**Group 5 Sub-Report: PCA Observations and Outcomes**

Definition of Principal Component Analysis according to Wikipedia:

“Principal component analysis is a popular technique for analyzing large datasets containing a high number of dimensions/features per observation, increasing the interpretability of data while preserving the maximum amount of information, and enabling the visualization of multidimensional data.”

Here, we take a visualization-based data mining approach to PCA, plotting numerical attribute columns in a series of cross-plots to observe correlations and interesting data clusters throughout the training data frames collected from the ‘collisions’, ‘parties’, and ‘victims’ tables of the switrs.sqlite database file.

We begin with a cross-comparison of the numeric attributes under two alternate data preparation schemes: “as-is”, and “min-max normalized”. “As-is” attribute comparisons reveal general data trends, as well as correlation patterns between attributes with the units presented as they exist in the source data. “Min-max normalization” standardizes numeric attribute columns with respect to the minimum and maximum values in each attribute column.

We begin by viewing cross-plots of non-standardized data to observe general correlations and patterns, then view an identical set of cross-plots produced with min-max normalized data to reveal meaningful clusters highlighted by the normal vector space.

Note: Please disregard the yellow points, they represent a code artifact and are not semantically meaningful under this analysis.

BEGIN: AS-IS NUMERIC PCA CROSS-PLOTS

**Plot:**

A diagram of a party

Description automatically generated with medium confidence

**Observations:**

**Action:**

**Plot:**

**A graph of blue dots

Description automatically generated**

**Observations:**

**Action:**

**Plot:**

A diagram of a number of blue dots

Description automatically generated

**Observations:**

**Action:**

**Plot:**

A diagram of a graph

Description automatically generated with medium confidence

**Observations:**

**Action:**

**Plot:**

A graph with blue dots

Description automatically generated

**Observations:** (CLUSTERING)

**Action:**

END AS-IS NUMERICAL PCA ON COLLISIONS SET

BEGIN MIN-MAX NORMALIZED PCA ON COLLISIONS DATA

**Plot:**

A graph of blue dots

Description automatically generated

**Observations:**

**Action:**

**Plot:**

A graph with blue dots

Description automatically generated

**Observations:** (CLUSTERING)

**Action:**

**Plot:**

A graph of blue dots

Description automatically generated

**Observations:** (GOOD CLUSTERING)

**Action:**

**Plot:**

A graph with blue dots

Description automatically generated

**Observations:** (GOOD CLUSTERING)

**Action:**

**Plot:**

A diagram of a graph

Description automatically generated with medium confidence

**Observations:** (GOOD CLUSTERING)

**Action:**

**Plot:**

A diagram of a graph

Description automatically generated with medium confidence

**Observations:** (GOOD CLUSTERING)

**Action:**

**Plot:**

A graph with blue dots

Description automatically generated

**Observations:** (INTERESTING SUBJECT)

**Action:**

**Plot:**

A graph with blue dots

Description automatically generated

**Observations:**

**Action:** collisions at intersections tend to involve more injuries.

**Plot:**

A graph of blue dots

Description automatically generated

**Observations:**

**Action:**

**Plot:**

**Observations:**

**Action:**

**Plot:**

**Observations:**

**Action:**

**Plot:**

**Observations:**

**Action:**

**Plot:**

**Observations:**

**Action:**

**Plot:**

**Observations:**

**Action:**

**Plot:**

**Observations:**

**Action:**

**Plot:**

**Observations:**

**Action:**