

# Data Science Lab 2

<b>Team nr:</b> 2	<b>Student 1:</b> Åmund Grimstad	<b>IST nr:</b> 1116675
	<b>Student 2:</b> Arthur de Arruda Chau	<b>IST nr:</b> 1116090
	<b>Student 3:</b> Benjamin Raymond Kuhn	<b>IST nr:</b> 1115778
	<b>Student 4:</b> João Rafael Freitas Lourenço	<b>IST nr:</b> 425699

## CLASSIFICATION

### 1 DATA PROFILING

#### *Data Dimensionality*

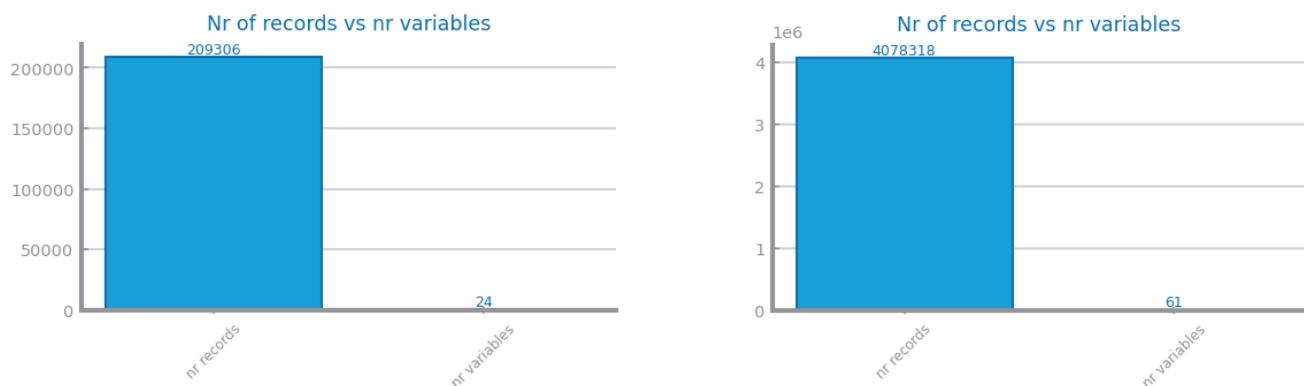


Figure 1: Nr Records x Nr variables for Traffic Accident Data (left) and Flight Cancellation Data (right)

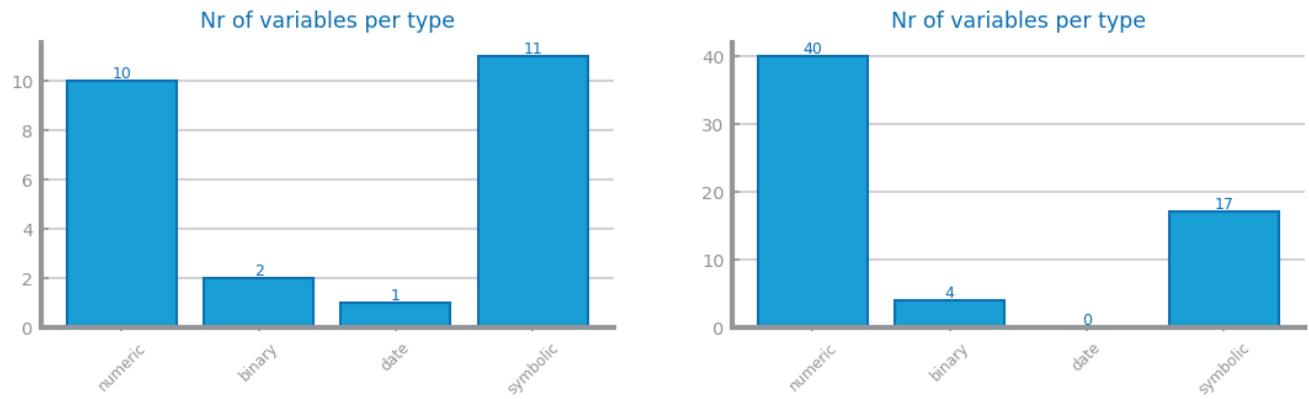


Figure 2: Nr variables per type for Traffic Accident Data (left) and Flight Cancellation Data (right)

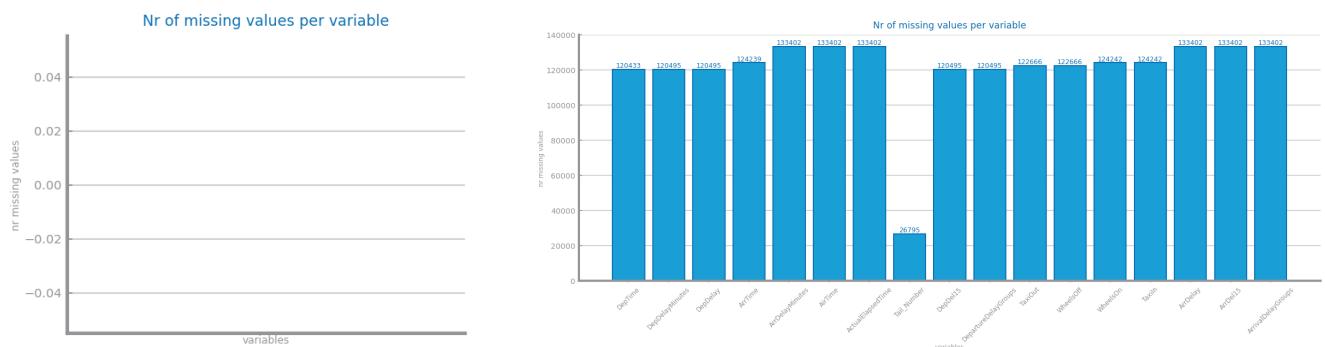


Figure 3: Nr missing values for Traffic Accident Data (left) and Flight Cancellation Data (right)

## Data Distribution

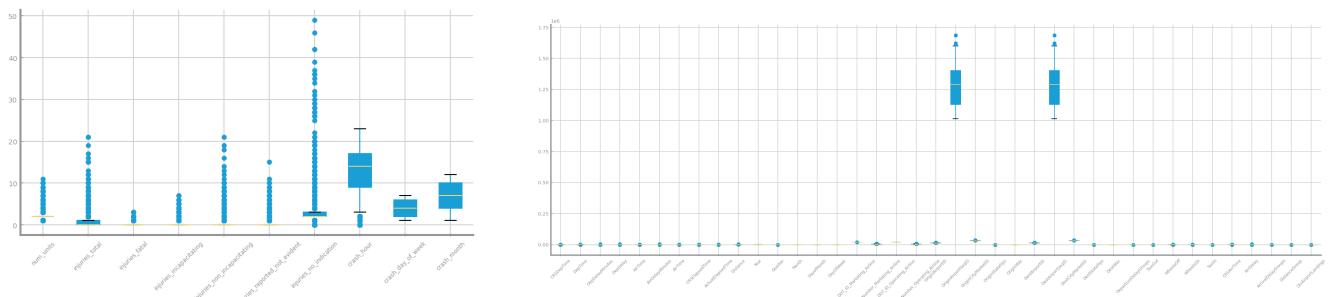


Figure 4: Global boxplots for Traffic Accident Data (left) and Flight Cancellation Data (right)

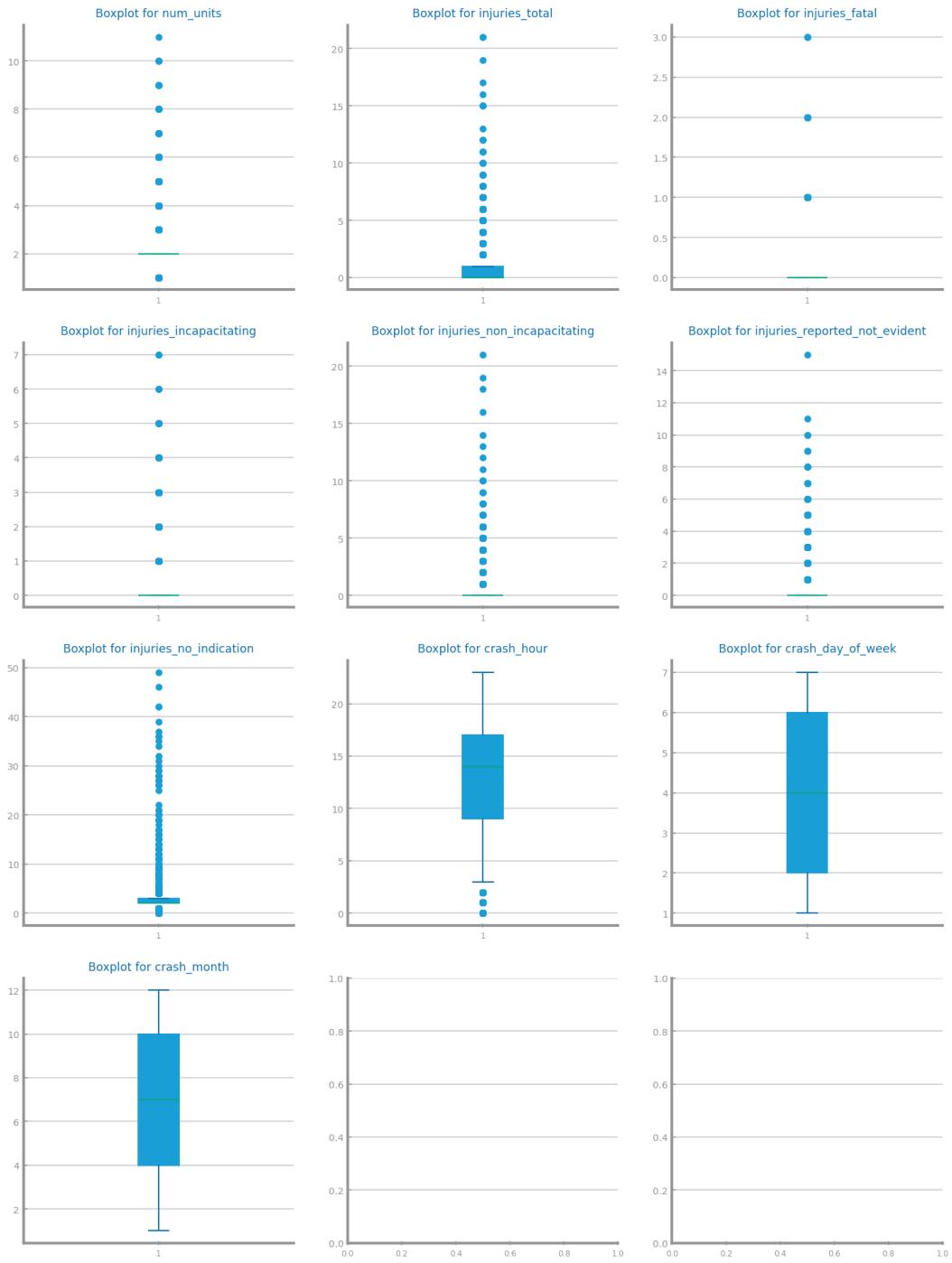


Figure 5: Single variables boxplots for Traffic Accident Data



Figure 6: Single variables boxplots for Flight Cancellation Data

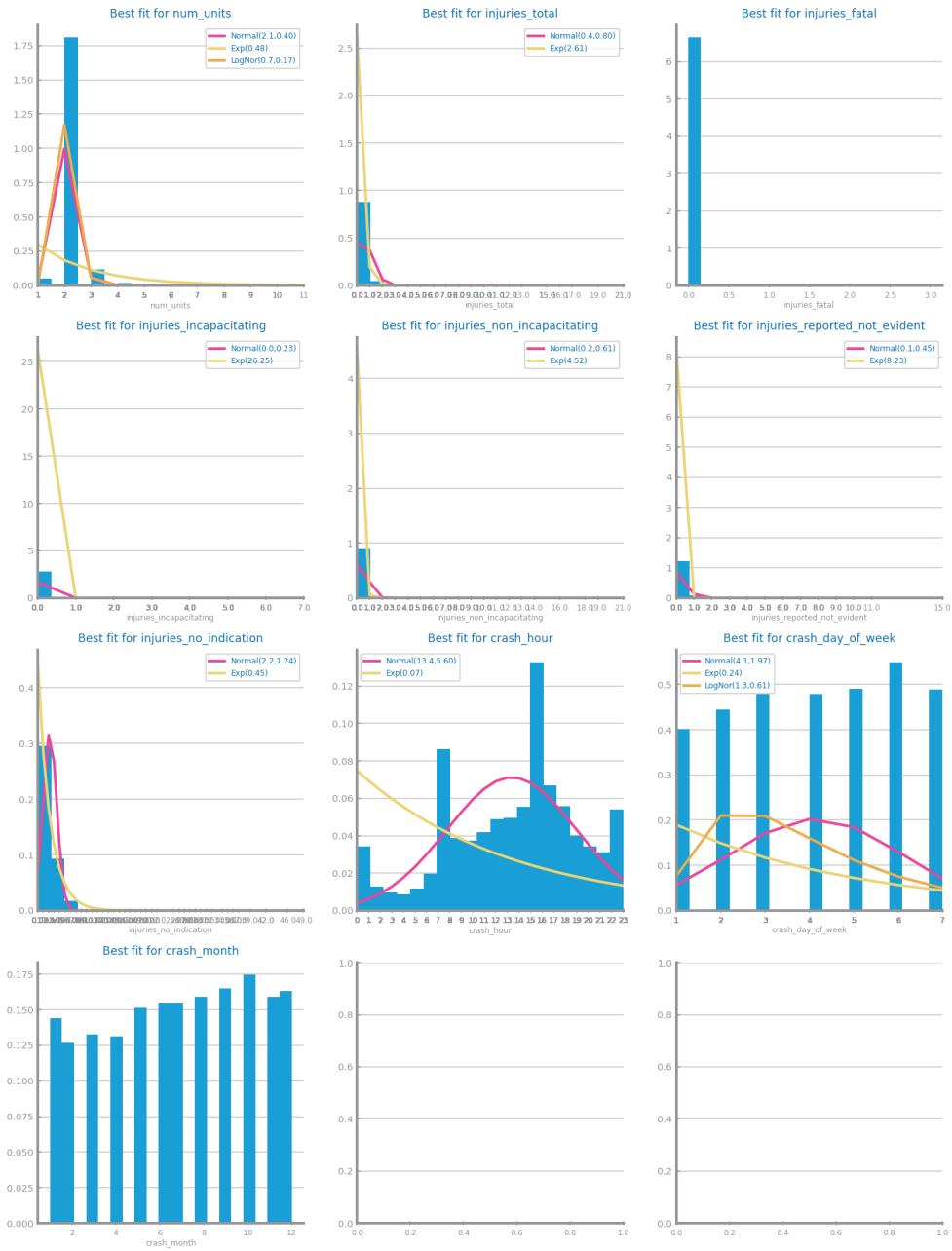


Figure 7: Histograms for Traffic Accident Data - numeric

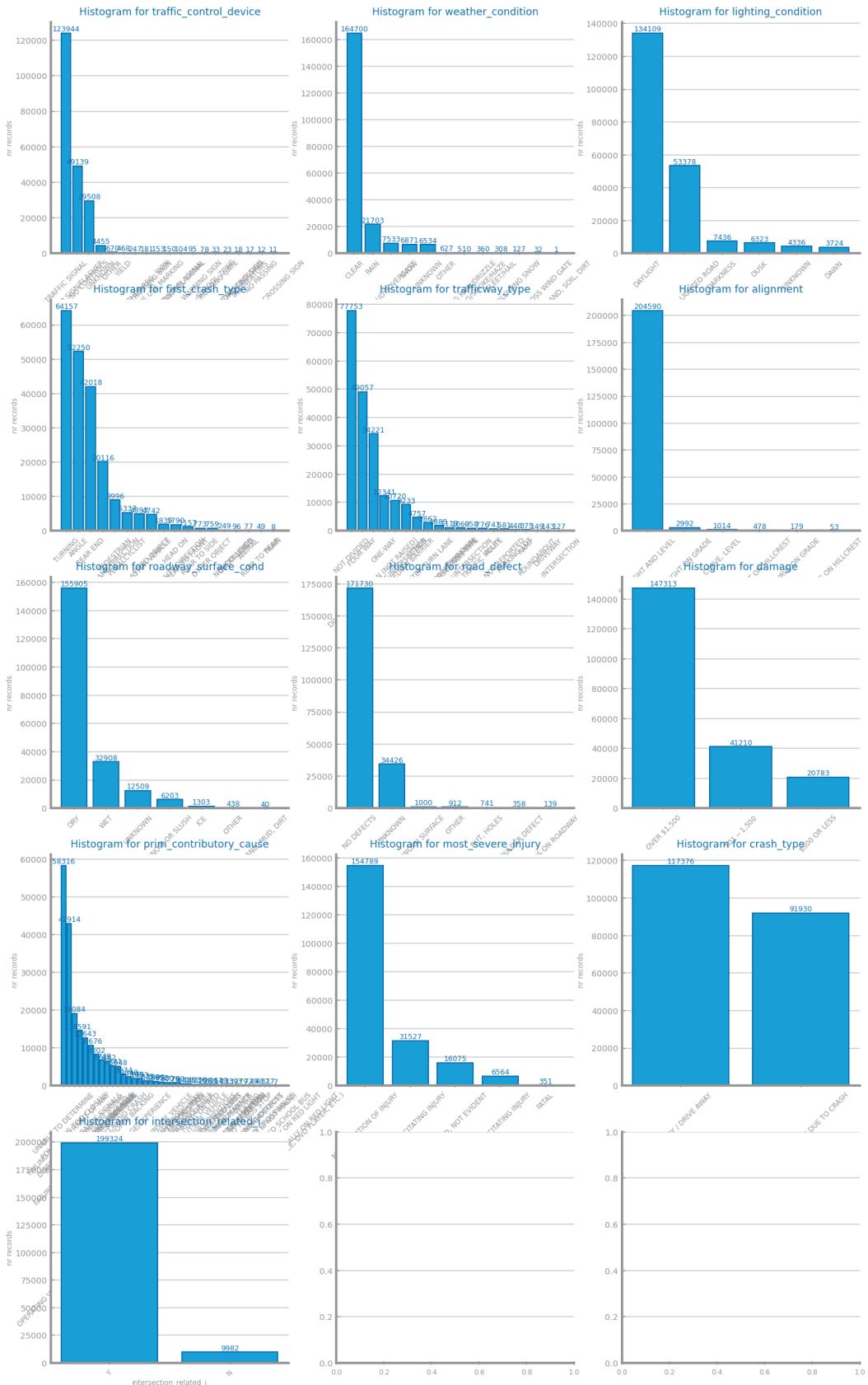


Figure 8: Histograms for Traffic Accident Data - symbolic

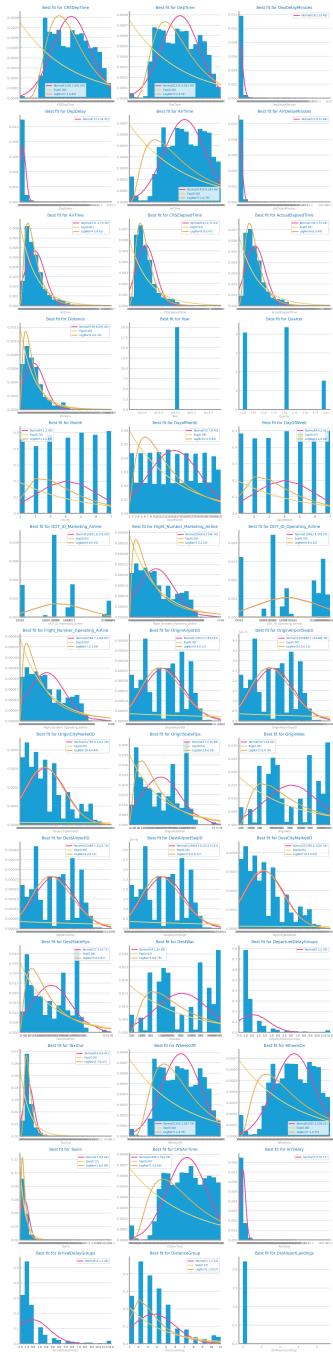


Figure 9: Histograms for Flight Cancellation Data - numeric

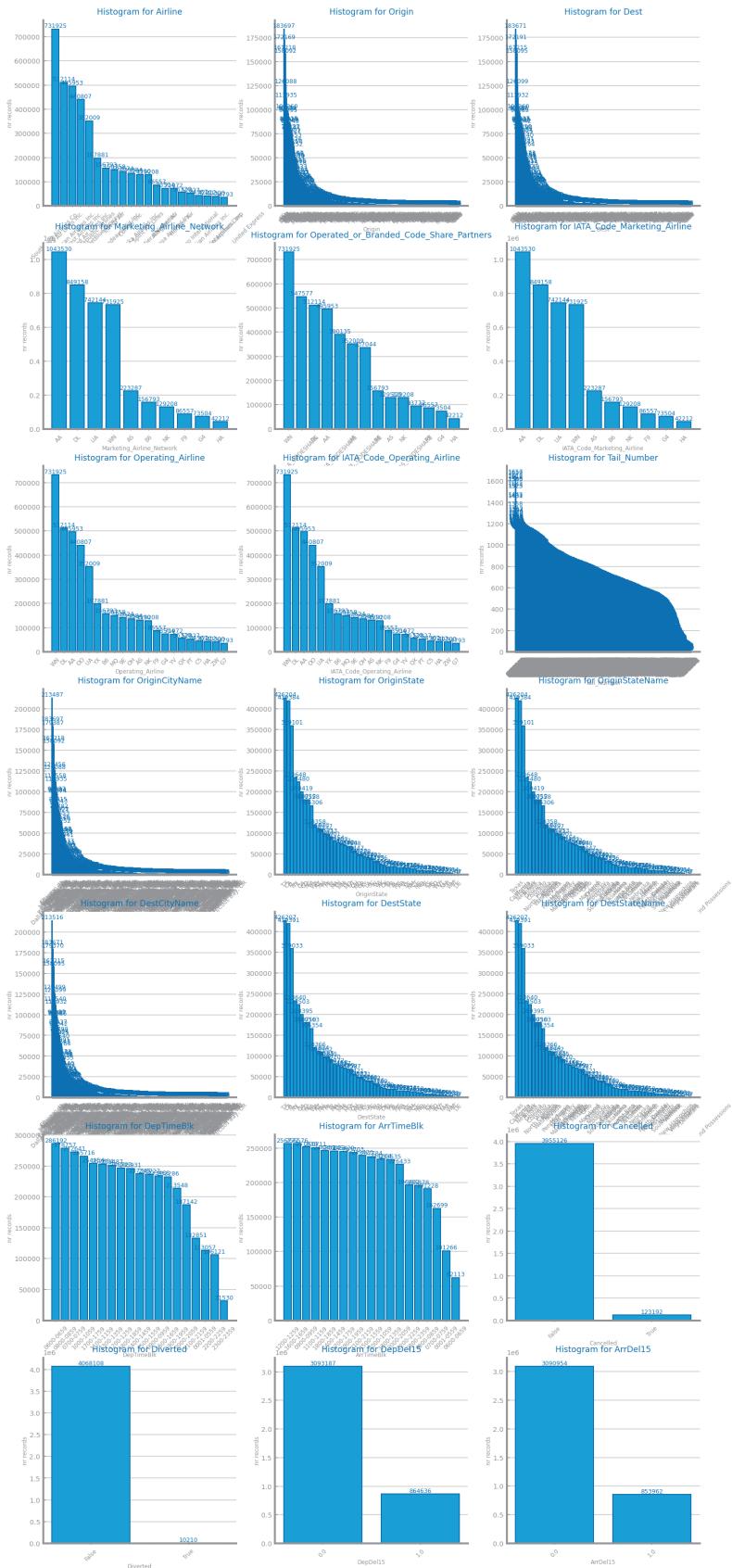
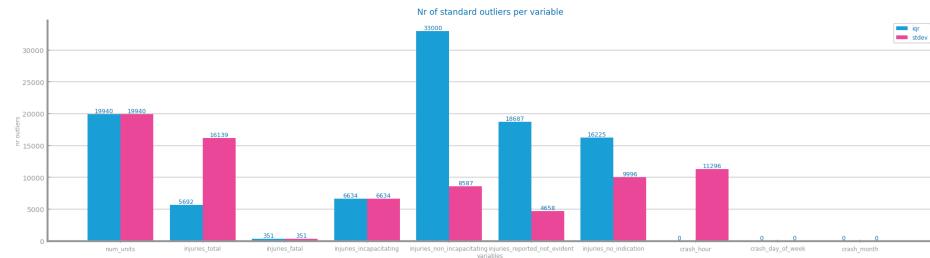
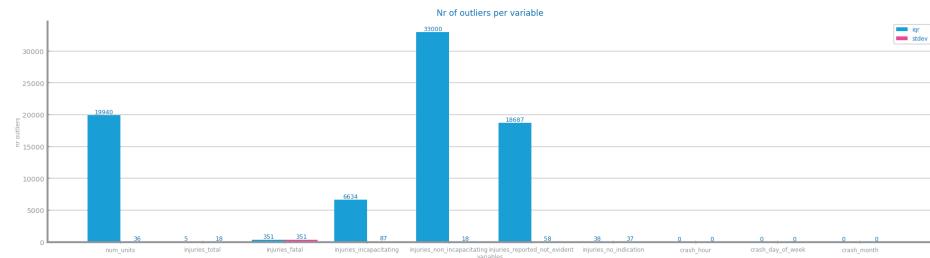


Figure 10: Histograms for Flight Cancellation Data - symbolic

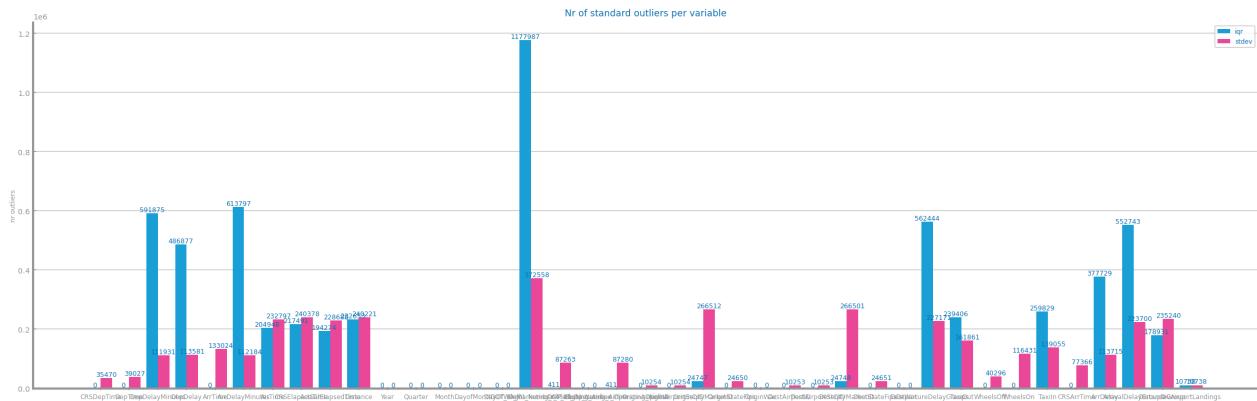


(a) standard

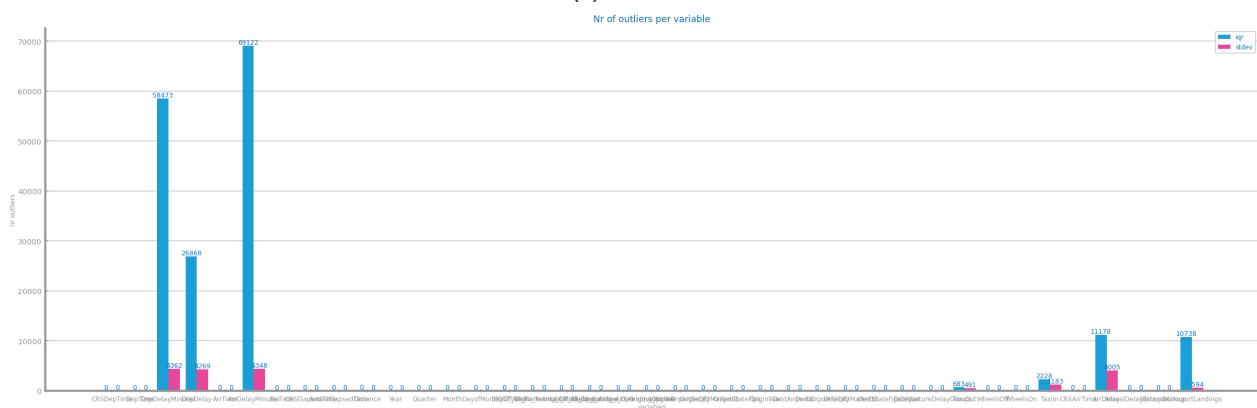


(b) tuned

Figure 11: Outliers study Traffic Accident Data



(a) standard



(b) tuned

Figure 12: Outliers study Flight Cancellation Data

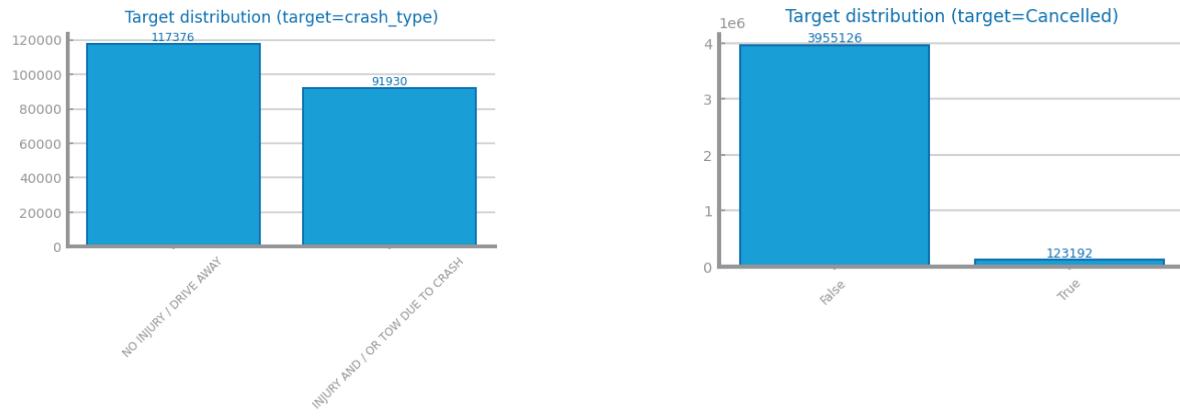


Figure 13: Class distribution for Traffic Accident Data (left) and Flight Cancellation Data (right)

## *Data Granularity*

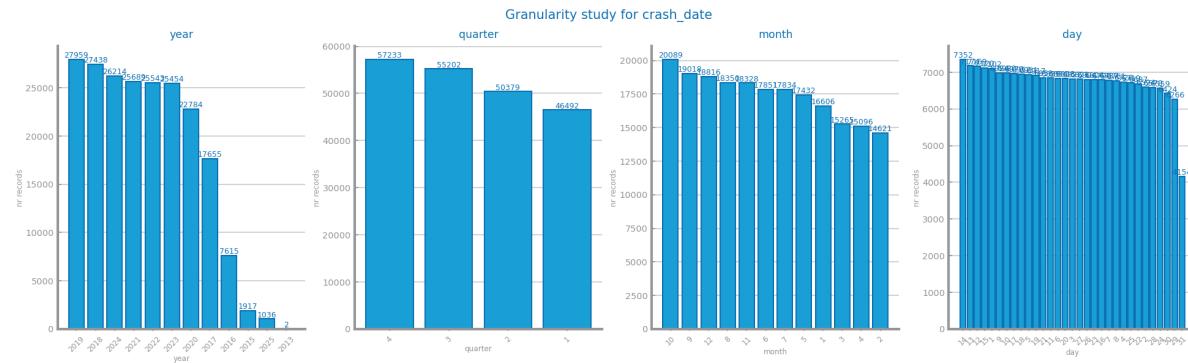


Figure 14: Granularity analysis for Traffic Accident Data

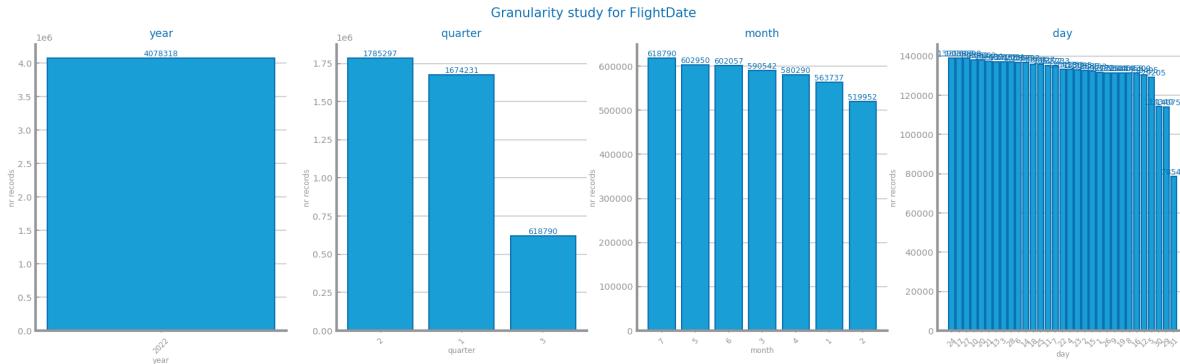


Figure 15: Granularity analysis for Flight Cancellation Data

## *Data Sparsity*

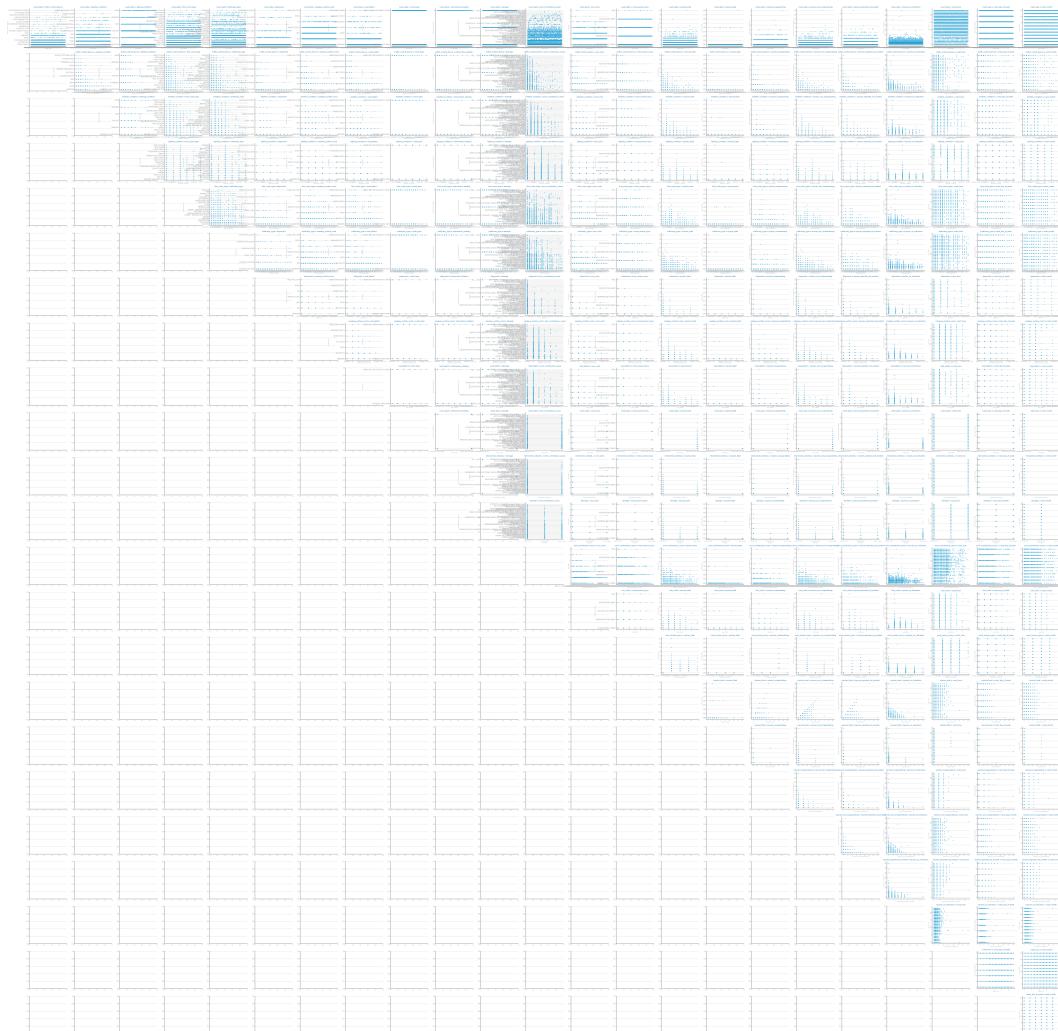


Figure 16: Sparsity analysis for Traffic Accident Data



Figure 17: Sparsity analysis per class for Traffic Accident Data

#### Sparsity analysis for Flight Cancellation Data

[https://github.com/ch4uz/ds-project/blob/organized-structure/classification/ds2\\_flight\\_cancellation/charts/lab2\\_profiling/flight\\_cancellations\\_sparsity\\_study.png](https://github.com/ch4uz/ds-project/blob/organized-structure/classification/ds2_flight_cancellation/charts/lab2_profiling/flight_cancellations_sparsity_study.png)

#### Sparsity analysis per class for Flight Cancellation Data

[https://github.com/ch4uz/ds-project/blob/organized-structure/classification/ds2\\_flight\\_cancellation/charts/lab2\\_profiling/flight\\_cancellations\\_sparsity\\_per\\_class\\_study.png](https://github.com/ch4uz/ds-project/blob/organized-structure/classification/ds2_flight_cancellation/charts/lab2_profiling/flight_cancellations_sparsity_per_class_study.png)

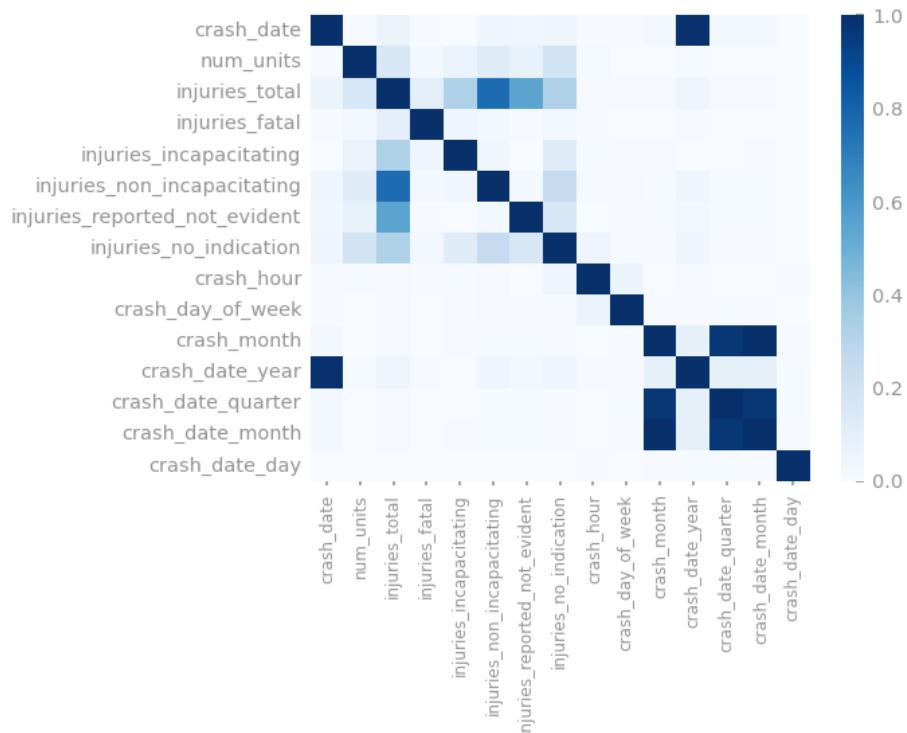


Figure 18: Correlation analysis for Traffic Accident Data

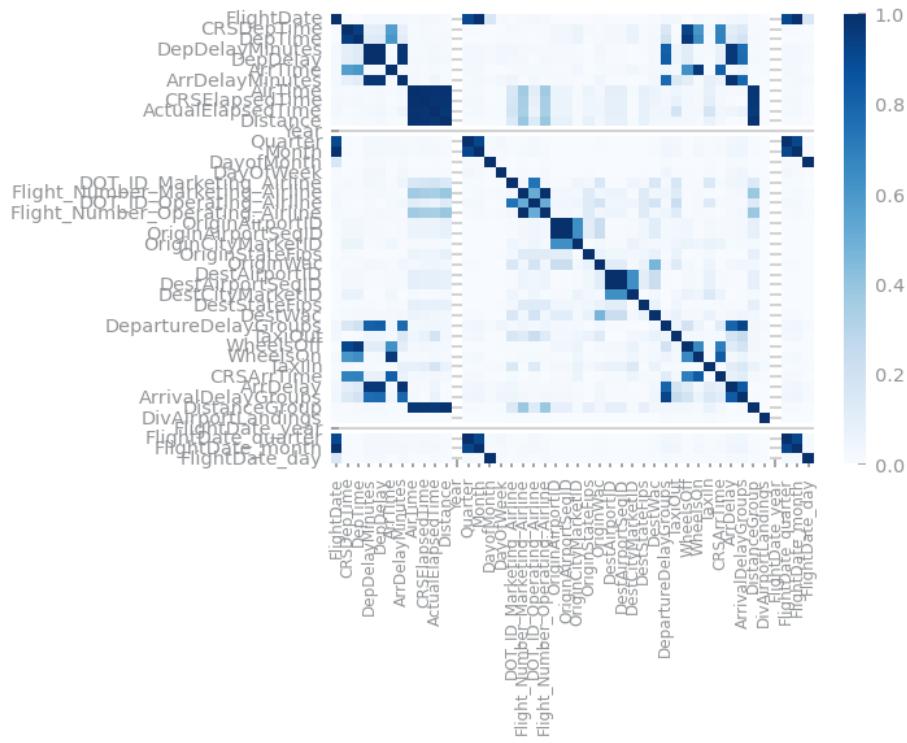


Figure 19: Correlation analysis for Flight Cancellation Data