**Final Project – Data Analysis Project Proposal**

UIUC: Su’21 STAT 420 - Statistical Modeling in R

**Analysis of Airbnb data for Sydney, Australia**

Team Members: Cyril Karpenko, Paul Whitely, Supriya Puri

Our proposal is to analyze Airbnb data for Sydney, Australia in order to understand how various predictors influence the rental price of the properties. `Prices` vary mostly on the basis of the types of ‘property’, ‘rooms’, ‘the neighborhood’, ‘proximity to tourist spots’ etc. However, we certainly cannot overlook other variables like ‘number of beds’, ‘cleaning fee’, ‘number of nights’ as these factors help in selecting a particular Airbnb listing.

**Why do we want look at Airbnb data?**

In recent years, we have observed how the use of various collaborative platforms like Airbnb, Uber, Zipcar have become prevalent and how these platforms present more unique, personalized way of experiencing the world in a budget friendly manner. As a potential guest, one might wonder what factors govern the price for a property - whether the listed price is fair according to the facilities provided or the neighborhood it is in. In a similar manner, the potential owners might want to determine what should be an ideal `rental price` for the property and how they can improve/upgrade the property to gain more monetary benefits.

Since its humble beginnings, Airbnb has made no secret of its heavy use of data science to build new product offerings, improve its service and capitalize on new marketing initiatives. Similarly, users and hosts rely on data science to gain better economic benefits.

We have acquired our data from open data on Kaggle at:

<https://www.kaggle.com/tylerx/sydney-airbnb-open-data>.

Our primary dataset will be acquired from the ***listings\_dec18.csv*** file. This data includes 96 columns and 36,663 rows of information. From the given variables, we believe that below mentioned variables are among the ones which have greatest importance:

* neighbourhood - where in city is the Airbnb located, closer to offices or to tourist places as per the guest’s requirement
* property\_type - is it an apartment, a townhouse or a house
* room\_type - whether the listing is for a private room or the entire property
* accommodates - how many people can stay at a time in the property
* beds – how many beds are available to accommodate guests
* bathrooms – number of bathrooms vary per the guests visiting
* amenities – internet, parking, washer included
* host\_is\_superhost – hosts and property need to qualify under various matrices to be a superhost
* cleaning\_fee - since it adds on to the total final price
* number\_of\_reviews - the properties with more reviews are generally booked early.

We will pre-process the data to remove unused fields, missing or invalid information before analyzing with R. We chose Sydney, Australia Airbnb data because it sounded like an interesting region to study and also the Airbnb data is sufficiently large, accurate and readily available to be both useful and timely to study.

We’ve created an R markdown document ***BNBexplore.rmd*** to learn about the primary and reference datasets that we will be using. The following is a screen shot showing a load and header view of the data using `price` as a response variable.

Graphical user interface, text, application

Description automatically generated

Table

Description automatically generated

We will use additional data retrieved from Kaggle to help us analyze the primary ***“Listings”*** data.

We intend to study data from the following three datasets to assess their suitability to provide predictor variables:

* Points of Interest: Entertainment, Leisure, Sporting, Shopping
* Transportation: Train Stations, Docks or Airports
* Neighborhoods

We will use the ***“Explore”*** R-markdown document to explore and learn about the datasets before cleansing them to discard the unneeded data. Our overall objective will be to determine how Airbnb prices vary depending not only on the listed features of a rental, but when correlated against predictors from the surrounding cities.

Kaggle provides a visual tool to generate data for points of interest. The area selected was approximately the city-center of Sydney, Australia. (shown below)

Map

Description automatically generated

**Appendix**

Structure of Listing Data :

str(listingData)

## 'data.frame': 36662 obs. of 96 variables:

## $ id : int 12351 14250 15253 20865 26174 38073 39348 44545 45440 56842 ...

## $ listing\_url : chr "https://www.airbnb.com/rooms/12351" "https://www.airbnb.com/rooms/14250" "https://www.airbnb.com/rooms/15253" "https://www.airbnb.com/rooms/20865" ...

## $ scrape\_id : num 2.02e+13 2.02e+13 2.02e+13 2.02e+13 2.02e+13 ...

## $ last\_scraped : chr "2018-12-07" "2018-12-07" "2018-12-07" "2018-12-07" ...

## $ name : chr "Sydney City & Harbour at the door" "Manly Harbour House" "Stunning Penthouse Apartment In Heart Of The City" "3 BED HOUSE + 1 BED STUDIO Balmain" ...

## $ summary : chr "Come stay with Vinh & Stuart (Awarded as one of Australia's top hosts by Airbnb CEO Brian Chesky & key sharehol"| \_\_truncated\_\_ "Beautifully renovated, spacious and quiet, our 3 Bedroom, 3 Bathroom home is only a 10 minute walk to beaches "| \_\_truncated\_\_ "Penthouse living in a great central location: You will be staying in a unique apartment on the top floor of a c"| \_\_truncated\_\_ "Hi! We are a married professional couple with 2 kids. When making a booking, please tell us a little bit about"| \_\_truncated\_\_ ...

## $ space : chr "We're pretty relaxed hosts, and we fully appreciate staying with someone else, in their home home, is not for e"| \_\_truncated\_\_ "Our home is a thirty minute walk along the seashore promenade to Manly, one of Sydney's most beautiful beaches,"| \_\_truncated\_\_ "A charming two-level, two-bedroom, two-bathroom duplex apartment on the border of East Sydney, Darlinghurst and"| \_\_truncated\_\_ "HOUSE : \_\_\_\_\_\_\_\_\_ \* DUCTED AIR CONDITIONING IN ALL ROOMS \* BEDROOM 1 : QUEEN BED \* BEDROOM 2 : QUEEN BED \* BEDR"| \_\_truncated\_\_ ...

## $ description : chr "Come stay with Vinh & Stuart (Awarded as one of Australia's top hosts by Airbnb CEO Brian Chesky & key sharehol"| \_\_truncated\_\_ "Beautifully renovated, spacious and quiet, our 3 Bedroom, 3 Bathroom home is only a 10 minute walk to beaches "| \_\_truncated\_\_ "Penthouse living in a great central location: You will be staying in a unique apartment on the top floor of a c"| \_\_truncated\_\_ "Hi! We are a married professional couple with 2 kids. When making a booking, please tell us a little bit about"| \_\_truncated\_\_ ...

## $ experiences\_offered : chr "none" "none" "none" "none" ...

## $ neighborhood\_overview : chr "Pyrmont is an inner-city village of Sydney, only about 2kms from the Sydney CBD (Central Business District) / C"| \_\_truncated\_\_ "Balgowlah Heights is one of the most prestigious areas on the Northern Beaches. Filled with seaside character "| \_\_truncated\_\_ "The location is really central and there is number of things to do and see all within a few kilometres; Stanley"| \_\_truncated\_\_ "BALMAIN is an older inner city village / suburb with numerous cafes, restaurants, parks, walks around the harbo"| \_\_truncated\_\_ ...

## $ notes : chr "We've a few reasons for the 6.00pm arrival time / check in / bag drop time... One of them is that we're really "| \_\_truncated\_\_ "" "The apartment is on the top floor with 4 flights of stairs to walk and there is NO lift, it is great for workin"| \_\_truncated\_\_ "During your stay guests may use the landline for all local, national and 1300 calls to landlines or mobiles. T"| \_\_truncated\_\_ ...

## $ transit : chr "Our home is centrally located and an easy walk to everywhere (everywhere you want to go anyways). We also have "| \_\_truncated\_\_ "Balgowlah - Manly bus # 131 or #132 (Bus stop is 3 doors down, around the corner) takes you to Manly Ferry in "| \_\_truncated\_\_ "If walking is not your thing I am well serviced for public transport with a bus stop almost outside of door tha"| \_\_truncated\_\_ "Balmain is well serviced by buses (from Darling St, Mullins St and Victoria Rd), but you can also catch a ferry"| \_\_truncated\_\_ ...

## $ access : chr "We look forward to welcoming you just as we would welcome our family and friends. We'll provide you with a code"| \_\_truncated\_\_ "Guests have access to whole house except locked basement and storage room. (Rear ground level studio not included.)" "Guests have own key and access to use facilities of apartment including Roof Terrace." "Whole house (including studio/flat & courtyard gardens)" ...

## $ interaction : chr "As much or as little as you like. We live here. This is our home. We will be at home when you stay, and genera"| \_\_truncated\_\_ "" "You would be staying with me and my very cute, house trained small dog Charlie he is a 14 years old Pomeranian,"| \_\_truncated\_\_ "We will always be available should you require anything, but otherwise, not at all." ...

## $ house\_rules : chr "We look forward to welcoming you to stay you just as we would our family and friends. \"Farm Gate Rules\" - if "| \_\_truncated\_\_ "Standard Terms and Conditions of Temporary Holiday Accommodation Note: Variances can be agreed on but only by a"| \_\_truncated\_\_ "I am fairly easygoing and will try to accomodate guests reasonable requests. I ask that guest treat my home wi"| \_\_truncated\_\_ "PLEASE ENJOY YOURSELVES WITHOUT MAKING TOO MUCH NOISE AS WE HAVE VERY GOOD NEIGHBOURS. NO SMOKING INSIDE THE H"| \_\_truncated\_\_ ...

## $ thumbnail\_url : logi NA NA NA NA NA NA ...

## $ medium\_url : logi NA NA NA NA NA NA ...

## $ picture\_url : chr "https://a0.muscache.com/im/pictures/763ad5c8-c951-43e0-b926-4a98c25c45e8.jpg?aki\_policy=large" "https://a0.muscache.com/im/pictures/56935671/fdb8c0bf\_original.jpg?aki\_policy=large" "https://a0.muscache.com/im/pictures/3643414/2ed84e0e\_original.jpg?aki\_policy=large" "https://a0.muscache.com/im/pictures/58785124/c690682a\_original.jpg?aki\_policy=large" ...

## $ xl\_picture\_url : logi NA NA NA NA NA NA ...

## $ host\_id : int 17061 55948 59850 64282 110561 103476 168828 112237 201611 113874 ...

## $ host\_url : chr "https://www.airbnb.com/users/show/17061" "https://www.airbnb.com/users/show/55948" "https://www.airbnb.com/users/show/59850" "https://www.airbnb.com/users/show/64282" ...

## $ host\_name : chr "Stuart" "Heidi" "Morag" "Fiona" ...

## $ host\_since : chr "2009-05-14" "2009-11-20" "2009-12-03" "2009-12-19" ...

## $ host\_location : chr "Sydney, New South Wales, Australia" "Sydney, New South Wales, Australia" "Sydney, New South Wales, Australia" "Sydney, New South Wales, Australia" ...

## $ host\_about : chr "G'Day from Australia!\n\nHe's Vinh, and I'm Stuart...\n\nWe're big fans of AirBnB for nearly a decade having us"| \_\_truncated\_\_ "I am a Canadian who has made Australia her home for over ten years. I appreciate a good place to stay, having h"| \_\_truncated\_\_ "I am originally Scottish but I have made Sydney my home for the last 14 years, I love the city and the country"| \_\_truncated\_\_ "We are a professional, married couple with 2 children. " ...

## $ host\_response\_time : chr "within a few hours" "within a few hours" "within an hour" "within a day" ...

## $ host\_response\_rate : chr "100%" "90%" "100%" "100%" ...

## $ host\_acceptance\_rate : chr "N/A" "N/A" "N/A" "N/A" ...

## $ host\_is\_superhost : chr "f" "f" "f" "t" ...

## $ host\_thumbnail\_url : chr "https://a0.muscache.com/im/users/17061/profile\_pic/1425534501/original.jpg?aki\_policy=profile\_small" "https://a0.muscache.com/im/users/55948/profile\_pic/1327370146/original.jpg?aki\_policy=profile\_small" "https://a0.muscache.com/im/pictures/user/730ee903-bc77-40ad-b55e-966966de19bb.jpg?aki\_policy=profile\_small" "https://a0.muscache.com/im/users/64282/profile\_pic/1266145067/original.jpg?aki\_policy=profile\_small" ...

## $ host\_picture\_url : chr "https://a0.muscache.com/im/users/17061/profile\_pic/1425534501/original.jpg?aki\_policy=profile\_x\_medium" "https://a0.muscache.com/im/users/55948/profile\_pic/1327370146/original.jpg?aki\_policy=profile\_x\_medium" "https://a0.muscache.com/im/pictures/user/730ee903-bc77-40ad-b55e-966966de19bb.jpg?aki\_policy=profile\_x\_medium" "https://a0.muscache.com/im/users/64282/profile\_pic/1266145067/original.jpg?aki\_policy=profile\_x\_medium" ...

## $ host\_neighbourhood : chr "Pyrmont" "Balgowlah" "Darlinghurst" "Balmain/Birchgrove" ...

## $ host\_listings\_count : int 2 2 2 2 1 1 2 1 10 54 ...

## $ host\_total\_listings\_count : int 2 2 2 2 1 1 2 1 10 54 ...

## $ host\_verifications : chr "['email', 'phone', 'manual\_online', 'reviews', 'manual\_offline', 'offline\_government\_id', 'government\_id', 'work\_email']" "['email', 'phone', 'reviews', 'jumio', 'offline\_government\_id', 'government\_id']" "['email', 'phone', 'facebook', 'reviews', 'jumio', 'offline\_government\_id', 'government\_id']" "['email', 'phone', 'reviews', 'jumio', 'government\_id', 'work\_email']" ...

## $ host\_has\_profile\_pic : chr "t" "t" "t" "t" ...

## $ host\_identity\_verified : chr "t" "t" "t" "t" ...

## $ street : chr "Pyrmont, NSW, Australia" "Balgowlah, NSW, Australia" "Darlinghurst, NSW, Australia" "Balmain, NSW, Australia" ...

## $ neighbourhood : chr "Pyrmont" "Balgowlah" "Darlinghurst" "Balmain/Birchgrove" ...

## $ neighbourhood\_cleansed : chr "Sydney" "Manly" "Sydney" "Leichhardt" ...

## $ neighbourhood\_group\_cleansed : logi NA NA NA NA NA NA ...

## $ city : chr "Pyrmont" "Balgowlah" "Darlinghurst" "Balmain" ...

## $ state : chr "NSW" "NSW" "NSW" "NSW" ...

## $ zipcode : chr "2009" "2093" "2010" "2041" ...

## $ market : chr "Sydney" "Sydney" "Sydney" "Sydney" ...

## $ smart\_location : chr "Pyrmont, Australia" "Balgowlah, Australia" "Darlinghurst, Australia" "Balmain, Australia" ...

## $ country\_code : chr "AU" "AU" "AU" "AU" ...

## $ country : chr "Australia" "Australia" "Australia" "Australia" ...

## $ latitude : num -33.9 -33.8 -33.9 -33.9 -33.9 ...

## $ longitude : num 151 151 151 151 151 ...

## $ is\_location\_exact : chr "t" "t" "f" "t" ...

## $ property\_type : chr "Townhouse" "House" "Apartment" "House" ...

## $ room\_type : chr "Private room" "Entire home/apt" "Private room" "Entire home/apt" ...

## $ accommodates : int 2 6 2 8 2 2 2 2 11 4 ...

## $ bathrooms : num 1 3 1 2 1 1 1 1 3 1 ...

## $ bedrooms : int 1 3 1 4 1 0 1 1 5 2 ...

## $ beds : int 1 3 1 4 1 1 1 1 7 2 ...

## $ bed\_type : chr "Real Bed" "Real Bed" "Real Bed" "Real Bed" ...

## $ amenities : chr "{TV,Internet,Wifi,\"Air conditioning\",\"Paid parking off premises\",Breakfast,Heating,\"Smoke detector\",\"Car"| \_\_truncated\_\_ "{TV,Wifi,\"Air conditioning\",Kitchen,\"Pets live on this property\",Cat(s),\"Free street parking\",Heating,Was"| \_\_truncated\_\_ "{TV,\"Cable TV\",Internet,Wifi,\"Air conditioning\",Kitchen,\"Paid parking off premises\",\"Pets allowed\",\"Pe"| \_\_truncated\_\_ "{TV,Internet,Wifi,\"Air conditioning\",Kitchen,\"Pets allowed\",\"Pets live on this property\",Cat(s),\"Indoor "| \_\_truncated\_\_ ...

## $ square\_feet : int NA NA NA NA NA 0 323 NA NA 0 ...

## $ price : chr "$100.00" "$471.00" "$109.00" "$450.00" ...

## $ weekly\_price : chr "$800.00" "$3,000.00" "$800.00" "$2,804.00" ...

## $ monthly\_price : chr "" "" "$3,000.00" "$7,009.00" ...

## $ security\_deposit : chr "$0.00" "$900.00" "" "$800.00" ...

## $ cleaning\_fee : chr "$55.00" "$100.00" "" "" ...

## $ guests\_included : int 2 6 1 6 1 2 1 1 10 1 ...

## $ extra\_people : chr "$395.00" "$40.00" "$10.00" "$0.00" ...

## $ minimum\_nights : int 2 5 2 7 1 2 5 5 3 2 ...

## $ maximum\_nights : int 7 22 7 365 60 730 1125 365 365 200 ...

## $ calendar\_updated : chr "today" "9 months ago" "4 days ago" "7 weeks ago" ...

## $ has\_availability : chr "t" "t" "t" "t" ...

## $ availability\_30 : int 6 12 9 0 6 12 7 1 14 0 ...

## $ availability\_60 : int 16 16 22 17 20 16 7 1 30 0 ...

## $ availability\_90 : int 28 46 41 17 50 46 7 1 48 0 ...

## $ availability\_365 : int 187 321 316 69 140 321 194 1 301 0 ...

## $ calendar\_last\_scraped : chr "2018-12-07" "2018-12-07" "2018-12-07" "2018-12-07" ...

## $ number\_of\_reviews : int 493 1 300 15 42 63 6 53 54 17 ...

## $ first\_review : chr "2010-07-24" "2016-01-02" "2012-02-23" "2012-01-03" ...

## $ last\_review : chr "2018-11-21" "2016-01-02" "2018-11-16" "2018-01-07" ...

## $ review\_scores\_rating : int 95 100 88 96 94 97 87 98 95 80 ...

## $ review\_scores\_accuracy : int 10 10 9 9 10 10 8 10 10 8 ...

## $ review\_scores\_cleanliness : int 9 10 9 9 10 10 8 10 9 8 ...

## $ review\_scores\_checkin : int 10 10 9 9 10 10 9 10 10 8 ...

## $ review\_scores\_communication : int 10 8 9 10 10 10 10 10 10 9 ...

## $ review\_scores\_location : int 10 10 9 10 10 9 8 10 10 10 ...

## $ review\_scores\_value : int 10 10 9 9 9 9 8 10 9 8 ...

## $ requires\_license : chr "f" "f" "f" "f" ...

## $ license : chr "" "" "" "" ...

## $ jurisdiction\_names : logi NA NA NA NA NA NA ...

## $ instant\_bookable : chr "f" "f" "t" "f" ...

## $ is\_business\_travel\_ready : chr "f" "f" "f" "f" ...

## $ cancellation\_policy : chr "strict\_14\_with\_grace\_period" "strict\_14\_with\_grace\_period" "strict\_14\_with\_grace\_period" "strict\_14\_with\_grace\_period" ...

## $ require\_guest\_profile\_picture : chr "t" "f" "f" "t" ...

## $ require\_guest\_phone\_verification: chr "t" "f" "f" "t" ...

## $ calculated\_host\_listings\_count : int 2 2 2 1 1 1 2 1 10 53 ...

## $ reviews\_per\_month : num 4.83 0.03 3.63 0.18 0.45 0.64 0.77 0.54 0.88 0.24 ...