Airbnb: Feel Like Home Wherever You Go

I. Business Justification

I wanted to get a better understanding of the business in Seattle area. Airbnb is an online platform that connects people looking to rent their homes with people who are looking for accommodations. I found a dataset about Airbnb in Seattle on Kaggle. Using this dataset, I want to answer the following questions:

- When is the busiest time to visit Seattle?
- Who is the top performer among the host? Is there a correlation between the hosts, their average estimated earnings, location, and review score?
- What was the sale growth in the last 5 years? Future sales projection?
- How is the price market of Airbnb in Seattle? Which area has the highest average price?

I used Tableau to address these questions because it will provide me more insights from the visualizations and predictive analysis that I could create in the tool.

II. Data preparation

The data originally had three tables: Listings, Reviews, and Calendar with approximated 3000 records from 2009 to 2016. After examining the data and joining these tables, I decided not to use the calendar table. This table have four columns: lisitng_id, date, available, and price. All of the price is the same $85 with some null values. Kaggle does not have explanations for the attributes and I could not really understand the meaning of this table or what its relationship with the other two tables.
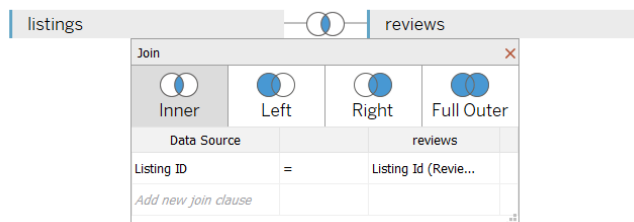
listings+ (SeattleAirbnb)

Fig 1: Inner Join on the listing_id

I joined Listing table and Reviews table by paring the listing_id. By doing this, I have each review matches with one listing. It is 1:M relationship because one listing can have multiple reviews. There was Date when the review was written and I interpreted that as the time series where the

transaction was completed between the host and the review. After joining these tables, the data dropped to 1001 rows.

There were originally 100 attributes and I discarded a lot of them. Many attributes were url pictures of the host, the reviewers and the places, which were not helpful for the project. There were many location attributes like state, city, smart location indicating Seattle, WA. I know this dataset is just about Seattle, so I hided them all. There were three different prices values: price per night, weekly price and monthly price. There were many null values in weekly and monthly prices so I decided not to use them. Here is the list of the dimensions and measures that I used for this project.

Dimensions
- Property Type
- Room Type
- Bed Type
- Neighborhood (Cleansed)
- Zipcode
- Availability for (30 days, 60 days and 90 days)
- Date (Review)

Measures
- Count of Listings
- Guest Included
- Price per night
- Minimum nights

Since the dataset did not have actual sales values, I created new measures with the following calculations:

Estimated Sales	×

[Guests Included] * [Minimum Nights] * [Price per night]

Estimated Profit	×

INT(SUM([Estimated Sales]-([Estimated Sales]*0.03) – ([Estimated Sales]*0.065) – ([Estimated Sales]*0.3)))

Estimated Monthly Earni	×

([Estimated Profit]/12)

Fig 2: Calculations for Estimated Sales, Estimated Profit, Estimated Monthly Earnings

- Airbnb fee is 3% of the transaction
- Washington Retail Sales Tax is 6.5%
- Income tax is 30%

III. Data Visualizations

I created three dashboards, including DB-SalesPerformance, DB-ListingFinder and DB-TopPerformance.
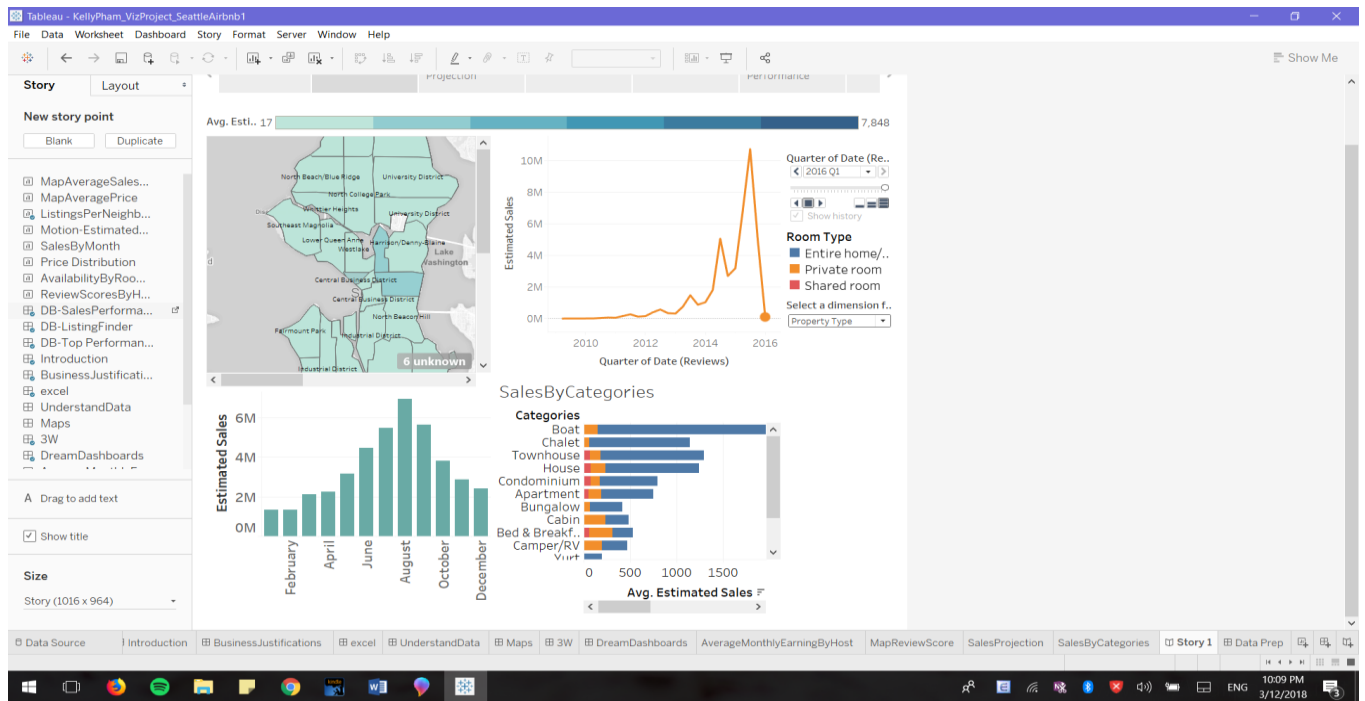
1. Dashboard # 1: Sales Performance

Fig 3: Dashboard #1 – Sales Performance

For dashboard #1, I created a map to demonstrate the average sales by different areas in Seattle. It shows that Beltown ($2012), North Beacon Hill ($636), and Arbor Heights ($2584) had the highest average sales from 2009 to 2015. For this visualization, I used the Zipcode, Estimated Sales, and Neighborhood to label the areas. I imported a map background from Mapbox in the "Light" style so that I can show the color effects of sorting different values in different neighborhoods.

There was a bar chart showed that the June, July, August and September were when Seattle had the most visitors. Over the 5 years period, August hit a total estimated sale of almost 7M. January and February had the lowest sales of approximately 1.3M.

The dashboard also had a motion line chart to show the sales growth from 2009 to 2016. The motion chart showed that there was a steady growth in sales from 2009 to 2014. The sales spiked in 2015 at approximately 10M in 2015. Airbnb was founded in 2009, so it probably picked up its popularity in 2014 and 2015. That was why there was such a big increase in 2015. The sales dramatically dropped in 2016 because the data only included two transactions in the first quart of 2016. It was a simple motion chart. I dragged the Date (Review) to Pages, used the continuous right lines to do the motion line for the estimated sales.

Dashboard #1 also had another bar chart to show sales by different categories, including Room Type, Bed Type, and Property Type. I created a parameter so that I could switch to different types of categories on the same bar chart
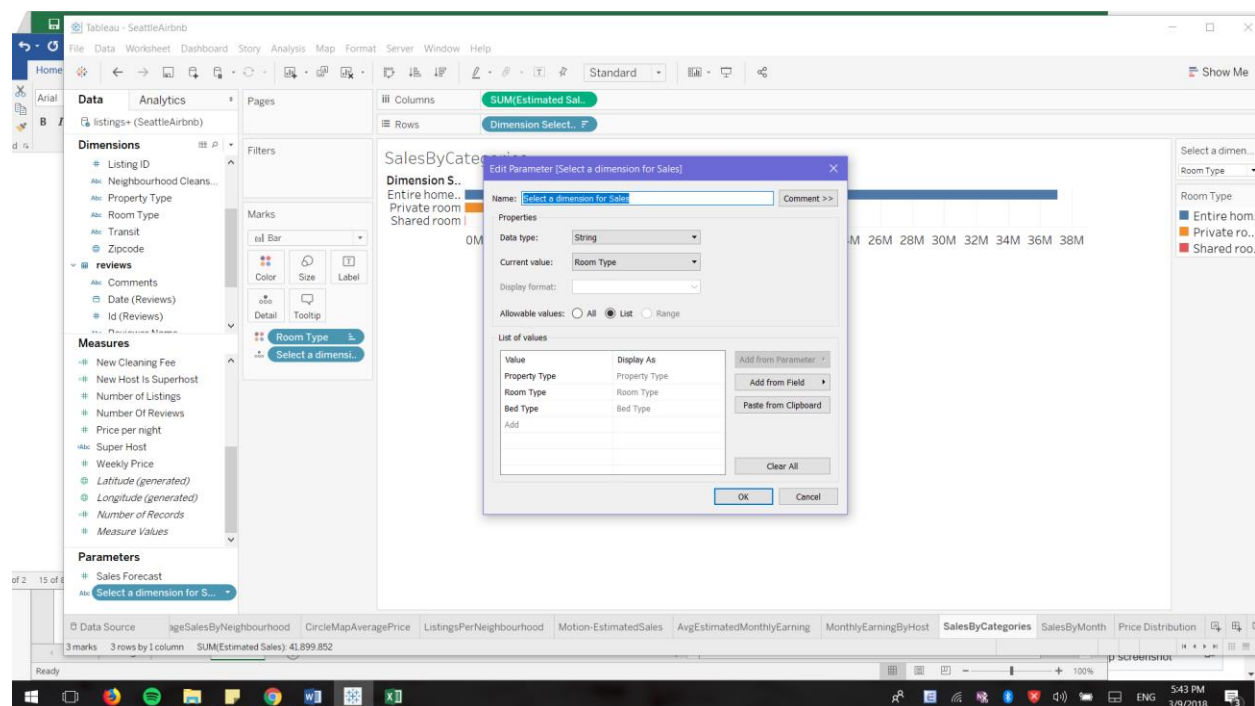

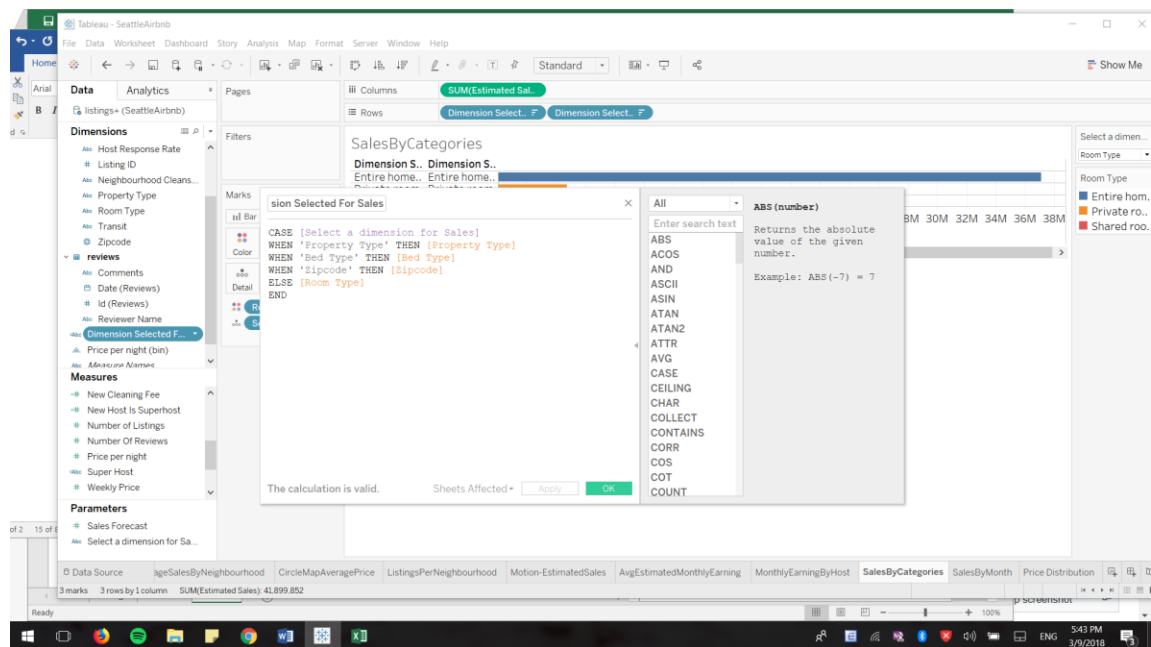*Fig 4: Parameter – Select a Dimension for Sales*

Fig 5: Calculated Field for Parameter

The results of this bar chart showed that Real Bed was the most popular option, following by: Airbed, Pull-out Sofa, Futon and Couch. In term of Room Type, Entire home/Apartment was the most favored as its average estimated sales was $707.20 over the 4 years period. This was followed by Private Room ($152.20) and Shared Room ($63.80). Under the category of Property Type, I was surprised to see Boat on top of the list with an average estimated sale of $6,772. I anticipated that House or Apartment would have the highest sales, but they came in third ($1014) and sixth ($557) places on the list. Townhouse was the second place with an average estimation of $1,115 in sales. I did not expect such a big difference between Townhouse and Boat by $5,657. Seattle has many boat docks, boats and boathouses and this is probably a big attraction for the visitors to get this experience in the summer time. The last two ranks in the list belonged to Tent ($65) and Dorm ($38). This bar chart was divided each Property Type in different colors to show the sales of Room Type within a certain Property Type.

I also created a visualization to project the sales, which was not included in the dashboard #1. This visualization was named "SalesProjection". The purpose of this visualization was that if I picked a specific date, I could predict the sales to increase or decrease by choosing a certain percentage for my prediction. This could be helpful for the sales manager at Airbnb to look at and adjust the marketing strategy accordingly. I created two parameters: one to pick a date, and one to choose a percentage for the forecast.
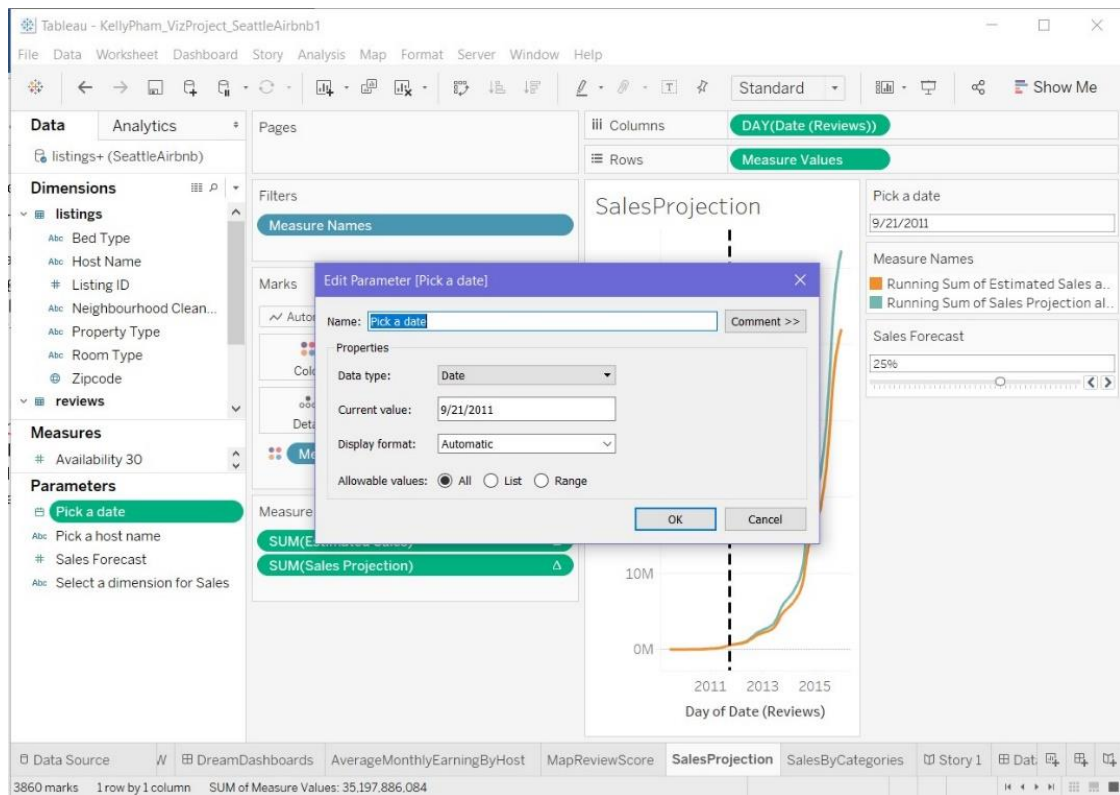
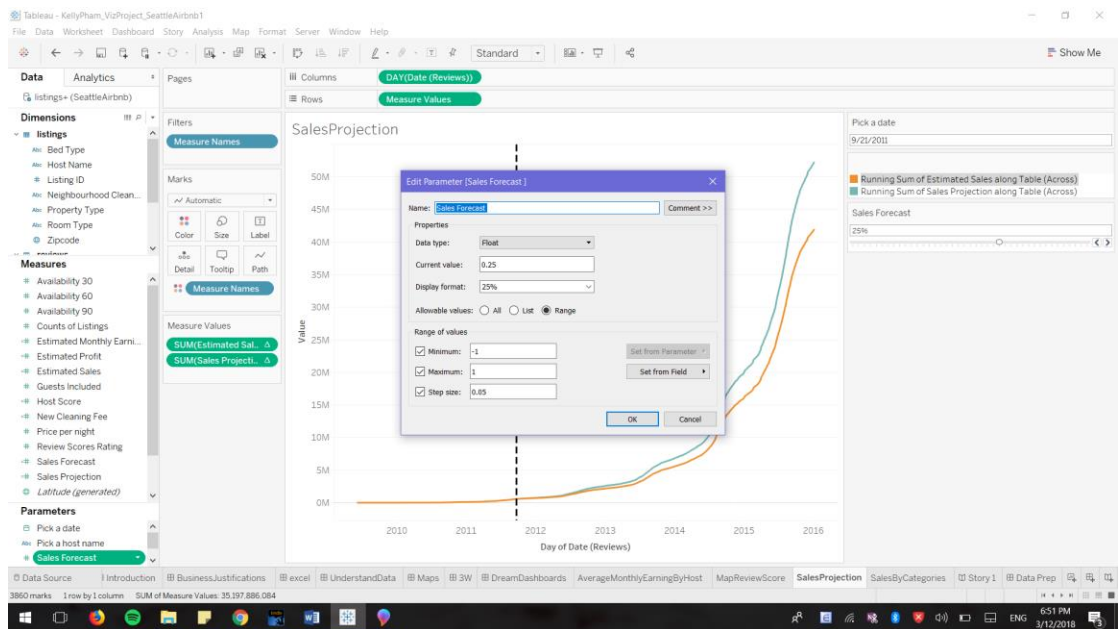*Fig 6: "Sales Forecast" Parameter*
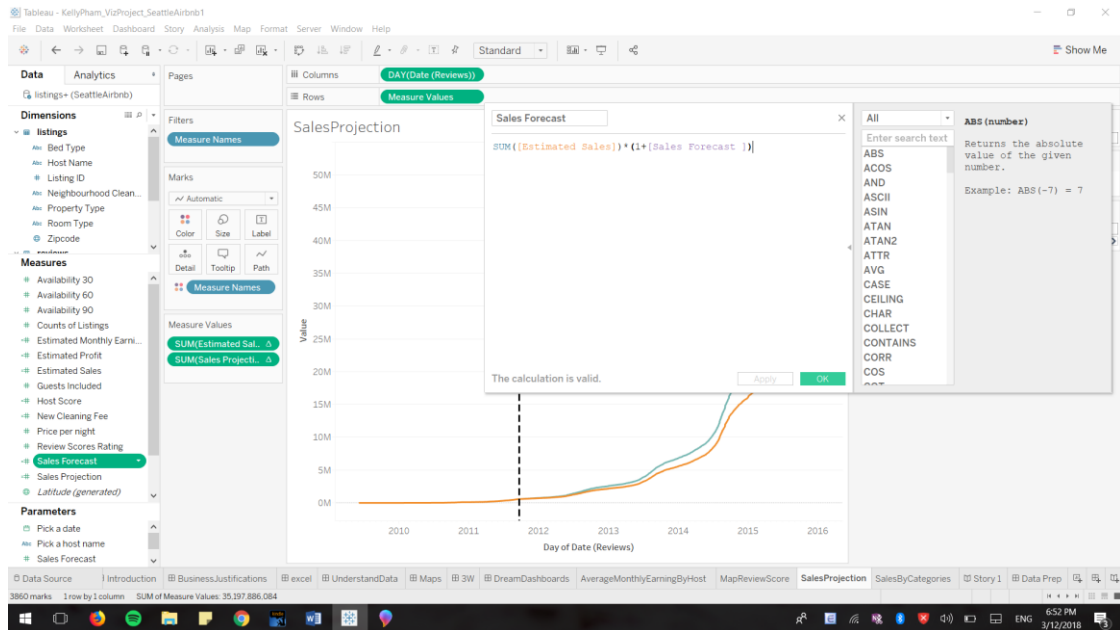

Fig 7: "Sales Forecast" Calculation

Fig 8: "Sales Forecast" or the reference line calculation.
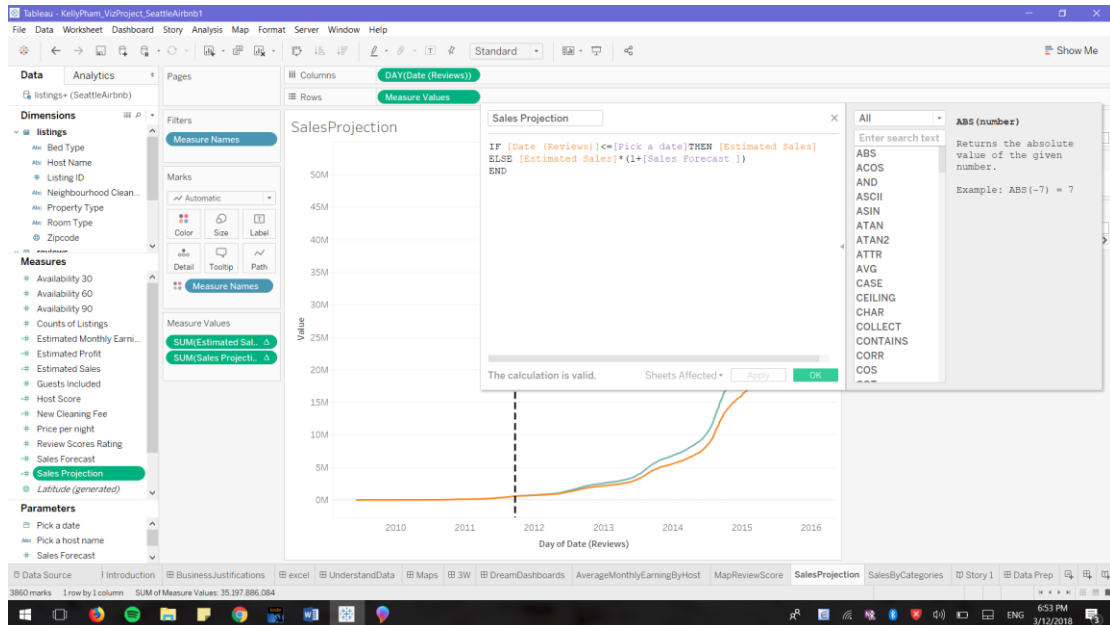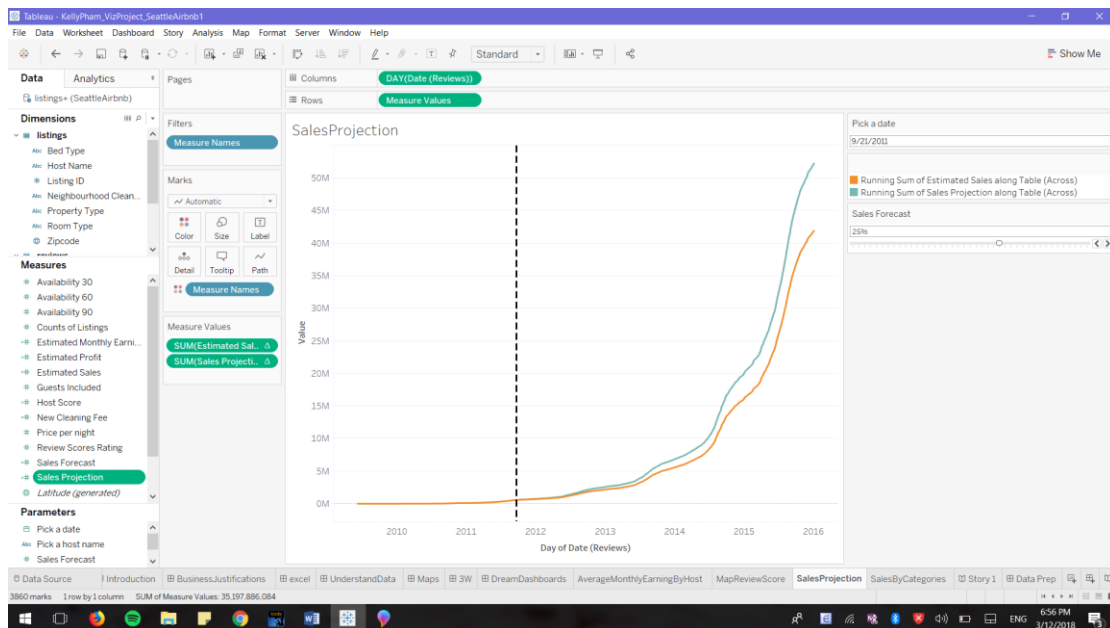

Fig 9: *"Sales Projection" or the reference line calculation.*

In this chart, the dotted black line is a reference line showing the date chosen. In this screenshot, it was 9/21/2011 and I was projecting that sales would increase 25% from this day. This created a dynamic view for the users to predict sales, and if they combined this with the bar chart "Sales By Categories", they could give out incentives to certain type of property owners to increase the number of supplies.

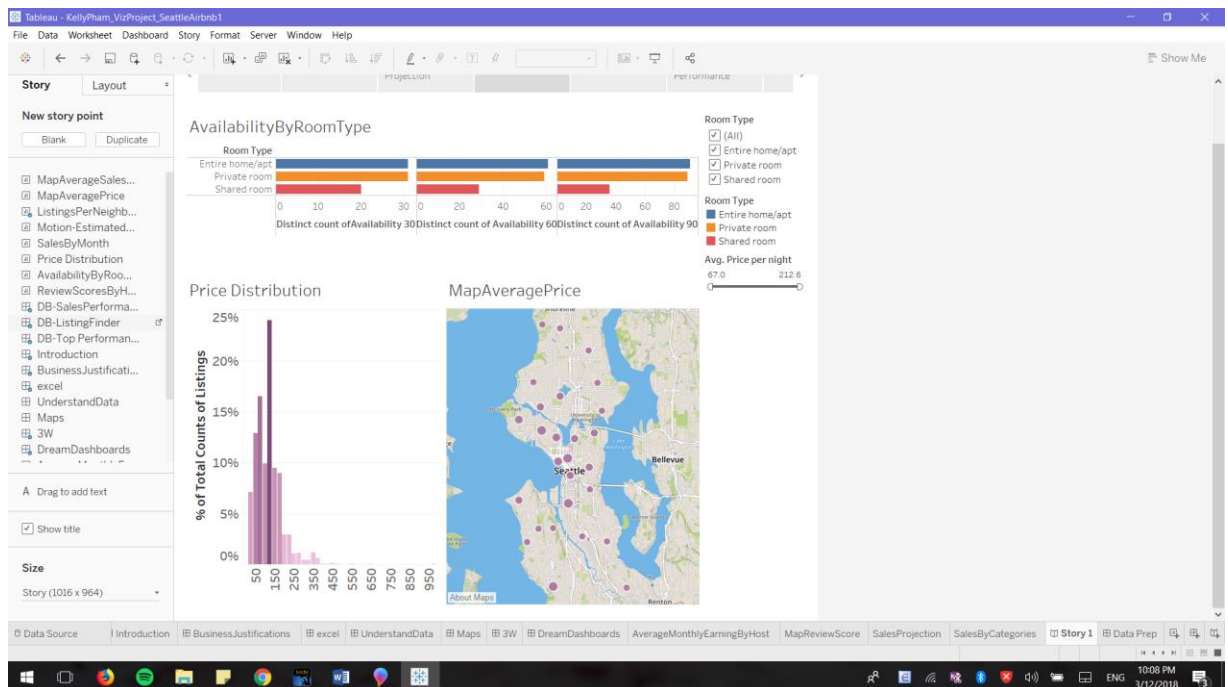2. **Dashboard #2 –Listing Finder**

s

Fig 10: Dashboard #2 – Listing Finder

This second dashboard showed us information about the prices of listings in different neighborhoods. I created price bins and then a price distribution. I did this by right clicking on "Price per night", chose "Create" and then "Bins"
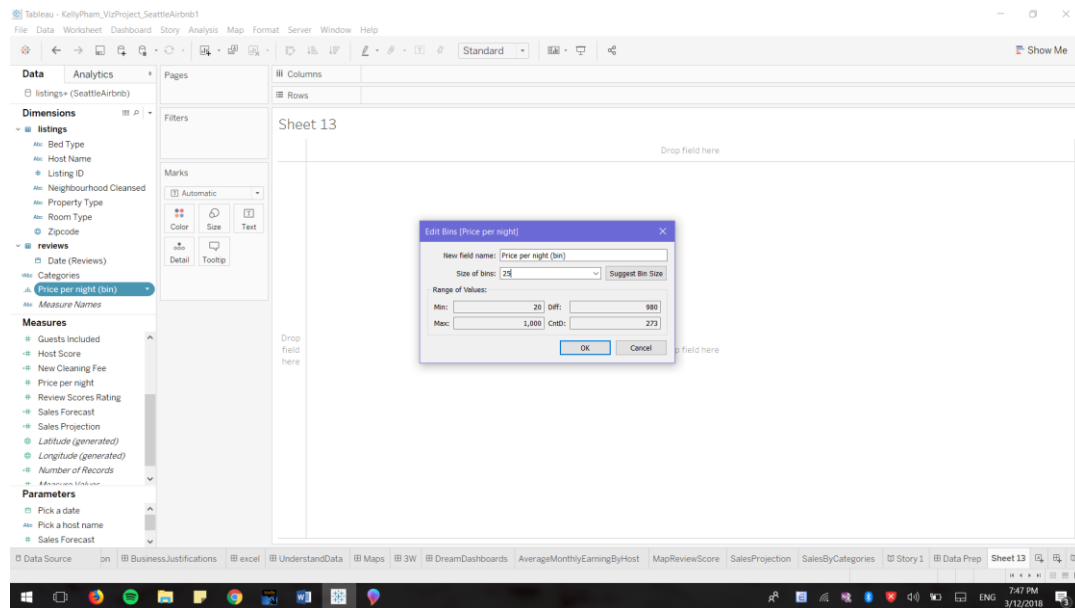

Fig 11: Create bins for pricing

Based on the Price Distribution chart, I could see that most of the listing range from $50 to $125 per night. The price of $125 accounted for 23.95% of total sum of listings. This dataset had a couple outliers of $550, $650, $975. These values are very were only 0.04% of the total listings

I also created a map with a price filter to see the average price per night in different neighborhoods. The average price of $125 focused mostly in the zipcode 98199, Magnolia neighborhood.

I also created a bar chart to show the different listings based on the availability. Some hosts provided their places for 30 days, 60 days and 90 days. I added a filter "Room Type" for this chart so that users can look for the listing based on the type of room. I did a distinct count of the listings because some listings were repeated multiple times in the data since one listing can have multiple reviews. Many listings were available from 30 days to 90 days.

In addition, I created a treemap visualization to show the total number listing available in different neighborhoods. Belltown and Broadway had the most listings of 16, and it was followed by Central Business District (14), University District (10) and Wallingford (10).

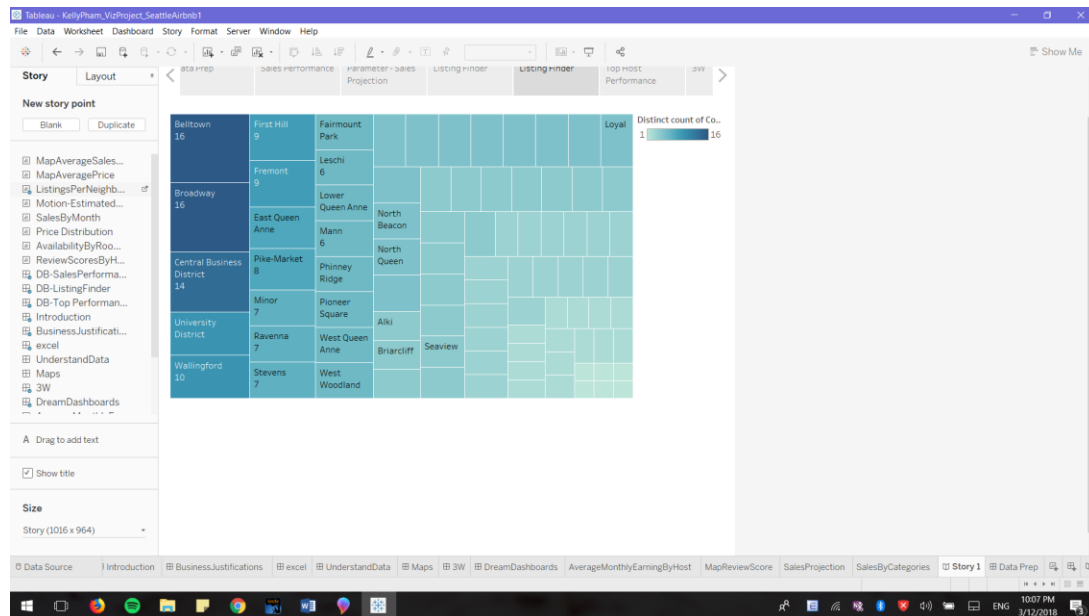Other areas that had a good number of listings were First Hill, Freemont, Queen Anne, Ravenna and Stevens.

Fig 12: The Treemap visualization

IV. Closing thoughts

1. What went well?

- Tableau was a very friendly-user and intuitive tool to work with. There were many tutorials available and many learning discussion boards to help debug the issues.
- Once I learned a feature in Tableau it was easy to recreate and build on top of it. For example, once I was able to get motion working on one spreadsheet it was not that much harder to get it coordinated with two or three. This was exciting, and made for a more interactive experience. Additionally, there were often times when I would accidentally mess something up and I would have to recreate it from scratch. I am quite proficient at creating custom icons on charts now.

2. What went wrong?

- Tableau made my computer super slow, especially after downloading the beta version, which caused quite a few headaches!
- I created all my dashboards to look fantastic, and then put them in a story. When I watched the story, I was so upset because all of my formatting I had just worked so hard to get perfect had to be redone with the sheet at the "Automatic" size. This caused for a lot of rework that ultimately didn't even translate into the online publishing. I wish that Tableau had a better process for this.

3. **What to improve?**
   - Explore more with custom images. I wanted to get images into a tooltip – and even downloaded a Viz with an example of this but I was unable to get it to work.
   - Pull in global data- It would have been interesting to see how Seattle market is compared to others, or if the trends I discovered in popular type of property was consistent.
   - Have a web link connected with the map or Viz within a Viz. It could be interesting to have a web link of weather data to see if there is a correlation between the weather and the numbers of sales.