



Airbnb Seattle Analysis

Team 2 :

- Ankita Mittal
- Avinav Dixit
- Kumar Saket
- Shweta Malla
- Umang Sharma

INDEX

- Kaggle Competition Overview
- Exploratory Analysis
- Predictive Analysis
- Observations
- Recommendations
- COVID - 19



Kaggle Overview



VARIABLES

Availability_365, bathrooms, bed_type, bedrooms, host_identity_verified, host_is_superhost, host_listings_count, price, cleaning_fee, minimum_nights, amenities_count, host_age, host_response_rate, instant_bookable, is_location_exact, property_type, requires_license



PERFORMANCE AUC

89.25%



FINAL METHOD USED

Random Forest

mtry=3,
ntree=1000,
importance=TRUE,
na.action=na.roughfix



DATA PROCESSING

Replaced missing values,
removed extreme outliers,
created more variables to
improve model accuracy

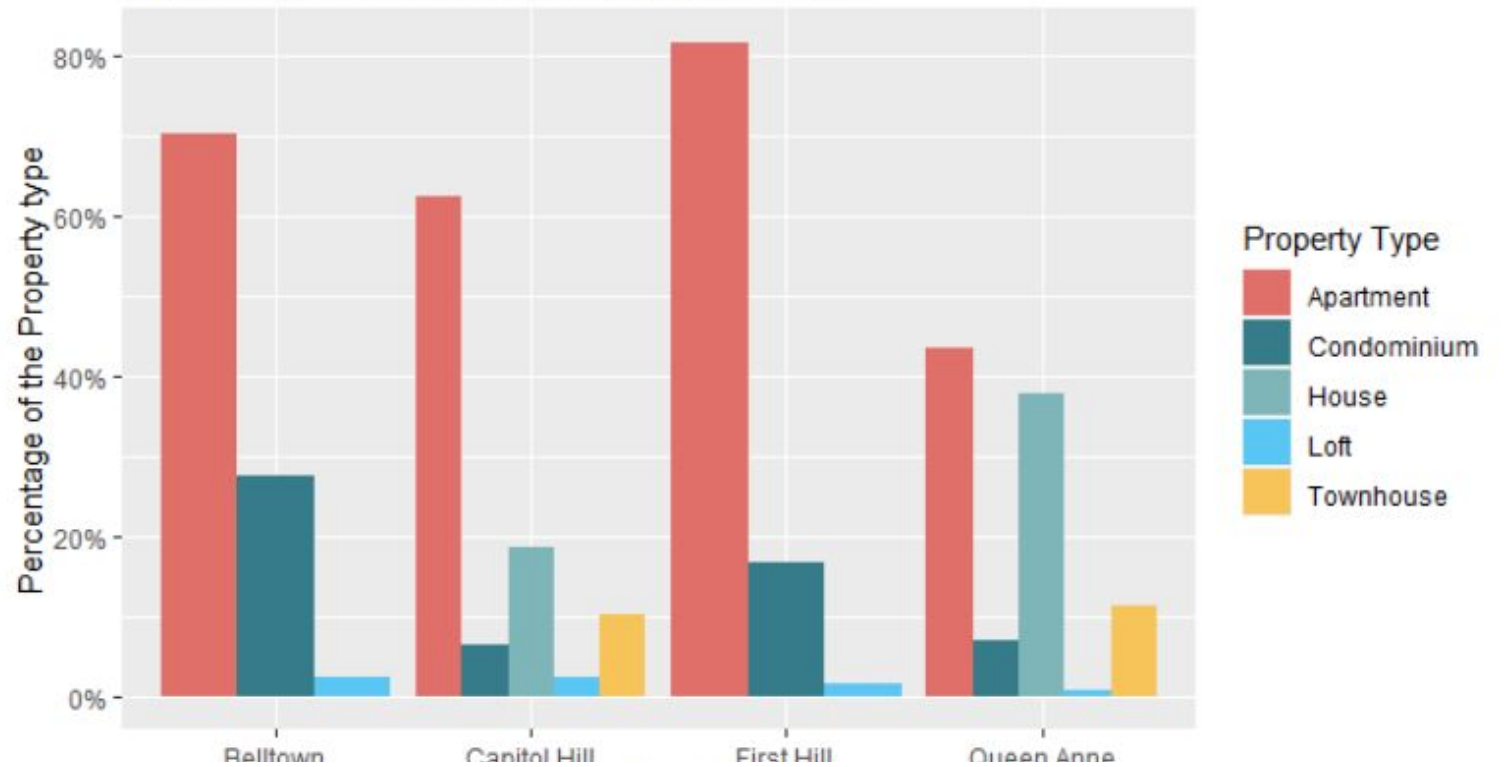


Exploratory Data Analysis

Types of Listings in Seattle

Which property type is listed the most in Seattle?

Property Types in Top Neighborhoods in Seattle



Popular Property & Host Characteristics

Host has profile pic -> TRUE

Bed Type -> "Real Bed"

Property Type -> "Apartment"

Host Identity Verified -> TRUE

Host Is Superhost -> FALSE

Host response time -> "Within hour"

Mean Monthly Price -> \$2300-\$2600

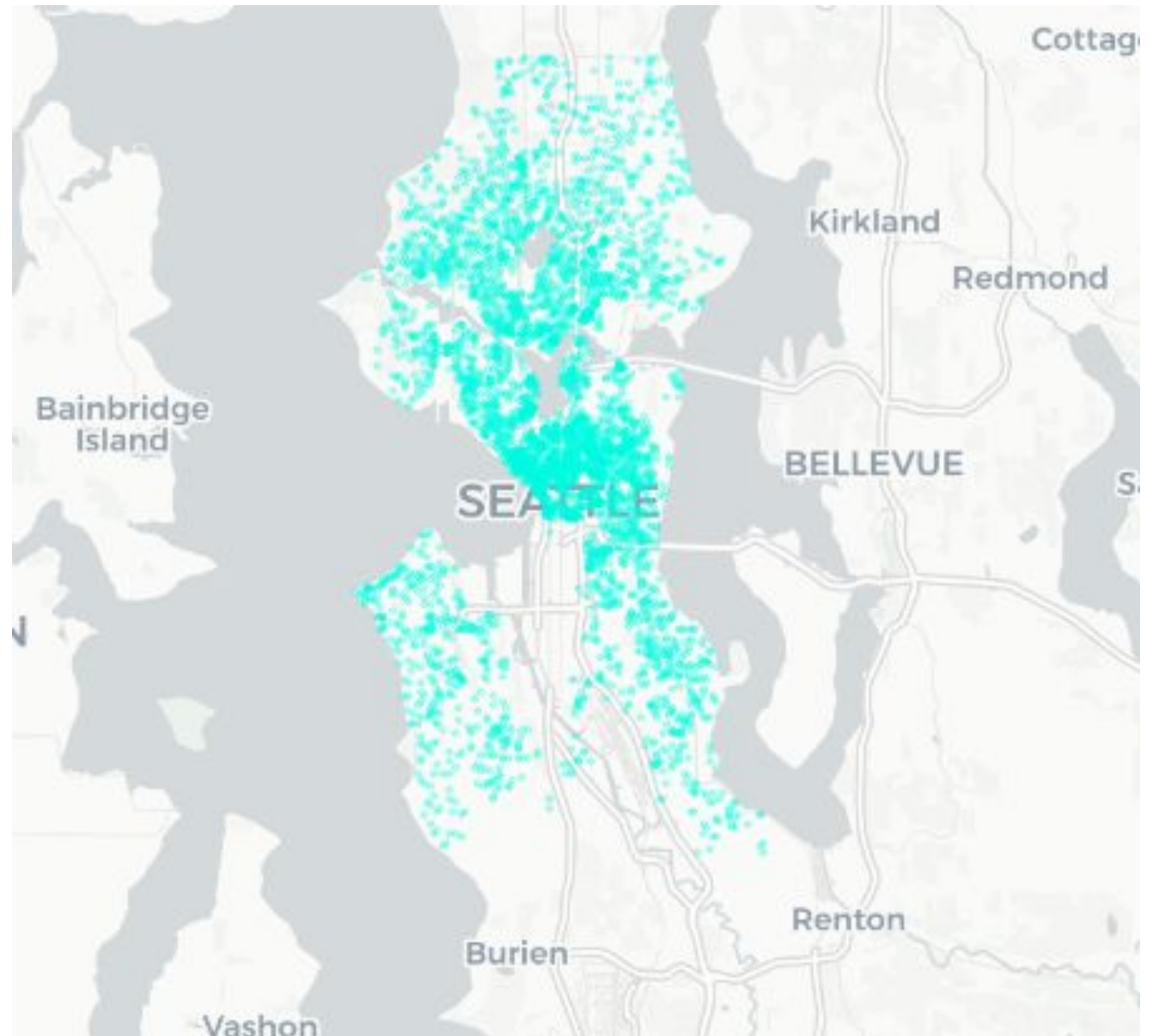
Mean Availability within 30 Days -> 10-18 Days

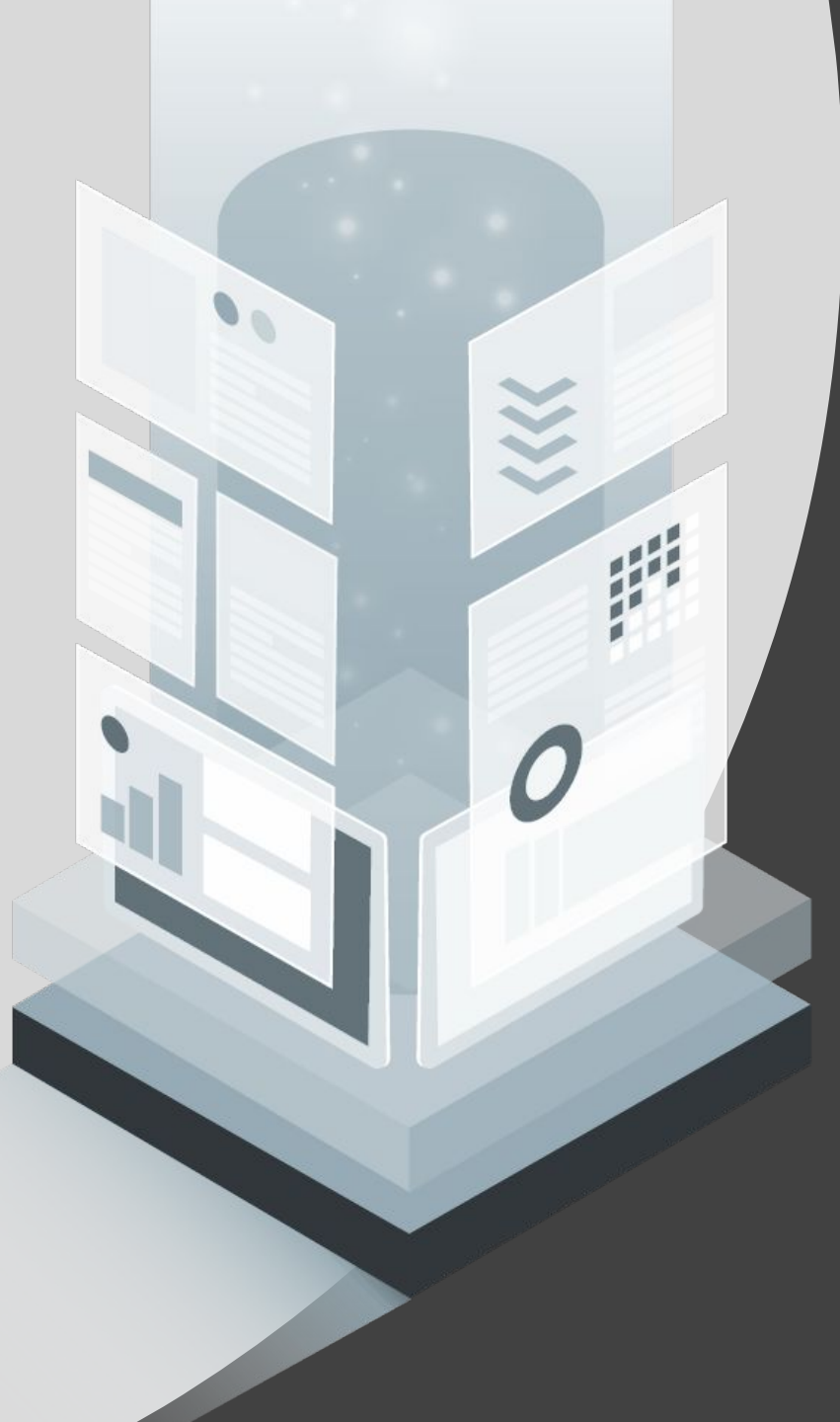
Mean Availability within 60 Days -> 24-42 Days

Mean Minimum Nights -> 2-12 Nights

Spread of Airbnb Listings

- The listings are almost evenly spread across Seattle, with higher concentration in Central Seattle.





Predictive Analysis


```
Call:
glm(formula = high_booking_rate ~ availability_365 + bathrooms +
    bed_type + bedrooms + bed_type + price + host_identity_verified +
    host_is_superhost + host_listings_count + maximum_nights +
    minimum_nights + amenities_count + host_verification_count +
    host_age, family = "binomial", data = airbnbSeattleTrainSplit)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7885  -0.7478  -0.3210   0.7858   3.6397
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -4.391e+00  1.182e+00  -3.716 0.000203 ***
availability_365  1.727e-03  3.019e-04   5.721 1.06e-08 ***
bathrooms      -2.082e-01  7.443e-02  -2.797 0.005161 **
bed_typeCouch   -8.229e+00  2.286e+02  -0.036 0.971283
bed_typeFuton    2.043e+00  1.262e+00   1.619 0.105481
bed_typePull-out Sofa  3.050e+00  1.366e+00   2.232 0.025613 *
bed_typeReal Bed  2.335e+00  1.164e+00   2.006 0.044908 *
bedrooms       -1.482e-01  5.320e-02  -2.786 0.005336 **
price          -2.418e-03  4.451e-04  -5.433 5.55e-08 ***
host_identity_verifiedTRUE  2.978e-01  8.469e-02   3.516 0.000438 ***
host_is_superhostTRUE    1.387e+00  7.998e-02  17.346 < 2e-16 ***
host_listings_count    -2.123e-03  2.621e-04  -8.100 5.50e-16 ***
maximum_nights        -2.931e-04  7.309e-05  -4.011 6.05e-05 ***
minimum_nights        -5.280e-02  6.566e-03  -8.041 8.92e-16 ***
amenities_count        5.339e-02  3.976e-03  13.427 < 2e-16 ***
host_verification_count -9.520e-03  1.843e-02  -0.517 0.605381
host_age             1.853e-03  3.735e-04   4.961 7.02e-07 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 5815.4 on 4326 degrees of freedom
Residual deviance: 4187.0 on 4310 degrees of freedom
(4 observations deleted due to missingness)
AIC: 4221
```

```
Number of Fisher Scoring iterations: 11
```

Logistic Regression for High Booking Rate

- Higher booking rate was closely related to availability, number of bedrooms & bathrooms and host details & price.
- Minimum nights & Maximum nights played a good part in determining whether a person would book an airbnb or not.

Linear Regression on Price

- Used variables that were statistically significant and added our domain knowledge to run the models and compared RMSE.

Call:

```
lm(formula = price ~ availability_365 + bathrooms + bed_type +  
  bedrooms + bed_type + host_identity_verified + host_is_superhost +  
  host_listings_count + maximum_nights + minimum_nights + amenities_count +  
  host_verification_count + host_age + extra_people + guests_included +  
  host_response_time + instant_bookable + beds + cancellation_policy +  
  maximum_nights + room_type + security_deposit, data = airbnbSeattleTrain)
```

Residuals:

Min	1Q	Median	3Q	Max
-444.0	-55.2	-11.5	34.4	9752.9

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.231e+01	7.418e+01	1.244	0.213389
availability_365	2.130e-01	2.285e-02	9.322	< 2e-16 ***
bathrooms	2.198e+01	5.746e+00	3.825	0.000132 ***
bed_typeCouch	5.937e+01	1.158e+02	0.513	0.608289
bed_typeFuton	-1.036e+01	7.273e+01	-0.142	0.886728
bed_typePull-out Sofa	-1.154e+01	7.670e+01	-0.150	0.880406
bed_typeReal Bed	-3.562e+00	6.242e+01	-0.057	0.954487
bedrooms	2.829e+01	4.428e+00	6.389	1.81e-10 ***
host_identity_verifiedTRUE	-2.178e+01	6.340e+00	-3.436	0.000596 ***
host_is_superhostTRUE	1.991e-02	6.732e+00	0.003	0.997640
host_listings_count	2.933e-01	9.800e-03	29.928	< 2e-16 ***
maximum_nights	-1.172e-02	5.310e-03	-2.207	0.027377 *
minimum_nights	-3.938e-01	2.150e-01	-1.832	0.067000 .
amenities_count	1.336e-01	2.853e-01	0.468	0.639637
host_verification_count	-7.652e+00	1.367e+00	-5.598	2.28e-08 ***
host_age	-1.091e-02	2.793e-02	-0.391	0.696063
extra_people	-8.630e-01	1.464e-01	-5.895	3.98e-09 ***
guests_included	9.712e+00	2.207e+00	4.400	1.11e-05 ***
host_response_timeN/A	3.234e+01	4.323e+01	0.748	0.454407
host_response_timewithin a day	-1.720e+01	4.533e+01	-0.380	0.704304
host_response_timewithin a few hours	-2.527e+01	4.393e+01	-0.575	0.565131
host_response_timewithin an hour	-3.106e+01	4.314e+01	-0.720	0.471614
instant_bookableTRUE	3.602e+00	6.132e+00	0.587	0.556986
beds	1.278e+01	2.836e+00	4.506	6.75e-06 ***
cancellation_policymoderate	-3.017e-01	7.656e+00	-0.039	0.968566
cancellation_policystrict	-5.286e+02	2.643e+01	-20.003	< 2e-16 ***
cancellation_policystrict_14_with_grace_period	3.704e+01	8.014e+00	4.622	3.90e-06 ***
cancellation_policysuper_strict_30	-1.567e+02	3.246e+01	-4.829	1.41e-06 ***
cancellation_policysuper_strict_60	-6.267e+01	4.265e+01	-1.469	0.141786
room_typeHotel room	1.341e+02	2.552e+01	5.253	1.55e-07 ***
room_typePrivate room	-5.267e+01	7.227e+00	-7.288	3.60e-13 ***
room_typeShared room	-1.425e+02	2.104e+01	-6.771	1.41e-11 ***
security_deposit	2.646e-03	8.195e-03	0.323	0.746837

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 195.1 on 5375 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 0.3167, Adjusted R-squared: 0.3127

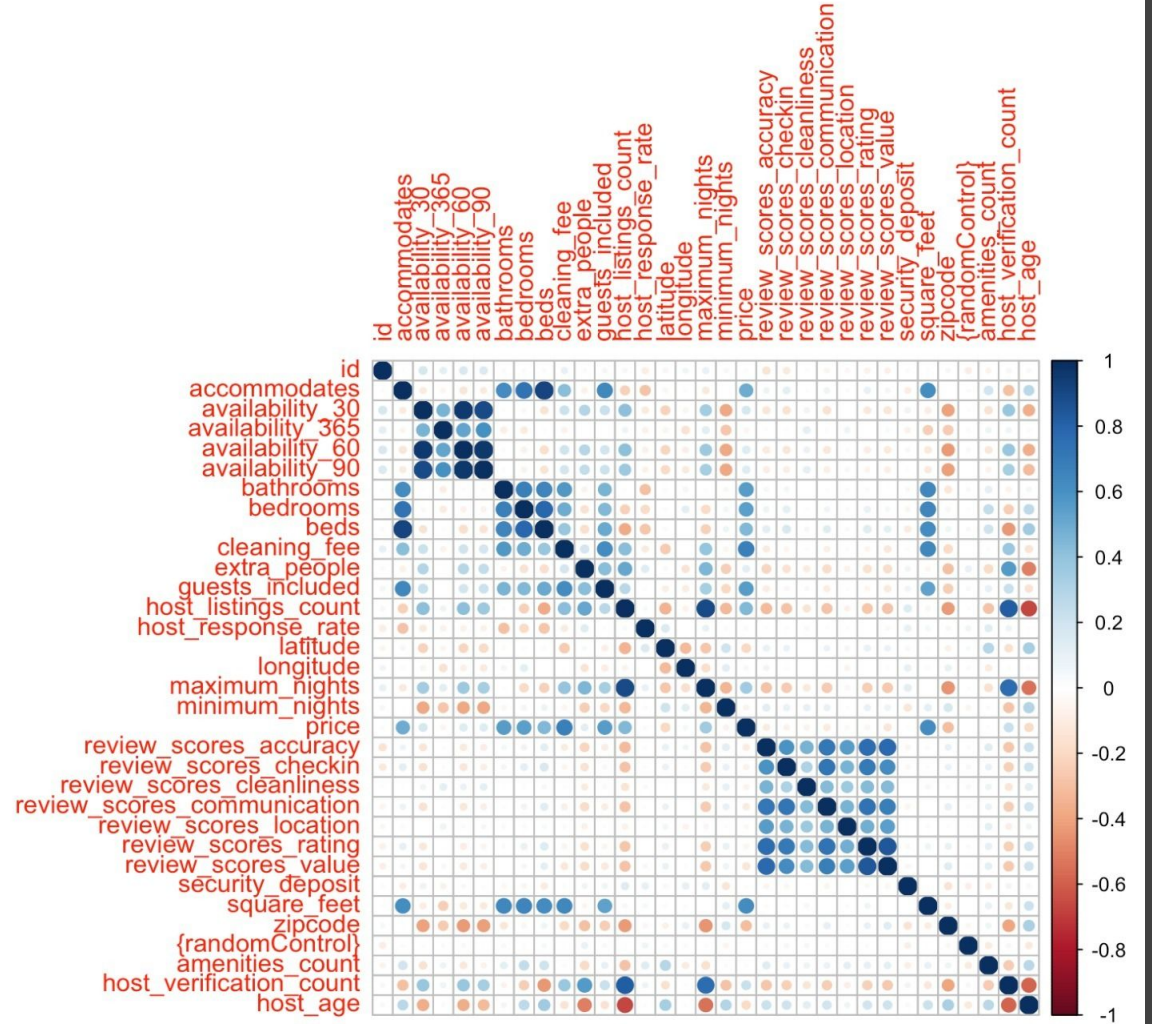
F-statistic: 77.86 on 32 and 5375 DF, p-value: < 2.2e-16

Correlation Plot for price

Important variables:

availability_365, bathrooms, bed_type, bedrooms ,
host_identity_verified, host_is_superhost,
host_listings_count, maximum_nights, minimum_nights,
amenities_count, host_verification_count, extra_people,
guests_included, host_response_time, instant_bookable,
beds

Using our domain knowledge we eliminated the less
significant variables.



Random Forest & kNN for Price

```
```\r}  
RMSE(resultsKNN$predictedPrice, resultsKNN$price)
RMSE(resultsRfHbr$predictedPrice, resultsRfHbr$price)
```\r}
```

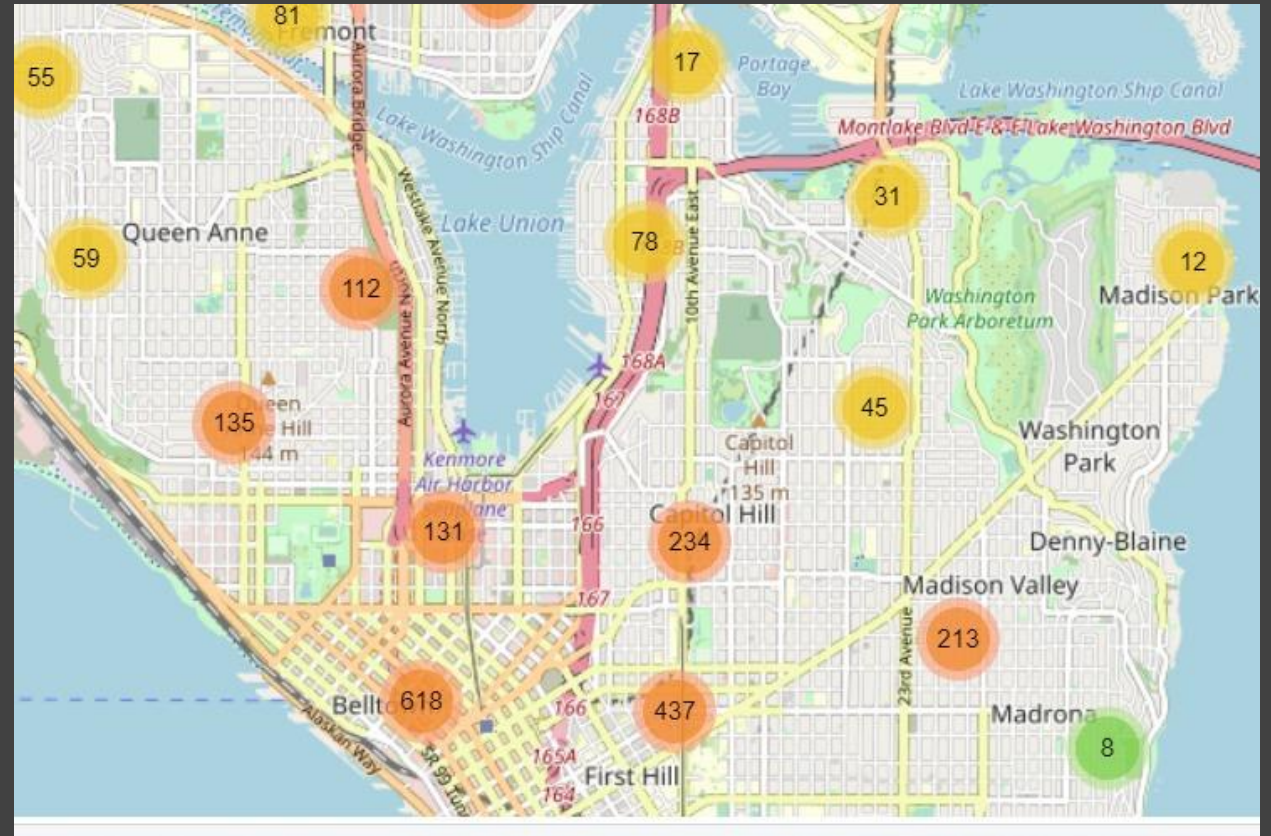
```
[1] 164.8818  
[1] 158.9436
```



Observations

Top 5 listed neighborhoods in Seattle

- Belltown
- First Hill
- Capitol Hill
- Queen Anne
- Madison Valley



Interesting Observations

- According to the given dataset 13% of the most Expensive rentals(monthly_price>5000) are located in Capitol hill.
- Capitol hill also tops the table for monthly prices below 100\$ with (~8% of the total 4975 listings).
- The maximum rent for a property listed is \$36000 in the Ballard neighbourhood.
- In Capitol Hill which has a variety of property types to be rented ~65 %of the listed properties are Apartments, with a contribution of ~29 % to high_booking_rate, when condos only contribute ~4%.

neighbourhood <chr>	pct <dbl>
Capitol Hill	13.043478
Bryant	8.695652
Magnolia	8.695652
Queen Anne	8.695652
Alki	4.347826
Ballard	4.347826

neighbourhood <chr>	pct <dbl>
Capitol Hill	7.69849246
Belltown	6.25125628
Queen Anne	6.21105528
Minor	5.18592965
Ballard	4.84422111

Interesting Observations

- Increased Apartment construction due to rising rental demands since 2010.
- Maximum hosts joined Airbnb (~ **19%**) in the months of December and November in the year **2015**.
- Approximately, according to HMA (Housing Market Analysis), 40% of units constructed were in the Downtown | Capitol Hill | Queen Anne market areas, which include the neighborhoods encompassing Amazon's various campuses.
- A rather large number of people joined as host in August 2013.

Month_Year2 <S3: yearmonth>	n <int>	pct <dbl>
2013 Aug	279	5.15616337
2015 Nov	198	3.65921271
2015 Dec	149	2.75364997
2016 Jul	147	2.71668823
2016 Jan	111	2.05137682
2014 Aug	100	1.84808723
2017 Feb	98	1.81112549
2014 Jul	97	1.79264461
2017 Jul	92	1.70024025
2015 Jul	84	1.55239327

year <int>	n <int>	pct <dbl>
2015	1042	19.2570689
2013	817	15.0988727
2014	792	14.6368509
2016	737	13.6204029
2017	479	8.8523378
2012	429	7.9282942
2018	368	6.8009610
2019	330	6.0986879
2011	278	5.1376825
2010	97	1.7926446

[illegible][illegible]

What's in a house?

Essentials

Shampoo

Hangers

Kitchen

Wifi

Iron

Stove

Heating

TV

Dryer

Internet

Hair dryer

Oven

Elevator

Dishwasher

Coffee maker

Self check-in

Bathtub

Cable TV

Indoor fireplace

Gym

Room-darkening shades

Smart lock

Lockbox

Safety card

Lake access

Wide hallways

Key pad

Doggie

Other

Hot tub

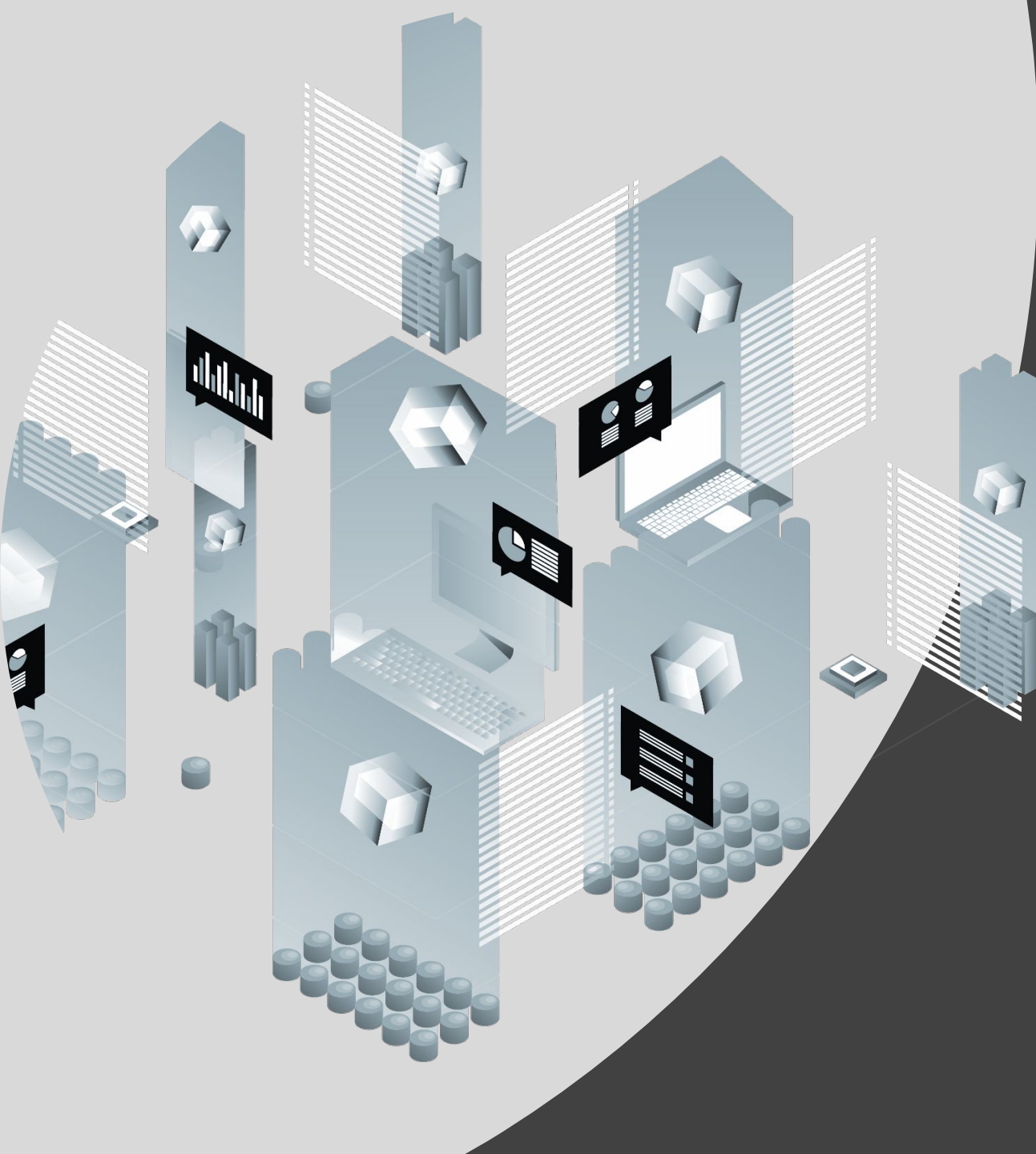
Breakfast

Word Clouds

Cleaning Fee vs High Booking Rate

- The median cleaning fee for establishments with low booking rate is higher than those with high booking rate - \$80 vs \$50.
- Establishments with low booking rate have exceptionally large number of outliers till \$555.





Recommendations

Things to consider while investing



Where and When to invest in Seattle?



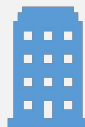
Festivals & major events are mainly between June – August; 2021



For a steady income, invest near University District or Near Capitol Hill



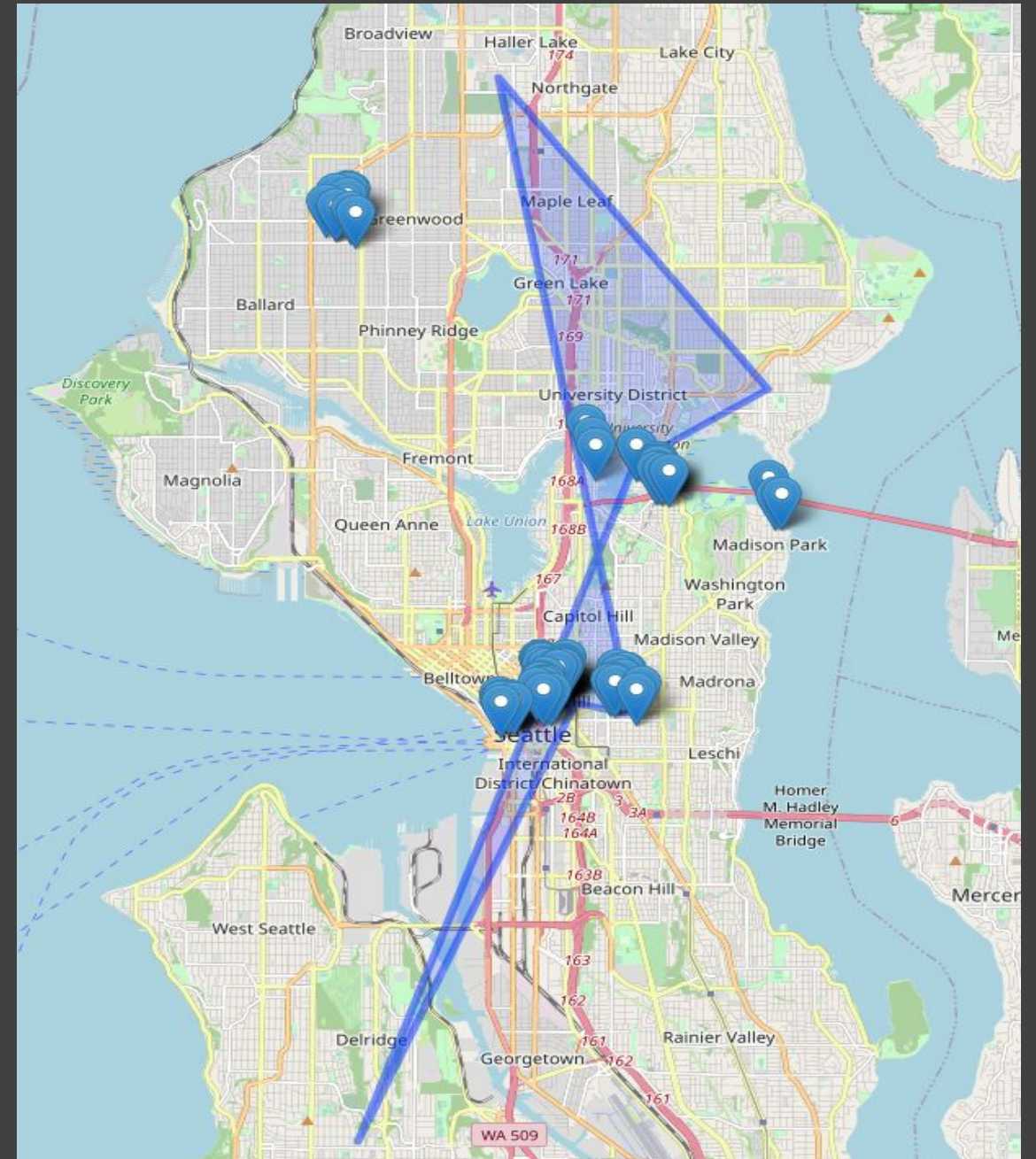
Home Prices slowing but rising



Near Business headquarters

What about COVID-19?

The marked listings can be rented by authorities to provide quality accommodation to front line workers & at the same time the best properties can keep their business running.



THANK YOU!
