Houston_Travel_Cost
Repository location for our group project.

Andrew Vestal, Elly Atwood, Cameron Medina

Project Outline
Project Title:

Summary:

We will look at data and analysis conducted on our subject to paint a picture of what the current consensus on cost of travel is within our ROI (Houston? Seattle? Los Angeles?) . Find a finite value for cost of travel from specific areas or average cost of travel within a city or state, to which we can compare our prediction model against.

Planned methodology and approach:

Determine an abstract of what our cost of travel will encompass. We will need one feature/value to which we gauge all other features' importance in our prediction model. Our objective is to determine what personal factors influence the travel time of individuals in Houston in order to illustrate what groups primarily shoulder the burden of long commute times and their associated side effects. With this data we hope to provide a greater understanding of what groups are at a disadvantage because of their proximity to their jobs. We will produce a model using potentially relevant variables such as age, sex, or job sector in order to answer our research question.

Expected outcomes:

'To create predictions with our machine learning model that show close results to our related metric we use to determine cost of commute. Though we don't expect perfect results, our goal is to create something that gets close. If our model holds, we can possibly determine areas or individuals that require more infrastructure and better means for travel.'

Potential dataset variables.

https://api.census.gov/data/2019/acs/acs5/variables.html
- Means of transportation by age
- Means of transportation by race (several datasets)
- Means of transportation to work by place of work
- Means of transportation to work by class of worker
- Means of transportation to work by industry
- Means of transportation to work by worker's earnings
- Sex of workers by vehicles available
- Sex of workers by travel time to work
- Sex of workers by means of transportation to work

Tools and packages we need:
- Pandas
- GeoPandas
- NumPy

- SkyLearn
- MatPlotLibrary
- Shapely

Any other relevant information:

References:

Notes:
- Use block groups
- Be aware that hispanic will be black or white, do not do black + white + hispanic

Articles:
https://www.nytimes.com/2015/05/07/upshot/transportation-emerges-as-crucial-to-escaping-poverty.html#:~:text=In%20a%20large%2C%20continuing%20study,there%20moving%20up%20the%20ladder.

https://nationalequityatlas.org/indicators/Commute_time#/

https://blog.arlingtontransportationpartners.com/what-are-your-employees-commutes-costing-you

https://www.metromile.com/blog/the-true-cost-of-commuting/

https://www.lendingtree.com/auto/cost-of-commuting/