#### Yongzhi Li, Siyao Zhang, Vanessa Lin, Kairui Liu

#### Project Description

This is a written, non-technical description of your project. Depending on the specifics of your project, you should outline the answers to these (and perhaps other) questions:

* What is the dataset you'll be working with? Please include background on who collected the data, where you accessed it, and any additional information we should know about how this data came to be.

We will be working on the Booking.com dataset. The Booking.com website collected the data, and we will access the datasets through Booking API (https://developers.booking.com/api/commercial/index.html?version=2.3&page\_url=possible-values). Booking API collects and reflects the real-time pricing and availability which is constantly changing. To provide the most accurate data, it is responsible for showing current information which also ensure the content is updated through Booking.com and might be removed in the case of Booking.com no longer having permission to use/collect the information from a property. The price returns the actual price that each property has entered into the system and the price which each booker pay/will pay for their staying at the property. It also contains static data, such as property description, facility information, locations(states, cities), hotel types and other information that changes less in comparison to availability and prices. Booking API does not alter or make changes on those information they collect and make sure that it represents the property and its features. Changing of information, such as prices, text description and property pictures are forbidden as it may misrepresent the property.

* Who is your target audience? Depending on the domain of your data, there may be a variety of audiences interested in using the dataset. You should home in on one of these audiences.

Our target audience would be travelers and partners(websites or property owners) It shows a wide variety of accommodation online and have the dataset prepared for a global audience to accommodate their needs. Travellers who wish to find a place to stay in any city. With everything from hotels to luxury villas, there’s always a place for everyone. For people who are interested in using the data on their website(bloggers, partner websites), they could choose which properties appear on their site, meaning they can offer accommodation worldwide or in a select area. For property owner, Booking API provide them with a deep analysis of their performance, which allows them to optimise and earn as much as possible.

* What does your audience want to learn from your data? Please **list out at least 3 specific questions** that your project will answer for your audience.

“What are the price trends in my neighborhood areas?”

“How does the price varies by month and season for each year and for each city depending on the number of stars for the hotel.”

“What are the most commonly seen customers’ types in my neighborhood areas?”

“What are the property types in my neighborhood and how many is each type in my neighborhood?”

“How many listings are in my neighborhood and where are they?"

“Which area in a city has the most booking orders?”

“How many nights do people usually stay for?”

“Is there a correlation between hotel type and hotel rating?”

#### Technical Description

This section of your proposal is an opportunity to think through the specific analytical steps you'll need to complete throughout the project.

* What will be the format of your final product (Shiny app, HTML page or slideshow compiled with KnitR, etc.)?

Our final product would be an HTML webpage with interactive visualizations.

* How will you be reading in your data (i.e., are you using an API, or is it a static .csv/.json file)?

We are using API to access Booking.com datasets.

* What types of data-wrangling (reshaping, reformatting, etc.) will you need to do to your data?

Reshaping, grouping, summarizing, joining data, structuring, data cleansing

* What (major/new) libraries will be using in this project (no need to list common libraries that are used in many projects such as dplyr)

KnitR, ggplot, httr, jsonlite, tidyr, data.table, magrittr, etc.

* What questions, if any, will you be answering with statistical analysis/machine learning?

According to past data, what might the prices in future years be? Prediction of future prices based on previous prices data.

* What major challenges do you anticipate?

Programming the interface of the interactive visualization.