Visualization Project:

Field/type:

1. Id – string, unique
2. Number of vehicles – int, continuous
3. Number of casualties – int, continuous
4. Date – date
5. Time – time
6. Speed limit – float, continuous
7. Road conditions – string, categorical
8. Weather conditions – string, categorical
9. Day – string, categorical
10. Road type – string, categorical
11. Gender – string, categorical
12. Severity – string, categorical
13. Age group – string, categorical

Data understanding:

the dataset contains information about bicycle accidents in the UK of the years 79-18, the data has 13 columns and 800k rows. The data showcases accidents with their respective elements, such as road conditions, speed limit, the date, time and light conditions as well as information about the people involved in the accident (only the drive).

Primary data preparation:

1. Road conditions – to int
2. Weather conditions – to int
3. Day – to int
4. Road type – to int
5. Gender – to int
6. Age group – binning, to int
7. Severity – binning, to int
8. Number of vehicles – binning, outliers
9. Number of casualties – binning, outliers
10. Light conditions – to int

Possible correlations & visualizations:

1. Severity by light conditions, road type, age group&gender
2. Casualties by Weather conditions, speed limit, road type, biker info
3. “ “ by number of vehicles