11/10   Analyzing NBHDS   Call w/ Coline

1 month to go w/ springboard

Predict # of visitors in each Census Block Group

Bands, regional likes, time to visit   Each Day + Each Hour

Where public transportation

#1 data   13 columns   #2 dataset   2 columns
                                               Lat + long
  CBG # ← Index
                                  [220,000 rows]
  Raw visit Counts
                                    Covers the entire U.S.
  1 month (March)
                                    Remove NA
  Unique # of visitors

  🔵 work versus home CBG
        CBG

     Distance from home → median distance for all visitors

     Brands → Top brands 10 per CBG

     Related same day versus same month

     popularity by hour

     SafeGraph & the US Census Bureau

        [�]

           ↳ Create better information from Census BG

  Data Wrangling → lots of work
  Raw Visit count Column

  Rate for each HR 1:24 + Day 1:7
  Max, min, & average → Visitors from CBG

Clusters   by geographic area by optics algorithm
            by visitor per day & per hour

k-means successfully w/ 5 clusters elbow
            inertia measure & method between 2 & 15

Weekdays for Raw Visit counts

Hourly differences didn't make sense
# of Visitors per month

Decision Tree to identify Representative Hour

Cluster 3 → smaller in terms of data

Business Development Manager → Link w/ Business needs
                                        &
                            Data Science

Coder.io