Predicting Airbnb Prices in Florence, Italy

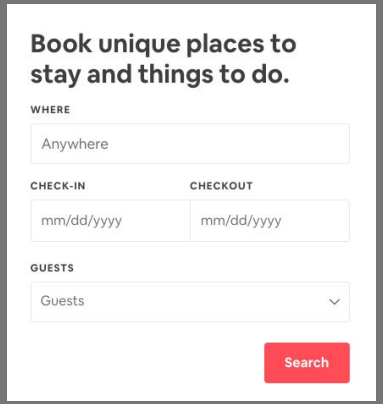
A Group 5 Production
Jyoti Sharma, Tanvi Hindwan,
& Jessa Henderson

December 4, 2019
DATS6103: Data Mining

About Airbnb

- Airbnb is the most demanding online marketplace where people seek and offer lodging as homestay.
  - Over 7 million accommodations listed
  - Over half a billion guest check-ins
  - In 191 countries & regions

- Renters list a variety of criteria to attract renters and increase their prices.
  - Neighborhood
  - Offered amenities like wifi or hair dryers
  - Number of bedrooms/bathrooms
  - Pictures



Research Question

What features predict airbnb prices in Florence, Italy?

Dataset Details

- Original Dataset
  - 58MB
  - 97 variables
    - Pictures
    - Reviews
    - Location information
    - Price information
    - Amenities offered
    - Counts of bedrooms & bathrooms
    - Among others
  - Total Observations in dataset - 12304
  - Dataset URL: http://insideairbnb.com/get-the-data.html

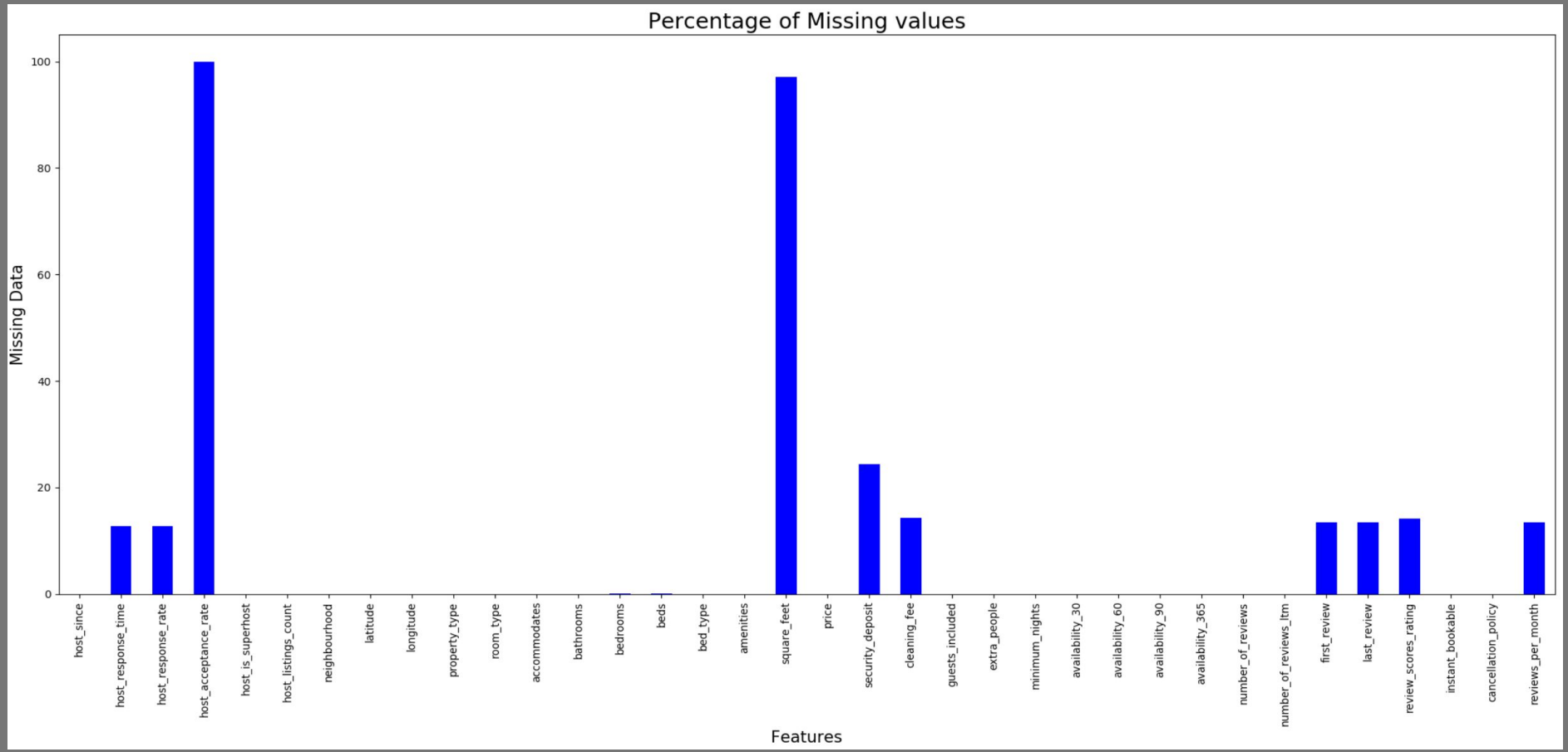
Data Preprocessing

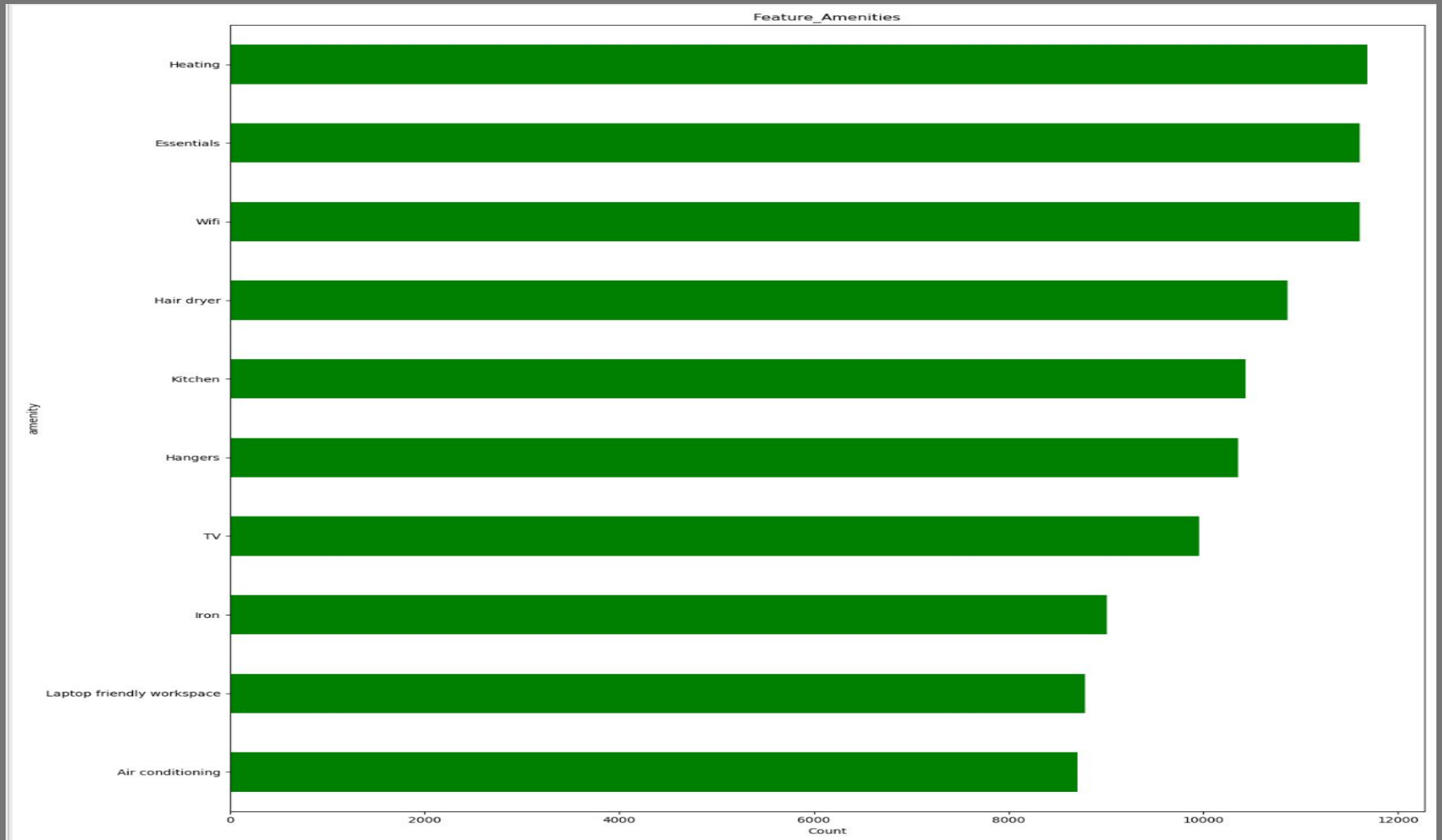
Missing Values Table

Table 1
Missing values and Data Type of Variables in Dataset

Target	NaNs	Dtype	Features	NaNs	DType
price	0	object	host_since	5	datetime64[ns]
			host_response_time	1566	object
			host_response_rate	1566	object
			host_acceptance_rate	12304	float64
			host_is_superhost	5	object
			host_listings_count	5	float64
			neighbourhood	0	object
			latitude	0	float64
			longitude	0	float64
			property_type	0	object
			room_type	0	object
			accommodates	0	int64
			bathrooms	5	float64
			bedrooms	7	float64
			beds	23	float64
			bed_type	0	object
			amenities	0	object
			square_feet	11940	float64
			security_deposit	3010	object
			cleaning_fee	1766	object
			guests_included	0	int64
			extra_people	0	object
			minimum_nights	0	int64
			availability_30	0	int64
			availability_60	0	int64
			availability_90	0	int64
			availability_365	0	int64
			number_of_reviews	0	int64
			number_of_reviews_ltm	0	int64
			first_review	1663	datetime64[ns]
			last_review	1663	datetime64[ns]
			review_scores_rating	1736	float64
			instant_bookable	0	object
			cancellation_policy	0	object

Missing Values




Feature_Amenities
Heating
Essentials
Wifi
Hair dryer
Kitchen
Hangers
TV
Iron
Laptop friendly workspace
Air conditioning
amenity
0
2000
4000
6000
8000
10000
12000
Count

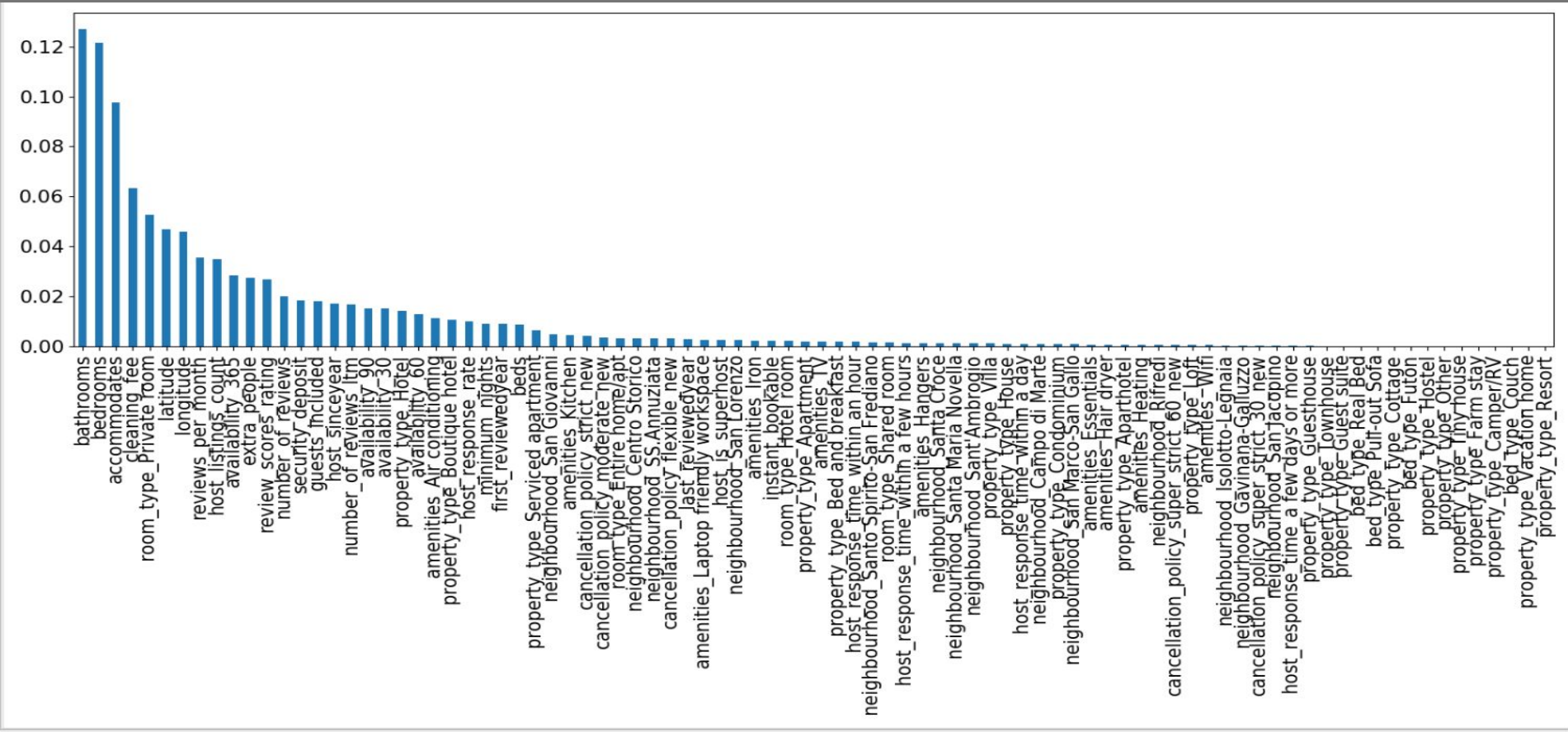
Experimental Setup

- Feature Selection was conducted using Random Forest
  - Results will be shown during the demo

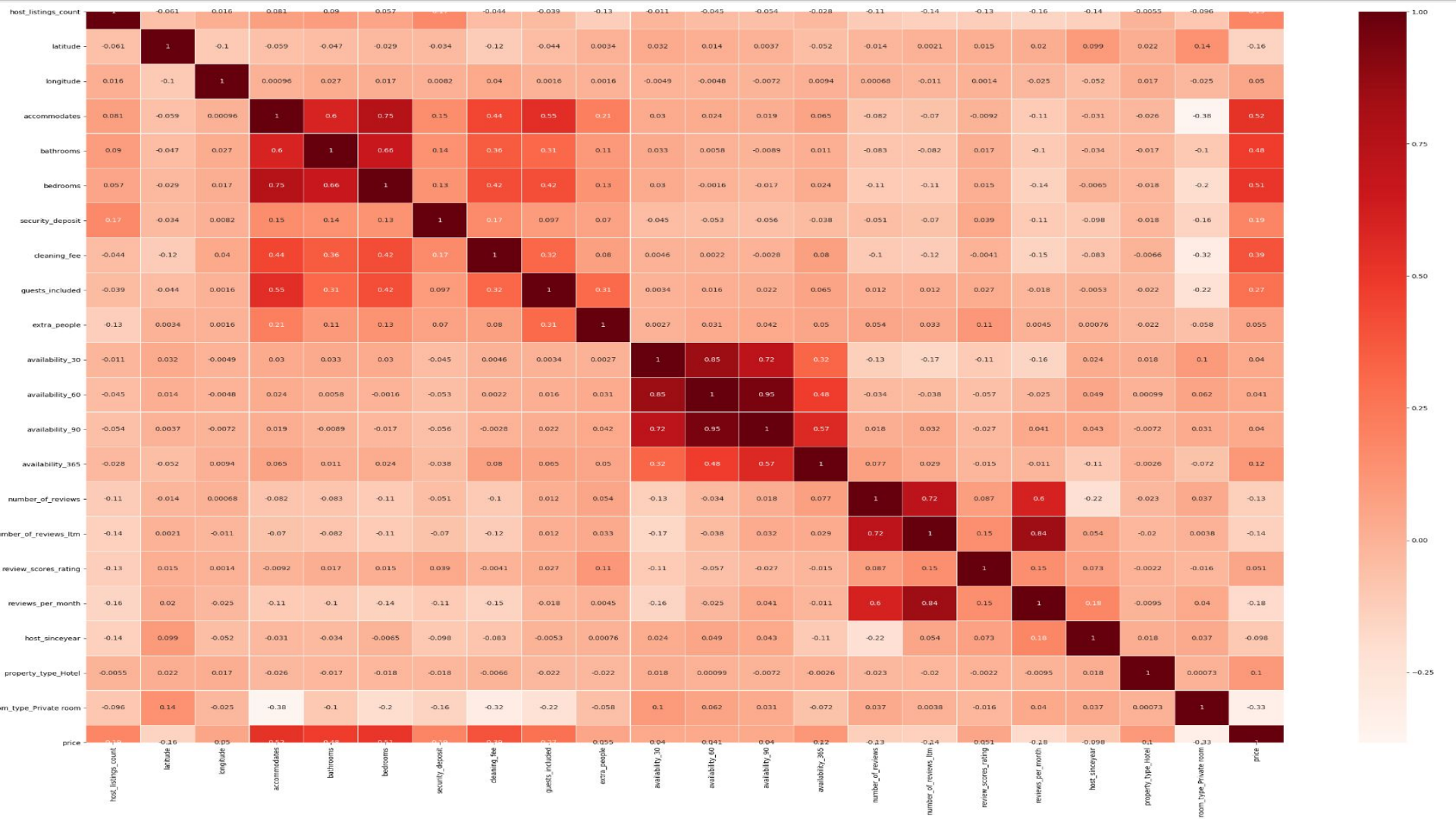
- Once features were paired down, the research question was answered using linear regression.
  - Results will be shown during the demo

Feature Selection

Overall Features



Linear Regression


host_listings_count
latitude
longitude
accommodates
bathrooms
bedrooms
security_deposit
cleaning_fee
guests_included
extra_people
availability_30
availability_60
availability_90
availability_365
number_of_reviews
number_of_reviews_ltm
review_scores_rating
reviews_per_month
host_sinceyear
property_type_Hotel
room_type_Private room
price
1.00
0.75
0.50
0.25
0.00
−0.25

Results

- We'll now use the GUI to demonstrate our findings from our dataset

Limitations

- Missing values

- User self-reported data - may contain social desirability bias