

An aerial photograph of a winding, two-lane asphalt road that snakes through a vast, arid desert landscape. The terrain is characterized by deep orange-brown soil and patches of white, possibly salt flats or sand. The road's path is highly irregular, with several sharp turns and loops. On the left side of the image, there is a dark grey rectangular overlay containing white text.

# COURSERA DATA SCIENCE CAPSTONE

FINDING THE IDEAL LOCATION FOR A  
BUSINESS IN A COVID-AFFECTED USA



# FINDING THE IDEAL LOCATION

- Covid has affected business globally and domestically across the world
- Does not make sense to open a business in an area that has been and will continue to be hard hit by Covid
- Task is to find a state that is amenable to business and has dealt with Covid successfully
- In addition, finding a location within the state that meets the business' needs
  - Physical activity and wellbeing venues nearby
  - Located in a busy business district, ideally



# DATA ACQUISITION AND CLEANING

- Covid-19 Data for the US:
  - Historical Data for US - <https://raw.githubusercontent.com/nytimes/covid-19-data/master/us.csv>
  - Historical Data for States - <https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states.csv>
  - Historical Data by County - <https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv>
- Geospatial data for Los Angeles:
  - [https://apps.gis.ucla.edu/geodata/sr\\_Latn/dataset/los-angeles-county-neighborhoods/resource/6cde4e9e-307c-477d-9089-cae9484c8bc1](https://apps.gis.ucla.edu/geodata/sr_Latn/dataset/los-angeles-county-neighborhoods/resource/6cde4e9e-307c-477d-9089-cae9484c8bc1)
- Location and venue data from Foursquare



# DATA ACQUISITION AND CLEANING

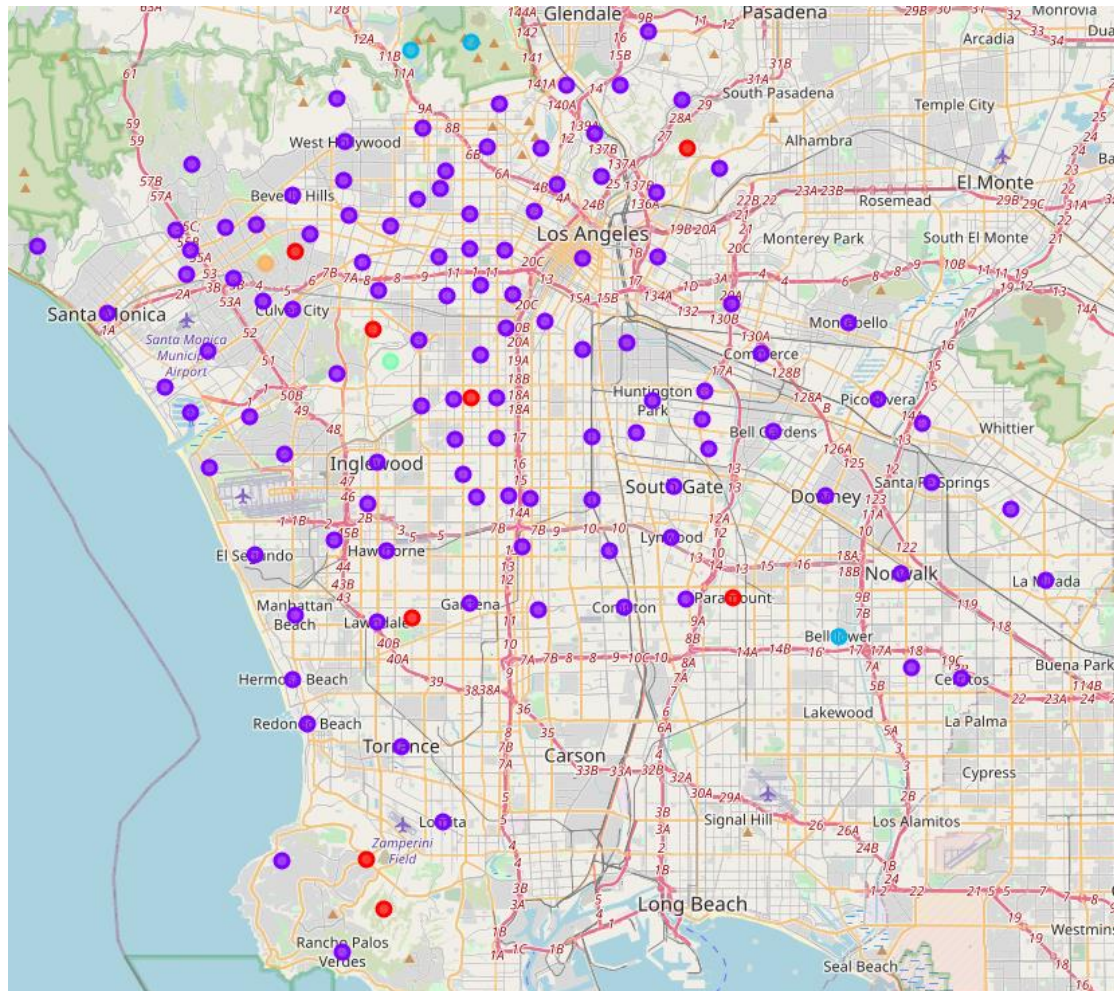
- Census population data for US States extrapolated using average population growth since last census
- State Codes added to dataset for compatibility with Plotly, courtesy of Roger Allen on Github
- Outlying neighbourhoods that are not suitable for big business locations were removed from the dataset
  - E.g. Antelope Valley, Santa Monica Mountains

## IDEAL STATE – STATES WITH THE MOST/LEAST NEW CASES AS A PERCENTAGE OF POPULATION FROM APR-JUN 2021

	state	% new cases
<b>4</b>	CA	0.302988
<b>36</b>	OK	0.364726
<b>3</b>	AR	0.366039
<b>11</b>	HI	0.367477
<b>16</b>	KS	0.397597

	state	% new cases
<b>22</b>	MI	2.397393
<b>23</b>	MN	1.409687
<b>7</b>	DE	1.397857
<b>5</b>	CO	1.383533
<b>38</b>	PA	1.357156

# LOS ANGELES MAP AFTER BEING CLUSTERED



## MOST COMMON VENUES IN CLUSTER 1

	Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Alondra Park, CA	Home Service	Park	Yoga Studio	Farmers Market	English Restaurant
6	Baldwin Hills/Crenshaw, CA	Playground	Park	Yoga Studio	Farmers Market	English Restaurant
14	Beverlywood, CA	Business Service	Park	Yoga Studio	Farm	English Restaurant
47	Harvard Park, CA	Park	Yoga Studio	Farmers Market	English Restaurant	Escape Room
78	Montecito Heights, CA	Food	Park	Yoga Studio	Farmers Market	English Restaurant
83	Paramount, CA	Mexican Restaurant	Park	Business Service	Burger Joint	Flower Shop
92	Rolling Hills Estates, CA	Bank	Business Service	Farm	Park	Yoga Studio
93	Rolling Hills, CA	Business Service	Yoga Studio	Electronics Store	Escape Room	Ethiopian Restaurant



# CONCLUSION

- Neighbourhoods in cluster 1 most suitable for businesses that seek physical activity locations
- Proximity to Yoga Studios, Parks, Playgrounds
- Large number of neighbourhoods clustered in cluster 2 – symptom of problem with model
- Model can be improved by modifying k-clustering algorithm or using another algorithm entirely
- The model does not take into account other factors that may be important, such as:
  - Cost of rent for business space
  - Proximity to competitors that could affect business
  - Ease of access for parking, or public transport options