



# Airbnb Seattle

## – Host Rental Market –

### -- Host Recommendations --

## Project Report

---

- RYAN TIMBROOK

# Airbnb – Seattle Project Report

---



Research  
Objective



Research Plan and  
Methodology



Data Findings



Conclusion



Recommendations

# Research Objective

---



Using the Seattle Airbnb data set, identify key listing and host characteristics of the top income earners to provide insights for hosts to be more profitable and give potential hosts recommendations on how to be in the top income earning pool.



Additionally, for Airbnb's business model, they can give recommendations to hosts and develop a new "premium feature" that will help hosts learn how to be a top earner.

# Research Objectives

---

## Questions to answer:

- Which Hosts are running businesses with multiple listings, and where are they located in the city?
- How many listings are in each neighborhood and where are they? Identify where hosts have competition and what the ideal price range is for a given listing property type.
- How much are hosts making from renting to tourists? And what is their net yearly lifetime income of their potential? *(-hosts can select which days out of the year a listing property is available to be rented, this effects lifetime potential earnings.)*
- What are the top performing listing property characteristics? *(i.e., type of property, it's amenities, it's occupancy accommodations, it's rental policies, etc.)*
- How are customers rating their Hosts? What is the overall sentiment of their review comments on a Host and how does this reflect on the hosts earned income potential?

# About the Data

The primary data used for this analysis will be the Airbnb Seattle datasets. It will be used to profile Seattle Airbnb hosts and customers in efforts to understand the rental market in Seattle.

Zillow datasets will provide median home sale values and rental forecasts by zip code. This data will be used to provide insights into the cost of purchasing properties as potential investments opportunities that could be used as short-term rentals.

Walking score's, public transportation scores and travel time scores are all influential factors that influence how consumers of short-term rental housing choose where and what type of logging they prefer.



Airbnb Seattle:  
[Kaggle](#)

calendar.csv  
listings.csv  
reviews.csv



[Zillow Housing Data](#)

Home values  
Rentals  
Inventory,  
listings and  
sales



[Walkscore.com](#)

Walk Score API  
Public Transit  
API  
Travel Time API



# About the Data

## Context

Since 2008, guests and hosts have used Airbnb to travel in a more open, personalized way. As part of the Airbnb Inside initiative, this dataset describes the listing activity of homestays in Seattle, WA.

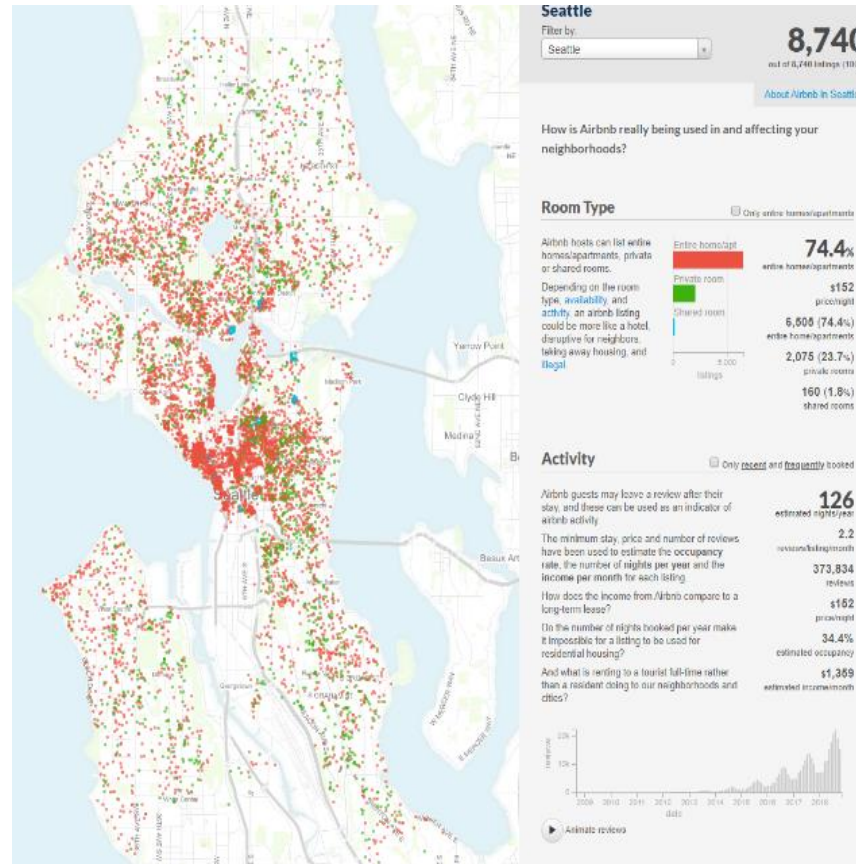
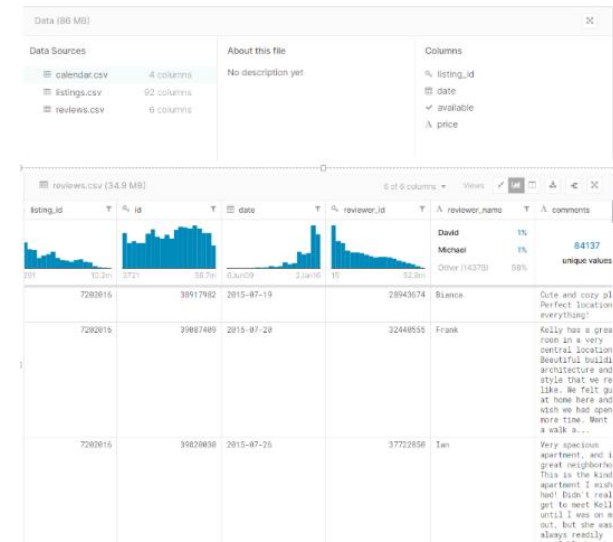
## Content

The following Airbnb activity is included in this Seattle dataset: \* Listings, including full descriptions and average review score \* Reviews, including unique id for each reviewer and detailed comments \* Calendar, including listing id and the price and availability for that day

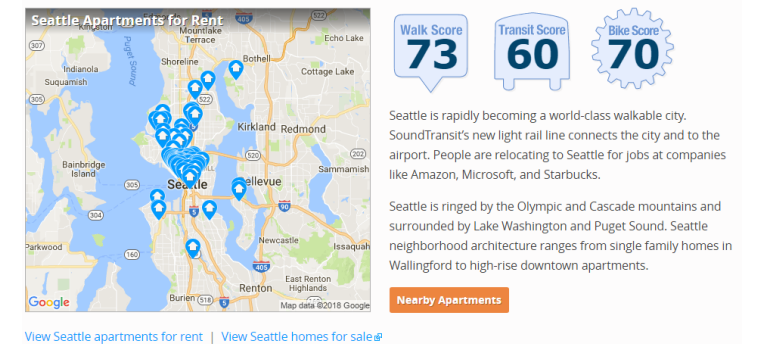
## Inspiration

- Can you describe the vibe of each Seattle neighborhood using listing descriptions?
- What are the busiest times of the year to visit Seattle? By how much do prices spike?
- Is there a general upward trend of both new Airbnb listings and total Airbnb visitors to Seattle?

For more ideas, visualizations of all Seattle datasets can be found [here](#).



## Living in Seattle



# Research Plan and Methodology

---



Exploratory Data Analysis | Sentiment Analysis | Conjoint analysis | Logistic Regression | Data Modeling



Python | SQL Server | Power BI

# Research Plan & Methodology

---

## Identify Hosts running businesses

- Some hosts have multiple listings.
- A host may list separate rooms in the same apartment, or multiple apartments or homes.
- Hosts with multiple listings are more likely to be running a business and are less likely to be living in the listing property.

**48%** of Hosts have multiple listings

## Derived Target Lifetime Income Variables

- Airbnb guests may leave a review after their stay, and these can be used as an indicator of airbnb activity.
- The **minimum stay**, **price** and **number of reviews** are used to estimate the **occupancy rate**, the number of **nights per year** and the **income per month** for each listing.
- Hosts can limit the available days a property listing can be rented, therefore to calculate lifetime income earning to potential ratios, **calendar available days** is factored into the equations.



# Estimating Lifetime Income

---

**Lifetime Yearly Income** = ( (Minimum nights \* Number of Reviews) / ((Last Review Date – First Review Date) / 365)) \* Price

**Lifetime Potential Income** = Calendar Available Days \* ((Last Review Date – First Review Date) / 365) \* Price

**Lifetime Yearly Potential Income** = Lifetime Potential Income / ((Last Review Date – First Review Date) / 365)

**Percent of Yearly Income of Potential** = (Lifetime Yearly Income / Lifetime Yearly Potential Income) \* 100

Lifetime Earners Group Categories:

**Lifetime Earner:**

- 0: Low – (Bottom 25%)
- 1: Mid – (Mid 50%)
- 2: High – (Top 25%)

Lifetime Earner is used as a target variable for predicting top earners and what listing characteristics are most influential in making the prediction.

# Host Reviews

---

- Analyze the hosts review data set for identifiable traits that are primary drivers of top performers.
  - What review scores have the most impact on predicting top performs such as cleanliness, communication, location, accuracy, checkin, value, sentiment polarity.
- Additionally:
  - Host Response Rate
  - Host Acceptance Rate
  - Host Response Time
  - Host Policies

# Data Findings

---

SENTIMENT ANALYSIS | TOPIC MODELING | REGRESSION  
ANALYSIS | POWER BI DYNAMIC DASHBOARDS



# Data Findings

## Host Reviews Sentiment Topic Modeling

---

### ➤ Latent Dirichlet Allocation Model

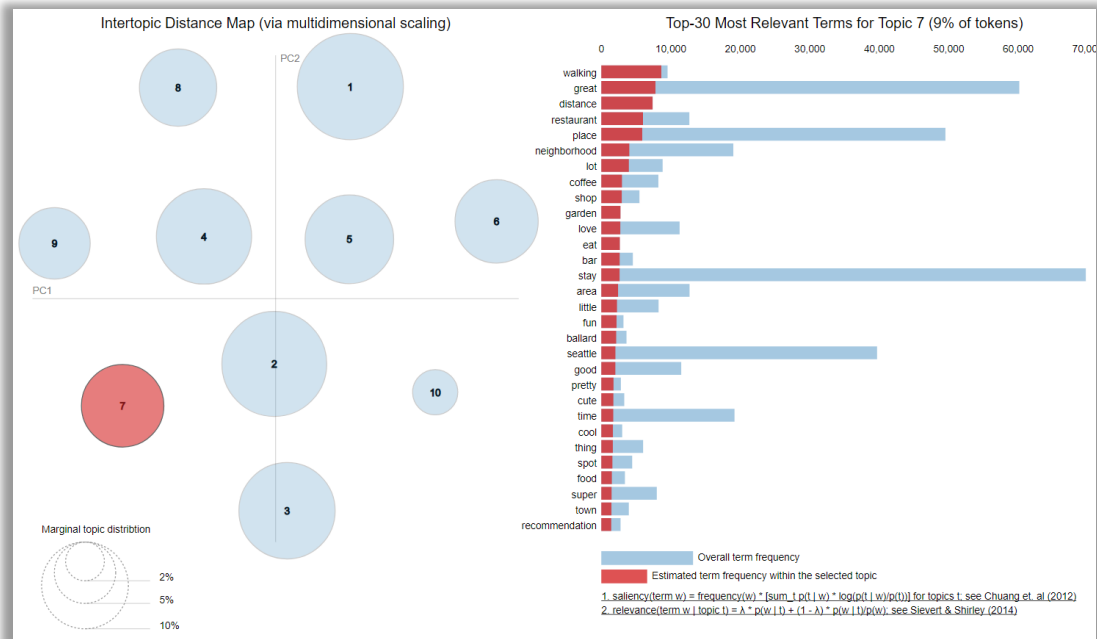
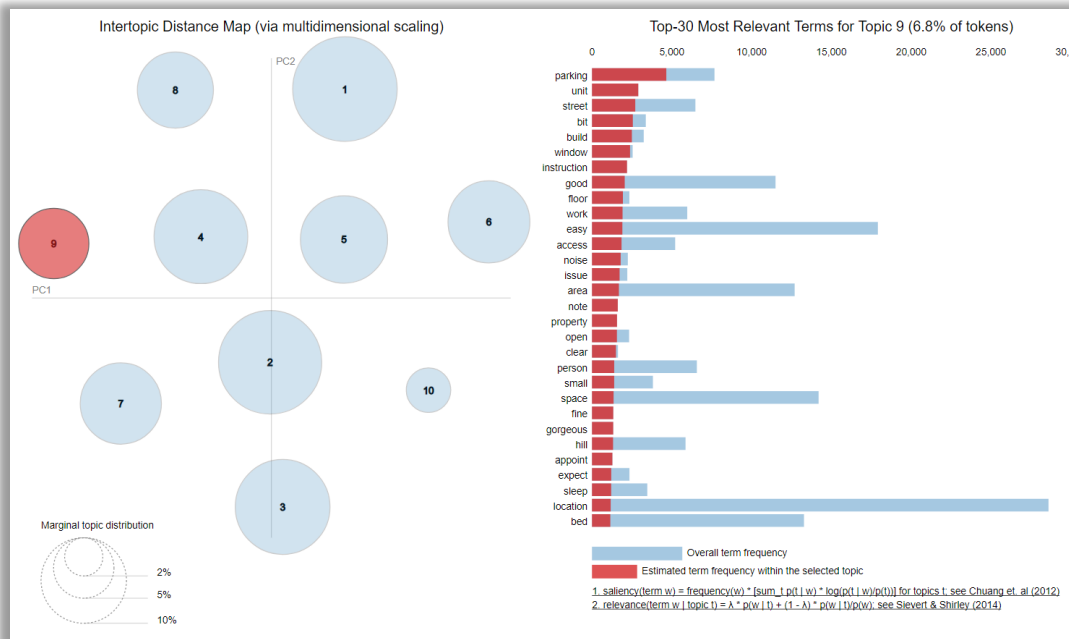
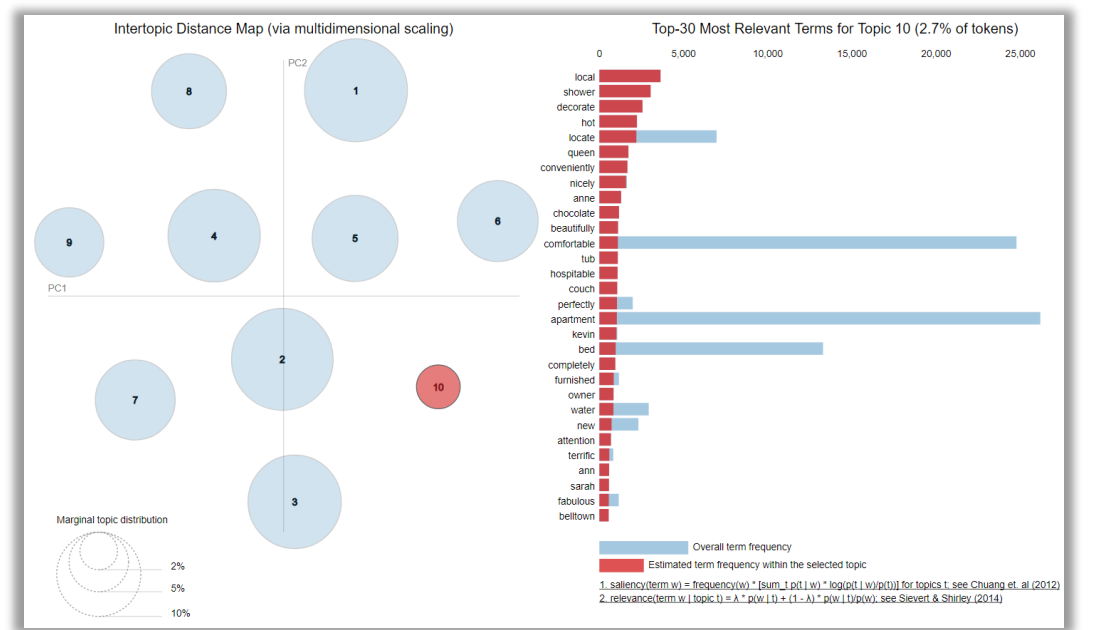
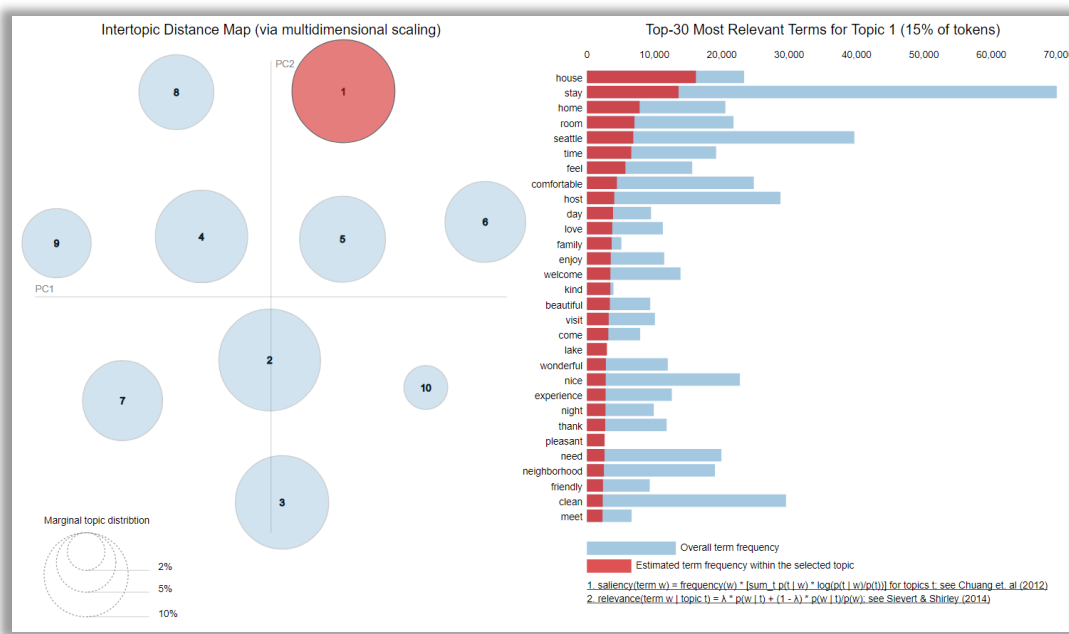
#### ➤ Top 10 **Positive** Feature Words assigned to 10 topics

Topic	Top_10_Features
0	great stay place apartment location check seattle need clean question
1	stay great nice clean place room comfortable host location bed
2	kitchen need apartment bed bathroom amenity bedroom nice door towel
3	parking unit street bit build window instruction good floor work
4	house stay home room seattle time feel comfortable host day
5	walk downtown bus seattle restaurant minute close great away place
6	home feel welcome host provide experience accommodation fresh coffee wonderful
7	stay great seattle place recommend host location highly perfect clean
8	walking great distance restaurant place neighborhood lot coffee shop garden
9	local shower decorate hot locate queen conveniently nicely anne chocolate

#### ➤ Top 10 **Negative** Feature Words assigned to 10 topics

Topic	Top_10_Features
0	stay apartment place room house host clean night time day
1	apartment bus away walk kitchen war restaurant walking seattle minute
2	et tr�� la le pour tout d�� il ce en
3	die sehr wir war da mit f�� ����r es man
4	een en het meet op huis niet maar voor van
5	better stay draw floor dirty clean need sehr experience bathroom
6	mit von wir stay man die instruction internet cottage pillow
7	sehr die wir f�� war haben ����r seattle da place
8	la el en para muy es por todo ����n casa
9	house room code wir stay car mit bit entrance bus

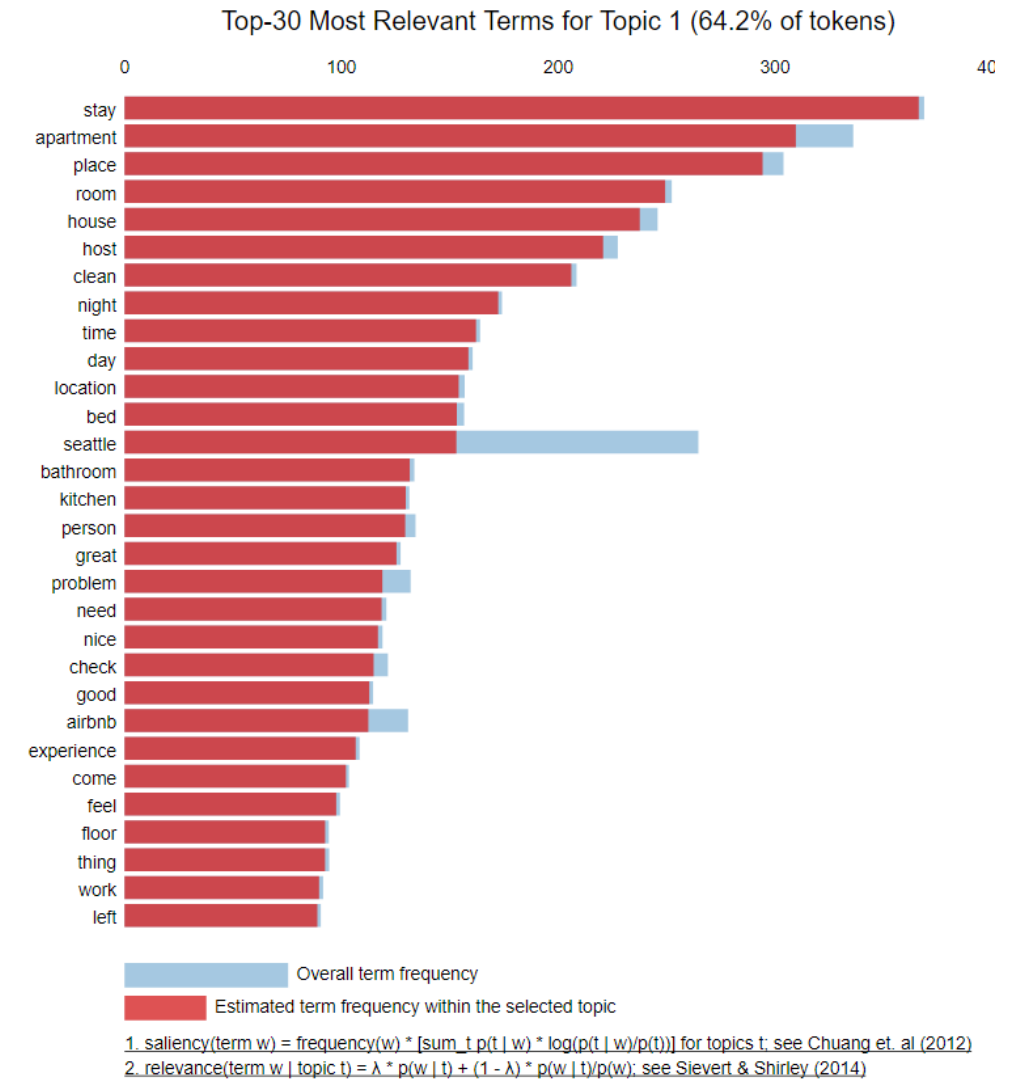
Positive word topic assignments



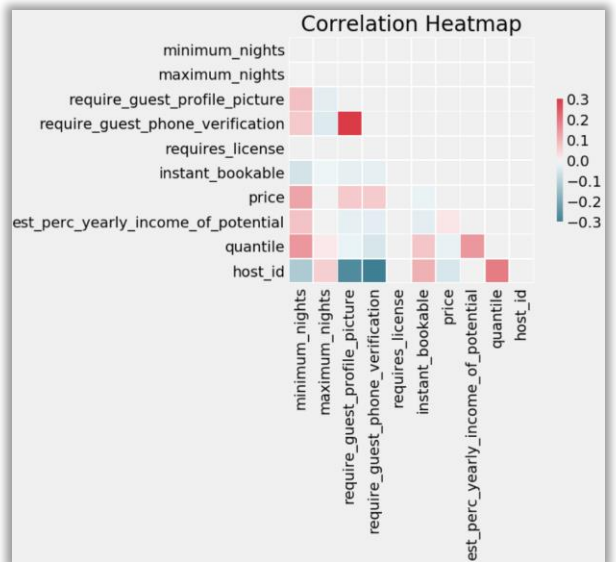
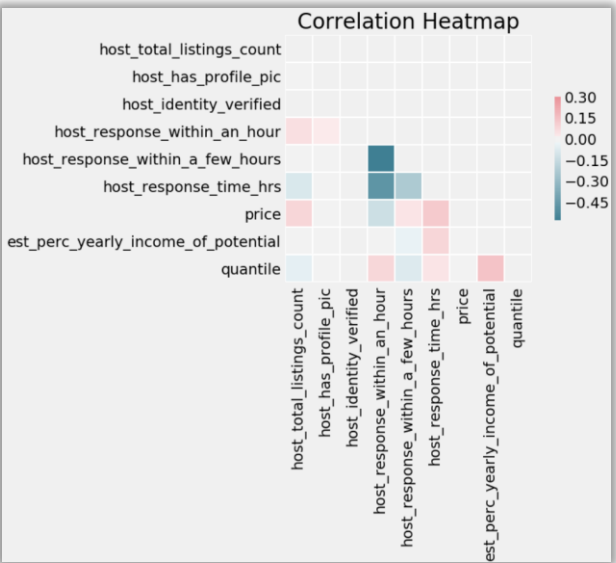
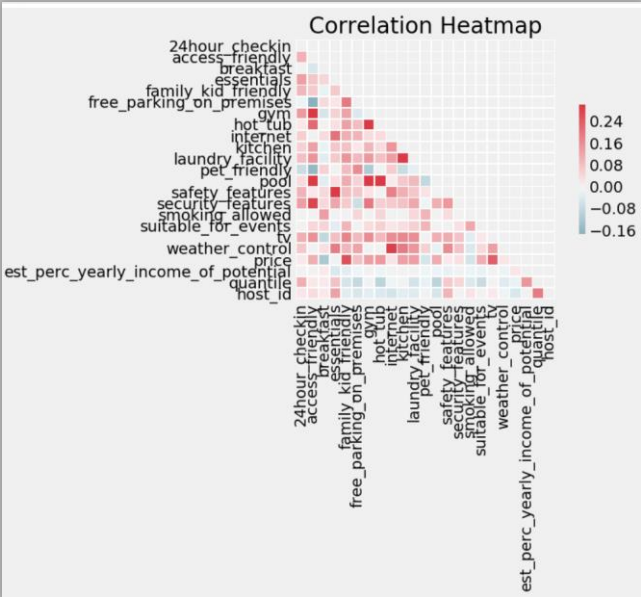
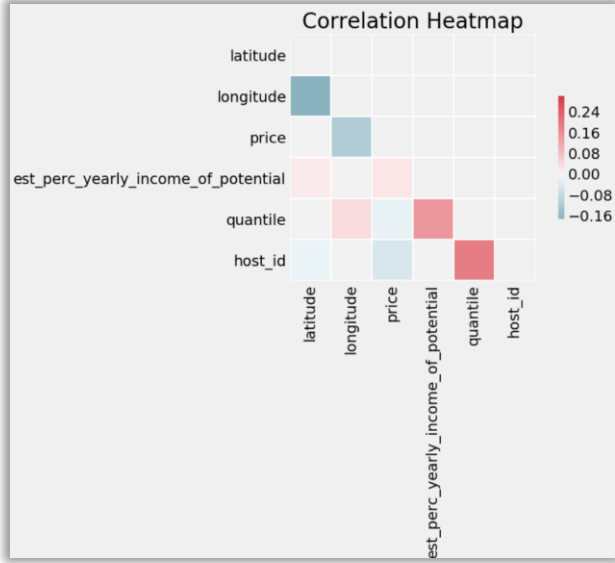
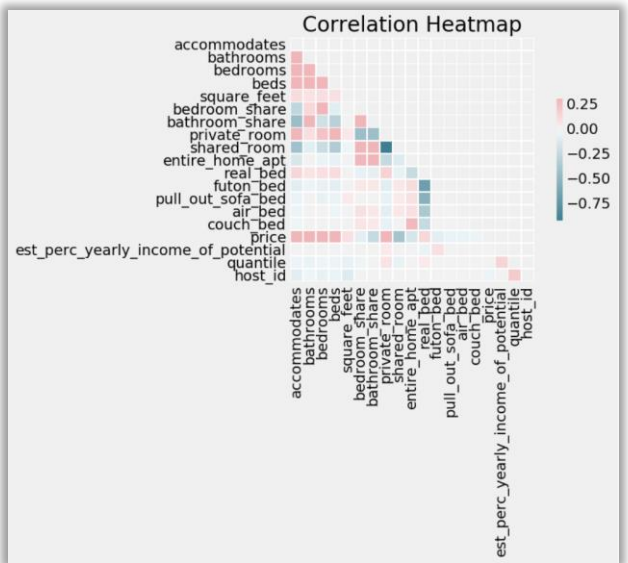
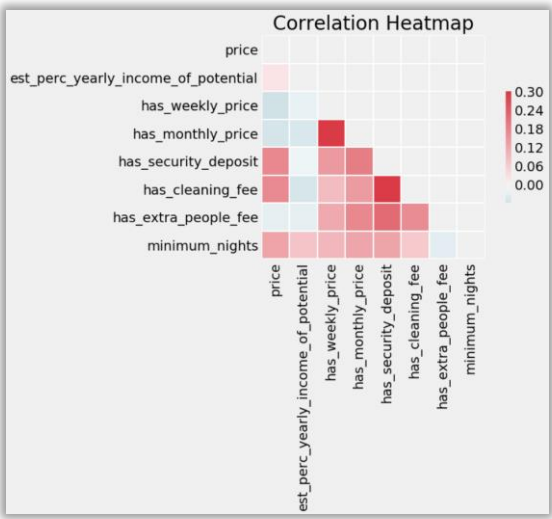


# Negative Review Sentiment Topic Assignment

- This model represents those things customers talked most negatively about in their review comments.



➤ Data set attribute correlation analysis (Used in determining optimal modeling coefficient selections)

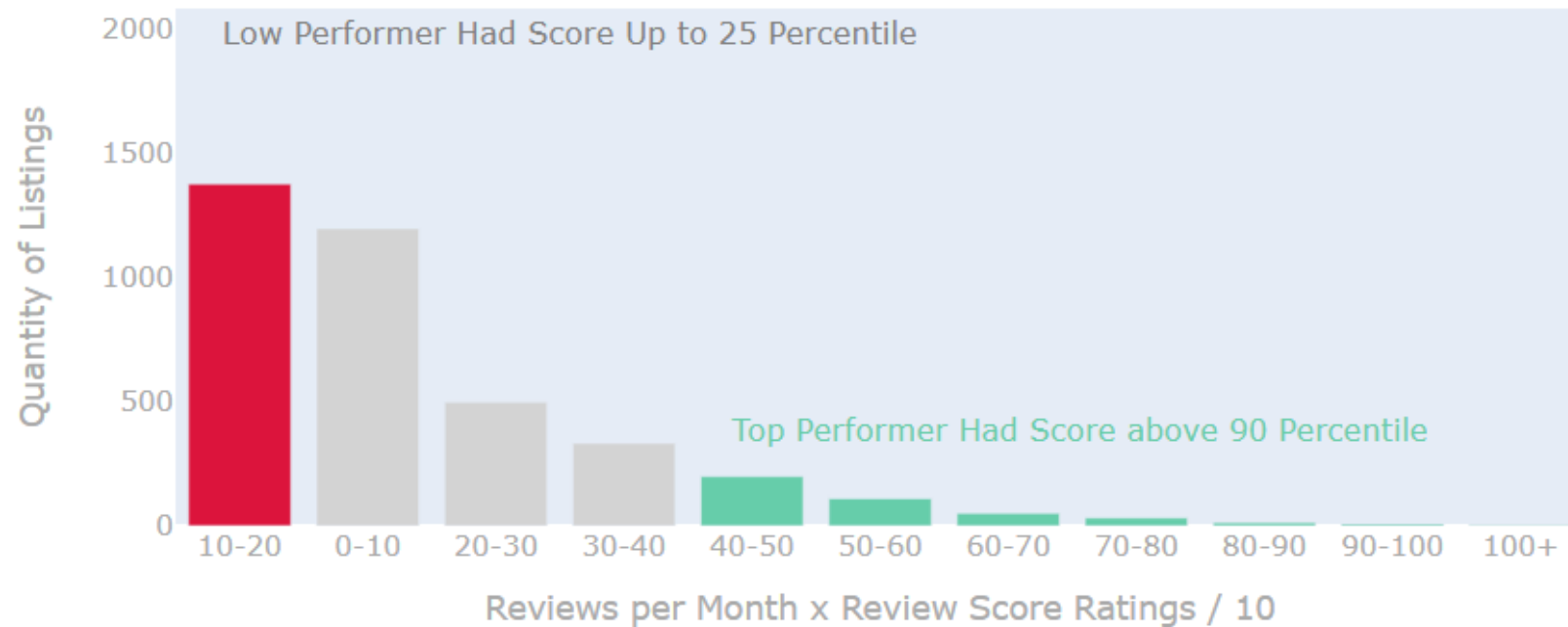


# Data Findings

## Comparing Listing Performance between Top and Bottom Income Generators

---

### Comparison Listings Performance between Top Performer and Low Performer



# Data Findings

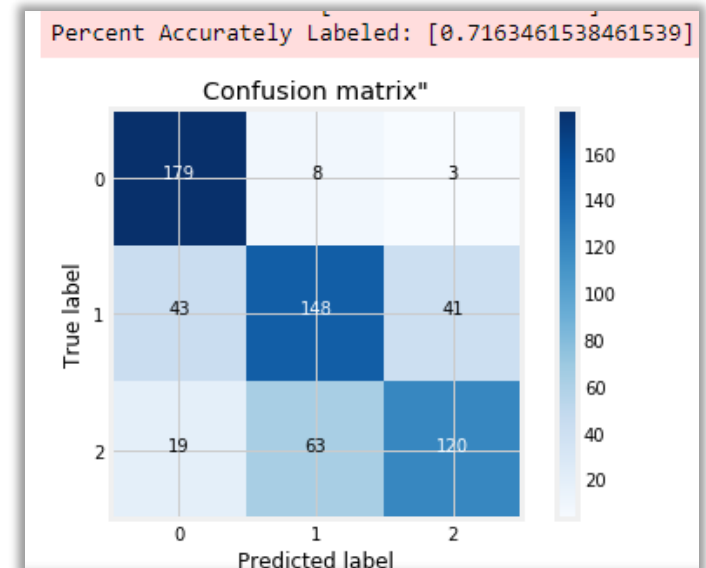
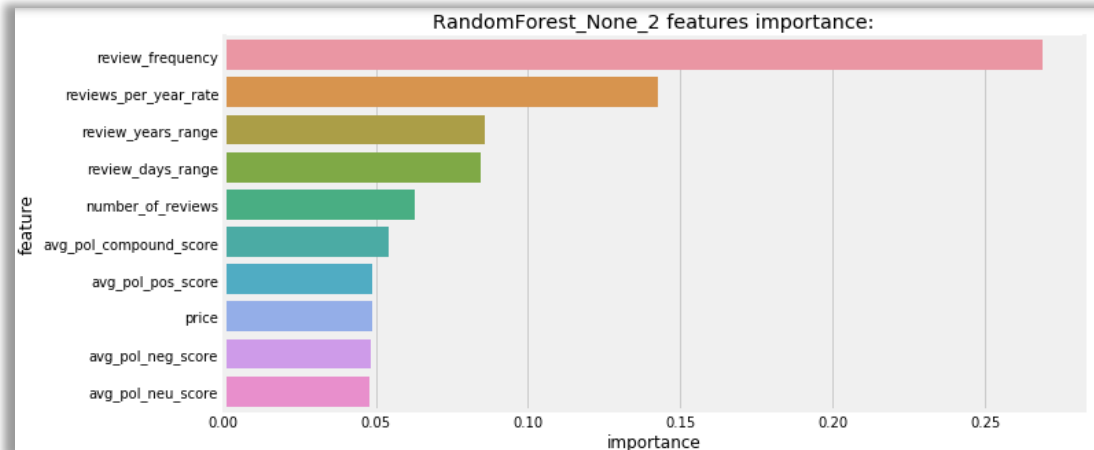
## Best Predictor of Lifetime Earner

**71%** -- Predicting Lifetime Earner using RandomForest Regression algorithm

- These attributes are expected – Lifetime Earner is a calculated value based on reviews, price, calendar available days.
- Many additional models were explored for each type of listing characteristic grouping to identify the most impactful attributes.

	precision	recall	f1-score	support
Class0	0.74	0.94	0.83	190
Class1	0.68	0.64	0.66	232
Class2	0.73	0.59	0.66	202
micro avg	0.72	0.72	0.72	624
macro avg	0.72	0.72	0.71	624
weighted avg	0.71	0.72	0.71	624

	feature	importance
19	review_frequency	0.269055
18	reviews_per_year_rate	0.142461
17	review_years_range	0.085814
16	review_days_range	0.084456
0	number_of_reviews	0.062600
12	avg_pol_compound_score	0.054232
9	avg_pol_pos_score	0.049033
20	price	0.048792
10	avg_pol_neg_score	0.048235
11	avg_pol_neu_score	0.047738



# Data Findings

## Using RandomForest to identify top listing characteristics

### Pricing Features

#### feature importance:

- price 0.223598
- cleaning\_fee 0.137062
- weekly\_discount\_rate 0.075858
- weekly\_discount\_offer 0.074313
- weekly\_price 0.073819
- security\_deposit 0.071463
- monthly\_discount\_rate 0.067102
- extra\_people\_fee 0.064742
- monthly\_discount\_offer 0.052550
- monthly\_price 0.051312

### Listing Policies

feature	importance
0	minimum_nights 0.828039
3	cancellation_policy_strict 0.074474
2	cancellation_policy_moderate 0.049105
1	cancellation_policy_flexible 0.048382
4	require_guest_profile_picture 0.000000
5	require_guest_phone_verification 0.000000
6	requires_license 0.000000
7	instant_bookable 0.000000

### Host Attributes

feature	importance
4	avg_pol_compound_score 0.876119
0	review_scores_cleanliness 0.041120
3	review_scores_location 0.029133
1	review_scores_checkin 0.024762
2	review_scores_communication 0.024093
6	avg_review_sentiment_positive 0.004602
5	avg_review_sentiment_negative 0.000171

feature	importance:
	host_response_rate 0.679550
6	host_response_time_hrs 0.082153
3	host_acceptance_rate 0.080296
7	host_identity_verified_t 0.073783
1	host_response_within_an_hour 0.042228
2	host_response_within_a_few_hours 0.033426
8	host_has_profile_pic_f 0.003132
4	host_has_profile_pic 0.002762
9	host_has_profile_pic_t 0.002670
5	host_identity_verified 0.000000

### Property Type Features

feature	importance
23	beds 0.199057
25	bathroom_share 0.158313
24	bedroom_share 0.114992
21	bathrooms 0.111810
20	accommodates 0.101871
22	bedrooms 0.060007
9	property_type_House 0.039505
0	property_type_Apartment 0.036890
7	property_type_Condominium 0.017948
13	property_type_Townhouse 0.017648

### Property Amenities

feature	importance
17	tv 0.095476
4	family_kid_friendly 0.095303
5	free_parking_on_premises 0.089218
11	pet_friendly 0.088012
14	security_features 0.086070
10	laundry_facility 0.061102
1	access_friendly 0.055839
9	kitchen 0.053870
2	breakfast 0.052264
0	24hour_checkin 0.047540

### Neighborhoods

feature	importance
9	neighbourhood_group_cleansed_Magnolia 0.129263
6	neighbourhood_group_cleansed_Downtown 0.111628
10	neighbourhood_group_cleansed_Northgate 0.092926
12	neighbourhood_group_cleansed_Queen Anne 0.084018
1	neighbourhood_group_cleansed_Beacon Hill 0.065903
0	neighbourhood_group_cleansed_Ballard 0.058906
16	neighbourhood_group_cleansed_West Seattle 0.058077
4	neighbourhood_group_cleansed_Central Area 0.052056
14	neighbourhood_group_cleansed_Seward Park 0.049179
11	neighbourhood_group_cleansed_Other neighborhoods 0.046223

# Data Findings

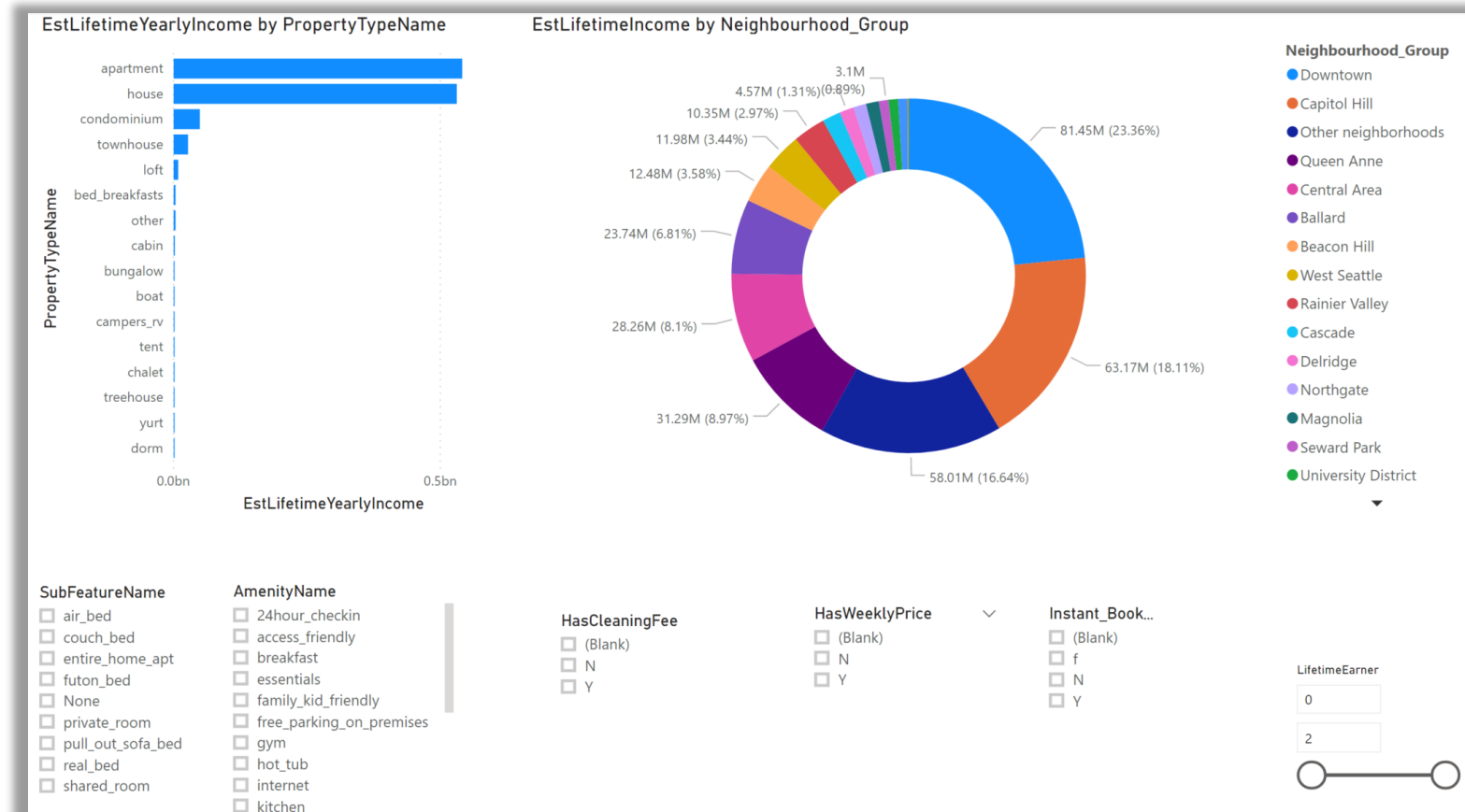
## Power BI Dashboard – Lifetime Earning by Neighborhoods and Property Characteristics

### Top Listing Property Types:

- Apartments
- Houses

### Top Neighborhoods:

- Downtown
- Capital Hill
- Other neighborhoods
- Queen Anne
- Central Area
- Ballard

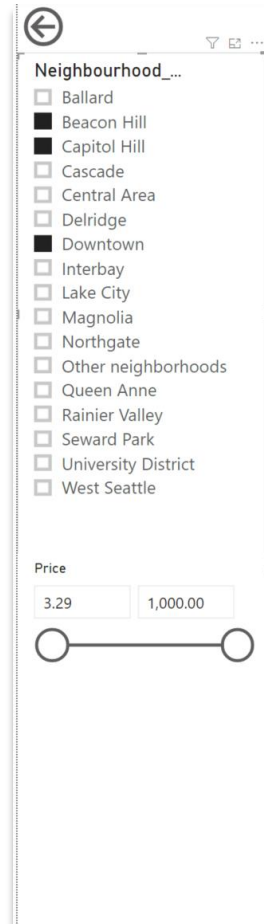




# Data Findings

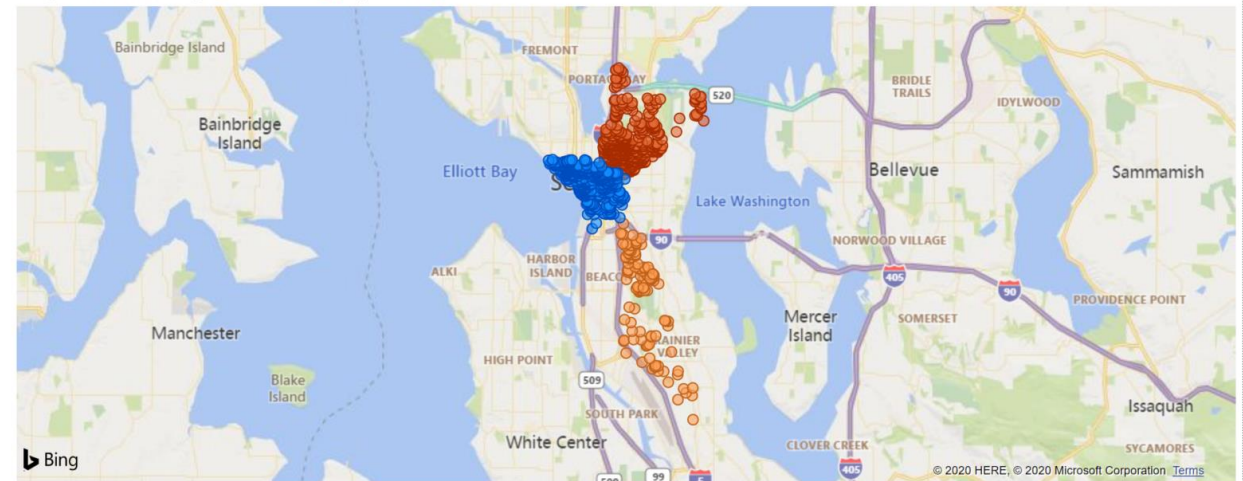
## Power BI Dashboard – Lifetime Earning by Neighborhoods and Property Characteristics

- Dashboard to identify price ranges by neighborhood and density of listings per neighborhood.

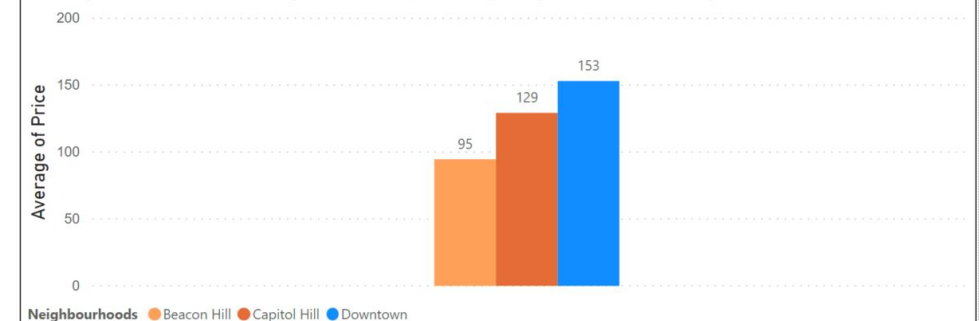


Neighbourhood\_Group, Latitude and Longitude

Neighbourhood\_Group ● Beacon Hill ● Capitol Hill ● Downtown



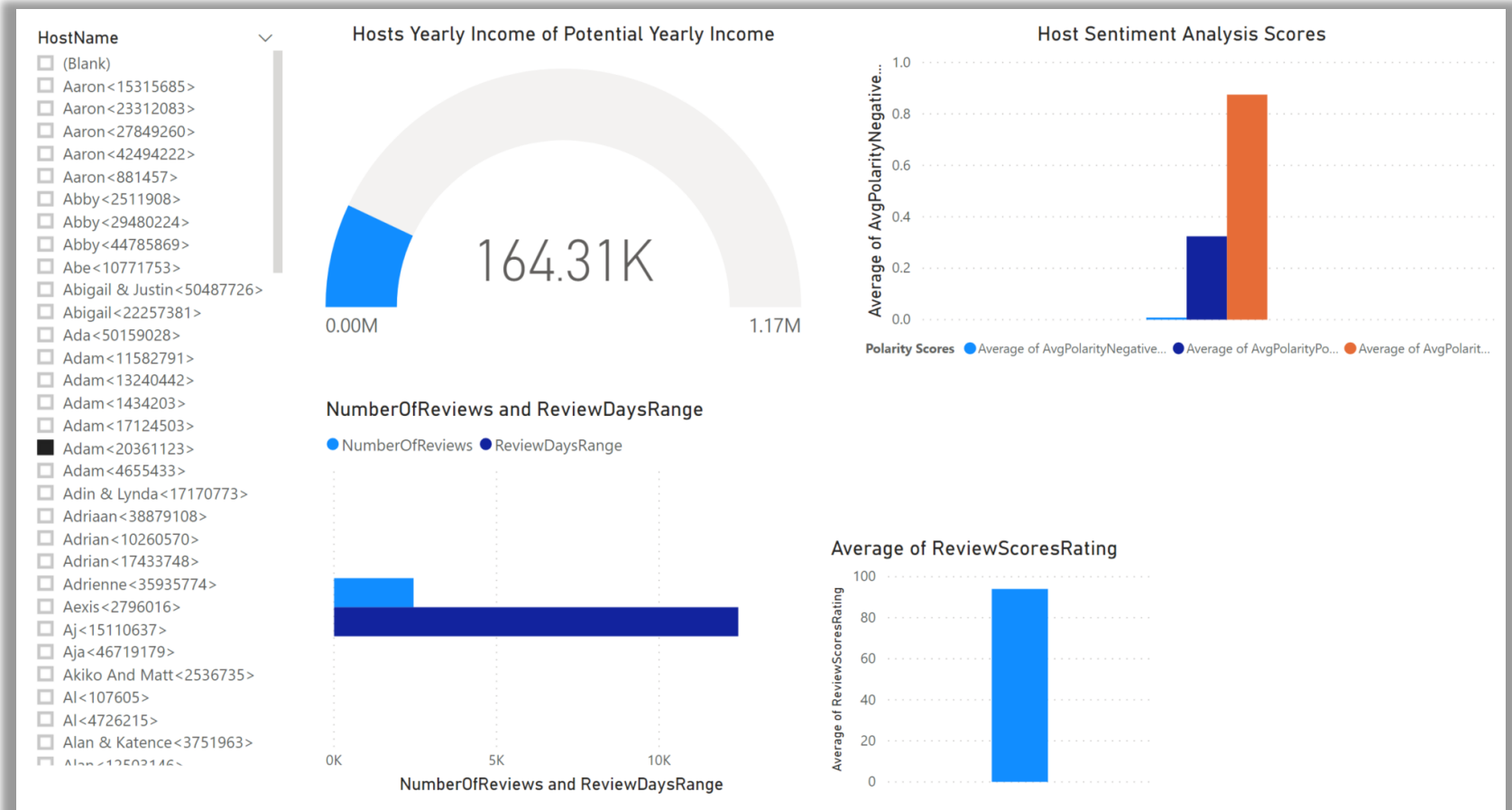
Average of Price and Average of MonthlyPrice by Neighbourhood\_Group



# Data Findings

## Power BI Dashboard – Lifetime Earning by Neighborhoods and Property Characteristics

- Dashboard to analyze host income performance while observing customer sentiment and review score ratings of that host.

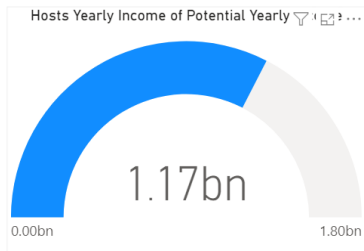


# Conclusion

---

Which Hosts are running businesses with multiple listings, and where are they located in the city?

- **48%** of hosts appear to be running business through their listings. The BI Dashboard provided identifies each of the neighborhoods these business run in and the average price ranges.



How much are hosts making from renting to tourists? And what is their net yearly lifetime income of their potential?

- The lifetime income for all the hosts in Seattle combine for 1.17 billion out of a possible 1.80 billion. Individual hosts lifetime income of potential can be searched using the Power BI Dashboard provided.

What are the top performing listing property characteristics?

- Apartments
- Houses
  - Located in Downtown or Other neighborhoods

How are customers rating their Hosts

- Overall they rate their hosts highly, with strong positive sentiment comments

# Recommendations – Hosts

---

- **Increase the acceptance rate of the rental.** Top performer almost never reject the orders, not like low performers around 45% of the listings that did not accept any bookings.
- **Be responsive** most of the top performer hosts always giving response within an hour, about 80% of all the time.
- **Be a super host.** It's the status and the recognition from the Airbnb because they provide amazing experience and great example for other hosts.
- **Always response the conversations.** More responsive the hosts, higher are their scores. Top performers always respond to every conversation.
- **Activate the instant bookable features;** They give better experience to the future guests.
- **Make your account verified by Airbnb.** Many of low performer accounts have not verified by Airbnb more than the high performer by 25%.
- Focus your attention to the **Downtown and Capital Hill** area.

# Airbnb – Seattle Project Plan & Research Design

## Assignment Questions:

- ☐ Does the popular opinion (sentiment) of customers' comments impact a host's price point? Does it have an affect on the number of days of the year that a property is booked? What are the positive and negative top 5 topics for each segment and how can this information be used to help hosts be more successful with their bookings?
- ☐ Determine the best property to purchase for use as a vacation rental.
- ☐ What are the attributes of a successful rental listing posting? An unsuccessful rental listing?
- ☐ Do property attributes or host rating affect the success of a listing more?

# Airbnb – Seattle Project Plan & Research Design

Analysis:

□ Types of analysis that will be utilized are:

- Text Mining - Sentiment Analysis
- Text Mining - Topic Modeling
- Cluster Analysis
- Logistic Regression
- Perceptual Maps



# Airbnb – Seattle Project Plan & Research Design

## Key Variables:

- ❑ Key variables that will be used are customer reviews of the rental properties along with attributes of the properties themselves.
- ❑ In addition we will pull outside data sources in to provide additional details about each property that are not included in the original data set through the use of API's.
  - Walk Score/Bike Score/Transit Score
  - MLS data from real estate sites such as Zillow or Redfin
  - Demographic data based on Zip Code of the rental property

### Key Variables

**reviews** [reviewer\_name, comments, date] (keys to link the datasets will also be included such as 'listing\_id', 'reviewer\_id')

**listings** [price, weekly\_price, monthly\_price, security\_deposit, cleaning\_fee, number\_of\_reviews, first\_review, last\_review, review\_scores\_rating, review\_scores\_accuracy, review\_scores\_cleanliness, review\_scores\_checkin, review\_scores\_communication, review\_scores\_location, review\_scores\_value, reviews\_per\_month, host\_since, host\_response\_time, host\_responase\_rate, host\_acceptance\_rate, host\_is\_superhost, host\_listing\_counts, host\_total\_listings\_count, property\_type, room\_type, accommodations, bathrooms, bedrooms, beds, bed\_type, amenities, square\_foot, availability\_30, availability\_60, availability\_90, avaiability\_365]

# About the Data

## Context

Since 2008, guests and hosts have used Airbnb to travel in a more open, personalized way. As part of the Airbnb Inside initiative, this dataset describes the listing activity of homestays in Seattle, WA.

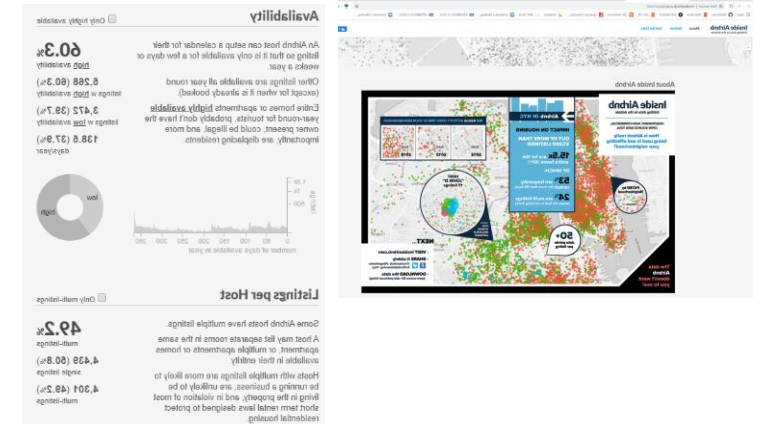
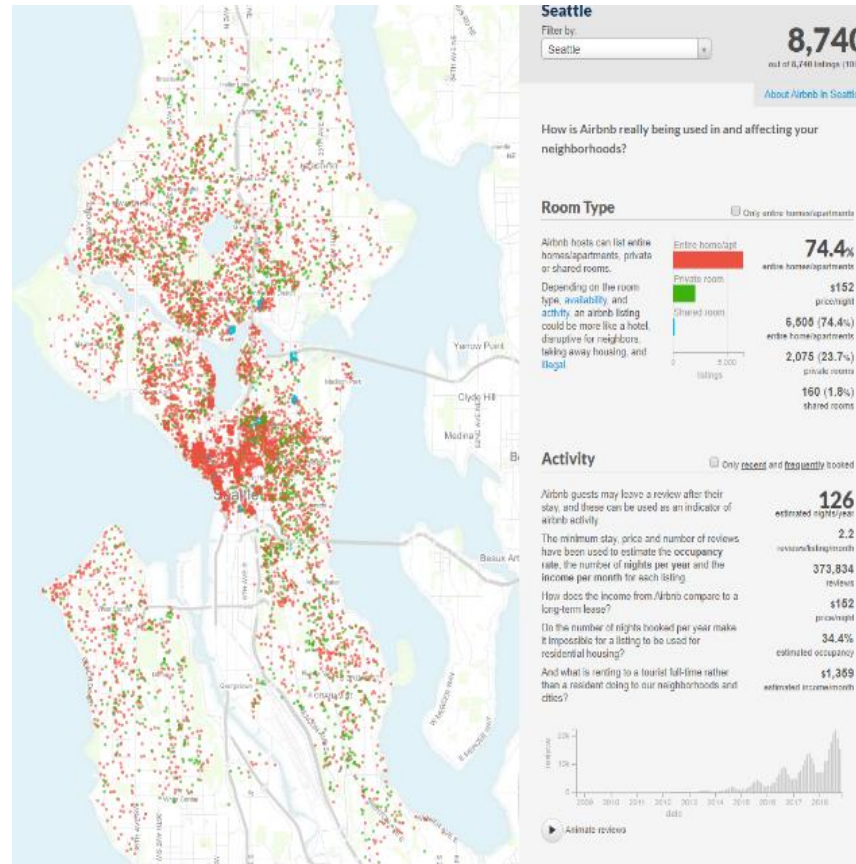
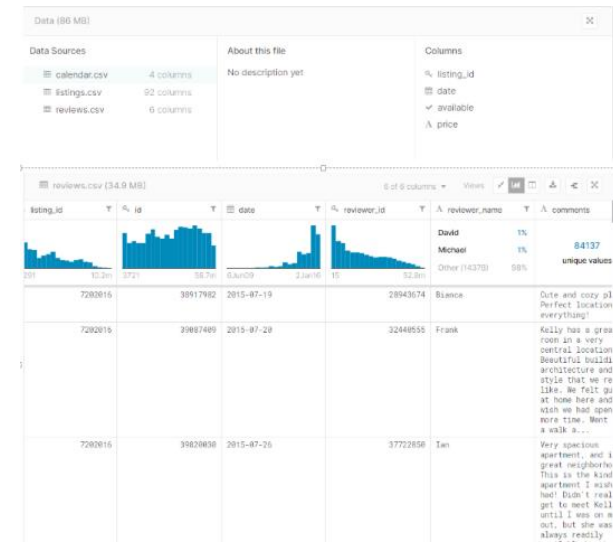
## Content

The following Airbnb activity is included in this Seattle dataset: \* Listings, including full descriptions and average review score \* Reviews, including unique id for each reviewer and detailed comments \* Calendar, including listing id and the price and availability for that day

## Inspiration

- Can you describe the vibe of each Seattle neighborhood using listing descriptions?
- What are the busiest times of the year to visit Seattle? By how much do prices spike?
- Is there a general upward trend of both new Airbnb listings and total Airbnb visitors to Seattle?

For more ideas, visualizations of all Seattle datasets can be found [here](#).



## Scores for Seattle

The Walk Score, Transit Score, and Bike Score of Seattle are based on a weighted average of the scores of many addresses in the city.



Walk Score	Transit Score	Bike Score
Walk Score measures the walkability of any address based on the distance to nearby places and pedestrian friendliness.		
90-100	<b>Walker's Paradise</b> Daily errands do not require a car	
70-89	<b>Very Walkable</b> Most errands can be accomplished on foot	
50-69	<b>Somewhat Walkable</b> Some errands can be accomplished on foot	
25-49	<b>Car-Dependent</b> Most errands require a car	
0-24	<b>Car-Dependent</b> Almost all errands require a car	

## Scores for Seattle

The Walk Score, Transit Score, and Bike Score of Seattle are based on a weighted average of the scores of many addresses in the city.



Walk Score	Transit Score	Bike Score
Transit Score measures how well a location is served by public transit based on the distance and type of nearby transit lines.		
90-100	<b>Rider's Paradise</b> World-class public transportation	
70-89	<b>Excellent Transit</b> Transit is convenient for most trips	
50-69	<b>Good Transit</b> Many nearby public transportation options	
25-49	<b>Some Transit</b> A few nearby public transportation options	
0-24	<b>Minimal Transit</b> It is possible to get on a bus	

## Scores for Seattle

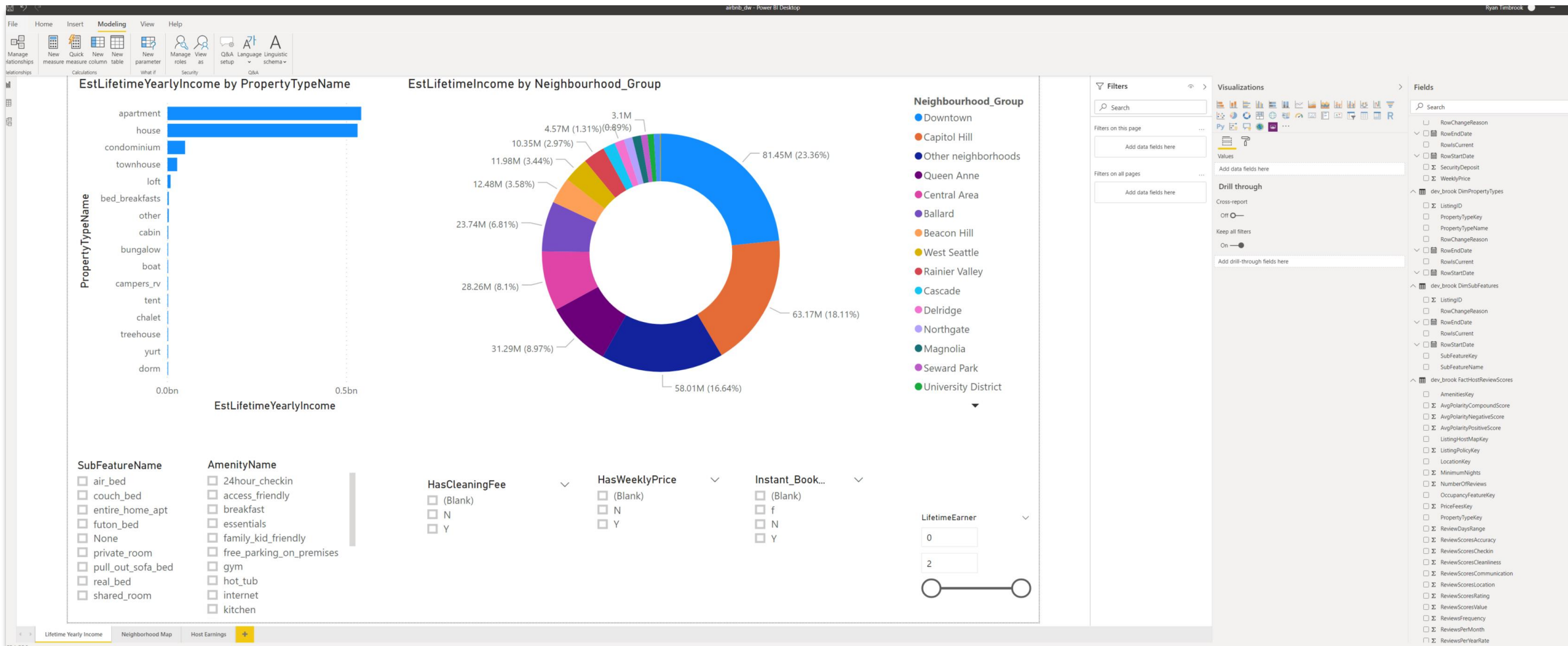
The Walk Score, Transit Score, and Bike Score of Seattle are based on a weighted average of the scores of many addresses in the city.



Walk Score	Transit Score	Bike Score
Bike Score measures whether an area is good for biking based on bike lanes and trails, hills, road connectivity, and destinations.		
90-100	<b>Biker's Paradise</b> Daily errands can be accomplished on a bike	
70-89	<b>Very Bikeable</b> Biking is convenient for most trips	
50-69	<b>Bikeable</b> Some bike infrastructure	
0-49	<b>Somewhat Bikeable</b> Minimal bike infrastructure	



## MAR 653 Marketing Analytics Project


WBS NUMBER	TASK TITLE	TASK OWNER	START DATE	DUE DATE	DURATION	PCT OF TASK COMPLETE	PHASE ONE							PHASE TWO							PM							
							WEEK 1		WEEK 2		WEEK 3		WEEK 4		WEEK 5		WEEK 6		WEEK 7									
1	Project Conception and Initiation						M	T	W	R	F	M	T	W	R	F	M	T	W	R	F	M	T	W	R	F	M	T
1.1	Research Marketing Project Ideas	Ryan, Steve, Tim	1/13/20	1/21/20	8	100%																						
1.2	Milestone 1: Project Idea Deck	Steve	1/21/20	1/22/20	1	100%																						
1.3	Project Initiation	Ryan, Steve, Tim	1/23/20	1/24/20	1	100%																						
2	Project Definition and Planning																											
2.1	Scope and Goal Setting	Ryan, Steve, Tim	2/3/20	2/4/20	1	100%																						
2.2	Communication Plan - Weekly Live Touchpoint	Ryan, Steve, Tim	2/3/20	2/4/20	1	100%																						
2.3	Research Questions Definition	Ryan, Steve, Tim	1/22/20	2/3/20	11	100%																						
2.4	Create Project Plan Worksheet	Ryan	2/1/20	2/4/20	3	100%																						
2.5	Define Research Design Specifications	Ryan, Steve, Tim	2/3/20	2/4/20	1	100%																						
2.4	Milestone 2: Project Plan and Research Design Deck	Ryan	2/4/20	2/5/20	1	100%																						
3	Project Conception and Initiation																											
3.1	Status and Tracking	Ryan	2/5/20	3/13/20	38	5%																						
3.2	Obtain Research Datasets	Ryan	1/31/20	2/1/20	1	100%																						
3.2.1	Perform Initial Data Exploratory Analysis	Ryan, Steve, Tim	2/3/20	2/5/20	2	20%																						
3.2.2	Define Marketing Research Models	Ryan, Steve, Tim	2/3/20	2/7/20	4	0%																						
3.3	Milestone 3: Market Research Design Specification	Ryan, Steve, Tim	2/6/20	2/10/20	4	0%																						
4	Project Performance / Monitoring																											
4.1	Project Objectives	Ryan	2/10/20	3/13/20	33	0%																						
4.2.1	Project Plan Updates	Ryan	2/10/20	3/13/20	33	0%																						
4.3	Implementation - Research Question 1 - Marketing Technique: xxx	Ryan	2/10/20	3/2/20	22	0%																						
4.3.1	XXX Model Data EDA	Ryan	2/10/20	2/15/20	5	0%																						
4.3.2	XXX Model Experimentation & Results	Ryan	2/17/20	2/22/20	5	0%																						
4.3.3	XXX Model Interpretation	Ryan	2/24/20	2/26/20	2	0%																						
4.3.4	XXX Model Recommendations	Ryan	2/26/20	2/27/20	1	0%																						
4.3.5	Milestone 4.1: Business Model XXX Report	Ryan	2/26/20	3/2/20	6	0%																						
4.4	Implementation - Research Question 2: XXX	Steve	2/10/20	2/15/20	5	0%																						
4.4.1	XXX Model Data EDA	Steve	2/10/20	2/15/20	5	0%																						
4.4.2	XXX Model Experimentation & Results	Steve	2/17/20	2/22/20	5	0%																						
4.4.3	XXX Model Interpretation	Steve	2/24/20	2/26/20	2	0%																						
4.4.4	XXX Model Recommendations	Steve	2/26/20	2/27/20	1	0%																						
4.4.5	Milestone 4.2: Business Model XXX Report	Steve	2/26/20	3/2/20	6	0%																						
4.5	Implementation - Research Question 3: XXX	Tim	2/10/20	2/15/20	5	0%																						
4.5.1	XXX Model Data EDA	Tim	2/10/20	2/15/20	5	0%																						
4.5.2	XXX Model Experimentation & Results	Tim	2/17/20	2/22/20	5	0%																						
4.5.3	XXX Model Interpretation	Tim	2/24/20	2/26/20	2	0%																						
4.5.4	XXX Model Recommendations	Tim	2/26/20	2/27/20	1	0%																						
4.5.5	Milestone 4.3: Business Model XXX Report	Tim	2/26/20	3/2/20	6	0%																						
4.6	Milestone 4: Final Project Report Presentation Deck	Ryan, Steve, Tim	3/6/20	3/13/20	7	0%																						













# Generated Data Modeling Table Examples

---

  dev\_brook.EstimatedOccupancy

 Columns

-  listing\_id (PK, int, not null)
-  est\_lifetime\_occ (smallint, not null)
-  est\_lifetime\_occ\_daily\_rate (float, not null)
-  est\_lifetime\_occ\_yearly\_rate (float, not null)
-  est\_lifetime\_income (float, not null)
-  est\_lifetime\_yearly\_income (float, not null)
-  est\_lifetime\_potential\_income (float, not null)
-  est\_lifetime\_potential\_yearly\_income (float, not null)
-  est\_perc\_yearly\_income\_of\_potential (float, not null)
-  quantile (tinyint, not null)



*Your best quote that reflects your approach... “It’s one small step for man, one giant leap for mankind.”*

- NEIL ARMSTRONG