Investigation into Multi-Modal Risk Assessment In Helsingborg

Final report from project with support from Vinnova (DNR 2022-02900) as part of the call "Innovations for sustainable mobility system", November 2022 to August 2023.

Project Partners Research Institutes of Sweden (RISE), City of Helsingborg and Vianova.

Author: Alex Pazuchanics, Vianova.
Co-authors: Kent Eric Lång, RISE
Niklas Tengheden and Oscar Grönvall, City of Helsingborg

Abstract

Safety for urban road users is an important and recurring theme in Sweden and the world. As cities grow and new means of transport and solutions are added, safety becomes even more important as more people and vehicles share the same space. Cities and mobile operators often find it difficult to understand which measures have the greatest positive effects and how different infrastructure and policy updates affect the safety of those on our roads. The growing amount of data produced in our cities, such as telematics from sensors and IoT applications, has proven to have great potential to help cities, mobility operators and vehicle owners understand where there are potential safety risks, what measures can prevent them and how implemented initiatives affect the level of safety on roads.

The purpose of this project was to understand how to use data and new technologies to increase safety on roads and do better planning of infrastructure and policy work. The project manager was RISE, Vianova was responsible for delivering the technology, data and a final report. The City of Helsingborg acted as a subject expert on road safety, infrastructure and policy.

The use of connected vehicle data (connected cars, commercial vehicles and a connected version of "vulnerable road users" such as bikes, scooters, and pedestrians) have potential to fulfil use cases. Methodologically, connected vehicle data can add some insights, particularly in identifying areas where collisions have not yet occurred, but where the elements of risk are present.

Vianova works with a number of providers to generate floating car data. However, several data sources proved unsuitable in the Helsingborg region, and so Vianova was only able to rely on a single source of data. Longer time periods, more robust data sets (including multiple vehicle types), and complementary sources of data could enhance the value of the tool in order to make it more impactful. Additionally, the partners iterated on strong user experience improvements in the Vianova platform to improve the legibility of the experiences.

The use cases identified are relevant for the city, and the potential of using a mix of data sources has genuine value. More data sources are needed to produce insights sufficient to complement the existing traffic safety data generated from collision reports.

If it is possible to identify relevant correlation between personal injuries and Floating Vehicle Data (FVD), a tool of this kind can be an important support in the cities' analysis work. If the data can be updated in real time, there are great opportunities to supplement today's planning methods with current situation analyses and thus be even more agile in our road safety work.

Content

Background

In 1997, the Swedish Parliament adopted a new long-term goal and strategy for road safety, Vision Zero. The goal is for no people to be killed or seriously injured in a traffic accident. The Swedish government's website for road safety, roadsafetysweden.com, states "*Sweden has succeeded in moving from vision to action. For more than two decades, various stakeholders in Sweden have worked proactively towards the goal of zero fatalities and serious injuries on the roads. For example, many roads have been provided with a central barrier to prevent frontal accidents; the police can check the speed of cars with cameras; Many four-way intersections have been replaced with roundabouts; the automotive industry invests in safer cars; and when designing the urban traffic environment, the focus is on vulnerable road users. The work has paid off. When Vision Zero was adopted by the Swedish Parliament, the number of deaths in traffic accidents was 7 per 100 0000 inhabitants. At the time, it was a low figure from a global perspective and many were sceptical that it would be possible to reduce the figure further. Since then, the number of road deaths in Sweden has more than halved, while the traffic volume has increased dramatically. However, the vision has still not been fully implemented. In recent years, the reduction in deaths has levelled off and new efforts are now needed.*"

There is a need for new innovative solutions to ensure that the number of road accidents continues to decrease in Sweden and that the curve does not stop horizontally or, worse still, go backwards in development. Given that there are more and more people, vehicles and traffic services to share the same. One of the potential solutions for this is to make use of better data and insights that can be created from connected vehicles and other devices. This data can help us understand the final pieces of the puzzle in Vision Zero.

There are several companies and projects that have carried out security analyses and improvement projects using telematics data. Both OEMs such as Volvo, Renault and Daimler as well as technology companies such as Here use telematics data to develop their own security products that will help cities improve safety. This project is unique in that it will analyze telematics data from several types of vehicles: private cars, trucks, bicycles and electric scooters. This data will go through Vianova's AI and ML based models to make risk assessments of all roads in the city.

Purpose of the project

The purpose of the project was to help the City of Helsingborg (as well as other cities) to understand how they can work more data-driven with security issues and better prevent accidents. Success will be measured by the following criteria:

- Number of unknown safety risks in the road network identified
- Number of commitments identified and implemented to reduce the number of accidents , e.g. infrastructure and policy changes
- Reduction in the number of accidents on Helsingborg's roads

More specifically:

- Combine multiple data sources to better understand vulnerable road user (VU) vulnerabilities.
- Improve the city's infrastructure to increase safety for all road users.
- Predict dangerous events for VRUs.

Methodology

The methodology in this project is unique in that telematics data from several types of vehicles was analysed: private cars, trucks, bicycles and electric scooters. This data went through Vianova's AI and ML based models to make risk assessments of all roads in the city. The project will also include other data sources such as geo and infrastructure data.

Data sources to collect:

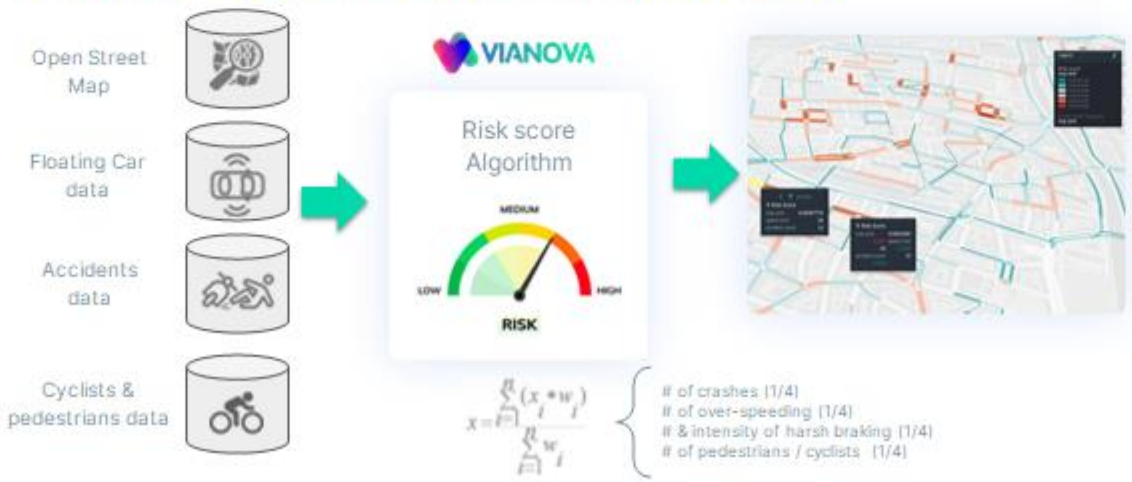
- Shared mobility: cars, bikes, e-scooters
- Connected vehicle data (FCD and ADAS) from OEMs
- Pedestrian and bicycle counting
- Crash Data from Road Authorities
- Infrastructure Data (bike lanes, intersections) + construction work

The project set out with three global use cases for a road safety practitioner to test, see chapter below.

A digital "road safety dashboard" was developed in the Vianova platform. With the help of this platform, the parties made analyses with the purpose to reduce the number of traffic accidents the city of Helsingborg.

Multi-dataset / Combining

Combining data to produce risk score per road segment



How it works

1. Data integration & aggregation — ***Abnormal driving events*** *(harsh braking, acceleration, over-speeding, harsh-cornering)*
2. Data enrichment — ***Combining other data-sets for explanation:*** *road conditions, pedestrian & cyclists flows, traffic flows*
3. Data contextualisation & visualisation — ***Time of day, day of week, weather,*** *cycling lanes or pedestrian crossings, speed limit*
4. Risk score & algorithm — ***Multi-factor risk factor calculation*** *- risk factor will be adjustable in the near-future*
5. Prioritisation, intervention & tracking — ***Prioritisation of risk hotspots with thresholds methodology*** *(risk score, # of events, etc.)*

Examples of what Vianova's Road Safety project has looked like in the past.

Global risk score

A global risk score was calculated, per segment, based on the following mix of criteria:

-The total count of collisions of moderate or major severity
-The total number of harsh braking events
-The total number of overspeeding observations
-The percentage intensity of the average overspeeding event, compared to the posted speed of the road
-The total number of trips taken by shared micro-mobility

These counts were normalized by the number of trips identified on each road in the data set of connected vehicles, in order to ensure representativeness. Due to the road segment aggregation technique, road segments were of approximately equivalent length.

A risk score is created for each individual component. The combined risk score is taken as an average (equally weighted) of the individual risk score. In this manner, risk can either be generalized (taken across the combined set of variables) or localized (identified for a particular type of risk, such as the risk of over-speeding).

It is mathematically complex to “reweight” individual components to produce new risk scores based on the weighting significance of each component. But, by filtering roads based on individual component scores, a user can “reweight” the importance of various factors, or exclude a factor altogether. This technique effectively allows for a user to re-evaluate riskiest roads based on a set of new factors beyond those determined by Vianova.

A risk score can be quantified in three equivalent, but differently descriptive manners. The table below summarizes the “rank”, percentile”, and “category” approaches to individual or overall risk.

	Rank	Percentile	Category
Description	The absolute ranking of a street segment, when compared to other street segments.	A normalization of street segments, compared to the total universe of segments	A one to five normalization of the street segments
Rationale	The rank is a constant sorting of street segments which remains consistent even as certain segments are excluded through filtering	Provides moderate precision about the relative presence of a variable	Serves as a quick shorthand for the relative presence of a variable compared to others
Risk is found with...	Lower values	Higher values	Higher values
Minimum Value	1	1	1
Maximum Value	Dependant on the number of road segments	99	5

Data Sources

The project used multiple different data sources from a mix of providers.

Road Information

The project relied on a set of road segments identified in the Helsingborg area via Open Street Maps. OpenStreetMap (OSM) is a collaborative, open-source mapping project that aims to create a free and editable map of the entire world. It utilizes crowd-sourced data from a global community of contributors who gather, verify, and update geographic information. OSM provides a platform where individuals can add and edit various map features such as roads, buildings, points of interest, and more.

With its comprehensive and detailed mapping data, OpenStreetMap has become a valuable resource for a wide range of applications, including navigation systems, urban planning, disaster response, and research. Its open nature allows for continuous improvement and customization, making it a powerful tool for both individuals and organizations seeking accurate and up-to-date geographic information. OpenStreetMap serves as a basemap for several popular mapping systems in areas where private sector data collection is inadequate.

Helsingborg has a total of 2,253 km of roadway covered, meaning a set of segments and intersections totaling 16,481.

OpenStreetMap provided the initial road segment information, as well as information about the speed limits of the various road segments. However, review by city partners indicated that the OpenStreetMap speed limit information was at times inaccurate, reflecting the weaker crowd-sourcing presence in a small city compared to in large urban areas.

Historic Collision Data

Information about historic collisions was provided by Helsingborg City, based on the national traffic database (STRADA). The collision count (excluding fatalities) was 1,564 over the past 5 years, and collisions were ranked with a 1 to 3 scale for severity.

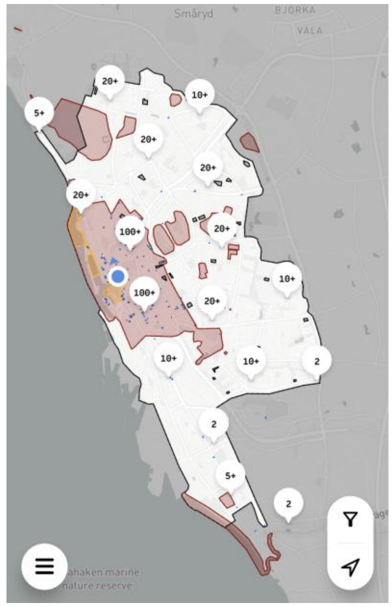
Shared Micro-Mobility Data

Many traditional transport data sources leave out the travel behavior of cyclists and pedestrians. As these travellers are not easily counted using traditional data collection techniques, it can be difficult to understand their presence, especially across an entire city network.

Vianova’s experience with shared micro-mobility operators gives us an understanding of the data sources which can be used to understand relative density of trips taken on e-scooter. Shared micro-mobility data was collected from the electric scooter companies operating in Helsingborg. Six months of data was collected from providers Voi, Tier, and Bolt.

Data collected was primarily “mid trip telemetries”, ie, gps traces of vehicles as they travelled across a particular road segment. These traces, which are recorded typically between every 4 and every 12 seconds (depending on the operator) can be used to record the route of a trip from its origin to destination. Aggregated together, these traces can be used to show a relative density of trips taken by shared modes.

Shared e-scooters are heavily used in the dense urban areas of Helsingborg, but are not available in the entire zone of Helsingborg explored in the other data sources. Therefore, there is a risk that ridership of shared modes is not as reflective of general micro-mobility patterns (including privately owned bikes and scooters) than in other cities where services are available across the city.



Floating Vehicle Data

Vianova works with a number of providers to generate floating car data. However, several data sources proved unsuitable in the Helsingborg region, and so Vianova was only able to rely on a single source of data, Bridgestone Mobility Solutions (Bridgestone, or BMS).

Bridgestone collects data using a set of aftermarket sensors in order to help fleet managers improve the performance of their vehicles. As a result, the dataset is skewed to fleet managed vehicles, such as vans, trucks, and buses. Approximately 15-20% of the vehicles are also classified as “cars”, which are typically driven either as part of motor pools, or operated as privately driven “company cars”.

Bridgestone data included approximately 10.6 million data points, covering a date range from January 2022 to March 2023.Given the road network, this resulted in an average of 644 data points per road segment per year. However, the median count of data points per road segment per year was only 15 observations. This wide disparity suggests that the vast majority of the road network (by kilometerage) in Helsingborg was only lightly covered by the data set (though these road segments are also likely underused by all travellers). Data observations were concentrated in the urban core, along major arterials, and disproportionately located in industrial areas near the harbour, where commercial vehicles may be more likely to appear than in other locations.

Use Case Results

The project set out with three global use cases to test, which written in standard use case formula, would be the following:

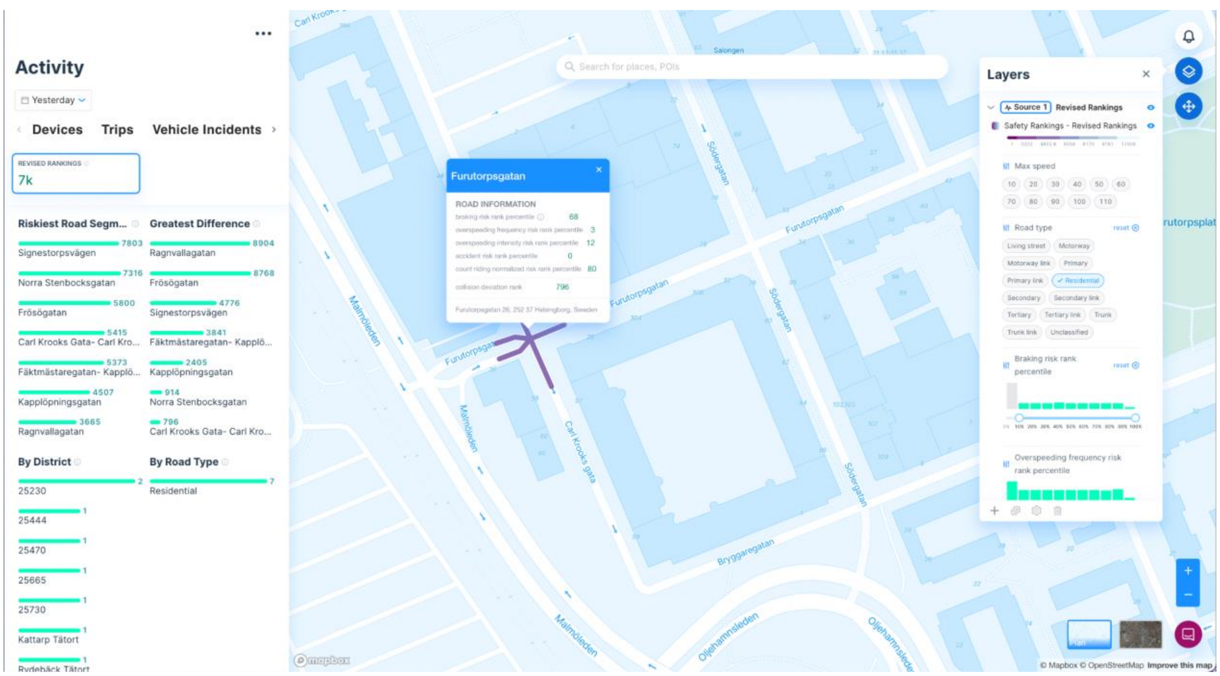
Use Case 1: “As a road safety practitioner, I want to identify road segments with particular risk, and compare the relative risk of two road segments in my city, in order to better prioritize my road safety investments”.

Use Case 2: “As a road safety practitioner, I want to understand the relative components of risk for any given street, and particularly to understand roads where collisions are low but other risks are high (potentially signalling a spot of future collisions), in order to supplement the knowledge I already has from collision history data”.

Use Case 3: “As a road safety practitioner, I want to evaluate the effect of before and after interventions to slow down traffic and reduce the risk of collisions on a street, in order to determine the effectiveness of my interventions”.

Use Case 1: Global Understanding of Road Risk

The technique implemented provides the city the ability to understand the global risks of particular road segments. The following illustration provides an example of the data presented in the dashboard:



The highest ranked road segments given the particular filtering of the data is available on the left-hand side of the platform. This data can then be further filtered using filters such as the road type, the posted speed limit, whether or not the road segment was an intersection (as opposed to a unique stretch of road) and which geographic subdistrict of the city the road sits in.

For each individual road, the component risk percentiles have been displayed, making it easier to understand contributing factors to its overall rank. For example, in this intersection, road risks are most pronounced when associated with harsh braking of connected vehicles, as well as a large number of shared micromobility devices.

Of the top 25 road segments with the greatest risk, 20 of them were located on the Malmöleden. Despite having no vulnerable road users, the road ranked extremely high, with virtually all floating car observations exceeding the posted speed limit of 40kph.

Several road segments with relatively low total counts of trips are ranked highly in this assessment. When the number of heavy braking or speeding observations are normalized against the total volume of trips, it is still possible that a single triggering observation can create a “false positive” for a road segment.

Use Case 2: Collision Data vs. Floating Vehicle Data

Given the relative paucity of data about collisions (due largely to the lack of collisions), the city was particularly interested in understanding road segments where the data from connected vehicles “told a different story” than the collisions.

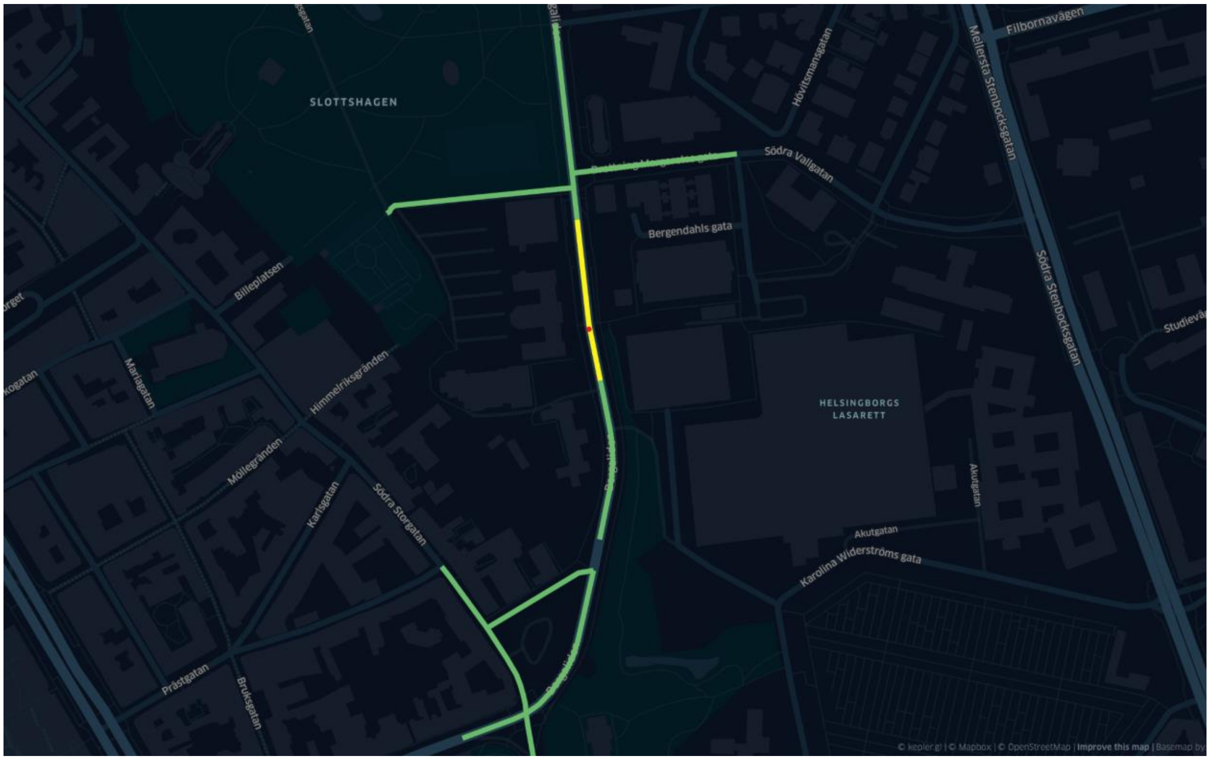
Vianova was able to calculate a “Collision differential Score” which evaluated the relationship between collision data and connected vehicle data (braking, overspeeding frequency, and overspeeding intensity). This collision differential score, like other scores, can be represented in terms of a rank, a category, or a percentile.

The collision differential score does not indicate which of the two possibilitieis causes the divergence of the two groups of elements (ie, whether there were a small number of collisions but a large number of other risk factors, or whether there were a large number of collisions, but a lack of other risk factors. Nevertheless, streets with high collision differentials have a possibility of serving as a guide for resource investments.

Use Case 3: Before and After Evaluation of Road Interventions

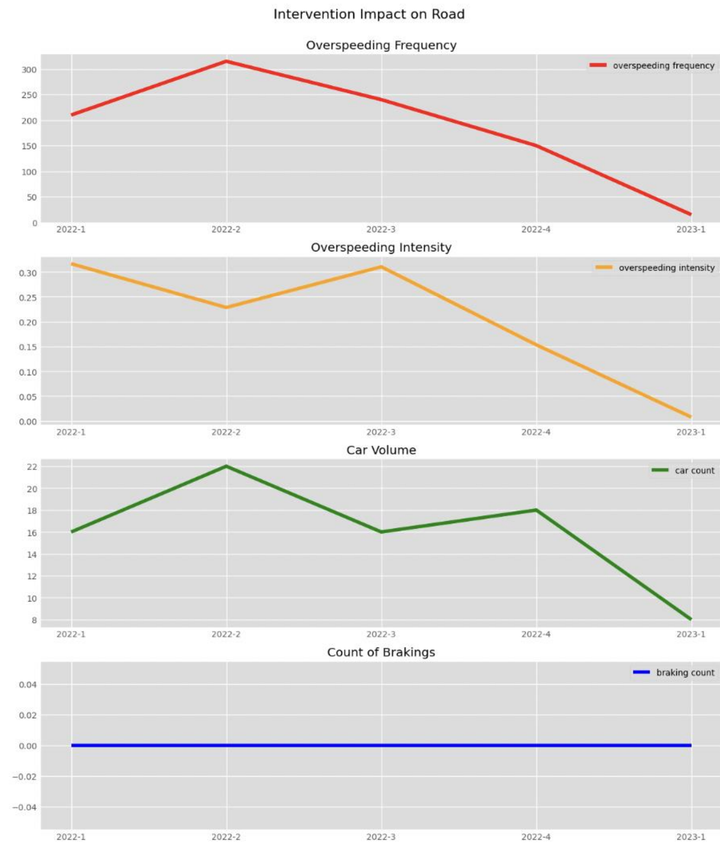
Over the course of the time period for which data was available the city of Helsingborg made several interventions into the road network in order to improve safety outcomes. These physical changes, such as adding speed humps or cyclepaths, can result in changes to the observed data on the street.

Vianova attempted to collect data over the geography that the intervention was made on, as well as the roads that form a nearby part of the network. In the analysis conducted by Vianova, the road network was extended to 200 m from the identified point of the intervention. The hope was to be able to identify unintended consequences on these roads resulting from the intervention.



In this regard, the hypothesis was again hampered by the relatively small amount of observations. Many of the interventions were made on more suburban roads, leading to relatively low numbers of observations. When grouping observations by quarter, the volume was diluted on all but the most major roads. An example analysis is below, showing declines in speeding, but over a very small number of observations (appx 20 vehicles each quarter).

The investigation demonstrates the validity of the technique for further development, particularly over a longer time window. But it also suggests the setting of a minimum threshold of representativity either in relative terms (a road segment must be no more than X% off of the median number of observations) or in absolute terms (a road segment must have at least 100 observations) to be considered as part of the model.



Feedback on User Experience

During the pilot with Helsingborg and RISE, several opportunities to optimize the experience were identified by the city users, which can support future product development for ease of use.

- There could be a more clear way to manipulate the layer of data in order to more search for the appropriate street segment.
- Many of the streets in Helsingborg with the greatest risk share the same name (ie, are different segments of the same road). Therefore, the visual experience identifying which particular segment of road is the one on a certain ranking proved somewhat difficult.
- Open Street Map data was not completely reliable for pulling information about the street (such as the street type or the speed limit) in a small city such as Helsingborg compared to larger cities with a more active community.
- Use Case 3 (the before and after analysis) requires manual implementation from the data science team at Vianova to implement, rather than being available directly on the platform itself. This situation arises from the architecture of the road safety product, as data is initially aggregated without consideration for time (ie, all available points go into the initial road scoring algorithm) and then must be dis-aggregated for particular time ranges.

Conclusions and Future Investigations

From Vianova

Overall, data accessibility in Helsingborg represented the greatest challenge to delivering a high quality experience for fulfilling the use cases identified. Below, we can find the average and median data points available per road segment per year in three pilots which Vianova was running concurrently.

	Zurich	**London**	**Helsingborg**
Average Data Points Per Road Segment Per Year	1186	8352	644
Median Data Points Per Road Segment Per Year	273	1730	15

The lack of connected vehicle data for the geography poses significant limits to the potential for robust analysis. From a product development perspective, this reality will likely force Vianova to focus on developing road safety applications for urban areas with more than 100,000 residents until the number of connected vehicle data sources increase.

Over time, as the overall share of connected vehicles increases and OEMs with high market share in Sweden such as Volvo, VW, and Toyota make connected vehicle data available to third parties such as Vianova, it will be possible to continue to increase the potential of connected vehicle insights in additional cities. These additional data providers will need to be harmonized (ie, the approaches each data provider will use to calculate harsh braking or swerving may vary), but the overall increase in connected vehicles will increase reliability of the sampling and lead to a higher degree of confidence in the representativeness of the data.

The use cases identified are relevant for the city, and the potential of using a mix of data sources (connected vehicles and a connected version of “vulnerable road users” such as bikes, scooters, and pedestrians) has genuine value. However, more time and a greater array of data sources is needed to produce insights sufficient to complement the existing traffic safety data generated from collision reports.

Within the short term, the most effective technique for incorporating the data into the city’s operations would be to track the segments where planned interventions are constructed, and then monitor their change in rankings (and the changes in the component risks) on a quarterly basis.

From City of Helsingborg

The municipality currently uses accident statistics from the national accident database STRADA. Road safety planning is done by analysing data from the national accident statistics, but also through views and cases from citizens. Accident statistics, which are an important basis for decisions on road safety improvement measures, are usually analysed five years back to be considered representative.

Through the project, the municipality has gained a greater insight into the possibility of using "floating vehicle data" (FVD). Lessons learned from this are that it is important that the number of data sources is sufficiently large and that the city can feel confident that the database is representative of traffic as a whole. It should therefore be required that a supplier can convey enough data for the particular geography requested. An important lesson is that depending on the data source, the number of data points may differ and be deficient in some countries and cities but sufficient in others. It requires a lot of effort to familiarise yourself with and understand and ensure that the data is representative.

In our comparisons between FVD and accident data, we can note that identified locations differ from each other. An initial analysis suggests that road user behaviour at FVD cannot necessarily be linked to places where personal accidents occur, which is what we mainly study. With a larger amount of FVD, one can probably identify the correlation that exists between the two data sets, but then needs to be studied more deeply to ensure this.

During the project, grading of safety has been discussed and how important different indicators such as overspeeding, braking and accidents are in relation to each other. It is challenging to determine what is unsafe traffic and what is "normal driving behavior". A continued development could mean working on standardizing this. A further development could involve designing so that the system alerts when it detects that something changes or a negative trend on a street. This could be an important complement to the municipality's traditional way of working.

The automatic loading of underlying map data provides the opportunity to drill down into other data that is relevant in the context. We have noted that there are shortcomings in the data warehouse on speed where we see that some streets indicate large amounts of speeding, but where it has rather been a matter of the speed regulation not being correct in the data warehouse.

If it is possible to identify relevant correlation between personal injuries and FVD, a tool of this kind can be an important support in the cities' analysis work. If the data can be updated in real time, there are great opportunities to supplement today's planning methods with current situation analyses and thus be even more agile in our road safety work.