

# Data Science Project

<b>Team nr:</b> 16	<b>Student 1:</b> Antero Morgado <b>IST nr:</b> 1119213 <b>Student 2:</b> David Ferreira <b>IST nr:</b> 1107077 <b>Student 3:</b> José Fernandes <b>IST nr:</b> 1103727 <b>Student 4:</b> Olha Buts <b>IST nr:</b> 1116276
--------------------	---

## CLASSIFICATION

### 1 DATA PROFILING

#### *Data Dimensionality*

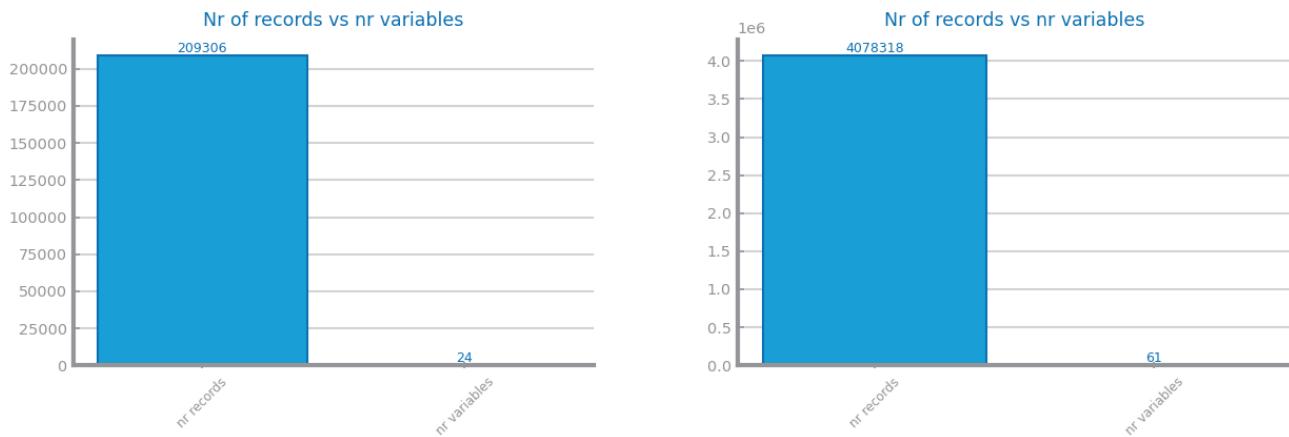


Figure 1: Nr Records x Nr variables for dataset 1 (left) and dataset 2 (right)

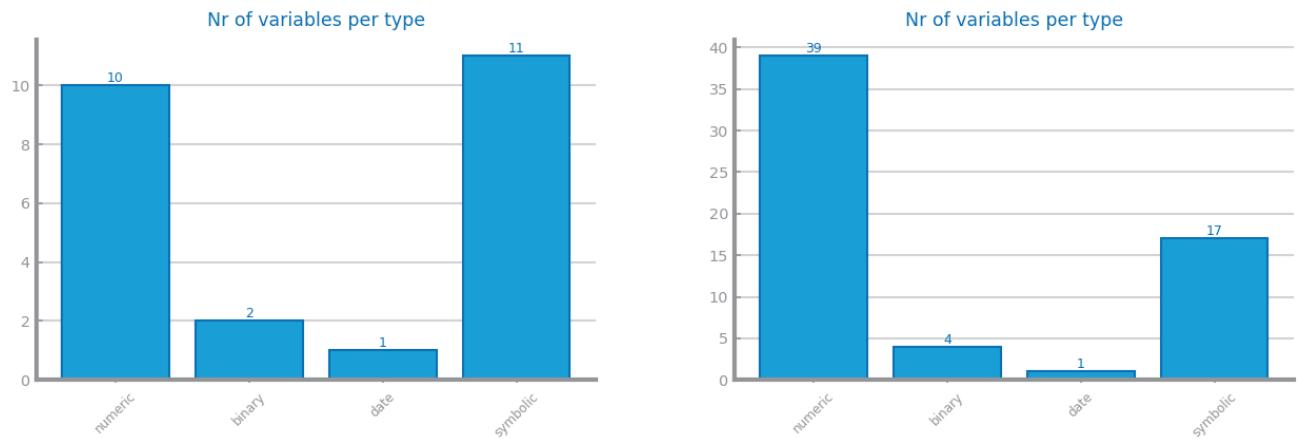


Figure 2: Nr variables per type for dataset 1 (left) and dataset 2 (right)

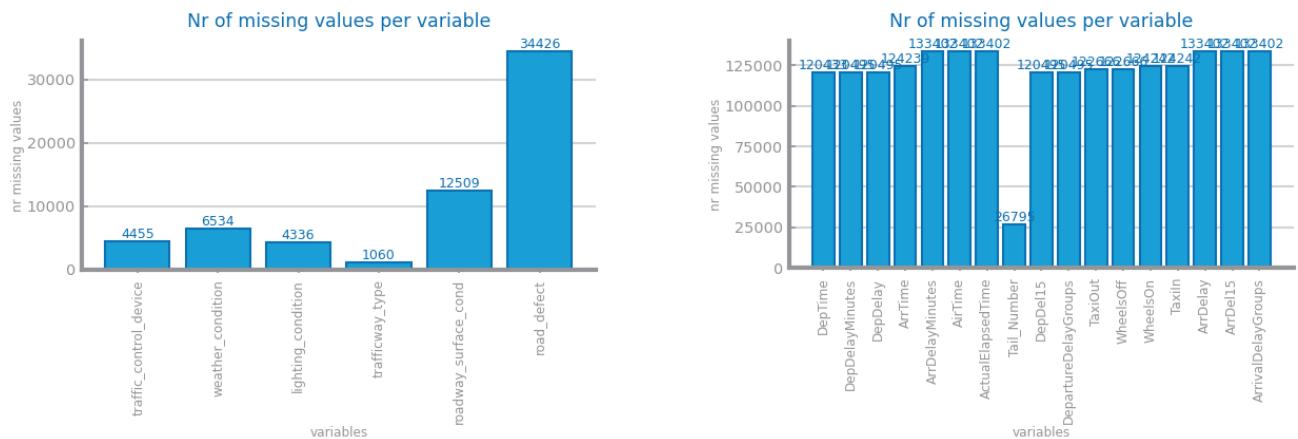


Figure 3: Nr missing values for dataset 1 (left) and dataset 2 (right)

## Data Distribution

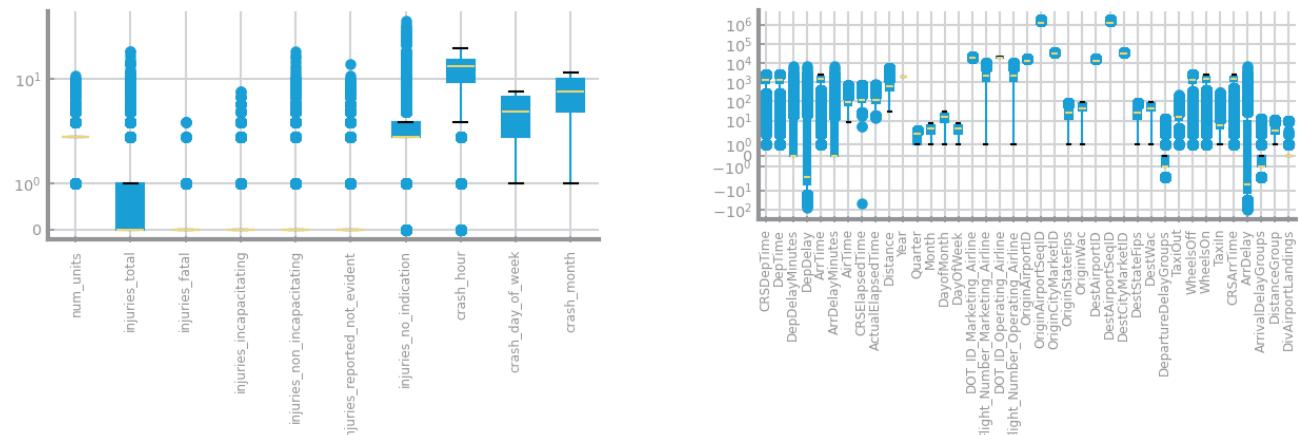


Figure 4: Global boxplots dataset 1 (left) and dataset 2 (right)

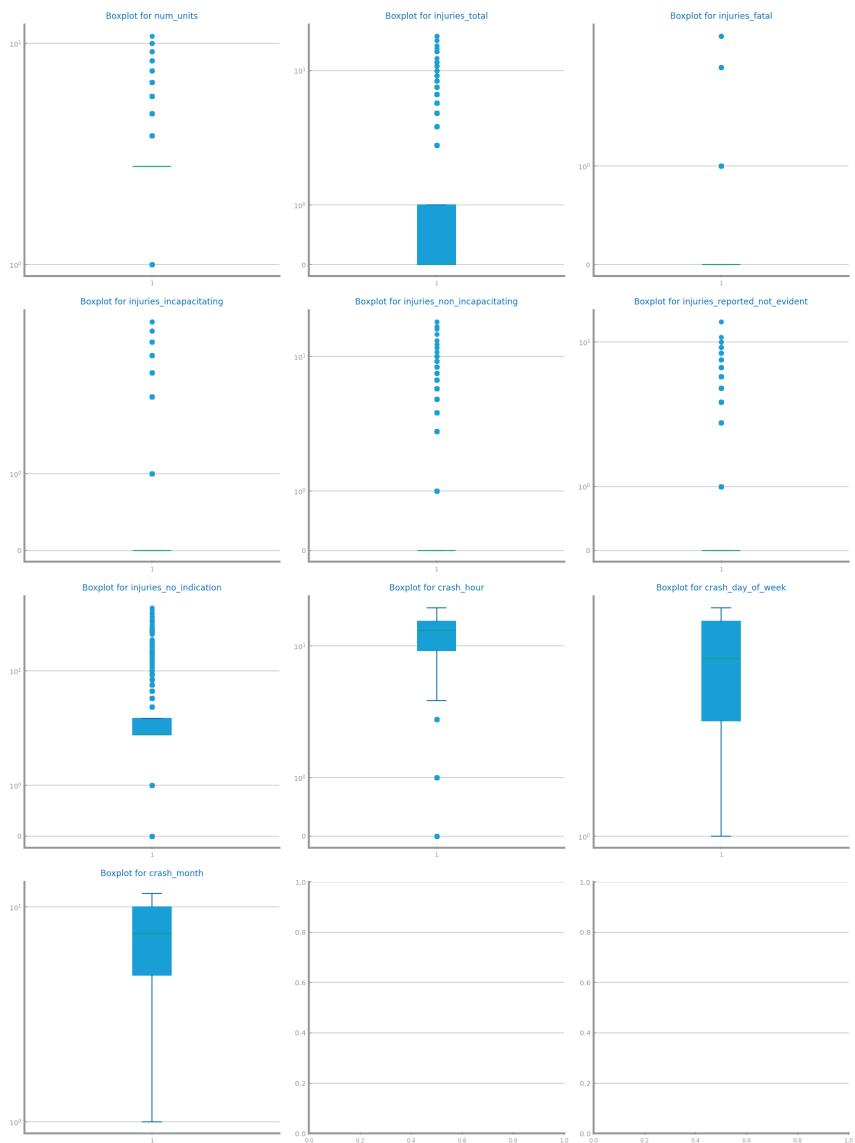


Figure 5: Single variables boxplots for dataset 1

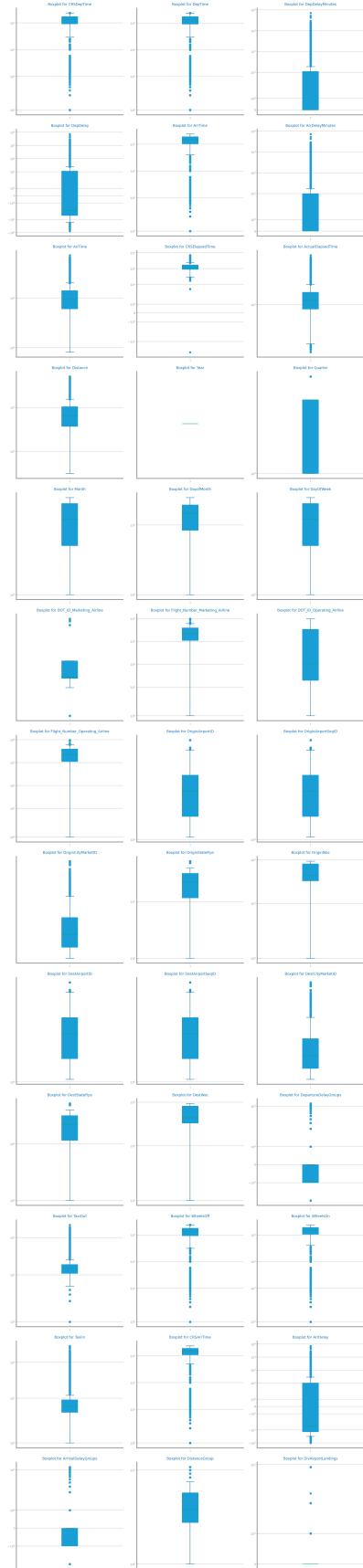


Figure 6: Single variables boxplots for dataset 2

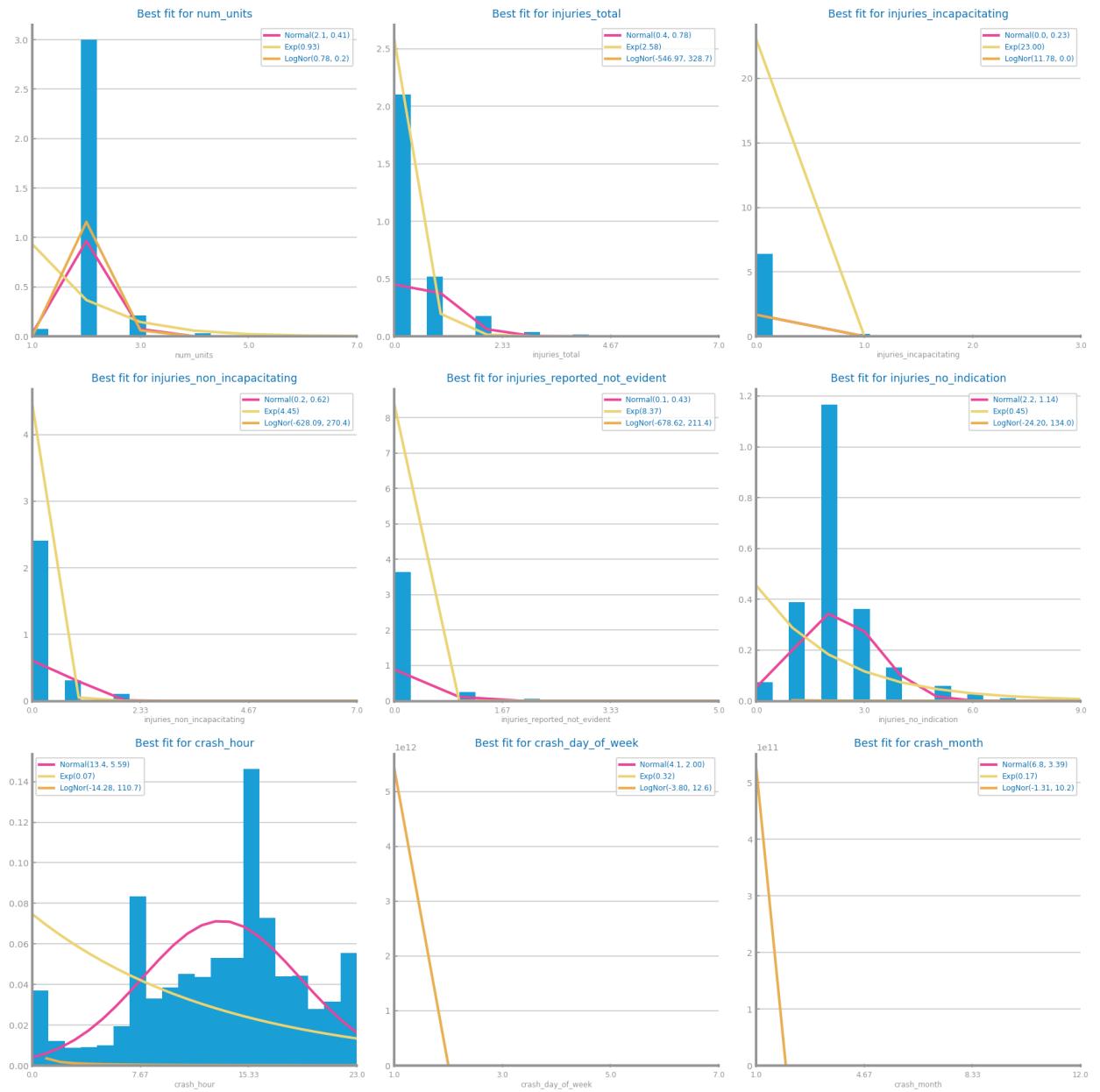


Figure 7: Numeric Histograms for dataset 1

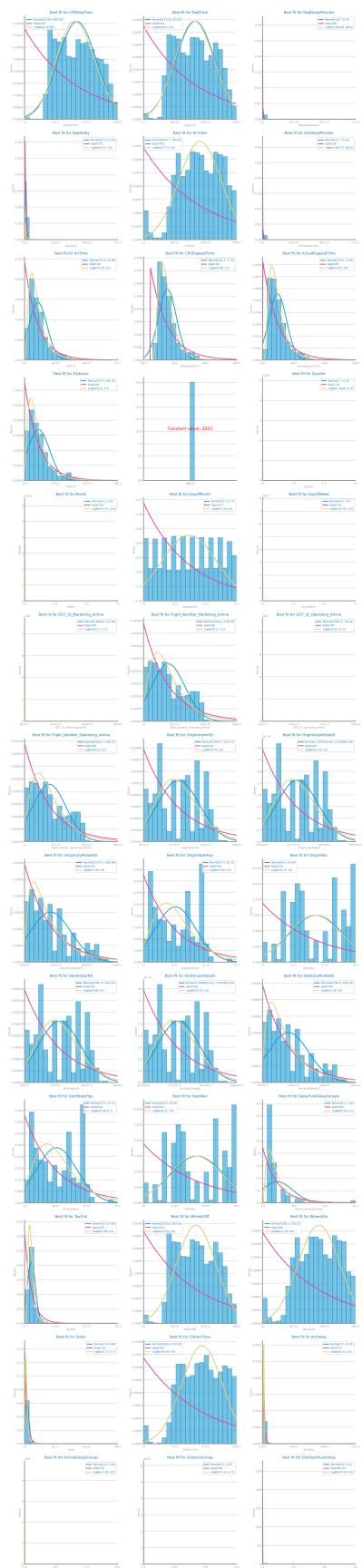


Figure 8: Numeric Histograms for dataset 2



Figure 9: Symbolic Histograms for dataset 1

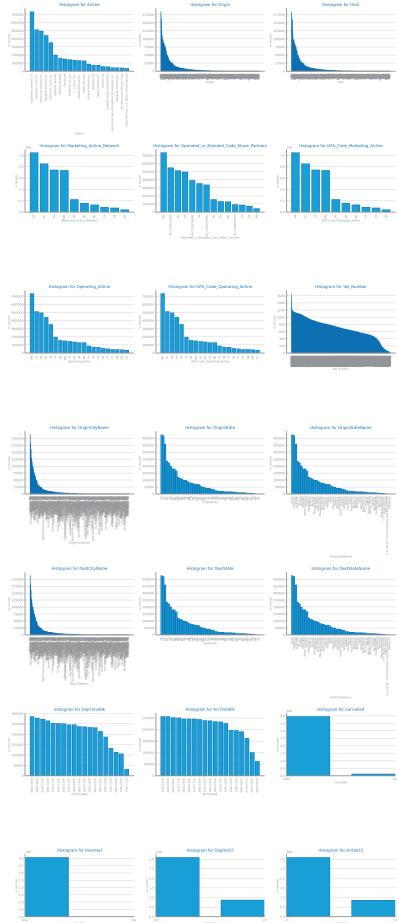


Figure 10: Symbolic Histograms for dataset 2

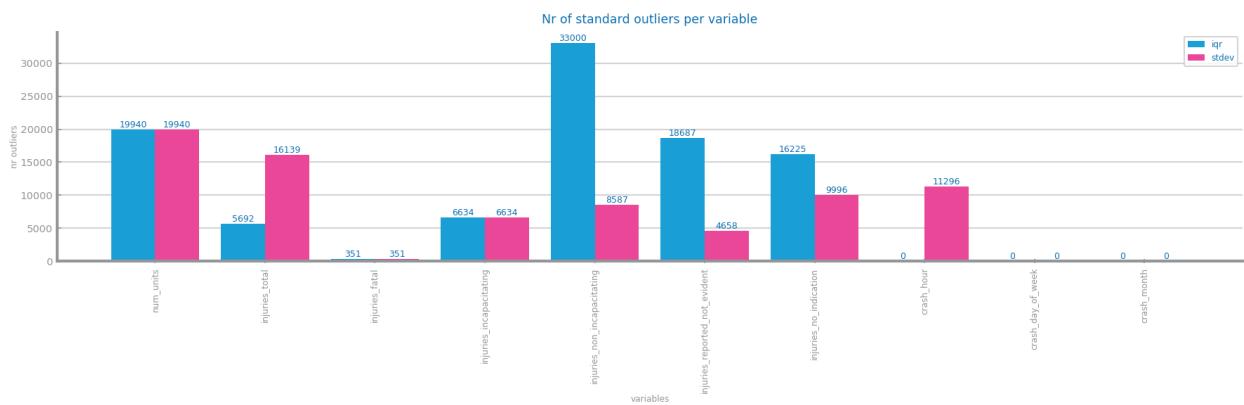


Figure 11: Outliers study dataset 1

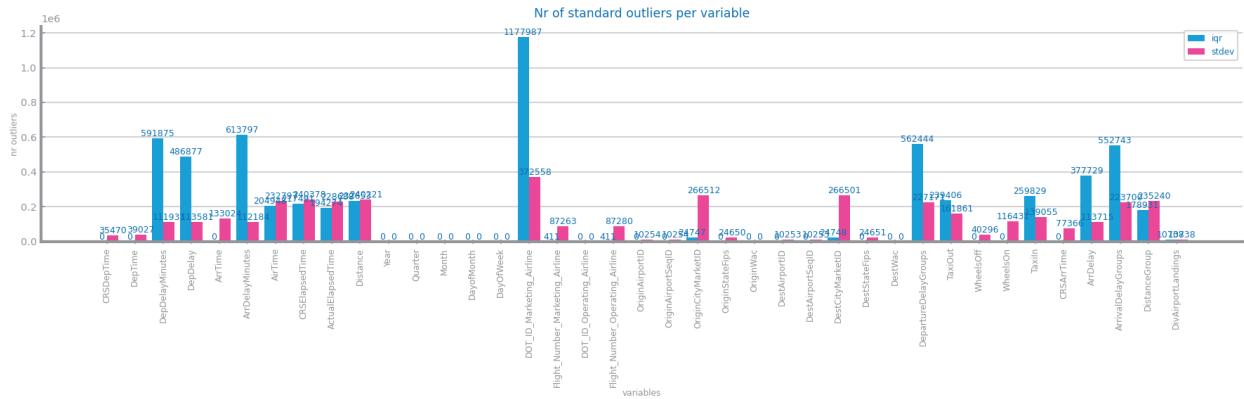


Figure 12: Outliers study dataset 2

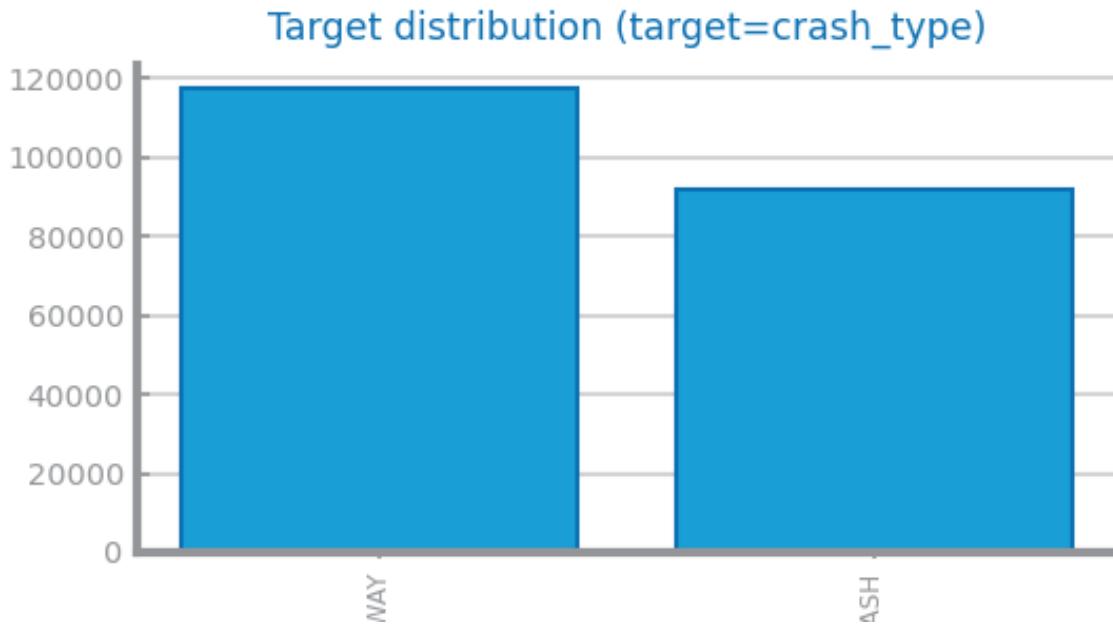


Figure 13: Class distribution for dataset 1

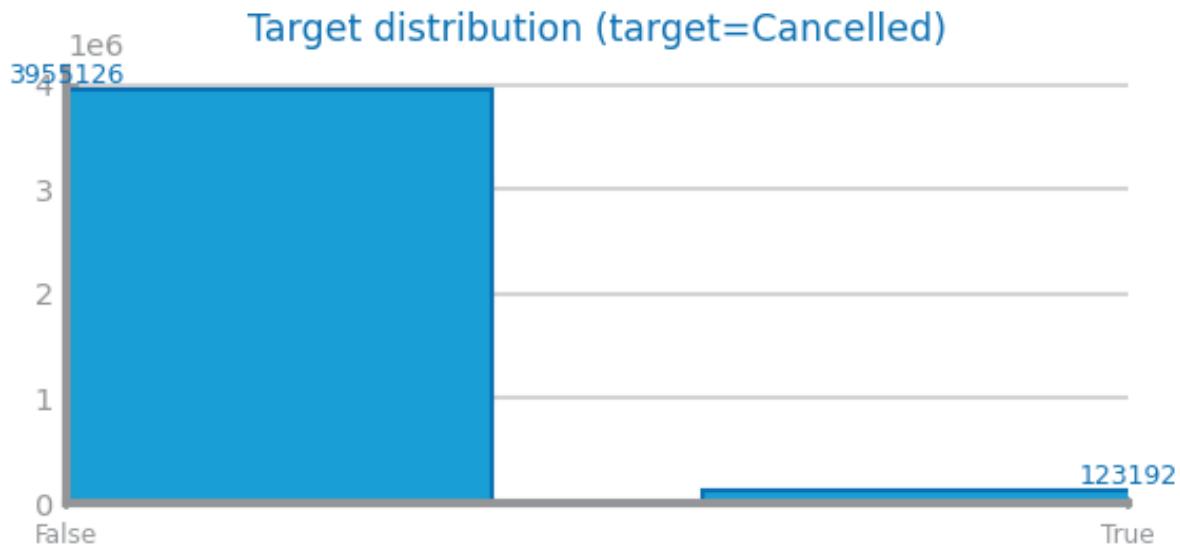


Figure 14: Class distribution for dataset 2

### *Granularity*

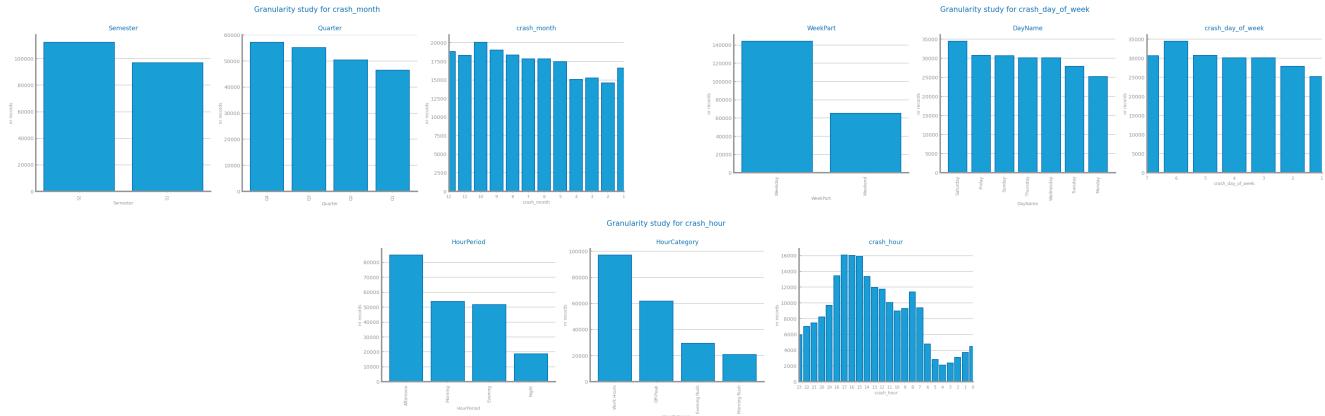


Figure 15: Granularity analysis for dataset 1

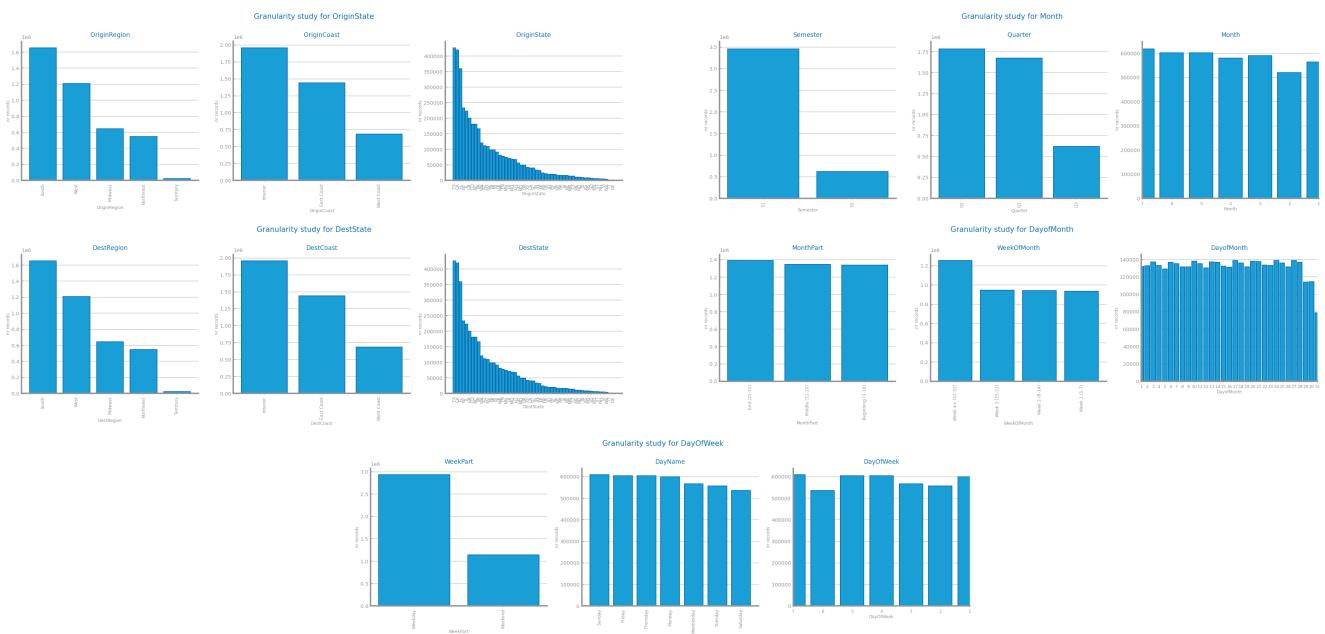


Figure 16: Granularity analysis for dataset 2

## Data Sparsity

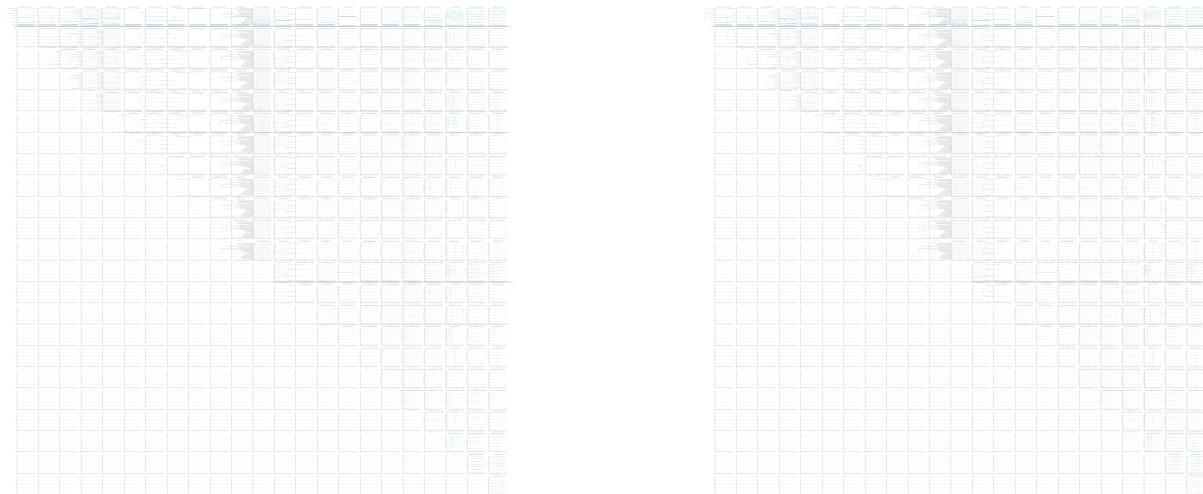


Figure 17: Sparsity analysis for dataset 1

[Click here to open the image](#)

Figure 18: Sparsity analysis for dataset 2

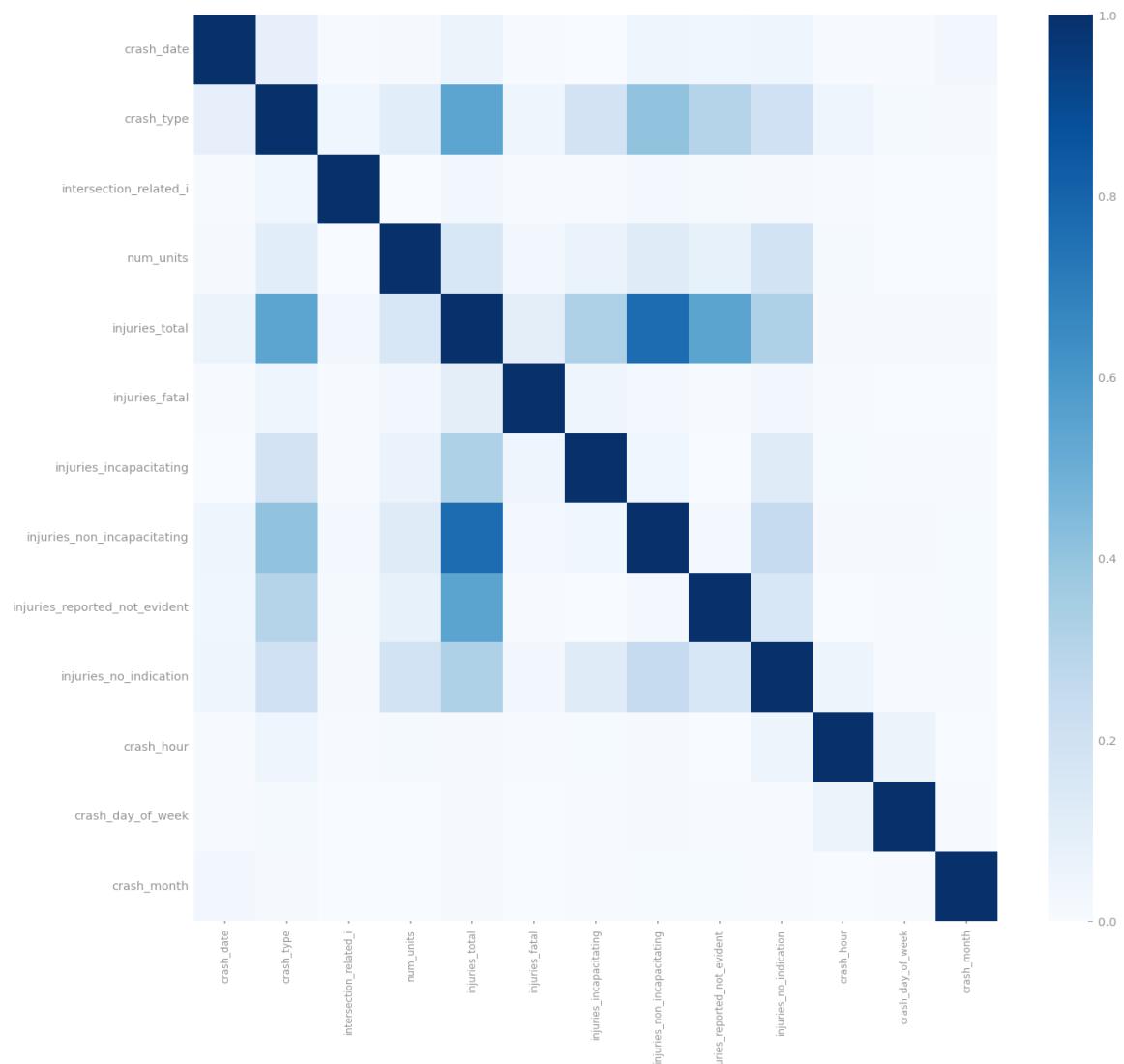


Figure 19: Correlation analysis for dataset 1

Figure 20: Correlation analysis for dataset 2