Project 1, “Data Analysis” at General Assembly, Fall 2018

By Andrew Percival

ANALYZING THE VIABILITY OF AIR BNB IN THE WASHINGTON DC MARKET

THE QUESTION

- *Should our investor invest in an Airbnb hotel in Washington, D.C.? If so, in which neighborhood should they invest?*

  - Prompt 1: Host revenue — How much revenue do successful hosts generate?
  - Prompt 2: Property reviews — Which property types receive the most positive reviews?
  - Prompt 3: Neighborhood popularity — Which neighborhoods host the most listings?
  - Prompt 4: Neighborhood sentiment — Which neighborhoods receive the most positive reviews?

DATA CLEANING

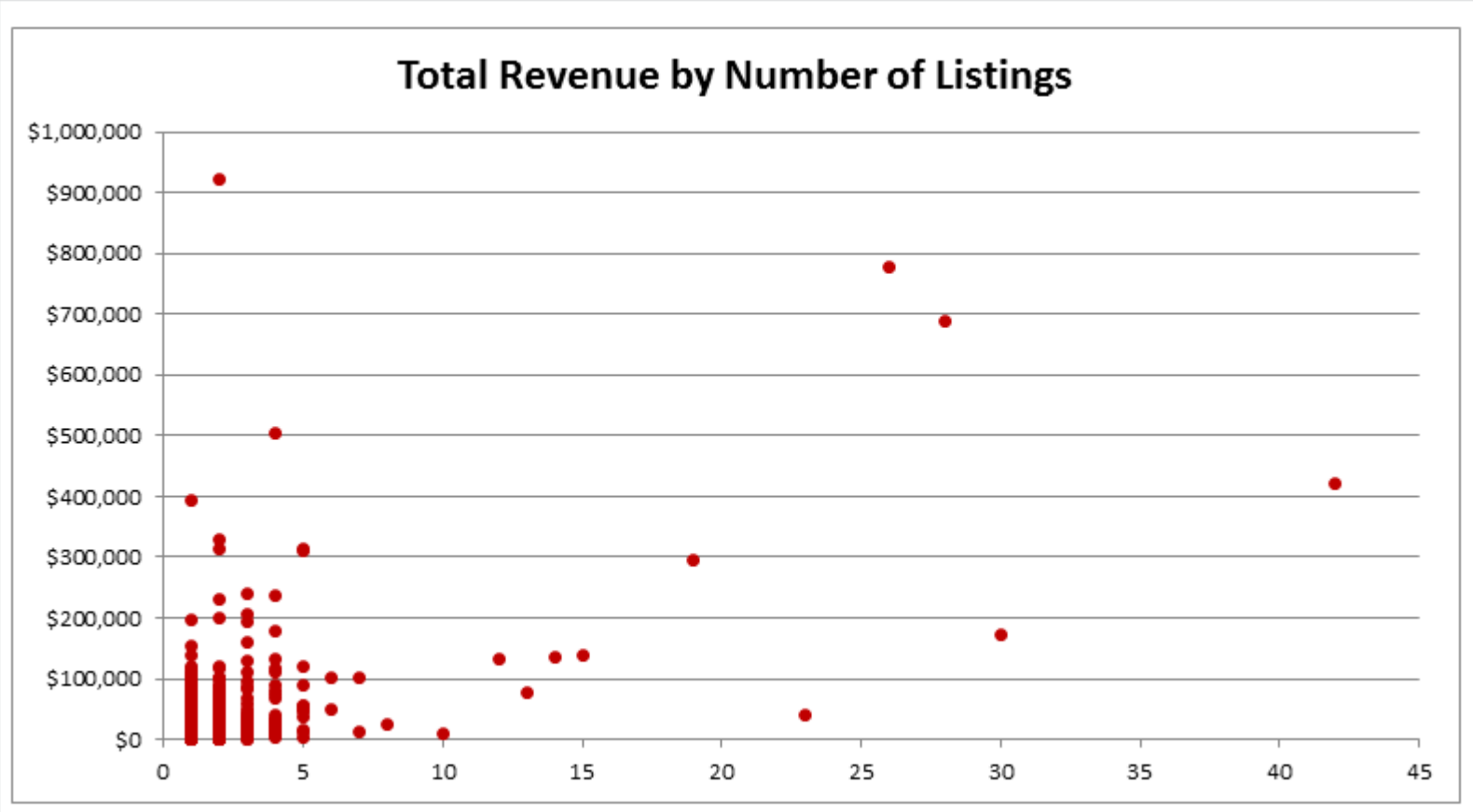
Data Cleaning Steps

Copied data tab, all data from here on out derived from "Project 1_DC AirBnB (2)"
Deleted column AL titled "neighbourhood_group_cleansed" because it had no data in it (dangerous blank column)
Sorted by Reviews, deleted all with 0 (830 total, took 3723 listings to 2893 listings...cut out 22.3% of the rows)
Inserted a column in B, entered function searching for current row ID in any rows above it, Ex "COUNTIF(A$2:A3,A4)", Paste-Special those results
Filtered new data and "0" was the only value displayed, therefore there are no duplicate IDs
Deleted all qualitative columns that will not help with the analysis (based on question "Are descriptions of the property and summaries of the neighborhood going to help your analysis?")
Standardized entries for state & city (were about 5 properties inside DC city limits, but without a simple "Washington" entry in city and "DC" entry in state)
    Were about 5 properties inside DC city limits, but without a simple "Washington" entry in city and "DC" entry in state)
    Were a handful of neighborhoods that needed to be standardized, i.e. "Mt Vernon Square" > "Mount Vernon Square"
    Noticed there were also a few mismarked "smart_location" entries, corrected those
Standardized entries for zip code (were a handful that were "20002-1234" types
Standardized Longitude & Latitude columns by formatting all as Numbers w/ 10 decimal places
Added columns CG:CJ titled "Guests_Adj", "Days_Adj", "BookingsAdj", "Est_Daily_Revenue", "EstRevenuePerBooking", "Est_Total_Revenue" and entered the formulas as indicated in the instructions

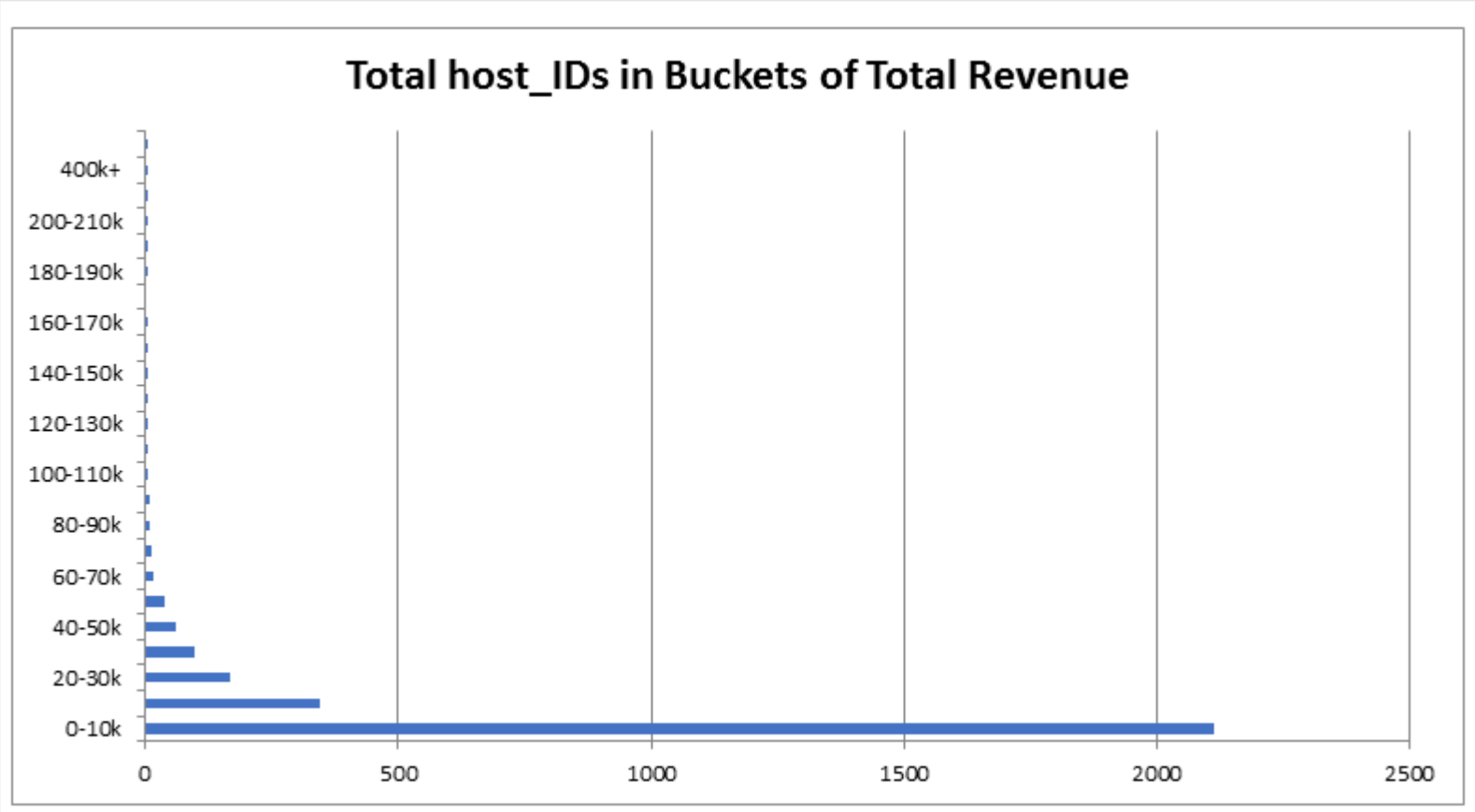
PROMPT 1: HOST REVENUE

- How much revenue do successful hosts generate?
  - The *most* successful hosts are capable of generating over 100k total *between all their properties*, but this is rare
    - Only 25 host_IDs (0.9% of all) generated more than 100k total
    - Only 109 host_IDs (3.8% of all) generated more than 50k total
    - 75th percentile total revenue was $8,700
  - Rarer still is the ability to generate over 100k *per property*
    - Only 16 host_IDs (0.6% of all) generated more than 100k per property
    - Only 72 host_IDs (2.5% of all) generated more than 50k per property
    - 75th percentile total revenue per property was $7,112
  - The average host_ID generated $14,505 of total revenue, $9,905 per property, but that was driven by a couple of large outliers
  - The *median* host_ID generated $3,876 of total revenue, $3,360 per property

PROMPT 1: HOST REVENUE



PROMPT 1: HOST REVENUE



PROMPT 1: HOST REVENUE (CIRCLING BACK)

- Who is the single-most lucrative host?
  - Defined by both Sum of Estimated Total Revenue and Estimated Total Revenue Per Listing, the most lucrative is Host_ID 180061…
  - *According to the data as we were instructed to manipulate it, they have hosted 416 separate bookings, each at a minimum stay of 28 days…that equates to 11,648 days of hosting guests (that's almost 32 years!)…the series of assumptions we made is likely incorrect…reckless even*

Verified info

Hey, I'm Hermosa!

Washington, District of Columbia, United States · Joined in July 2010

Report this user

We love to travel and our dogs. We like good food and good TV shows. We take spontaneous trips and prefer flexible travel plans. We like to make the most of our trips so we are usually out all day while on vacation.

332 Reviews 1 Reference Verified

Reviews (332)

PROMPT 2: PROPERTY

- Which property types receive the most positive reviews?
  - Technically, the Bungalow and Cabin property type received the best ratings, but that is a low sample (1!)
  - Condominium & Townhouse each garnered a decent sample of ratings and saw average scores above 95
  - The two “big” categories were very close in average rating…House (93.6) and Apartment (93.3)
  - The only real outliers were Other (84.5) and Dorm (60.0), though they too were low sample size
  - *In my opinion, there is nothing here…there’s nothing indicating that one property type is liable to receive more positive reviews than another*

PROMPT 2: PROPERTY

PROMPT 3: NEIGHBORHOOD POPULARITY

- Which neighborhoods host the most listings?

Neighborhoods By Listings

Host_ID	Listings
Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View	357
Dupont Circle, Connecticut Avenue/K Street	290
Capitol Hill, Lincoln Park	245
Shaw, Logan Circle	243
Union Station, Stanton Park, Kingman Park	238
Edgewood, Bloomingdale, Truxton Circle, Eckington	198
Kalorama Heights, Adams Morgan, Lanier Heights	185
Brightwood Park, Crestwood, Petworth	140
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street	139
Howard University, Le Droit Park, Cardozo/Shaw	116

Zip Codes By Listings

Zip Code	Listings
20009	562
20001	435
20002	405
20003	222
20010	215
20011	154
20007	134
20005	130
20037	95
20008	89

PROMPT 4: NEIGHBORHOOD SENTIMENT

- Which neighborhoods receive the most positive reviews?

Neighborhoods By Average Rating

Host_ID	Listings	Avg Rating
Woodland/Fort Stanton, Garfield Heights, Knox Hill	3	98.7
Deanwood, Burrville, Grant Park, Lincoln Heights, Fairmont Heights	3	97.7
Cleveland Park, Woodley Park, Massachusetts Avenue Heights, Woodland-Normanstone Terrace	41	97.1
Capitol View, Marshall Heights, Benning Heights	7	96.7
Douglas, Shipley Terrace	6	96.0
Spring Valley, Palisades, Wesley Heights, Foxhall Crescent, Foxhall Village, Georgetown Reservoir	29	95.9
Woodridge, Fort Lincoln, Gateway	14	95.6
Near Southeast, Navy Yard	12	95.3
Capitol Hill, Lincoln Park	245	95.0
Mayfair, Hillbrook, Mahaning Heights	6	95.0
Cathedral Heights, McLean Gardens, Glover Park	45	94.5
Union Station, Stanton Park, Kingman Park	238	94.3
Shaw, Logan Circle	243	94.0
Edgewood, Bloomingdale, Truxton Circle, Eckington	198	93.9
Georgetown, Burleith/Hillandale	87	93.8
Dupont Circle, Connecticut Avenue/K Street	290	93.4
Brightwood Park, Crestwood, Petworth	140	93.4
Kalorama Heights, Adams Morgan, Lanier Heights	185	93.3
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street	139	93.2
Howard University, Le Droit Park, Cardozo/Shaw	116	93.0
West End, Foggy Bottom, GWU	99	93.0
Congress Heights, Bellevue, Washington Highlands	5	92.8
Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View	357	92.8
North Cleveland Park, Forest Hills, Van Ness	23	92.5
Southwest Employment Area, Southwest/Waterfront, Fort McNair, Buzzard Point	63	92.3

Zip Codes by Average Rating

Zip Code	Listings	Avg Rating
20268	1	100.0
20743	1	100.0
20782	2	100.0
20004	3	96.7
20006	13	95.8
20003	222	95.2
20008	89	94.2
20007	134	94.1
20002	405	94.1
20032	8	94.0
20005	130	94.0
20036	54	93.9
20001	435	93.4
20009	562	93.3
20011	154	93.2
22209	1	93.0
20037	95	92.8
20010	215	92.4
20017	60	92.4
20020	58	92.4
20016	48	92.4
20024	62	92.3
20019	28	92.2
20018	36	91.8
20012	35	91.7

*** ADDED PROMPT: NEIGHBORHOOD $ ***

- Which neighborhoods generate the most daily revenue?

Neighborhoods By Average Daily Revenue

Statistically relevant sample size (proven market) that averages over $150 per day

Host_ID	Listings	Avg Daily Revenue
Georgetown, Burleith/Hillandale	87	$225
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street	139	$184
Spring Valley, Palisades, Wesley Heights, Foxhall Crescent, Foxhall Village, Georgetown Reservoir	29	$172
West End, Foggy Bottom, GWU	99	$166
Shaw, Logan Circle	243	$161
Capitol Hill, Lincoln Park	245	$158
Howard University, Le Droit Park, Cardozo/Shaw	116	$151
Dupont Circle, Connecticut Avenue/K Street	290	$143
Cleveland Park, Woodley Park, Massachusetts Avenue Heights, Woodland-Normanstone Terrace	41	$133
Union Station, Stanton Park, Kingman Park	238	$133
Southwest Employment Area, Southwest/Waterfront, Fort McNair, Buzzard Point	63	$131
Kalorama Heights, Adams Morgan, Lanier Heights	185	$129
Friendship Heights, American University Park, Tenleytown	23	$126
Near Southeast, Navy Yard	12	$125
Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View	357	$123
Congress Heights, Bellevue, Washington Highlands	5	$122
Edgewood, Bloomingdale, Truxton Circle, Eckington	198	$118
Brightwood Park, Crestwood, Petworth	140	$114
Ivy City, Arboretum, Trinidad, Carver Langston	52	$111
North Cleveland Park, Forest Hills, Van Ness	23	$111
Hawthorne, Barnaby Woods, Chevy Chase	15	$110
Cathedral Heights, McLean Gardens, Glover Park	45	$109
Brookland, Brentwood, Langdon	44	$108
Sheridan, Barry Farm, Buena Vista	7	$106

Zip Codes by Average Daily Revenue

Zip Code	Listings	Avg Daily Revenue
20268	1	$337
20004	3	$220
20007	134	$191
20052	3	$179
22209	1	$175
20037	95	$174
20005	130	$165
20008	89	$157
20003	222	$157
20036	54	$154
20001	435	$149
20009	562	$134
20016	48	$133
20024	62	$131
20002	405	$128
20010	215	$118
20011	154	$114
20015	21	$112
20006	13	$110
20032	8	$108
20018	36	$101
20017	60	$94
20012	35	$88
20020	58	$83

*** ADDED PROMPT: NEIGHBORHOOD $ ***

- Which neighborhoods generate the least daily revenue?

Neighborhoods By Average Daily Revenue				Zip Codes by Average Daily Revenue		
Statistically relevant sample size (proven market) that averages under $80 per day						
Host_ID	Listings	Avg Daily Revenue		Zip Code	Listings	Avg Daily Revenue
Eastland Gardens, Kenilworth	5	$43		20782	2	$37
Deanwood, Burrville, Grant Park, Lincoln Heights, Fairmont Heights	3	$59		20743	1	$48
Woodland/Fort Stanton, Garfield Heights, Knox Hill	3	$62		20064	1	$50
Capitol View, Marshall Heights, Benning Heights	7	$72		20712	2	$71
Historic Anacostia	24	$73		20019	28	$73
Fairfax Village, Naylor Gardens, Hillcrest, Summit Park	4	$77		20910	4	$74
Douglas, Shipley Terrace	6	$81		20912	3	$76
Lamont Riggs, Queens Chapel, Fort Totten, Pleasant Hill	10	$81		20020	58	$83
Colonial Village, Shepherd Park, North Portal Estates	16	$82		20012	35	$88
North Michigan Park, Michigan Park, University Heights	26	$84		20017	60	$94
Woodridge, Fort Lincoln, Gateway	14	$85		20018	36	$101
Twining, Fairlawn, Randle Highlands, Penn Branch, Fort Davis Park, Fort Dupont	19	$86		20032	8	$108
Takoma, Brightwood, Manor Park	48	$95		20006	13	$110
Mayfair, Hillbrook, Mahaning Heights	6	$95		20015	21	$112
River Terrace, Benning, Greenway, Dupont Park	6	$99		20011	154	$114
Sheridan, Barry Farm, Buena Vista	7	$106		20010	215	$118
Brookland, Brentwood, Langdon	44	$108		20002	405	$128
Cathedral Heights, McLean Gardens, Glover Park	45	$109		20024	62	$131
Hawthorne, Barnaby Woods, Chevy Chase	15	$110		20016	48	$133

*** ADDED PROMPT: PROPERTY TYPE $ ***

- Which property types generate the most daily revenue?

Property Type by Avg Daily Revenue		
property_typ	Coun	Average of Daily Revenue
Townhouse	48	$168
Loft	19	$146
House	967	$142
Bed & Breakfast	38	$139
Other	8	$138
Apartment	1757	$134
Condominium	51	$131
Cabin	1	$110
Dorm	2	$83
Bungalow	1	$65

*** ADDED PROMPT: TARGETS ***

Neighborhood_Cleansed	Count of id	Average of Est_Daily_Revenue
Capitol Hill, Lincoln Park	238	157.9
Apartment	129	140.8
House	109	178.1
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street	134	184.5
Apartment	124	185.9
House	10	167.0
Georgetown, Burleith/Hillandale	83	223.4
Apartment	50	186.5
House	33	279.3
Howard University, Le Droit Park, Cardozo/Shaw	103	151.4
Apartment	58	133.9
House	45	174.0
Shaw, Logan Circle	221	157.7
Apartment	166	153.4
House	55	170.8
West End, Foggy Bottom, GWU	94	166.6
Apartment	88	161.6
House	6	239.8
Grand Total	873	168.3

Focus on Apartments & Houses

- This collection of high revenue neighborhoods has 258 houses and 615 apartments

*** ADDED PROMPT: TARGETS ***

Neighborhood_Cleansed	Count of id	Average of Est_Daily_Revenue
Colonial Village, Shepherd Park, North Portal Estates	15	76.9
Apartment	3	45.7
House	12	84.8
Historic Anacostia	24	72.8
Apartment	9	77.9
House	15	69.7
Lamont Riggs, Queens Chapel, Fort Totten, Pleasant Hill	8	91.3
Apartment	3	66.7
House	5	106.0
North Michigan Park, Michigan Park, University Heights	24	84.4
Apartment	10	91.7
House	14	79.1
Takoma, Brightwood, Manor Park	47	95.0
Apartment	14	88.1
House	33	97.9
Twining, Fairlawn, Randle Highlands, Penn Branch, Fort Davis Park, Fort Dupont	18	86.9
Apartment	3	85.0
House	15	87.3
Woodridge, Fort Lincoln, Gateway	13	84.6
Apartment	2	70.0
House	11	87.3
Grand Total	149	85.8

Focus on Apartments & Houses

- This collection of low revenue neighborhoods has 105 houses and 44 apartments
- It is possible that the gap between houses & apartments in aggregate is driven primarily by the neighborhood (that there are more apartments in the higher-priced neighborhoods)
- It looks like a house is usually preferable to a single-family dwelling

GRAND CONCLUSIONS

- I am nervous in making any recommendations to an investor based on this data, for these reasons:
  - More time needs to be dedicated to the data cleaning/gathering process
  - Major assumptions made as we attempted to speculate revenue…the bookings adjustment (Reviews x 2) is the most troublesome, as it makes it impossible to compare oft-rented properties to ones rented less frequently…the investor would have control over how frequently/infrequently to rent the property
  - Lack of overlaid geography/economic data, i.e. we know neighborhoods, but we don't know how many people or housing units are in them which would be needed information in trying to identify areas that are over/under saturated with Air BnBs
  - More time needs to be dedicated to the data cleaning/gathering process

GRAND CONCLUSIONS (CONT)

- I am nervous in making any recommendations to an investor based on this data, for these reasons:
  - We are only considering half of the formula, we are ignoring the acquisition cost
    - Maybe buying a place in Anacostia with the expectation that we could rent it out as an Air BnB for $75 per day would be a great investment, if we can get the property at the right price?
    - Maybe buying a place in Georgetown would be a stretch, even if we could rent it out for more than $200 per night?

GRAND CONCLUSION

- But if absolutely forced to make a recommendation, and being sensitive to my investor's desire to go into only proven markets…
  - I would pay a lot of attention to location, and look for affordable real estate in the areas proven to go for over $175 per night, or perhaps even *more* affordable real estate in the $125-175 range
  - I would pay some attention to property type, knowing that houses are, when controlling for location, the best bet to rent for a high daily amount…BUT knowledge that neighborhoods are the driver, that an apartment in one part of town can be expected to outperform a house in another
    - Maybe just develop a proxy for square feet?
  - I would pay no attention to reviews
    - Junk data
    - Not much variation anyway
    - Money is the ultimate review (If they were at one point willing to pay for it, what does it matter what their review is? Is this more cleanliness/customer service-related than it is neighborhood/property-related? Is repeat-business a thing in this industry?)