**Exploratory Data Analysis**

**Julian**

In general, check the influence of environmental factors on accidents. If they are large we can expect good results from a score. If not, we have to set expectations right.

Visualisations:

* Accidents per month (see over year trends)
* Accidents per weather condition
* Accidents per weather/road condition
* PCA/MCA to see which variables correlate most
* Dataset distribution over the chicago area (mapbox.com)
* Distribution of hotspots, how much is left when excluding these

**Mathieu**

Visualisations:

* Accidents per Hours in a day of a month/ in general
* Accidents per trafficway type (the road is divided, not divided, in one way)
* Accidents per traffic control devices (traffic signs, stop signs)

Group these analysis by category :

* Look at the number of accidents with a financial cost of “less than 500$”, “between 500-1500$” and “over 1500$” according to the month of the year, the road condition, the weather condition, the localisation of the accident,...
* Look at the number of accidents with fatal, serious, slight or superficial injury according to the month of the year, the road condition, the weather condition, the localisation of the accident,...

**Venkatesh**

Feature Engineering:

* Removal of the columns(features) that are irrelevant to our mission.
* Handling of missing values in the remaining features.
* Creation of new columns based on the existing in order to have a comprehensive feature set.
* Descriptive statistics (I will update the specific tasks as I dig into the dataset, usually, some results induce me to try out a new one).

**Louis**

* Calculate the dangerosity of road types (intersections, 2 way road)
* Correlation between the viariables
* Maybe a heat\*map
* Are people more dangerous at specific hours
* Does traffic change the danger ? (need traffic data of chicago)
* Are there accidents types with few casualties ?