**Project Outline/Plan**

1. Research Proposal
   1. By analyzing car crash reports from nyc, I will use PSM and EB to see if there is any casual inference on bigger cars and car crashes
2. Question Research
   1. Do bigger car cause more crashes on the road?
3. Cleaning data
   1. [Columns of the dataset](%20https:/data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about_data)
      1. ~~Things we can remove:~~ 
         * + ~~time~~
           + ~~date~~
           + ~~zip code~~
           + ~~latitude~~
           + ~~longitude~~
           + ~~location~~
           + ~~anything related to street~~
           + ~~remove every specific person killed, only have the overall number (eg. remove pedestrian, cyclist, motorist)~~
      2. ~~Only include frequency that happen often~~
      3. ~~Create dummy variables for the following~~
         1. ~~Truck~~
         2. ~~Van~~
         3. ~~Sedan~~
         4. ~~Motorcycle~~
         5. ~~Bike~~
         6. ~~Bus~~

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. ~~Create a dummy variable if a person was killed, and another one if they were killed~~
    2. ~~Create a dummy variable that represents 1 if the crash was from a truck, sedan, or bike~~

1. ~~Implementation of models~~
   1. ~~PSM~~
   2. ~~Entropy Balancing~~