Look into the negative values

* If the household have e.g (-9, 20.2) -> 20.2
* Appropriate skip -> 0
* Mean/Median for thos who don’t know
* Look at those with big proportion (20%)
* **> 15% drop**
* **Check the variables and do aggregation and categorisation**

**MSASIZE**

**Jiaqi**

**BIKESHARE** Count of Bike Share Program Usage – average

**CARRODE** Count of People in Vehicle to Work - average

**CARSHARE** Count of Car Share Program Usage – average

**CNTTDTR** Count of person trips on travel day - average

**DISTTOSC17** Road network distance, in miles, between respondent's home address and school address – average, have to check if got student in the family

**FLEXTIME** Flex Time – ratio (1-yes, 2 – no) – **remove**

**GT1JBLWK** More than One Job - COUNT

**HEALTH** Opinion of Health – average (higher is better)

**HHSIZE** Count of household members

**MCUSED** Count of Motorcycle or Moped Trips - AVERAGE

**NBIKETRP** Count of Bike Trips - AVG

**WKFMHMXX** Count of Days Worked From Home in Last Month

**Wei Ji**

**~~NOCONG~~** ~~Trip Time in Minutes to Work without Traffic - AVG~~

**~~TIMETOWK~~** ~~Trip Time to Work in Minutes - AVG~~

**NUMADLT** Count of adult household members at least 18 years old

**~~NWALKTRP~~** ~~Count of Walk Trips - AVG~~

**~~OCCAT~~** ~~Job Category – COUNT~~

**~~PTUSED~~** ~~Count of Public Transit Usage - AVG~~

**~~PUBTIME~~** ~~Minutes Spent Transferring to Work – AVG~~

**~~RIDESHARE~~** ~~Count of Rideshare App Usage - AVG~~

**~~R\_AGE\_IMP~~** ~~Age (imputed) – AVG - done~~

**~~R\_SEX\_IMP~~** ~~Gender (imputed – male%, female%(adults) – done~~

**WKRMHM** Option of Working from Home

**WRKTIME** Arrival Time at Work

**~~WKFTPT~~** ~~-> ratio~~

**~~DISTTOWK17~~**Road network distance, in miles, between respondent's home address and work address – average -> have to check if got worker

**~~EDUC~~** ~~Educational Attainment – average (higher is better) - done~~

Later – one hot encoding/bin