AIRnalysis: How does Airbnb impact our lives?

The online Google Doc can be accessed here.

Team Member

Shu Xu, shuxu@g.harvard.edu
Shuyuan Xiao, sxiao@g.harvard.edu
Zheyu Wu, zheyuwu@g.harvard.edu

Table of Content

Team Agreement

- Communication protocol:
    - Communication protocol:
    - All communications should be made through the project Slack Channel
    - All communications should be clear and frank but with respect. We value any ideas from anybody in and outside of the team.
- Meeting logistics
    - Every Wednesday 12:30 pm - 1:30 pm at Maxwell Dworkin 1st floor lobby
- Code collaboration
    - We will mainly use GitHub for code version control and storage.
    - GitHub Repository: https://github.com/wuzheyu/airnalysis
- Team roles
    - Decider: Shuyuan
    - Designer: Shu
    - Programming: Shuyuan, Zheyu, Shu
    - Even though roles are assigned differently, all team members are expected to involve all technical aspects of the project. Any exceptions should be communicated beforehand to allow effective task coordination.
    - Only one decider is listed here but the final decision should still be discussed among all the team members.
    - All the programming deliverables should be well-documented to allow future maintenance and smooth work transition between team members.
    - We will try to set project goals in a way that makes sure all team members have approximately equal contribution. All team members should be responsible for the assigned work.
- Non-performing team members
    - We will meet regularly to make sure all team members deliver assigned work on time. However, for non-performing team members, we could adjust the workload for future milestones accordingly. If future milestone is not still delivered on time by the team member, we will schedule a meeting with the course staff to discuss about next steps.

Signatures

Zheyu Wu	Shu Xu	Shuyuan Xiao
		

Date
11/9/2019

Week 10 (Nov 3 - Nov 9)

Draft Project Proposal (Nov 4)

Motivation

NOTE: our original project title is called “AIRnalysis: How can Airbnb benefit all three parties - guests, hosts and the company?”.

Airbnb is an online platform for offering lodging and tourism experiences. Its recent announcement of planning to go public in 2020, as well as its market value of over $31 billion have made the public wonder its exact business model and the path to success. Over the years, there is an abundance of Airbnb’s data online, and visualizing them is one of the most efficient ways for the public to fully and quickly appreciate Airbnb’s business model. There are existing visualization tools, for example, Airbnb’s own Smart Pricing Tool and Inside Airbnb. However, they only provide isolated data points from the host’s perspective. The lack of aggregated data and data from the guest’s perspective make it challenging to draw insightful conclusions that could potentially benefit all three parties - guests, hosts and the company.

Description

In this project, we will explore the business model of Airbnb and consider the benefits of both home-owners and guests. On the home-owners’ side, we will examine how the prices of homes on Airbnb compare with prices of apartments and houses for lease or rental. On the guests’ side, we will inspect how the expenses of using Airbnb compare with hotel prices.

Project Goal

In this project, we would like to use a data-driven approach to first visualize data from the company (Airbnb), the host and the guest perspective (in the scope of Boston and Cambridge area). Based on these visualizations, we would like to expose the underlying business model, and uncover data patterns that would be beneficial in:

- Helping guests locate cost-effective housing listings
- Helping hosts in defining the price range and maximizing profit
- Helping the company understand insightful data patterns and make business decisions

Project Task

To summarize, we have the following tasks in mind:

- Dataset collection
- Data cleaning
- Preliminary exploratory data analysis

- High-level website skeleton design
- Detailed view design with following views:
    - Airbnb's growth in the last decade from different perspectives (e.g. revenue, profit, number of employees etc)
    - Relationship of listing price and important features (location, #bedrooms, #beds, etc), by narrowing down to the Boston area. A choropleth will be presented.
    - Comparing Airbnb listing with traditional rental in host's view.
    - Comparing Airbnb listing with hotel room in guest's view.
- Formulating narratives
- Integrating visualizations and narratives
- Publishing the website

Dataset

There are a number of possible data sources. For data from Airbnb, we will use Smart Pricing Tool and Inside Airbnb, as mentioned above. With the purpose of comparing the prices of homes on Airbnb with normal rental prices, we will extract rental or leasing data from Zillow (https://www.zillow.com/research/data/) and ApartmentList (https://www.apartmentlist.com/rentonomics/rental-price-data/). Data files like csv files could be directly downloaded. For hotel prices, we will consider scraping data from tripadvisor's website. A possible guidance for scraping is here. https://www.scrapehero.com/how-to-scrape-tripadvisor/

Project Detailed Plan (Nov 10)

Active Team Members

All three of us have actively worked on the plan: Shu Xu, Shuyuan Xiao, and Zheyu Wu.

Project Goal

Airbnb is an online platform for offering lodging and tourism experiences. Its recent announcement of planning to go public in 2020, as well as its market value of over $31 billion have made the public wonder its exact business model and its impact to the society as a whole. Over the years, there has accumulated an abundance of Airbnb's data online, and visualizing them is one of the most efficient ways for the public to fully and quickly appreciate its impact. There are existing visualization tools such as Inside Airbnb. However, they only provide data points solely from the host's perspective to help adjusting listing prices. The lack of data from the other perspectives makes it challenging to draw insightful conclusions about Airbnb's growth path and its extensive impact to the society.

In this project, we would like to use a data-driven approach to visualize data from the lens of the company (Airbnb) itself, the host, the guest and other competitors to provide a glance of its impact to the society. Based on these visualizations, we would like to expose the underlying business model, and uncover data patterns that would help the audience to appreciate Airbnb's social impact and possibly extend the conclusion to other forms of shared economy.

Project Task

Our end goal is to build a website that composes of data visualizations, texts and animations to tell a complete story about the Airbnb ecosystem. To achieve this goal, we have divided the work in three domains: data cleaning and collection, visualization, and storytelling. To be more specific, we have the following detailed tasks in mind:

- Dataset collection/cleaning:
    - Download all the needed datasets and perform analysis to find any missing data and/or outliers.
    - Necessary cleaning and aggregation
- Preliminary exploratory data analysis to determine the best measures to be visualized as well as the corresponding appropriate chart type
- Individual visualizations including the following views:
    - Airbnb's growth in the last decade from different perspectives (e.g. revenue, profit, number of employees etc)
    - Airbnb listing price distribution with a map of Boston area and filter bars (location, #bedrooms, #beds, etc). A choropleth or map with bubbles will be presented.

    - Comparison of Airbnb listing with traditional rental in the host's view.
    - Comparison of Airbnb listing with hotel room in the guest's view.
    - How Airbnb has affected the hotel industry.
- Test, evaluate and improve.
- Formulate narrative story-telling.
- High-level webpage skeleton design
- Integrating visualizations and narratives to the webpage
- Adjust format according to C.R.A.P

Dataset

There are a number of possible data sources. For data from Airbnb, we will use Inside Airbnb as mentioned above. With the purpose of comparing the prices of homes on Airbnb with normal rental prices, we will extract rental or leasing data from Zillow (https://www.zillow.com/research/data/) and ApartmentList (https://www.apartmentlist.com/rentonomics/rental-price-data/). Data files like csv files could be directly downloaded. For hotel prices, we will consider scraping data from tripadvisor's website. A possible guidance for scraping is here. https://www.scrapehero.com/how-to-scrape-tripadvisor/

Individual dataset attributes can be found below:

Airbnb Dataset

- Inside Airbnb
    - We have datasets for calendar data, review data, neighborhood data, and GeoJson files. We will choose to use a few of them
    - GeoJson file contains the geography information needed to render the choropleth
    - Dataset for ratings (reviews.csv.gz)
        - Columns: id (int), listing_id (int), date (timestamp), reviewer_id (int), reviewer_name (string), comments (string)
        - We will use this dataset to render the visualization for comparison living experience and quality between hotels and Airbnb homes.
    - Calendar dataset contains the pricing information
        - Columns: listing_id (int), date (timestamp), available(boolean indicating if available), price, adjusted_price, minimum_price, maximum_price (all prices are doubles)
        - We would use this dataset to compare the price differences between hotels and Airbnb homes.

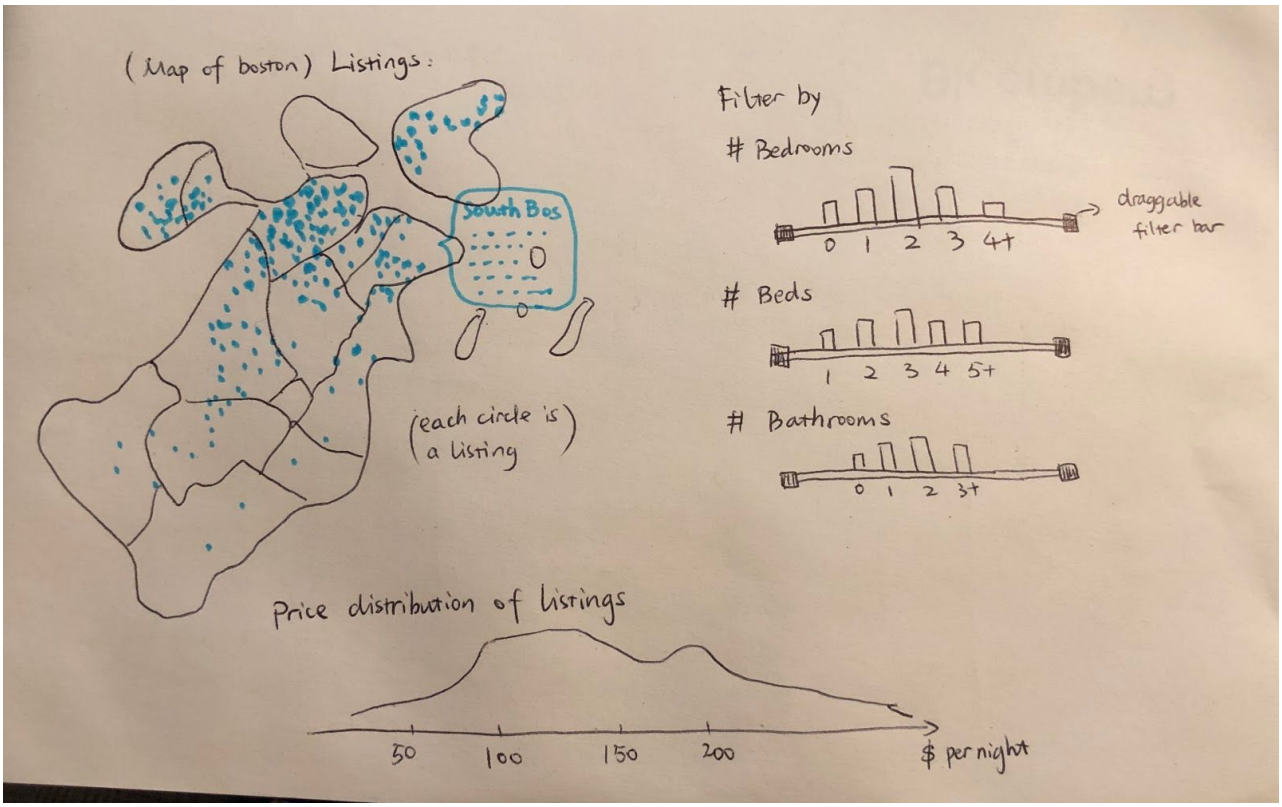
Rental or Lease Data:

- Zillow (https://www.zillow.com/research/data/)
    - Rental Data Set:
        - Columns: City, State, Metro, CountyName (all strings), Size (numeric), and Rental Prices from 2010-02 to 2019-09 (all doubles)
        - Rows: Different regions / areas
        - We will use a subset of the dataset: select recent data from Boston / Cambridge area
- ApartmentList (https://www.apartmentlist.com/rentonomics/rental-price-data/).
    - Rental Data set2:
        - Columns: Location (string), Year, Month, Price by types [Studio, 1br, 2br, 3br, 4br, etc.] (numeric)
        - Rows: Different regions / areas
        - We will utilize a subset of the dataset: select certain types of housings from Boston / Cambridge area in recent time.

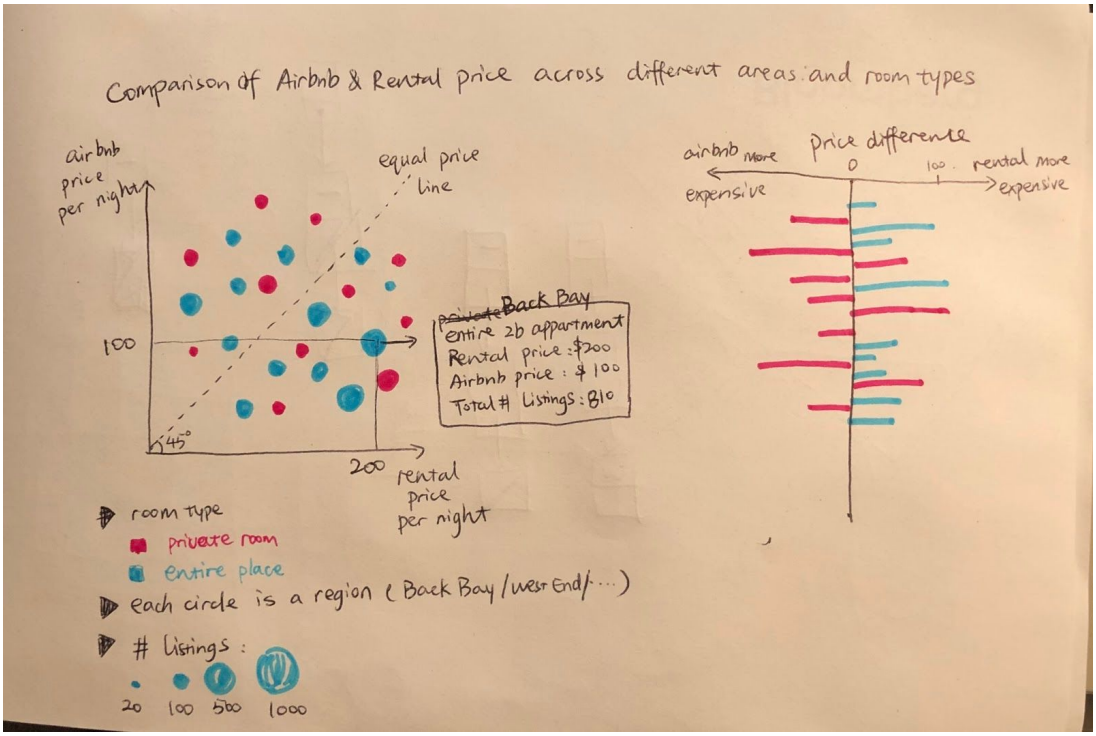
            - https://public.opendatasoft.com/explore/dataset/airbnb-ratings/export/?disjunctive.city&disjunctive.neighbourhood_cleansed&q=boston&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6ImFpcmJuYi1yYXRpbmdzIiwib3B0aW9ucyI6eyJkaXNqdW5jdGl2ZS5jaXR5Ijp0cnVlLCJkaXNqdW5jdGl2ZS5uZWlnaGJvdXJob29kX2NsZWFuc2VkIjp0cnVlLCJxIjoiYm9zdG9uIn19LCJjaGFydHMiOlt7ImFsaWduTW9udGgiOnRydWUsInR5cGUiOiJsaW5lIiwiZnVuYyI6IkFWRyIsInlBeGlzIjoiaG9zdF9saXN0aW5nc19jb3VudCIsInNjaWVudGlmaWNEaXNwbGF5Ijp0cnVlLCJjb2xvciI6IiNGRjUxNUEifV0sInhBeGlzIjoibGFzdF9yZXZpZXciLCJtYXhwb2ludHMiOiIiLCJ0aW1lc2NhbGUiOiJ5ZWFyIiwic29ydCI6IiJ9XSwiZGlzcGxheUxlZ2VuZCI6dHJ1ZSwiYWxpZ25Nb250aCI6dHJ1ZX0%3D
            - https://www.hotels.com/articles/ar000351/where-to-stay-in-boston-a-travel-guide-to-boston-s-neighborhoods/ (for hotel prices by neighborhood)

Sketches of visualization ideas:

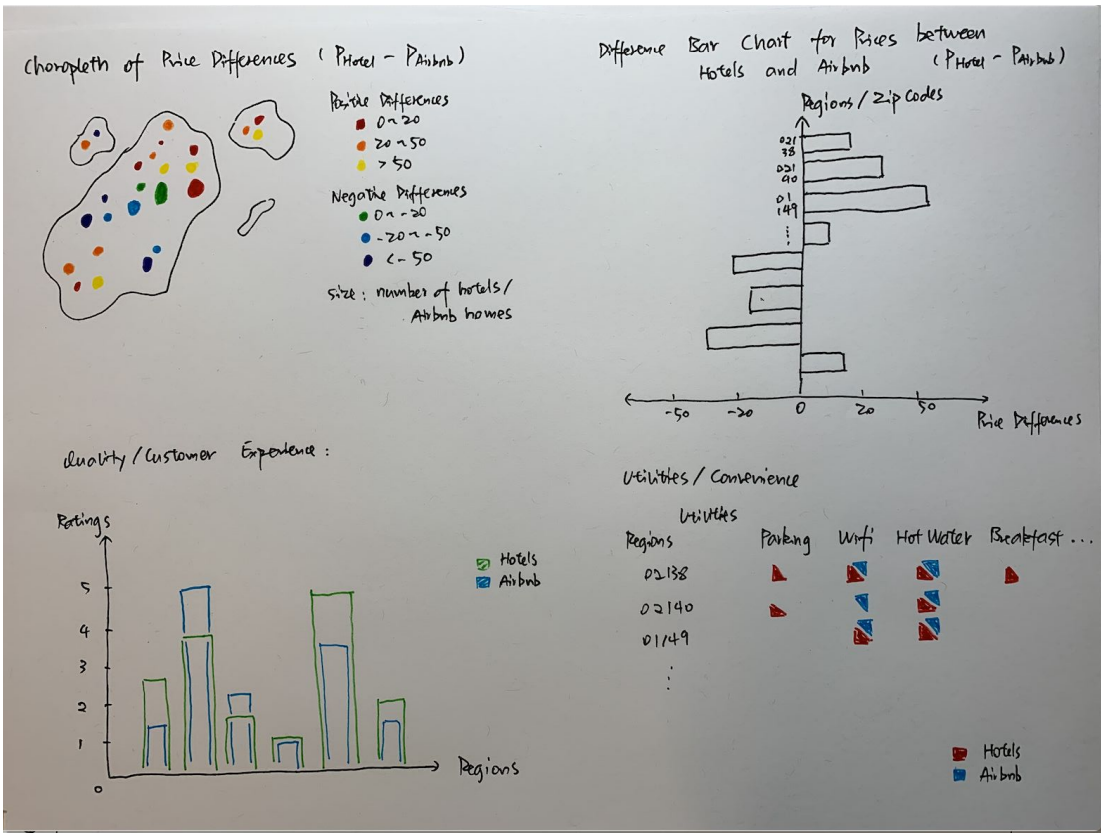
- Map Overview


(Map of boston) Listings:
South Bos
(each circle is a listing)
Filter by
Bedrooms
0 1 2 3 4+
draggable filter bar
Beds
1 2 3 4 5+
Bathrooms
0 1 2 3+
Price distribution of listings
50 100 150 200
$ per night

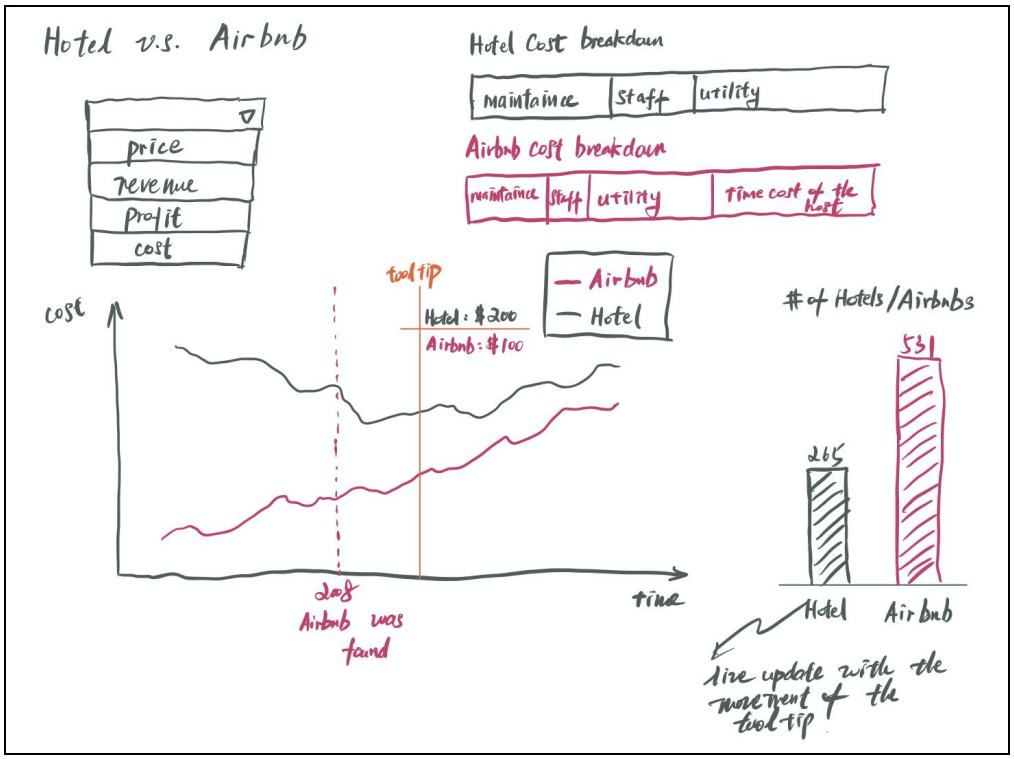
- Host


Comparison of Airbnb & Rental price across different areas and room types
airbnb price per night
equal price line
100
200
45°
rental price per night
~~private~~ Back Bay
entire 2b appartment
Rental price: $200
Airbnb price: $100
Total # Listings: 810
room type
private room
entire place
each circle is a region (Back Bay / West End / ...)
listings:
20 100 500 1000
price difference
airbnb more expensive
0
100
rental more expensive

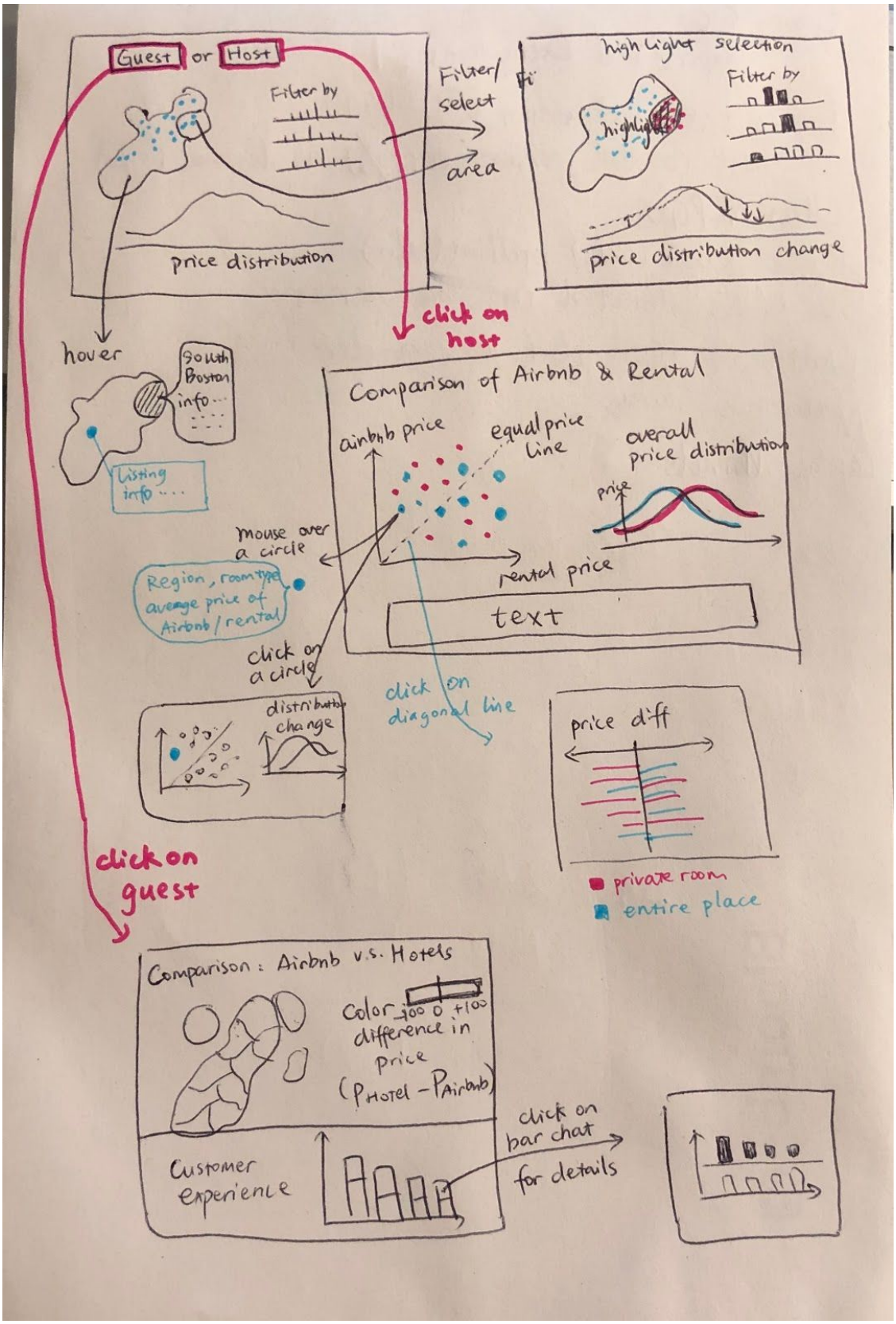
- Guest



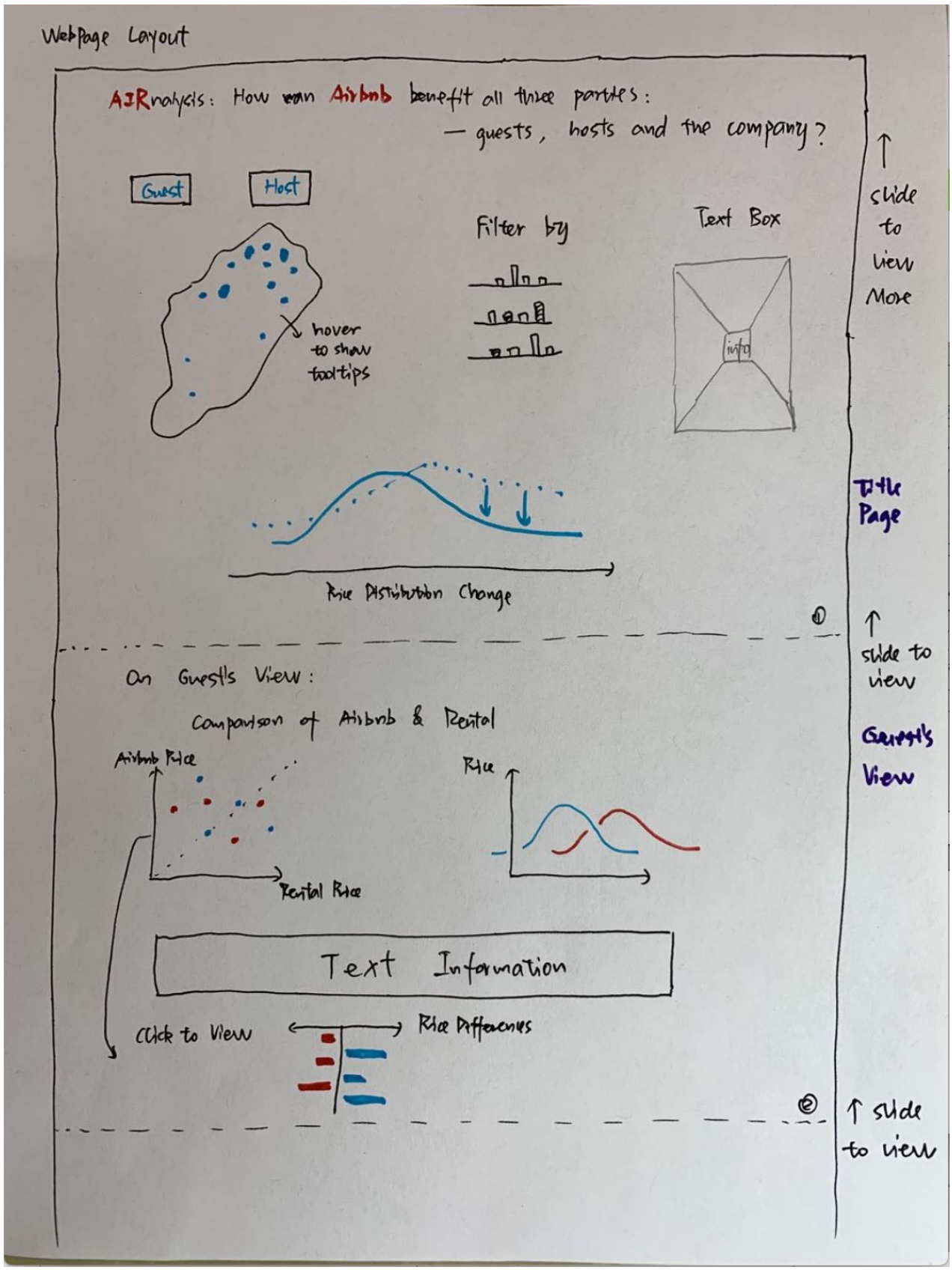
- Other hotels

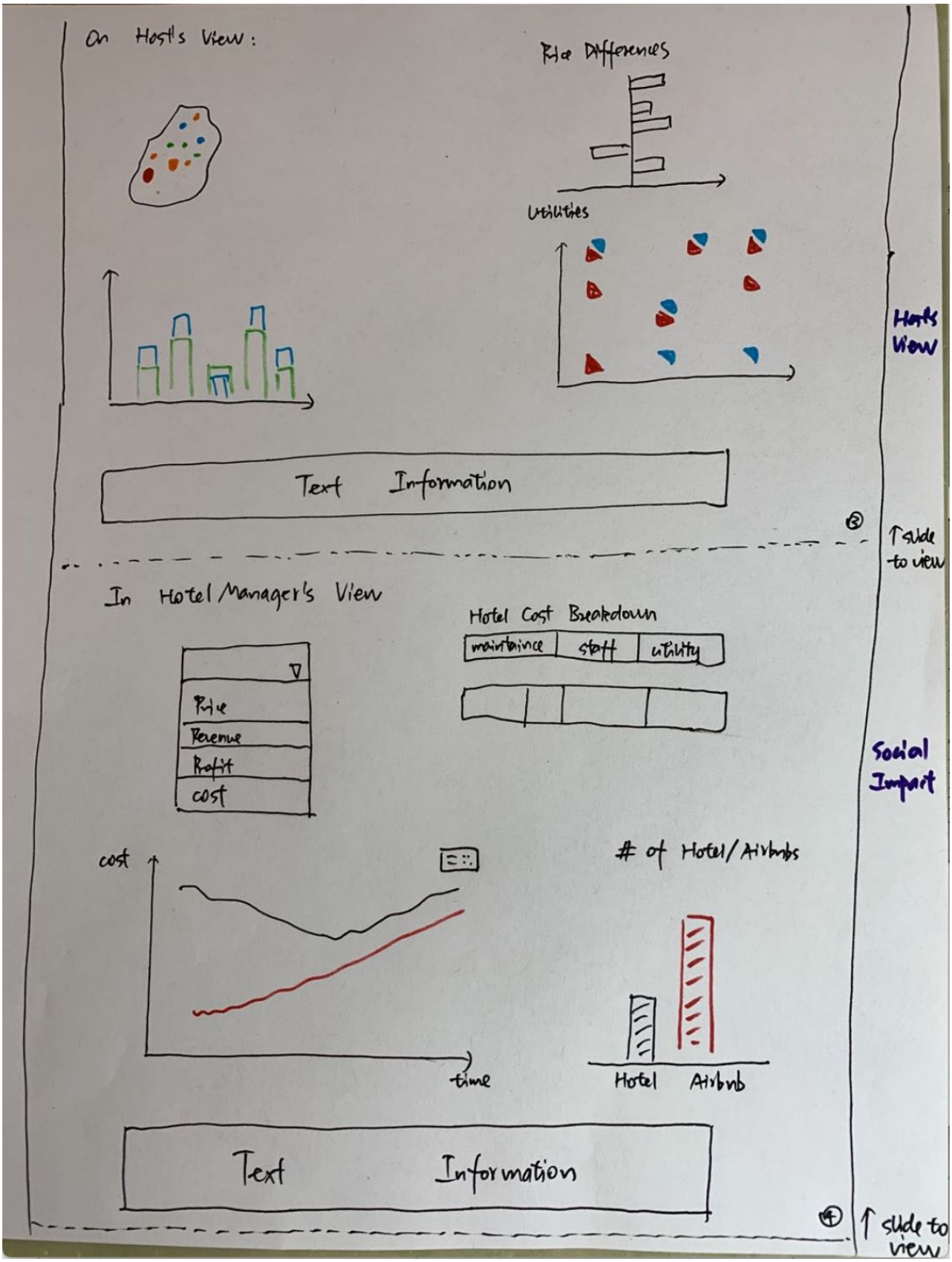

Hotel v.s. Airbnb
price
revenue
profit
cost
Hotel Cost breakdown
maintaince
staff
utility
Airbnb cost breakdown
maintaince
staff
utility
time cost of the host
cost
tooltip
Hotel: $200
Airbnb: $100
Airbnb
Hotel
time
2015
Airbnb was found
# of Hotels/Airbnbs
265
531
Hotel
Airbnb
line update with the movement of the tooltip

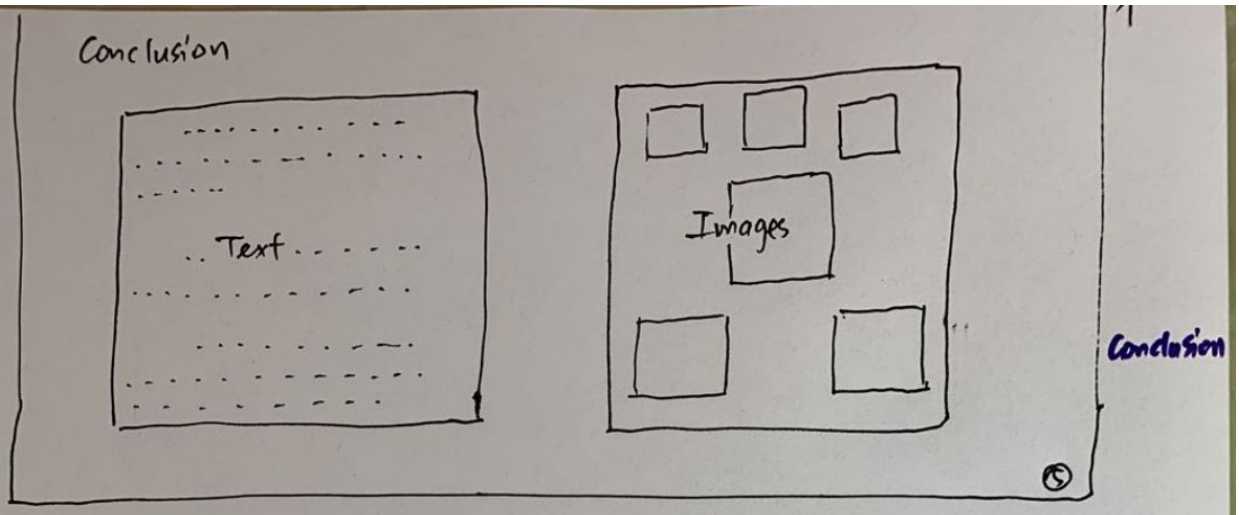
Sketch of the interaction storyboard:



Sketch of the **webpage layout/storytelling**




On Host's View:
Price Differences
Utilities
Text Information
Host's View
slide to view
In Hotel Manager's View
Hotel Cost Breakdown
maintaince
staff
utility
Price
Revenue
Profit
cost
Social Impact
cost
time
# of Hotel/Airbnbs
Hotel
Airbnb
Text Information
slide to view



The order of Page ② (On Guest's View) and Page ③ (On Host's View) could be switched, based on user's choice on Page ① (Title Page).

Project timeline (with milestones when you are planning to finish which feature)

Milestone Deadline	Deliverable
11/17/2019	● Data needs to be downloaded and fully cleaned (Shuyuan) ● A clear storyline must be formulated (Zheyu) ● With the cleaned data, some preliminary data visualization should be performed (Shu)
11/24/2019	● Any features under “must-have” bullet below must be implemented and fully tested (All) ● HTML skeleton should be up and running (Zheyu) ● Preliminary visualizations (see “Project Task” above) must be completed (All)
12/1/2019	● Any features under “Good-to-have” bullet below should be implemented and tested (All) ● The entire website, including narratives and layout design (“C.R.A.P.” rules), should be up and running (All)
12/8/2019	● Clean up the webpage, push the code (Zheyu) and if time permits, features under “Optional” bullet below should be implemented and tested (All)

Feature list (with must-have, good-to-have, and optional items)

- Must-have
    - A page with the project introduction and animations
    - A page with animated graph showing the global growth of Airbnb
    - A page with a selection box that allows the user to select to be a guest/host.
    - On the guest’s view page, map and filter bars to show what factors are related to airbnb listing price

    - On the host's view page, visualization of airbnb and traditional rental differences across different property types and regions.
    - A page with visualizations to show Airbnb's impact to the traditional hotel industry.
    - A conclusion page with any findings and takeaways
- Good-to-have
    - How price or other factors differ during different times of a year. (timeline bar for any or all of the visualizations)
    - Visualizations to compare the quality and living experience between hotels and Airbnb homes via ratings, customer feedback, as well as the convenience of living (utilities: parking lot, Wifi, hot water, etc. environment: bus, subway, train, restaurants, etc)
- Optional
    - Calculating results of host's net income compared with rental with invisible costs of hosting an airbnb listing taken into account.
    - Given host and guest statistics, calculation of the range of reasonable price with a machine learning model running in the back end
    - Semantic analysis of reviews provided on the Airbnb by guests

Description of team roles

For this project, we are assigning roles based on each team member's speciality. Shuyuan will be our decider when discussions couldn't reach a clear result. Only one decider is listed here but the final decision should still be discussed among all the team members. Shu will be our main designer. All three of us, Shuyuan, Zheyu, and Shu will need to accomplish a fair proportion of programming work. Zheyu will also be responsible for the story writing.

Even though roles are assigned differently, all team members are expected to involve all technical aspects of the project. Any exceptions should be communicated beforehand to allow effective task coordination.

Week 11 (Nov 10 - Nov 16)

To-do: Prototype V1 - Individual Visualizations (Due on Nov 17)

Meeting with the TF

- Communicate with each other frequently
- Team work. Everyone does similar things. Peer review process at the end of the project.
- Project is interesting. Look at different prices; state; country; distance to different facilities; scraping can take a long time;
- Bring novelty to the design/implementation. A good example is Lab 8. Take a look at glyph;

Active Team Members

All of us, Zheyu, Shu and Shuyuan contributed actively to prototype V1 submission.

Data

All of the data that will be used in this project have been refactored and cleaned. The cleaned data can be found in the “data” folder of our project.

Storyline

- Eric is a hotel manager in New York and he recently got married to Emily.
- They would like to travel to Boston for a short trip. They would like to try out Airbnb as one of their friends recommended it. They travelled to Boston and had a good time staying in the Airbnb (guest’s view).
- When they returned home in New York, Emily would also like to start an Airbnb business as there was an empty bedroom in their house (host’s view).
- As a hotel manager, Eric would also like to discover Airbnb’s impact on the traditional hotel industry (hotel competitors’ view).

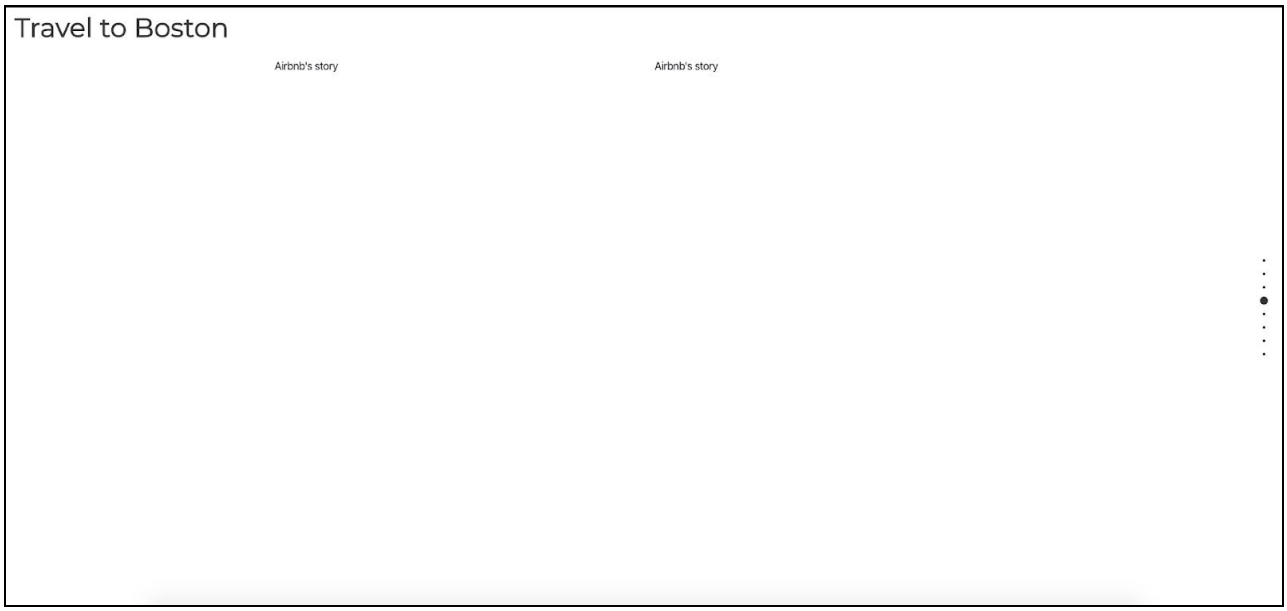
Rough Webpage Design

- The first page clearly states our project title. The background is a rental apartment which matches our topic. Notice the navigation panel on the right allows the user to quickly navigate through different pages.

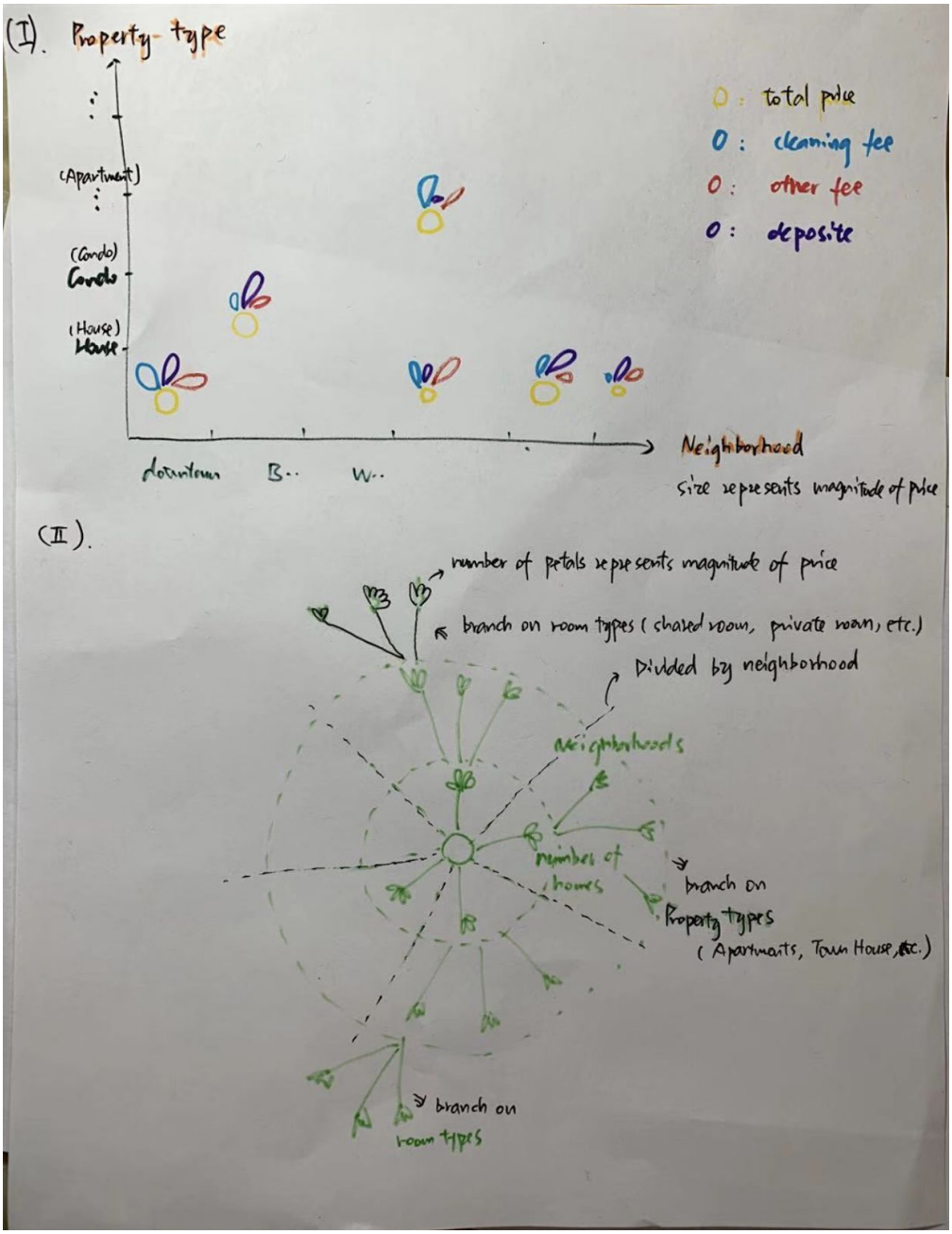
- The second introduces Airbnb’s global footprint. Notice that the logo will be gradually filled up with time of year displayed (still under construction). A corresponding Airbnb’s global

footprint will show up on the map.

- The third and following pages have a two-column layout that allows easier information organization.

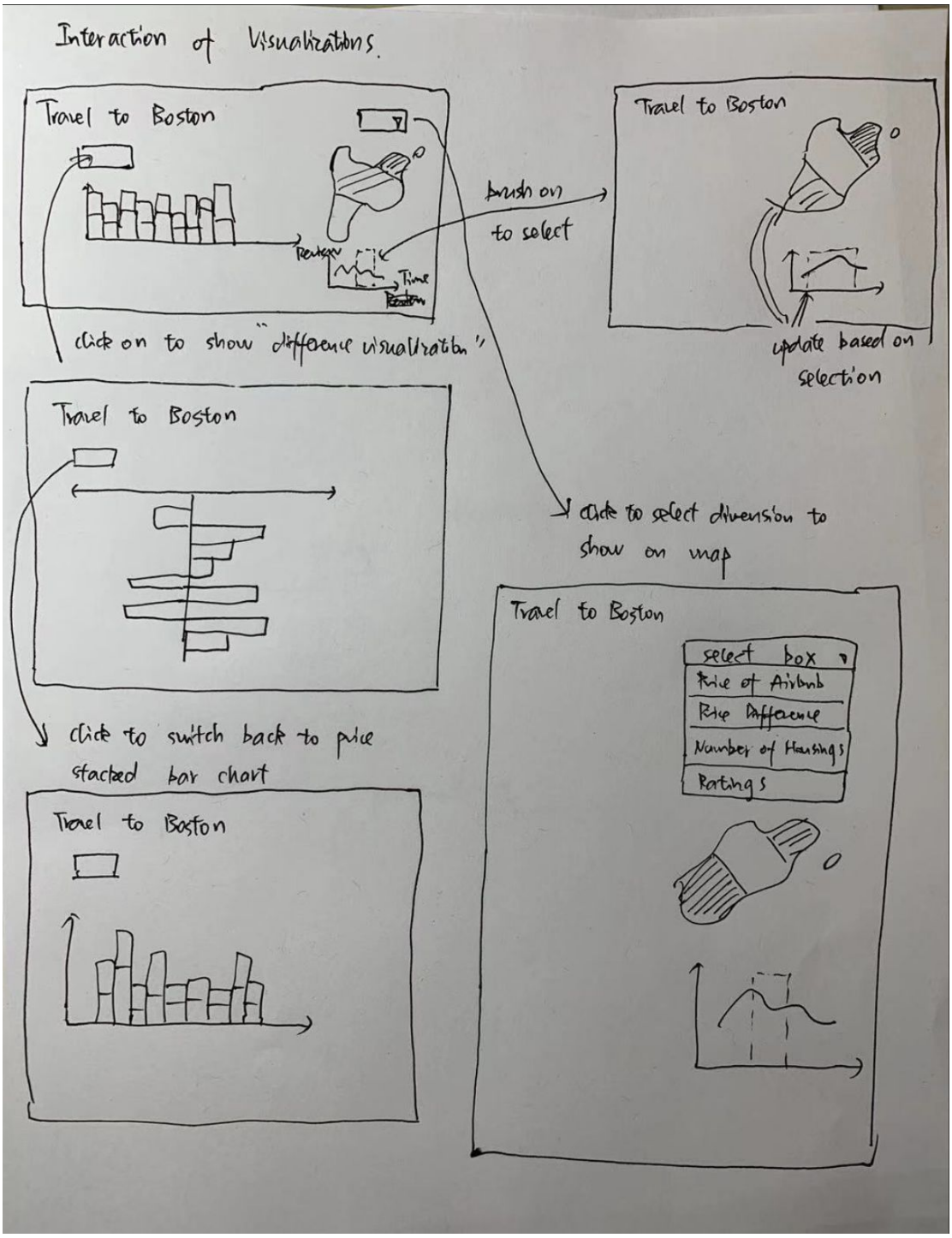


First Sketch of an Innovative View

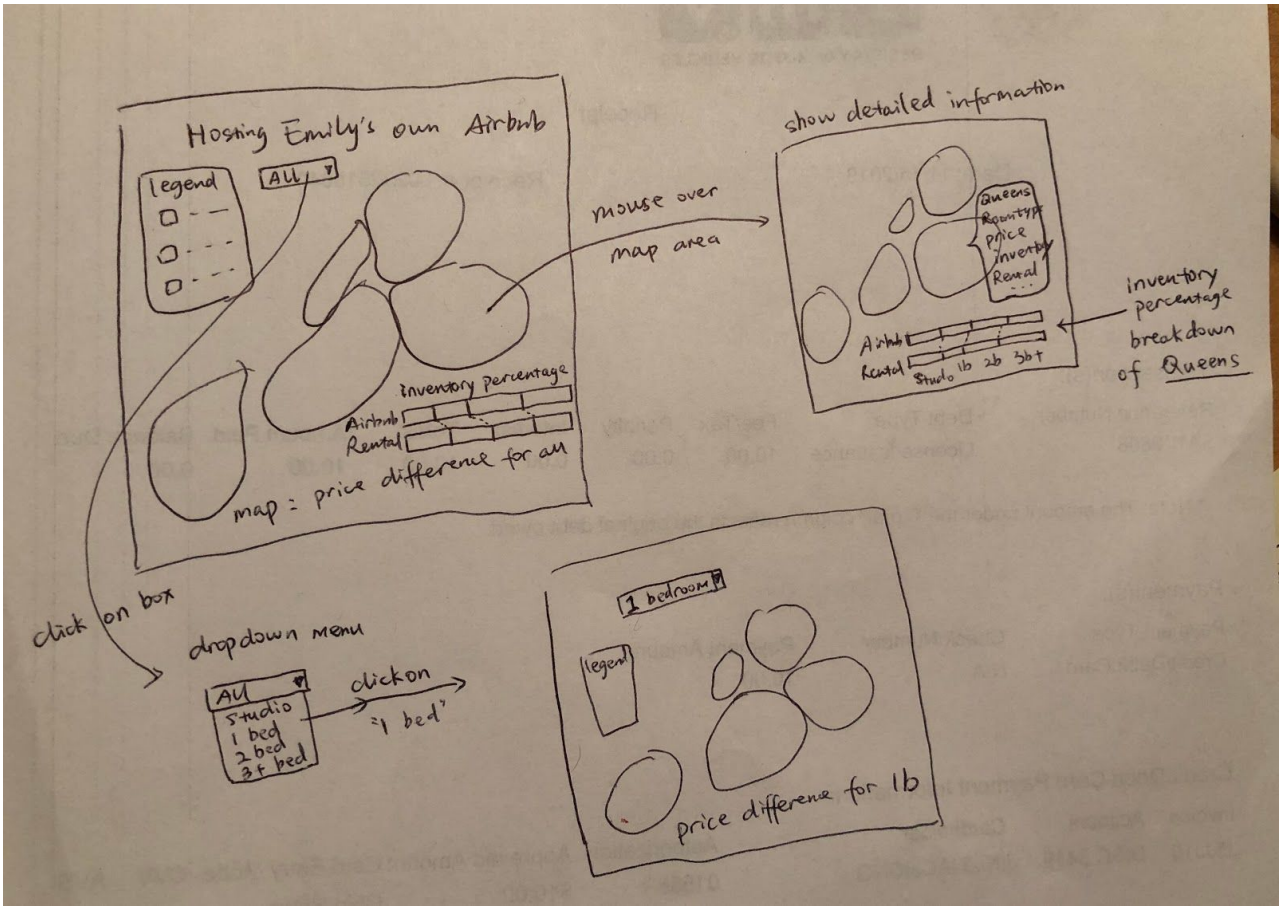


Interactions

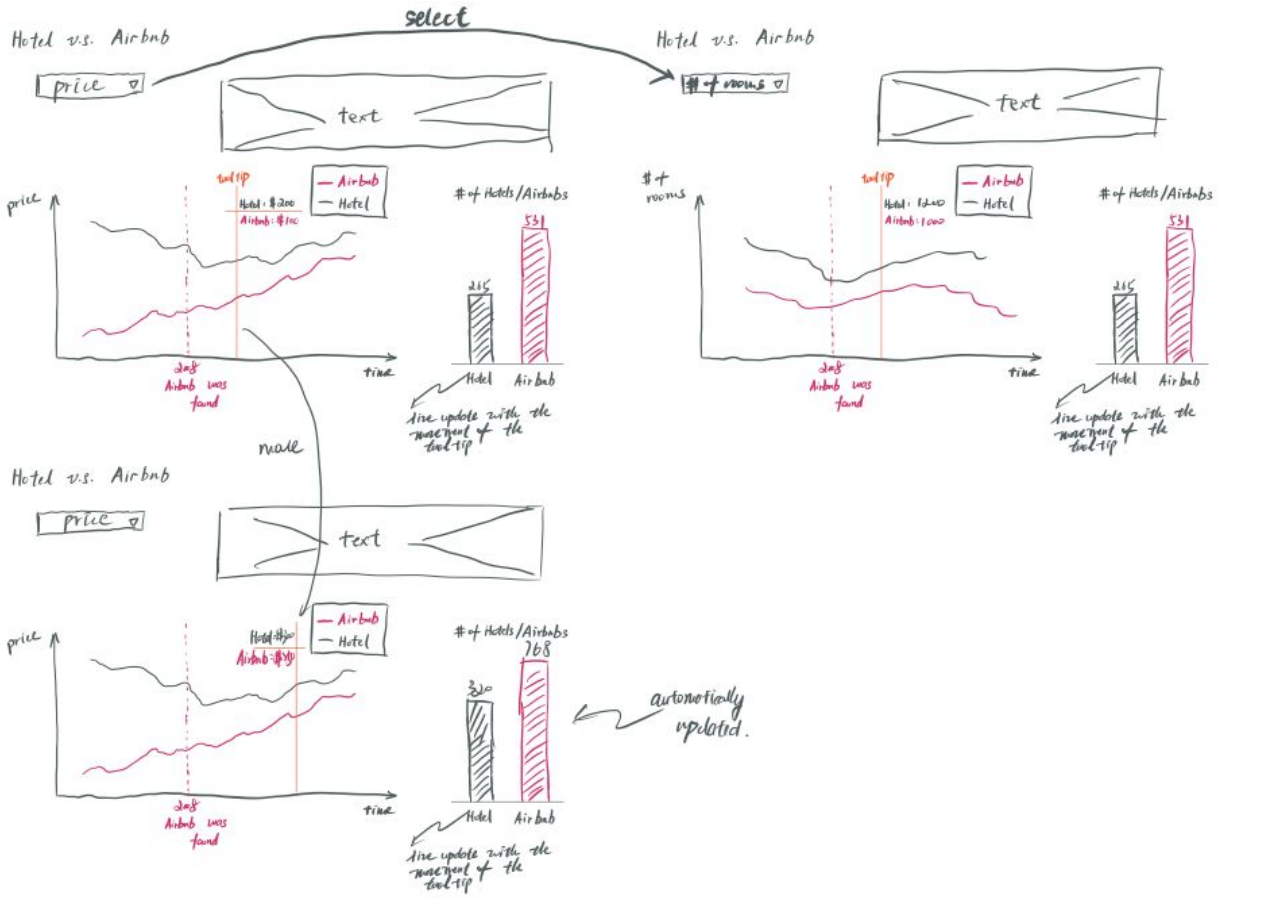
Travel to Boston



Hosting Emily's own Airbnb



Impact to traditional hotel industry



Week 12 (Nov 17 - Nov 23)

To-do: Prototype V2 - Web page Structure, Integration (Due on Nov 24)

Week 13 (Nov 24 - Nov 30)

Test and Evaluation

Week 14 (Dec 1 - Dec 7)

Final Version + Evaluation + Video (Due on Saturday Dec 7)