Pandemic in-> increased crash rate

From Washington state

Identify if fatal crashes are done by people of the same community so as to provide targeted measures to prevent crashes

Defining community area as multiple zip codes grouped together

High risk driver producing zip code

Person data and crash data

Crash event: vehicles, persons,

Analyze driver w

Person is driver hen person type equals 1

Coordinate and zip code mapping

A file that gives the central point of a zip code area?

Level 4: reversing geo coding, taking coordinate and yield a geo coding

Challenge: getting zip code using crash coordinate

Allowed to integrate public data

Machine learning model for crash prediction is welcomed.

Workflow:

* Step 1: take the information of each row (preferably its zip code; a crash may contain more than one zip code) and derive zip code
* Step 2: identify the proportion of crashes in which the involved persons are of the community where the crash happened
* Step 3: identify if there is significantly difference regarding the features of resident crash and non-resident crash
* Step 4: identify if there are communities that produce high risk drivers
* Step 5: identify the population demography of said risk driver producing communities

Rubic:

* Adding additional data
* Adding additional insights

Data

* Report number: crash level id
* May have records lacking zip code, better drop
* Only interested in issues since covid so don’t date back too much

Strategies:

* Working backward?
* Story telling
* pictures\ graphs\
* presentation: all students participated
* K.I.S.S. keep it simple stupid

Tech stack:

* Python +tableau

Others

* Cartopy a geospatial data processing lib
* Qgis gis package

Data EDA:

* Column par is the UID for an event.
* The whole table is yielded by joining person and crash table. So the tuple (par, vnumber) constitutes the UID for a row where par is the event id and vnumber is the person id