



# Build Better Models with Location Data

**Anirudh Shah**, Research Engineer  
[anirudh@askiggy.com](mailto:anirudh@askiggy.com)

[www.askiggy.com](http://www.askiggy.com)

# Agenda

1. Use Case
2. Geospatial Data Overview
3. Geospatial Boundaries
4. Demo
5. Feature Insights
6. Graph ML on Geospatial Data



# Use Case

- We are building a vacation rental recommender system
- We have user data on previous rentals
- We recommend rentals for a user's upcoming trip
- User's profile:
  - Mid 30s
  - Travelling with a group of friends
- We need geospatial data to build the system





# Geospatial Data Overview

# Geospatial Data at Iggy

- ‘What is nearby’
- Types of data included:
  - Points of interest (POIs)
  - Water bodies
  - Socio-economic data
  - And much more!

Data for recommender system:

- Nearby restaurants
- Nearby parks and recreation areas
- Nearby rivers and lakes
- Other nearby POIs
- Median age



# What datasets can be enriched with geospatial data?

- any data with a location column e.g. address, zipcode, city, latitude/longitude, geometry
- Examples in real estate, travel, marketing, utilities, public health and many more



station_id	name	status	address	council_district	modified_date
2576	Rainey @ River St	closed	64 Rainey St	9	2021-01-04T12:00:00Z
2712	Toomey Rd @ South Lamar	closed	1301 Toomey Road	5	2021-01-04T12:00:00Z
3381	East 7th & Pleasant Valley	closed	2772 E 7th	3	2021-01-04T12:00:00Z
3464	Pease Park	closed	1155 Kingsbury St	9	2021-01-04T12:00:00Z
1001	OFFICE/Main/Shop/Repair	active	1000 Brazos	1	2021-01-04T12:00:00Z
1002	6th & Navasota St.	closed	1308 W. 6th St.	3	2021-01-04T12:00:00Z
1003	8th & Guadalupe	closed	800 Guadalupe St.	9	2021-01-04T12:00:00Z

geo_id	zipcode	description	year	area_deprivation_index_percent
86000US01001	01001	ZCTA5 01001	2018	27.000
86000US01002	01002	ZCTA5 01002	2018	28.000
86000US01003	01003	ZCTA5 01003	2018	92.000
86000US01005	01005	ZCTA5 01005	2018	14.000
86000US01007	01007	ZCTA5 01007	2018	17.000
86000US01008	01008	ZCTA5 01008	2018	14.000
86000US01009	01009	ZCTA5 01009	2018	69.000





id	latitude	longitude	property_type	room_type	accommodates	bathrooms_text	bedrooms	beds	price	minimum_nights	maximum_nights
8214182	30.228	-97.725	Private room in rental unit	Private room	1	1 bath	1.000	1.000	27	1	5
4356661	30.255	-97.745	Entire rental unit	Entire home/apt	2	1 bath	1.000	1.000	75	1	1125
15035937	30.209	-97.896	Entire guesthouse	Entire home/apt	4	1 bath	1.000	3.000	150	3	1125
51170982	30.332	-97.725	Entire serviced apartment	Entire home/apt	4	1 bath	1.000	4.000	175	2	365
32207144	30.271	-97.749	Entire condominium (condo)	Entire home/apt	6	2 baths	2.000	3.000	1123	2	182
32207608	30.271	-97.746	Entire residential home	Entire home/apt	4	1 bath	1.000	2.000	685	2	182

# Why is geospatial data difficult to use in ML?



Sourcing



Cleaning



Computation



Engineering &  
Delivery





# Geospatial Boundaries

# Boundaries Overview

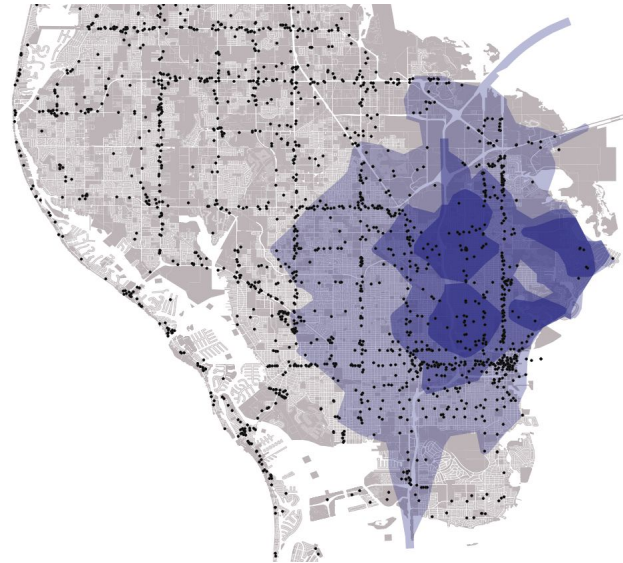
Geographic area over which data is aggregated

Administrative boundaries:

- Metro
- Zipcode
- Census Block Group (CBG)

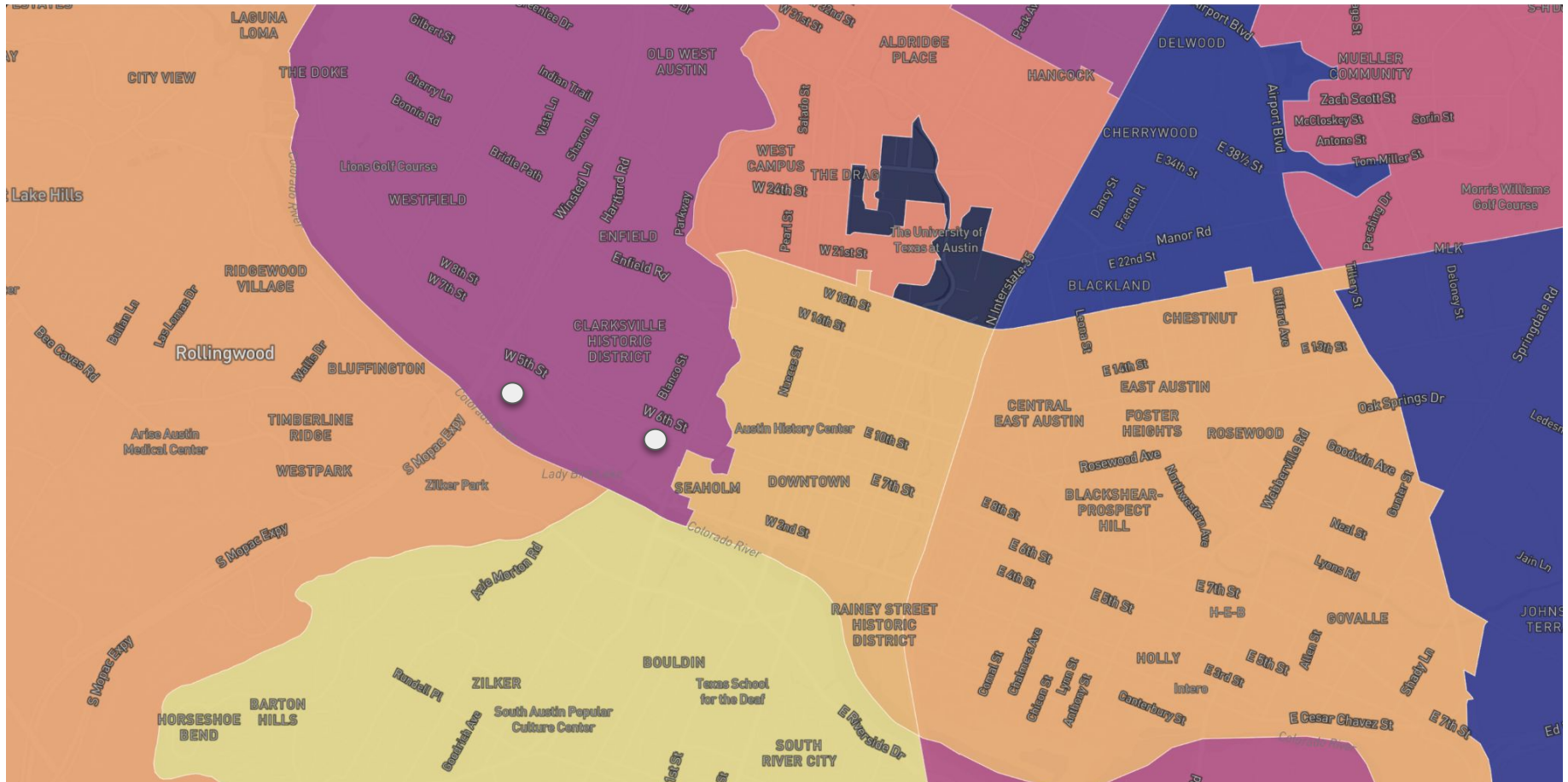
Derived boundaries:

- Isochrone 10 minute walk

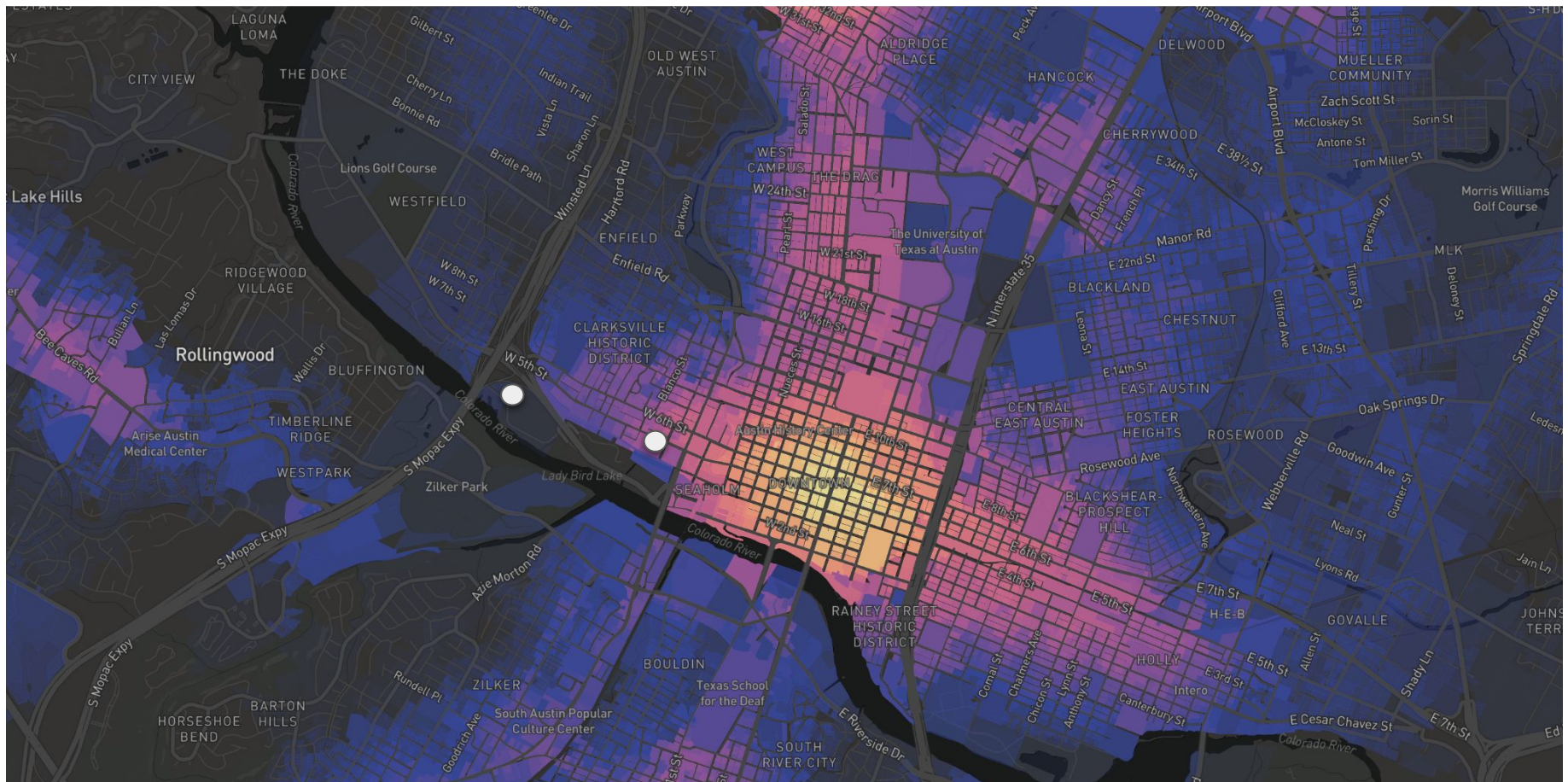


**'Asobranos flies'**









10-minute walk Isochrone boundaries





Demo







# Feature Insights

# Real Estate Insights

## Use Case:

- Predicting sales price
- Combination of property metadata and geospatial data
- **2% improvement in test set performance with geospatial features**

## Key geospatial features:

- Whether a coast intersects
- Number of education centers
- Number of gyms
- Number of manufacturing facilities

## Key property features:

- Square footage
- Floor finish
- Interior finish



# Vacation Rental Rating Prediction

## Use case:

- Management company that buys and rents vacation homes
- Buying a new home
- Predict customer's rating of a new home's location
- We need location data for this

## Key geospatial features:

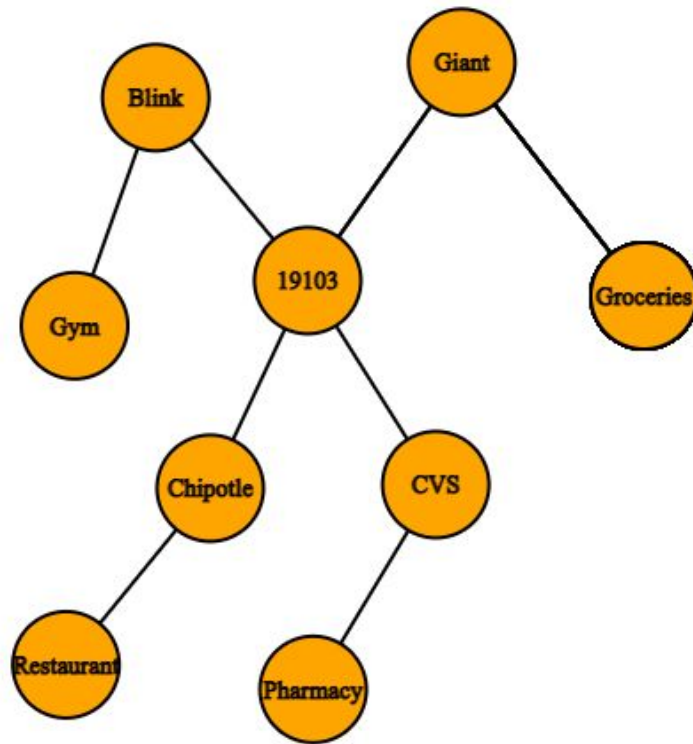
- Number of public parks per sqkm
- Number of gyms per sqkm/capita
- Number of POIs per capita
- Number of bars per sqkm



# Graph ML on Geospatial Data

# Graph ML on Geospatial Data

- Different ways to represent the data graphically
- How we are using graphs:
  - Quality control
  - Representation learning
  - Incorporate other data types e.g. satellite imagery and text



Thank you!



[www.askiggy.com](http://www.askiggy.com)