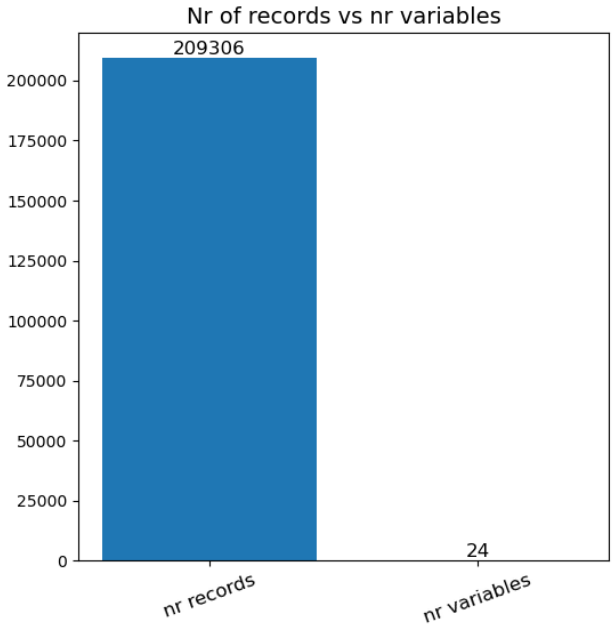
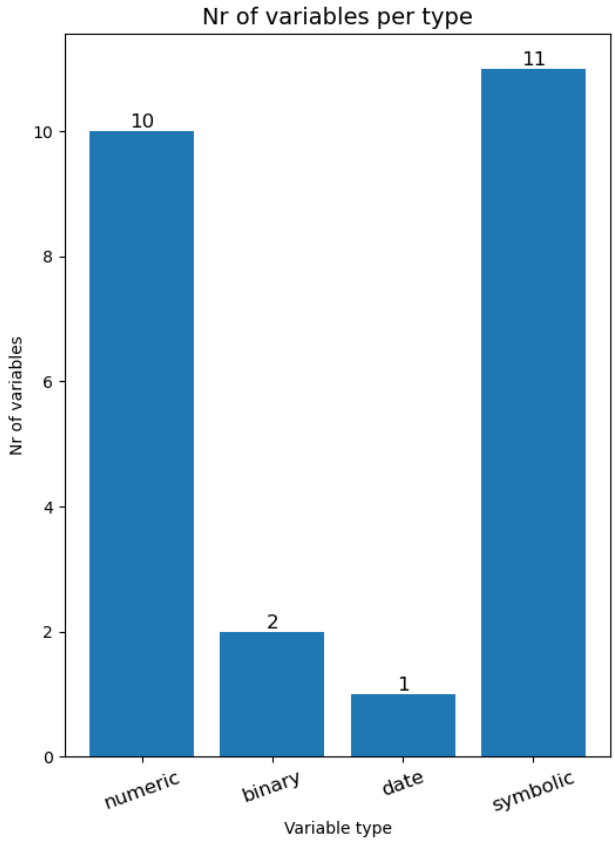
Lab 2 - Data Profiling Report

1 Accidents Dataset

1.1 Dimensionality



(a) Records vs Variables



(b) Variable Types

Figure 1: Dimensionality analysis for Accidents dataset

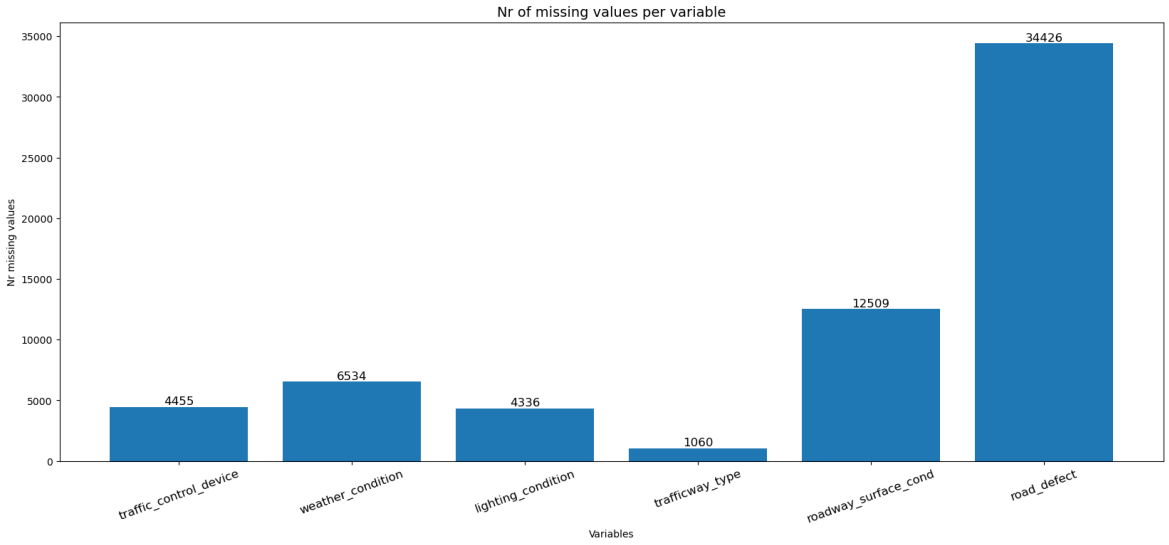


Figure 2: Missing values per variable - Accidents

1.2 Distribution

1.2.1 Global Overview

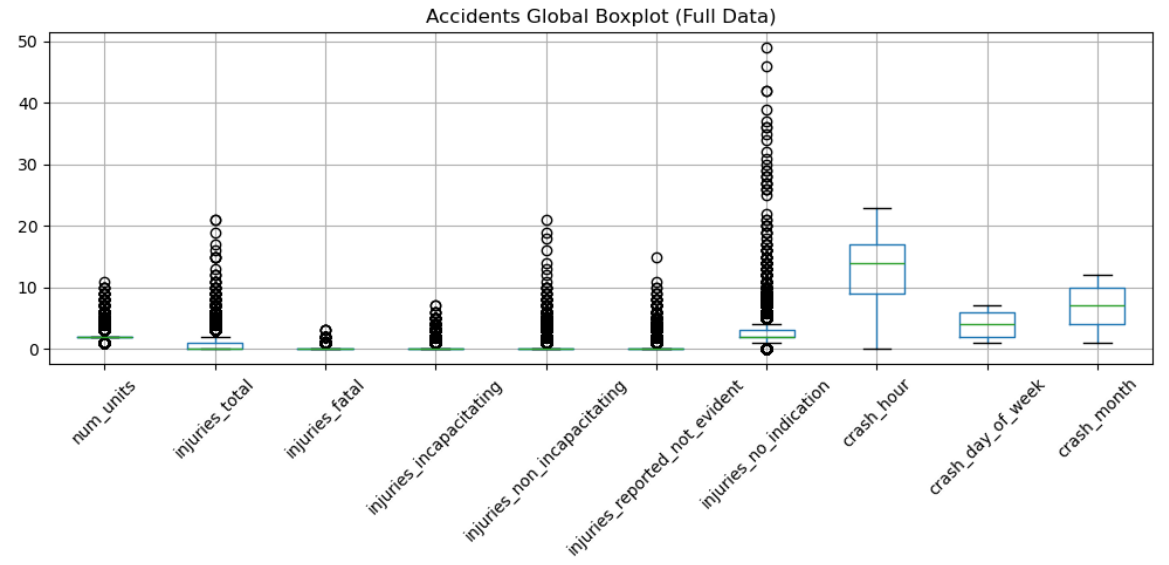


Figure 3: Global boxplot - Accidents

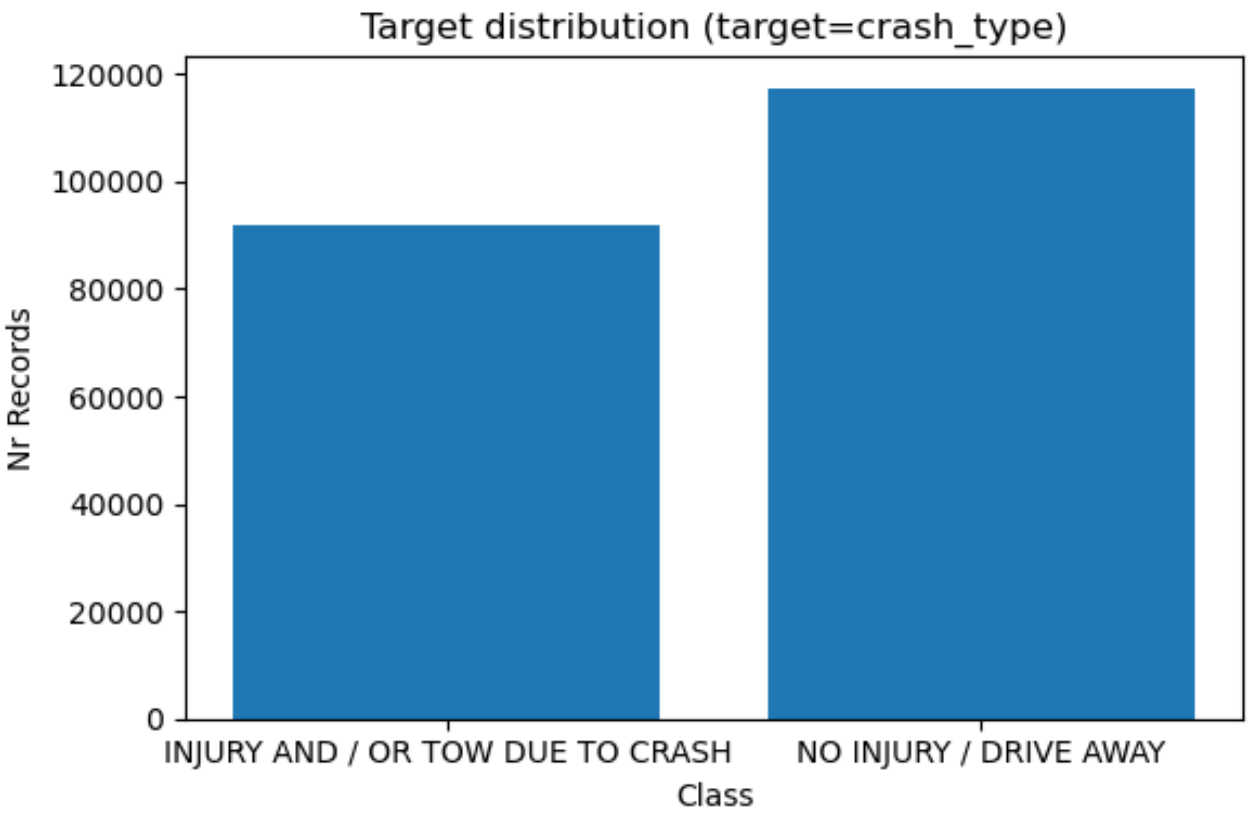


Figure 4: Class distribution - Accidents

1.2.2 Numerical Variables

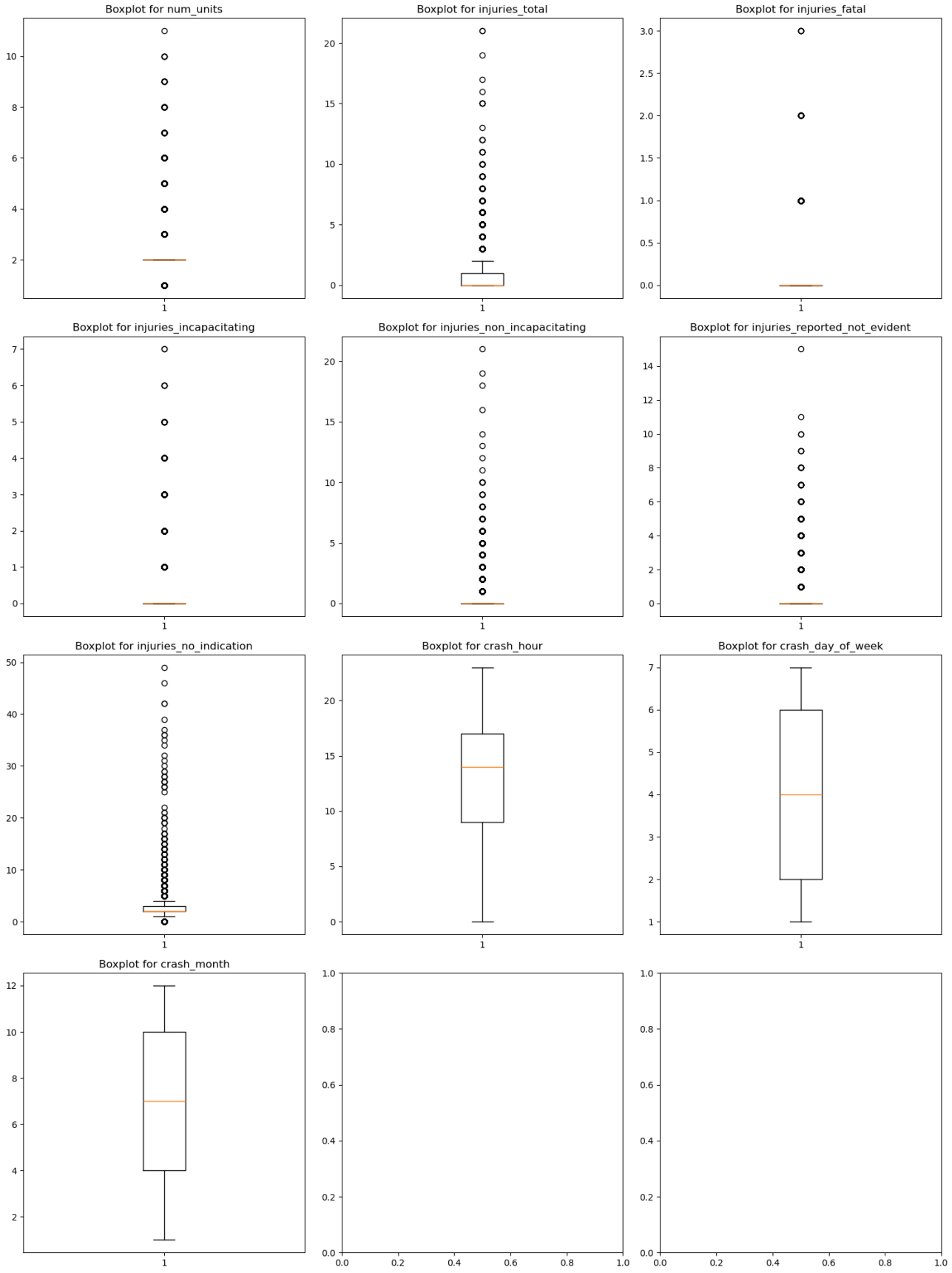


Figure 5: Individual boxplots - Accidents

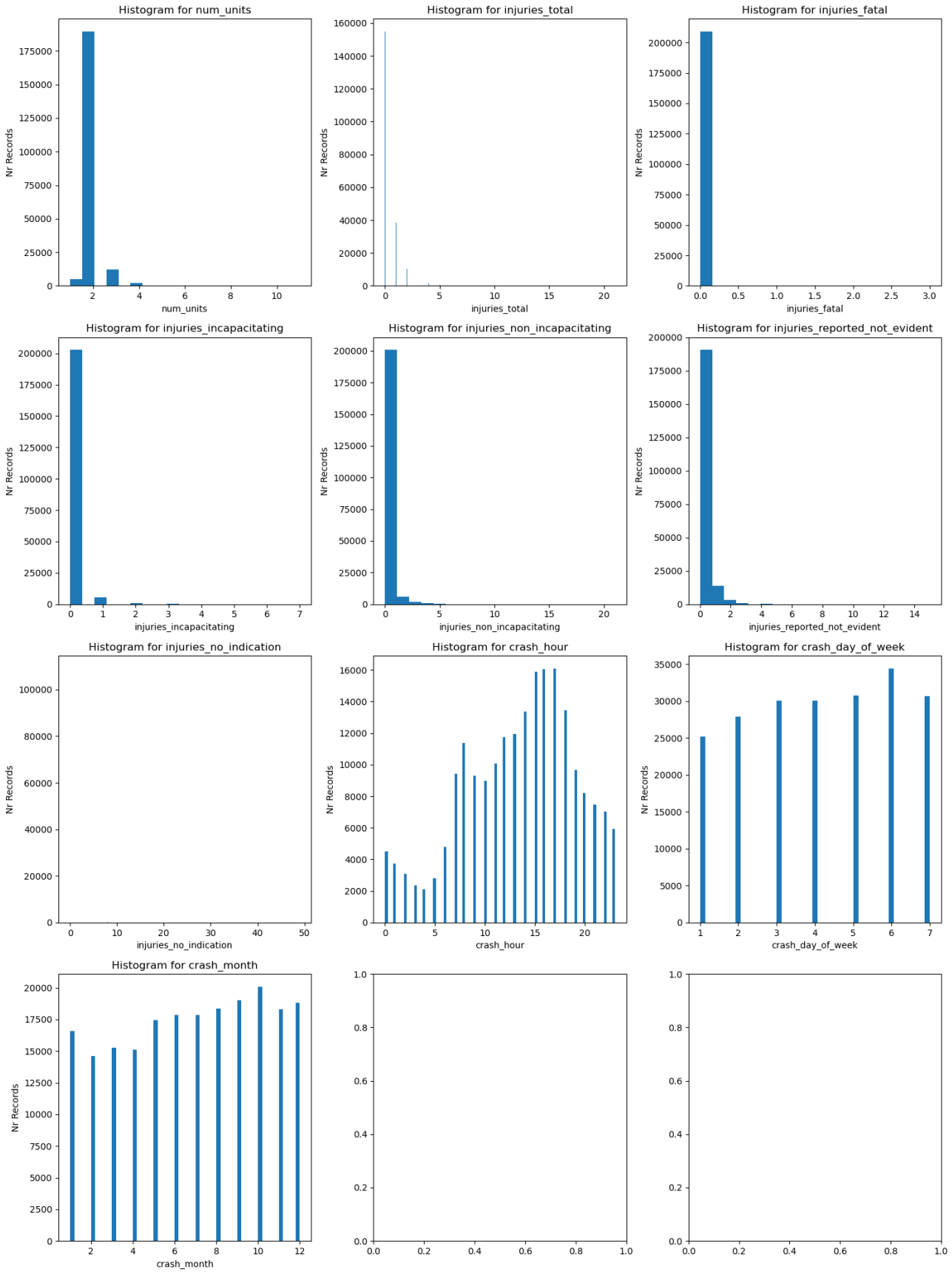


Figure 6: Histograms for numerical variables - Accidents

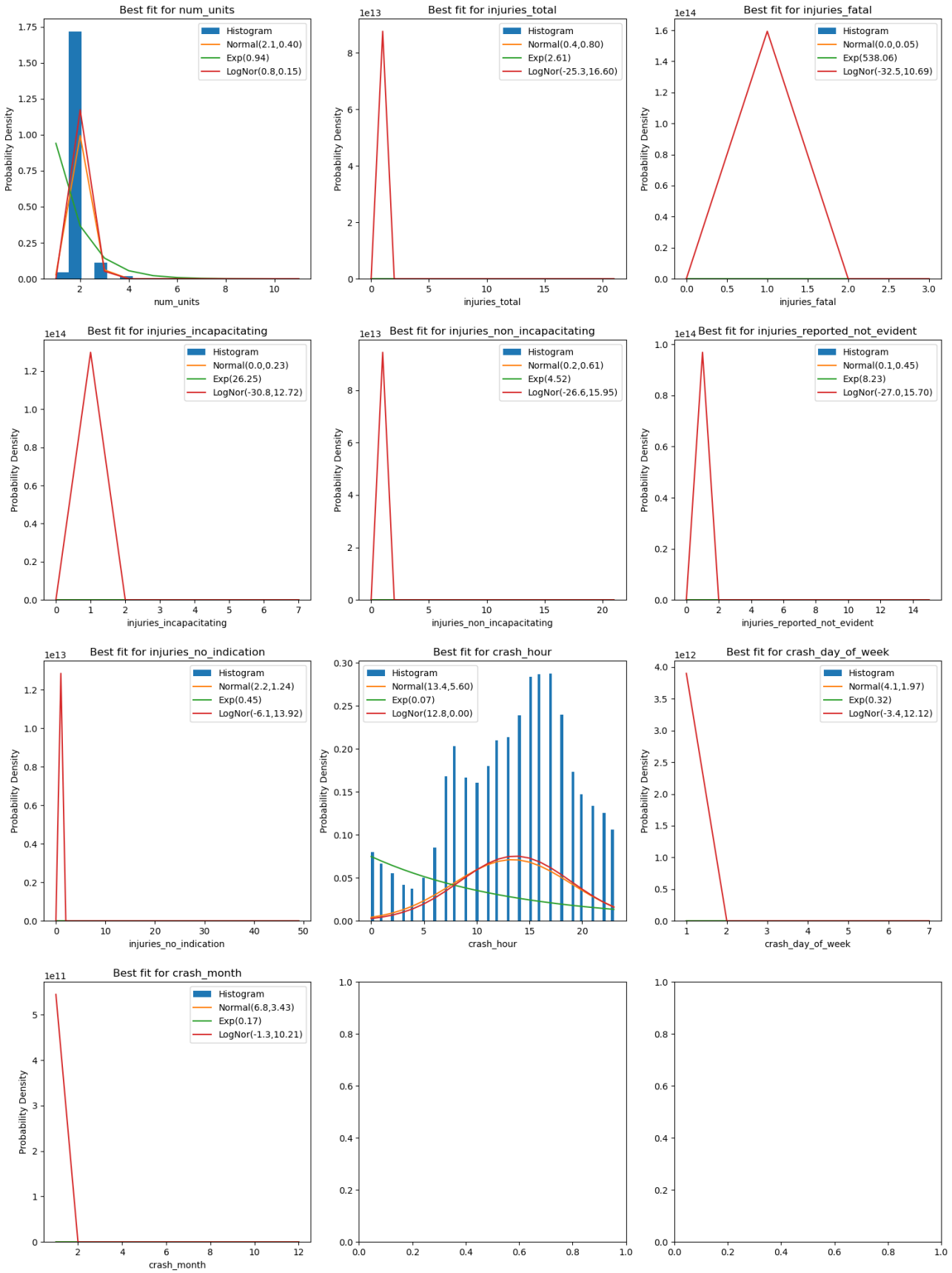


Figure 7: Probability density functions - Accidents

1.2.3 Outliers Study

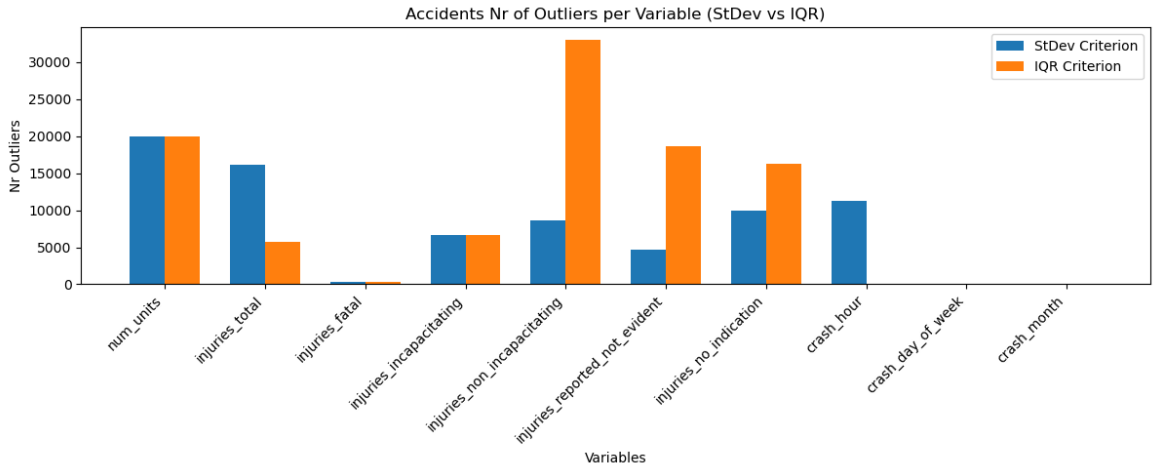


Figure 8: Outliers comparison (StDev vs IQR) - Accidents

1.2.4 Symbolic Variables

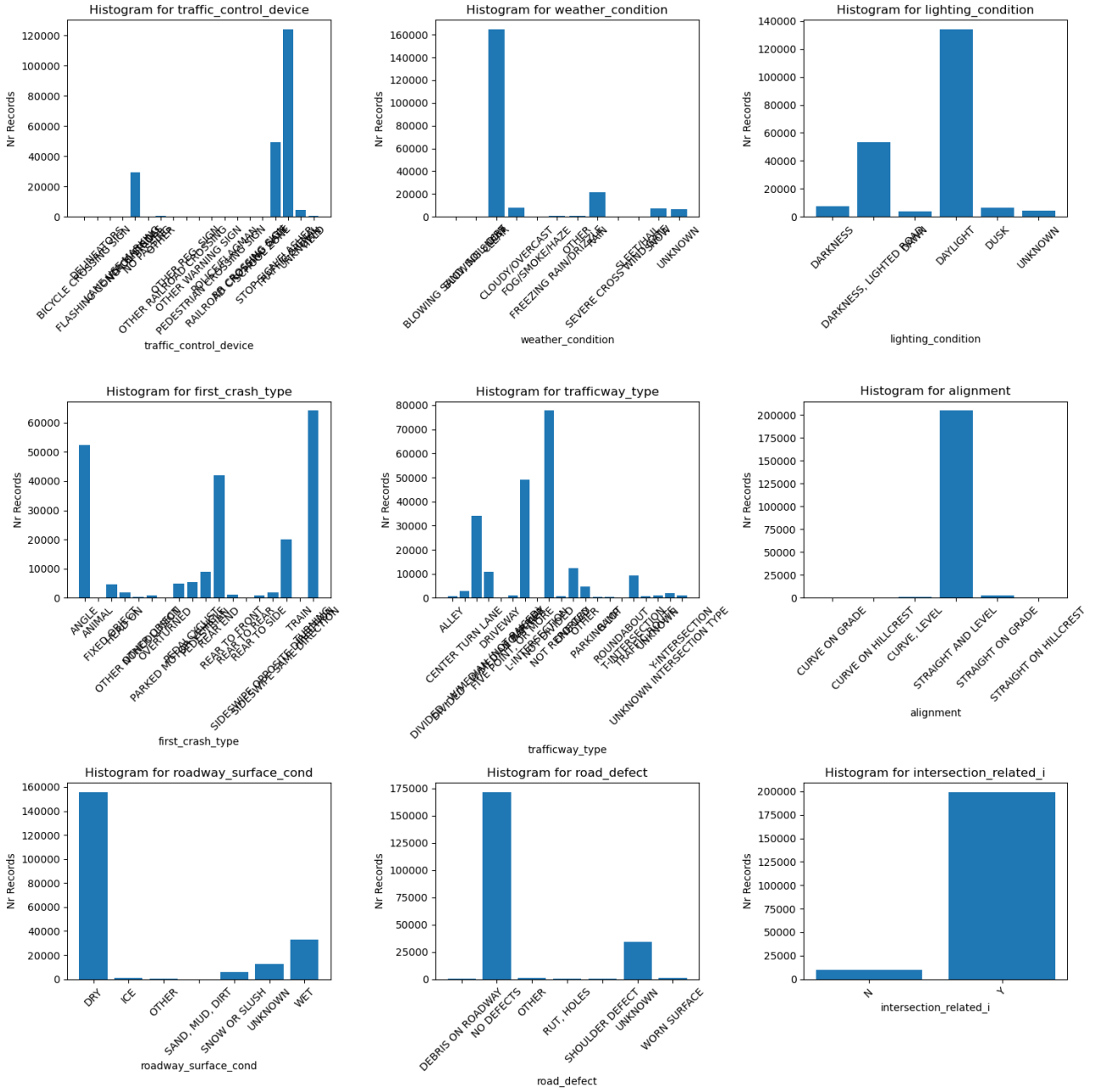


Figure 9: Histograms for symbolic variables - Accidents

1.3 Granularity

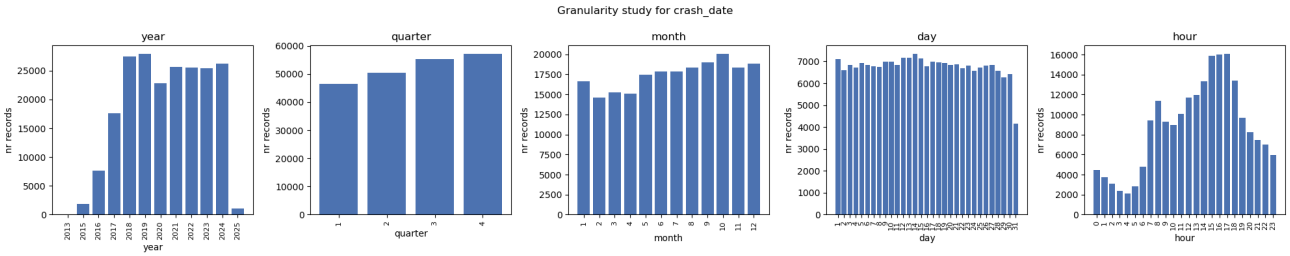


Figure 10: Temporal granularity for crash_date - Accidents

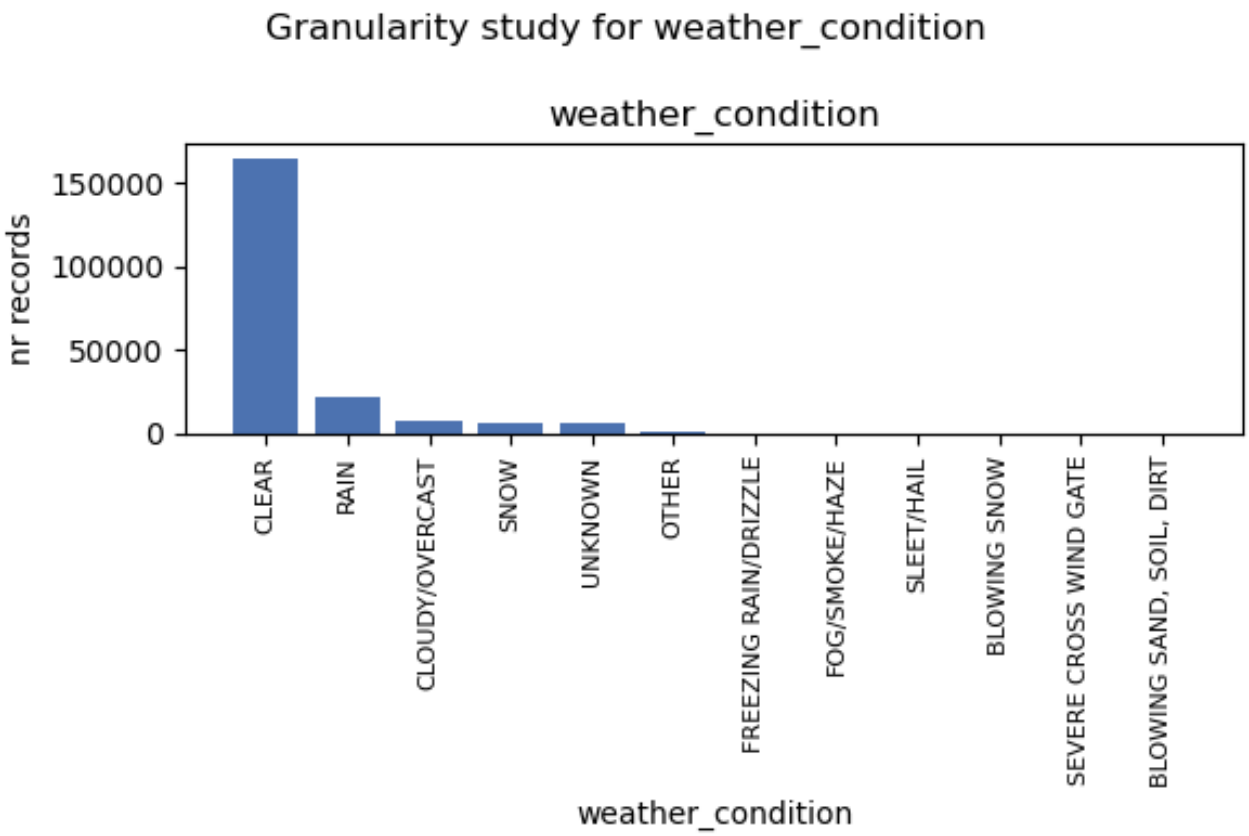


Figure 11: Granularity for weather_condition - Accidents

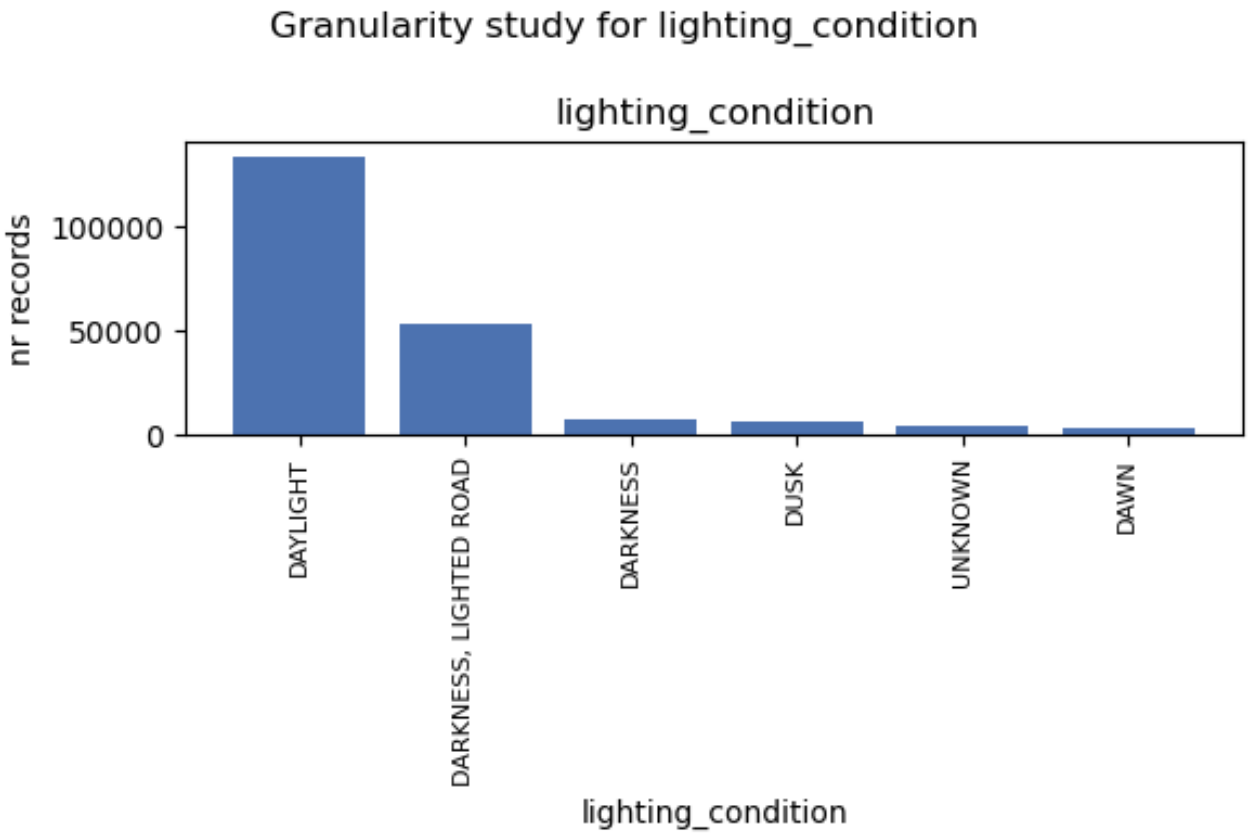


Figure 12: Granularity for lighting_condition - Accidents

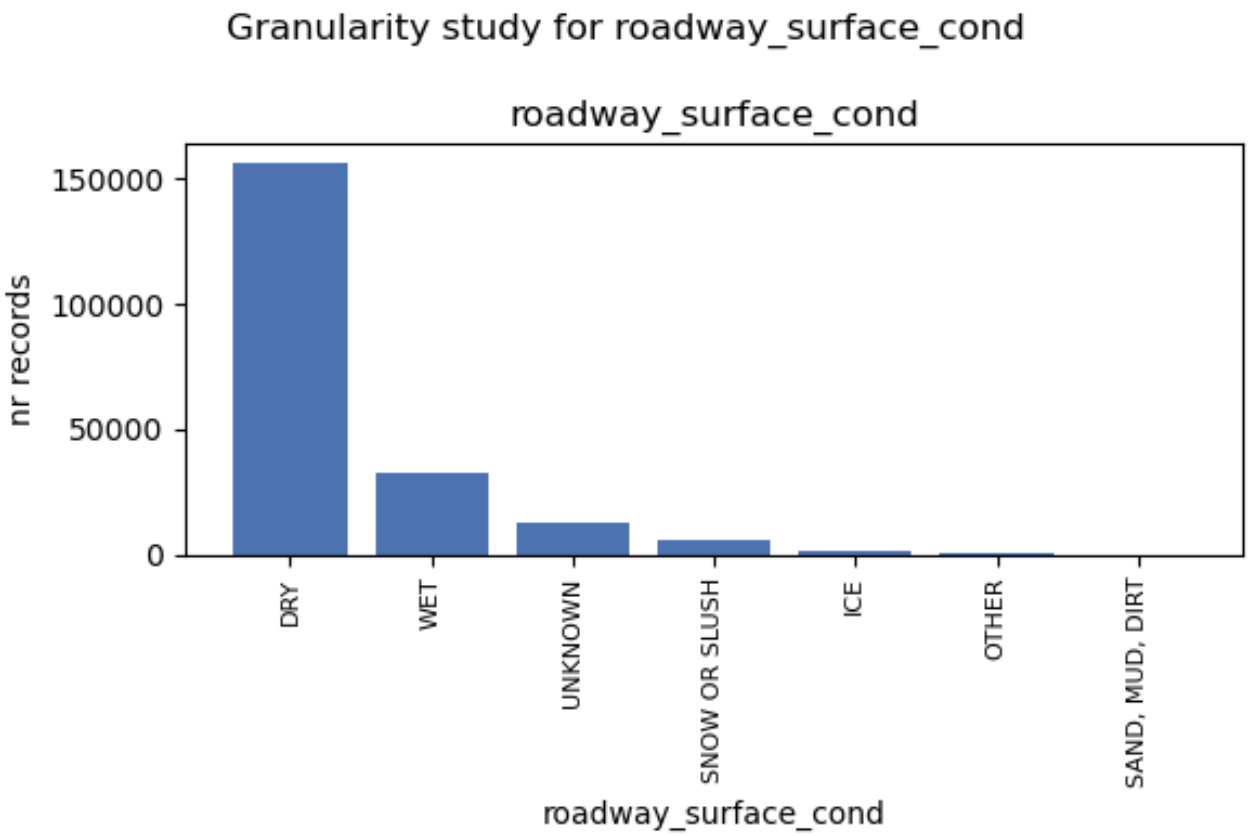


Figure 13: Granularity for roadway_surface_cond - Accidents

1.4 Sparsity

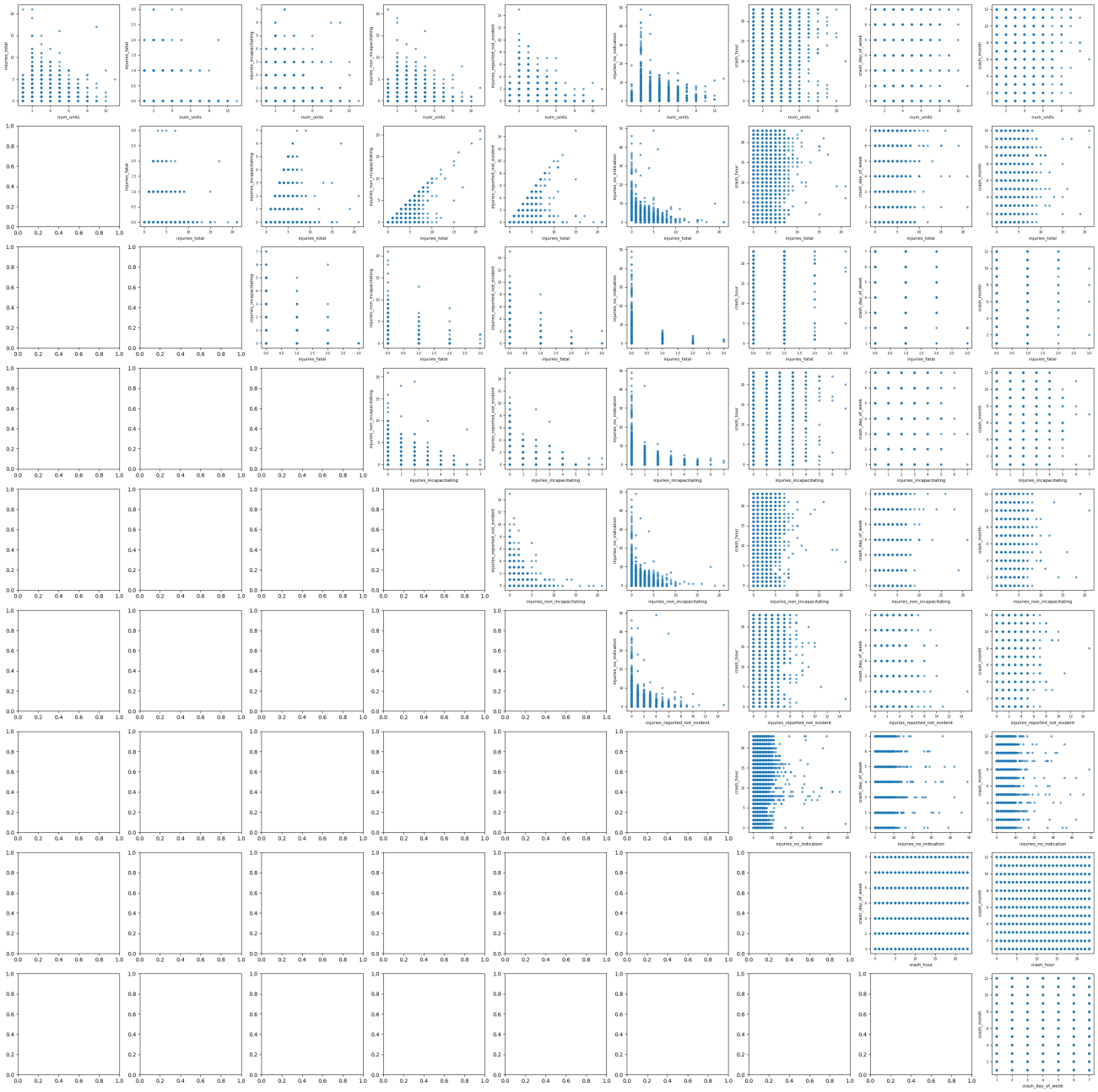


Figure 14: Sparsity study (scatter plots) - Accidents

Figure 15: Sparsity per class - Accidents

1.5 Correlation

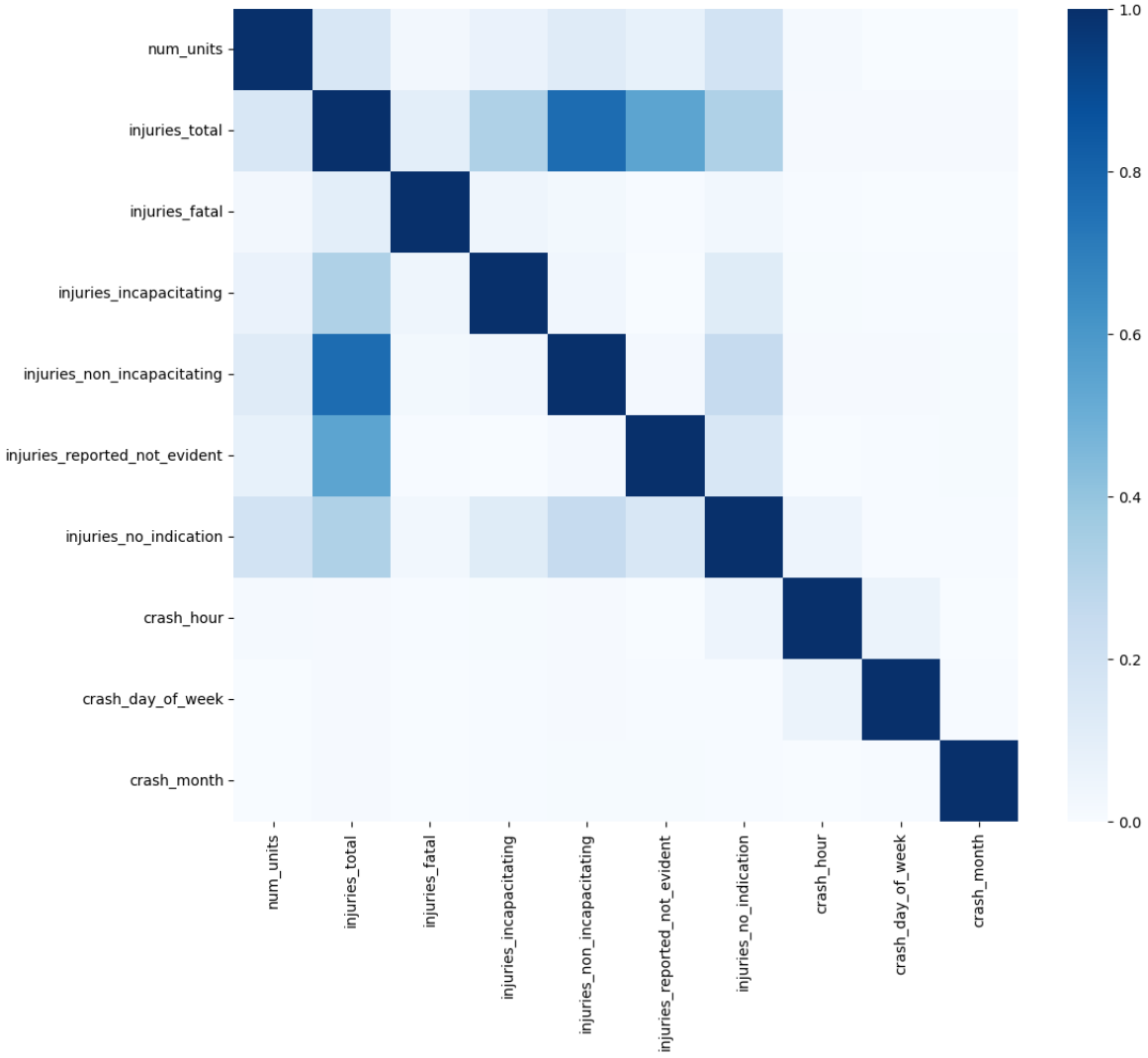
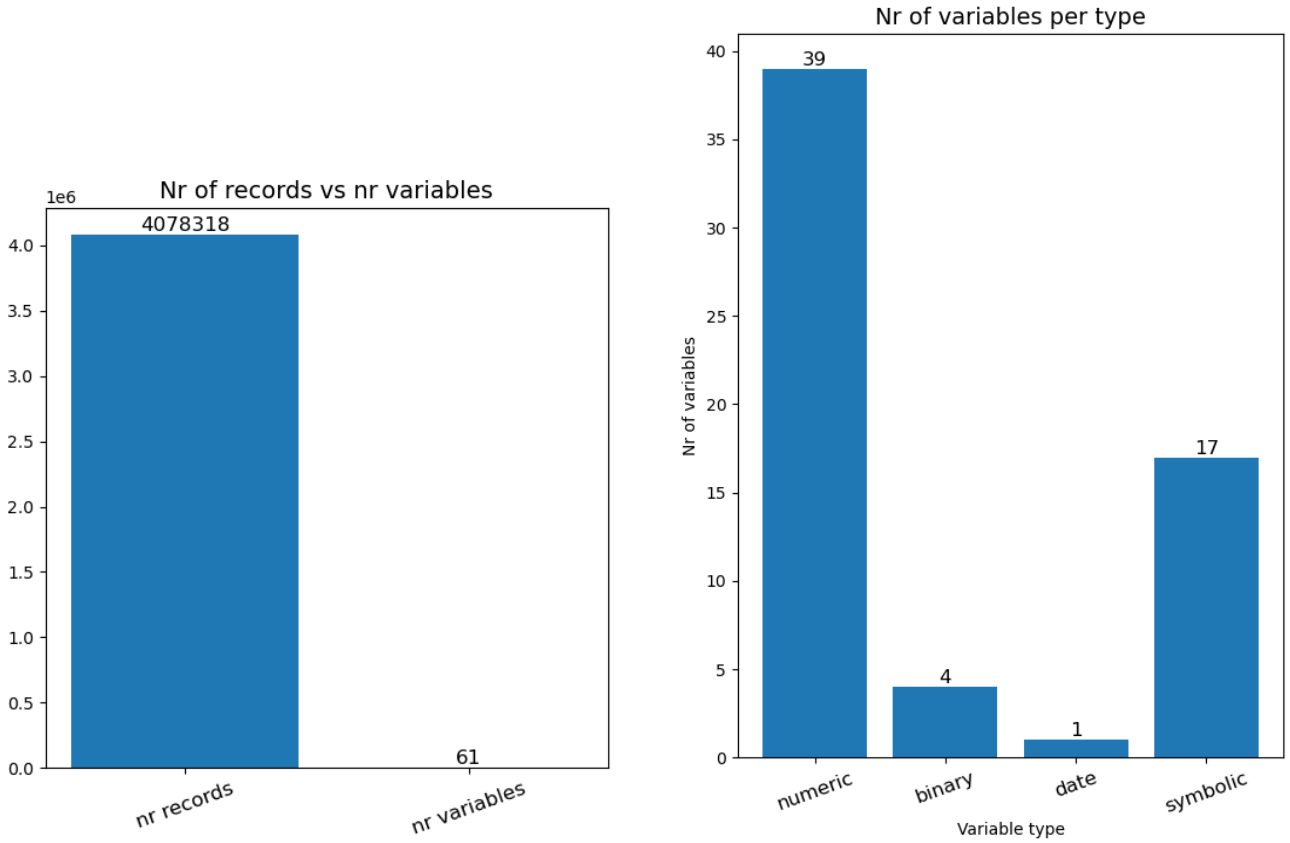


Figure 16: Correlation matrix - Accidents

2 Flights Dataset

2.1 Dimensionality



(a) Records vs Variables

(b) Variable Types

Figure 17: Dimensionality analysis for Flights dataset

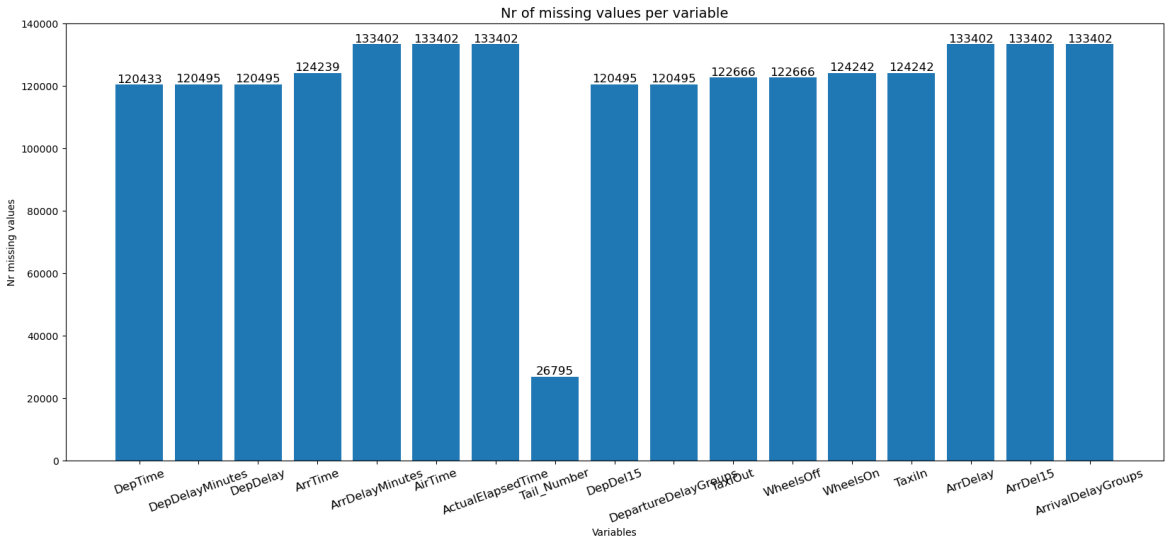


Figure 18: Missing values per variable - Flights

2.2 Distribution

2.2.1 Global Overview

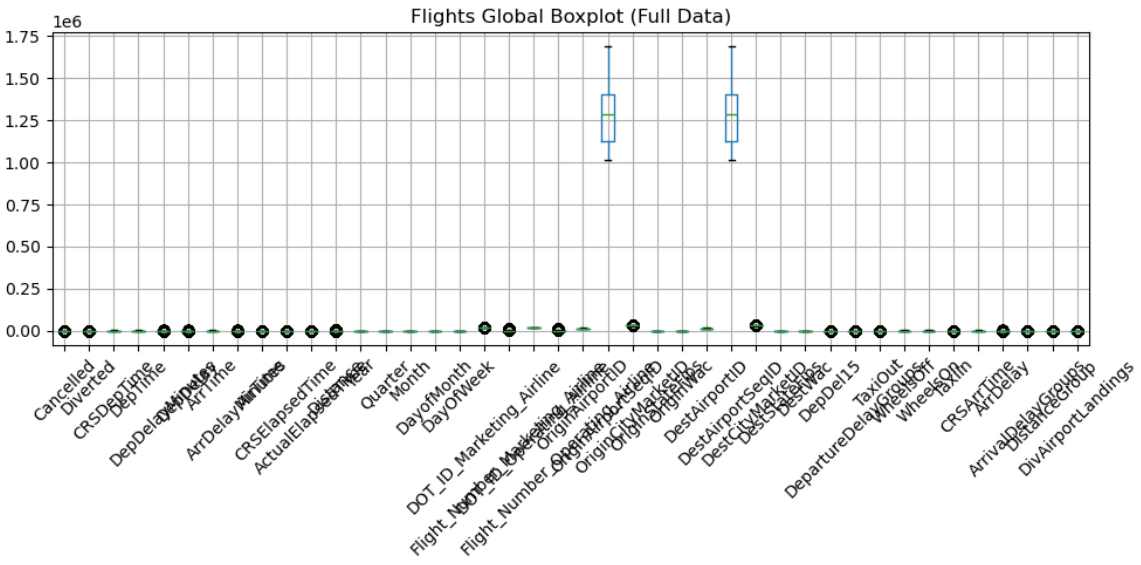


Figure 19: Global boxplot - Flights

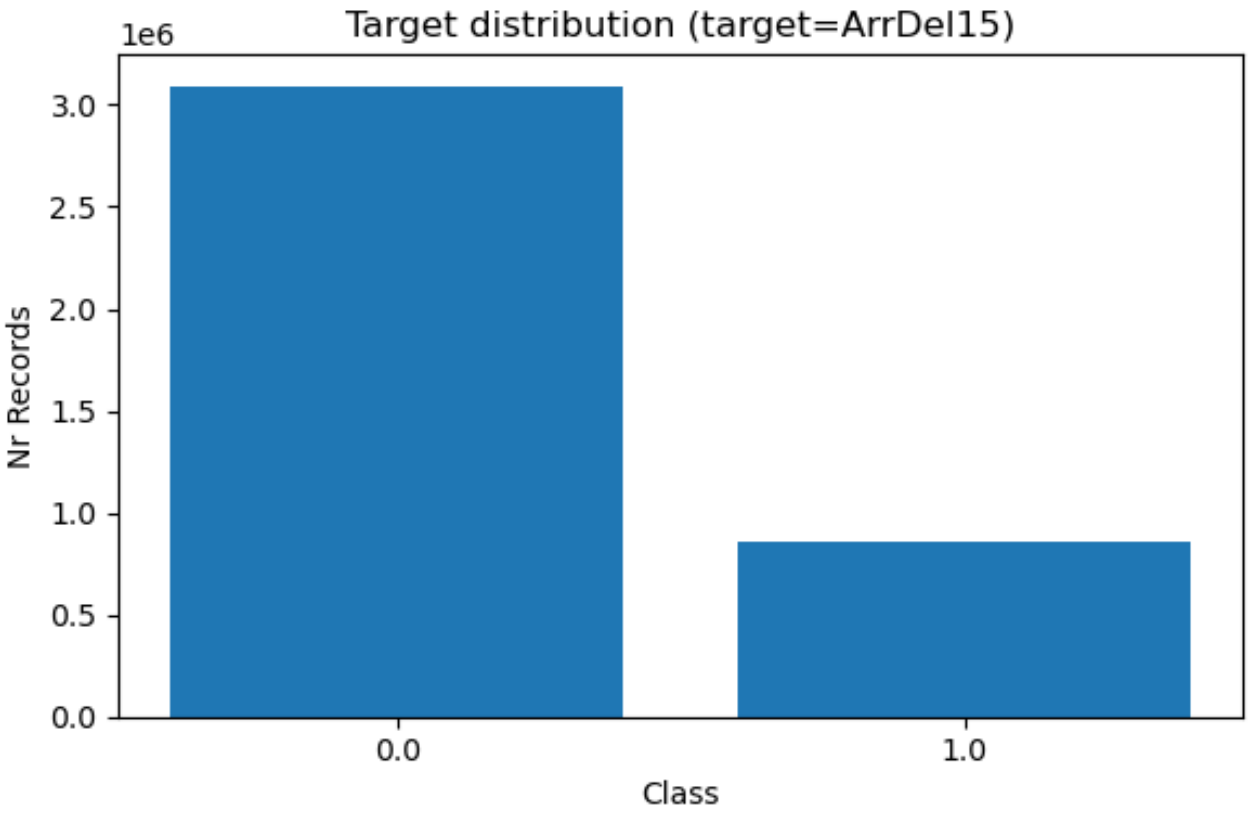


Figure 20: Class distribution - Flights

2.2.2 Numerical Variables

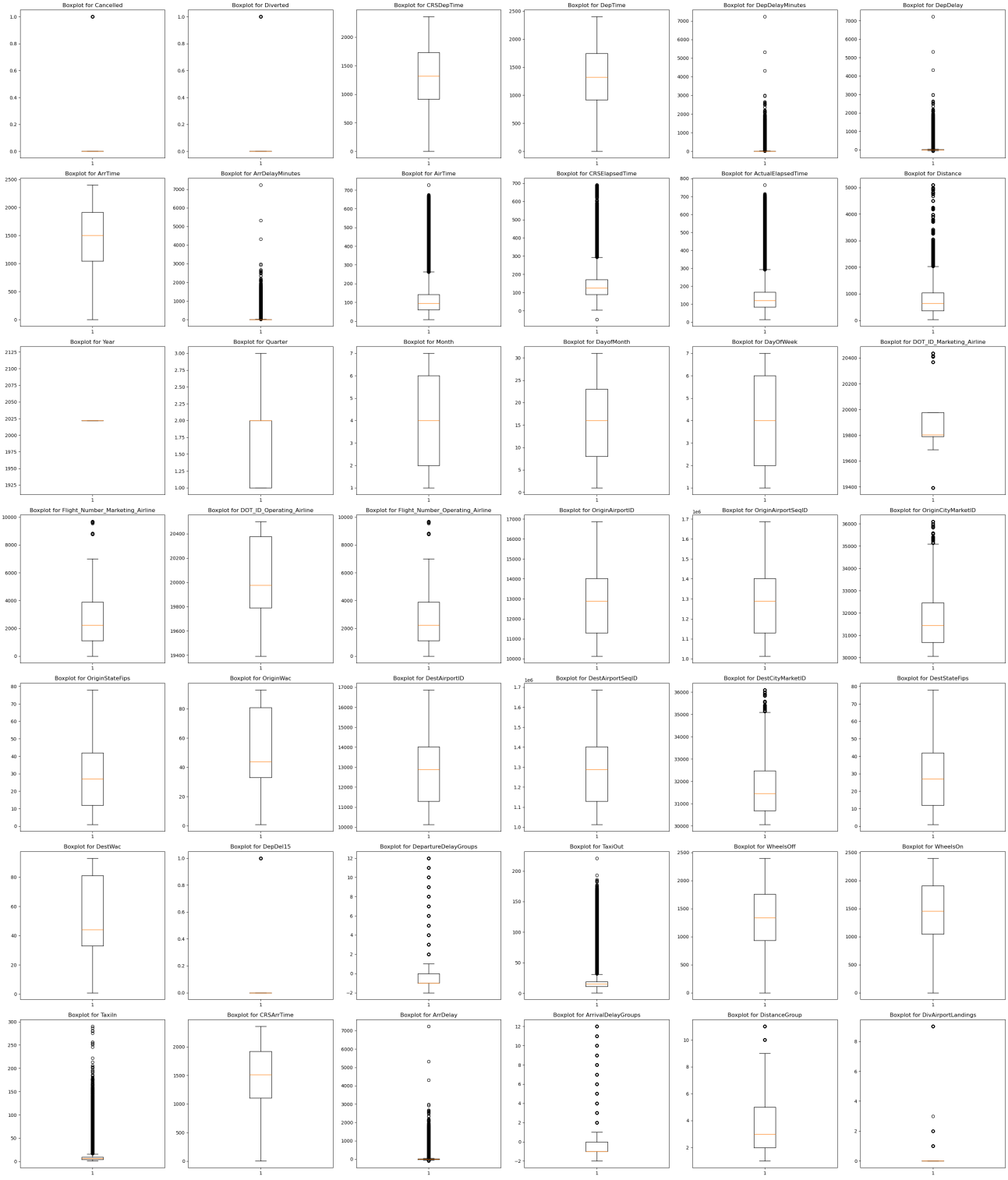


Figure 21: Individual boxplots - Flights

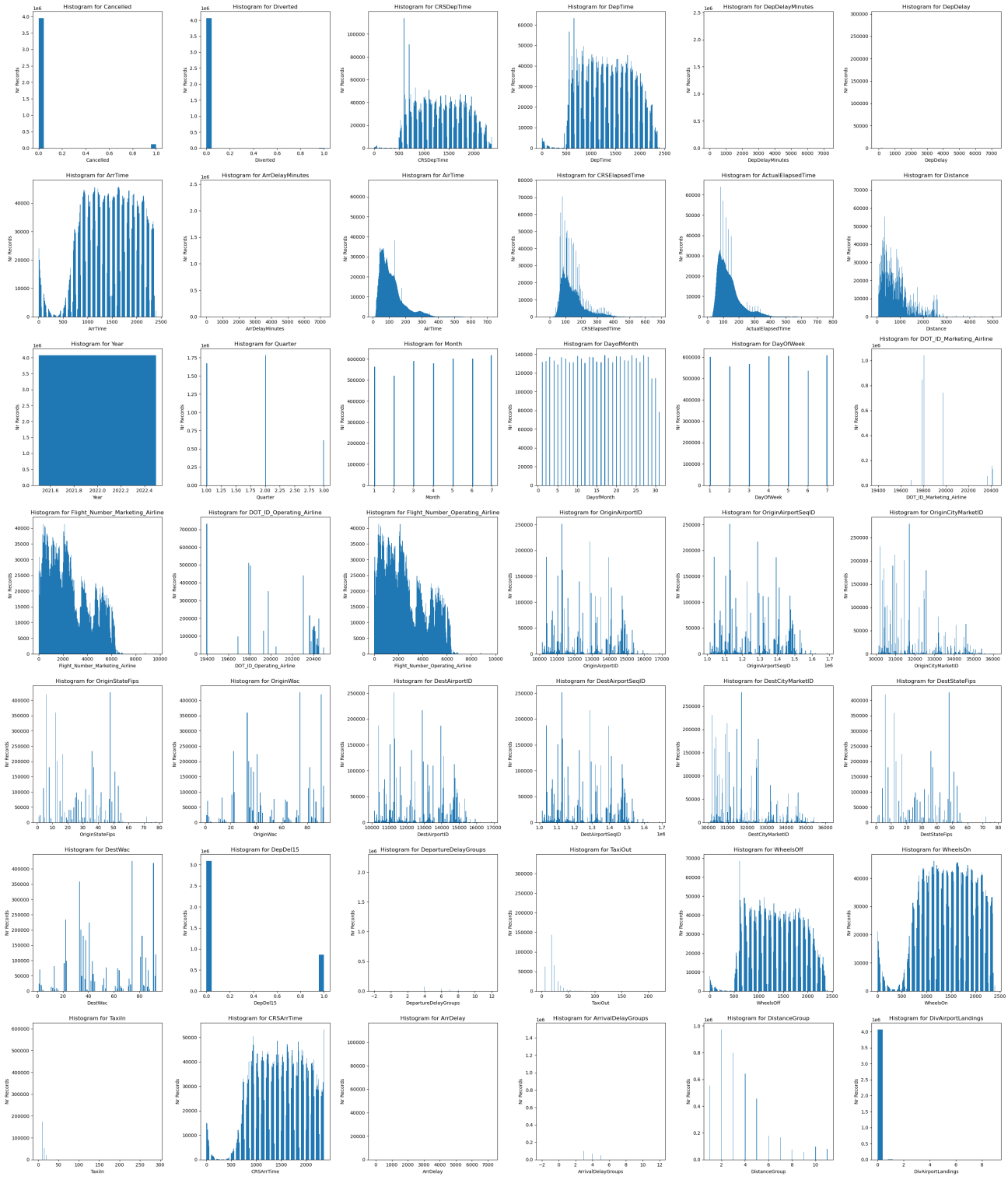


Figure 22: Histograms for numerical variables - Flights

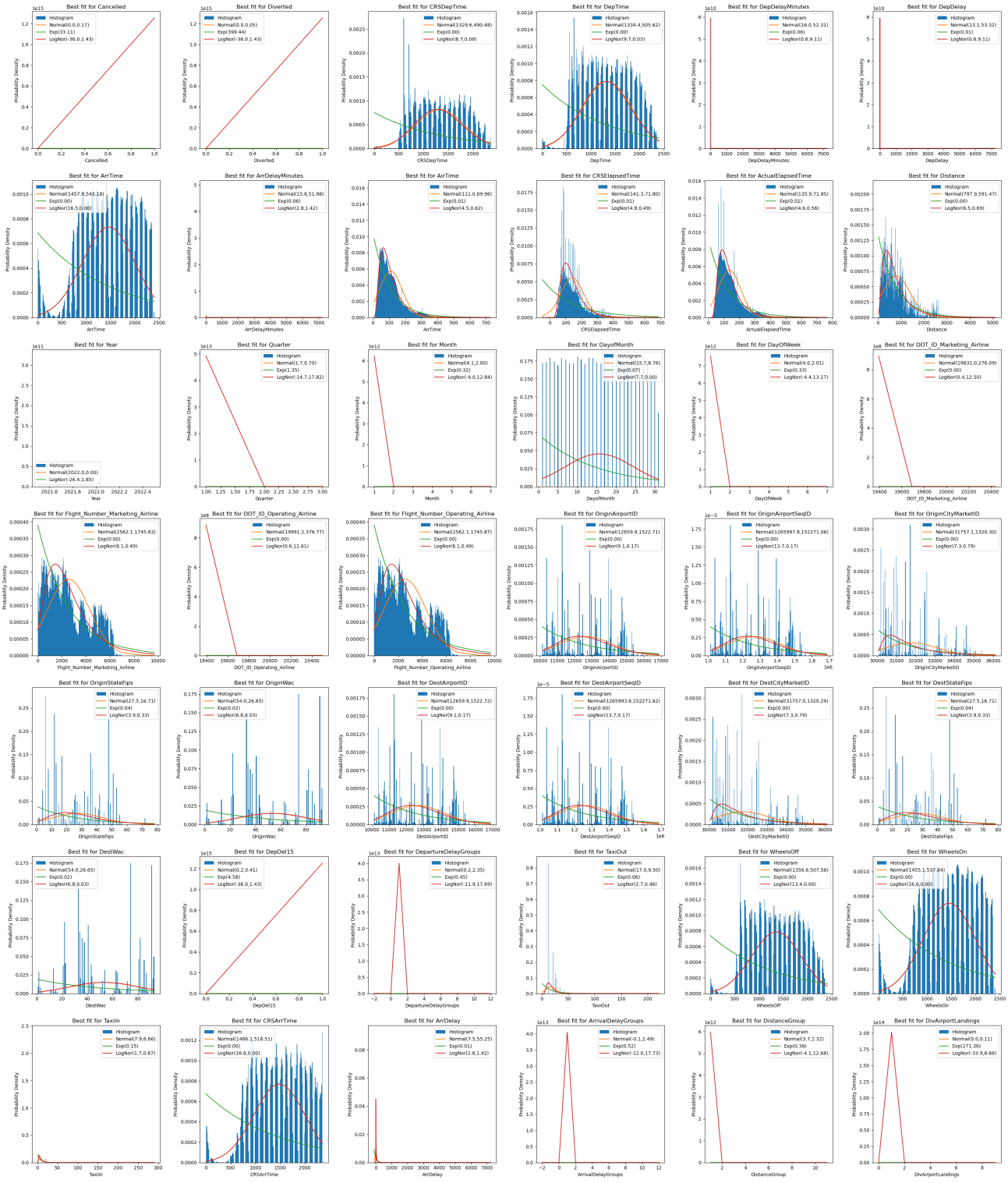


Figure 23: Probability density functions - Flights

2.2.3 Outliers Study

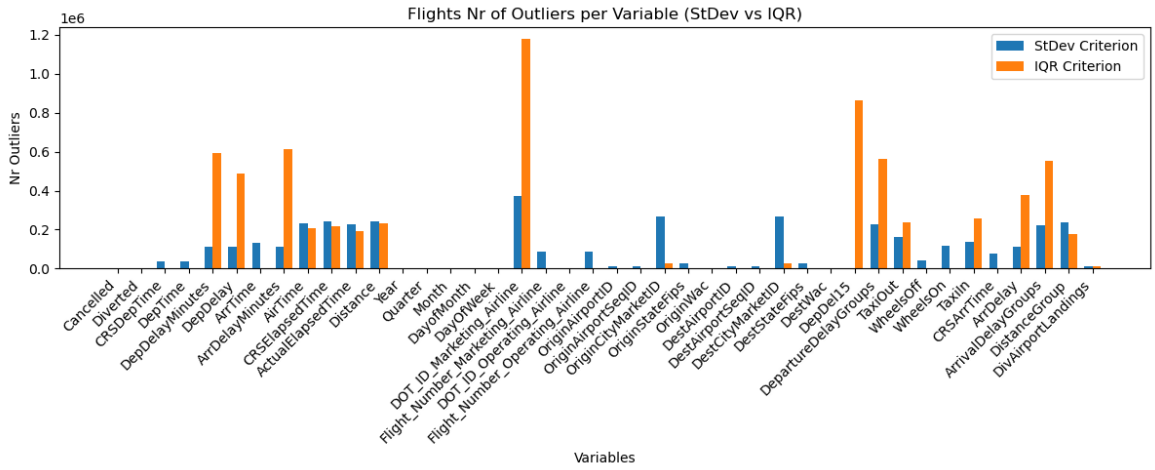


Figure 24: Outliers comparison (StDev vs IQR) - Flights

2.2.4 Symbolic Variables

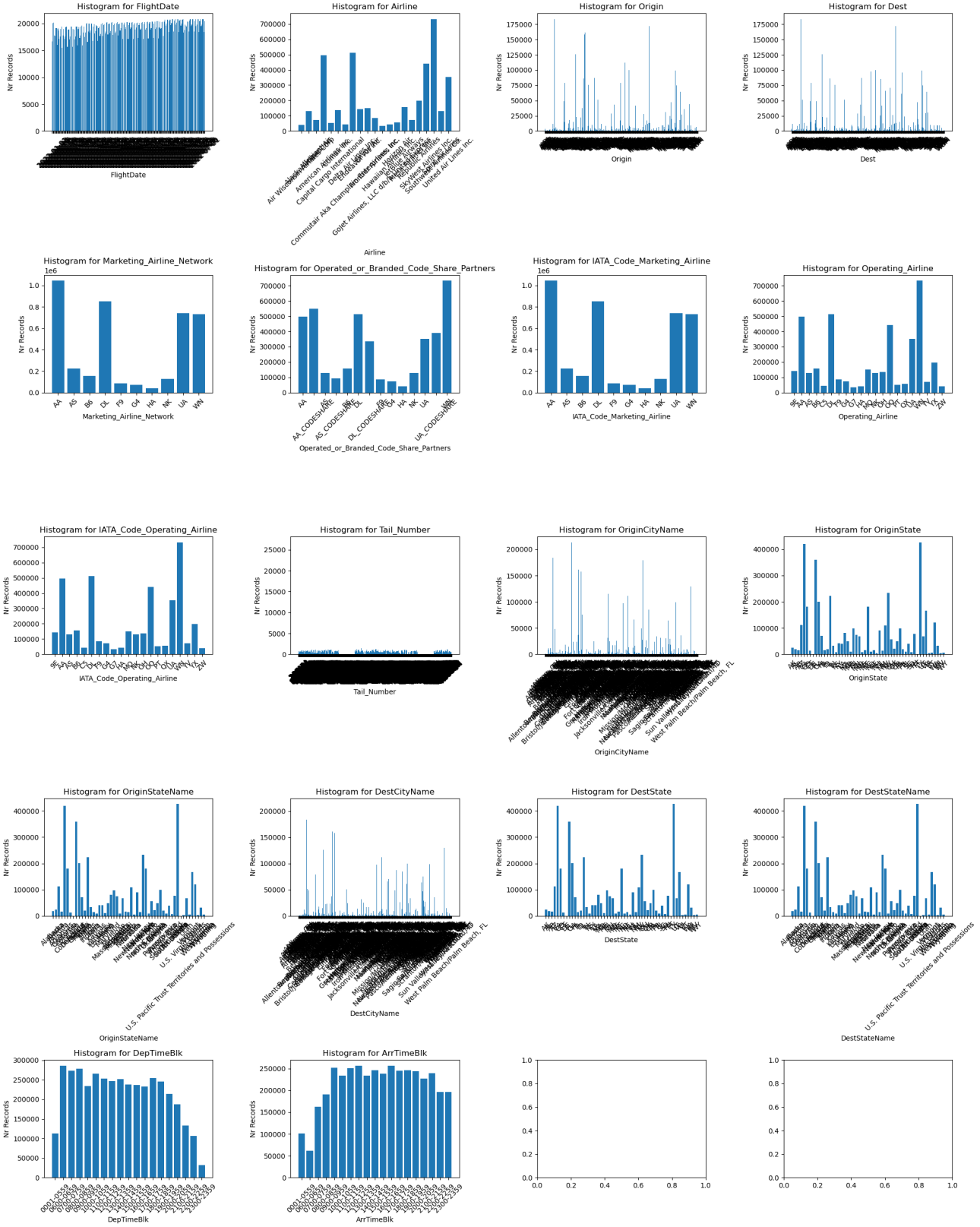


Figure 25: Histograms for symbolic variables - Flights

2.3 Granularity

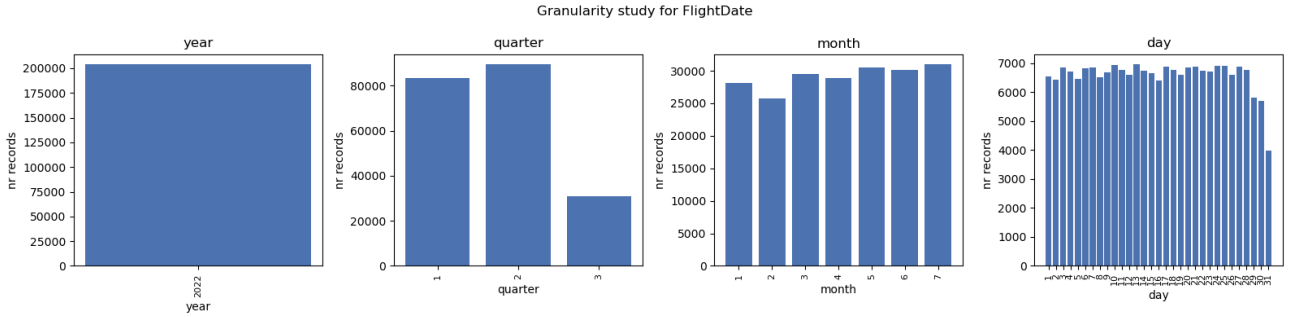


Figure 26: Temporal granularity for FlightDate - Flights

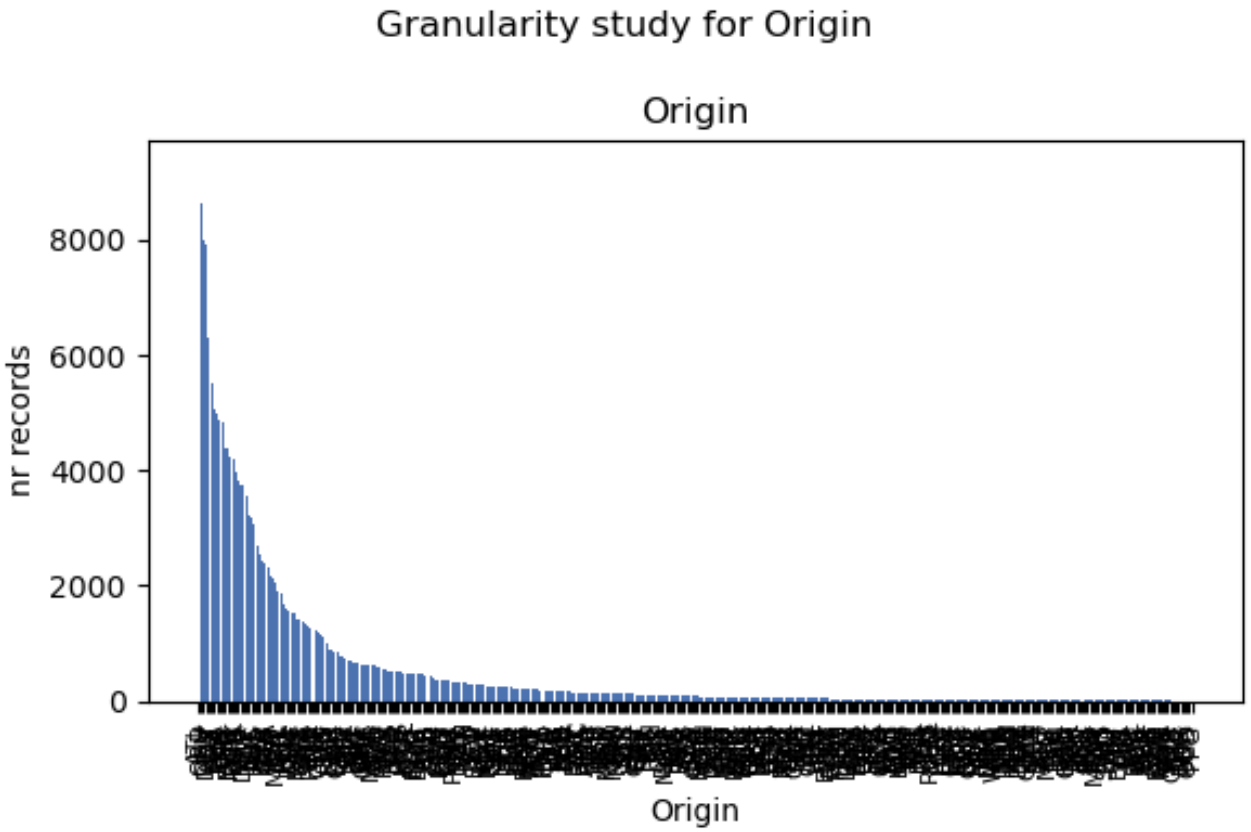


Figure 27: Granularity for Origin - Flights

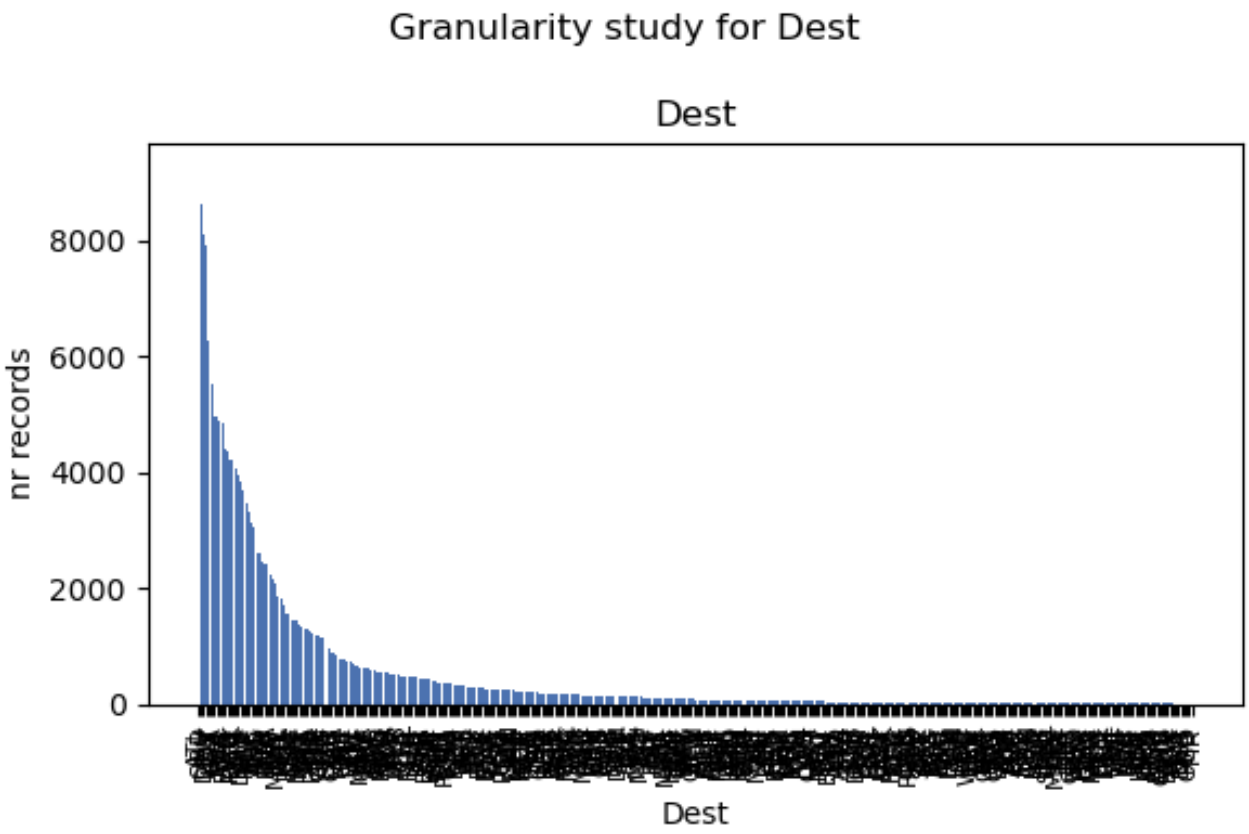


Figure 28: Granularity for Dest - Flights

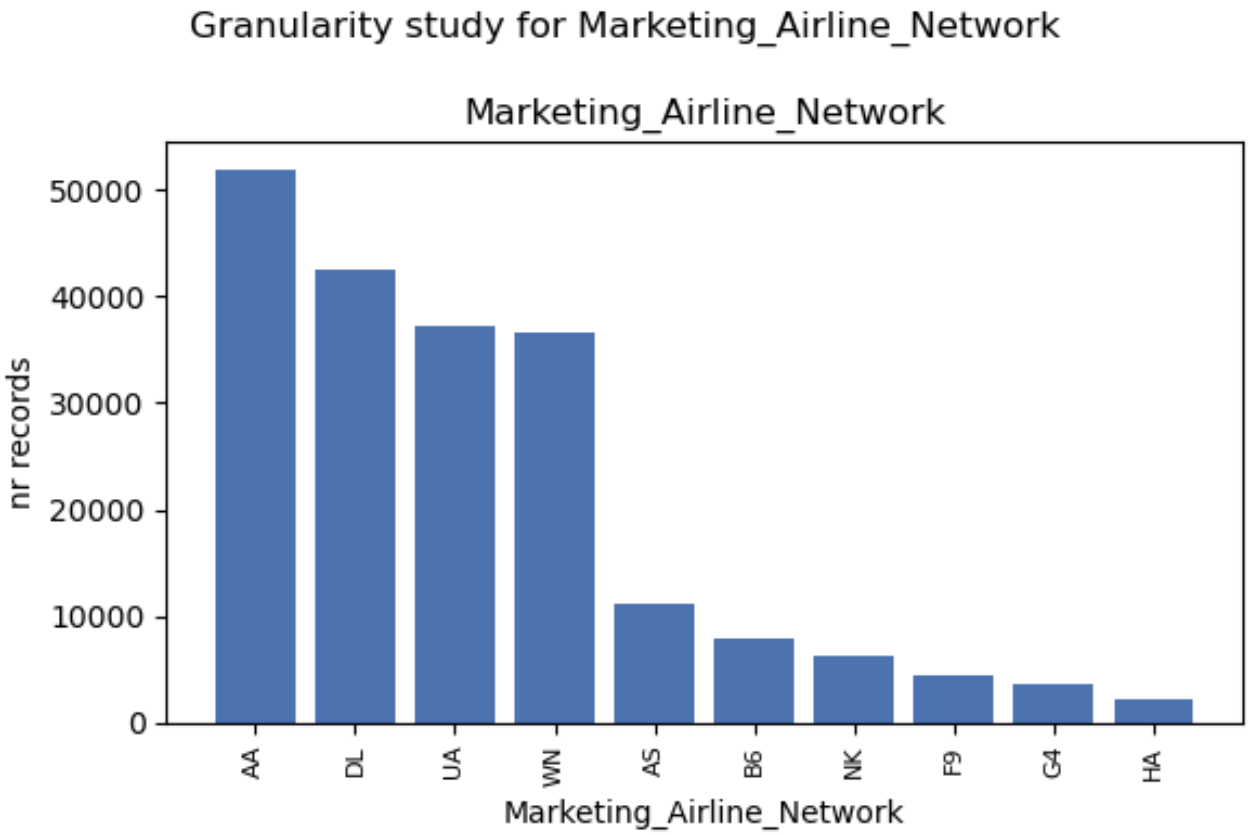


Figure 29: Granularity for Marketing_Airline_Network - Flights

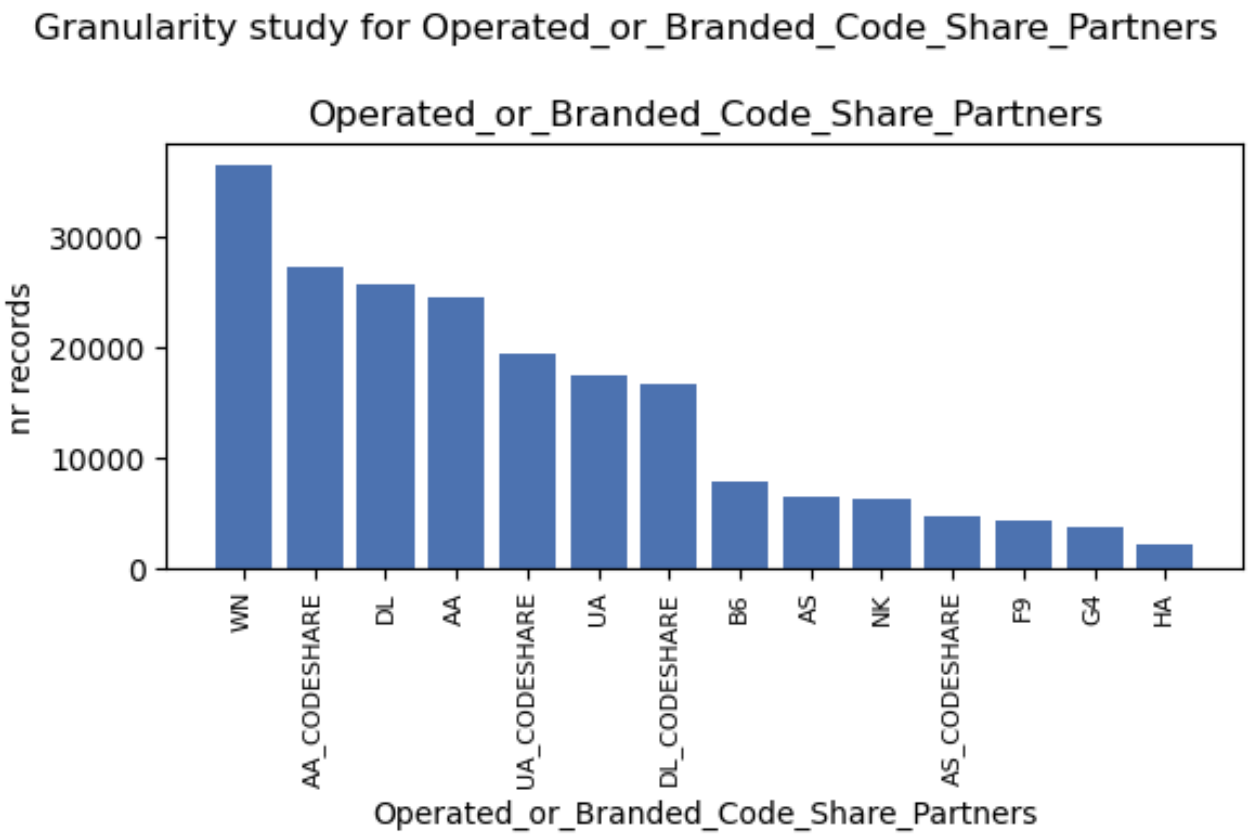


Figure 30: Granularity for Operated_or_Branded_Code_Share_Partners - Flights

2.4 Sparsity

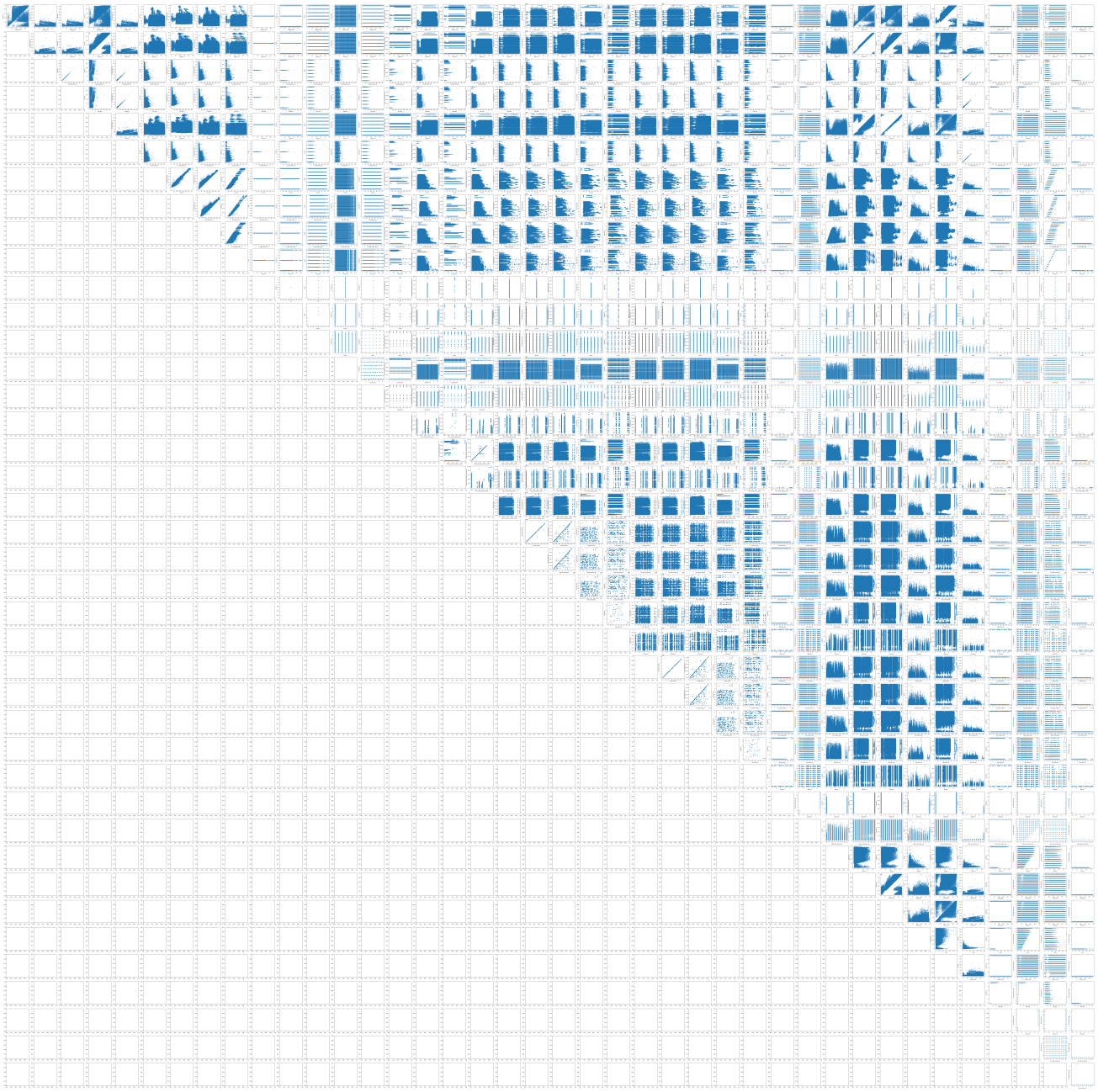


Figure 31: Sparsity study (scatter plots) - Flights

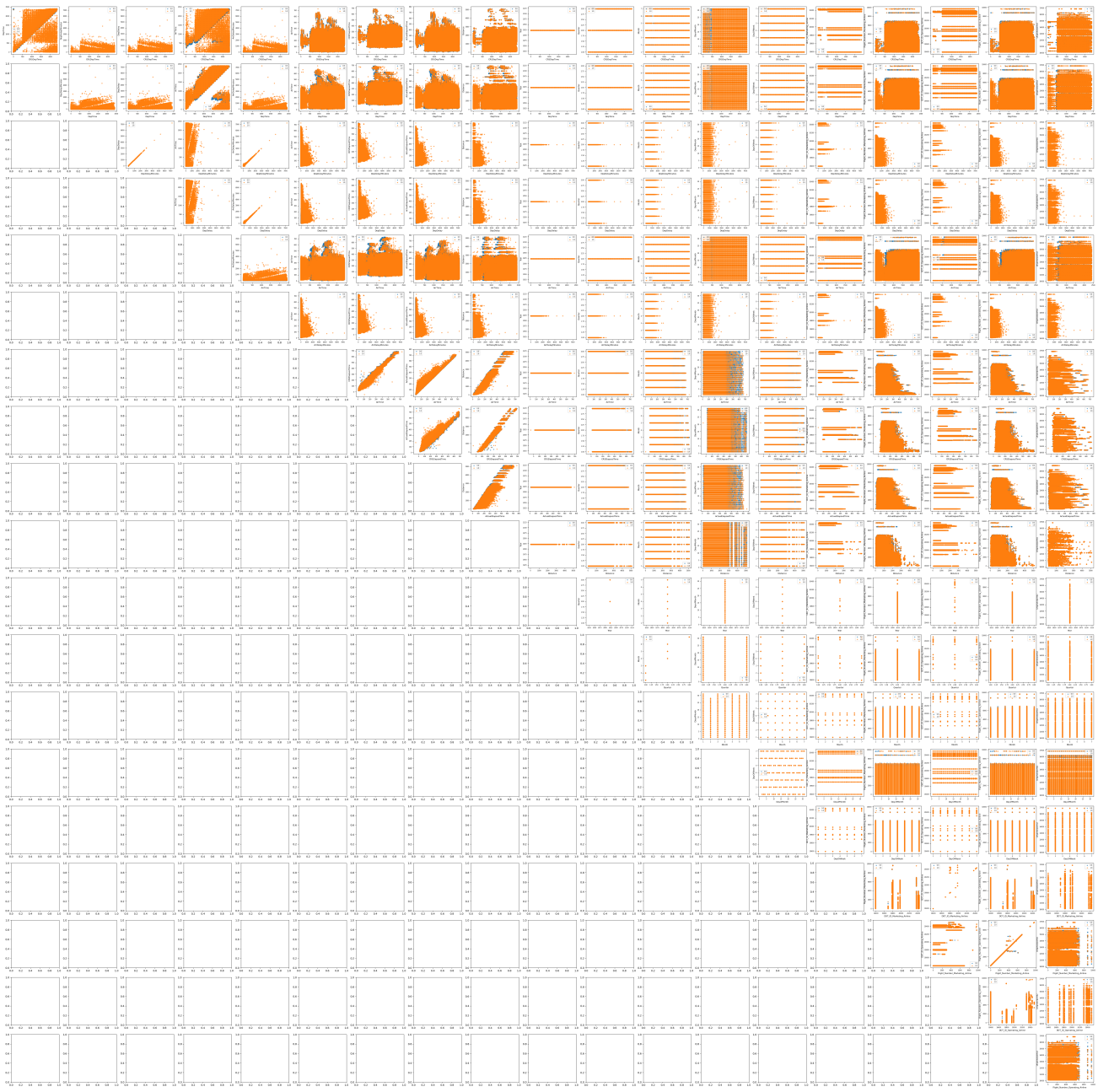


Figure 32: Sparsity per class - Flights (Chunk 1)

Figure 33: Sparsity per class - Flights (Chunk 2)

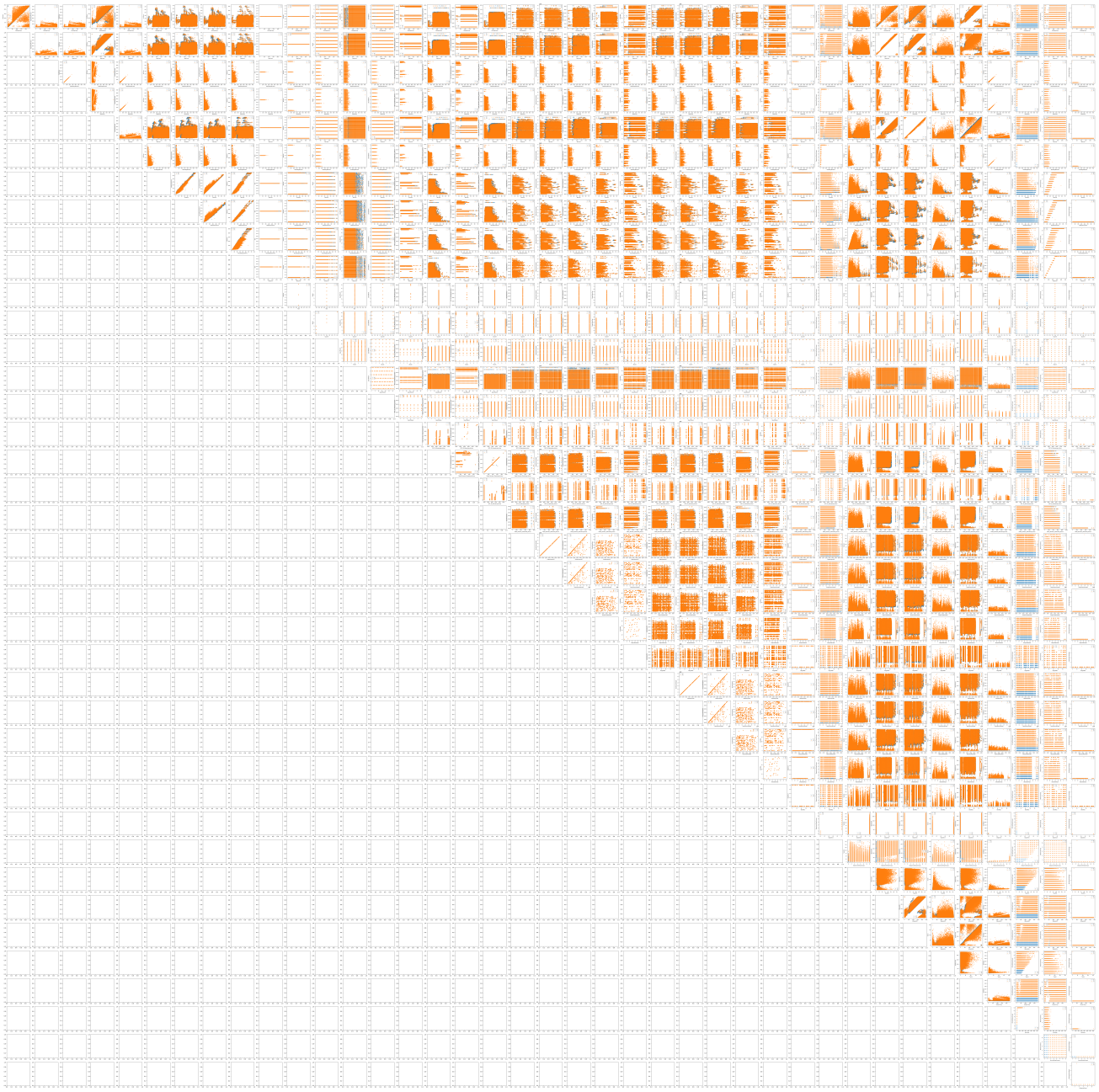


Figure 34: Sparsity per class - Flights

2.5 Correlation

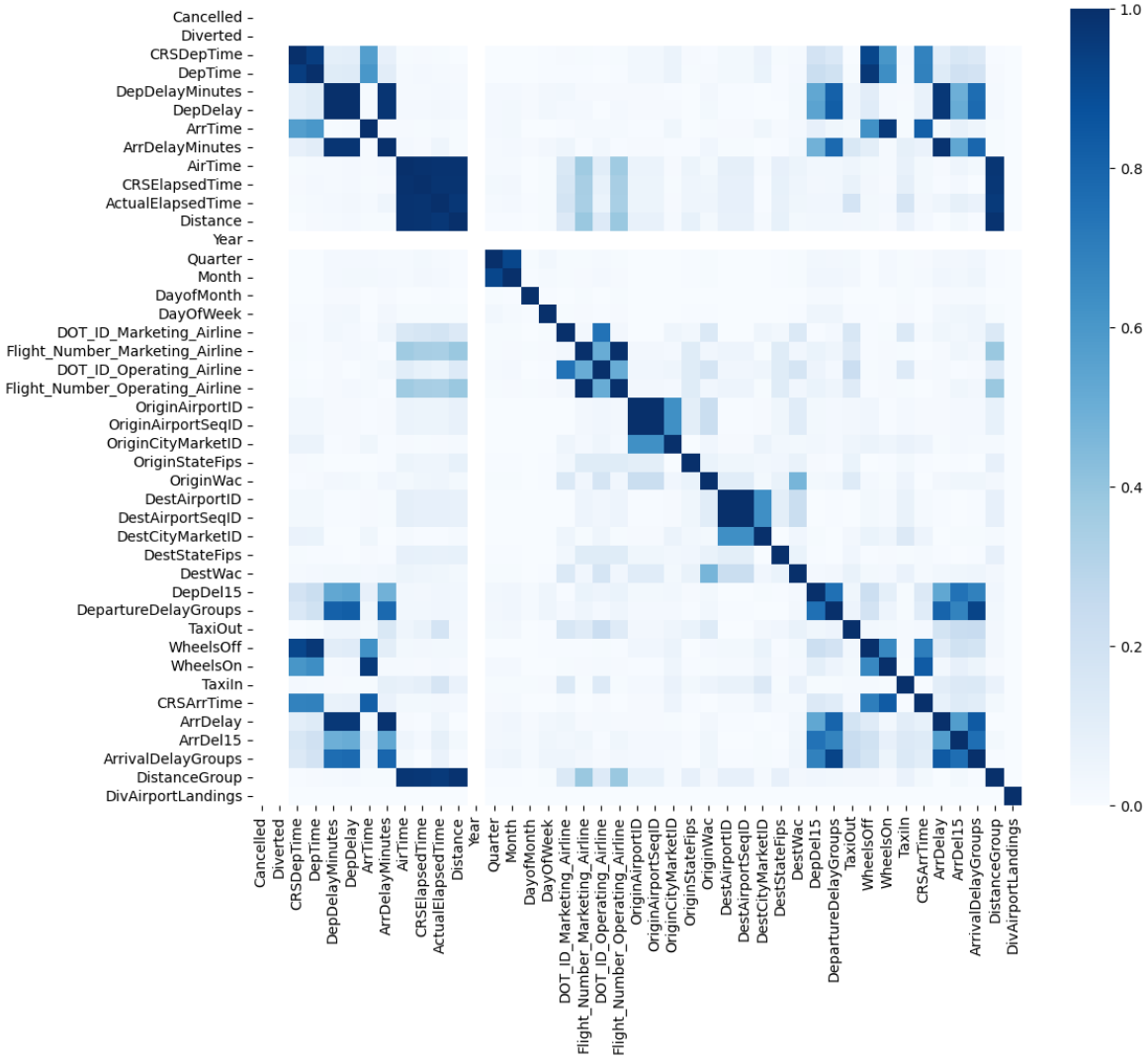


Figure 35: Correlation matrix - Flights