CS5224: Cloud Computing

Assignment 2

Final Report

CentsTrip:

Beautiful Singapore, Satisfactory Cost

By Group 5

Yu Yingchen (A0159338U)
Ding Wenjia (A0105515R)
Lu Yifan (A0159305E)
Zhu Longwei (A0148601L)
Huang Mengying (A0159188M)

Contents

1. Executive Summary

Over 16 million tourists travelled to Singapore during 2016, which are three times than the local population. Singapore is one of the hottest tourist city in Asia due to the pleasant city view and the open culture. Tourism contributed 4% of the whole annual income of Singapore in 2016, which is around 23 billion Singapore dollars. Internet product has a wonderful opportunity to get into such a big industry.

In fact, Singapore is the most expensive city according to the survey by EIU, 2016. Travel cost to most cities in Southeast Asia is just half or even less. This is a barrier for those people who travelled to Southeast Asia with limited budget before and wished to visit Singapore similarly. We studied popular trip advising websites and apps, discovering that they treat everyone indistinctly and people with limited budget can hardly find useful traveling information.

In our product, we target at those tourists who always want to enjoy the highest service quality with minimal cost and those who have specific budget for travel. The website we want to build contains two major parts: the cost efficiency analysis and recommendation system for users. Cost efficiency is to measure whether an airbnb or a tourist attraction has a lower price with same or even higher quality. In the first part, we plan to calculate the cost efficiency with our own algorithm on the Bluemix SaaS platform to evaluate accommodations and attractions. The data collected are from Singapore government website and other sources like Airbnb, which is the most famous company providing b&b service worldwide. In our own survey, there is no such websites or travel agents in Singapore dealing business with Airbnb. This means that we can provide a competitive price in accommodation for individual visitors. Then we'll let users choose their preferences on the website so as to recommend suitable plans for them. Finally, recommendation results will be represented back by a simple geometric graph, which is very user-friendly and brief.

2. Business Case Identification

2.1 Motivation

The general services those travel-related websites in Singapore provide are introducing attractions, booking hotels and selling tickets. With economic growing rapidly in Asia, more and more young people begin to travel abroad. Singapore, as a tourist city, has faced the visitor reduction problem since 2014.

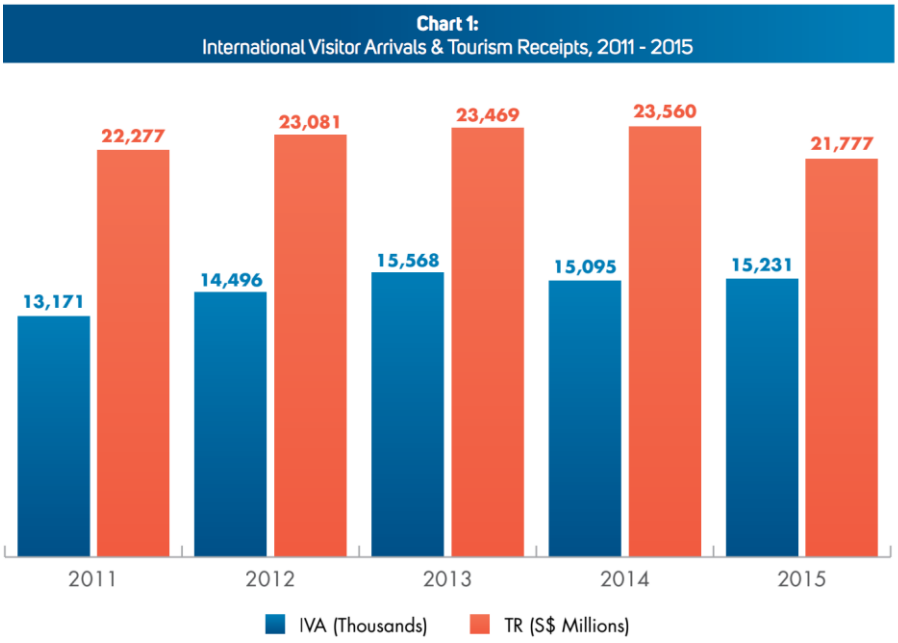


Figure 1: International Visitor Arrivals & Tourism Receipts in Singapore

2.2 Target users

As we mentioned previously, aimed users have limited budget in their trip. In our assumption, young couples and teenagers travel solely are potential consumers of our product. They prefer traveling individually but have clearly budget. These people are keen on searching online to find lower travel cost information while assure quality at the same time. For most travel advised sites, the great quality is based on the payment amount. However, our product will provide recommendations according to our distribution channels and algorithms which lower the cost and assure the quality.

3. Business Model

"CentsTrip" is a travel recommendation application for tourist who plans to travel to Singapore. It aims to provide the recommended travel plan with minimal expenditure for users based on their preference and budget.

3.1 Services to be developed

Two web-based SaaS services have been designed in this project:

a) **Cost Efficiency Analyzer**

It is important to recommend accommodations and attractions in good quality with rational price for users. The first service focuses on analyzing the cost efficiency of accommodations. The cost efficiency analysis of accommodations takes account of four factors, which are respectively rating, price, distance between accommodations and attractions, and distance between accommodations to transportation junctions like MRT stations. The basic idea of

this SaaS is to extract necessary data with each dimension from our database which is stored in Bluemix SaaS, Bluemix. Afterwards, it uses R script to fit those data into a weight regression model to calculate the cost efficiency of accommodations. More details will be covered in the following implementation part.

b) **Recommendation System**

The second service is a recommendation system which allows tourist to input their budget, travel duration and attraction preferences. Firstly, it will fetch tourist's input and pass them to its back-end. Afterwards, the recommendation results analyzed by back-end will be exhibited on the website. To increase the choice freedom of tourists, multiple accommodation and attraction recommendations and their information will be provided for tourist to explore and select. Meanwhile, the information of nearby food centers will also be offered to tourist. Lastly, a summary page will be displayed on the web page. In this page, it summarizes tourist's final choices about accommodation and attraction and provide associated information such as location and price.

3.2 Services Comparison

The existing traveling recommendation service providers focus more on integrating and exhibiting information, such as discounts and other tourists' review. In this case, they do not offer customized recommendation for target users. Unlike the existing ones, CentsTrip is a unique application that provides users the customized travel plan. The customized travel plan is generated by our recommendation system based on budget, traveling duration and preferences of each tourist, it consists of recommended accommodation and attractions. The detailed comparison between CentsTrip and well-known traveling service providers are tabulated in Table1.

	Accommodation Search	Attractions Information	Flight Search	Food Information	Customized Travel Plan
TripAdvisor	✓	✓	✓	✓	✗
Booking	✓	✓	✗	✗	✗
Agoda	✓	✗	✓	✗	✗
Ctrip	✓	✓	✓	✗	✗
Airbnb	✓	✗	✗	✗	✗
CentsTrip	✓	✓	✗	✓	✓

Table 1: Services Comparison

3.3 Revenue Model

As a web-based service, CentsTrip will be totally free for users. It attracts potential users by providing a good quality recommendation service. Subsequently, with the website traffic generated by users, customized advertisement such as accommodation discounts or ticket sales of attractions could be exhibited on CentsTrip. Generally, the revenue model of CentsTrip is to attract users and charge the advertisement fee.

3.4 Cost Analysis

As a SaaS service, CentsTrip has much less costs compared with other on-premise services. Firstly, the service is implemented on cloud platform such as Amazon Web Service or IBM Bluemix, which means up-front cost of setting up the services can be eliminated or significantly reduced. Moreover, implementing service on cloud platform leads to measured operational expenditures instead of anticipated capital expenditures. Secondly, due to the elasticity of cloud computing, our SaaS service can rapidly respond to the business cycle, which can help to reduce cost without limiting the capacity. However, those on-premise services do not possess this elasticity, it will take much more costs and time for them to scale up their resources to meet the demand. Lastly, in this competitive market, it is crucial to develop and launch our application fast enough to take the first mover advantage. Using cloud-based platform offers the chance to greatly shorten the development cycle, because its virtualization characteristic allows our SaaS service to be developed without considering setting up hardware, infrastructure and etc.

Above all, the business model of CentsTrip aims to use the minimal cost to provide users with recommendation service and benefit from website traffic and advertisements.

4. SaaS Architecture & Implementation

4.1 Architecture Overview

Figure 2 below is the overall architecture design of our CentsTrip program. The application is a web-based application hosted on the IBM Bluemix cloud platform. Users can access the website through cents-trip.mybluemix.net.

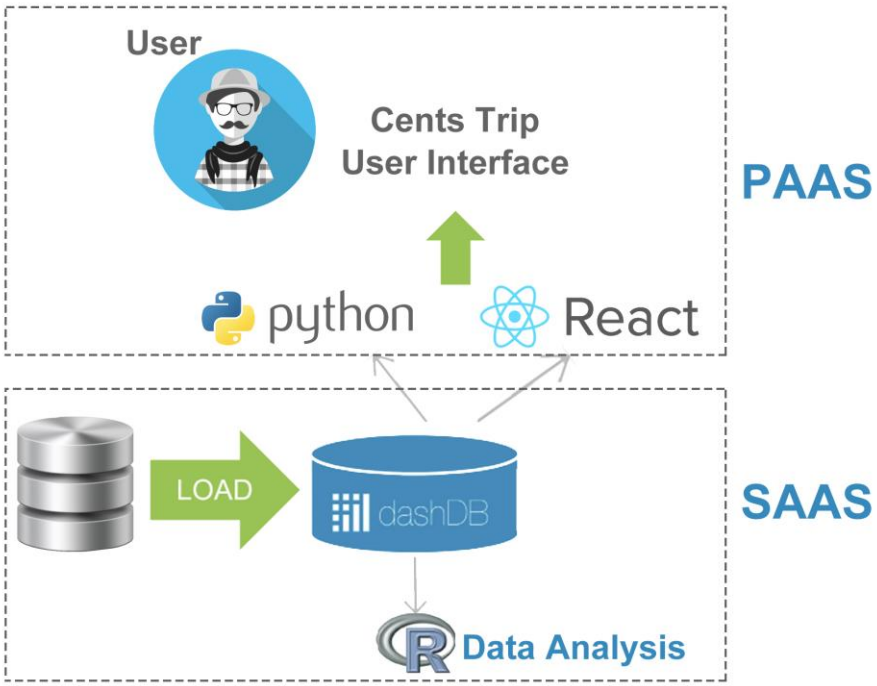


Figure 2: High Level Architecture

In the first application, we retrieve five sourced data (i.e., hawker center, food court, tourist attractions, Airbnb and MRT data). We preprocess and load them into the SaaS of IBM Bluemix - DashDB, and uses R script for analyzing. In the second part, Python and React library are adopted to design and implement the front and back end of the website.

4.2 Data Acquisition and Preprocessing

As stated above, two main datasets of CentsTrip are Airbnb and Tourist Attractions, we also collect other dataset from https://data.gov.sg/ to enhance the value of our application.

(1) We use the unofficial APIs for Airbnb (http://airbnbapi.org/) to crawl the items that we need, including the roomid, name, location, etc.
(2) We downloaded the tourist attractions data in KML format which is converted into three attributes in Google Fusion table. Namely, name, longitude, latitude.

The following points are the additional fields in the Tourist Attractions table:

- Category: represent the type of attractions, including architecture, heritage, arts, museums, memorials, places-to-see, nature-wildlife and recreation-leisure.
- Ticket-price: the price for an adult in each place.
- Popular Attractions as in Lonely Planet: there are only two values, 1 corresponds to the popular, otherwise, the value is 0.
- Rating: Tourists' review on Google Map.
- Opening_hours: an important attribute for the next recommendation system.
- Description: abstract introduction for attractions.

(3) Python script is used to crawl the name, longitude, latitude of MRT, food court, hawker center respectively.

4.3 Data Analysis and Transformation

4.3.1 Cost Efficiency of Airbnb:

Considering that tourists might concern the price and location when they choose a airbnb, we measure the cost efficiency which is to represent a airbnb has a lower price with the same or even higher quality by primarily taking these two attributes into account. In order to find the important factors among multiple attributes, we use **weighted least square estimation regression model** to select the influential parameters, which statistically use 1/|residual|^2 as the weight for each observation.

To obtain the predicted price of airbnb, price of airbnb is treated as a response and four attributes are fitted in our regression model, including the distance between airbnb and several famous attracted places, the distance with the nearest MRT, the rating of airbnb itself, the popularity of airbnb which is stated in Section 4.2. After obtaining the predicted price as shown on Figure 3, we calculate its difference with the real price of airbnb, and this result is corresponding to the cost efficiency of the airbnb, which means the higher absolute value stands for a good cost efficiency, and vice versa.

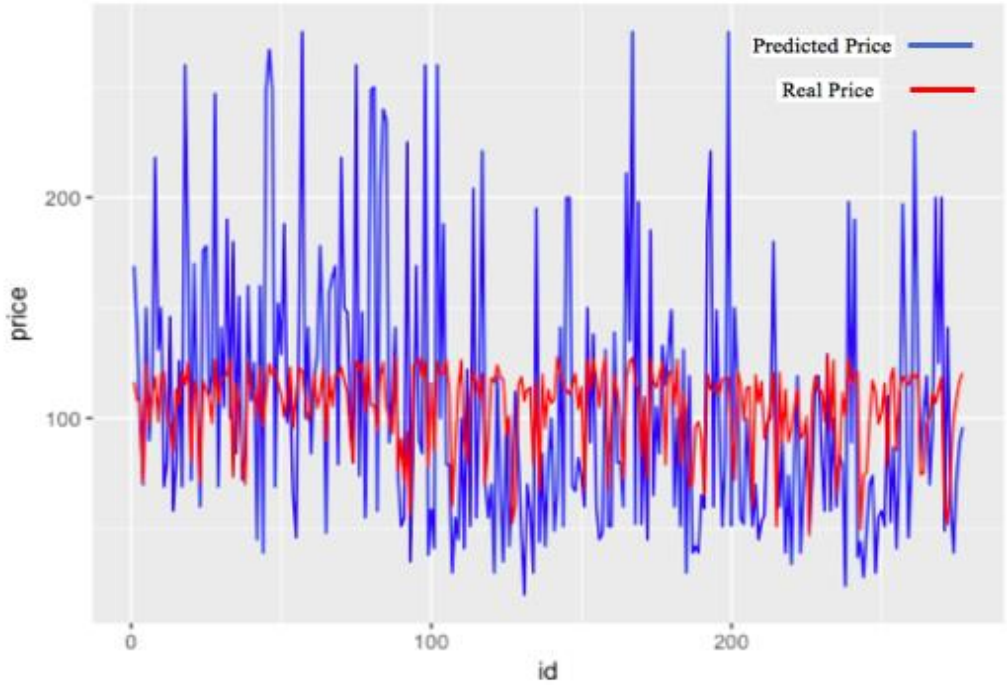


Figure 3: Price vs Predicted Price of Airbnb Accommodations

4.3.2 Similarities among Tourist Attractions:

In our recommendation system, users can select their preferred tourist attractions, then, the program will recommend the similar places based on their input. **TF-IDF (**Term Frequency–Inverse Document Frequency) is a popular and effective algorithm in text mining field to

calculate the text similarities. It is used for the "description" attribute in tourist attraction dataset. We consider the output as a new factor, and combine it with the price, popularity and rating attributes to measure the similarity between tourist attractions by calculating **cosine similarity**, whose results land between 0 and 1. The flow of this algorithm is presented on Figure 4. Based on the preferred tourist attractions of tourist, our system would recommend other tourist attractions that have the higher similarity to the tourist.

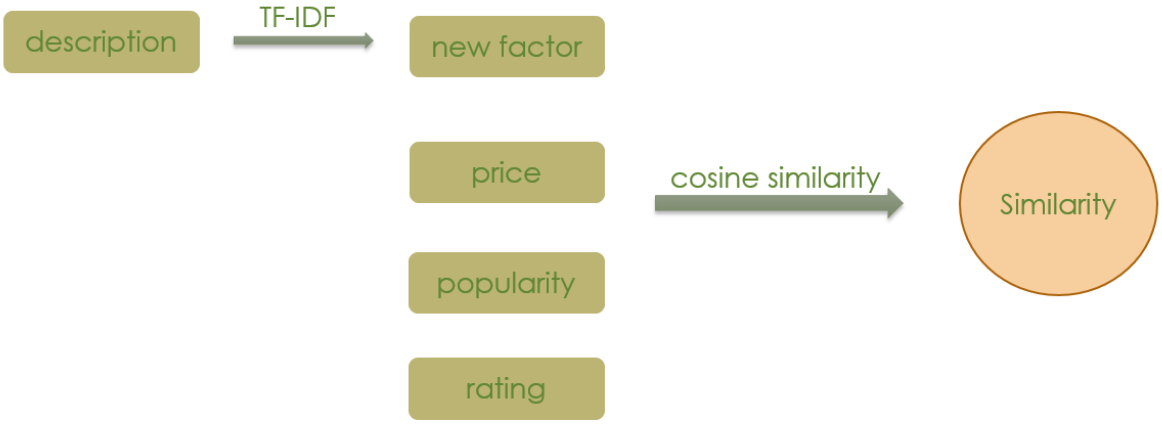


Figure 4: Flowchart of Calculating Similarity

4.4 Recommendation System

4.4.1 Front-end

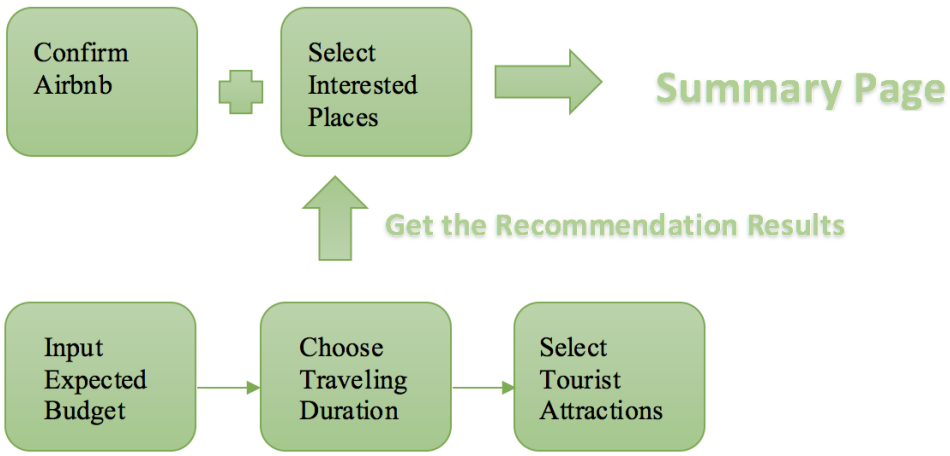


Figure5: Front-end Flow Diagram

Before the recommendation process, the web application retrieves all relevant information in each dataset by sending a GET request to the back end. Python script will proceed to acquire these data from DashDB and return it in JSON (JavaScript Object Notation) format.

Figure 5 shows the flow of the front-end of CentsTrip application. After getting user's input,

our web application will output and display the recommendation result to users. Based on it, users can confirm their desired Airbnb and other possible interested tourist attractions. Finally, the summary page will show all selected and recommended results and their corresponding location and price information to users. Figure 6 demonstrate the GUI (Graphical user interface) of the implemented web application when user's receiving recommendations and interact with them, and GUI of summary page is shown on Figure 7.

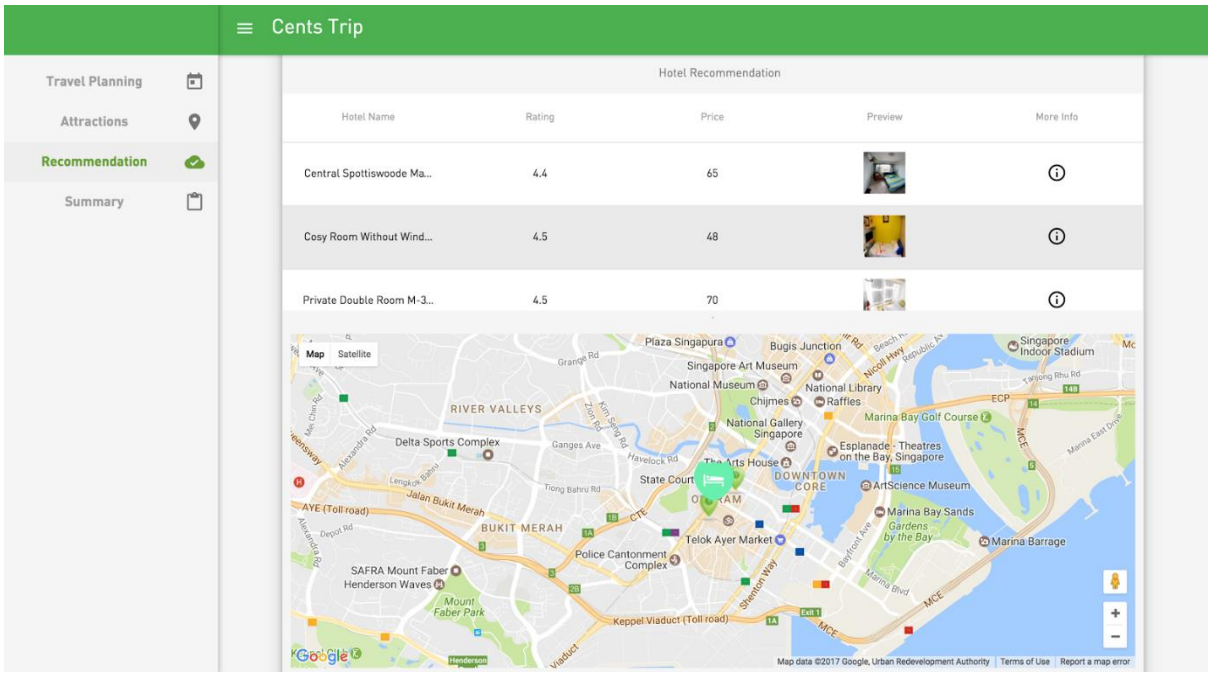


Figure 6. GUI of Airbnbs Recommendations

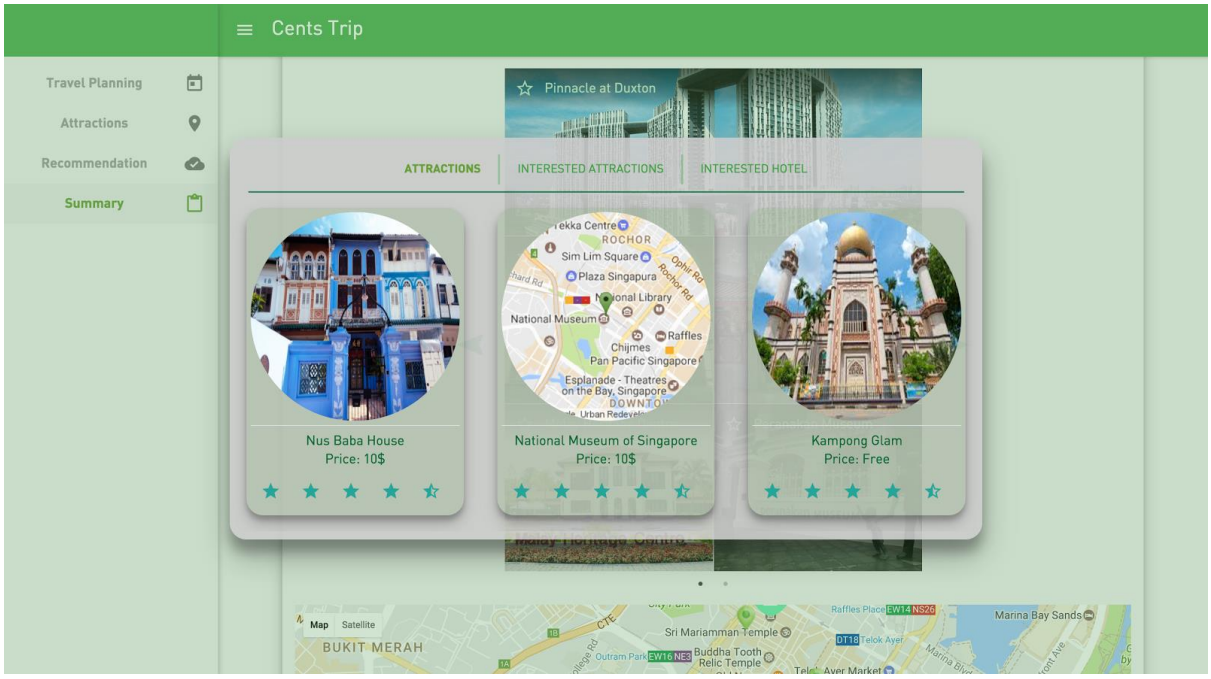


Figure 7. GUI of Summary Page

4.4.2 Back-end

The back-end service consists of the two main portions: tourist attraction along with the nearest several hawker centers recommendation, and the recommendation for Airbnb and hawker centers.

- Find the nearest 20 airbnbs from the selected tourist attractions, filter the airbnbs when its price satisfies the condition $airbnb\_price \times Number\ of\ days + ticket\_price) \times 120\% \leq budget$, number of days are the duration of travelling, ticket_price is the whole ticket prices of selected places. Recommend the 5 airbnbs with the highest cost efficiency value analyzed in the analysis part, along with the nearest 3 hawker centers or food courts from each airbnb.
- For the chosen tourist attractions, we recommend the nearest 3 hawker centers as above. Then we recommend similar places based on the calculated similarities value, and find the corresponding several hawker centers again.

POST function is the bridge to connect the front and back end, which receives the input contents from the front end, invoke the recommendation method, get the recommended result and return it for displaying on the website.

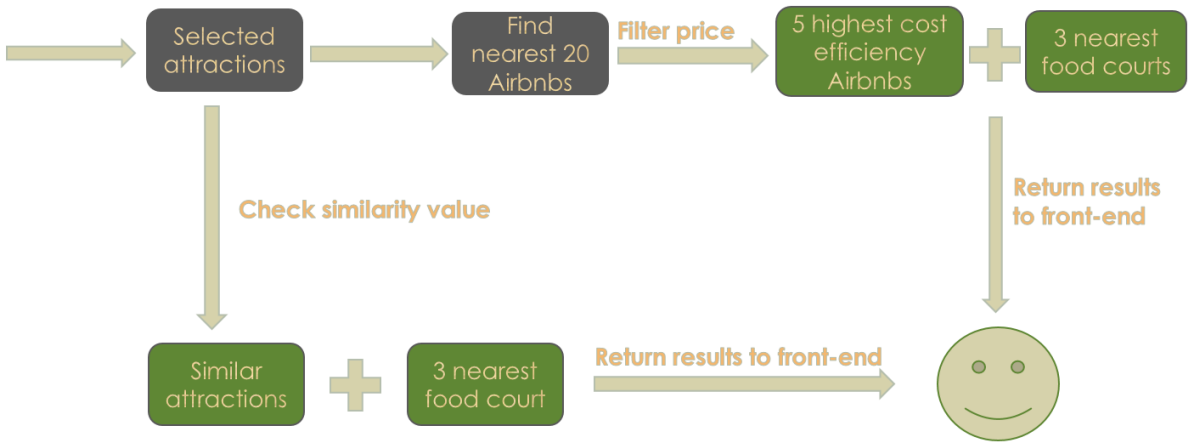


Figure 8. Back-end Flow Diagram

4.5 Limitation and Further Improvement

Being limited by the unofficial Airbnb API, we crawl the required data and store it into the DashDB but we cannot update it automatically. So we cannot get the real-time price during the recommendation phase. If we can commercialize our business model CentsTrip, we can collaborate with Airbnb to provide a more friendly and convenient service to users.

5. Economic Factors

5.1 Economic Benefits and Key Consideration

There are many economic benefits carried out by our application. Since the application supports users to search for accommodations and tourist attractions as the travelling plan, it contributes to the tourism activity in Singapore associated to economy and improves the flow of spending in the region. We analyze the related data from government statistics, and model the plan according to the users' preferences and requirements. In this measurement process, the population size, the distance among destinations, the budget quantity, and the duration of time are the key factors that change our estimations and plan. When the users access our website, there will be profit generating from the website due to the website traffic. The placement of advertisements on the website could also gain economic benefits, such as discount information and attractions service, so that we can generate income and revenue from them.

Depending on the different seasons in a year, the number of users' transactions may turn out to be different, for example, summer holiday is a peak season for people to travel and relax. Therefore, there will be a request fluctuation for the website, and we need to deal with the increase of demand. Accordingly, the price of different facilities may also fluctuate, the more people coming in, the more expensive the price may be. The availability of the travel agencies is also a consideration factor of our system. Furthermore, we have to analyze the similarities among different attractions and nearby food courts in order to promote a rational travel plan for better efficiency.

5.2 Pricing Models

We adopt dashDB entry plan in our project which cost zero dollar per month since our database contains less than 1GB data. As the amount of data increases, the cost may grow to $52.50 per month. The variable cost like license fee using IBM Bluemix platform is considered. In terms of fix cost, it mainly comes from the development cost. Compared to the on-premise services, our service reduces the upfront cost and does not spend any money on facilities, maintenance and administration. Considering the time and number of people put in, the revenue generated from our service will be used to support the cost with Bluemix. We aim to adopt flexible pricing model plans to satisfy the demand changes and workload fluctuation. Since the demand of service at this initial stage is small and uncertain, we would like to apply on-demand pricing model for it. However, subsequently, as the service is deployed successfully and the service demand increases, we may consider reserved pricing model as an alternative for reserving instances for the users.

5.3 Tradeoff between Cost and SLA

At current stage, we provide our service for free. Different from other on-premise services,

we reduce the cost to buy the IT resources for building the website and run the application. We guarantee our service quality using IBM Bluemix as the platform to develop, which saves time and money spending on preliminary stage setup and machine supply. We use dashDB Entry plans which cost zero dollar to fit the user's needs. The availability of our service can reach 99.95% on the basis of IBM Bluemix platform. In terms of response time, we adopt advanced searching algorithm to analyze the function and retrieve data from database, so that the website can display the results in a fast and robust way. Furthermore, since we already obtain the preference of accommodations according to their cost efficiency ranking in our initial service, we are able to quickly achieve the accommodation information with daily plans for users in the recommendation process. Additionally, we collect data from reliable sources, mostly from government's statistics, which guarantees the reliability of our services and results. Nevertheless, in the latter stages of development, we may want to charge a considerably amount of fee from users for the cost of updating the data storage, and providing multiple deployment of the application in order to increase the availability and reliability based on SLA.

6. Contribution

In this project, everyone was involved in acquiring and preprocessing data. Thanks to Longwei and Wenjia, the front-end framework and interface was designed and implemented successfully. Back-end recommendation logic and algorithms are explored and implemented by Yingchen, Yifan and Mengying. After the implementation of our web services, each teammate contributed on compiling the report of this project.

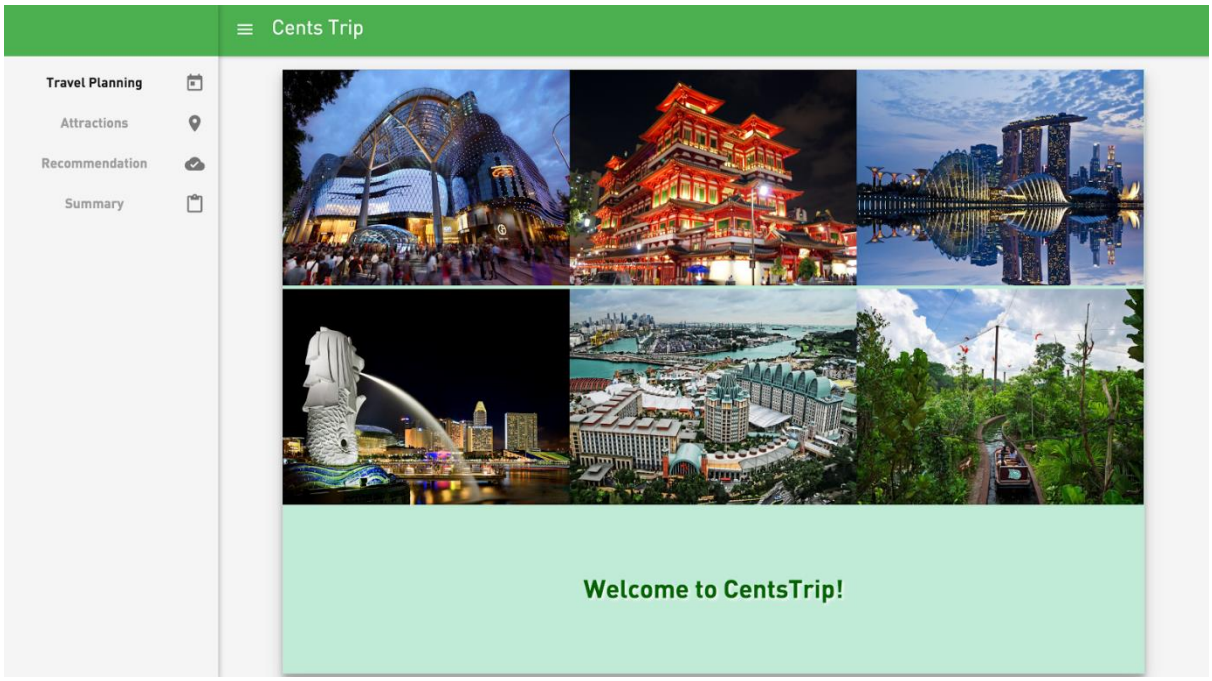
7. Conclusion

CentsTrip is an intelligent assistant which aims to recommend the most suitable accommodations and tourist attractions for users within their travel budget and based on their own travel preferences. The whole web application is implemented on IBM Bluemix Cloud, providing mainly two SaaS services, cost efficiency analyzer and recommendation system and showing results on geometric graph. The services take user's preference into consideration and yield to the result based on statistical analysis and similarity analysis. The application takes advantages of the elasticity of cloud platform, uses the minimal cost to provide recommendation service and benefits from website traffic and advertisements. In our future work, the application will focus on considering the visit duration of tourist attractions and providing personalized travel routines for each user. Besides, different pricing model may be considered to provide better recommendations. In general, CentsTrip focuses on providing better recommendations of accommodations and attractions according to user's specific need with the aid of cloud platform.

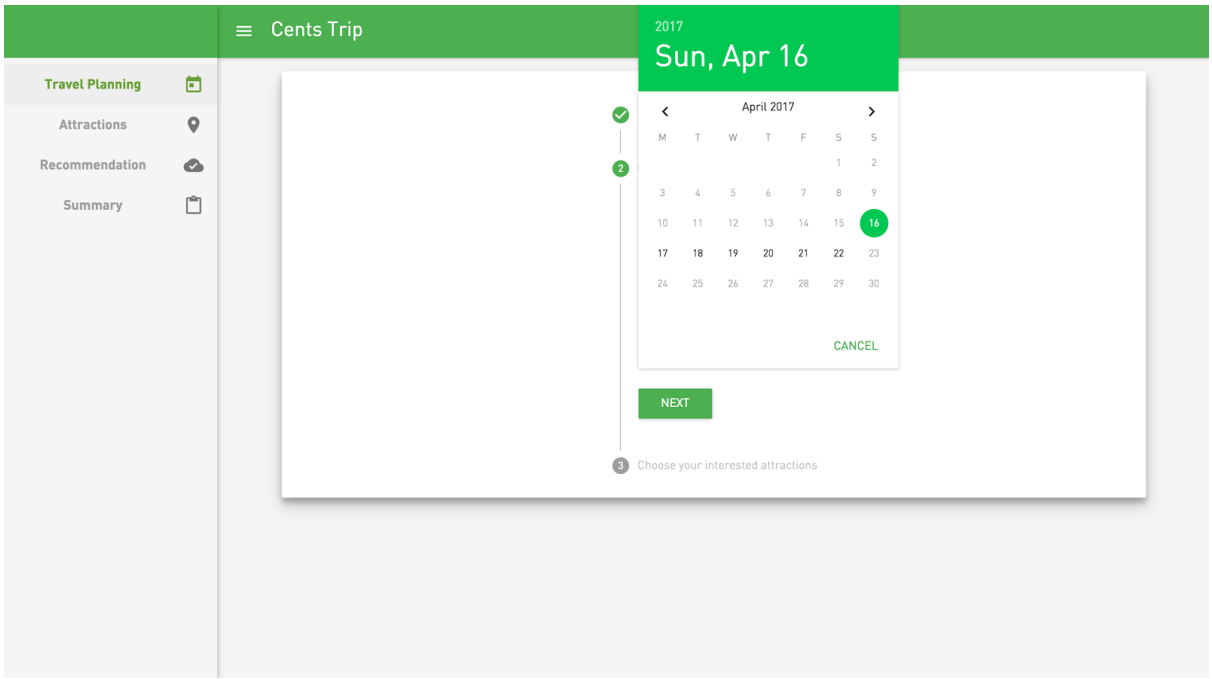
Appendix

CentsTrip GUI Screenshots:

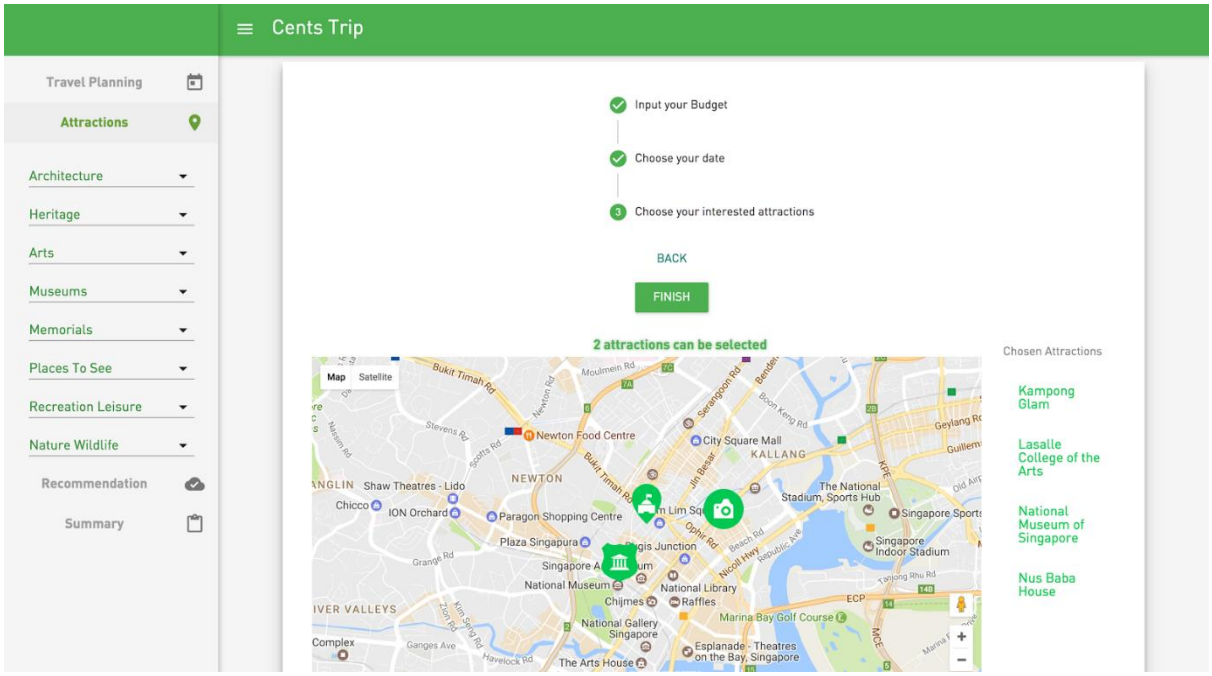
1. Welcome Page



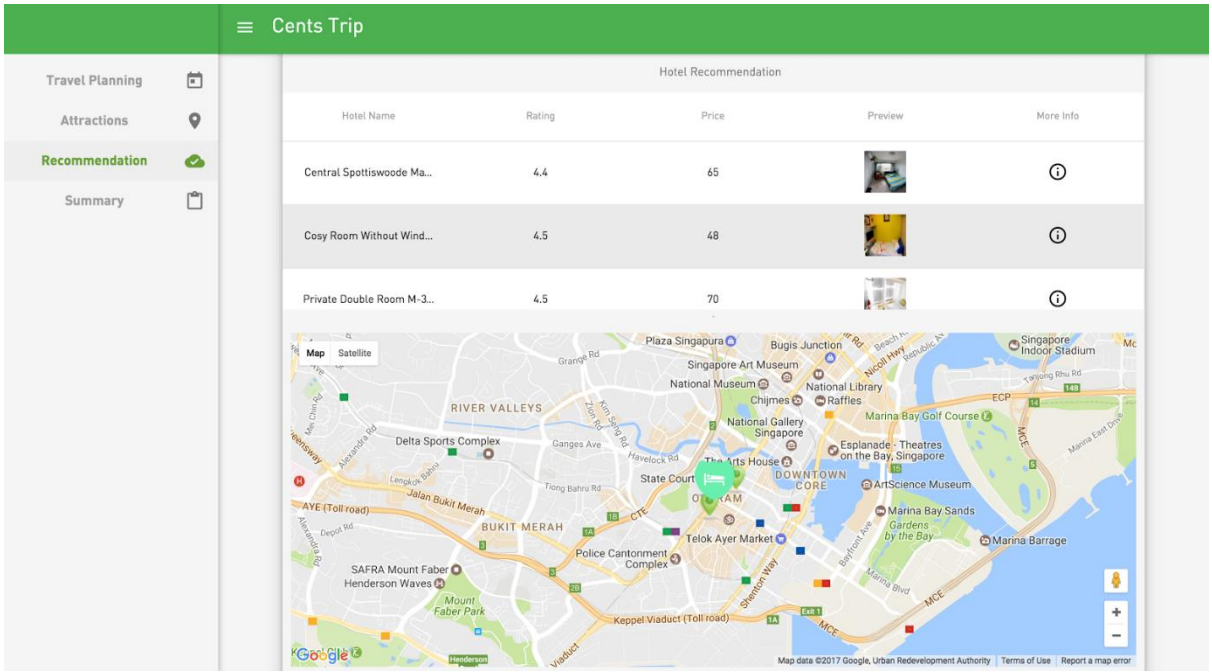
2. User input their duration.



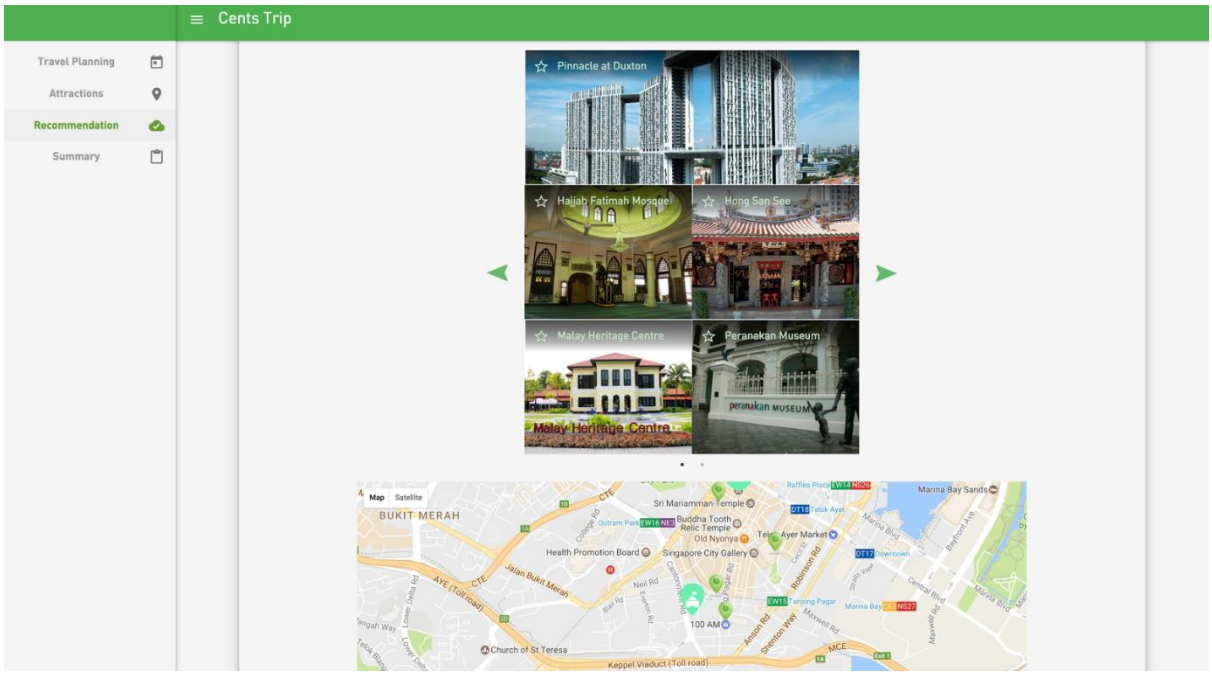
3. User input their preferred attractions.



4. Airbnb recommendations display and interact.



5. Tourist attractions recommendations display and interact.



6. Summary Page