PANAM

airbnb in Washington D.C.

PANAM Consulting

Max Pedersen – z5164270

Prerita Mehta – z5162933

Anderson Wong – z5076423

Noumik Thadani – z5246273

Andrew Rong – z5059252

UNSW SYDNEY | Australia's Global University

UNSW Business School
Information Systems and Technology Management

ASSESSMENT COVER PAGE

Title of Assignment: INFS3603 Team Assignment

Date Due: 4 April 2019 **Date Submitted:** 4 April 2019

I/We declare that this assessment item is my/our own work, except where acknowledged, and has not been submitted for academic credit elsewhere, and acknowledge that the assessor of this item may, for the purpose of assessing this item:(i)Reproduce this assessment item and provide a copy to another member of the University; and/or,II)Communicate a copy of this assessment item to a plagiarism checking service (which may then retain a copy of the assessment item on its database for the purpose of future plagiarism checking).

I/We certify that I/we have read and understood the University Rules in respect of Student Academic Misconduct.

ID	Name	Signature	Date
z5059252	Andrew Rong		30/3/2019
z5164210	Max Pedersen		30/3/2019
z5162933	Prerita Mehta		30/03/2019
z5246273	Noumik Thadani		30/30/19
z5076423	Anderson Wong		30/03/19

This cover sheet has to be completed and signed by ALL members in the group for the assignment submitted. **Note: 10 percent of the marks available for the assessment will be deducted for assessments submitted without a fully completed and signed cover page.**

Table of Contents

1. Background and Objectives

1.1 Key Objectives

The primary purpose of this report is to determine how Airbnb can subsequently expand market share, given its current position in the Washington D.C. market. Researching Airbnb's positioning within the rental apartment space revealed that leveraging the insights produced from their datasets could improve their overall position within Washington D.C. This principal goal also furthers Airbnb's vision of providing a "people-to-people" platform for travel bookings and experiences, to benefit all stakeholders including hosts, guests, and communities. Two underlying objectives stem from this primary purpose:

1. Where to play?
   Focusing on the customer environment, the main aim is to determine guest interaction with current geographic segments, and an optimal method of expansion
2. How to play in the most attractive segments?
   Focusing on the host environment, the main aim is to determine the elasticity of hosts to support an expansion, through conducting host segmentation analysis.

1.2 Data Science Profiles

Key analytic strengths were established prior to conducting the exploration and reflected that the group has a sound balance between the possible six data science attributes – namely: data wrangler (Noumik), modeler (Max), programmer (Anderson), visualizer (Andrew) and communicator (Prerita) (Appendix A).

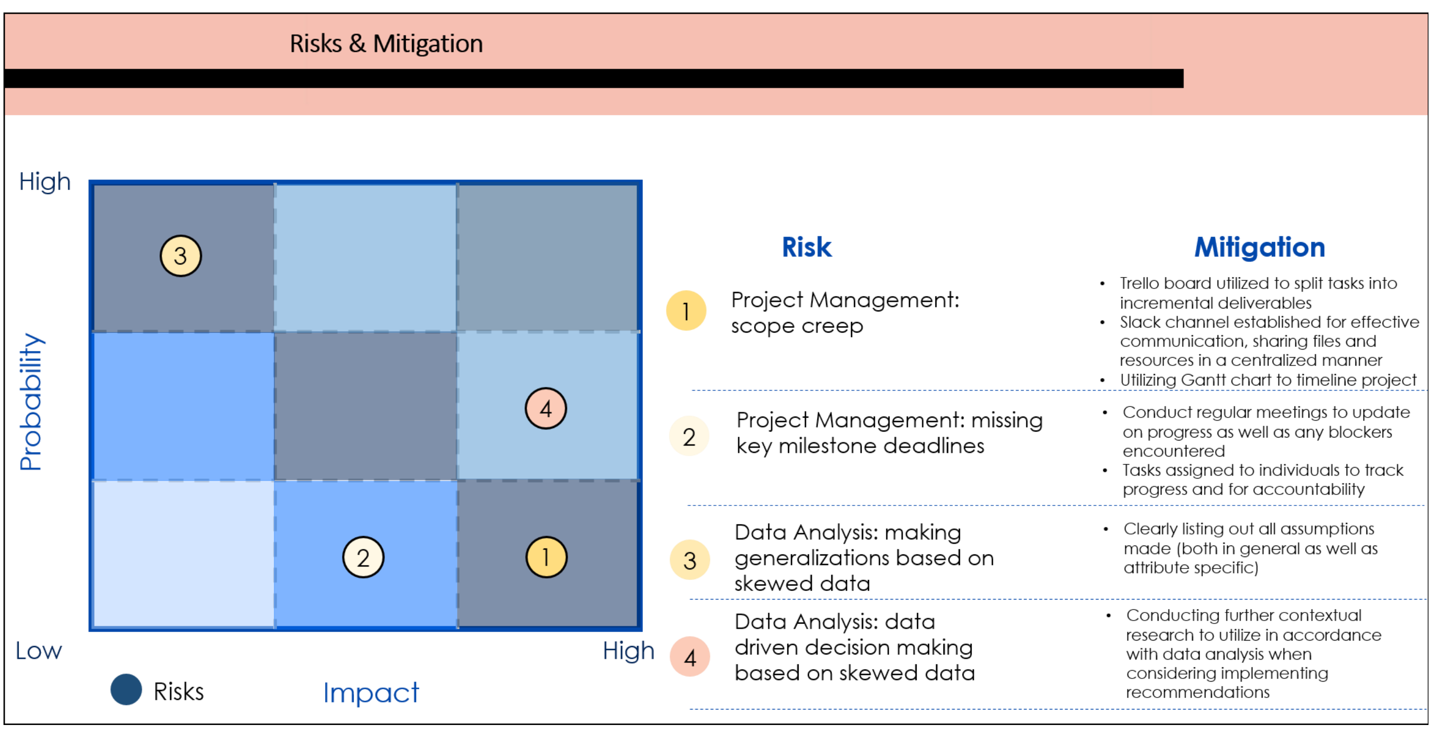
This, in turn, resulted in leveraging individuals' core skills and defining roles and responsibilities to conduct exploration and analysis in the most efficient and effective way. Noumik capitalized on his strong problem skills by extracting, cleansing and manipulating data into a useable format. This allowed Max to derive meaningful insights from the data by paying close attention to detail, particularly underlying assumptions and limitations in the data to predict and optimize decision-making. Anderson then utilized various software and his proficiency in technical languages to maximize the accuracy of the findings, while Andrew creatively conceptualized the data using graphical tools to create a cohesive story from the data. Prerita drew upon her deep understanding of the business context and implications of the findings to subsequently effectively communicate data complexities and technical understanding in an intelligible manner.

1.3 Mitigating Risks

Devising a strong project management plan from the outset was fundamental to mitigating the risk of scope creep or slow progress. The group utilized Trello, an organizational tool, in accordance with regular project meetings to split tasks into incremental deliverables, assign them to individuals, and ensure accountability for deadlines. A Slack channel was also established, being particularly effective as a centralized medium of communication to share information, files, and updates.

In terms of data analysis, clearly listing out all assumptions is crucial to alleviating the risk of making generalizations based on a relatively small sample of data. Similarly, conducting further research and appropriately extending the given data set is vital in providing informed recommendations for how Airbnb can expand optimally.

Figure 1: Risks & Mitigation Matrix



2. Analysis

2.1 Overview

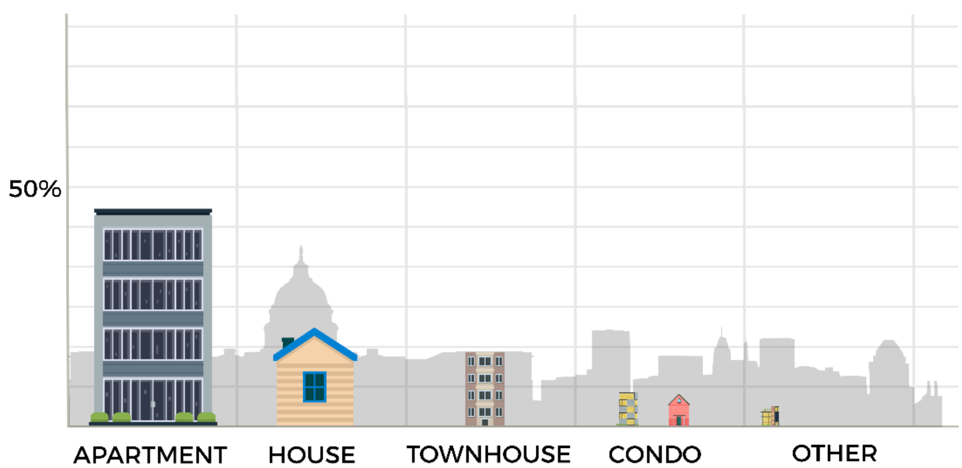
Washington D.C. receives more than 20 million overseas visitors annually, making it the 8th most popular destination within the US, at 5.56% of market share. As the US capital, it is both a political stronghold and prime tourist destination, having a strong inflow of government and business-related travel. Through analyzing key trends across various elements that comprise of the Airbnb environment including properties, hosts, guests, and sentiment, it becomes apparent that Washington D.C. is an emerging and lucrative market for Airbnb to further expand in.

2.2 Current Market Positioning & Expansion Strategy

Property Listing Analysis

For Airbnb to increase revenue, property listings must be determined to consequently derive consumers' preferences and their likelihood of renting a property. There are approximately 125 neighborhoods included in the dataset of Airbnb listings in Washington DC. Figure 2 reveals apartments and houses collectively comprise 63.7% of the Washington DC market, hence forming the predominant focus on Airbnb's market expansion for property preferences.

Figure 2: Property Types



Geographical Clusters Analysis

It is imperative for Airbnb to determine the concentration of its listings as it indicates where Airbnb holds the greatest competitive advantage and understanding of neighborhoods.

Figure 3: Concentration of Listings

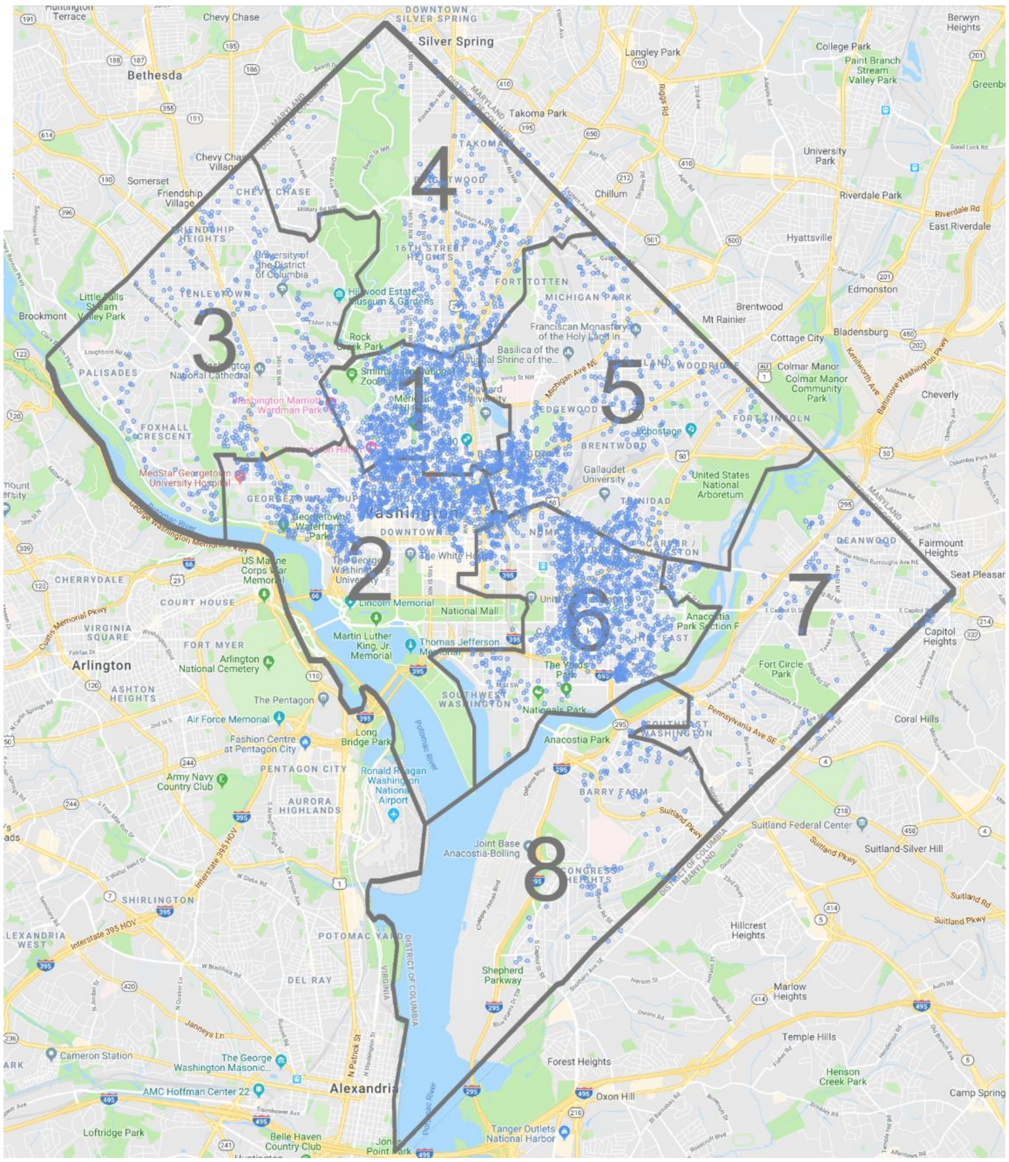


Figure 3 indicates that currently, Airbnb’s listings are primarily concentrated within Wards 1, 2, and 6. Wards 4, 7, and 8 have a lesser concentration, which could potentially be attractive for Airbnb to capitalize on in order to expand market share.

Guest Analysis

Of the 125 neighborhoods, the markets for each vary significantly, with Airbnb's markets consequently being segmented into established and emerging regions based on maturity and revenue, to realize where the most lucrative markets lie. Established markets have been classified as neighborhoods that have both a total inferred revenue of greater than $100,000 and 10+ listings in that respective segment (Appendix E). All other geographic segments not meeting these specific criteria are defined as emerging markets. Figures 4-7 depict inferred total revenue and number of listings for both emerging and established markets respectively.

Figure 4:

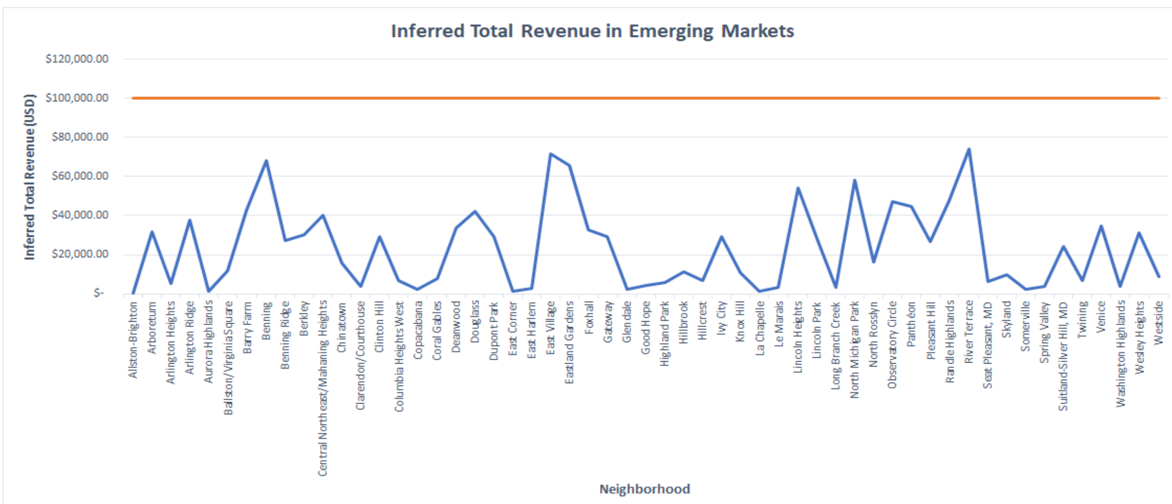


Figure 5:

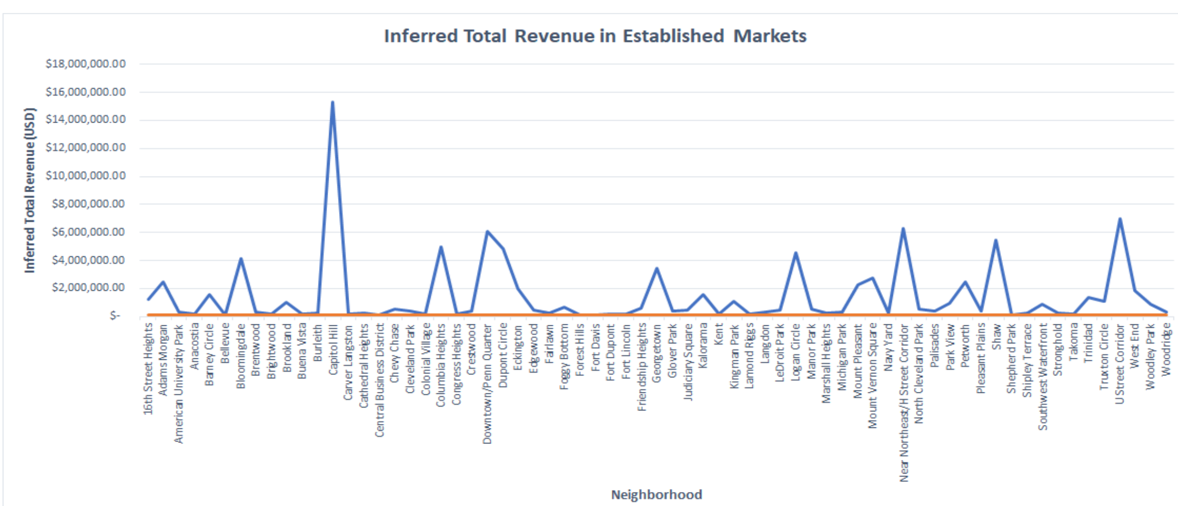


Figure 6:

Figure 7:

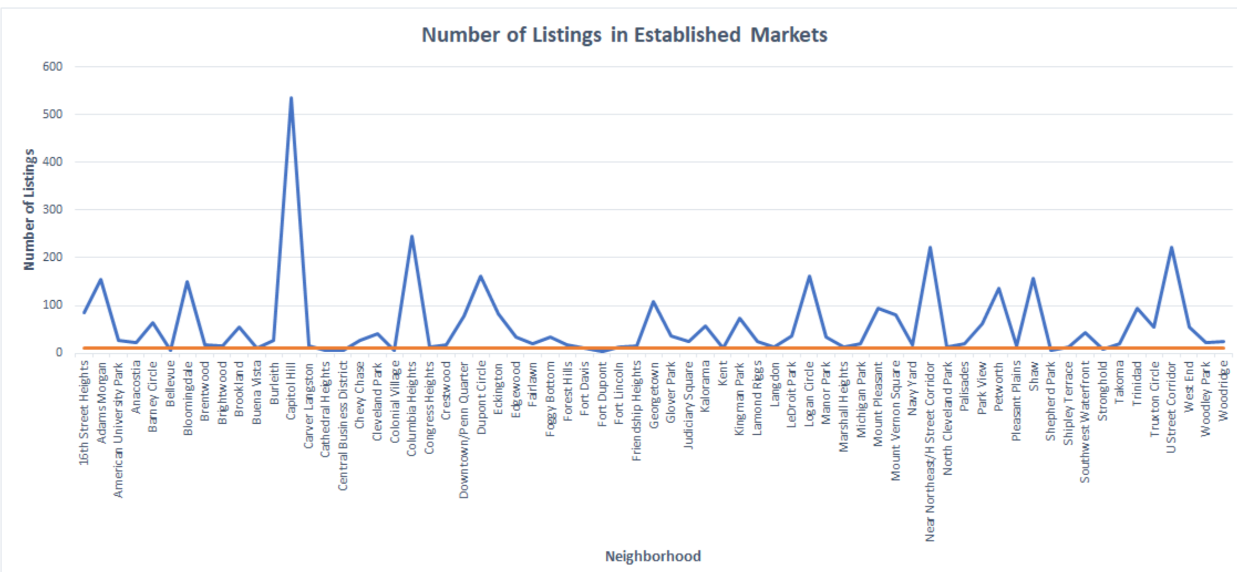
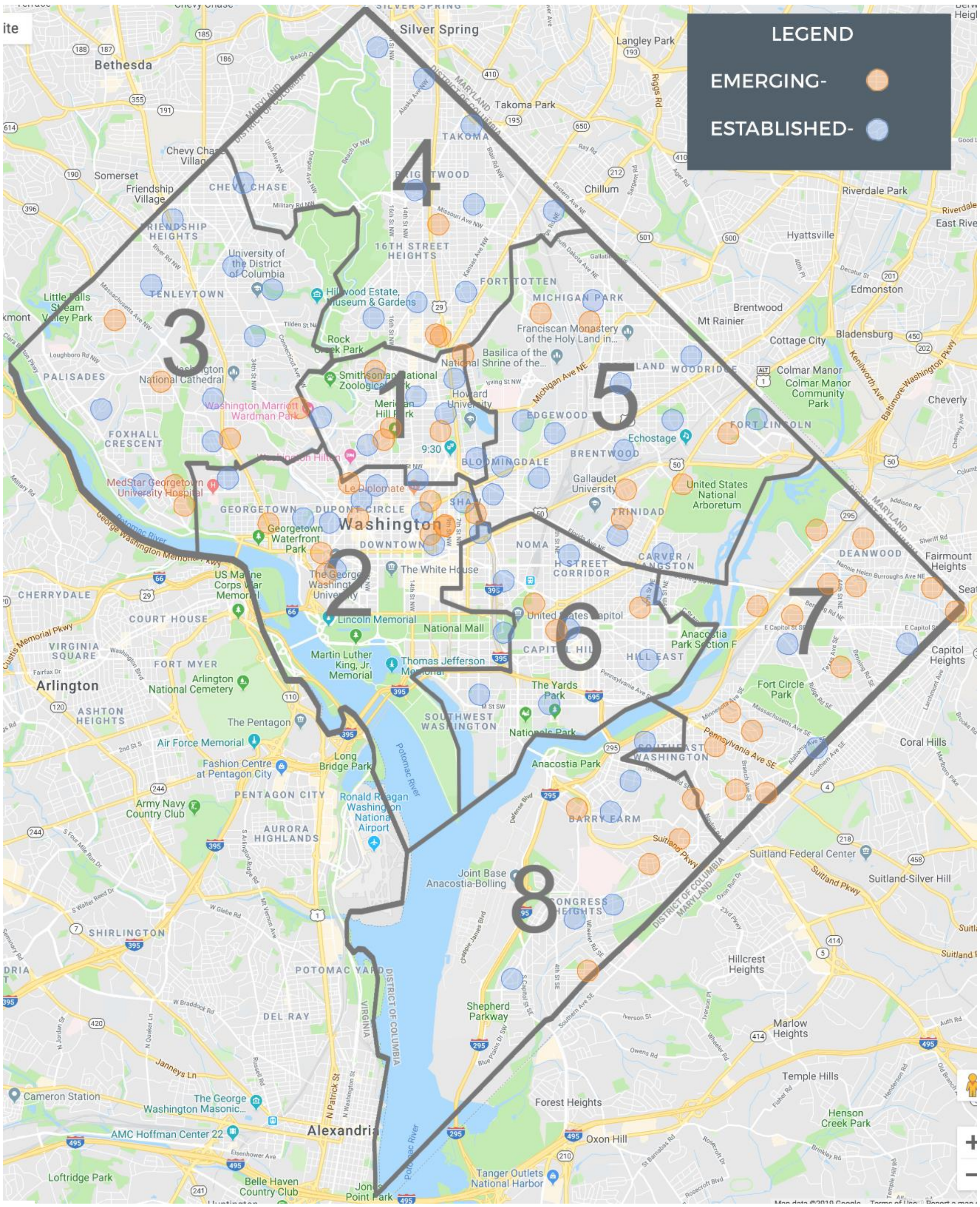


Figure 8 reveals the geographic concentration of both emerging (orange) and established (blue) markets in Washington DC. Wards 1, 2, 5 and 6 are more heavily concentrated with established markets, with Wards 7 & 8 having a larger proportion of emerging markets.

Figure 8:



Guest Review Analysis

Segmenting the markets into ‘emerging’ and ‘established’ was performed to enable analysis of the established market segment, and evaluation of property attributes based on guest reviews. Appendix E contains further explanation of calculations and logic, but some details are:

- **Price per Guest (PPG):** A new metric was established called price per guest (PPG) of a night to determine incremental revenue from an individual. This would account for differing distributions in property type across suburbs:

$$PPG = \frac{Price\ of\ listing}{Total\ Guests\ accommodated\ in\ the\ listing}$$

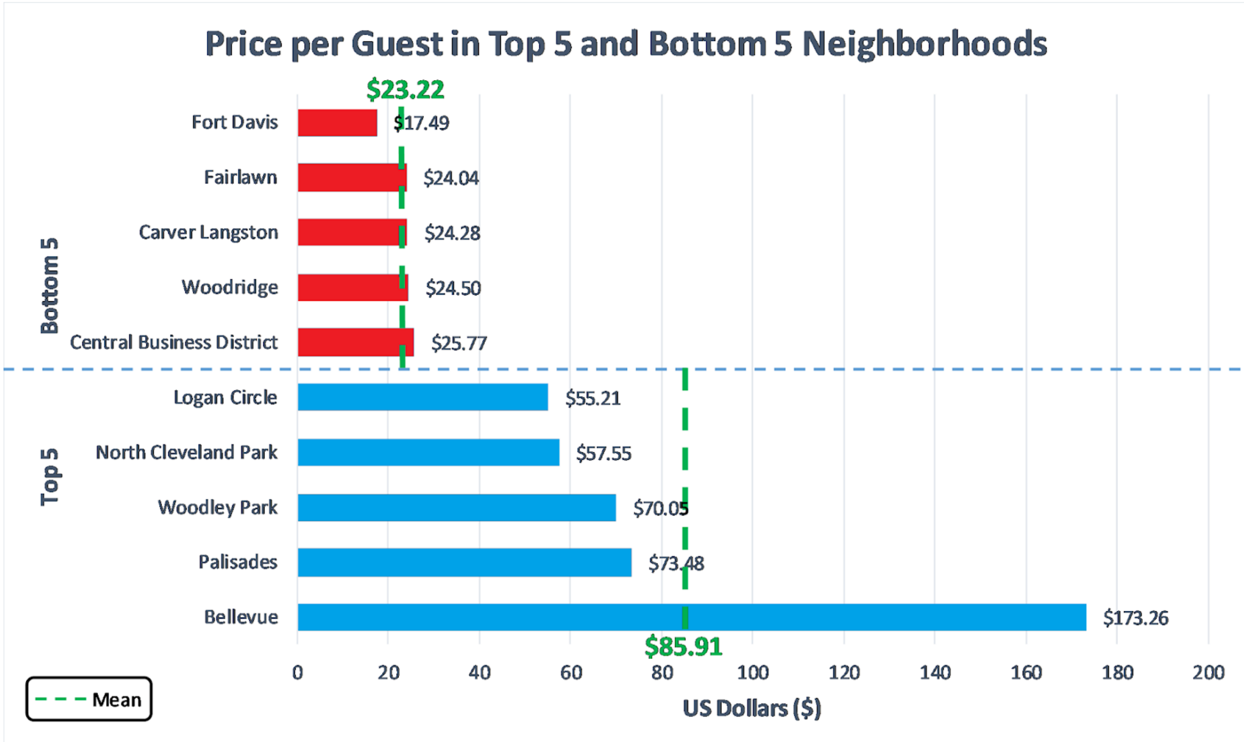
- **Value Ratings:** The value rating indicates a guest’s evaluation of the property’s value for money; hence, the distribution of these ratings across different price points can provide clearer insight into the effectiveness of the current pricing environment within D.C.

The analysis is focused on established markets, as there are wider spreads in price and a significantly larger size of properties and bookings that have occurred. Thus, data relating to customer feedback is likely to be more accurate, and inferences on customers’ willingness to pay certain prices can be made more easily.

Guest review analysis will evaluate Airbnb’s adjustments to price and value in the established markets, and determine if current levels are suitable for market expansion.

Average PPG Analysis

Figure 9:



As seen above, Airbnb's most attractive guests from a revenue perspective lie within Bellevue, Palisades, Woodley Park, North Cleveland Park, and Logan Circle. These neighborhoods have the highest PPGs, with a combined mean nightly PPG of $85.91. Airbnb's least attractive guests from a revenue perspective lie within Colonial Village, Central Business District, Woodridge, Carver Langston, & Fort Davis. These have the lowest PPGs, with a combined mean nightly PPG of $23.22. This equates to a 370% difference between the 5 least expensive neighborhoods and 5 most expensive neighborhoods, which is very significant.

Figure 10:

Average Value Review Score Analysis

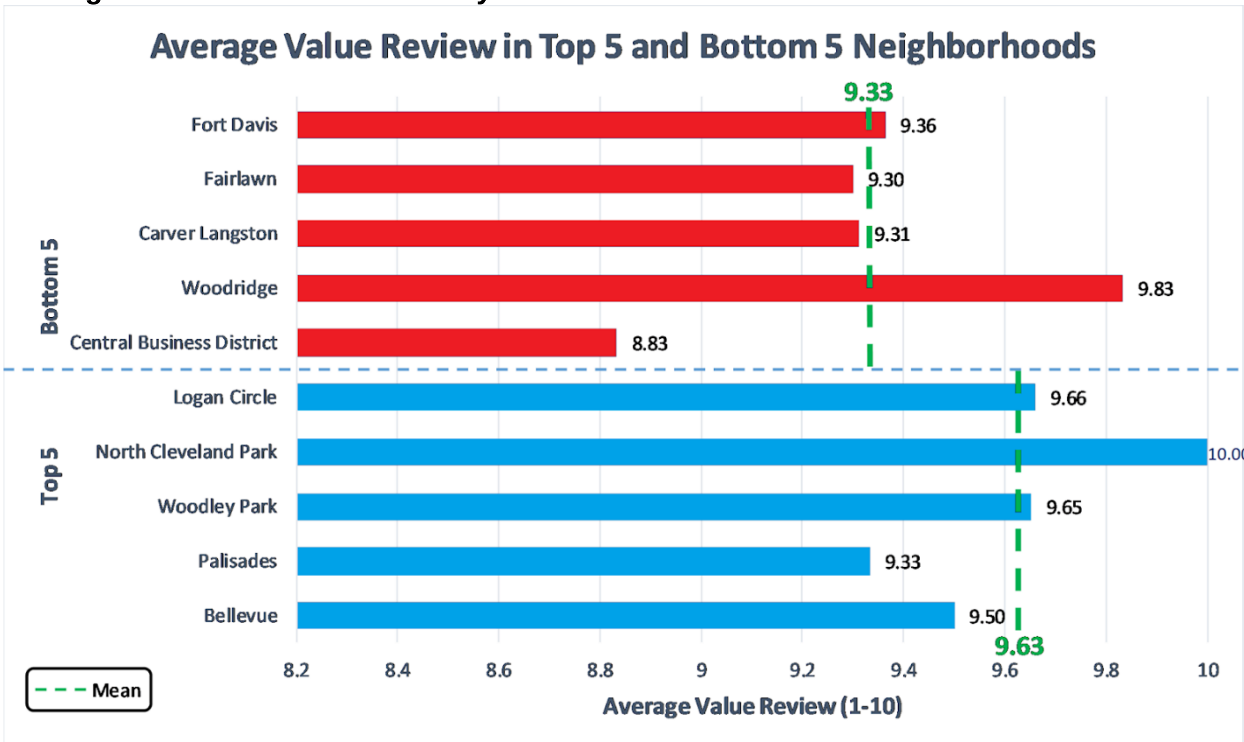


Figure 10 indicates the average value review for top 5 suburbs is 9.63, while the average value review for bottom 5 suburbs is 9.33. There is only a 0.086 difference in the mean value review for the 5 most expensive suburbs, and 5 least expensive suburbs. With a minimal variation in mean value scores between high-priced and low-priced segments, and an overall mean value review score higher than 9, it suggests that properties seem to be priced accordingly across geographical segments. Despite the significant difference in mean PPG demonstrated previously, there is no indication of significant overpricing occurring, with guests booking in specific neighborhoods based on their propensity to pay. De-aggregating this to an individual listing level, there is virtually no correlation between price and value rating (Pearson coefficient of -0.05) (Appendix H).

These findings are suggestive that the current range of pricing across geographical segments in Washington DC accommodates for differing guest propensities to spend, and that their value expectations from Airbnb properties at specific price points are being satisfied. As Airbnb considers market expansion options, pricing appears currently to be customer-driven at an optimal level (Collins et at. 2006), and so pricing adjustments to listings are unlikely to be a key component of an expansion strategy.

Location Ratings Analysis

Location ratings are the next consideration as an indicator of engagement with the current market. The percentage of established markets with a mean review score value of 9 or higher is also high at 79%, but less than value reviews (Appendix H). This reduction can be partially attributed to a group of suburbs in Wards 7 & 8, specifically Anacostia, Barry Farm and Fairlawn. Whilst Figure 2 would suggest their distance from key tourist areas is likely to influence lower reviews, these areas are lower in socioeconomic status and have higher crime rates in comparison to the D.C. average (Statistical Atlas 2019). Their mean location review scores are 7.45, 6.5, and 7.6 respectively, hinting a potential relationship between safety and location review.

Further evaluating the influence of exact location versus other factors including personal safety on location ratings, Google Maps' Distance Vector API was reverse-queried to determine the distance of each suburb's midpoint from Capitol Hill (Appendix H). Capitol Hill was chosen for its positioning as the largest established market by inferred revenue (Figure 5), hypothesized as indicative of strong demand for its location (Statistical Atlas 2019). Figures 11 & 12 show the lack of identifiable relationship between distance from Capitol Hill and average location review score, for established and emerging markets respectively.

Figure 11: Average Review Score by Distance, with # of listings (Established markets)

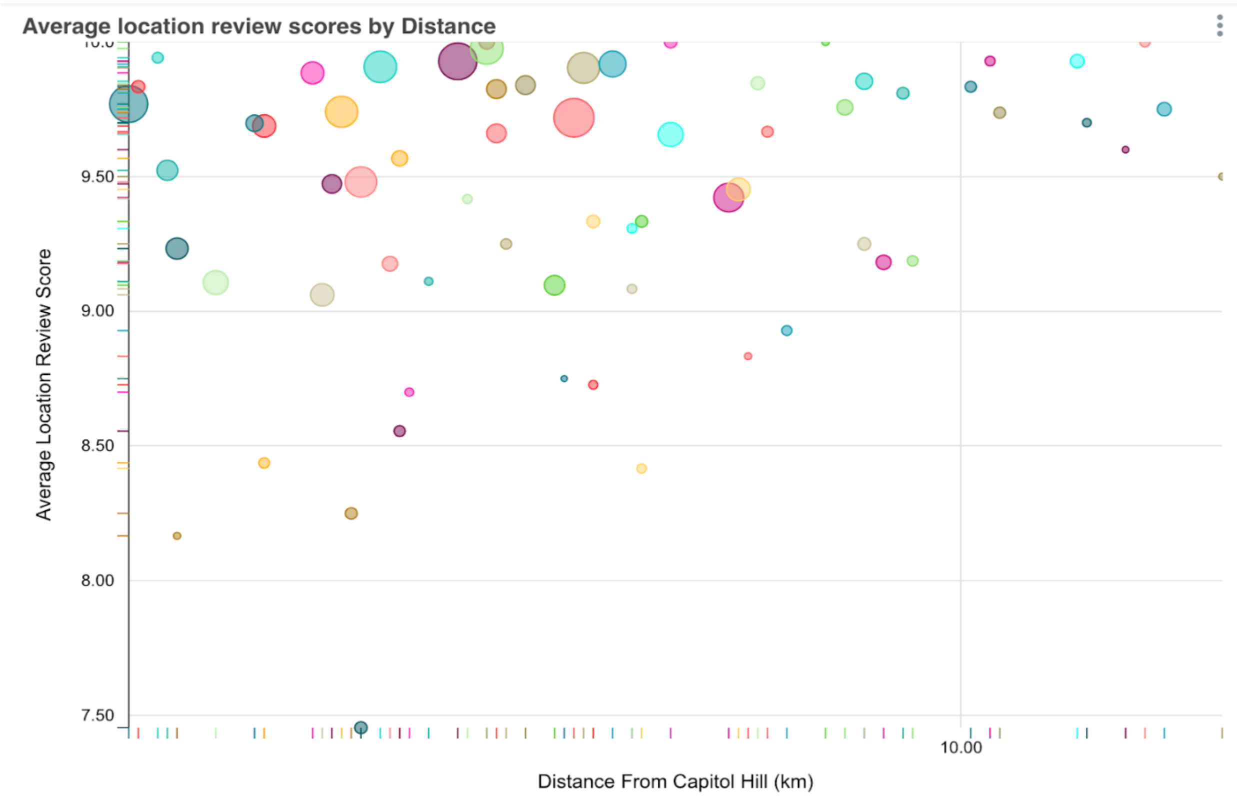
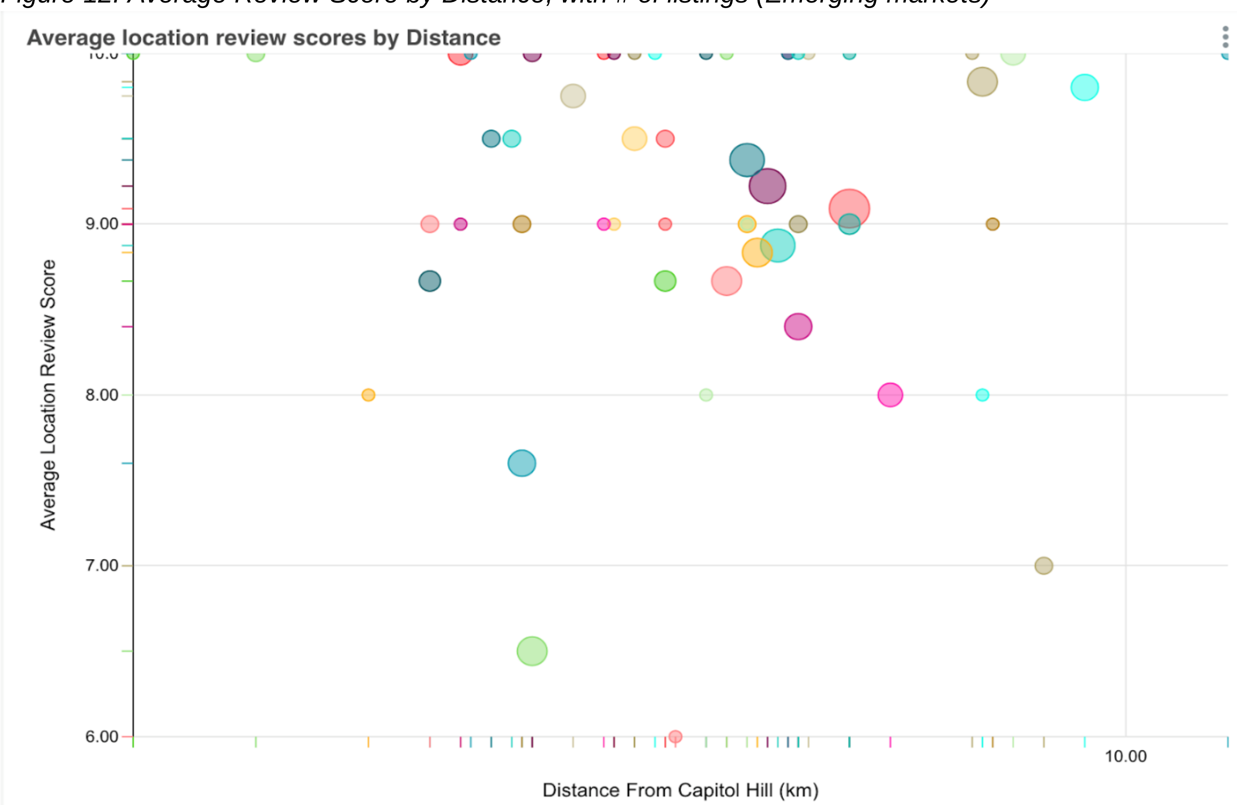


Figure 12: Average Review Score by Distance, with # of listings (Emerging markets)



A correlation coefficient of 0.124887 between average location review and distance from Capitol Hill indicated a very weak linear relationship (Appendix H), implying location review scores are not strongly linked with absolute location.

Given this, and that price reviews were overarchingly positive across price points, expanding geographically into emerging markets is viable. Guest engagement, feedback and spending in these emerging markets would not be fully dependent on their price level or absolute distance from the most established markets. The lack of particular offering types in established markets could then be identified and supplied by emerging markets.

The geographical proximity of these emerging markets would need to be carefully considered in the expansion strategy, since cannibalization is a genuine risk. However, by ensuring product differentiation of offerings in emerging markets as mentioned above, the effects of cannibalization could be minimized. Appendix G contains an example of how such product differentiation could be achieved, through varying the types of properties offered. Differentiation could also be achieved through offering listing features such as business-ready, instant-bookable and other elements not currently satisfied in certain established segments. Airbnb thus can focus on an expansion strategy into these emerging markets, in which there is a range of listings catered to various guests, between the two types of segments, to minimize cannibalization.

2.3 Incorporating Hosts into Possible Expansion Strategy

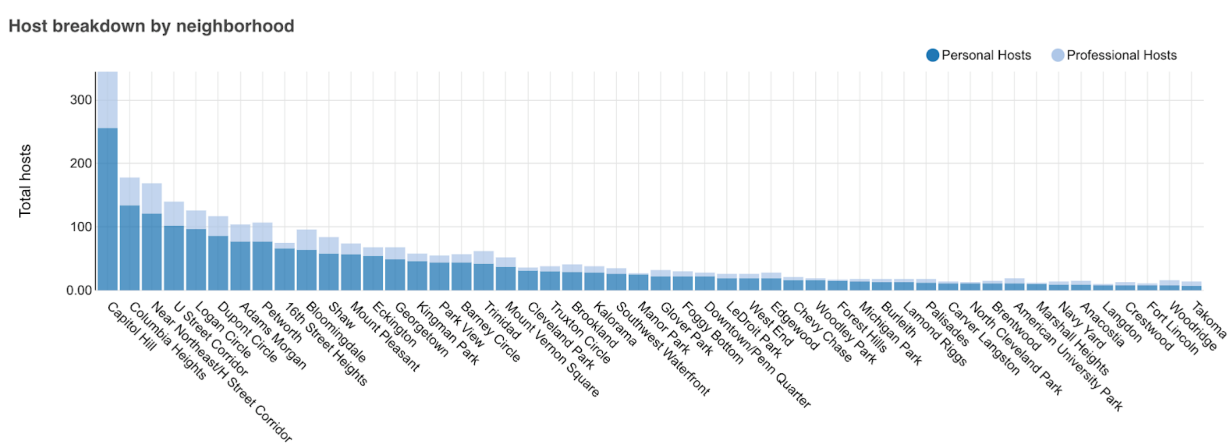
As identified above, the optimal expansion strategy focuses on expansion into emerging markets. However, for Airbnb to best achieve such expansion, they would need to optimize host adaptability, and potentially shift hosts from their most attractive geographic segments into these emerging segments. At the same time, shifting too many hosts could result in failure to maintain competitive advantage and operational efficiency (Statistical Atlas 2019). Hence, also applying segmentation to hosts across both emerging and established markets is imperative.

A frequent criticism of Airbnb is that whilst operating under the guise of local hosts earning supplemental income, professional property management companies or business operators are often key users. “Professional” Airbnb hosts, whose activities feature characteristics of a “business”, are assumed to offer more than single accommodation listings on Airbnb, with a higher likelihood of them also being in violation of most short-term rental laws protecting residential housing (Inside Airbnb, 2017).

The makeup of professional hosts in DC is crucial in determining the extent to which existing hosts can fulfil expansion into emerging suburbs, and the consequent need for new hosts on Airbnb. This hinges on the hypothesis that professional hosts have greater ability to adjust their uptake of properties across specific suburbs within DC. Whilst personal hosts typically list a property they have a direct connection to, thus constricting their ability to branch out into the emerging markets, or rapidly acquire new properties to list.

Inferring a host’s category as professional or personal was based on their listing count, an approach that proved successful in previous studies (DC Working Families, 2017). Across both emerging and expanded, a total of 18.7% of hosts are classified as professional, with the remaining 81.3% labelled personal hosts. A breakdown of host makeup by top 50 neighborhoods is provided below in Figure 13.

Figure 13: Host breakdown by neighborhood (Top 50)



There is currently no neighborhood where professional hosts outweigh personal hosts, and the top 5 neighborhoods contribute a significant proportion of the 18.7% of professionals. The higher proportion of professional hosts in these top 5 may have contributed significantly to the success of these neighborhoods in revenue generation, so incentivizing professional hosts to move away from these areas could cause significant financial issues. Whilst further data is needed to make greater conclusions, it is highly probable that Airbnb would need to consider attracting new hosts to the platform to support an expansion into the emerging neighborhoods.

3. Evaluations and Recommendations

3.1 Summary of Objectives and Insights

It has been identified that within Washington DC, there are both emerging and established markets across neighborhoods.

The analysis has identified that despite strong PPG variation, value reviews from guests are overarchingly positive, and have low correlation with PPG. Location reviews continue the positive trend, and were determined to have a very weak relationship with distance. These two observations justified a focus on emerging markets for expansion, with product differentiation to avoid cannibalization.

The next stage of the analysis evaluated the feasibility of such expansion, given optimal differentiations had been established. It analyzed trends for the categories of hosts – ‘personal’ and ‘professional’, and found it likely that there are not currently enough professional hosts to support expansion. Thus, Airbnb would need to attract new hosts to the platform when supporting an expansion strategy into the emerging markets.

3.2 Exploring Dataset (Open/Structured/Social Media)

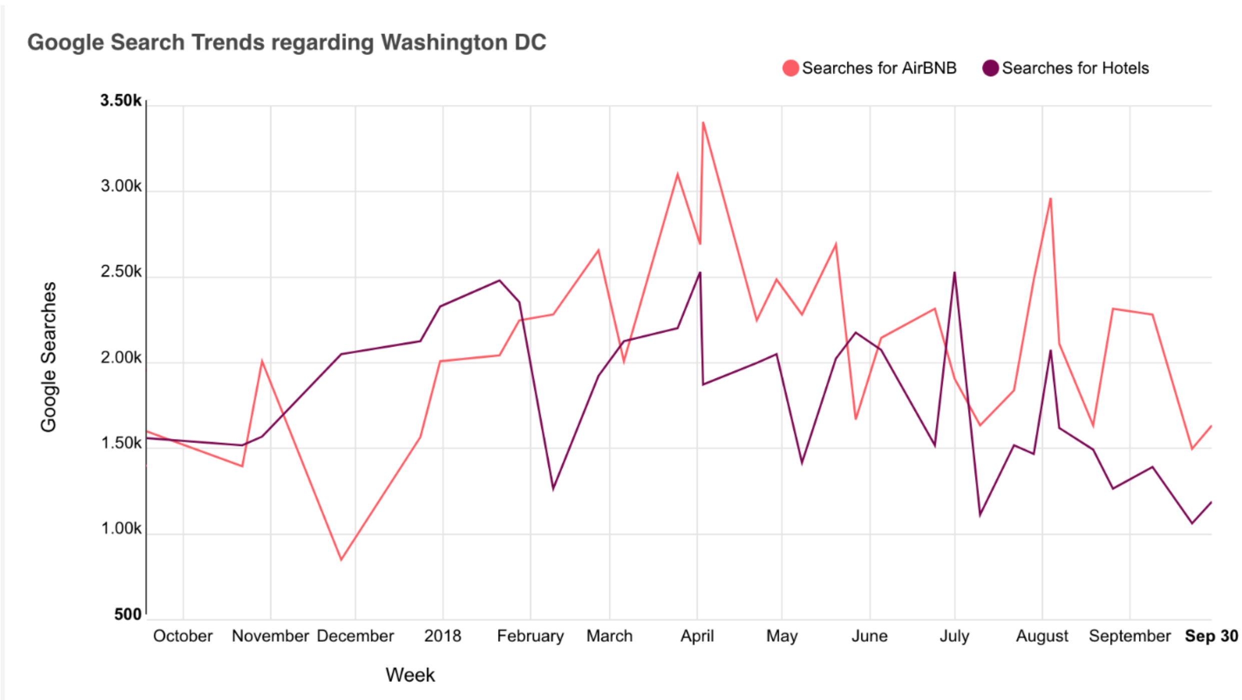
Whilst substantial insights have been derived above, the current dataset is incomplete and can be improved through joining it with **open data**.
Structured and unstructured data can also be used to enable further understanding of Airbnb's positioning within the Washington DC market as a whole.

The original source of the dataset is Inside Airbnb, which scrapes listing properties from Airbnb itself to provide open data. By using Inside Airbnb's original datasets, the initial dataset can be joined with an extended range of attributes, improving the quality and information range of the dataset. An example of a joined dataset in Appendix H shows additional features such as host profile description, their level of verification, and their expectations-from which a more accurate segmentation of whether a host is professional or personal can be made.

Structured data can be leveraged to analyze the external competitive environment of Airbnb. Whilst the initial dataset was limited to Airbnb's hosts in geographic segmentation, the District of Columbia provides location and attributes of all hotels within Washington DC (Opendata 2019), who collectively provide 30,919 rooms (Appendix H). Airbnb could perform geographical analysis of key competitors in the hotel industry- identifying neighborhoods with lower hotel density, the available hotel rooms per neighborhood, and the split of Airbnb to hotel rooms. This would facilitate greater understanding of its positioning within Washington D.C.

Social media data, a key type of unstructured data, has enabled implicit real-time inference of consumer opinions, trends, and behaviors to gain qualitative understanding of consumer feedback (Xu, Y. et al. 2016). Airbnb can utilize Twitter through the scraping framework provided (Appendix H), to gather tweets relating to itself and the hotel industry in Washington DC. Tweets can be filtered on keywords, location posted and dates, from which sentiment analysis can be applied. This would enable benchmarking and identification of guest sentiment trends over time.

The limitation of such Twitter data is that it only captures user-generated content. To understand user actions, a combination of Google Trends & Words Everywhere has been used in Appendix H, to determine absolute search count from Google Trends for Airbnbs and hotels in Washington DC. Figure 11 shows their tracking in the year prior to the capture of the dataset, from which Airbnb can further understand its positioning relative to the traditional hotel industry.



Word Count: 2965

4. References

Airbnb. (2019). *How do star ratings work?*. [online] Available at: https://www.Airbnb.com.au/help/article/1257/how-do-star-ratings-work [Accessed 1 Apr. 2019]

Airdna. *Shares of Full Time Airbnb Operators and their Revenue.* [online] Available at: https://i.pinimg.com/originals/83/c2/72/83c272260603e8bead24cb83b550714e.png [Accessed 31 Mar. 2019]

Areavibes (2019). *Anacostia, DC Crime Rates & Crime Map*. [online] Available at: https://www.areavibes.com/washington-dc/anacostia/crime/ [Accessed 29 Mar. 2019].

CBRE (2017), *Hosts with Multiple Units – A Key Driver of Airbnb Growth.* [ebook] Available at: https://www.ahla.com/sites/default/files/CBRE_AirbnbStudy_2017.pdf [Accessed 1 Apr. 2019]

Collins, M & Parsa, H.G. (2006). *Pricing strategies to maximize revenues in the lodging industry*. [ebook] Hospitality Management. Available at: https://pdfs.semanticscholar.org/b084/5675c228988d647976da5ae7b81df63d8d17.pdf [Accessed 29 Mar. 2019]

Dc Atlas (2019). *DC Crime Cards*. [online] Available at: https://dcatlas.dcgis.dc.gov/crimecards/all:crimes/all:weapons/1:year%20to%20date/citywide:heat [Accessed 30 Mar. 2019].

Fradkin, A., Grewal, E. and Holtz, D. (2018). *The Determinants of Online Review Informativeness: Evidence from Field Experiments on Airbnb*. [ebook] MIT Sloan School of Management. Available at: https://andreyfradkin.com/assets/reviews_paper.pdf [Accessed 28 Mar. 2019].

Jet, J. (2017). *Are Business Travelers Using Airbnb.* [online] Available at: https://www.forbes.com/sites/johnnyjet/2017/08/22/are-business-travelers-using-Airbnb/#615975a44ddf [Accessed 31 Mar. 2019].

Ke, Q. (2017). *Sharing Means Renting?: An Entire-marketplace Analysis of Airbnb.* [ebook] Indiana University, Bloomington. Available at: https://arxiv.org/pdf/1701.01645.pdf [Accessed 30 Mar. 2019]

Kwow, L & Xie, K. (2018). *Pricing strategies on Airbnb: Are multi-unit host revenue pros?.* [online] International Journal of Hospitality Management. Available at: https://www.researchgate.net/publication/327954155_Pricing_strategies_on_Airbnb_Are_multi-unit_host_revenue_pros [Accessed 29 Mar. 2019]

Learn Airbnb. (2016). *The State of Airbnb Hosting.* [ebook] Available at: https://learnAirbnb.com/wp-content/uploads/2017/08/LearnAirbnb.com-Airbnb-Home-Sharing-Report-v1.4.pdf [Accessed 1 Apr. 2019]

Metropolitan Police Department (2017). *2017 Annual Report* **|** *Metropolitan Police Department*. [online] Available at: https://mpdc.dc.gov/sites/default/files/dc/sites/mpdc/publication/attachments/MPD%20Annual%20Report%202017_lowres.pdf [Accessed 1 Apr. 2019].

Opendata (2019). *Hotels*. [online] Available at: http://opendata.dc.gov/datasets/hotels [Accessed 28 Mar. 2019].

Priceonomics. (2017). *The Rise of the Professional Airbnb Investor*. [online] Available at: https://priceonomics.com/will-real-estate-investors-take-over-Airbnb/ [Accessed 27 Mar. 2019].

Samaan, R. (2015). *Airbnb, rising rent, and the housing crisis in Los Angeles.* [ebook] Available at: https://www.ftc.gov/system/files/documents/public_comments/2015/05/01166-96023.pdf [Accessed 29 Mar. 2019]

Statistical Atlas. (2019). *Household Income in Anacostia, Washington, District of Columbia*. [online] Available at: https://statisticalatlas.com/neighborhood/District-of-Columbia/Washington/Anacostia/Household-Incomehttps://statisticalatlas.com/neighborhood/District-of-Columbia/Washington/Anacostia/Household-Income [Accessed 1 Apr. 2019].

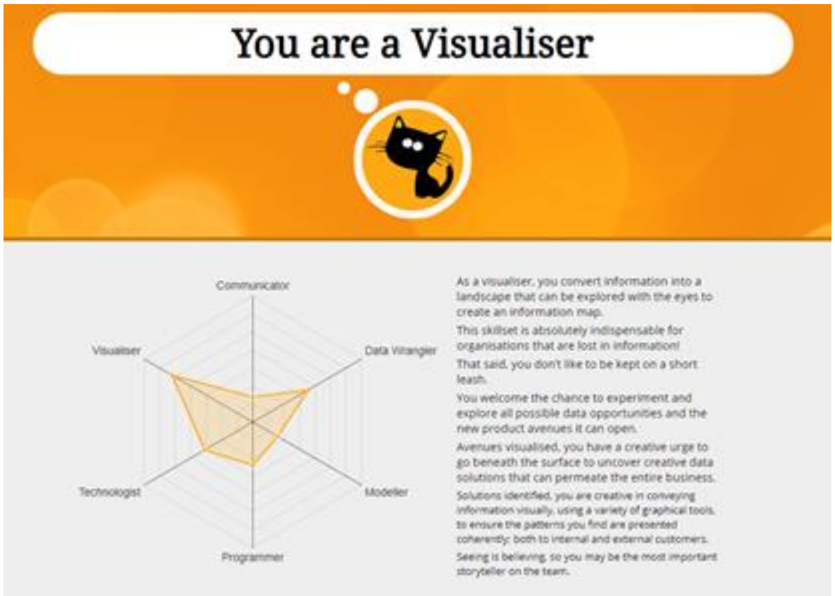
Working Families. (2017). *Selling the District Short.* [ebook] D.C. Working Families. Available at: http://dcsharebetter.org/wp-content/uploads/2017/03/D.C.-Housing-Report_Web.pdf [Accessed 2 Apr. 2019]

Xu, Y., Zhou, D., and Lawless, S. (2019). [ebook] *Inferring Your Expertise from Twitter: Integrating Sentiment and Topic Relatedness*, p.https://www.scss.tcd.ie/seamus.lawless/papers/WI-2016.pdf. Available at: https://www.scss.tcd.ie/seamus.lawless/papers/WI-2016.pdf [Accessed 29 Mar. 2019].

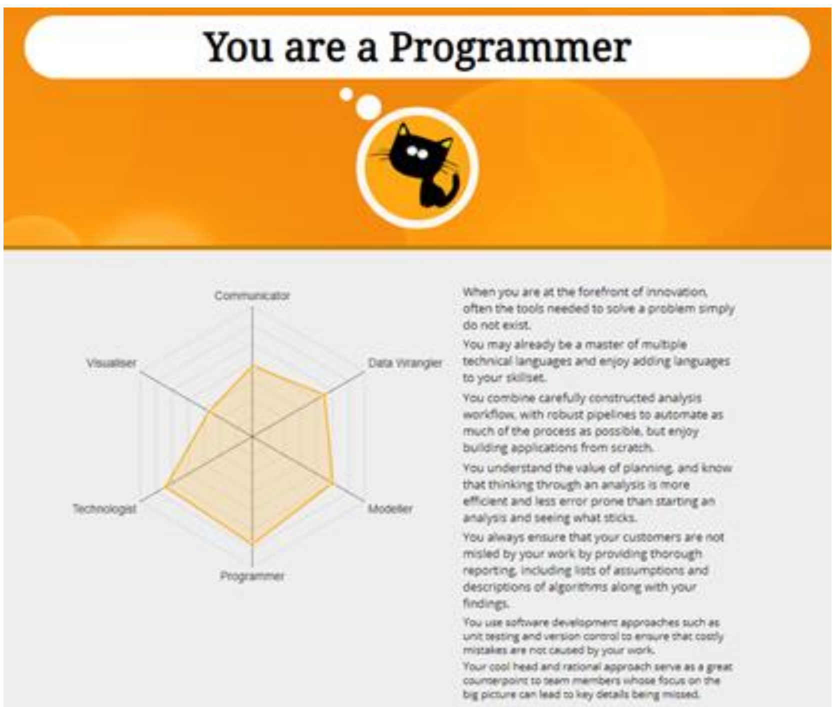
5. Appendix

Appendix A: Data Science Profiles

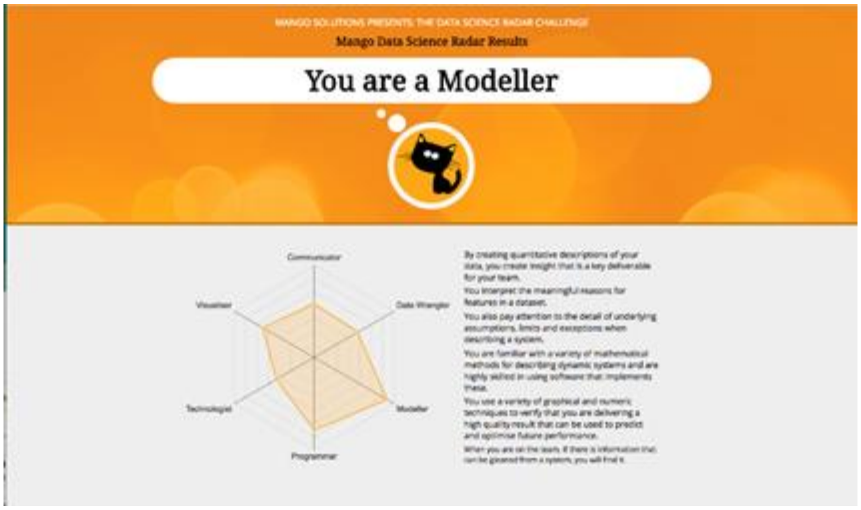
Andrew:



Anderson:



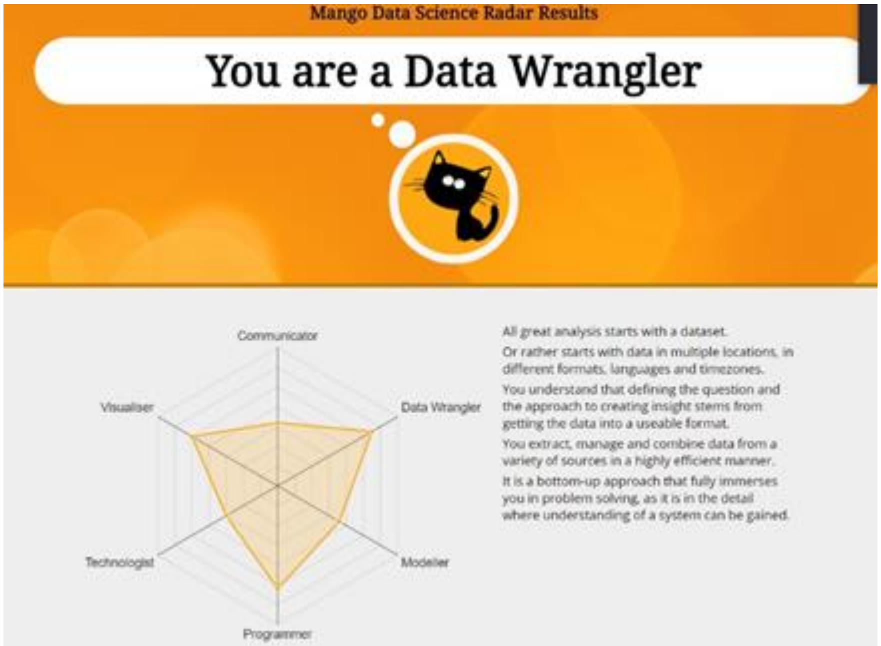
Max:


MANGO SOLUTIONS PRESENTS: THE DATA SCIENCE RADAR CHALLENGE
Mango Data Science Radar Results
You are a Modeller
Communicator
Data Wrangler
Modeller
Programmer
Technologist
Visualiser
By creating quantitative descriptions of your data, you create insight that is a key deliverable for your team.
You interpret the meaningful reasons for features in a dataset.
You also pay attention to the detail of underlying assumptions, limits and exceptions when describing a system.
You are familiar with a variety of mathematical methods for describing dynamic systems and are highly skilled in using software that implements these.
You use a variety of graphical and numeric techniques to verify that you are delivering a high quality result that can be used to predict and optimise future performance.
When you are on the team, if there is information that can be gleaned from a system, you will find it.

Prerita:

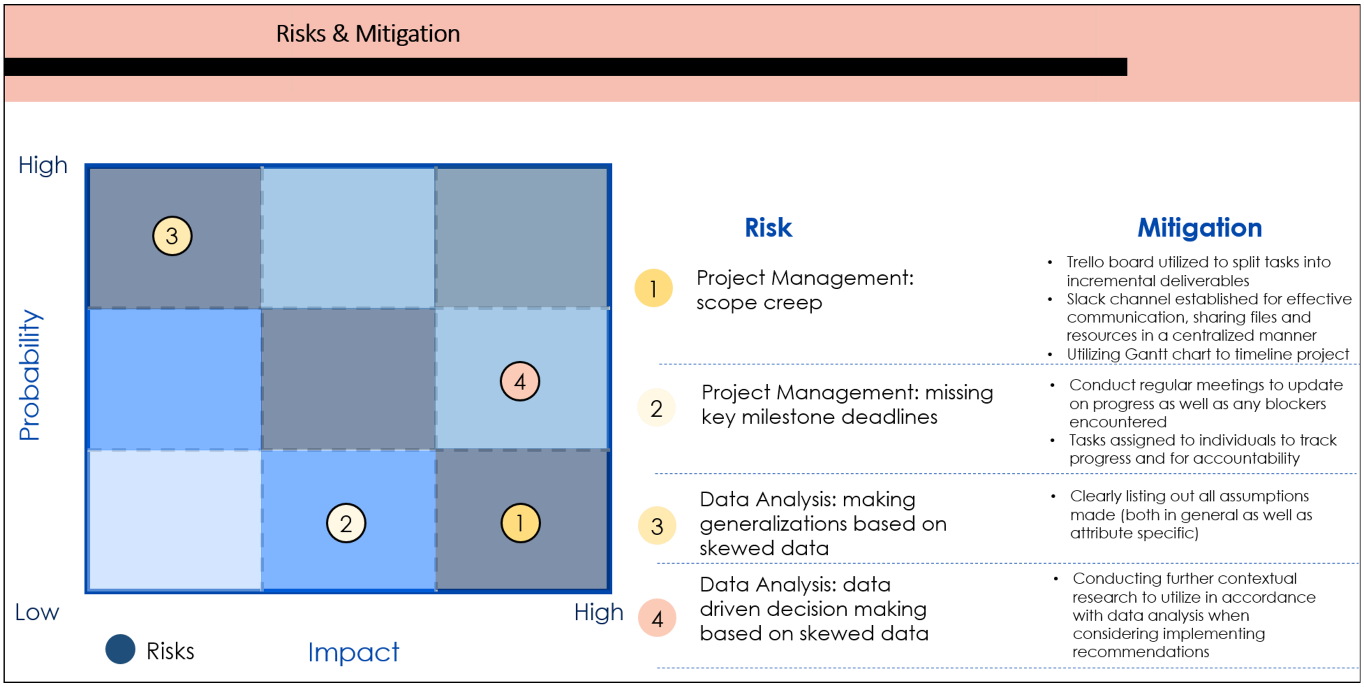

You are a Communicator
Communicator
Data Wrangler
Modeller
Programmer
Technologist
Visualiser
All great data scientists are master communicators.
After all, data doesn't sell itself; it needs a communicator to guide the way.
You are able to lead key business decision makers into an ongoing conversation with data, rather than carrying out ad hoc analyses.
You have a natural ability to communicate complex technical details to non-technical audiences.
You also have an understanding of the wider implications of the project, and so convey the key analysis insights to influence new business directions.
As a communicator you also understand that communication is a two-fold process: explain well, listen well.
You listen for business challenges, define requirements and clarify how data analysis can help.
Even on projects using highly technical software and mathematical methods, you are able to speak a language that each of your stakeholders understands!

Noumik:



Appendix B: Gantt Chart

Appendix C: Risk Mitigation Matrix



Appendix D: Assumptions

- The wards will consistently stay the same – wards 1, 2 and 6 will remain the most concentrated according to our implementation plan (one year horizon)
- Consumer trends are still stable – they still have a preference for Airbnb, (privacy) over hotels and we do not expect that to change
- The aim of growth is still specific to Washington DC and we do not expect other cities to take over DC in terms of rental focus
- Hosts are still available next year (they do not delist)
- Hosts willingness to move is accurately measured
- Cities are not expected to expand
- This is a limited data set, and more data is required in order to gain a holistic understanding of the Airbnb environment, which would in turn inform more accurate recommendations
- The original dataset has a historical trawl as of 12th October 2018, which is the basis of the analytical findings which may not be an accurate reflection of real time conditions
- The review scores are ratings and are subjected to noise and bias

Appendix E: Calculations and Logic

PPG (Price per Guest) - Price of listing / accommodates (number of guests) - to scale metrics based on price which may be misleading if there is a skew of properties accommodating only 1 person.

Inferred revenue was determined through the following equation:

(1/0.72 * number of reviews) * (3 * price)

This is based on a metric of 72% of guests leaving reviews (Fradkin, A., Grewal, E. and Holtz, D. 2018). The average stay in the Airbnb has been evaluated to be roughly 3 nights (Learn Airbnb 2016). There is a variety of assumptions underlying this analysis – namely that the price does not typically change, which is unlikely to hold up in real life. However, the metric aims to be directional and assumes that price changes will occur across groupings of properties.

Appendix F: Risks & Limitations

There were inherent limitations identified within the initial dataset used for exploration. By conducting extensive contextual research for information, it was inferred that the original dataset originated from Inside Airbnb and was later modified to include approximately 4300 out of the 8000+ existing listings. Furthermore, government research clearly highlights there are more listings in Washington D.C. than in the original dataset (DC Working Families, 2017), with the total listings count for individual hosts not matching the number of specific listings in the dataset. This emphasizes the amount of missing data with which analysis was derived. This could have a significant result on the findings, particularly in terms of validity, and raises the question of whether the analysis truly painted a holistic picture of the Airbnb environment in Washington D.C. Validity in particular, could be compromised, given that key insights as well as subsequent evaluations and recommendations were all derived from a relatively small sample size of data.

Moreover, the dataset appears to have a historical trawl as of October 12, 2018. It is consequently important to note that analytical findings are based off the specified timeframe, coupled with over half the data excluded in the original dataset and may not entirely reflect the current Airbnb climate in Washington D.C. Thus, whilst the following recommendations are informed by vast amounts of data analysis, actually implementing these suggestions must be informed by both data as well as intuition and experience.

Appendix G:

Example of Cannibalism (Section 2.2 – Risk to Analysis)

Capitol Hill has 534 listings and generates the most revenue of any established markets. Lincoln Park, by comparison, has just 1 listing in the dataset, and is 0.6 miles away. Given that apartments, houses and townhouses make up 82.3% of all listings within Capitol Hill, if Lincoln Park's makeup was focused largely on serviced apartments, B&Bs and condominiums, potential guests indifferent between the locations could purchase their desired listing types (B&Bs and condos) within Lincoln Park.

Appendix H:

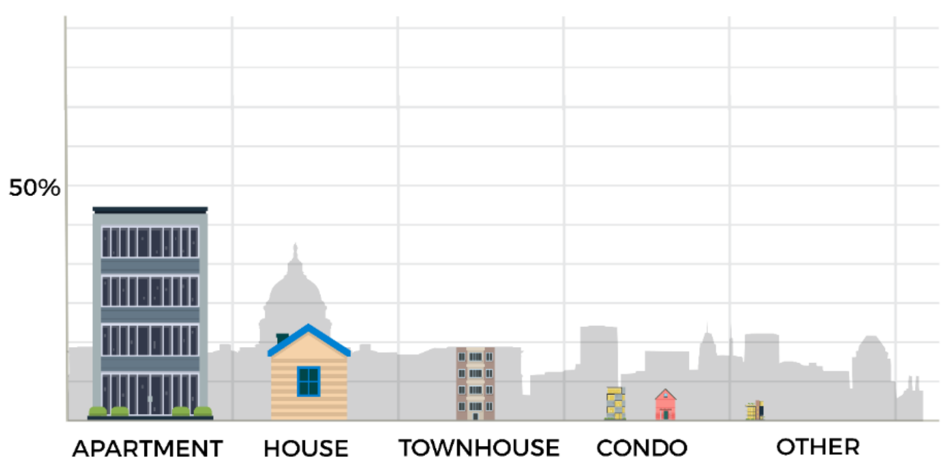
https://github.com/max-pedersen/infs3603project/ contains the relevant files relating to the initial data approach that was taken (see Dataset exploration & workings.ipynb). It also contains folders with samples of the open, structured and unstructured data that was mentioned.

https://github.com/max-pedersen/infs3603project/tree/master/extendingdata - Contains hotel data, and contains Google Trends & Words Everywhere data, to infer search counts.

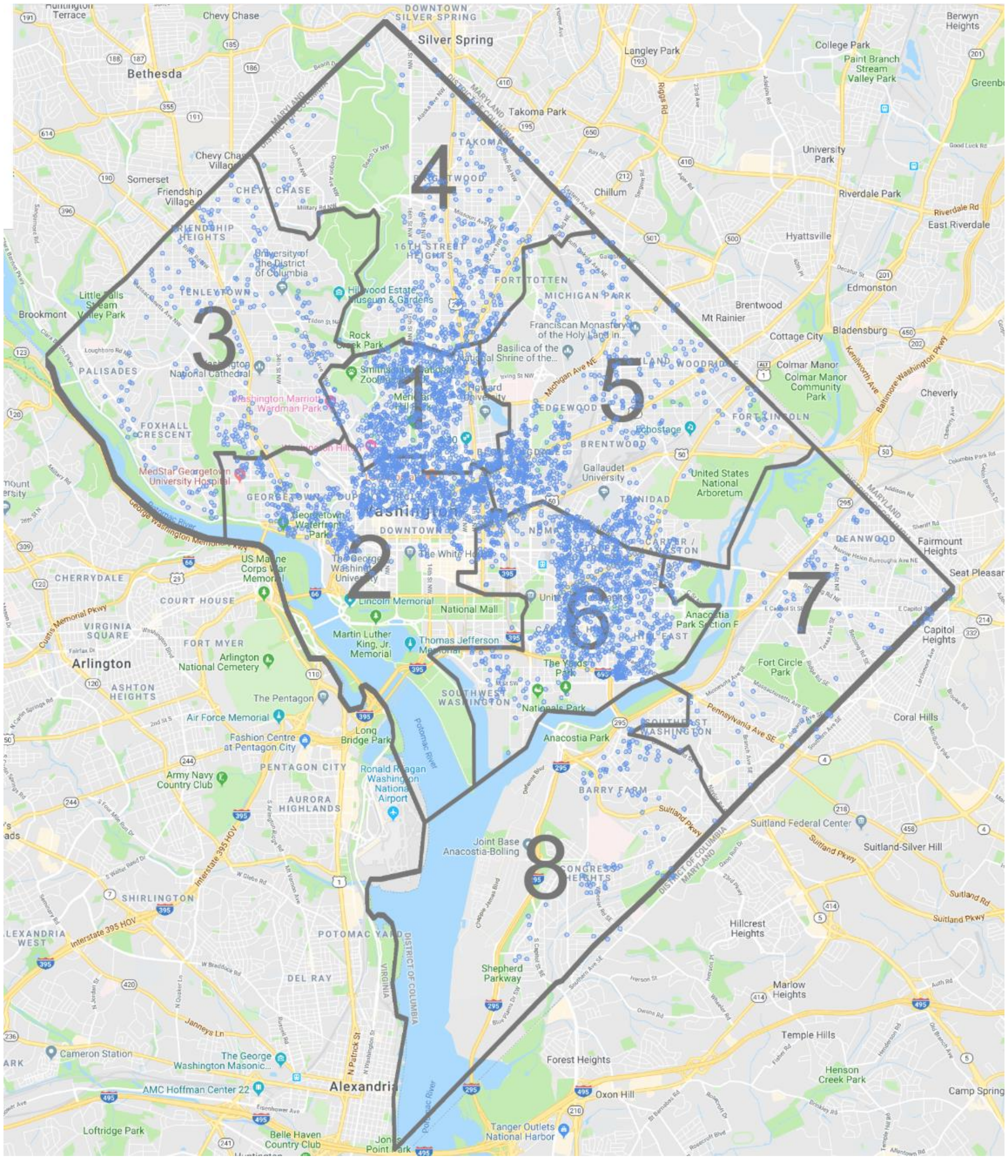
https://github.com/max-pedersen/infs3603project/tree/master/GetOldTweets-python-
Scraping module for Twitter, and examples of scraped Airbnb/hotel tweets

https://github.com/max-pedersen/infs3603project/tree/master/open-data-usage-inside-airbnb- Contains example of joined data with the open data source from Inside Airbnb

Appendix I:

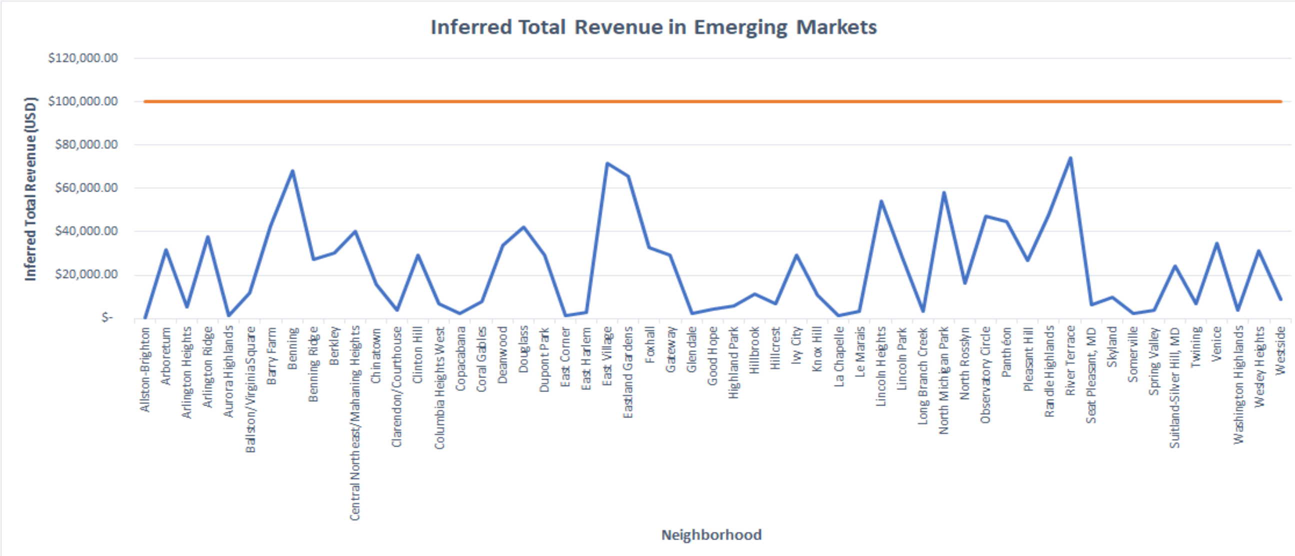


Appendix J:

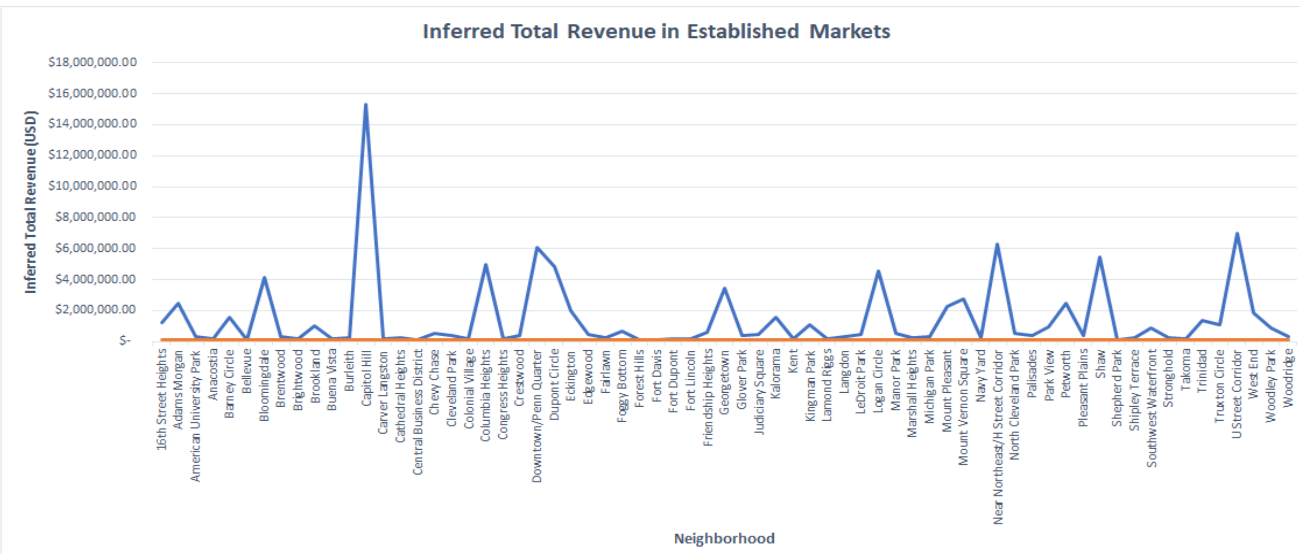


1
2
3
4
5
6
7
8
Silver Spring
Bethesda
Chevy Chase
Takoma Park
College Park
University Park
Riverdale Park
Hyattsville
Brentwood
Mt Rainier
Cottage City
Bladensburg
Cheverly
Arlington
Alexandria
Washington
The Pentagon
Arlington National Cemetery
Lincoln Memorial
National Mall
Thomas Jefferson Memorial
United States National Arboretum
Gallaudet University
Anacostia Park
Joint Base Anacostia-Bolling
Ronald Reagan Washington National Airport
Potomac River
Forest Heights
Oxon Hill
Suitland-Silver Hill
Capitol Heights
Coral Hills
Marlow Heights
Temple Hills
Hillcrest Heights
Fairmount Heights
Seat Pleasant
Camp Springs

Appendix K:

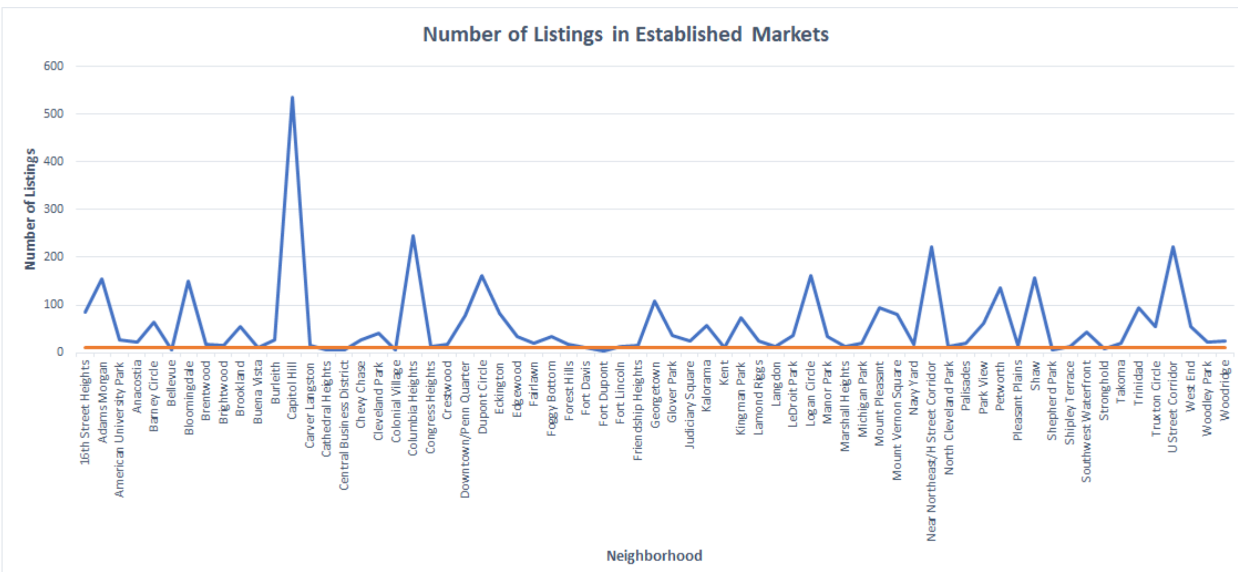


Inferred Total Revenue in Emerging Markets
$120,000.00
$100,000.00
$80,000.00
$60,000.00
$40,000.00
$20,000.00
$-
Inferred Total Revenue (USD)
Allston-Brighton
Arboretum
Arlington Heights
Arlington Ridge
Aurora Highlands
Ballston/Virginia Square
Barry Farm
Benning
Benning Ridge
Berkley
Central Northeast/Mahaning Heights
Chinatown
Clarendon/Courthouse
Clinton Hill
Columbia Heights West
Copacabana
Coral Gables
Deanwood
Douglass
Dupont Park
East Corner
East Harlem
East Village
Eastland Gardens
Foxhall
Gateway
Glendale
Good Hope
Highland Park
Hillbrook
Hillcrest
Ivy City
Knox Hill
La Chapelle
Le Marais
Lincoln Heights
Lincoln Park
Long Branch Creek
North Michigan Park
North Rosslyn
Observatory Circle
Panthéon
Pleasant Hill
Randle Highlands
River Terrace
Seat Pleasant, MD
Skyland
Somerville
Spring Valley
Suitland-Silver Hill, MD
Twining
Venice
Washington Highlands
Wesley Heights
Westside
Neighborhood



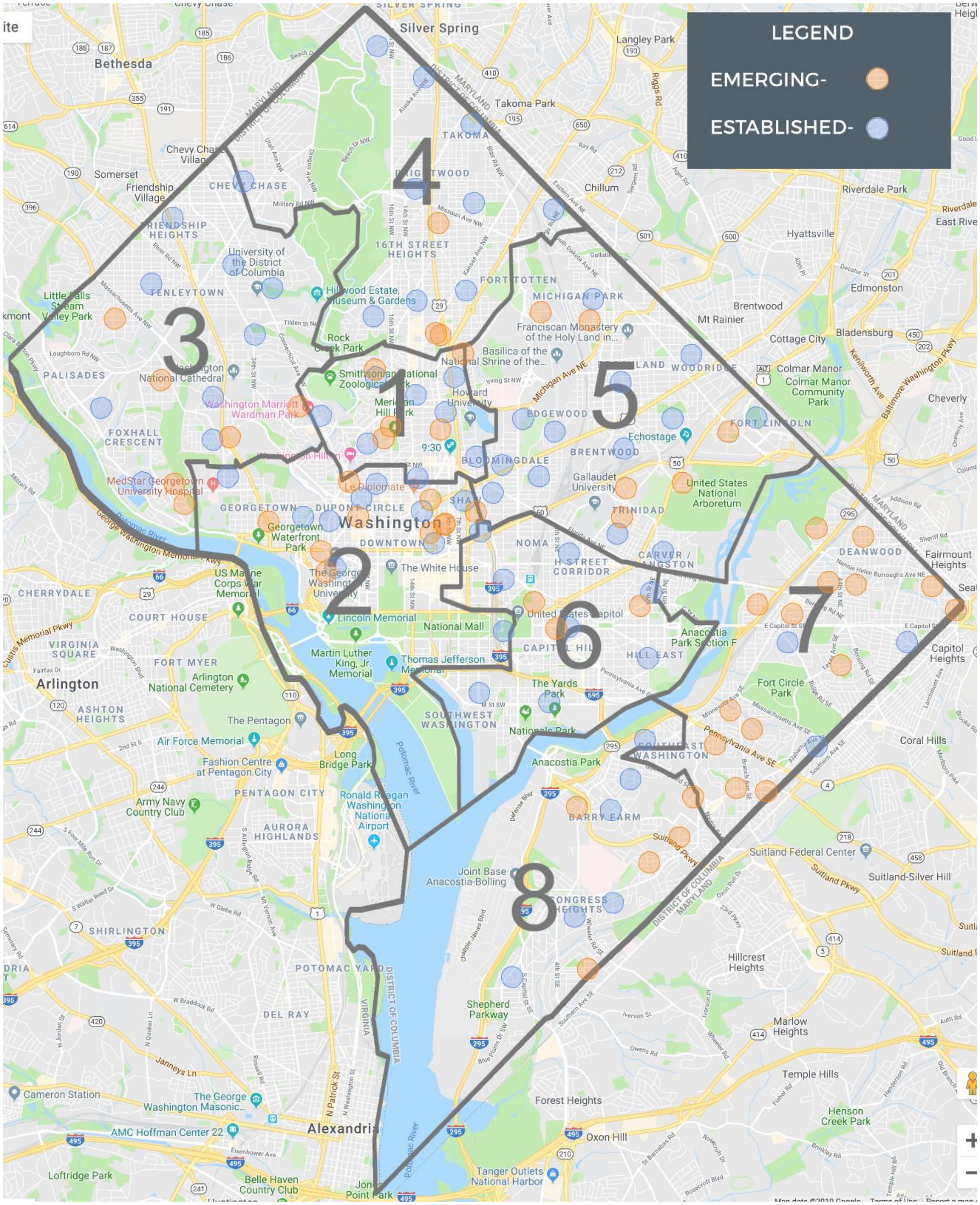
Inferred Total Revenue in Established Markets
$18,000,000.00
$16,000,000.00
$14,000,000.00
$12,000,000.00
$10,000,000.00
$8,000,000.00
$6,000,000.00
$4,000,000.00
$2,000,000.00
$-
Inferred Total Revenue (USD)
16th Street Heights
Adams Morgan
American University Park
Anacostia
Barney Circle
Bellevue
Bloomingdale
Brentwood
Brightwood
Brookland
Buena Vista
Burleith
Capitol Hill
Carver Langston
Cathedral Heights
Central Business District
Chevy Chase
Cleveland Park
Colonial Village
Columbia Heights
Congress Heights
Crestwood
Downtown/Penn Quarter
Dupont Circle
Eckington
Edgewood
Fairlawn
Foggy Bottom
Forest Hills
Fort Davis
Fort Dupont
Fort Lincoln
Friendship Heights
Georgetown
Glover Park
Judiciary Square
Kalorama
Kent
Kingman Park
Lamond Riggs
Langdon
LeDroit Park
Logan Circle
Manor Park
Marshall Heights
Michigan Park
Mount Pleasant
Mount Vernon Square
Navy Yard
Near Northeast/H Street Corridor
North Cleveland Park
Palisades
Park View
Petworth
Pleasant Plains
Shaw
Shepherd Park
Shipley Terrace
Southwest Waterfront
Stronghold
Takoma
Trinidad
Truxton Circle
U Street Corridor
West End
Woodley Park
Woodridge
Neighborhood

Number of Listings in Emerging Markets
Number of Listings
Neighborhod

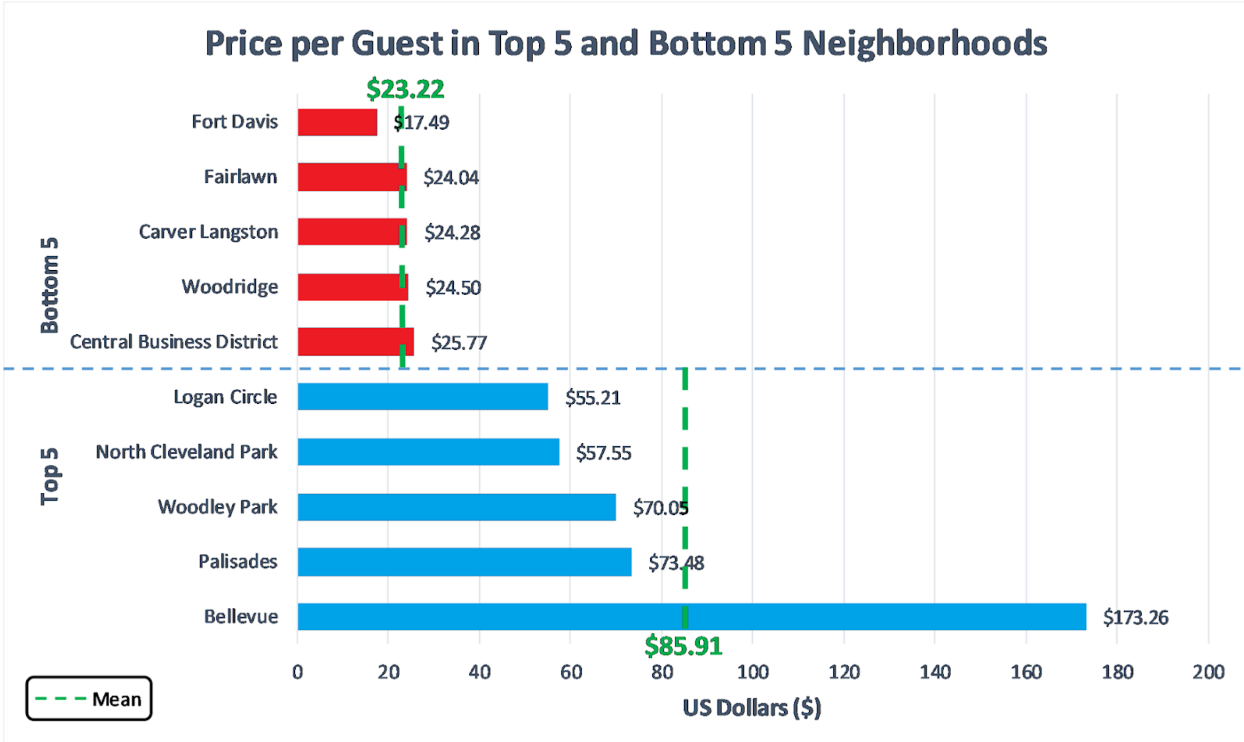


Number of Listings in Established Markets
Number of Listings
Neighborhood

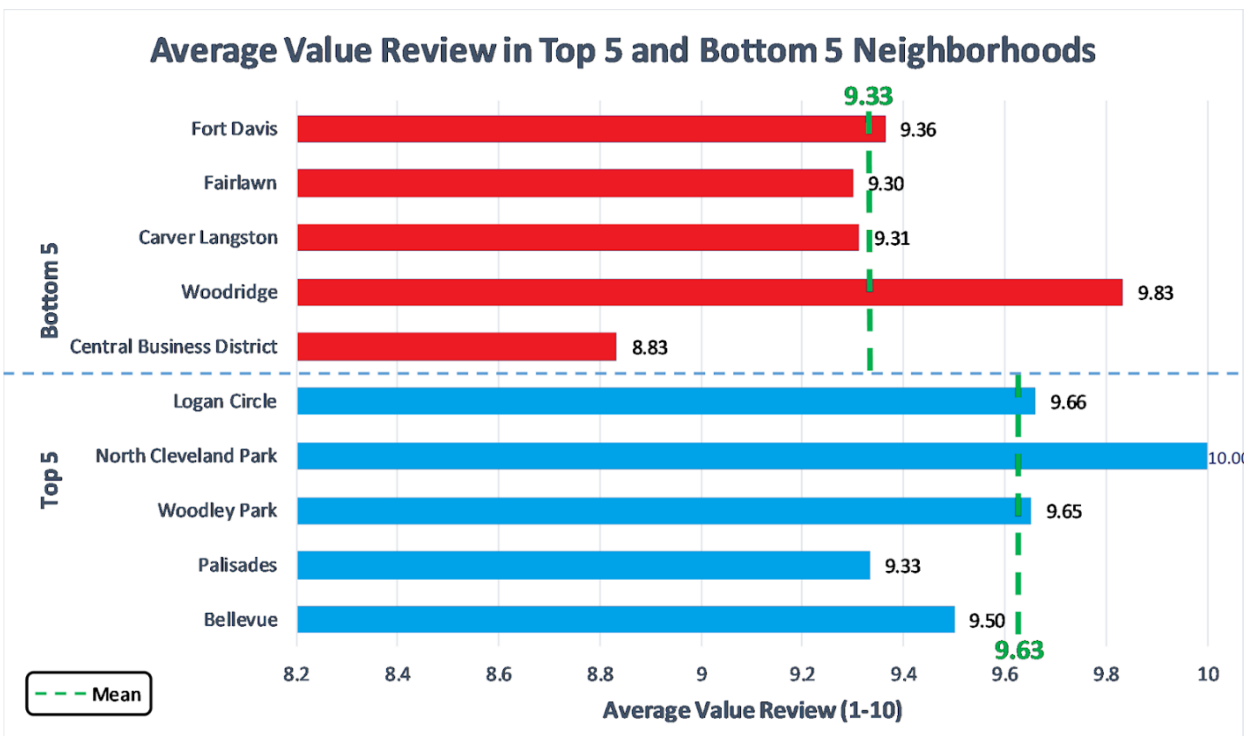
Appendix L:


LEGEND
EMERGING-
ESTABLISHED-
1
2
3
4
5
6
7
8
Washington
Silver Spring
Bethesda
Arlington
Alexandria

Appendix M:



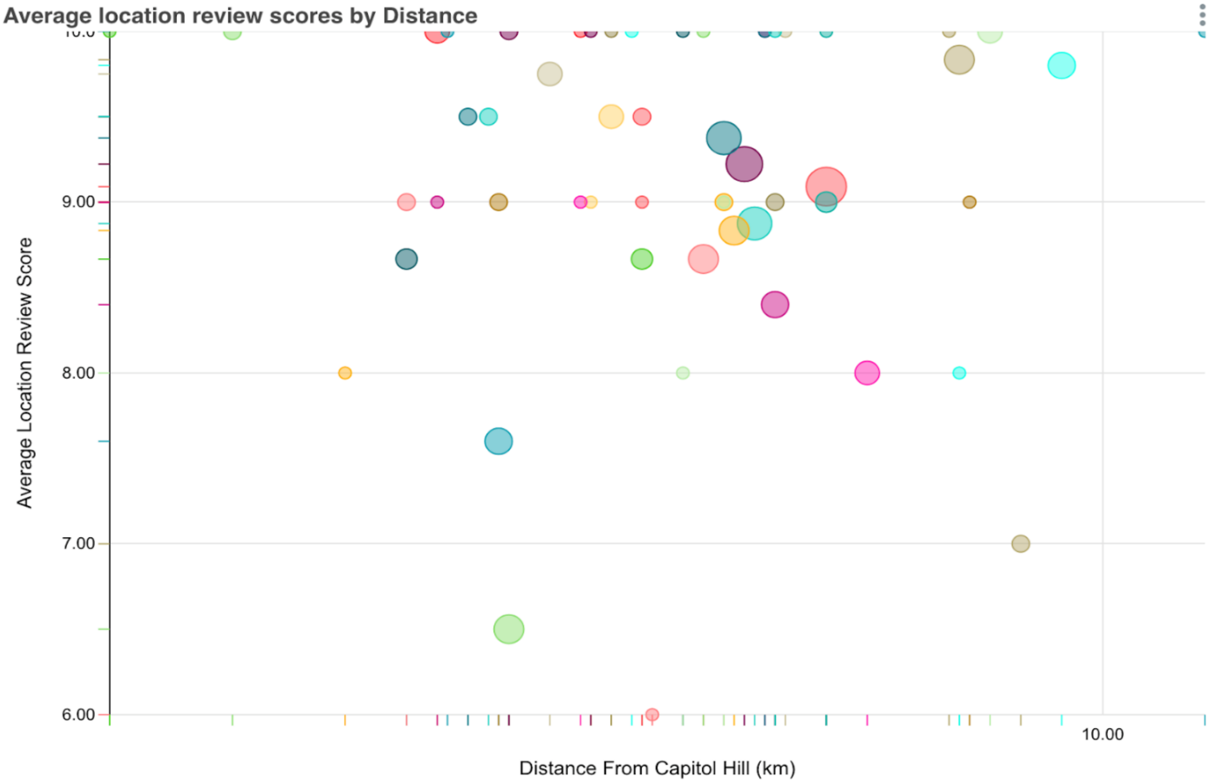
Price per Guest in Top 5 and Bottom 5 Neighborhoods
$23.22
Fort Davis $17.49
Fairlawn $24.04
Carver Langston $24.28
Woodridge $24.50
Central Business District $25.77
Bottom 5
Logan Circle $55.21
North Cleveland Park $57.55
Woodley Park $70.05
Palisades $73.48
Bellevue $173.26
Top 5
$85.91
0 20 40 60 80 100 120 140 160 180 200
US Dollars ($)
Mean

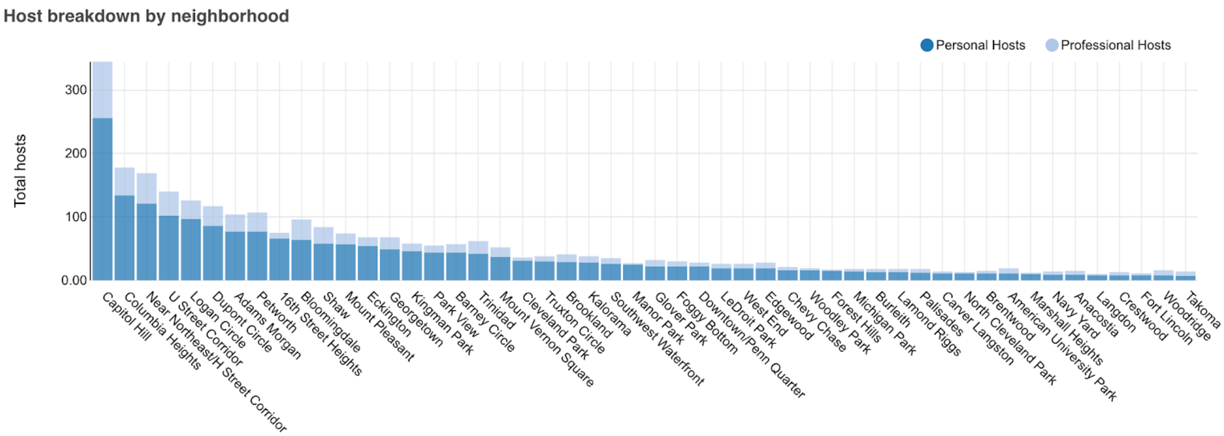


Average Value Review in Top 5 and Bottom 5 Neighborhoods
9.33
Fort Davis 9.36
Fairlawn 9.30
Carver Langston 9.31
Woodridge 9.83
Central Business District 8.83
Bottom 5
Logan Circle 9.66
North Cleveland Park 10.00
Woodley Park 9.65
Palisades 9.33
Bellevue 9.50
Top 5
9.63
8.2 8.4 8.6 8.8 9 9.2 9.4 9.6 9.8 10
Average Value Review (1-10)
Mean

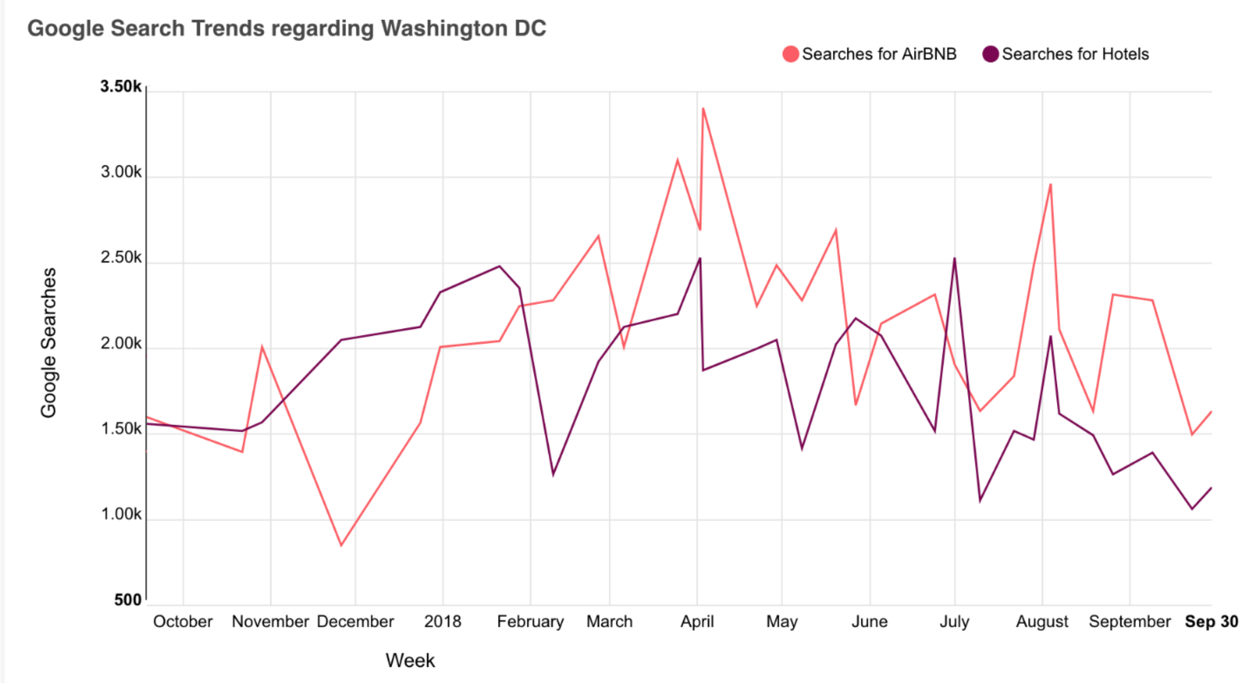
Average location review scores by Distance

Average location review scores by Distance




Host breakdown by neighborhood
Personal Hosts
Professional Hosts
Total hosts
300
200
100
0.00
Capitol Hill
Columbia Heights
Near Northeast/H Street Corridor
U Street Corridor
Logan Circle
Dupont Circle
Adams Morgan
Petworth
16th Street Heights
Bloomingdale
Shaw
Mount Pleasant
Eckington
Georgetown
Kingman Park
Park View
Barney Circle
Trinidad
Mount Vernon Square
Cleveland Park
Truxton Circle
Brookland
Kalorama
Southwest Waterfront
Manor Park
Glover Park
Foggy Bottom
Downtown/Penn Quarter
LeDroit Park
West End
Edgewood
Chevy Chase
Woodley Park
Forest Hills
Michigan Park
Burleith
Lamond Riggs
Palisades
Carver Langston
North Cleveland Park
Brentwood
American University Park
Marshall Heights
Navy Yard
Anacostia
Langdon
Crestwood
Fort Lincoln
Woodridge
Takoma

Appendix N:


Google Search Trends regarding Washington DC
Searches for AirBNB
Searches for Hotels
Google Searches
3.50k
3.00k
2.50k
2.00k
1.50k
1.00k
500
October
November
December
2018
February
March
April
May
June
July
August
September
Sep 30
Week