Traffic Accidents in the Netherlands

SAS Case Study

Abstract

This case study has been designed to expand the development of skills within the analytics lifecycle. Skills addressed include exploring new data, creating data visualizations, interpreting insights, and communicating key conclusions.

SAS Software

This case study targets the use of SAS® Visual Analytics on the SAS® Viya for Learners platform – an online resource, free to use for academics, that requires no set-up or infrastructure.

Industry Alignment

This case study aligns with the traffic collision and safety industry. Although the focus is the traffic sector, the skills learned are easy transferrable to other industries.

Table of Contents

Notes and Requirements

Learning Objectives

This case study will challenge learners to:
- Employ basic statistics to answer business related questions
- Build foundational data analysis and business intelligence skills

Students will be able to use SAS Visual Analytics to:
- Explore Data: examine key variables and look for pertinent information in the data
- Data analysis: exploration, insight gathering, and prioritization of information
- Develop an Interactive Report: present results in a structured and meaningful design

Estimated Completion Time

This case study can be completed as an individual or as a team-based assignment. While completion time can vary, the average estimated amount of time to complete this case study is approximately 1-3 hours.

Experience Level

The intended audience is not limited to students with backgrounds in mathematics or statistics. Undergraduate or postgraduate students across a broad range of disciplines can complete this case study, where sufficient exposure to SAS Visual Analytics can be provided in advance of the case study.

Prerequisite Knowledge

Software
Students should be familiar with the foundational skills of SAS Visual Analytics, similar to those taught in the SAS Visual Analytics 1 for SAS Viya: Basics course.

Content Knowledge
For best success, students should have a basic understanding of:
- Data visualization
- Dashboard design theory
- Entry level statistics

Additional Notes

This case study is separated into two sections: an *Intermediate Level* and an *Advanced Level*. While both sections contain the same content, the Advanced Level is designed with less step-by-step instruction to navigate the case study. Faculty should choose one level to give to their students to complete.

Notes to instructor regarding the three phases of this case study:

1. **Data Preparation:** Investing time exploring and getting to understand the data before commencing the work on the tasks is invaluable. What are the values or range of values held in each field? Are there implicit hierarchies within the data? Do we have the data we need to conduct the analysis or assist report design? What should be the default aggregations for the numeric fields? Exploring these questions will help in the following analysis and report development work.

2. **Data Analysis**: When considering the requirements for the executive summary, there is significant scope both in terms of the data that can be presented, or the methods used to communicate the insight gained.

   It is critical to decide what visuals and methods are important from a strategic perspective. What would the department team really find of value and how can this be communicated concisely?

3. **Interactive Report Design:** The objective of this element of the case study is to develop a report that allows the department to view the data relating to their region.

   The report will need to be able to quickly provide answers to questions raised at an executive level, whilst also ensuring there is visibility on any problem areas the department needs to take actions. General guidance would be to include (but is not limited to) the topics highlighted in the requirements.

   Overall, there should be an engaging, elegant, and logical structure to the dashboard design that allows the student to identify these things at a glance.

   Students could either be asked to present their findings or submit as coursework.

Required Setup
Access to SAS Viya for Learners platform.

Data Source

Introduction
The case study has been developed using the **ONGEVALLEN2016_EN** dataset available on SAS® Viya for Learners.

The path to the dataset location is as follows: *Cas-v4e065-default > USECASES >* **ONGEVALLEN2016_EN**

Description of Variables
The variables used for this exercise are:

Variable Name	Variable label	Type

VKL_End	Accident ending	Character
Max_Age	Age oldest casualty	Numeric
Min_Age	Age youngest casualty	Numeric
RH	Amount precipitation (0.1mm)	Numeric
FH	Average wind speed (0.1 m/s)	Numeric
xyCounty_Lat	County latitude	Numeric
xyCounty_Lon	County longitude	Numeric
County_Name	County name	Character
DR	Duration precipitation (0.1 hrs)	Numeric
SQ	Duration sunshine (0.1 hrs)	Numeric
FX	Highest wind gust (0.1 m/s)	Numeric
Holiday_General	Holiday general	Numeric
Holiday_School	Holiday schools	Numeric
VKL_Number	Identification accident	Numeric
ID_Location	Identification location	Numeric
xyLocation_Lat	Location latitude	Numeric
xyLocation_Lon	Location longitude	Numeric
Max_Speed	Maximum speed indicated	Numeric
Num_Bike	Number bikes	Numeric
Num_Casualties	Number casualties	Numeric
Num_Hit_and_Run	Number hit-and-run	Numeric
Num_Injured_Hospital	Number injured hospital	Numeric
Num_Injured_Misc	Number injured miscellaneous	Numeric
Num_ART8	Number intoxicated (liquor/drugs)	Numeric
Num_Lethal	Number lethal casualties	Numeric
Num_Motor	Number motorcycles	Numeric
Num_no_APK	Number no APK (vehicle testing)	Numeric
Num_Parties	Number parties involved	Numeric
Num_Passenger_Car	Number passenger cars	Numeric
Num_Pedestrian	Number pedestrians	Numeric
Num_Truck	Number trucks	Numeric
Num_not_Insured	Number uninsured	Numeric
Num_Van	Number vans	Numeric

xyProvince_Lat	Province latitude	Numeric
xyProvince_Lon	Province longitude	Numeric
Province_Name	Province name	Character
ROAD_NUM	Road number	Character
ROAD_DESC	Road situation	Character
T	Temperature (0.1 Celsius)	Numeric
Time	Time accident	Character
END_DESC	Type accident	Character
BEBKOM	Urban/built-up area	Character
VV	View horizontal	Numeric
WGD_CODE_1	Weather Condition	Character

Traffic in the Netherlands: Intermediate Level

The department for 'Traffic Safety' investigates the nature and causes of traffic accidents. The general objective is to evaluate the claims filed by the insured, but also to investigate how to prevent traffic accidents.

In this case study there are several scenarios to choose from. Of all the scenarios listed below, choose <u>one</u> to investigate and report on:
- Where were the accidents?
- When were the accidents?
- Who were offenders (e.g., Type of offenders: no APK, uninsured, hit-and-run)?
- Who were casualties (e.g., Type of casualty (material, injured, lethal), age victims)?
- What types of accidents happened?

Part 1: Exploration

Investigate the information of the accidents according to the scenario you selected above. Try to find interesting information you want to share with the department using the following steps:

1. Check how many observations and measures there are.
2. Check to see if you will need to filter the data in advance (data source filter) before data analysis.
3. Decide if you will need to hide data items you will not use.
4. Consider which categories (categorical variables) will be needed. Decide if you will need hierarchies and/or geographies.
5. Change any measure to categories if needed.
6. Check for statistics of interest (e.g., minimum, maximum, missing values).
7. Perform univariate analyses of the measures of interest (e.g., create histograms etc...).
8. Conduct bivariate analyses to investigate relationships between measures and categories.
9. Conduct multivariate analyses to check associations between measures.

Part 2: Report by Scenario

Consider the following steps to analyze the data and construct a report according to the scenario you selected at the beginning of the case study (you do not have to analyze the data for every scenario, just for the scenario you chose):

Scenario 1 – Where?
1. Duplicate the *Province name* variable and rename this item *Province*. Change the classification to geography.
2. Select Latitude and Longitude under Geography Data.

3. Select *Province latitude* and *Province longitude* for the latitude and longitude slots.
4. Use the World Geodetic System under for the coordinate space.
5. Calculate a new data item named *Ratio lethal*.
   a. Click <u>Text</u> and enter the following:
   Sum [_ByGroup_] ('Number lethal casualties'n) / Count [_ByGroup_] ('Number accidents'n)

6. Rename *Frequency* to *Number accidents.*
7. Create a Geo Map by *Province* using *Ratio Lethal* for color and *Number accidents* for size.
8. Calculate a new data item named *Ratio casualties*.
   a. Click <u>Text</u> and type the following:
   Sum [_ByGroup_] ('Number casualties'n) / Count [_ByGroup_] ('Number accidents'n)

9. Create a List Table using the following variables and add it to the same page as the Geo map:
   > *Province*
   > *Number accidents*
   > *Number casualties*
   > *Ratio casualties*
   > *Number lethal casualties*
   > *Ratio lethal*
10. Create a Bar Chart with *Number accidents* by *Road situation* and add it to the report page.

Scenario 2 – When?
1. Make a duplicate of *Date accident* and rename this item *Month accident.* Change the format to a month formatting of your choice.
2. Rename *Frequency* to *Number accidents.*
3. Create a Dual Axis Time Series Plot with *Month accident* by *Number accidents* and *Number casualties*.
4. Make a duplicate of *Date accident* and rename this item *Weekday accident*. Change the format to Day of Week.
5. Create another Dual Axis Time Series Plot with *Weekday accident* by *Number accidents* and *Number casualties* on the report page.
6. Calculate a new data item named *Ratio casualties*.
   a. Click <u>Text</u> and type the following:
   Sum [_ByGroup_] ('Number casualties'n) / Count [_ByGroup_] ('Number accidents'n)
7. Create a Time Series Plot of *Ratio casualties* per *Month accident* on the report page.
8. Make a duplicate of *Time accident* data item and rename it *Hour accident*. Change the format to an hour formatting of your choice.
9. Create a Dual Axis Line Chart with *Hour accident* by *Number accidents* and *Number casualties* on the report page.

10. Lastly, add a Button Bar above the Dual Axis Line Chart. Use *Weekday accident* as the category and add an action so that the Button Bar filters the Dual Axis Line Chart.

Scenario 3 – Offenders?
1. Duplicate the *Province name* variable and rename this item *Province*. Change the classification to geography.
2. Select Latitude and Longitude under Geography Data.
3. Select *Province latitude* and *Province longitude* for the latitude and longitude slots.
4. Use the World Geodetic System under for the coordinate space.
5. Rename *Frequency* to *Number accidents.*
6. Add a stacking container to the report page.
7. Add a Geo Map by *Province* to the container using *Number accidents* for size and *Number uninsured* for color.
8. Add another Geo Map by *Province* to the container using *Number accidents* for size and *Number no APK (vehicle testing)* for color.
9. Add a Geo Map by *Province* to the container using *Number accidents* for size and *Number intoxicated (liquor/drugs)* for color.
10. Add a final Geo Map by *Province* to the container using *Number accidents* for size and *Number hit-and-run* for color.
11. Rename each Geo Map with a more descriptive name.

Scenario 4 – Casualties?
1. Rename *Frequency* to *Number accidents.*
2. Create a custom category named *Youngest casualty* based on *Age youngest casualty* with the following names and their corresponding values:
   a. **0-4**
   b. **5-10**
   c. **11-15**
   d. **16-20**
   e. **21-30**
   f. **31-60**
   g. **61-75**
   h. **76-older**
   i. Missing values should display as missing
3. Create a Bar Chart of *Number accidents* by *Youngest casualty*. Exclude missing values of *Youngest casualty* and sort the chart with ascending *Youngest casualty*.
4. Make a duplicate of the *Date accident* variable and change the format to a weekday formatting of your choice. Rename the duplicate variable *Weekday accident*.
5. Using dual axes, add another Bar Chart to the report page with *Number casualties* and *Number accidents* by *Weekday accident*.
6. Make a duplicate of *Time accident* and rename the item *Hour accident*. Change the format to an hour formatting of your choice.

7. Add a Dual-Axis Line Chart to the report page with *Hour* accident as X, *Number accidents* as Y1, and *Number casualties* as Y2.
8. Use the first bar chart (Step 3 above) to filter the other visualizations.

<u>Scenario 5 – Type of Accidents?</u>
1. Rename *Frequency* to *Number accidents.*
2. Create a Hierarchy item named *Type accident hierarchy* using *Road situation* and *Type accident*, respectively.
3. Calculate a new data item named *Ratio casualties*.
   a. Click <u>Text</u> and type the following:

   Sum [_ByGroup_] ('Number casualties'n) / Count [_ByGroup_] ('Number accidents'n)

4. Create a Treemap using the *Type accident hierarchy* item with *Number accidents* for size and *Ratio casualties* for color.
5. Rename *Frequency Percent* to *Percent accidents*.
6. Add a Bar Chart of *Percent accidents* by *Accident ending* to the report page.
7. Appy a display rule to *Accident ending* in the bar chart with Red for **Lethal**, Blue for **Material only**, and Orange otherwise.
8. Apply a custom sort to *Accident ending* in the following order: **Material only** – **Injury Hospital** – **Injury First-Aid** – **Injury Misc** – **Lethal**. Update the bar chart to ensure it is sorted in the proper order.
9. Create a custom category named *Weather* based on *Weather condition* using the following labels:
   i. **D** = Dry
   ii. **H** = Strong wind
   iii. **M** = Mist
   iv. **R** = Rain
   v. **S** = Snow/Hail
   vi. All remaining values = Unknown
10. Add a Pie Chart of *Number accidents* by *Weather*. Ensure all values are being displayed.
11. Filter the Bar Chart with the Pie Chart.
12. Filter the Bar Chart and the Pie Chart with the Treemap.

Traffic in the Netherlands: Advanced Level

The department for Traffic Safety investigates the nature and causes of traffic accidents. The general objective is to evaluate the claims filed by the insured, but also to investigate how to prevent traffic accidents.

In this case study there are several scenarios to choose from. Of all the scenarios listed below, choose <u>one</u> to investigate and report on:

- Where were the accidents?
- When were the accidents?
- Who were offenders (e.g., Type of offenders: no APK, uninsured, hit-and-run)?
- Who were casualties (e.g., Type of casualty (material, injured, lethal), age victims)?
- What types of accidents happened?

Part 1: Exploration

Investigate the information of the accidents according to one or more of the scenarios mentioned above. Try to find interesting information you want to share with the department. Thoroughly examine the variables in the dataset, consider variables that may need to be created or modified, and conduct basic univariate, bivariate, and multivariate analyses to better understand the data.

Part 2: Report by Scenario

Given the scenario(s) you choose, construct a report with at least two sections or pages: an overview and a detailed one, making the latter available on demand by clicking an object on the 1st page. Incorporate a mixture of graphs, charts, tables, and interactive filters to convey key insights.

Regardless of which scenario you decide to focus on, include answers to the following general questions in your report:

1. Where are the majority of accidents?
2. Where are quite a number of (fatal) casualties?
3. What are the circumstances at the scene (road conditions, weather conditions, etc.)
4. What times are more dangerous in terms of number of (lethal) accidents?
5. In what kind of accidents (head/tail, straight road, etc.) are casualties?
6. What offending parties are involved?
7. How is the age of the casualties distributed?

Appendix

Appendix A: Access Software

Steps to accessing SAS Viya for Learners for the first time:
1. Navigate to the SAS Viya for Learners webpage,
   https://www.sas.com/en_us/software/viya-for-learners.html
2. Click the "Access for Educators" or "Access for Students" button based on your role.
3. Log in with your SAS Profile that is linked to an academic affiliation. If you don't have a SAS Profile, click here to set one up.
4. Access the software by clicking on "Launch SAS Viya for Learners 3.5"

For future sessions with VFL, you can access VFL by visiting https://vle.sas.com/vfl.

Contact us at academic@sas.com if you have any questions about access.

Appendix B: Helpful Documentation

Here are two options for additional help and guidance:
1. **SAS Help Center** in SAS VFL
   a. It's been lurking in the upper-right-hand corner all long!
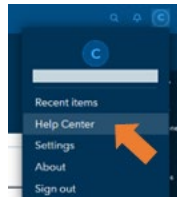
   b. Click the button to go the **SAS Help Center** and the **SAS Documentation**.
2. **SAS Video Library**
   a. Prefer videos instead? We've got you covered here:
   https://video.sas.com/
   b. Check out the **How To Tutorials** or simply **Search Videos**.

Appendix C: Recommended Learning

The SAS Global Academic Program offers free e-learning courses for students to learn SAS through the Student Skill Builder. The following e-learning courses and paths available are recommended to help with this activity:
- Create Visualizations of Analytics E-Learning Path
- SAS Visual Analytics 1 for SAS Viya: Basics (8.5)
- SAS Visual Analytics 2 for SAS Viya: Advanced (8.5)

Alternatively, the SAS Learning Subscription grants you access to an extensive library of SAS eLearning courses. Sign up for a free 30-day trial.

Appendix D: Data Information

The data set contains 124,992 rows. Each row is a traffic accident, identified by data item Identification accident. Each column is an attribute of the traffic accident.

Due course, there will be need for new data items and structures to be built. In the text of the case study those will be underlined per scenario.

Most of the names and values of the data items are self-explanatory, except for these below:
Weather condition (WGD_CODE_1)

D	Dry
H	Strong wind
M	Mist
R	Rain
S	Snow/Hail
O	Unknown

Urban/built-up area (BEBKOM)

BI	Inside urban area
BU	Outside urban area

Appendix E: Solutions

The following are possible solutions.
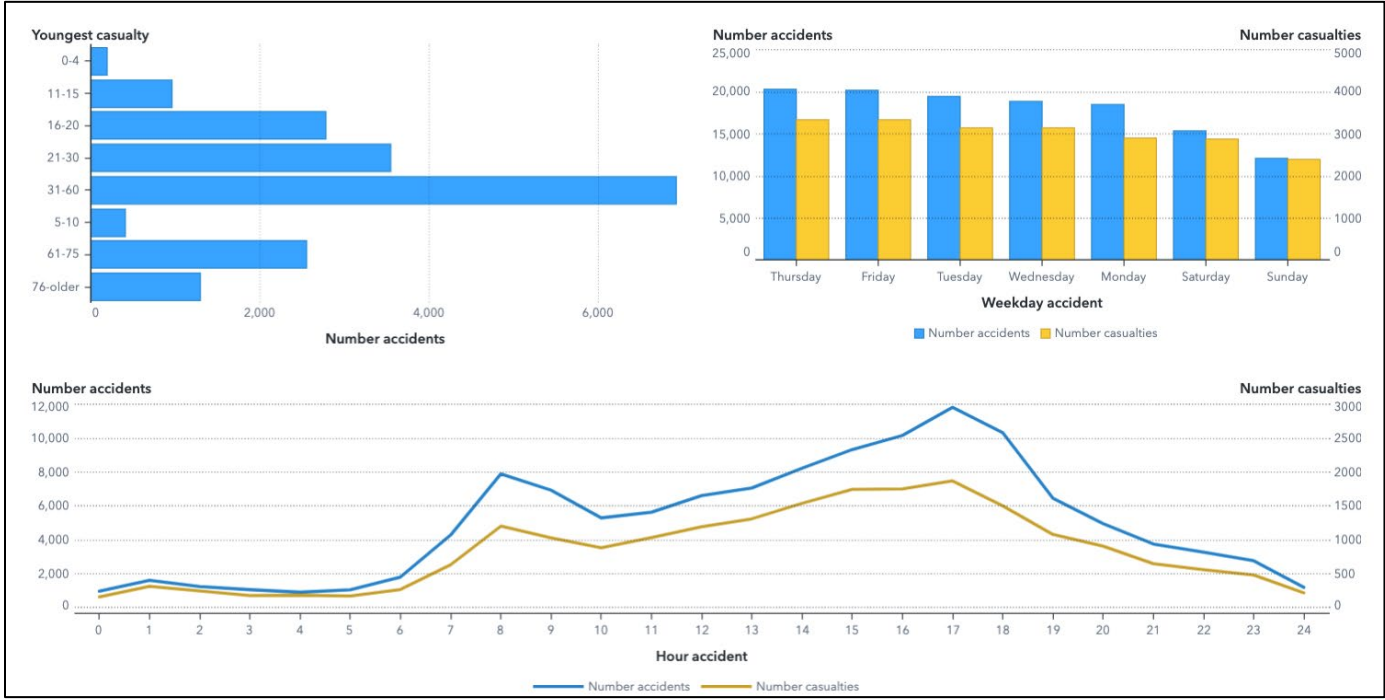
Scenario 1 Example Solution
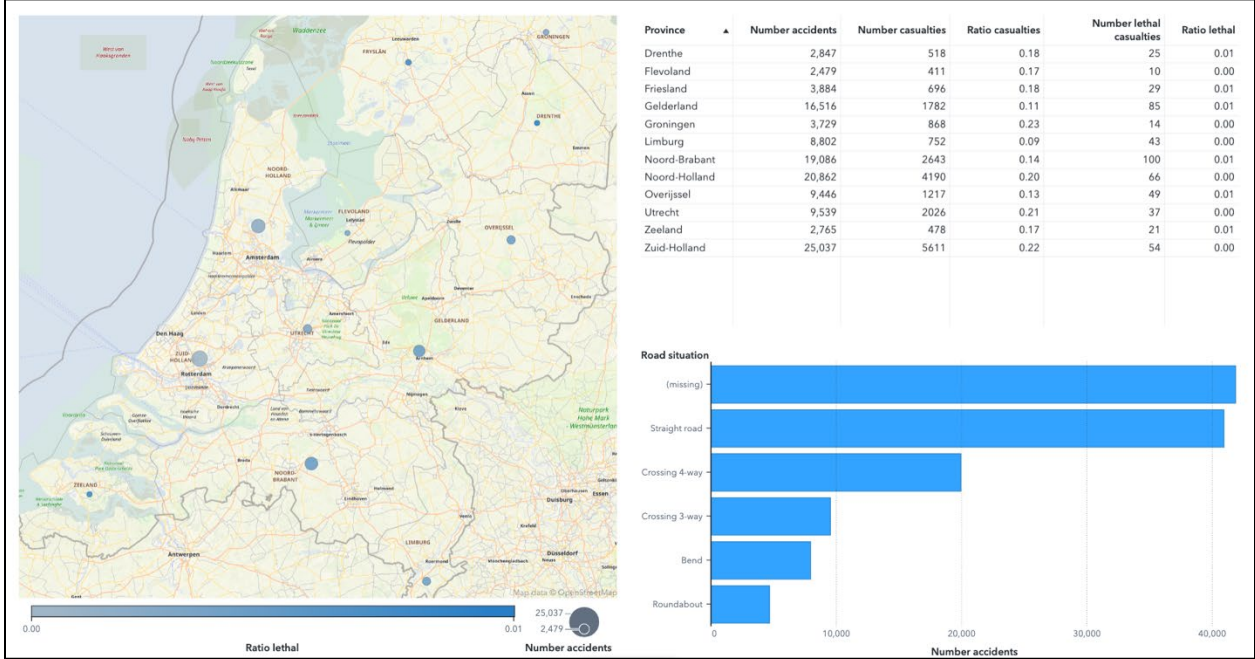
Scenario 2 Example Solution

Scenario 3 Example Solution

Scenario 4 Example Solution

Scenario 5 Example Solution