Who rides Uber Anyway? Analysis of ride-shares in Chicago during 2020

Shakil Rafi[1], Arna Nithila[2], & Anindya Debnath[3]
University of Arkansas, Fayetteville, AR
[1]Department of Mathematical Sciences, University of Arkansas
[2,3]Department of Civil Engineering, University of Arkansas

INTRODUCTION

With the advent of recent legislation, we have had an unprecedented access to ride-sharing data for large cities.

We take the data for the year 2020 and aim to segment the data a census tract level. We take inspiration of Soria, Chen, Stathoupolos, 2020, and use K-Means clustering to segment our data.

We notice that our data can be split into three clusters. We name these clusters rich, medium, and poor.

Most tracts are in the medium, followed by poor followed by rich.

METHOD

We drew data from two sources, the City of Chicago Data Portal and the Smart Locations Database by the EPA.

Data was cleaned and descriptive statistics achieved.

Exploratory data analysis was conducted, including pair-plots and a correlation heat-map constructed

An elbow method analysis was done optimizing for within-cluster-sum-of-squares against *k*.

A marked change in slope was noticed at k=3, indicating an “elbow”.

Richer neighborhoods in Chicago take more **frequent** but **shorter** trips than poorer neighborhoods.

Car-ownership is **not** a very good predictor of frequency of ride-shares.

Lower income neighborhoods tend to generate **longer trips.**

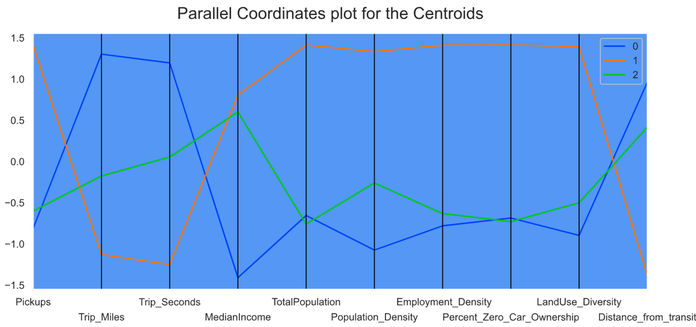
Employment density and **land use mix** is higher in richer neighborhoods.

This material is based upon work supported by the National Science Foundation under Award No. OIA-1946391. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

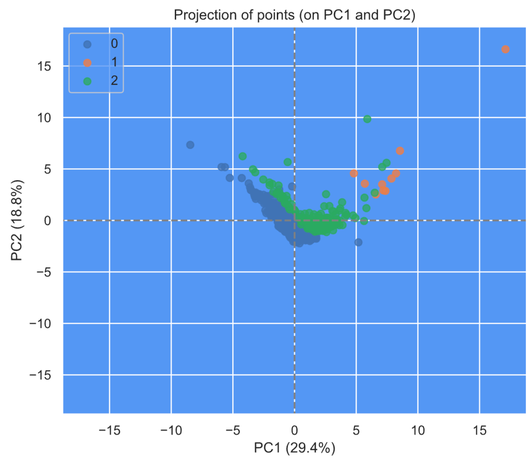
SCAN FOR PRE-PRINT

RESULTS

Three clusters were found with the following characteristics:



PCA shows reasonable explanatory power for the clustering:



DISCUSSION

It will be interesting to do a rudimentary time series analysis to see which census tracts were more resilient in recovery after the pandemic.

The relation between car-ownership and number of pick-ups in each census tract deserves special attention. Also, we may explore how ride-shares dropped in March 2020.