Alan Nguyen

Kartheek Sunkara

Luigi Penaloza

IST 652

Prof. Ben Nichols

**Project Proposal: Expedia Service Recommendation**

**Topic of Investigation:**

For our choice of topic, we would like to explore the data from Expedia to determine the best travel and the best hotels. We will then pipeline the data to recommend users through email and/or text messages. For our final project we are going to be looking into obscure data from a JSON web scraping and csv data from the API and Expedia. This data is representative of customer behavior on the site and could be considered log data to show information such as sessions, hotel data, and flight destination data. Information included will have hotel and flight data and metrics. We will also be determining the criteria and conditions that make a good hotel, flight, and combo deal. At the end of analysis of data, we will be attempting to imitate a subscriber news feed based email with this data.

**Dataset:**

The dataset will be a combination between both data from Kaggle and the Expedia API. Expedia’s data will contain more unstructured data where Kaggle will have the data cleansed and ready to work with but the information is based more around user activity as opposed to data provided directly on the site. On top of this, the Expedia API has updated and realtime data based on their site, so we will be using that to make a “deal” prediction model or the best value oriented purchases.

Sample of unstructured API Data:

"ratings": [

{

"score": 4,

"maxScore": 5,

"source": "AAA",

"description": "STARS"

}

],

"contacts": {

"Property": {

"phoneNumbers": [

{

"phoneNumberType": "Phone",

"countryAccessCode": "1",

"areaCode": "123",

"number": "1234567"

}

]

},

Sample of structured Kaggle Data:

|  |  |
| --- | --- |
| **Column** | **Metadata** |
| **hotel\_continent** | Designated continent number for hotel |
| **hotel\_country** | Designated country number for hotel |
| **month** | Month of the year in number |
| **srch\_destination** | Code for internet search destination |
| **user\_location** | Number code for user location |

**Acquisition Methodology:**

We will be acquiring part of the data from the Expedia API via simple web scraping and conducting data acquisition to automate recommendations for Expedia customers and potential service users. The other datasets will be acquired via data repositories such as Kaggle and data world.

**Big Things:**

The purpose of the investigation is to gather Expedia hotel and flight data to provide an analysis of the data, inform and even recommend to users the best deals based on analysis and potentially machine learning models.

**Small Things:**

From the data and analysis we’ll write a program that helps disseminate deals and other information via email or text message.

**References:** [https://expediaconnectivity.com/apis/product-management/property-api-contact-expedia](https://expediaconnectivity.com/apis/product-management/property-api-contact-expedia-group-before-adopting-/api-definition.html)

[-group-before-adopting-/api-definition.html](https://expediaconnectivity.com/apis/product-management/property-api-contact-expedia-group-before-adopting-/api-definition.html)

<https://www.kaggle.com/c/expedia-hotel-recommendations/data>