

Data Science Lab 2

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

CLASSIFICATION

1 DATA PROFILING

Data Dimensionality

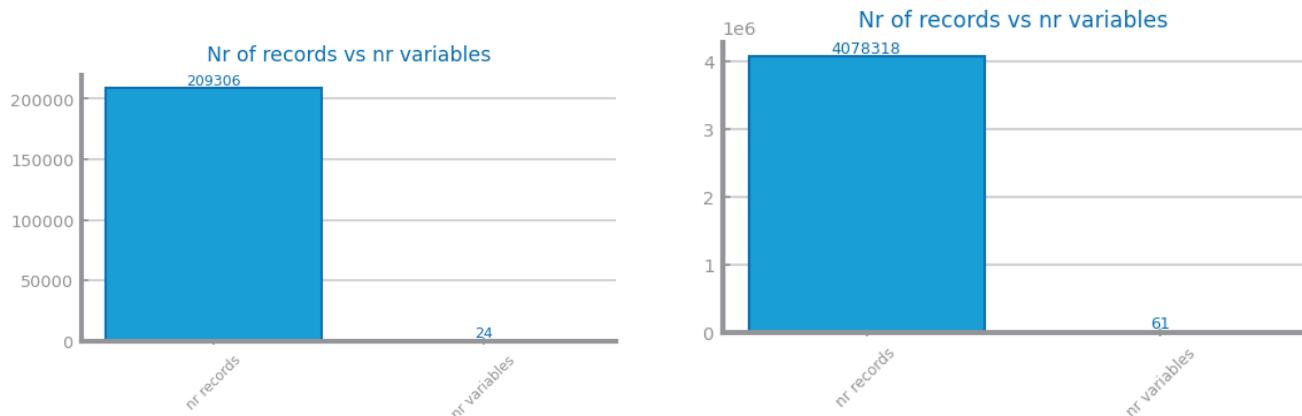


Figure 1: Nr Records x Nr variables for Traffic Accident Data (left) and Flight Cancellation Data (right)

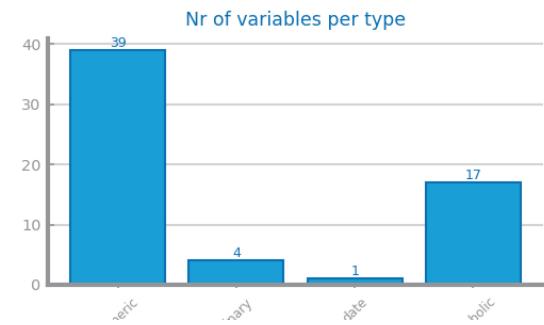
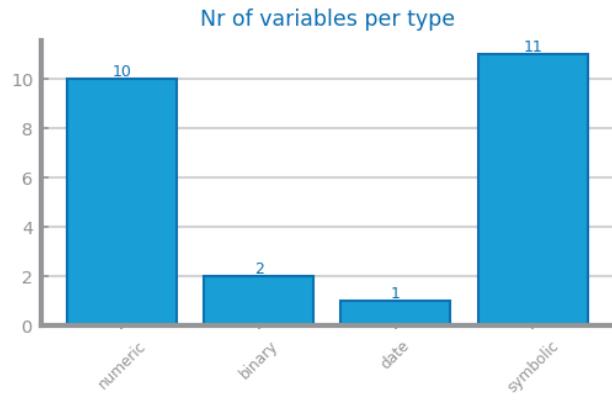


Figure 2: Nr variables per type for Traffic Accident Data (left) and Flight Cancellation Data (right)

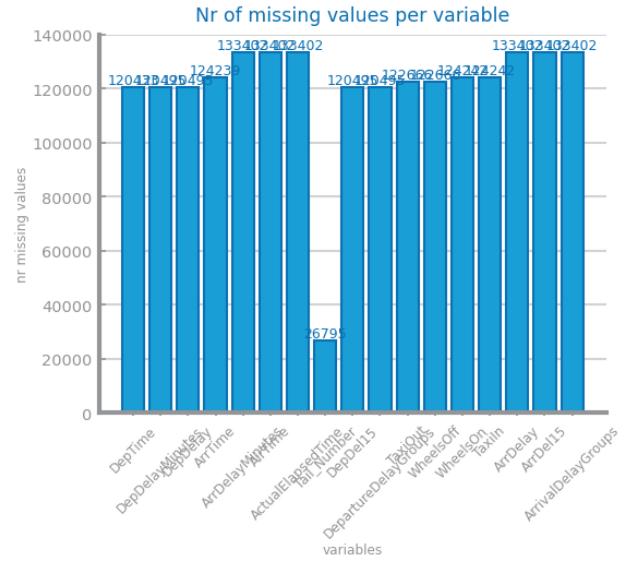
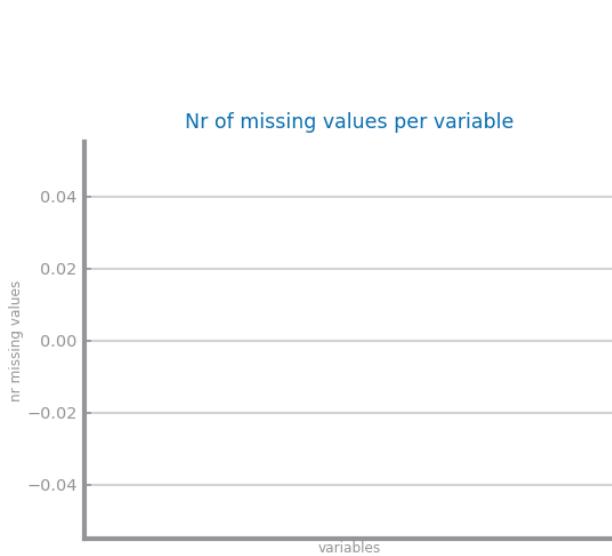


Figure 3: Nr missing values for Traffic Accident Data (left) and Flight Cancellation Data (right)

Data Distribution

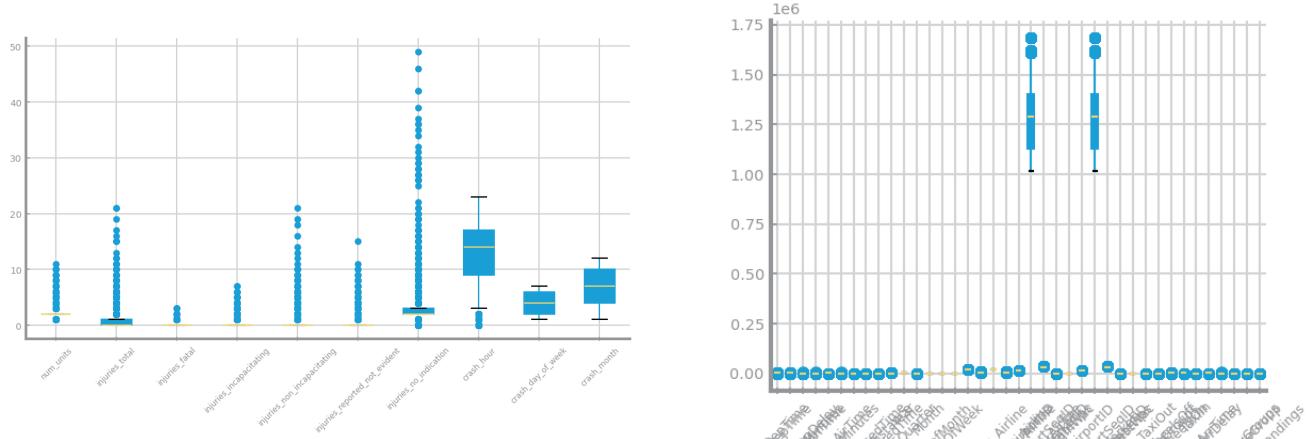


Figure 4: Global boxplots for Traffic Accident Data (left) and Flight Cancellation Data (right)

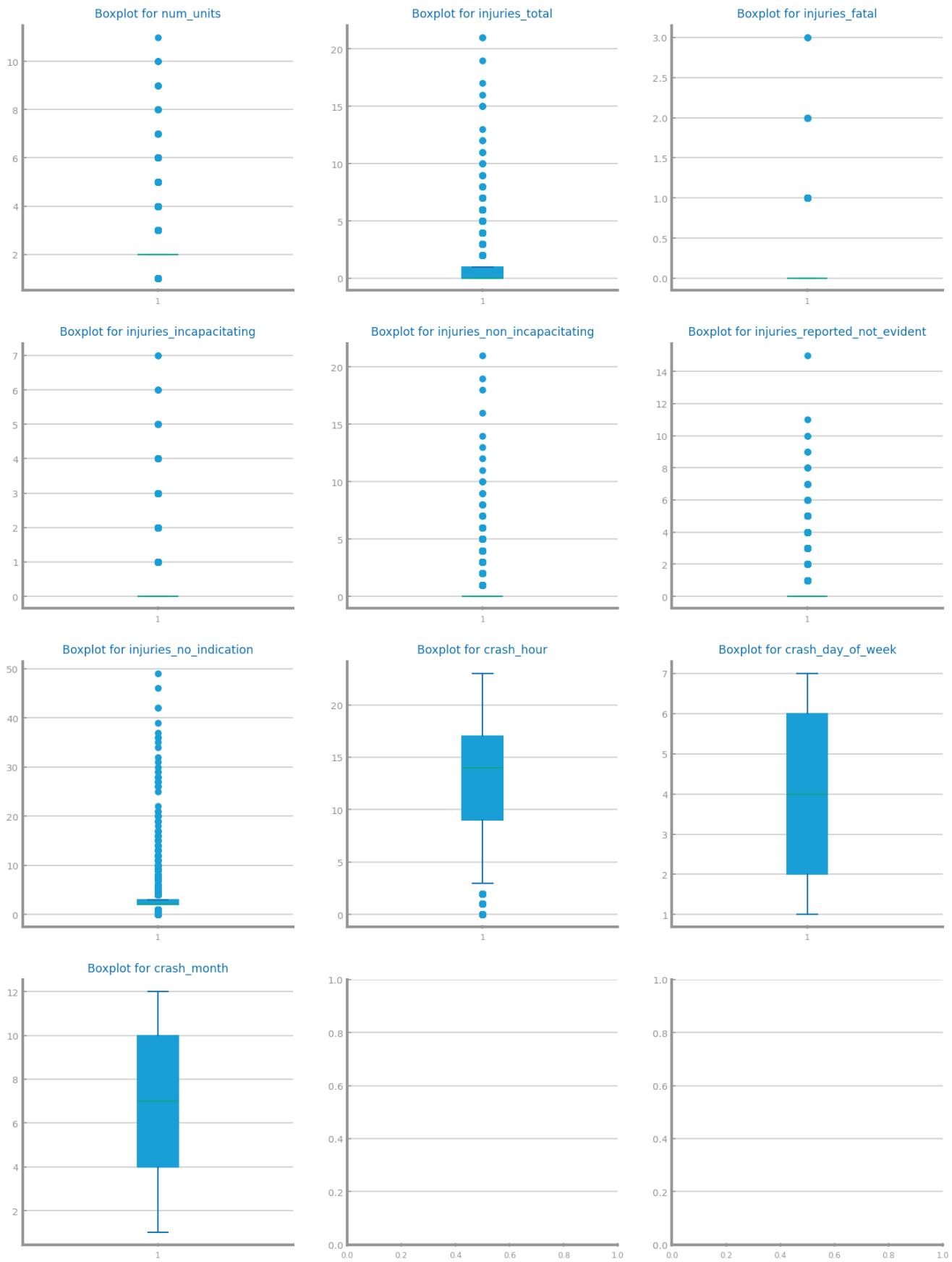


Figure 5: Single variables boxplots for Traffic Accident Data

Figure 6: Single variables boxplots for Flight Cancellation Data

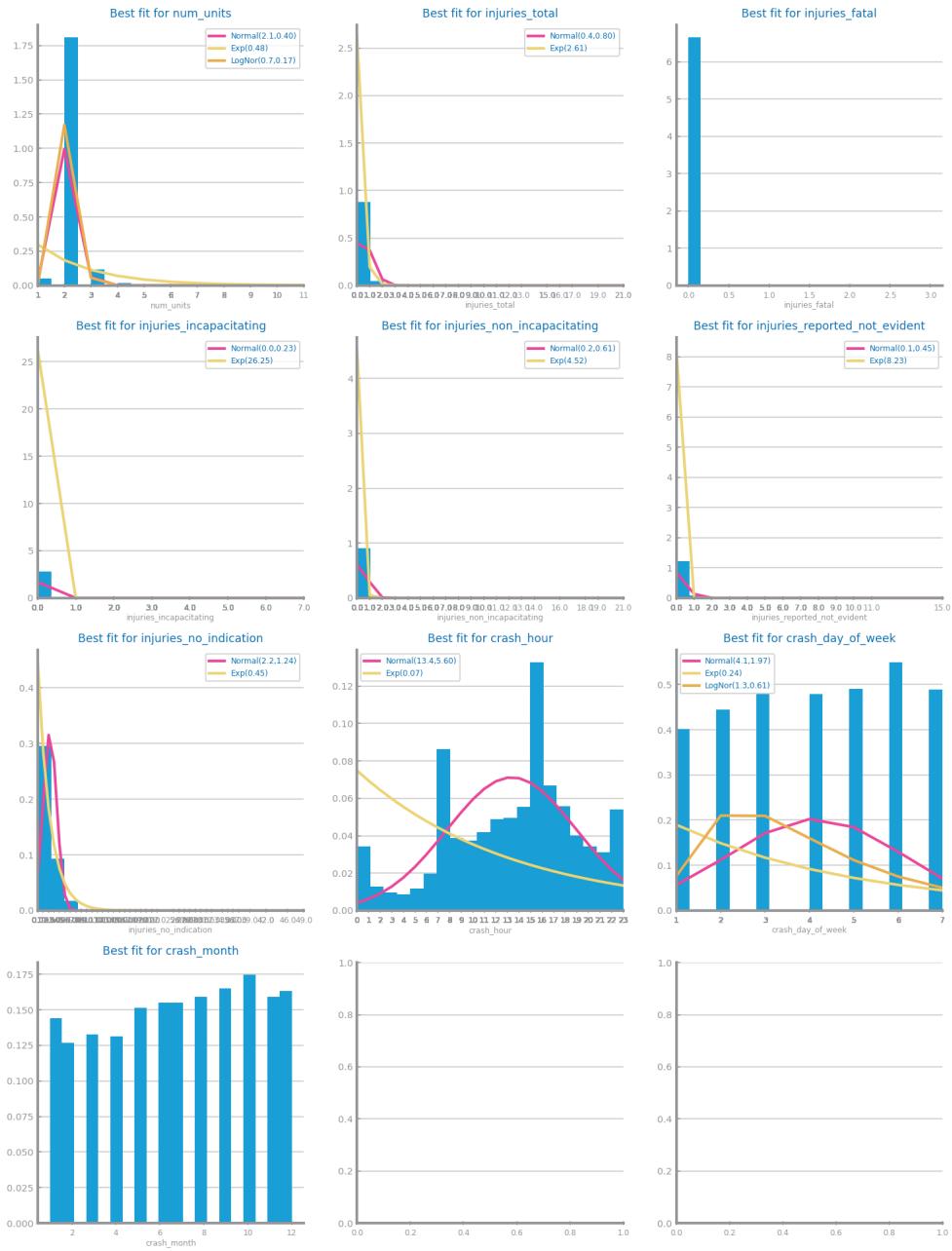


Figure 7: Histograms for Traffic Accident Data - numeric

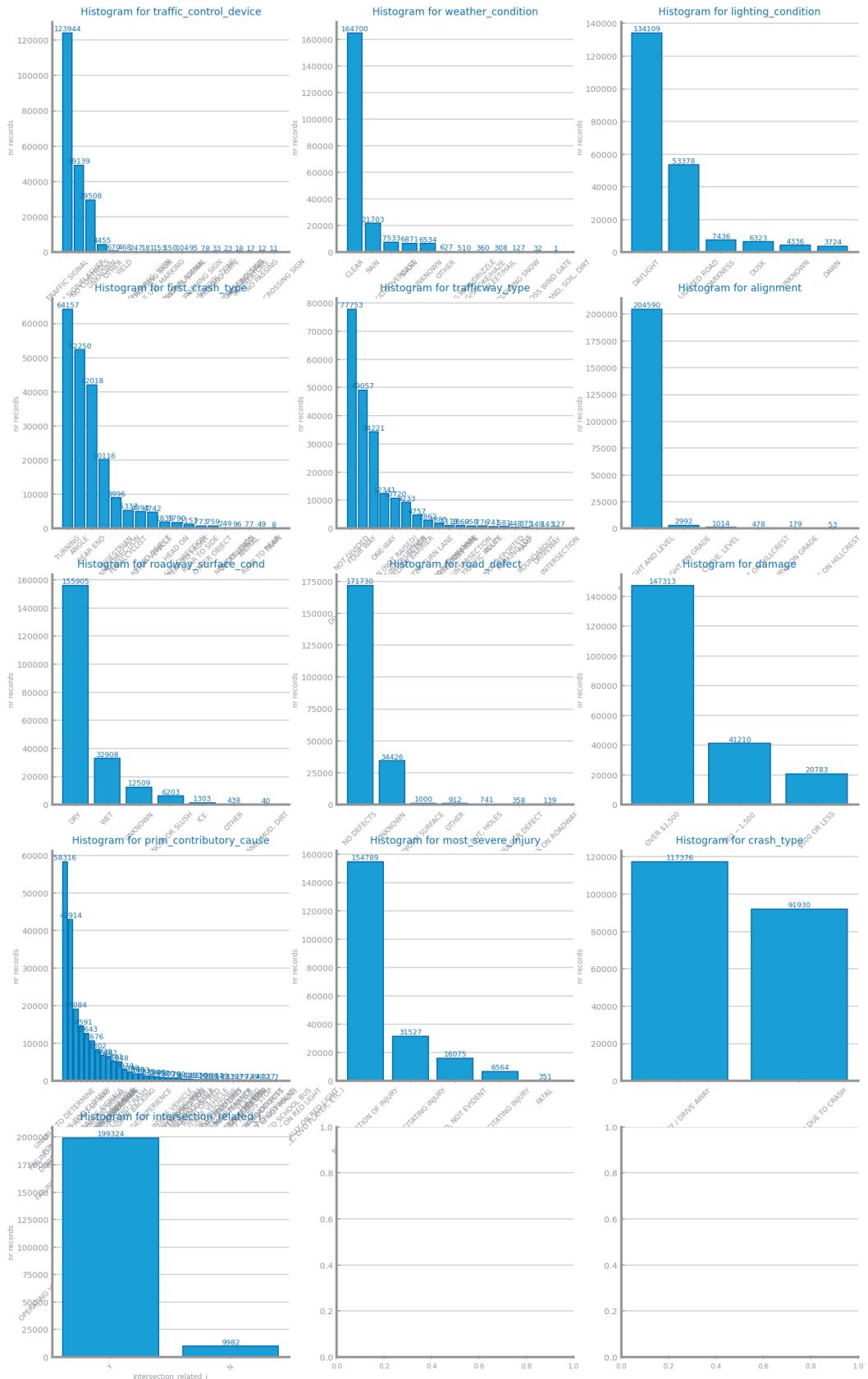
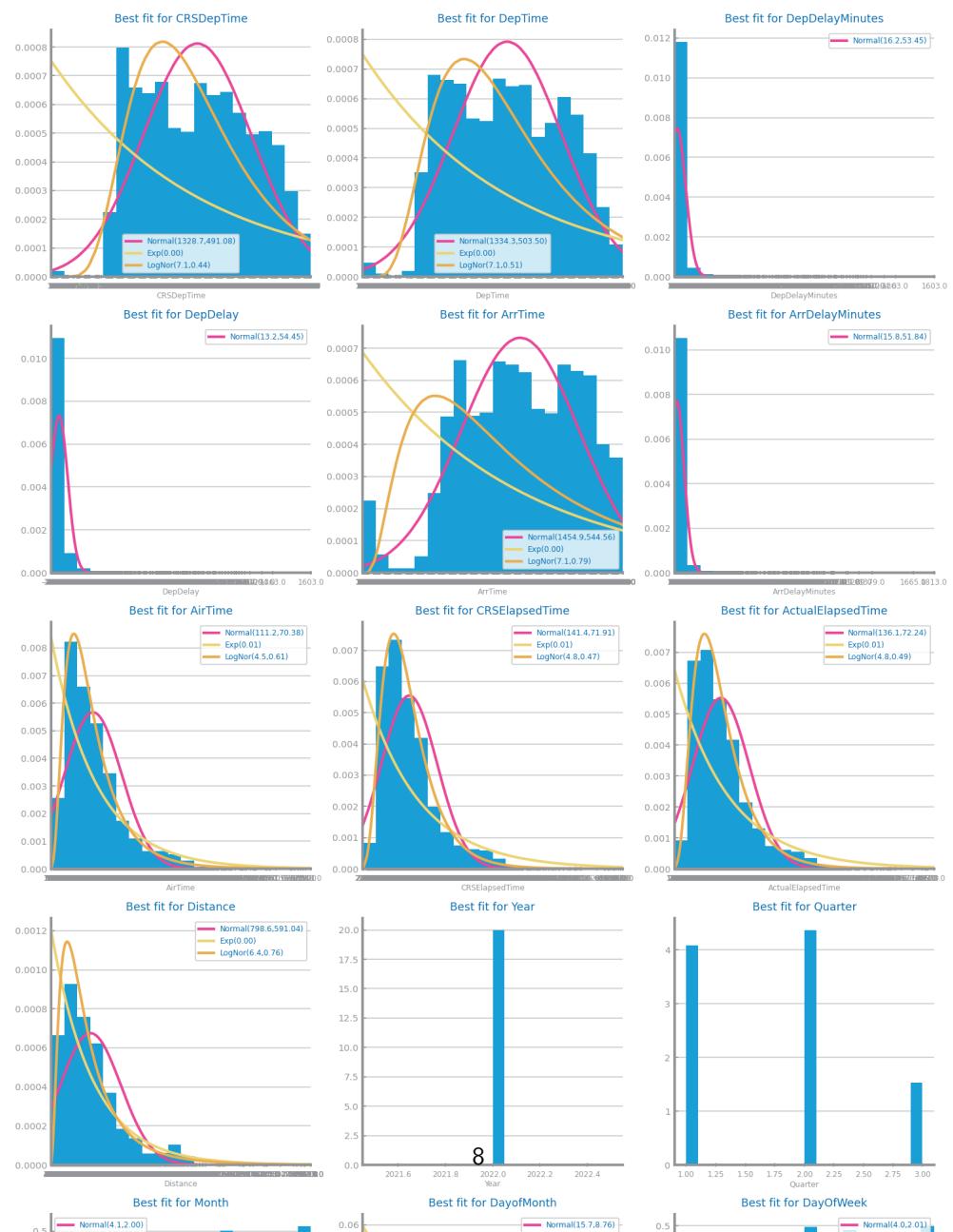
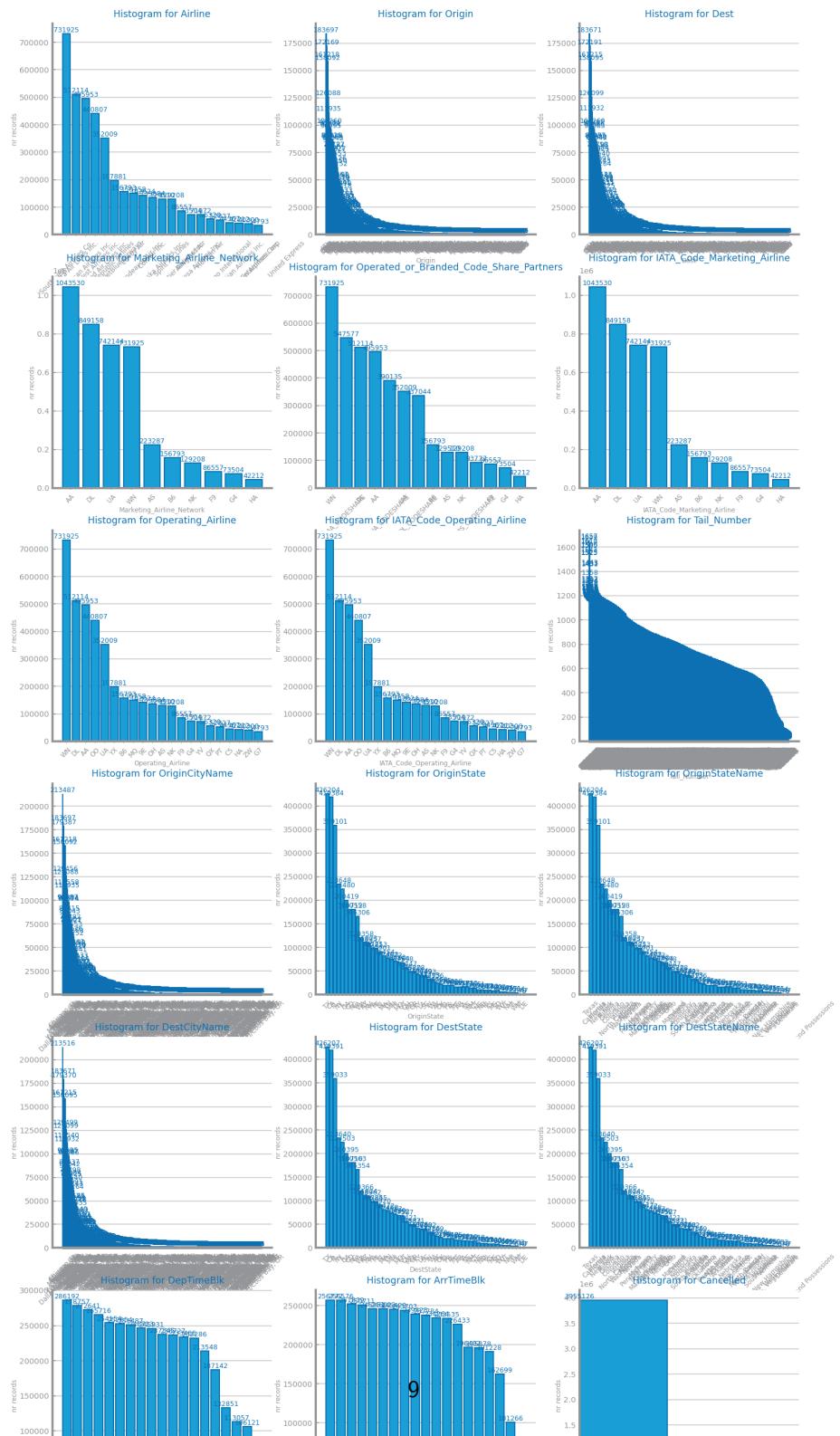


Figure 8: Histograms for Traffic Accident Data - symbolic





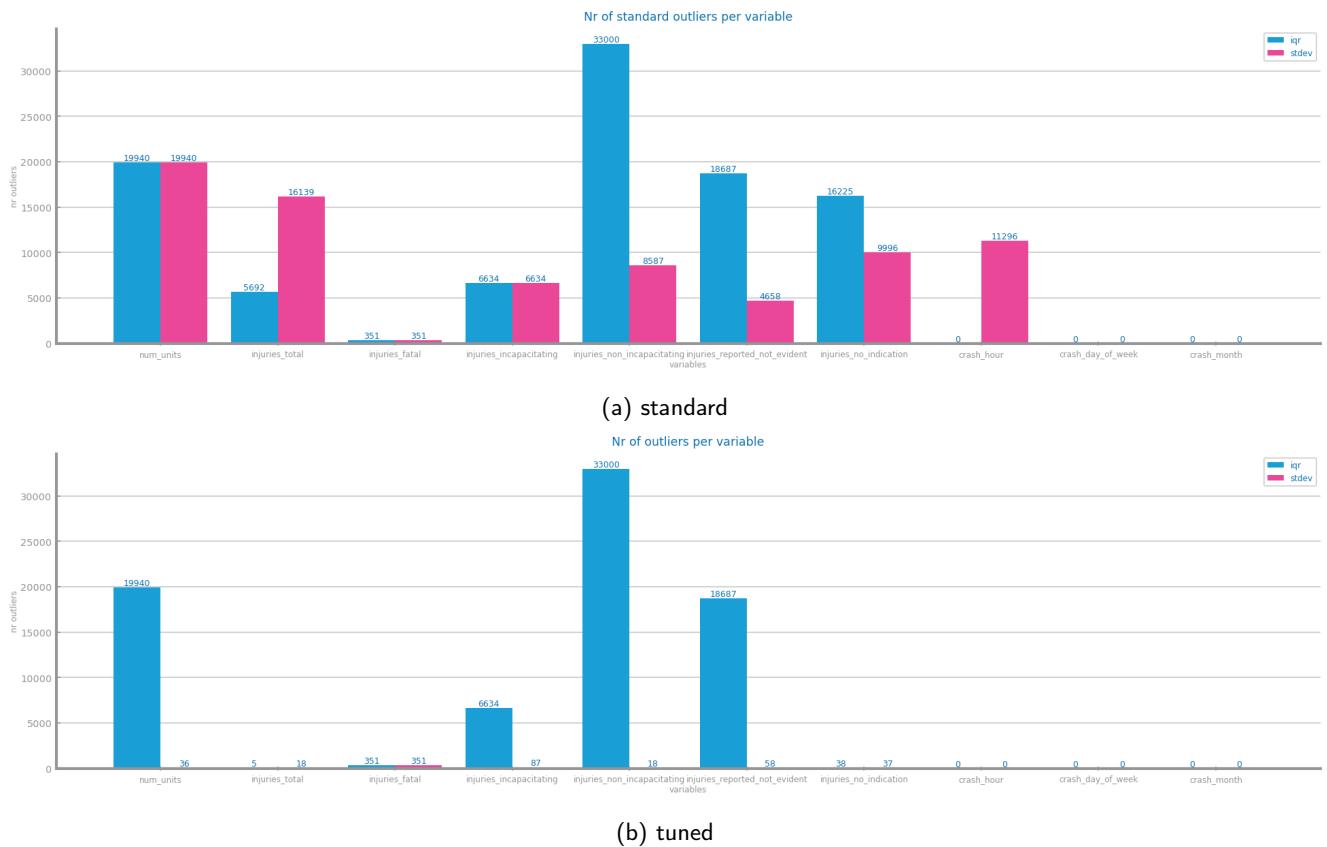
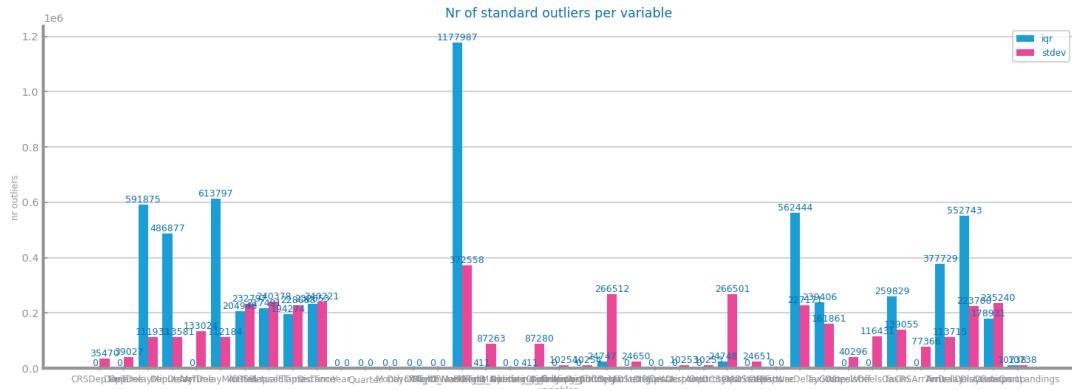
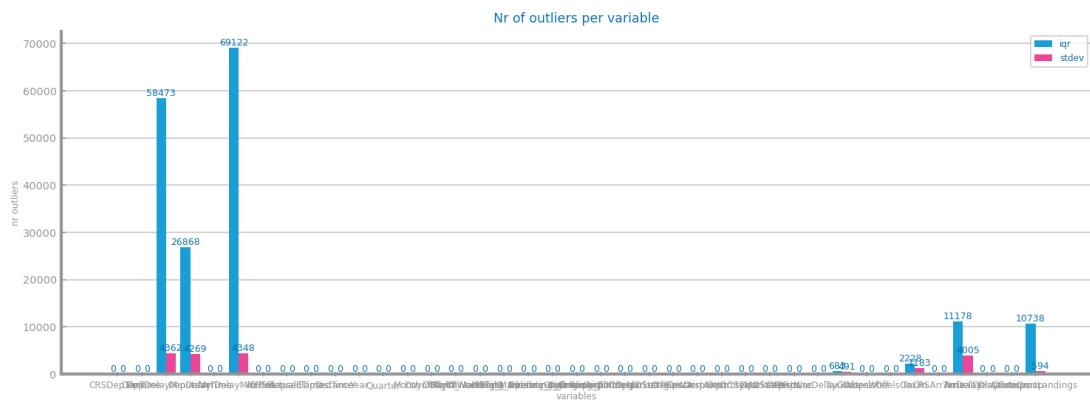


Figure 11: Outliers study Traffic Accident Data



(a) standard



(b) tuned

Figure 12: Outliers study Flight Cancellation Data

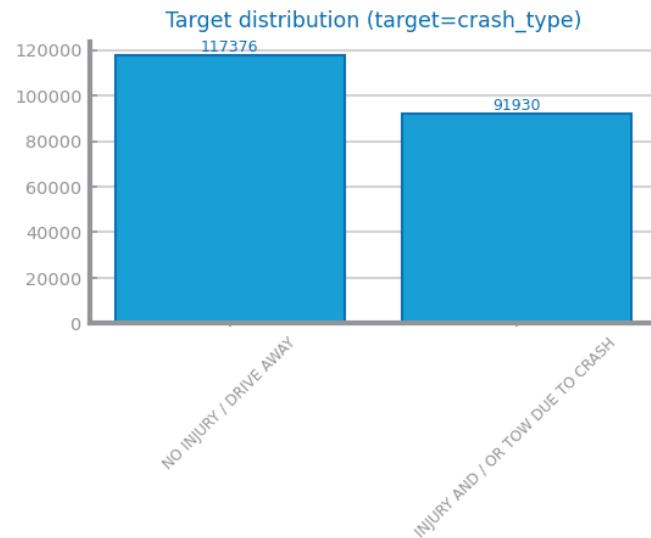


Figure 13: Class distribution for Traffic Accident Data

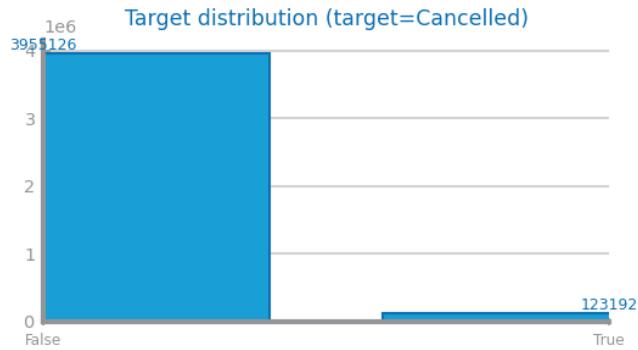


Figure 14: Class distribution for Flight Cancellation Data

Data Granularity

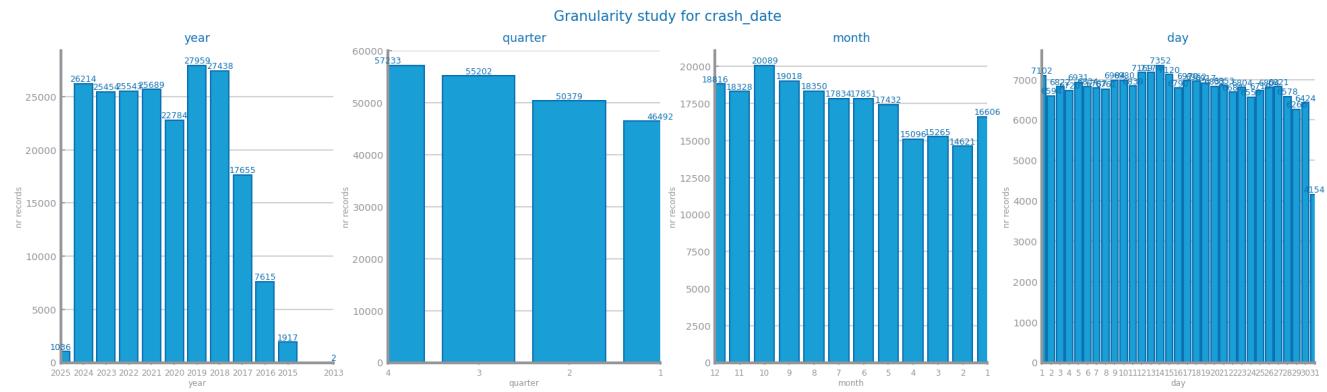


Figure 15: Granularity analysis for Traffic Accident Data

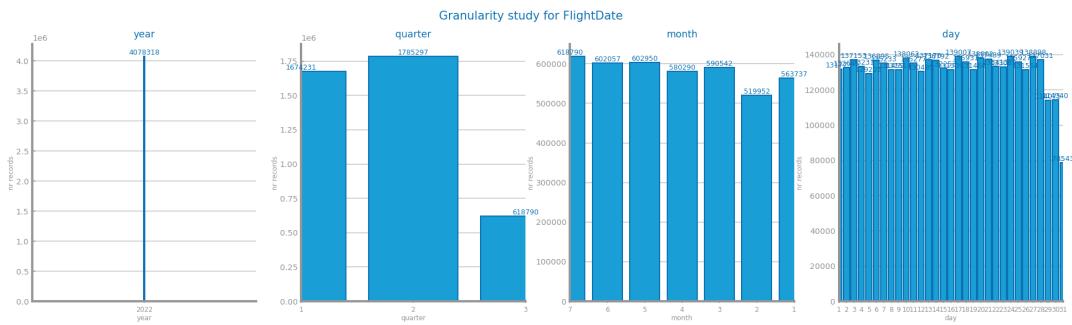


Figure 16: Granularity analysis for Flight Cancellation Data

Data Sparsity

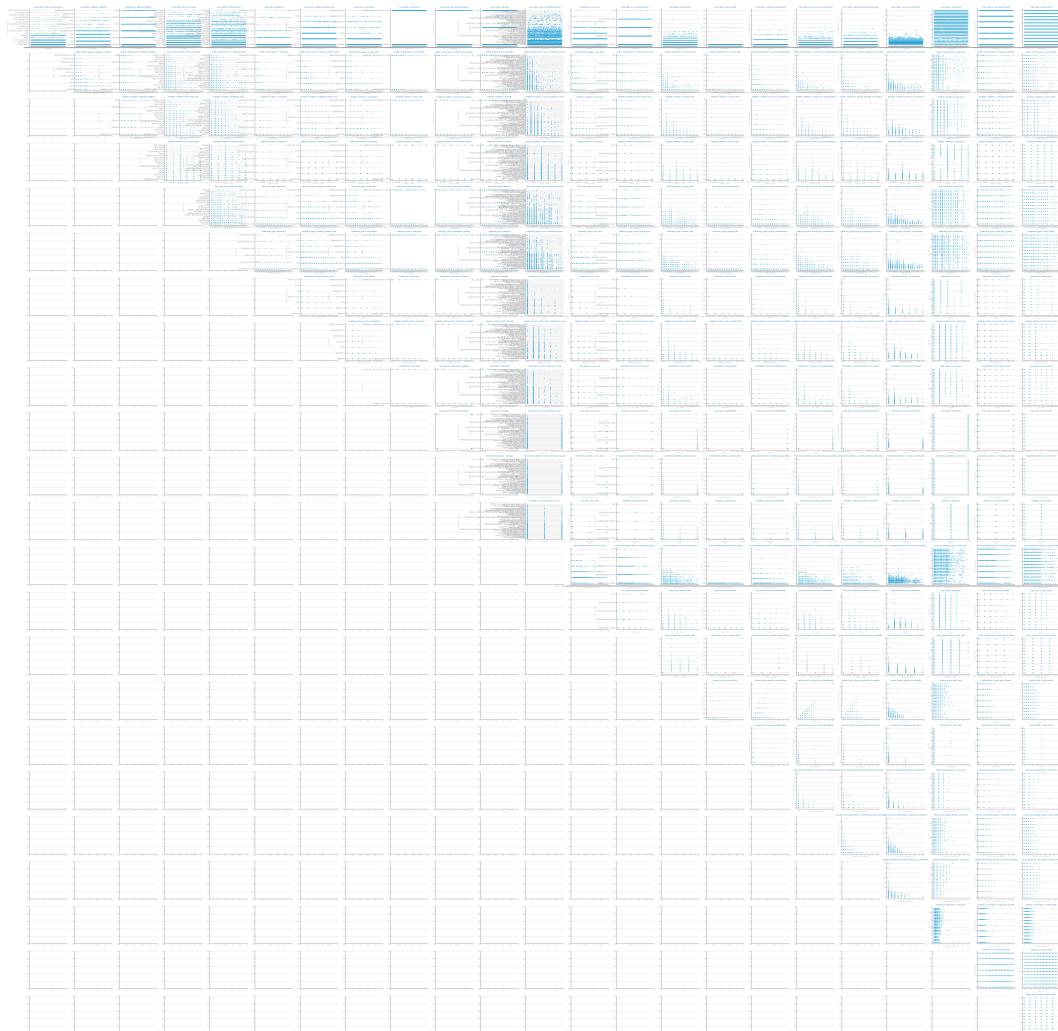


Figure 17: Sparsity analysis for Traffic Accident Data



Figure 18: Sparsity analysis per class for Traffic Accident Data

LINK

Figure 19: Sparsity analysis for Flight Cancellation Data

LINK

Figure 20: Sparsity analysis per class for Flight Cancellation Data

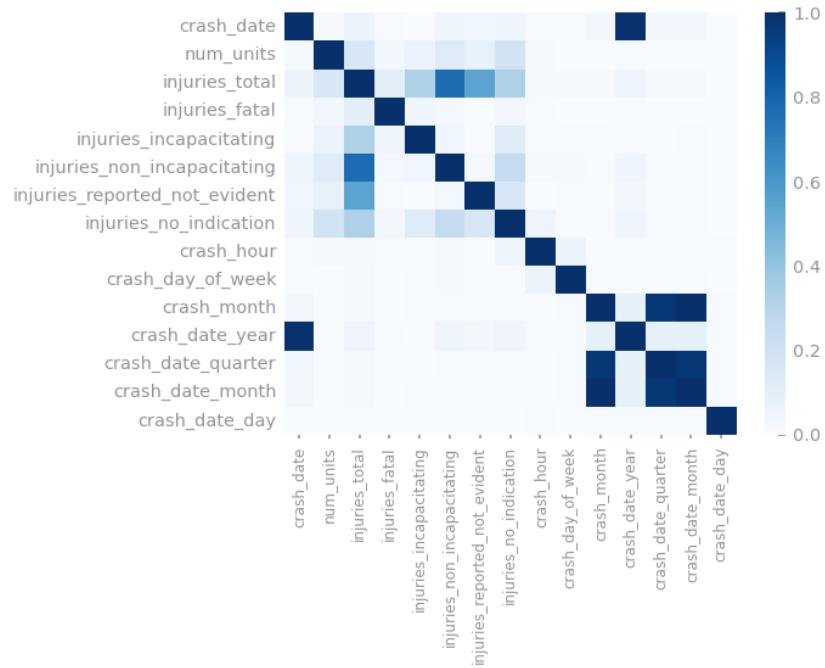


Figure 21: Correlation analysis for Traffic Accident Data

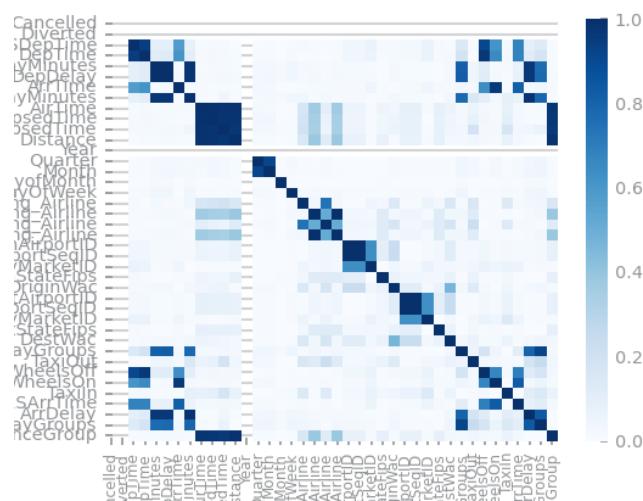


Figure 22: Correlation analysis for Flight Cancellation Data