SF – Transport Network Companies and Road Safety

Write one page describing the data set you are working with for your project, and explaining:

* Possible relationships in the data set that may be problematic
* How you could determine if they actually are a problem
* How you might mitigate the issue if you determine them to be a problem

Answer:

Possible relationships

1. There are 262 fields in my dataset. It is possible that out of these fields, few are subset of the other for combination of the other and hence there will be cases of correlation.
2. Zero crashes links - Second, as it was found and explained yesterday that the low speed local roads may have witnessed a very small number of crashes (fatality or injury) because they are the streets are narrower, speed of the vehicles cannot be greater than 15 miles and traffic on this type of segments is very limited to the neighborhood residents. However, local roads are the major percentage of any city’s road network and hence their rows in the dataset that would not have any traffic related casualty. This my have impact on our final output.
3. Multicollinearity – Since our analysis is like a before and after study and the base year and the target year is separated by 6 years the possibility of variables correlated is highly likely.

Determine if they are actually a problem

1. Zero crashes link are a problem because the expected number of crashes (and traffic related casualty) will go down as the % of these sub-types of road increases in the dataset.
2. Total number of crashes which is summation of total number of fatal and total number of injuries, may have corelated as they all three could be defined using total = fatal + injury relationship.

Mitigate these problems

1. I think by starting and building a basic model which includes the most obvious stuff like volume, capacity, and slowly improve the complexity by merging, summing or using derived fields
2. For zero-crashes link, maybe we would be likely use zero-inflated passion however if there exists some pre-requisite test to support its usage would be performed to justify its existence.