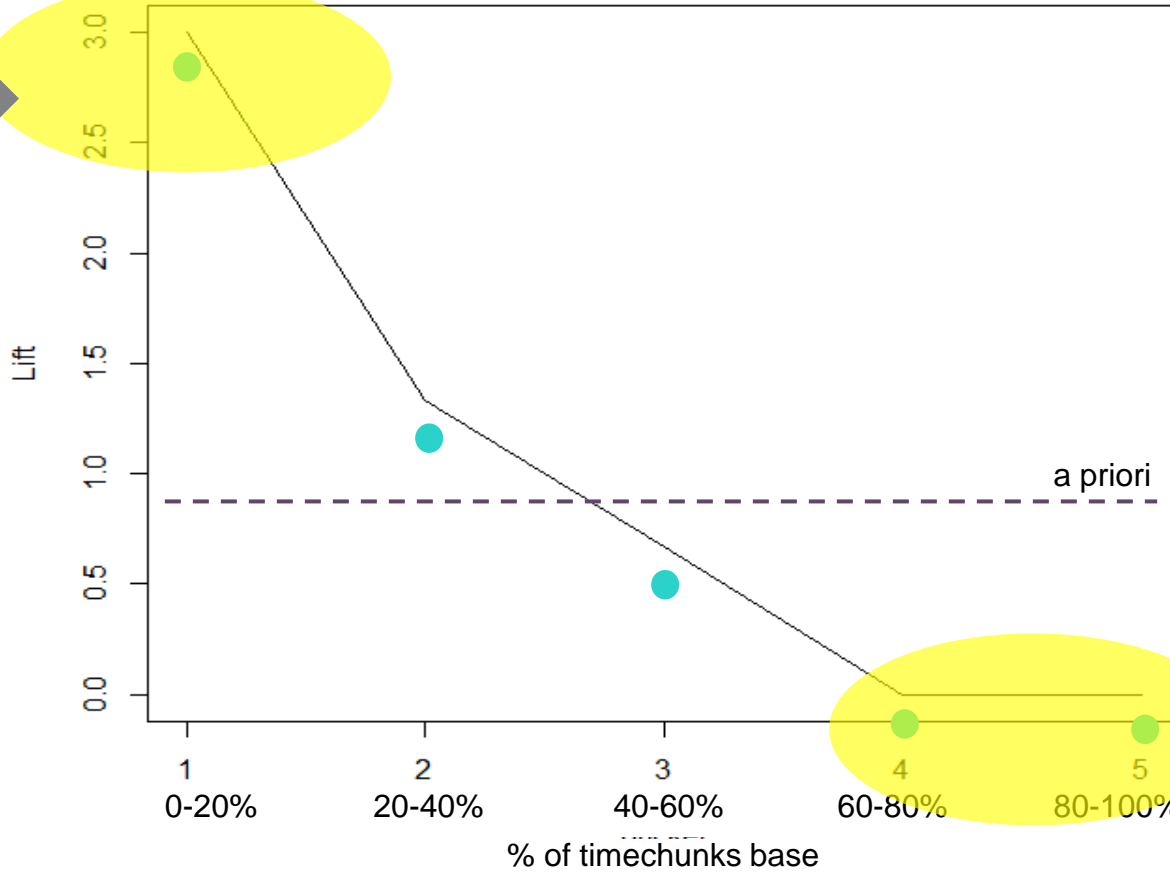
Aanpak

1. Voorbereiding Data Mart
2. Modelleren en Evaluatie
3. Herhaal 1 en 2



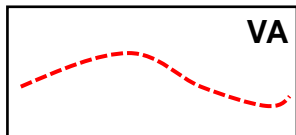
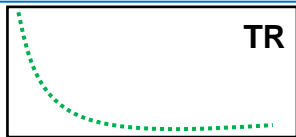
Een "lift chart" als evaluatie instrument

Voor de waarnemingen met de hoogste score is het model gemiddeld 3x beter dan een willekeurig patroon.

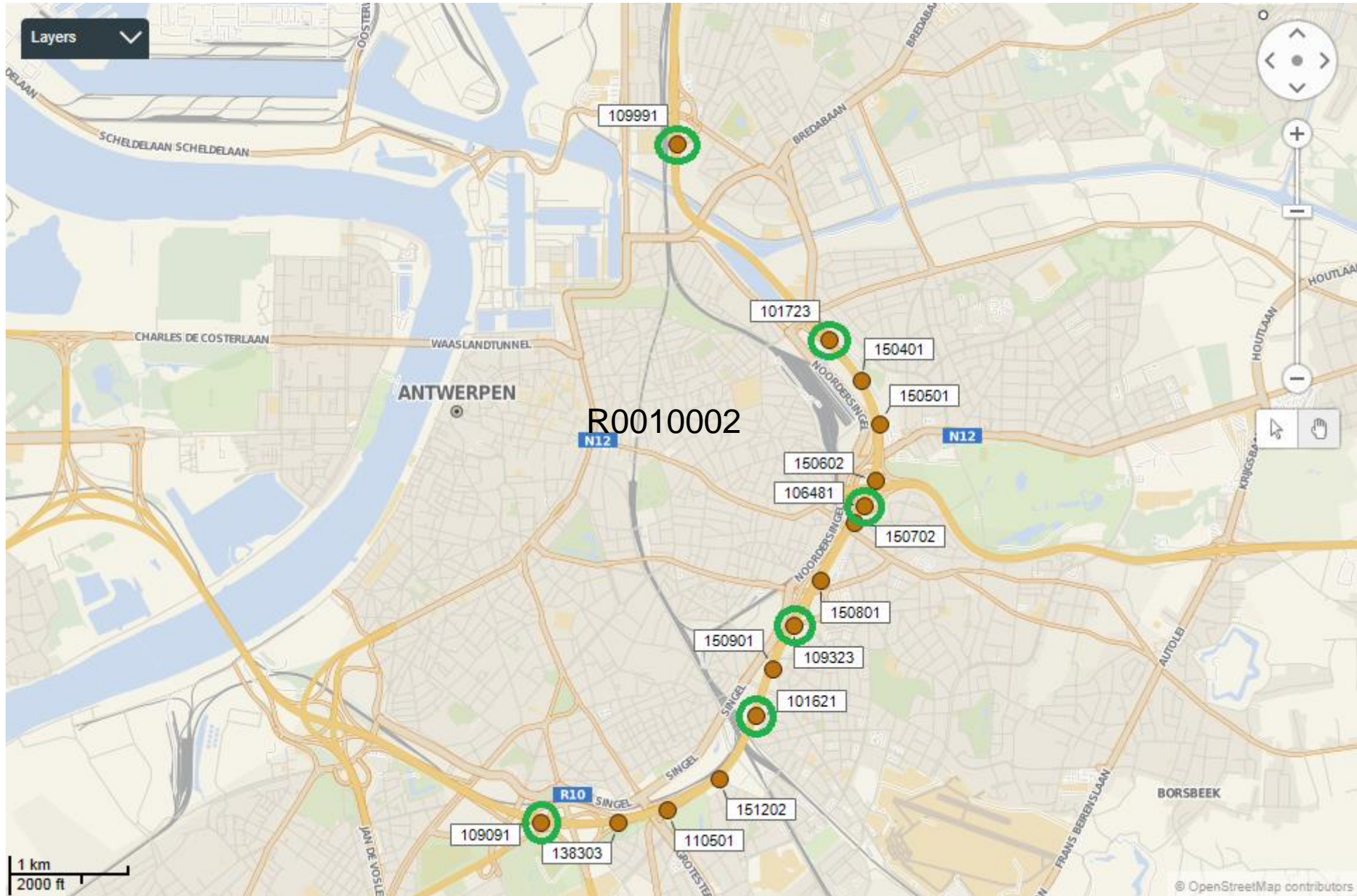


Voorwaarde Goed Model

Lift Chart Training
gelijk aan
Lift Chart Validatie model

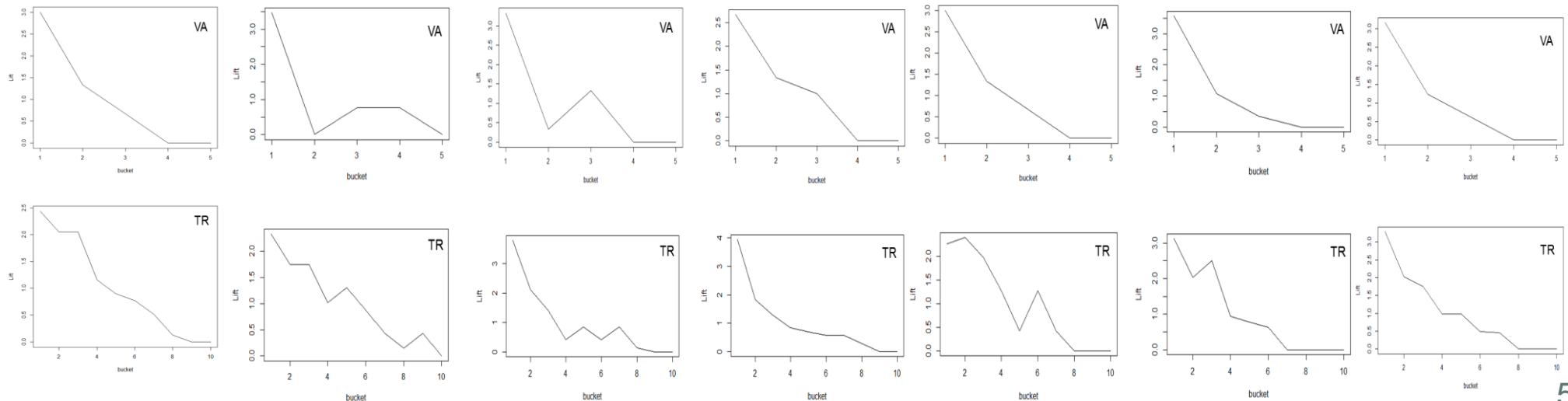


Voor 40% van alle waarnemingen, voorspelt het model zo goed als geen ongeval.



Modellering En Evaluatie – Antwerpse Ring (R0010002)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Tijdspanne	15 Minuten	15 Minuten	15 Minuten	15 Minuten	15 Minuten	15 Minuten	1 Minuut
Data	Alle Uren	Werkuren	Alle Uren	Alle Uren	Alle Uren	Alle Uren	Alle Uren
Ongevallen (TR/VA)	Alle 71/15	Alle 71/15	Alle 71/15	Alle 71/15	Alle 71/15	z. Alcohol 64/14	z. Alcohol 64/14
Locpost	1 Locpost 109091	1 Locpost 109091	6 locposts	6 locposts	6 locposts	6 locposts	109091 109991
Extra				Neuraal Netwerk Ekin	Neuraal Netwerk Tijdspanne	Ratio's Ekin Tijdspanne	Variance Variabelen



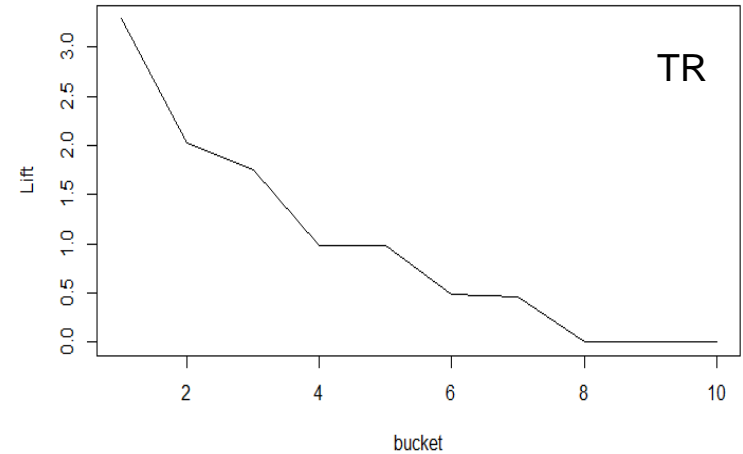
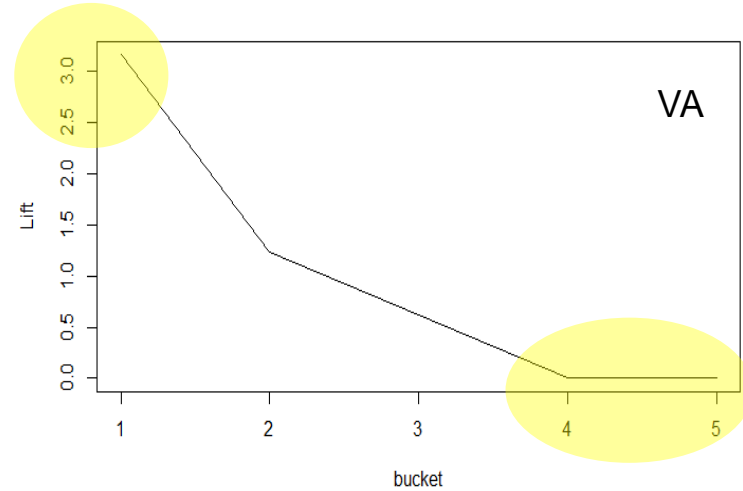
Model #7

- Datamart: 1-minuut
- Dag en Nacht
- Data van eerste en laatste locpost van de ring
- Enkel ongevallen zonder alcohol (64 in TR, 14 in VA)
- Extra Variance variabelen toegevoegd

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.889594	0.410626	-9.472	< 0.0000000000000002	***
avg_5min_speed_stddev_first	0.145422	0.028707	5.066	0.000000406897255009	***
avg_5min_timespan_min_last	-0.020530	0.003098	-6.627	0.000000000034228399	***
avg_5min_vehicle_above_120_ratio_first	-23.124283	2.843278	-8.133	0.000000000000000419	***
calendar_weekday_name_Sunday1	-2.006470	0.460028	-4.362	0.000012909913725001	***
momentum_sum_M_last	-0.069825	0.013604	-5.133	0.000000285423973089	***
stddev_5min_speed_min_last	-0.035590	0.013055	-2.726	0.00641	**
vehicle_0_20_ratio_first	1.222071	0.266986	4.577	0.000004710325202168	***
vehicle_61_80_count_last	0.063777	0.009078	7.026	0.000000000002132284	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



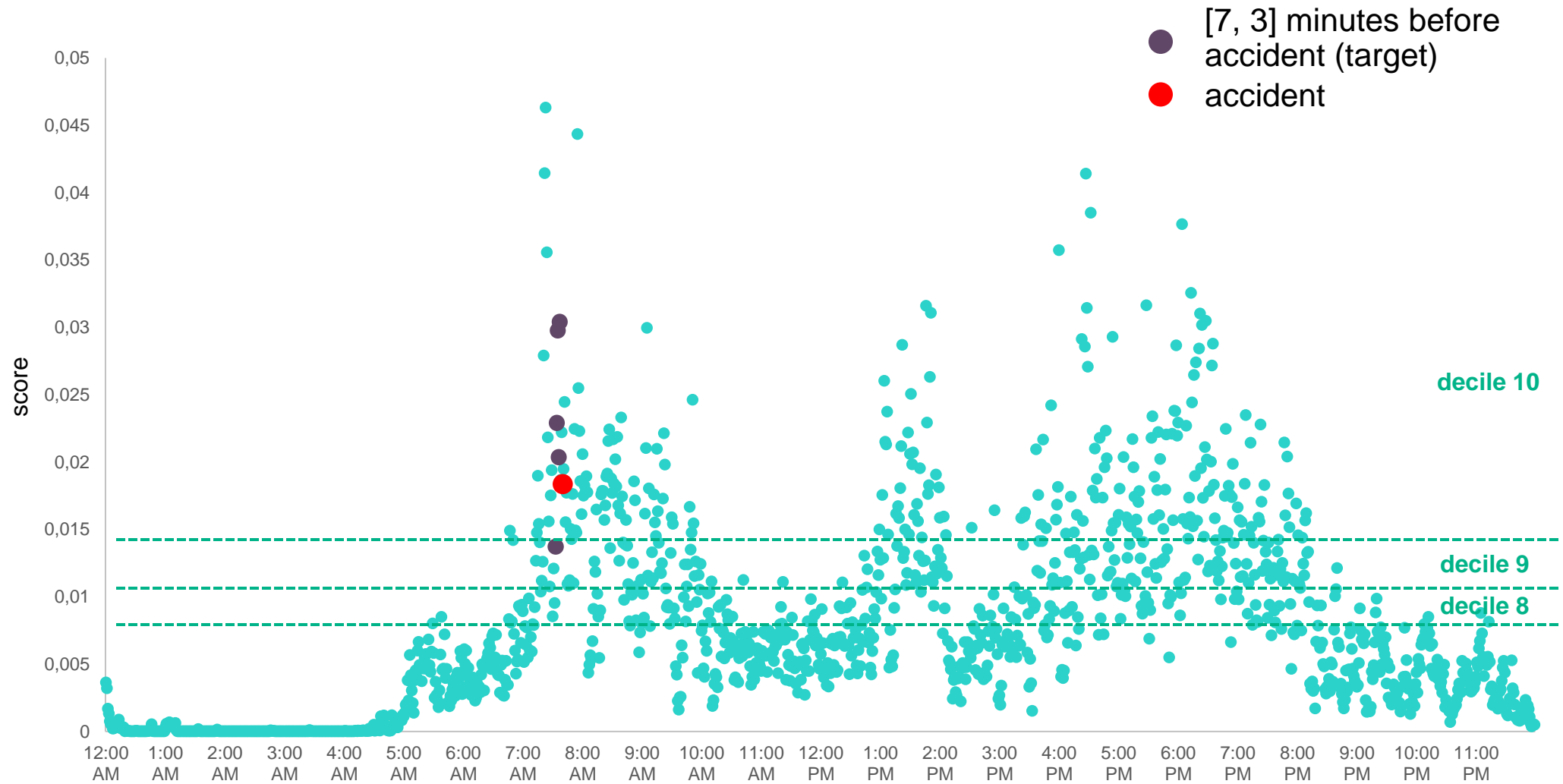
Model #7 – performance

Score decile	Score cut-off	Number of timechunks	Accidents	% of total accidents
10	0.0143817938611167	118085	29	33.7%
9	0.0104107194697205	118085	17	19.8%
8	0.00800501850636187	118085	11	12.8%
7	0.00611952288031388	118085	13	15.1%
6	0.00439352622243583	118085	10	11.6%
5	0.00255935166710759	118085	3	3.5%
4	0.00095185350680663	118085	3	3.5%
3	0.000159948008002032	118085	0	2.1%
2	0.0000000292151775831846	118085	0	1.0%
1	-	118085	0	0.0%

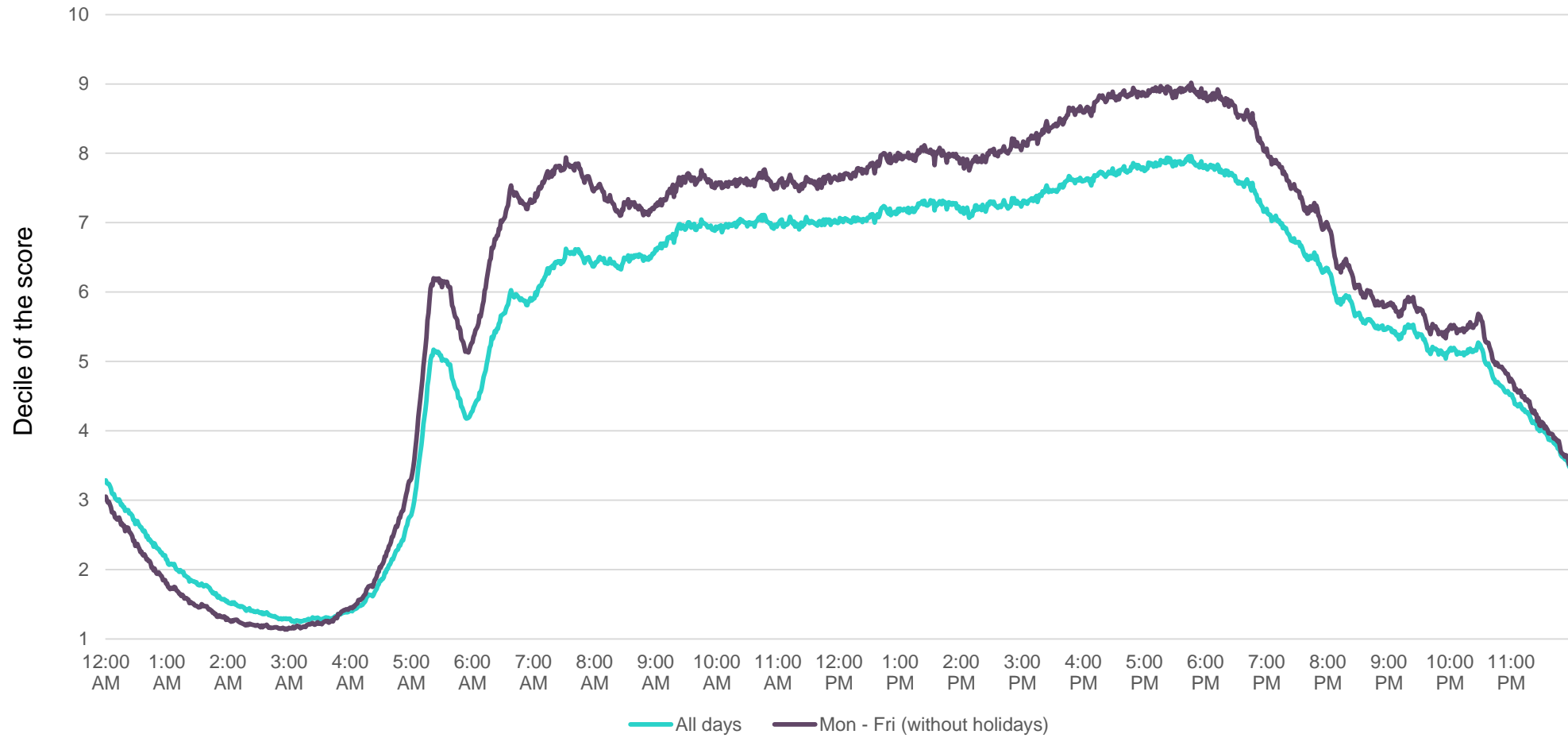
1 timechunk = 1 minuut

Model kan aangeven in welke 10% van de tijd zeker 1 ongeval op 3 (33,7%) gebeurt

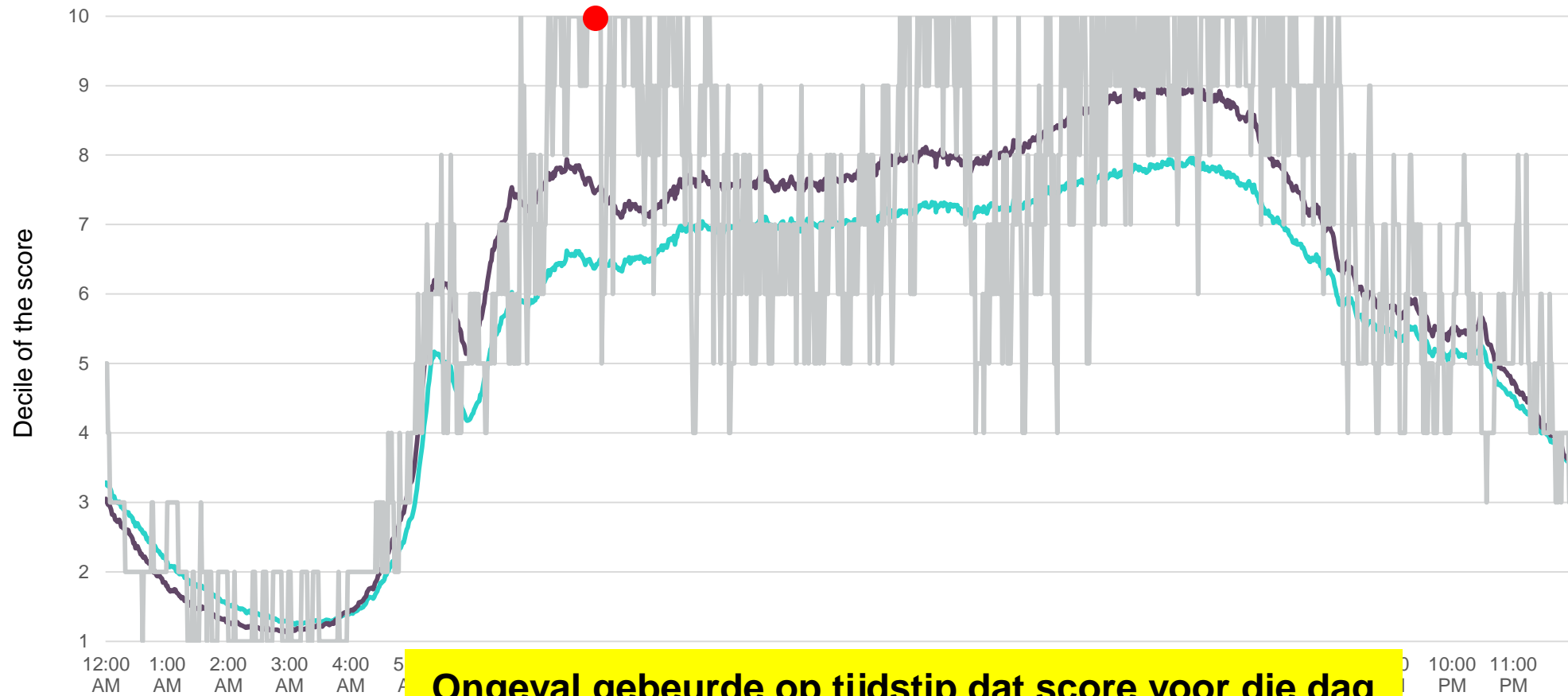
Model #7 – sample day (2015-10-22)



Model #7 – Gemiddelde score per minuut

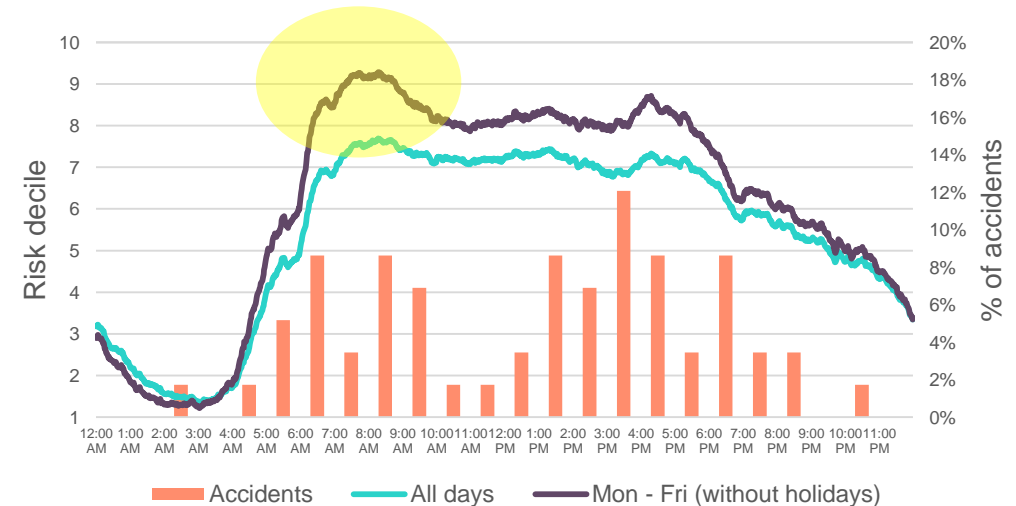
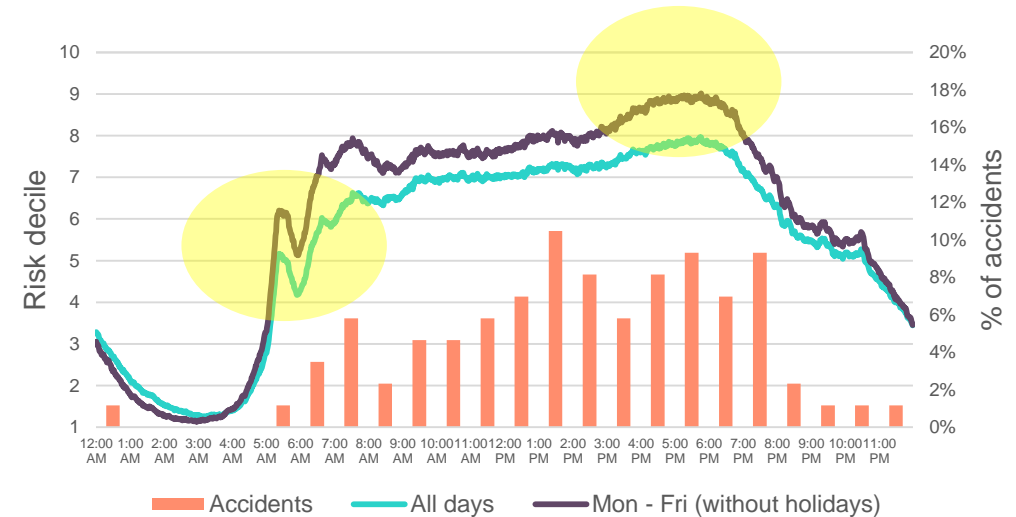
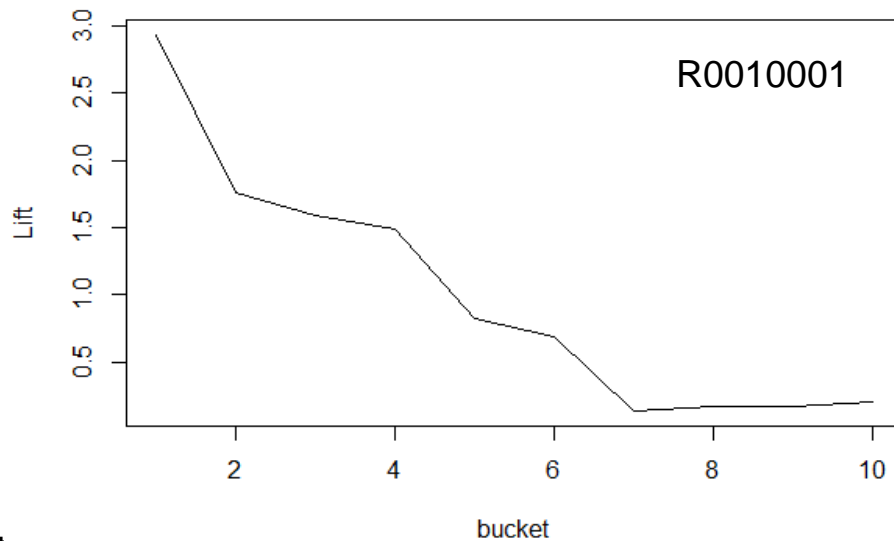
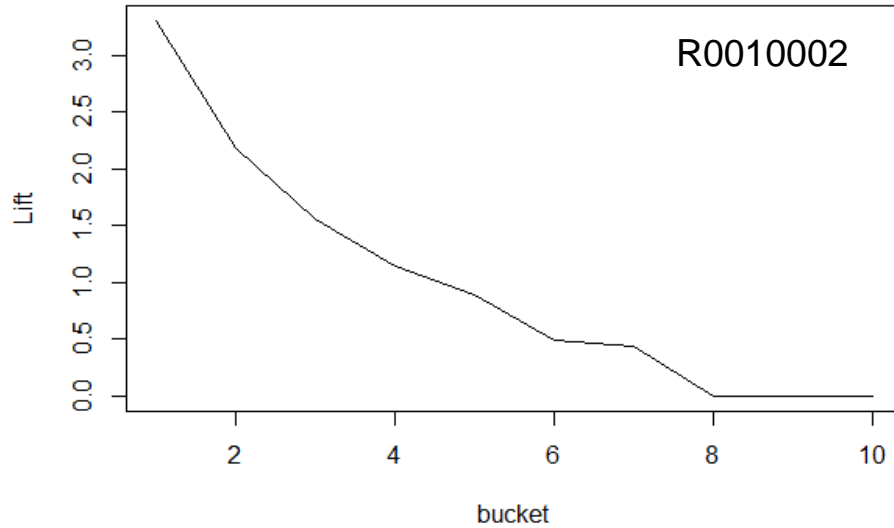


Model #7 – Gemiddelde score en effectieve score

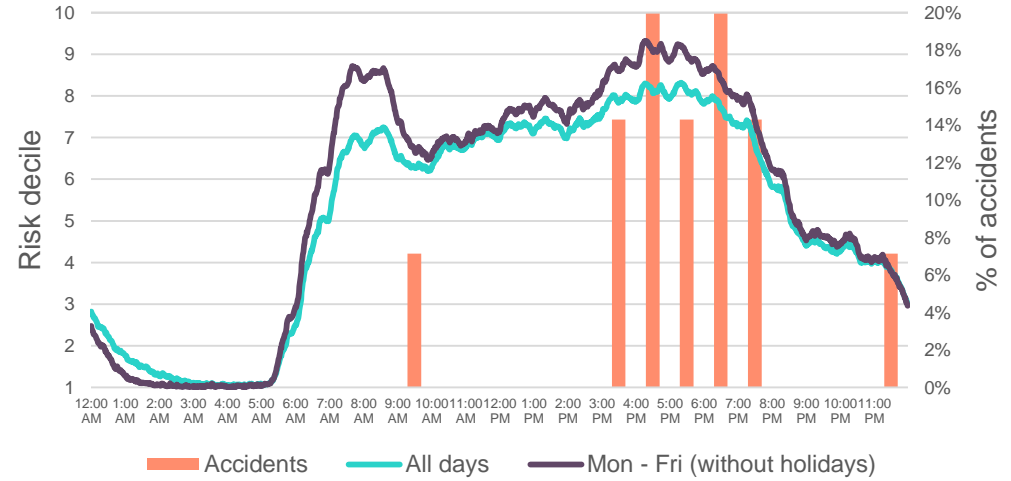
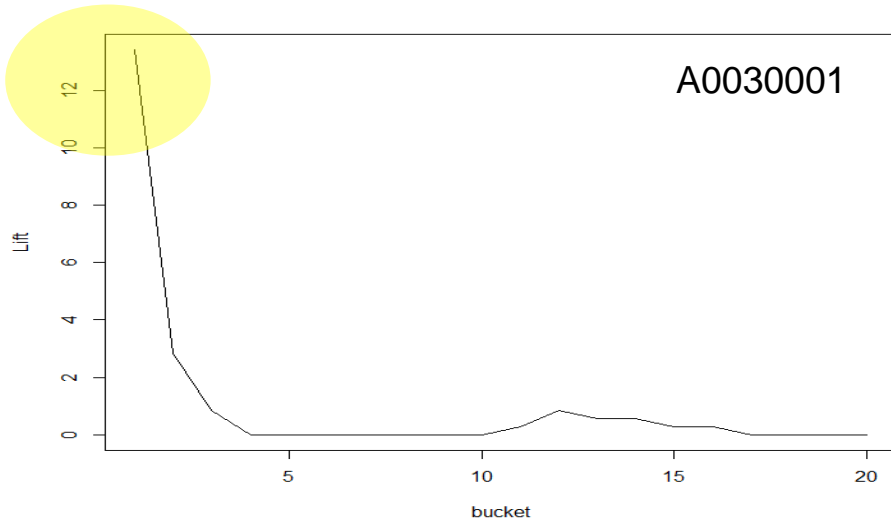


Ongeval gebeurde op tijdstip dat score voor die dag veel hoger is (10) dan gewoonlijk (7,5)

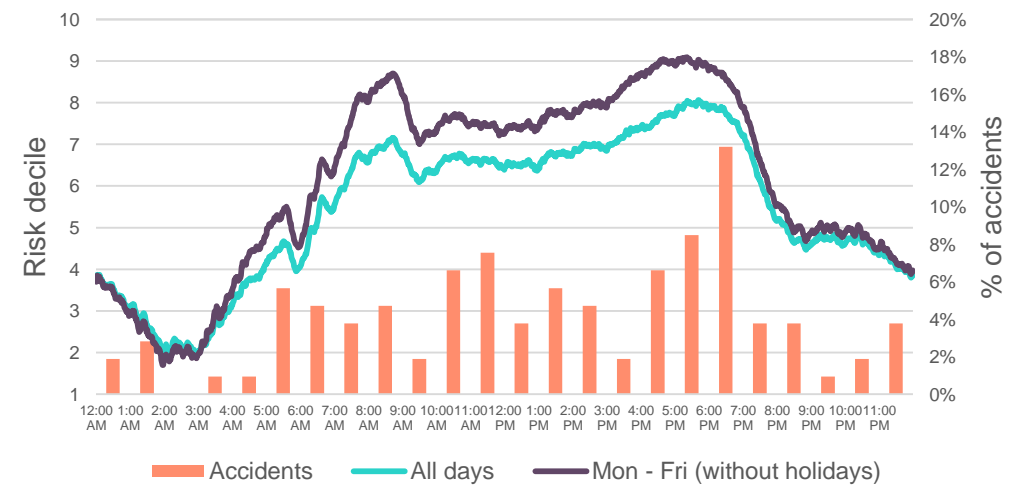
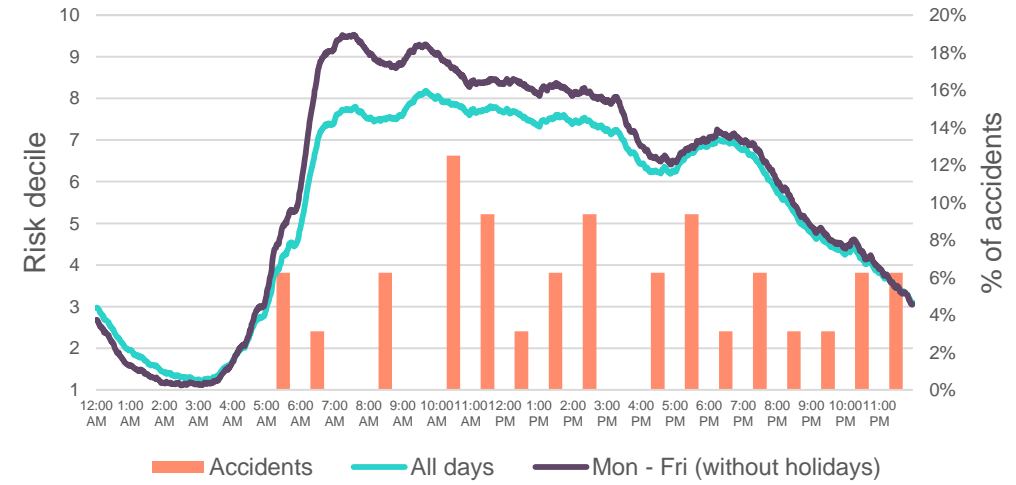
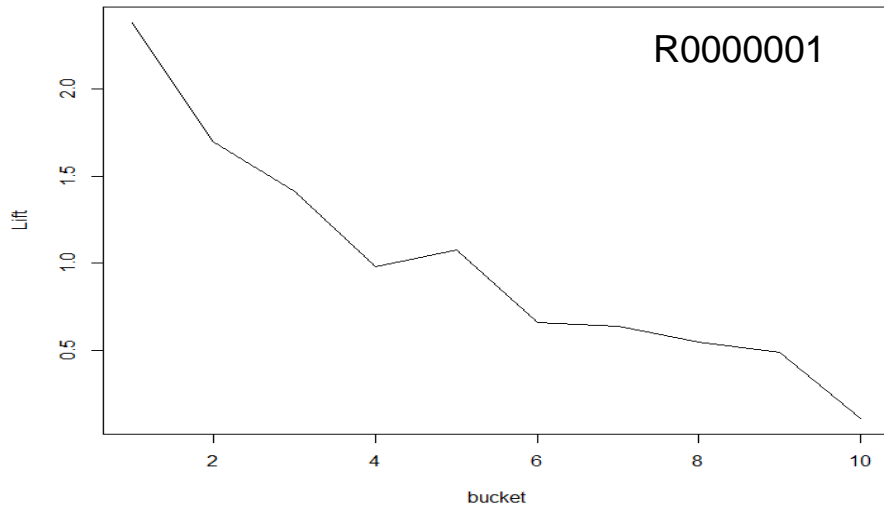
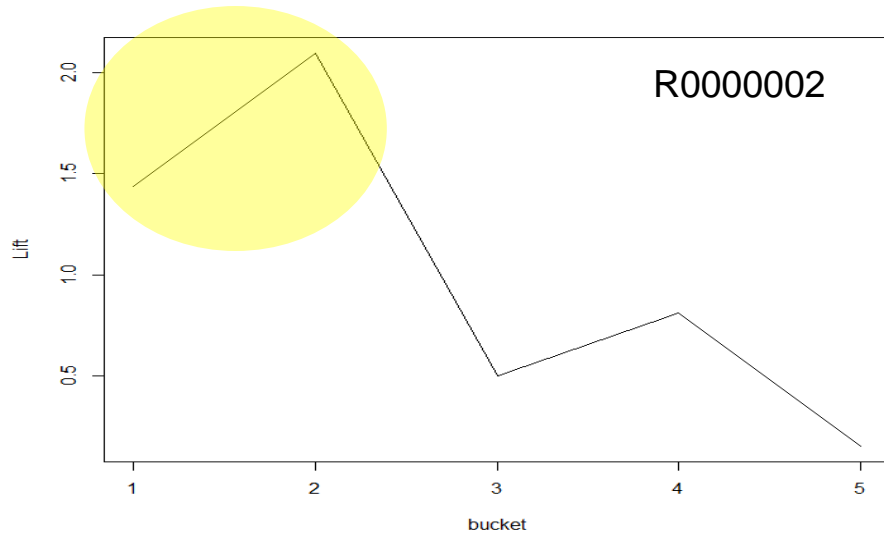
Model #7 – Validatie model op andere wegsegmenten



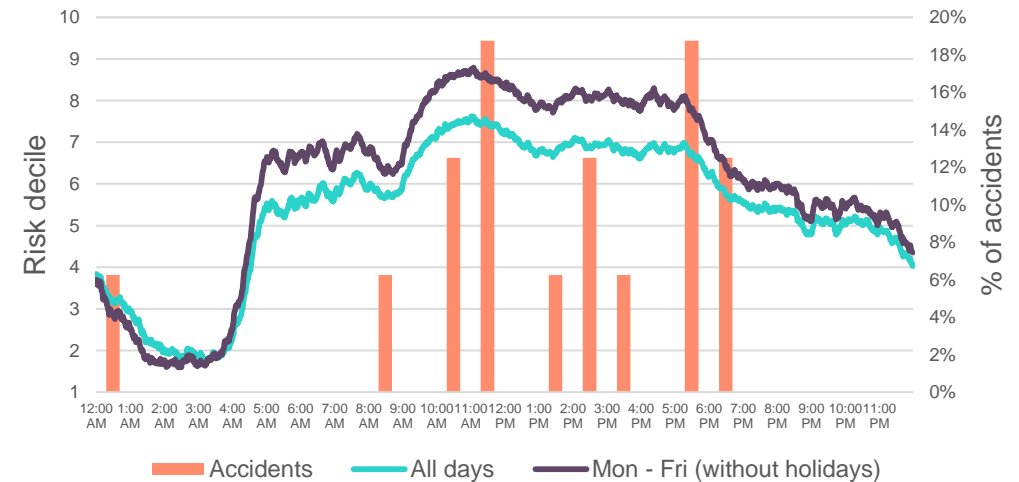
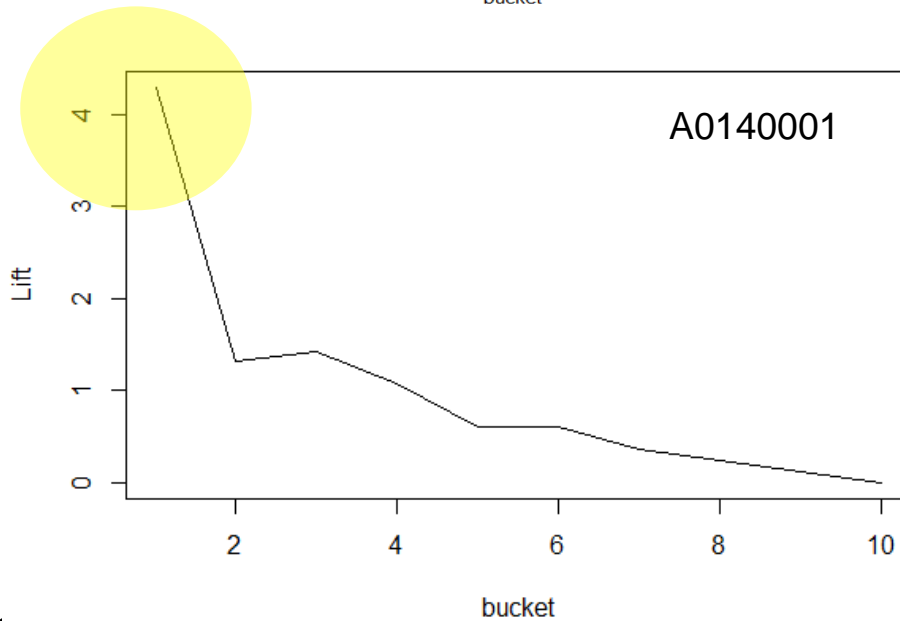
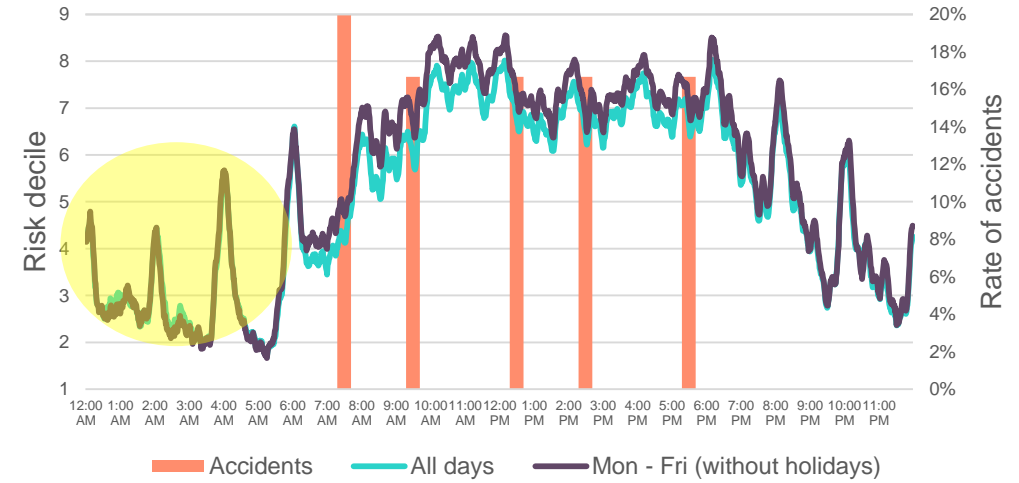
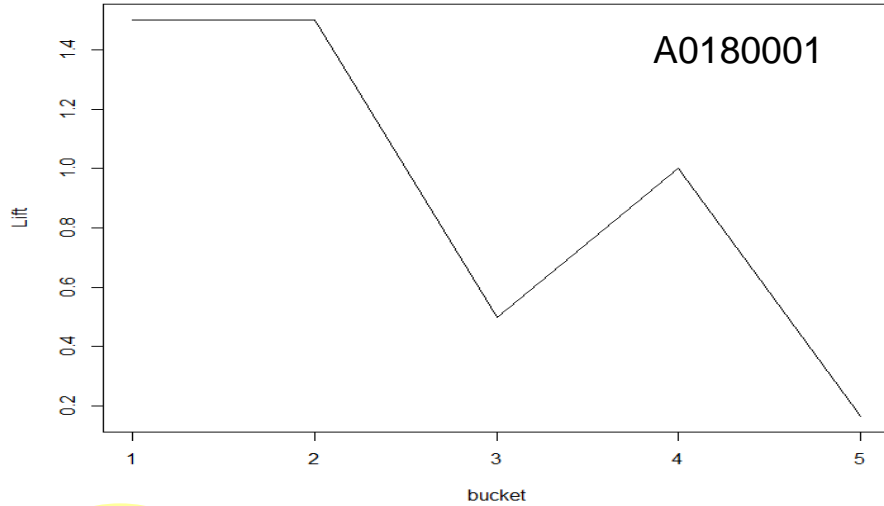
Model #7 – Validatie model op andere wegsegmenten



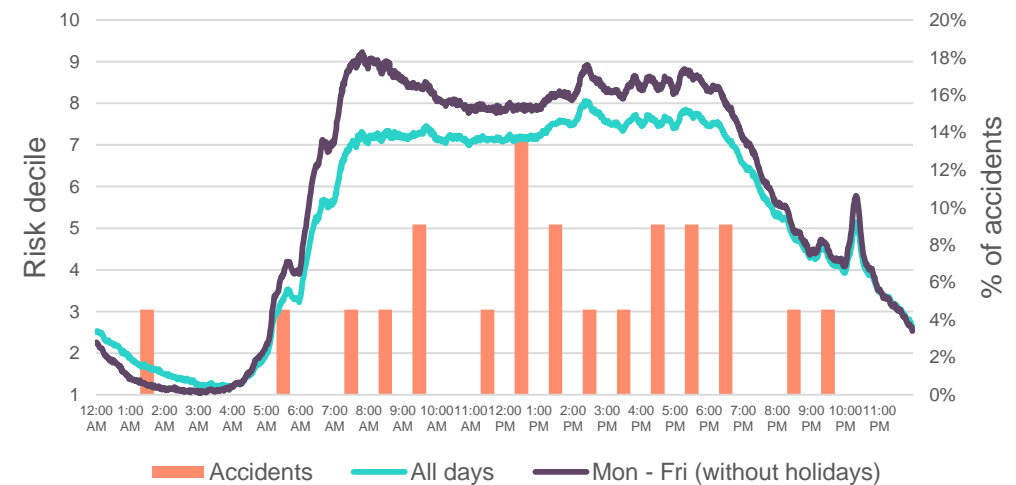
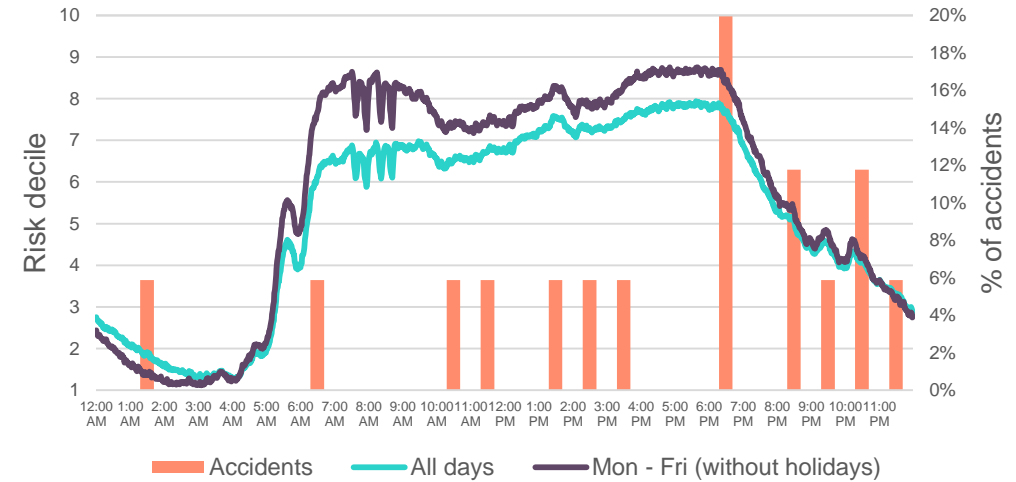
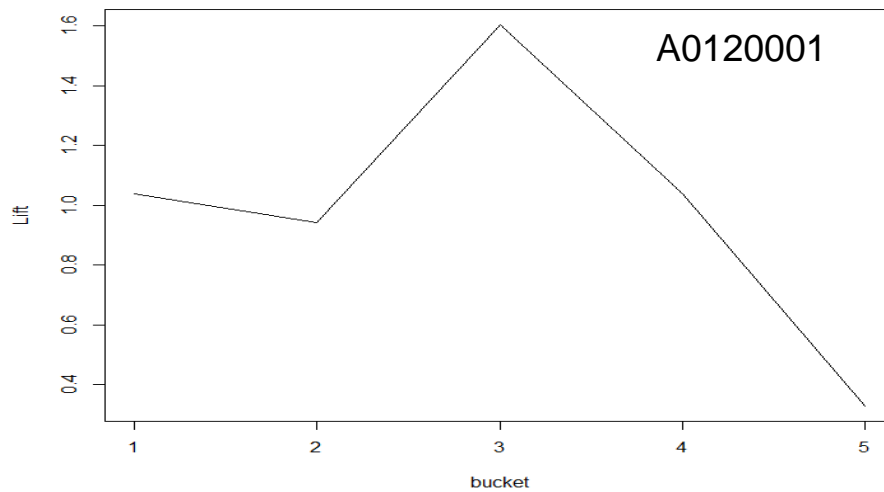
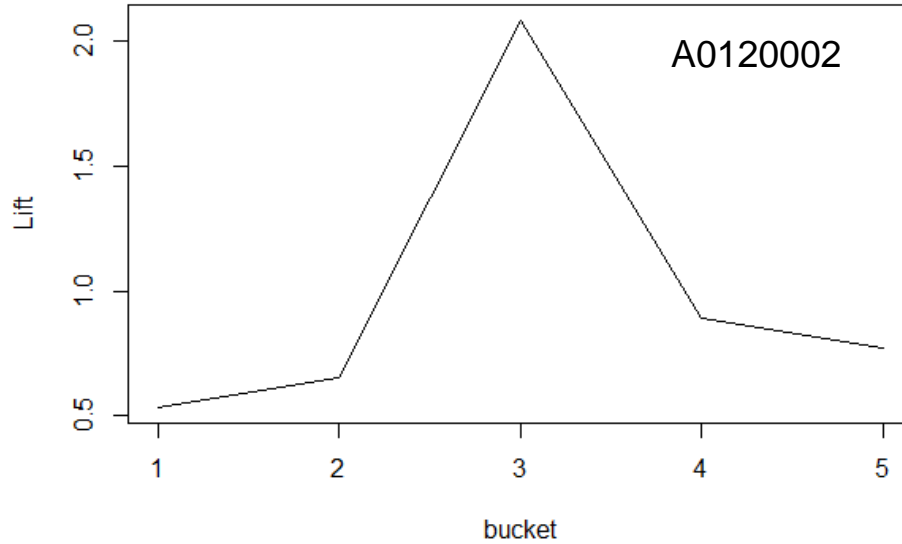
Model #7 – test on other road segments



Model #7 – test on other road segments



Model #7 – test on other road segments



Toepassen op A012 is niet goed.

Model for A012

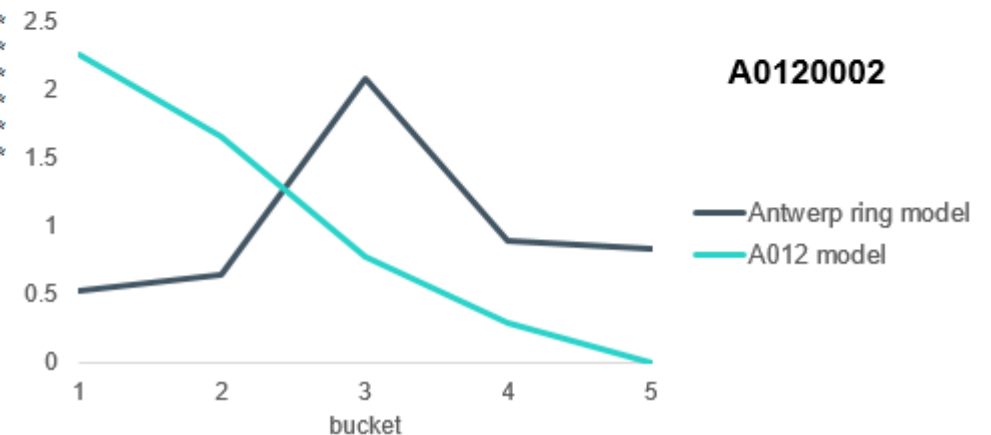
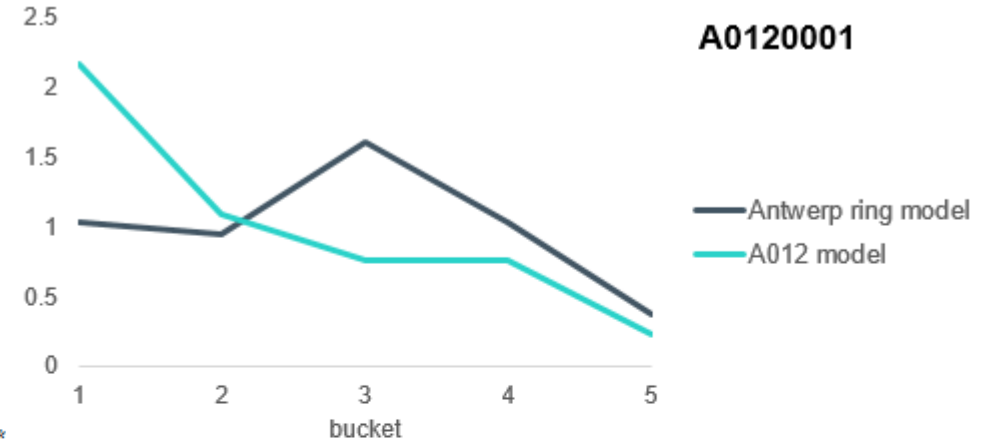
- Herberekenen model voor A012
- Training/validation sets bevat data van beide richtingen (A0120001 and A0120002)
- Data van eerste en laatste LocPost

coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.7339345	0.5467391	-6.829	0.000000000085233	*
avg_5min_kinetic_energy_min_k_first	-0.0036857	0.0007310	-5.042	0.0000004609788519	*
avg_5min_timespan_min_first	-0.0059162	0.0012657	-4.674	0.0000029499067725	*
stddev_5min_timespan_min_first	0.0047180	0.0008902	5.300	0.0000001157685250	*
vehicle_count_last	-0.0291770	0.0083394	-3.499	0.000468	*
vehicle_0_20_count_last	0.1906805	0.0252379	7.555	0.0000000000000418	*
avg_5min_speed_stddev_last	0.0767024	0.0255268	3.005	0.002658	*
avg_5min_vehicle_above_120_ratio_last	-26.3989163	4.2282642	-6.243	0.0000000004280489	*
stddev_5min_speed_max_last	0.0737763	0.0179570	4.108	0.0000398250486441	*

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Werkbaar model, voornamelijk andere variabelen en ook geen "Zondag".



Deel 3 - Modellerings

Besluiten & Relevantie

- **Ongevalsmodellerings** geven goede voorspellingen
 - Weerfactoren en/of kalender gerelateerde factoren komen niet stelselmatig terug in modellerings, het blijven de verkeersvariabelen en hun afgeleiden.
 - Gebruik van 1 minuut modellerings geeft beste resultaten
 - Modellerings zijn ongeveer 3x beter dan random.
 - DataMarts zorgen ervoor dat modellerings eenvoudig te bouwen zijn. Veralgemening is hiervoor niet echt nodig.

Wat bepaalt of een Locpost veilig is of onveilig?

Zijn bepaalde Locposts veiliger dan andere?

Wat zijn de bepalende factoren?

Kunnen we ongevallen voorspellen?

Welke informatie hebben we daarvoor nodig?

Hoe doen we die voorspelling het best?

Kunnen we modellen veralgemenen?

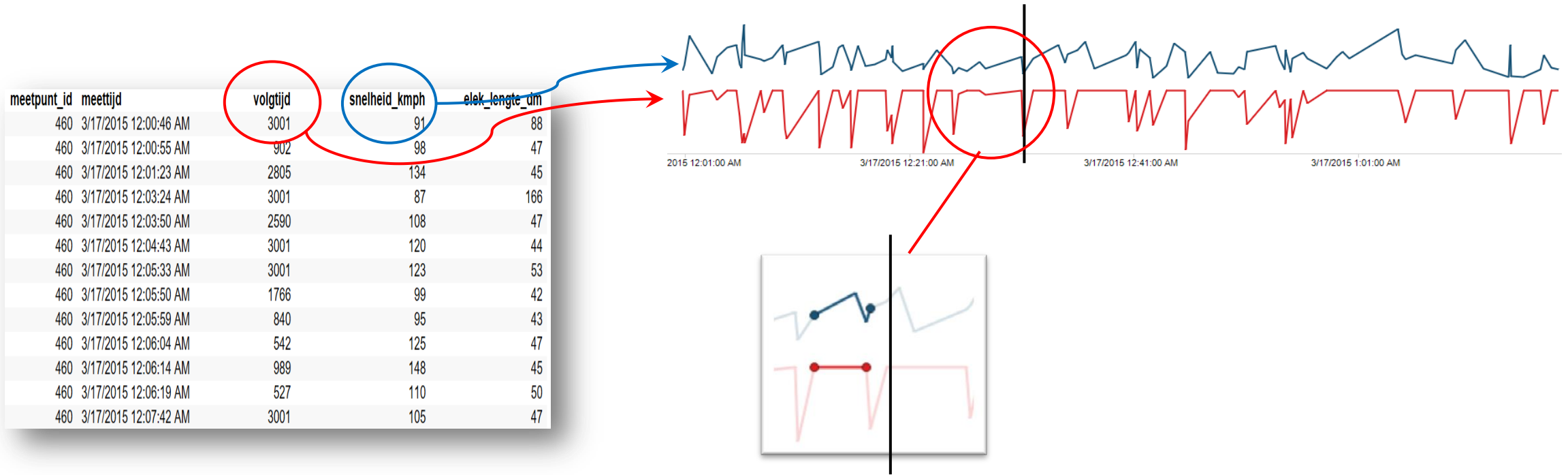
Zijn er patronen te herkennen die ongevallen veroorzaken?

Zijn er andere dimensies bepalend?



4. Tijdreeksanalyse - Patroonherkenning

Zijn er patronen te herkennen die ongevallen veroorzaken?



Verkeerspatroon = verloop van gemiddelde volgtijd en snelheid over de laatste 10 minuten

Aanpak

meetpunt_id	meettijd	volgtijd	snelheid_kmph	elek_lengte_dm
460	3/17/2015 12:00:46 AM	3001	91	88
460	3/17/2015 12:00:55 AM	902	98	47
460	3/17/2015 12:01:23 AM	2805	134	45
460	3/17/2015 12:03:24 AM	3001	87	166
460	3/17/2015 12:03:50 AM	2590	108	47
460	3/17/2015 12:04:43 AM	3001	120	44
460	3/17/2015 12:05:33 AM	3001	123	53
460	3/17/2015 12:05:50 AM	1766	99	42
460	3/17/2015 12:05:59 AM	840	95	43
460	3/17/2015 12:06:04 AM	542	125	47
460	3/17/2015 12:06:14 AM	989	148	45
460	3/17/2015 12:06:19 AM	527	110	50
460	3/17/2015 12:07:42 AM	3001	105	47



Gemiddelde Snelheid	14:00	Gemiddelde Volgafstand	14:00
Gemiddelde Snelheid	14:01	Gemiddelde Volgafstand	14:01
Gemiddelde Snelheid	14:02	Gemiddelde Volgafstand	14:02
Gemiddelde Snelheid	14:03	Gemiddelde Volgafstand	14:03
Gemiddelde Snelheid	14:04	Gemiddelde Volgafstand	14:04
Gemiddelde Snelheid	14:05	Gemiddelde Volgafstand	14:05
Gemiddelde Snelheid	14:06	Gemiddelde Volgafstand	14:06
Gemiddelde Snelheid	14:07	Gemiddelde Volgafstand	14:07
Gemiddelde Snelheid	14:08	Gemiddelde Volgafstand	14:08

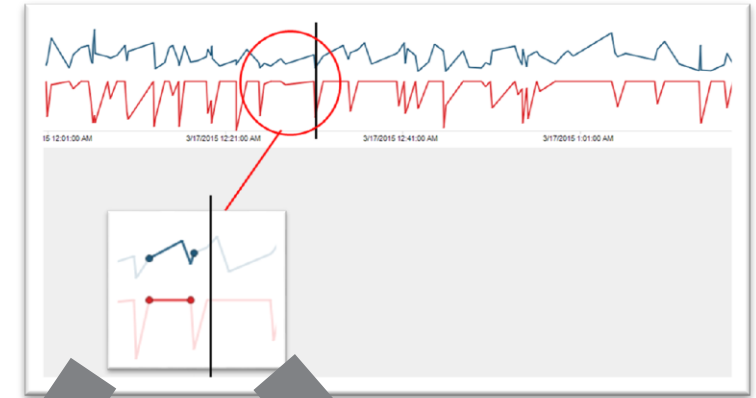
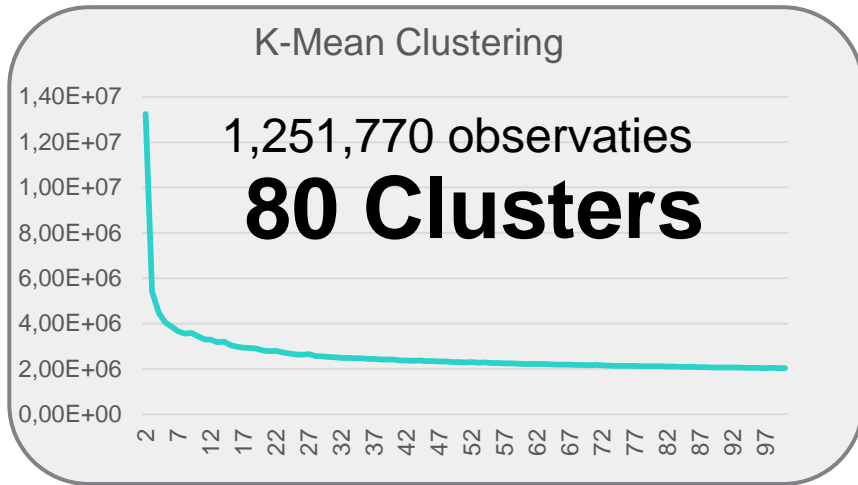
14:08

time_chunk	timespan_avg	timespan_avg_1	timespan_avg_2	timespan_avg_3	timespan_avg_4	timespan_avg_5
2015-03-07 12:04:00	336.8723404255	376.5306122449	448.7441860465	265.1818181818	331.2592592593	324.8958333333
2015-03-07 12:05:00	402.5238095238	336.8723404255	376.5306122449	448.7441860465	265.1818181818	331.2592592593
2015-03-07 12:06:00	336.94	402.5238095238	336.8723404255	376.5306122449	448.7441860465	265.1818181818
2015-03-07 12:07:00	304.2786885246	336.94	402.5238095238	336.8723404255	376.5306122449	448.7441860465
2015-03-07 12:08:00	366.2727272727	304.2786885246	336.94	402.5238095238	336.8723404255	376.5306122449
2015-03-07 12:09:00	298.6176470588	366.2727272727	304.2786885246	336.94	402.5238095238	336.8723404255
2015-03-07 12:10:00	251.5396825397	298.6176470588	366.2727272727	304.2786885246	336.94	402.5238095238
2015-03-07 12:11:00	346.1785714286	251.5396825397	298.6176470588	366.2727272727	304.2786885246	336.94
2015-03-07 12:12:00	325.9298245614	346.1785714286	251.5396825397	298.6176470588	366.2727272727	304.2786885246

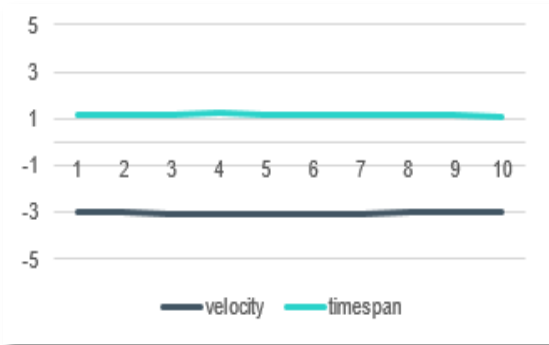
20 (genormaliseerde) variabelen per minuut

Voor visualisatie via PCA omgezet naar 2 variabelen, PC1 en PC2

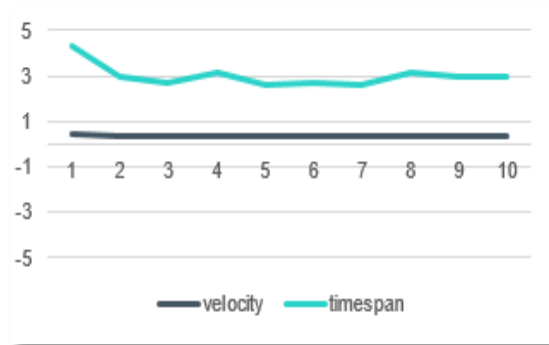
Clustering van alle "minuten" in 80 groepjes



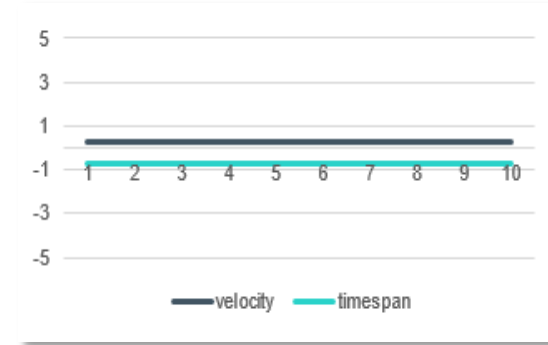
**Low speed
Average timespan**



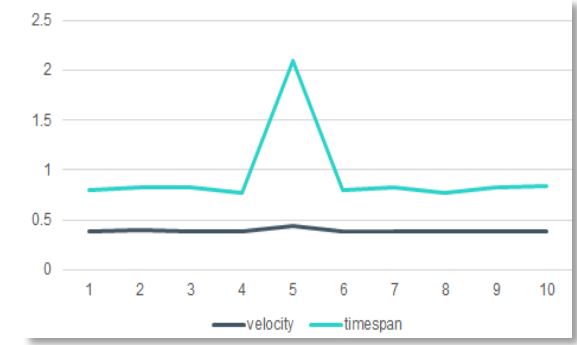
**High timespan
Average speed**



**Average speed
Average timespan**



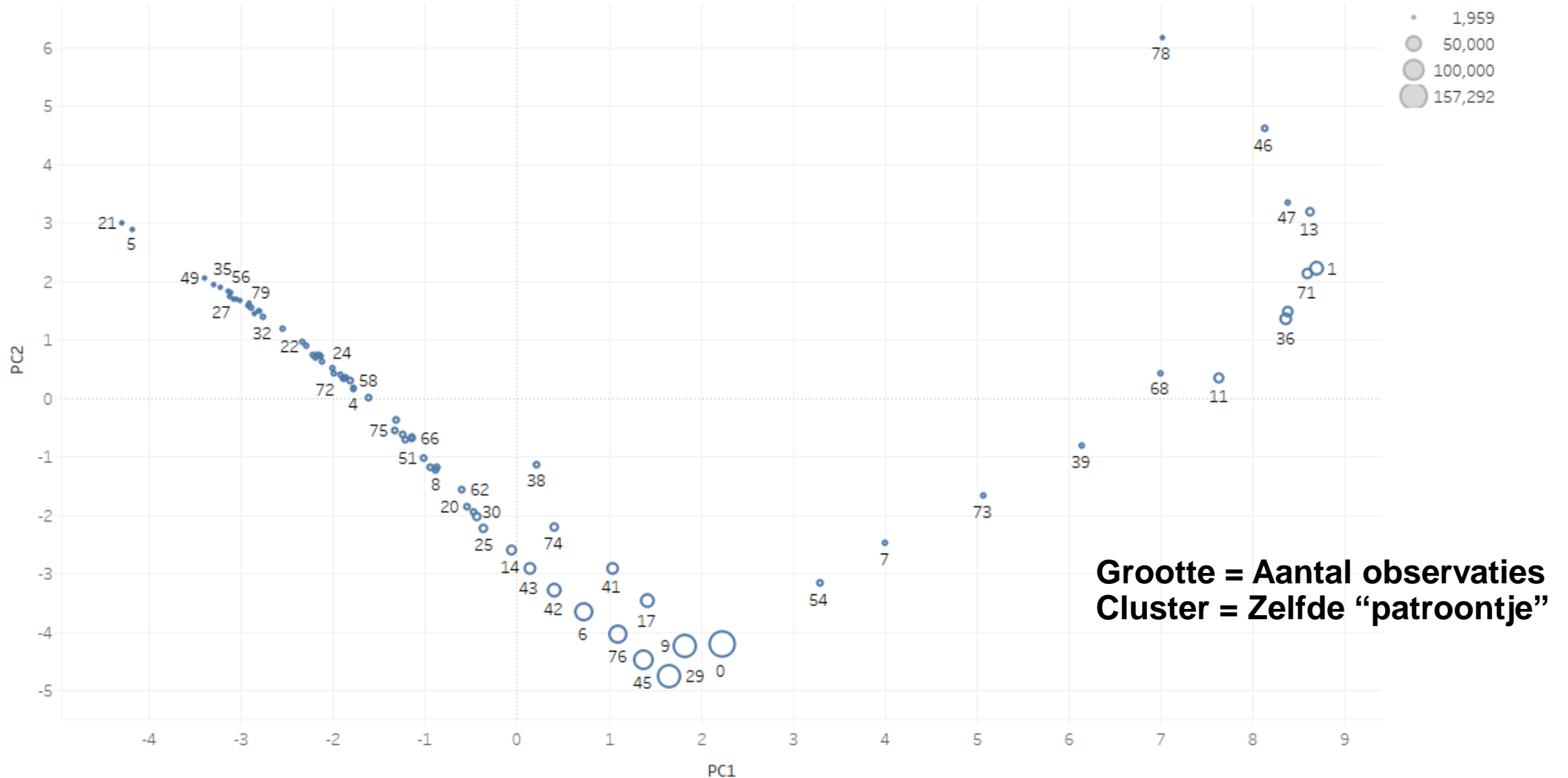
All kinds of peaks



Average speed: 85 [km/h]
Speed standard deviation: 26 [km/h]

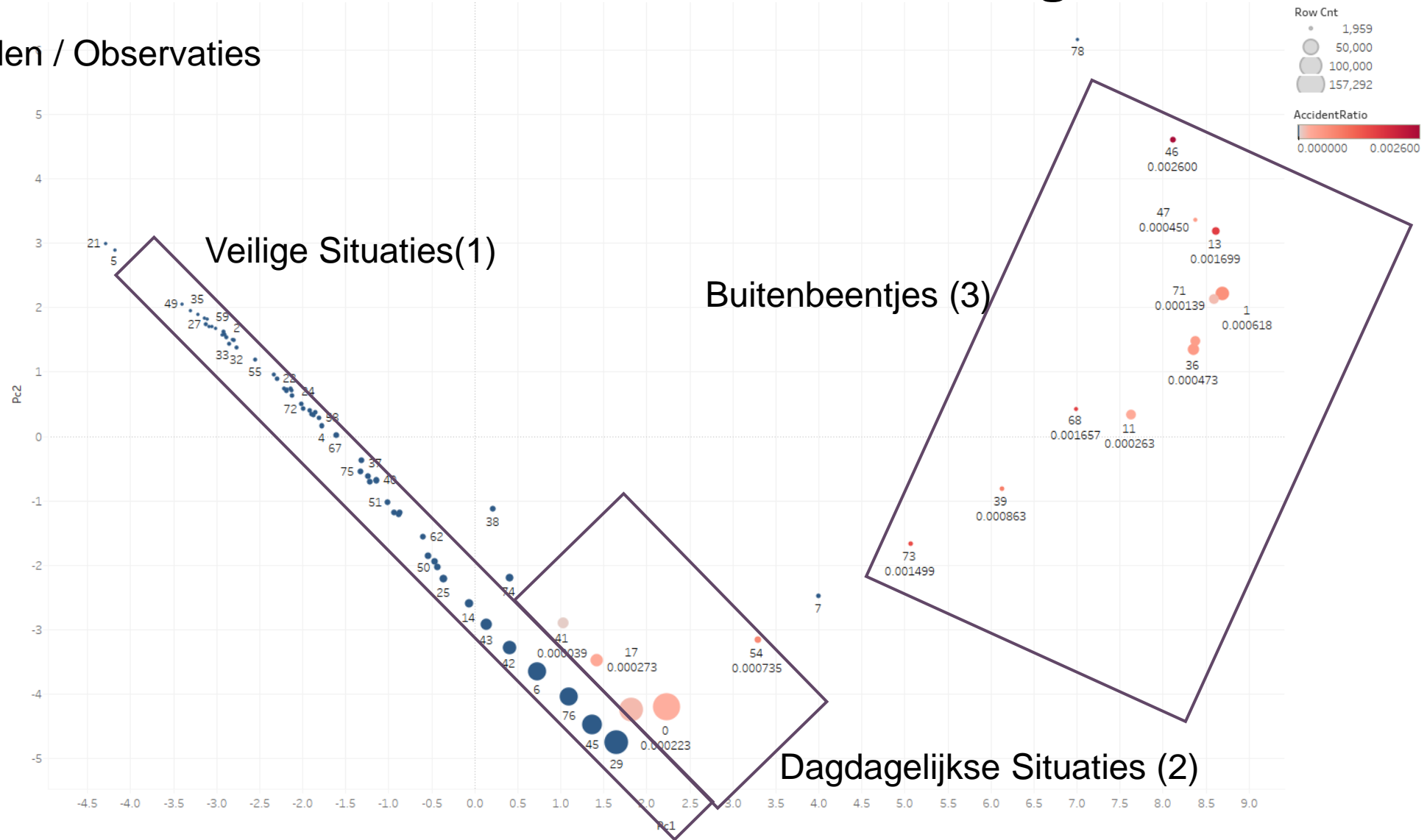
Average timespan: 666 [cs]
Timespan standard deviation 496 [cs]

Weergave van alle groepjes in 2 dimensies

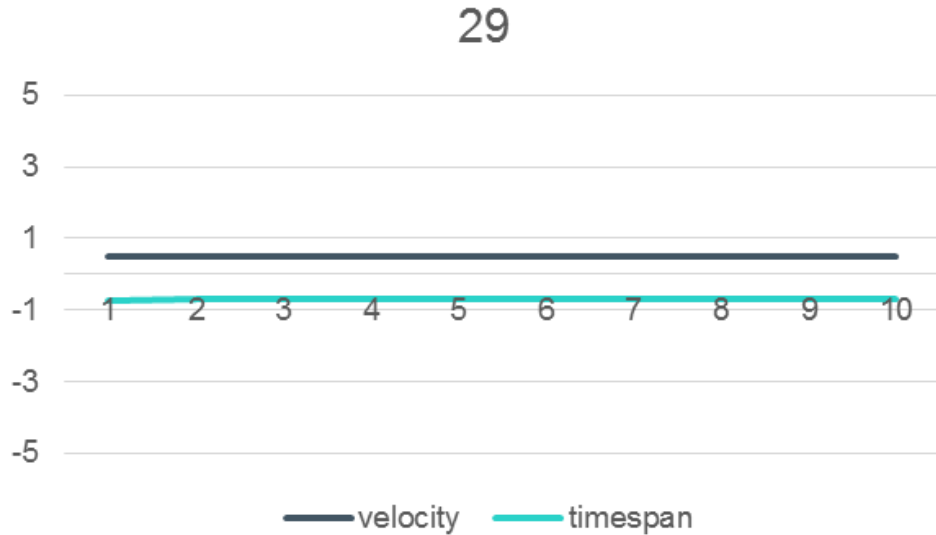


Is er een verband tussen deze clusters en ongevallen ?

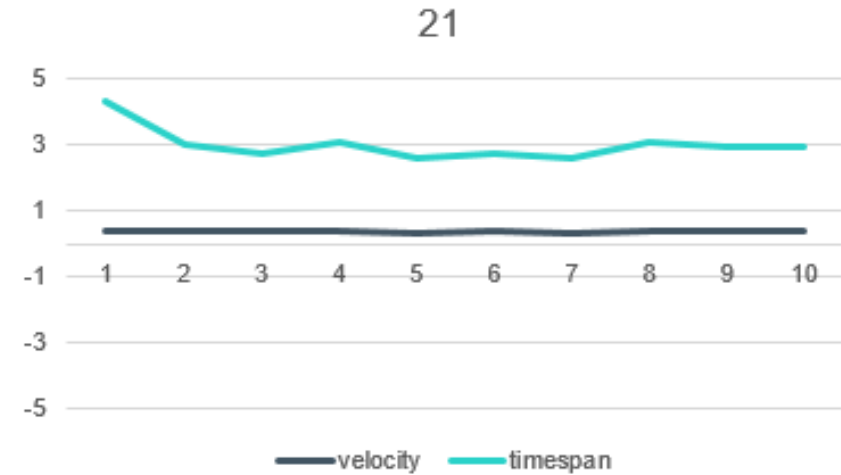
Ongevallen / Observaties



Veilige Situaties(1)

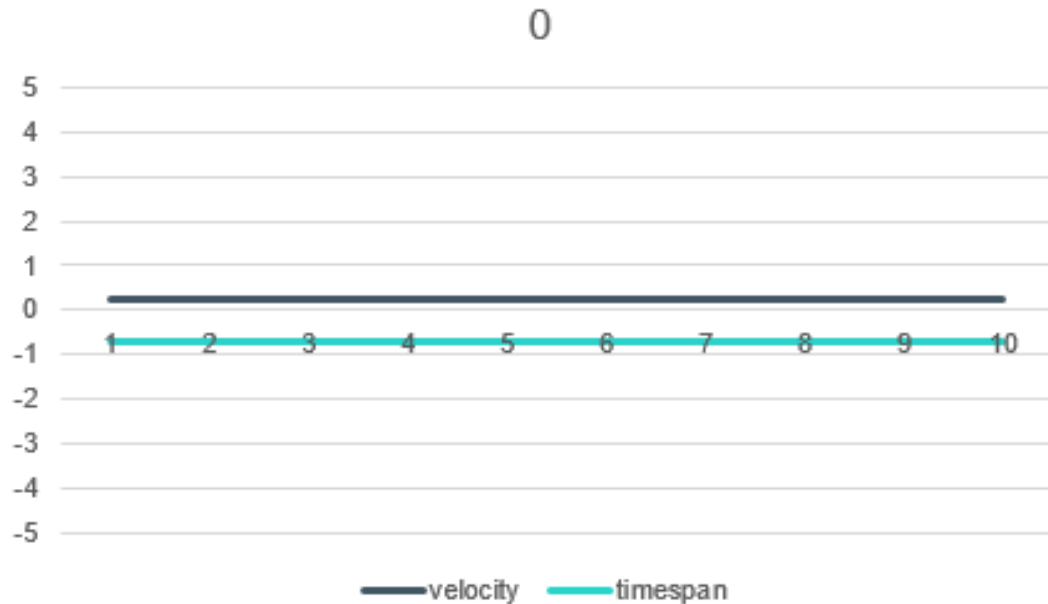


- Snelheid boven gemiddelde
- Tijdspanne korter dan gemiddeld
- „Snel, opeenrijdend verkeer ”
- **Meest voorkomende in deze cluster**
(10% observaties)



- Gemiddelde snelheid
- Tijdspanne veel hoger dan gewoon
- “Relaxed” Verkeer
- **Uitzonderlijk**

Dagdagelijkse Situaties (2)

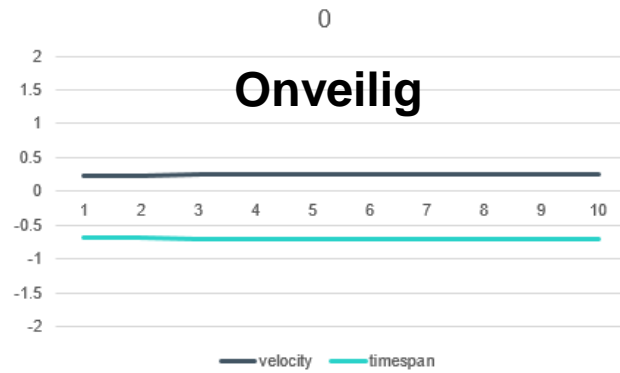
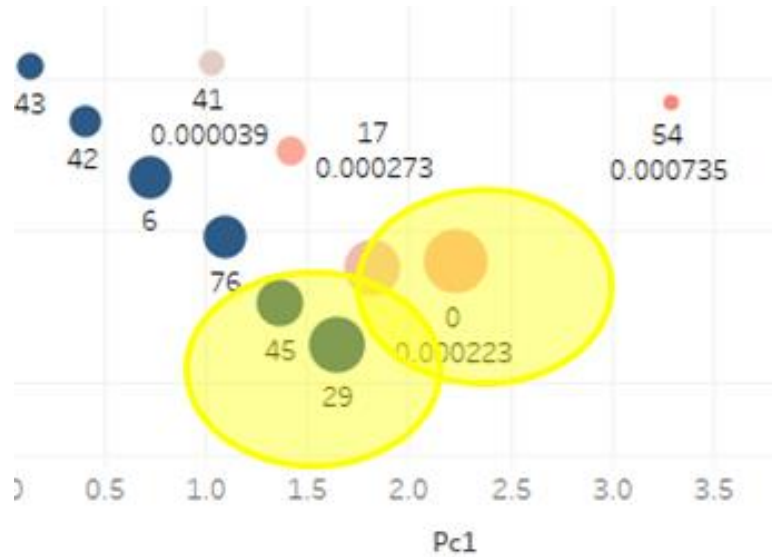


27% van alle observaties vallen in deze clusters
34% van alle ongevallen

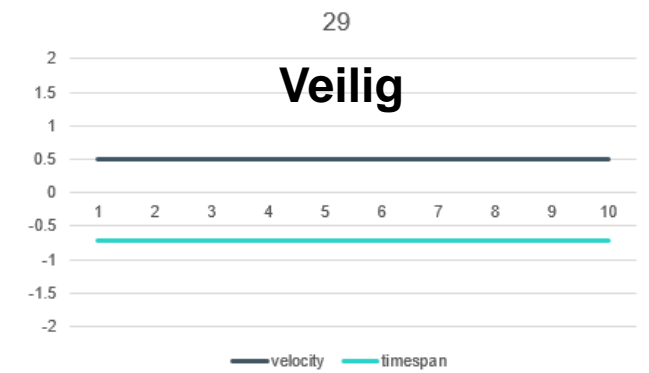
- Gemiddelde snelheid en tijdspanne
- **Cluster 0 = Meest Grootste**

12% van alle observaties
20% van alle ongevallen

Cluster 0 versus Cluster 29



- 12% van alle observaties
- 20% van alle ongevallen

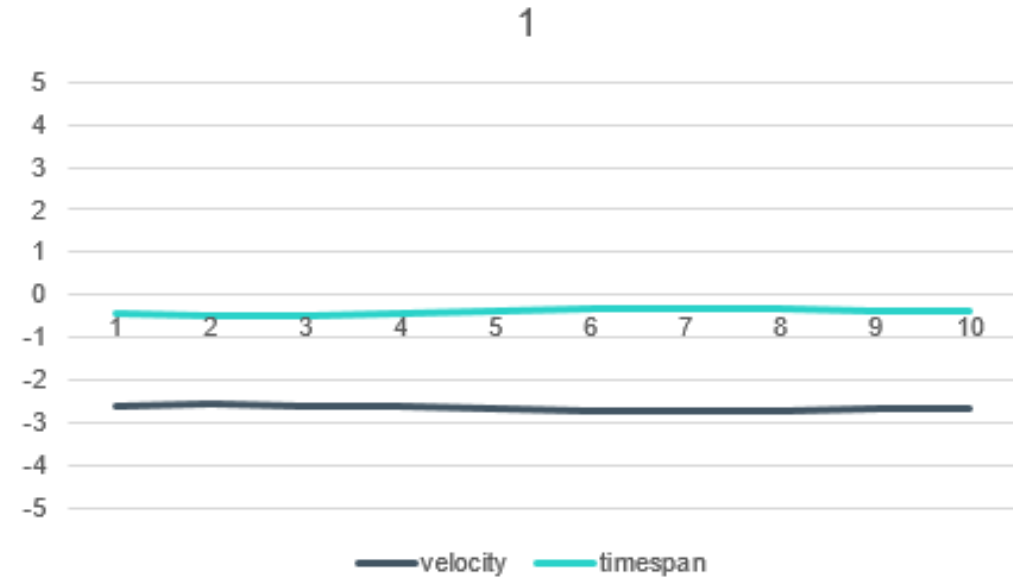


- 10% van alle observaties
- Geen ongevallen

Volgafstand korter dan gemiddeld – druk verkeer
Snelheid van veilige cluster is hoger

De buitenbeentjes (3)

- 16% van alle observaties
- 66% van alle ongevallen
- **Extreme situaties**
 - Volgtijd gemiddeld of licht lager
 - Heel lage snelheid
- Mogelijke verklaringen
 - Extreme weersituaties
 - File
 - Mogelijk invloed van verkeerde localisatie en tijdstip v.h. ongeval



Toepassen van score van Model #7

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.889594	0.410626	-9.472	< 0.000000000000000002	***
avg_5min_speed_stddev_first	0.145422	0.028707	5.066	0.000000406897255009	***
avg_5min_timespan_min_last	-0.020530	0.003098	-6.627	0.000000000034228399	***
avg_5min_vehicle_above_120_ratio_first	-23.124283	2.843278	-8.133	0.0000000000000000419	***
calendar_weekday_name_sunday1	-2.006470	0.460028	-4.362	0.000012909913725001	***
momentum_sum_M_last	-0.069825	0.013604	-5.133	0.000000285423973089	***
stddev_5min_speed_min_last	-0.035590	0.013055	-2.726	0.00641	**
vehicle_0_20_ratio_first	1.222071	0.266986	4.577	0.000004710325202168	***
vehicle_61_80_count_last	0.063777	0.009078	7.026	0.000000000002132284	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Deel 4 - Patroonherkenning

Besluiten & Relevantie

- **Tijdreeksanalyse van evolutie van 2 primaire verkeerfactoren** geven bevestiging van de resultaten van de modellering.
 - De verschillende patroontjes onderscheiden **veilige situaties** van **onveilige situaties**.
 - Extra analyse m.b.t. verschillen (outliers) binnen elke cluster gaf geen extra informatie.

Opmerking: Dit was een eerder een exploratieve studie op 1 locpost. Idealiter te herhalen op andere locposts vooraleer verder uit te werken.

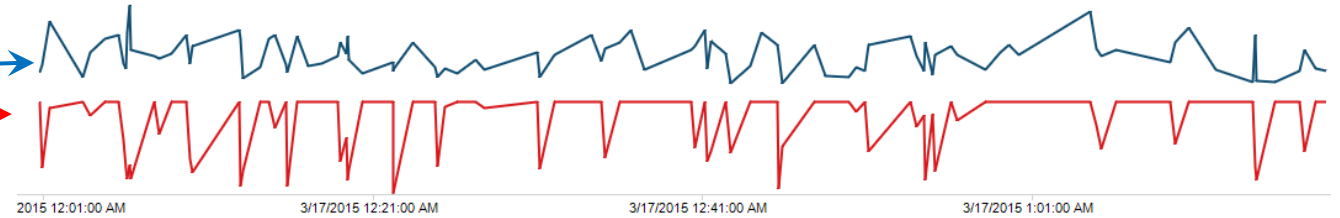
Wat bepaalt of een Locpost veilig is of onveilig?
Zijn bepaalde Locposts veiliger dan andere?
Wat zijn de bepalende factoren?
Kunnen we ongevallen voorspellen?
Welke informatie hebben we daarvoor nodig?
Hoe doen we die voorspelling het best?
Kunnen we modellen veralgemenen?
Zijn er patronen te herkennen die ongevallen veroorzaken?
Zijn er andere dimensies bepalend?



5. Frequentieanalyse & Multiscale- analysis

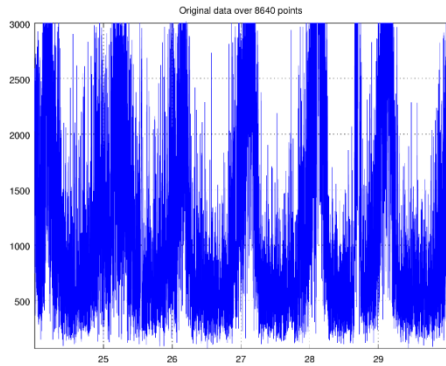
Zijn er extra dimensies die we nog niet onderzocht hebben ?

meetpunt_id	meettijd	volgtijd	snelheid_kmph	elek_lengte_dm
460	3/17/2015 12:00:46 AM	3001	91	88
460	3/17/2015 12:00:55 AM	902	98	47
460	3/17/2015 12:01:23 AM	2805	134	45
460	3/17/2015 12:03:24 AM	3001	87	166
460	3/17/2015 12:03:50 AM	2590	108	47
460	3/17/2015 12:04:43 AM	3001	120	44
460	3/17/2015 12:05:33 AM	3001	123	53
460	3/17/2015 12:05:50 AM	1766	99	42
460	3/17/2015 12:05:59 AM	840	95	43
460	3/17/2015 12:06:04 AM	542	125	47
460	3/17/2015 12:06:14 AM	989	148	45
460	3/17/2015 12:06:19 AM	527	110	50
460	3/17/2015 12:07:42 AM	3001	105	47



Frequentieanalyse A12

Time representation

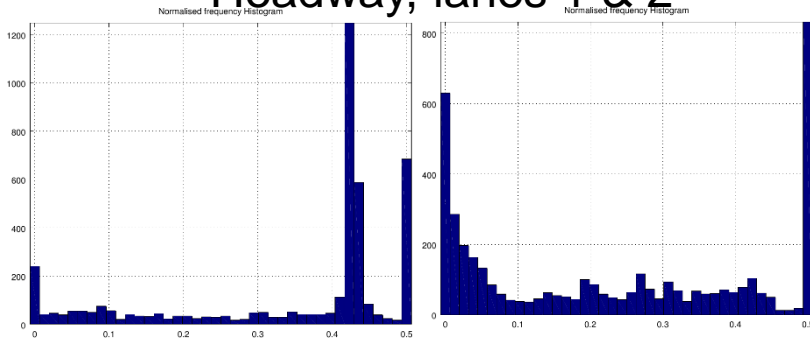


Fourier
Transform

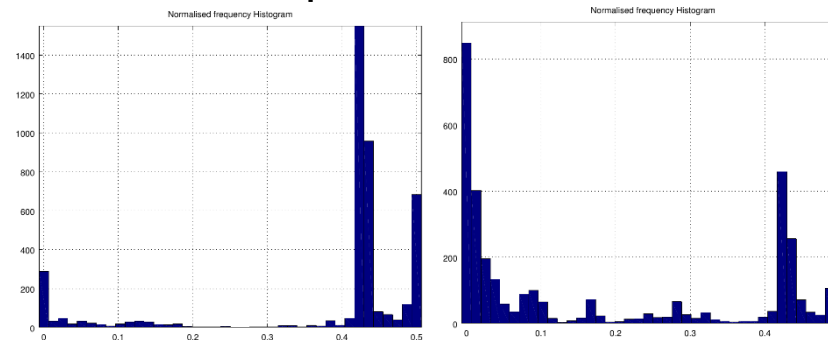
Time/Frequency representation



Headway, lanes 1 & 2



Speed, lanes 1 & 2

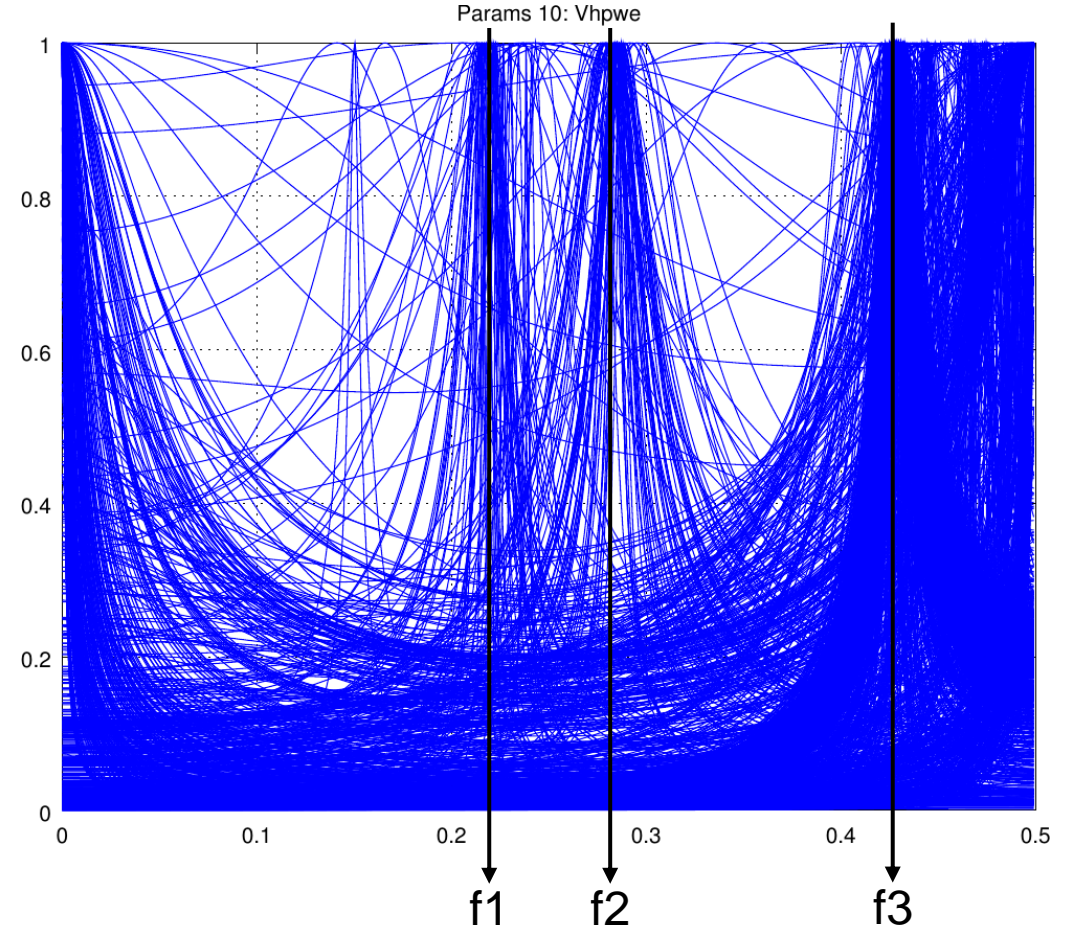
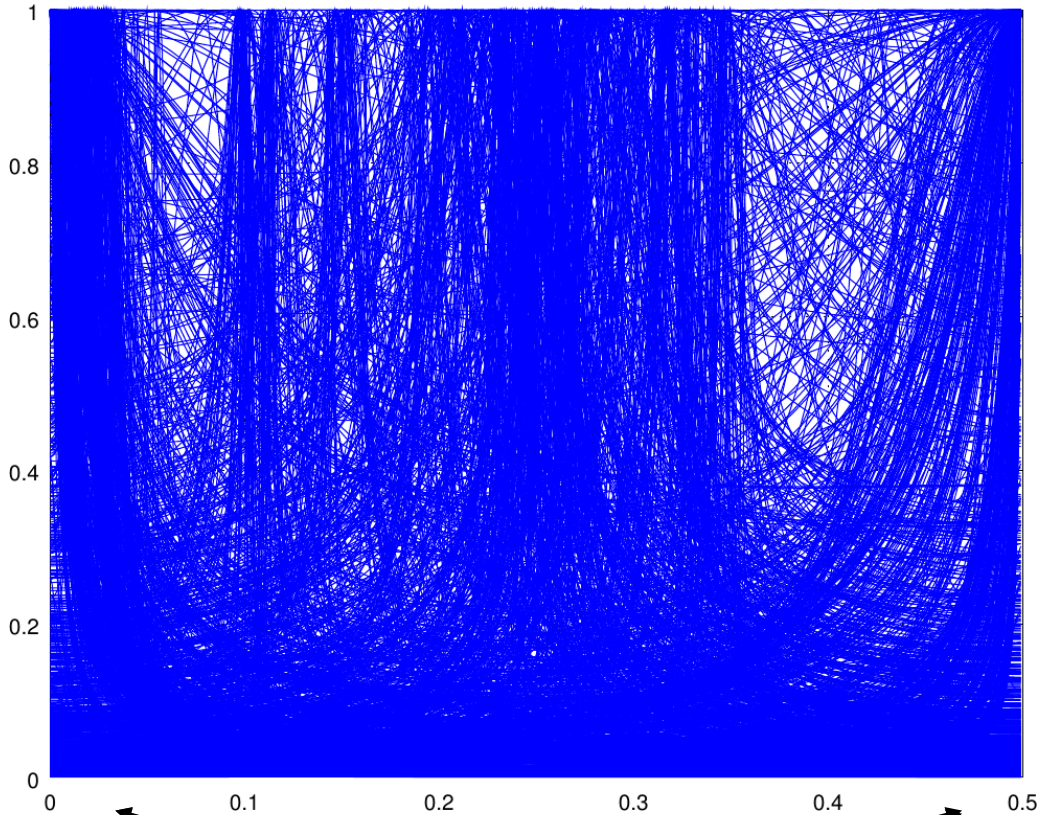


**Duidelijke dominante
frequentiemodulaties**

Frequentieanalyse A12

Observatie: 2 soorten van verkeersgedrag

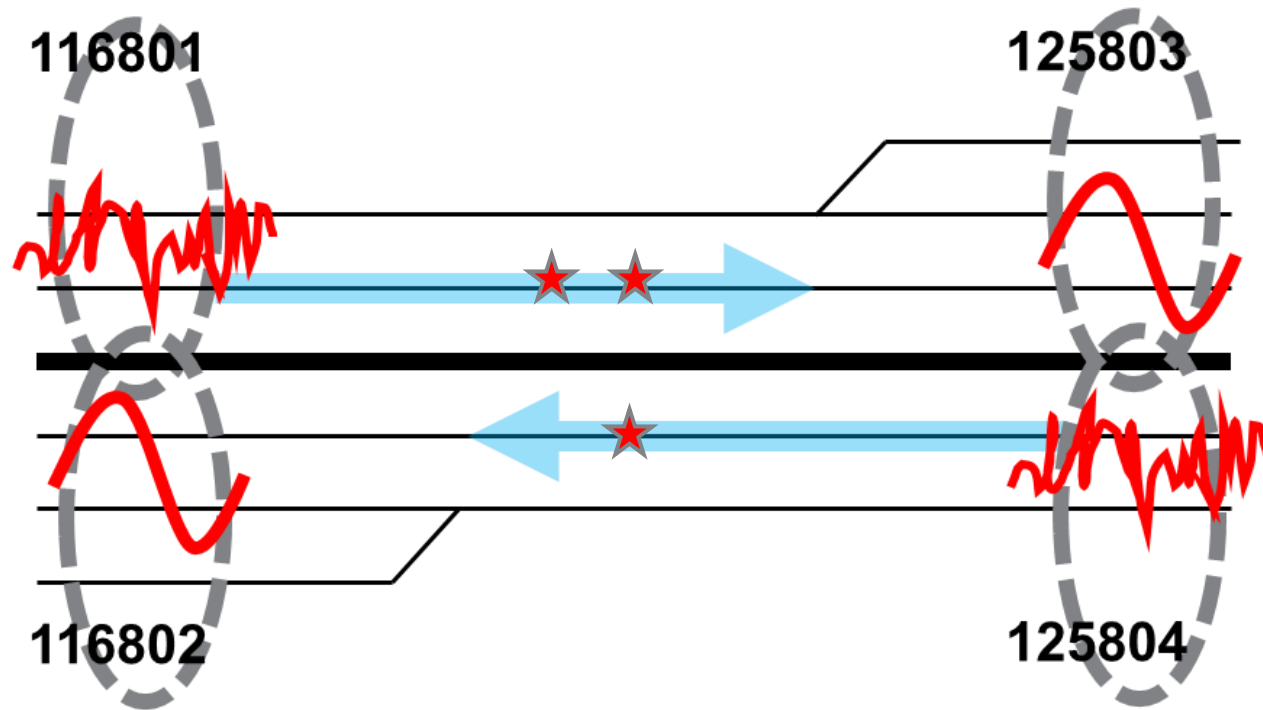
Params 10: Vhpwe



Vele verschillende frequenties
Variabel gedrag

Duidelijke dominante frequenties
Stabiel gedurende bepaalde tijd

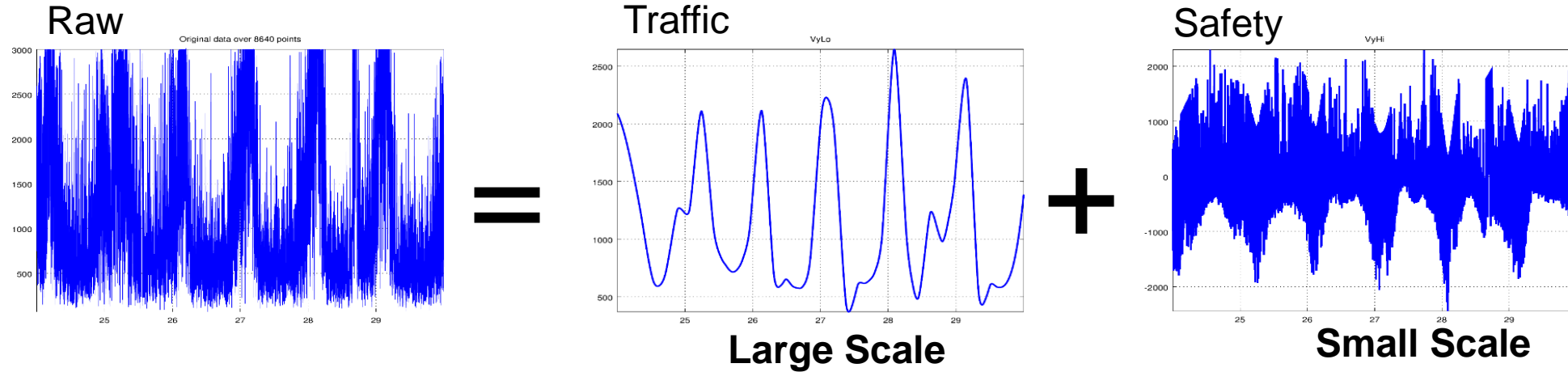
Frequentieanalyse A12



- ★ 3 accidents in between:
- ★ 26/07/2015 12:15, Dir 1
- ★ 26/07/2015 01:20, Dir 2
- ★ 29/07/2015 15:30, Dir 1

Oorzaak ? Gevolg ?

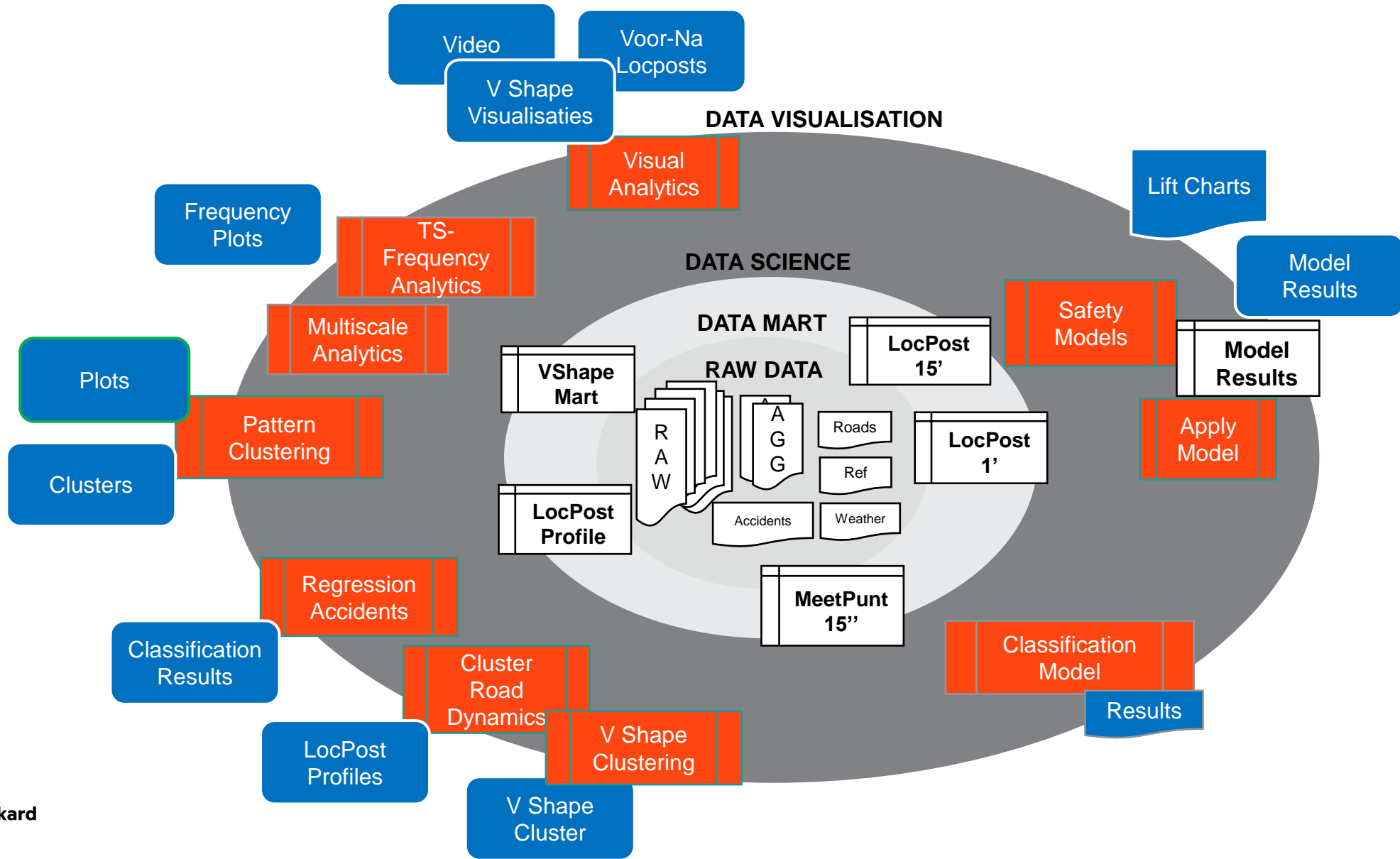
Multiscale analysis



Extra differentiatie mogelijk op "input" signalen zodat er meer nadruk komt op verandering t.o.v. normaal gedrag.



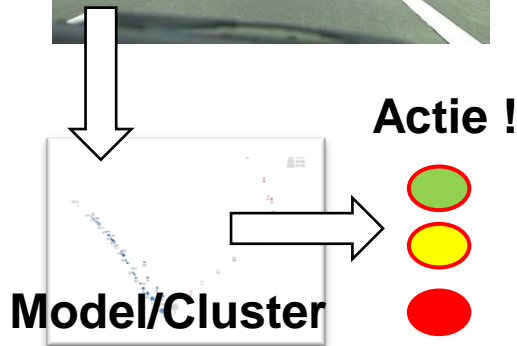
Eindconclusies en Volgende Stappen



Enkele denkpistes...

Implementatie

- Gebruik van **voorspellende modellen** en/of patroonherkenning om actie te nemen !
- Mogelijke acties van inschakelen extra monitoring, sturing verkeer tot notificatie hulpdiensten.



Uitbreiding Breedte

- Toepassen van patroon herkenning en clustering **op alle locposts** om zo meer aanpak te valideren en extra inzichten te verkrijgen
- Bouwen van **automatische modelgeneratie** om zo voor elke locposts een model te krijgen en dan die modellen te kunnen vergelijken.
- Toepassen van bijkomende correcties (tijd/plaats) op **ongeval gegevens** – eventueel via modellering.
- Bijkomende **opsplitsen van modelleringen** m.b.t. tijd en ruimte, en/of verkeersstroom.
- Uitvoeren **frequentieanalyse** op meerdere locposts.
- Hergebruik van DataMarts voor andere use-cases (afstandscontrole)

Uitbreiding Diepte

- Uitvoeren van **onderzoek** wat verschillen zijn tussen locposts in zelfde clusters.
- Onderzoek naar verschil in **voorspellende factoren** tussen locposts.
- Uitvoeren **modelleringen** met gebruik van **“2de orde” factoren** zoals bijvoorbeeld : gevonden clusters, berekende score van andere modellen, berekende dominante frequenties en multi-scale data.
- Uitvoeren **modelleringen** per rijstrook

Een data gedreven verkeersbeleid

Een blik op de toekomst ?

