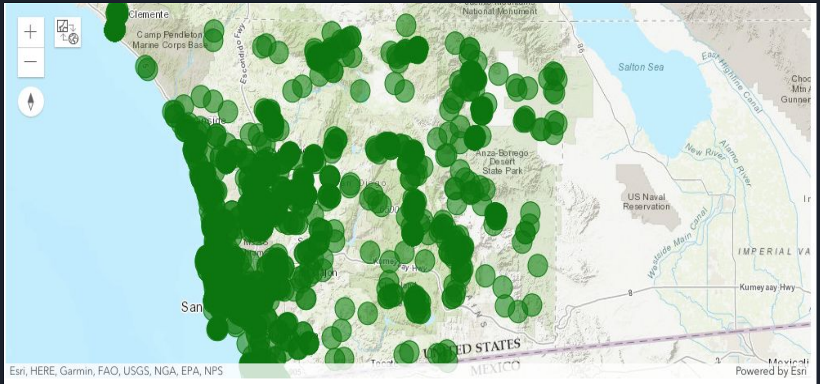
Analyzing Factors of Airbnb Pricing in San Diego

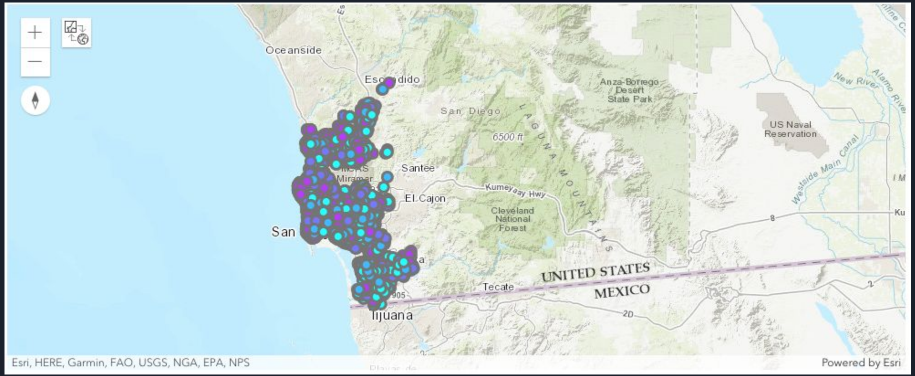
By Benson Duong, Tyson Tran

Question: What spatial and non-spatial factors are most likely to influence the listings price of an Airbnb?

Data

- OpenStreetMap - Tourist Attractions
  - Tourist Attraction Point data for North America
  - Updated every 5 mins
  - Places and things of specific interest to tourist
- Inside Airbnb
  - Scraped data of all Airbnb listings in San Diego on December 24, 2022



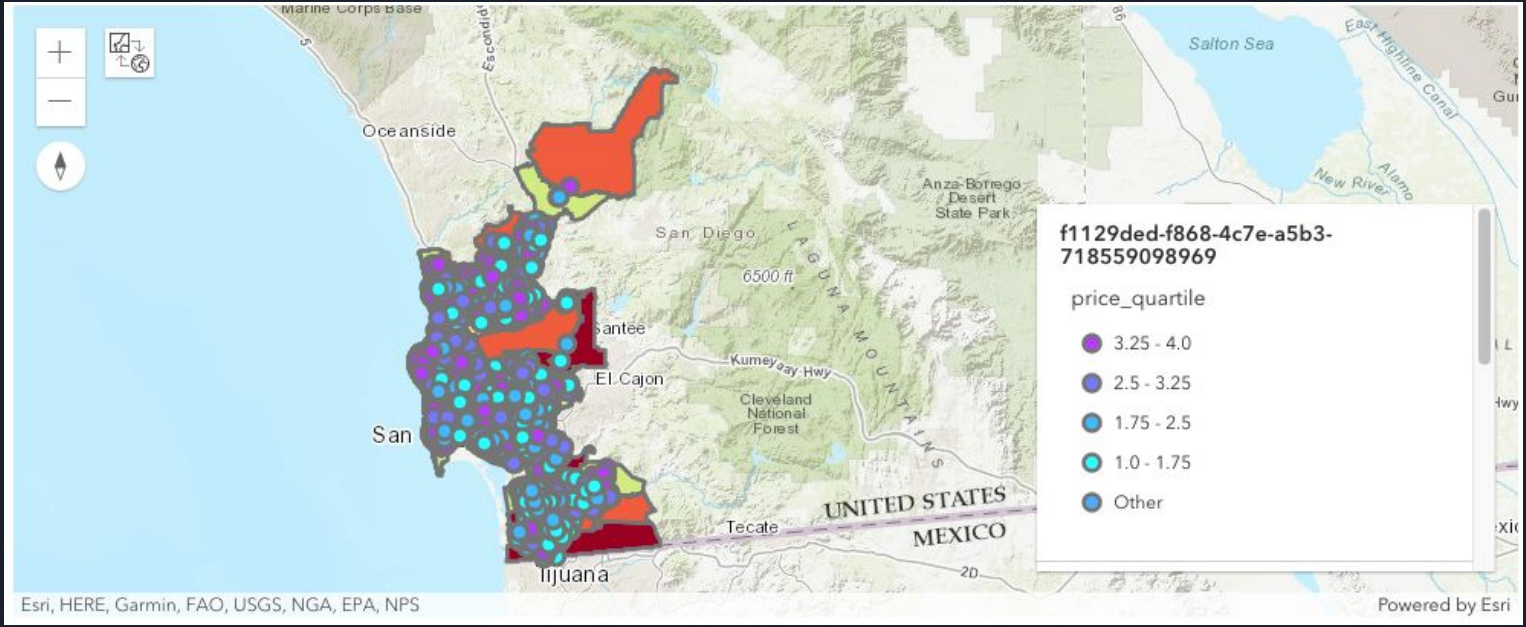


Geo-enrichment

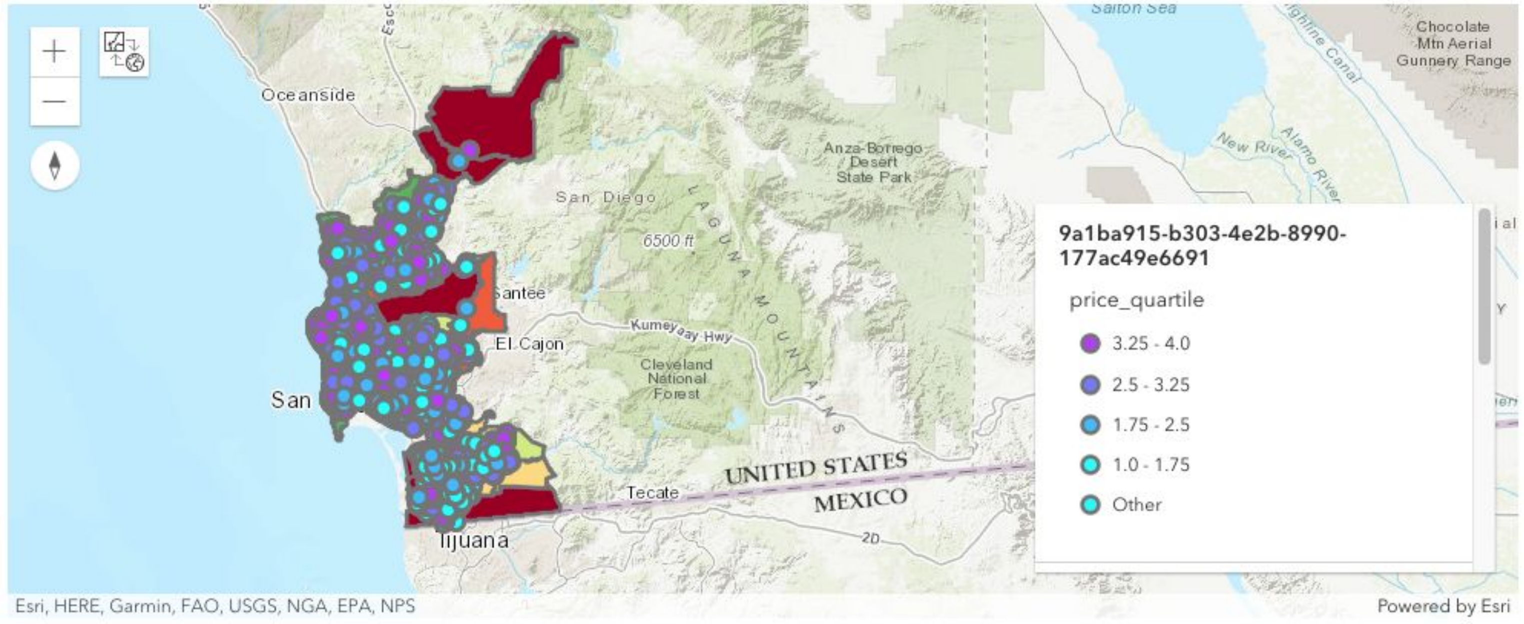
- Geo-enrichment to zipcodes
  - Population, and Population Density
  - Businesses:
    - Shopping, Recreation
    - Hospitality, Accomodation
  - Crime
  - Economic Data:
    - Median Household Income (And Projected Growth)
    - Median Home Values (And Projected Growth)
  - Households with Children

Part 1: Spatial Factors

Movie/Amusement Businesses vs Airbnb Listing Price

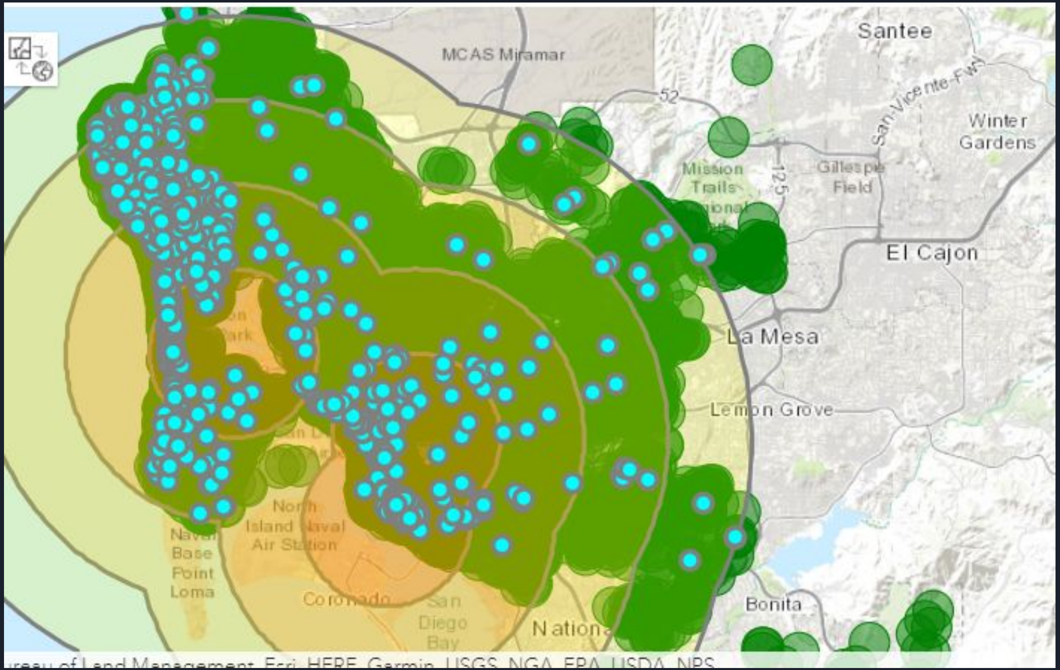


2022 Home Median Value vs Airbnb Listing Price



Part 2: Tourist Attraction Concentric Buffers

Buffers around Landmarks and Airbnb price



Concentric Buffers between specific San Diego Landmarks, specifically:

- SeaWorld
- Downtown and the Convention Center
- Balboa Park and Zoo

The Top 1000 Most Expensive Airbnbs in cyan (Others in green)

Part 3: Non Spatial Factors

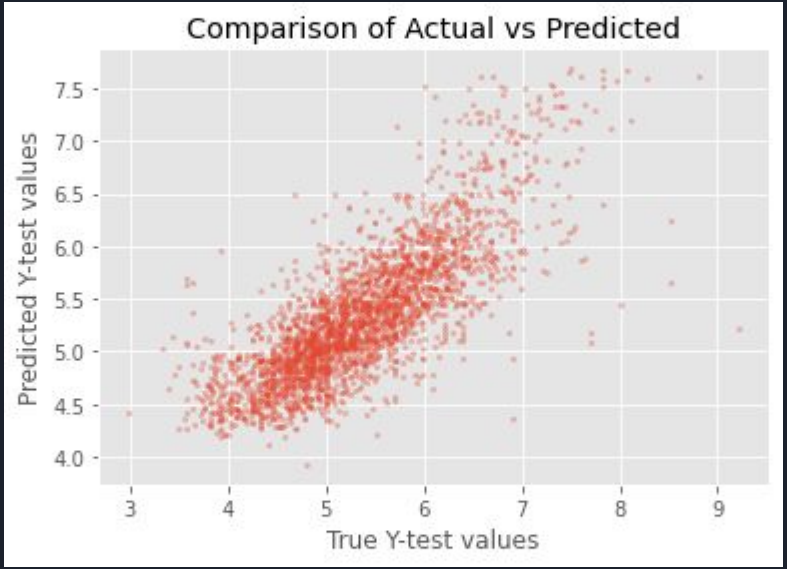
Non Spatial Factors

Non Spatial Factors (continued)

We can see that there is no correlation between the experience that a host has and the price that a property is listed at

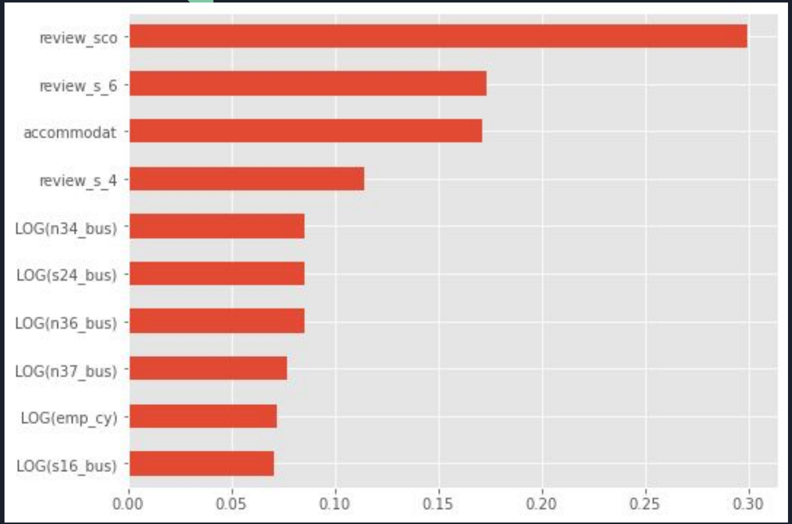
Part 4: Price Prediction Model

Linear Regression



- A prediction model for regression was done that predicts an airbnb's price (log-transformed)
  - given its airbnb data and geo-enriched data about the zipcode it resides in
- A very promising score at a corrcoef of 79%

Feature Importances



* Most Influential Features in predicting an Airbnb's price

*__review_sco__ - the review score rating

* review_s_6 - the review_scores_value

* **Accomodat** - the amount of people it can accomodate

* **review_s_4** - the review_scores_communication

* **Log(n34_bus)** - # of Arts/Entertainment/Rec Businesses in the zipcode

* **Log(n36_bus)** - # of Accommodation Businesses in the zipcode

* **Log(S24_BUS)** - # of Hotel/Lodging Businesses in the zipcode

* **Log(emp_cy)** - the employed population in the zipcode

* **Log(S16_BUS)**- # of Eating & Drinking Businesses in the zipcode

Closing Thoughts

- **Modifiable Areal Unit**: We have to note the potential biases in using zipcodes to geo-enrich our data, since zipcodes are unequally shaped and sized making problems like modifiable areal unit problem.
- Other biases: Spatial Autocorrelation, Population Bias, spurious correlations
- If further time and resources were available, this project can be expanded to see if Airbnb influences on housing availability or mark gentrification,
  - with data sources like **percentage of household families living in apartments** -
  - and their year-to-year income growth rate with respect to apartment rent increase rates,
  - and yearly numerical increases of airbnbs in a zipcode.
- We'd like to thank the instructors for their help in this project by providing a tourist attraction data source and overall teaching us these GIS tools this quarter.

Q & A