Trip Analyzer The Unbiased & Interactive View of Tourism Data to Tourists for Making Informed Decisions

Xiaoqi Zheng Si-Onn Kwok Brian Ng Youjung Kim Yiufung Leung Ankur Agarwal

Introduction

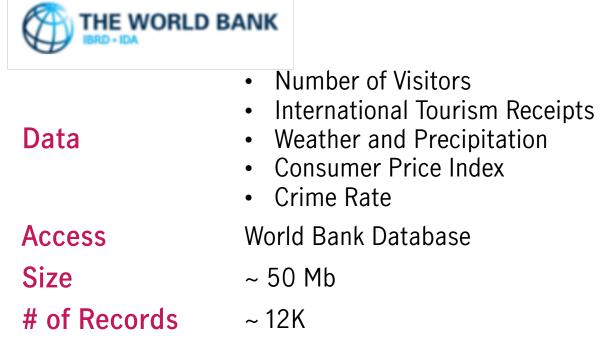
- There are two major challenges that tourists currently encounter while planning their next vacation.
- #1 Information on different travel destinations is distributed all over the internet with no single unified view.
- #2 Although tourist websites provide an overview of the travel destinations, the information is modelled to favour the underlying commercial interests.



Motivation

- Every tourist is different and different factors such as travel/vacation budget, tourist density, crime rate, safety, pollution levels, number of attractions, social sentiments etc. weigh differently for each.
- Our solution does not give travel recommendations but provides an unbiased interactive visual snapshot of travel related information allowing the tourists to shortlist travel destinations based on their personal preferences.

Trip Analyzer - Design & Data Source



Developers

Point of Interest & Categorization Data Sygic API Access Size Realtime Data Access ~ 200 cities (for project)

~ 20 POI x 10 categories / city

~ Images for each POI aqicn.org

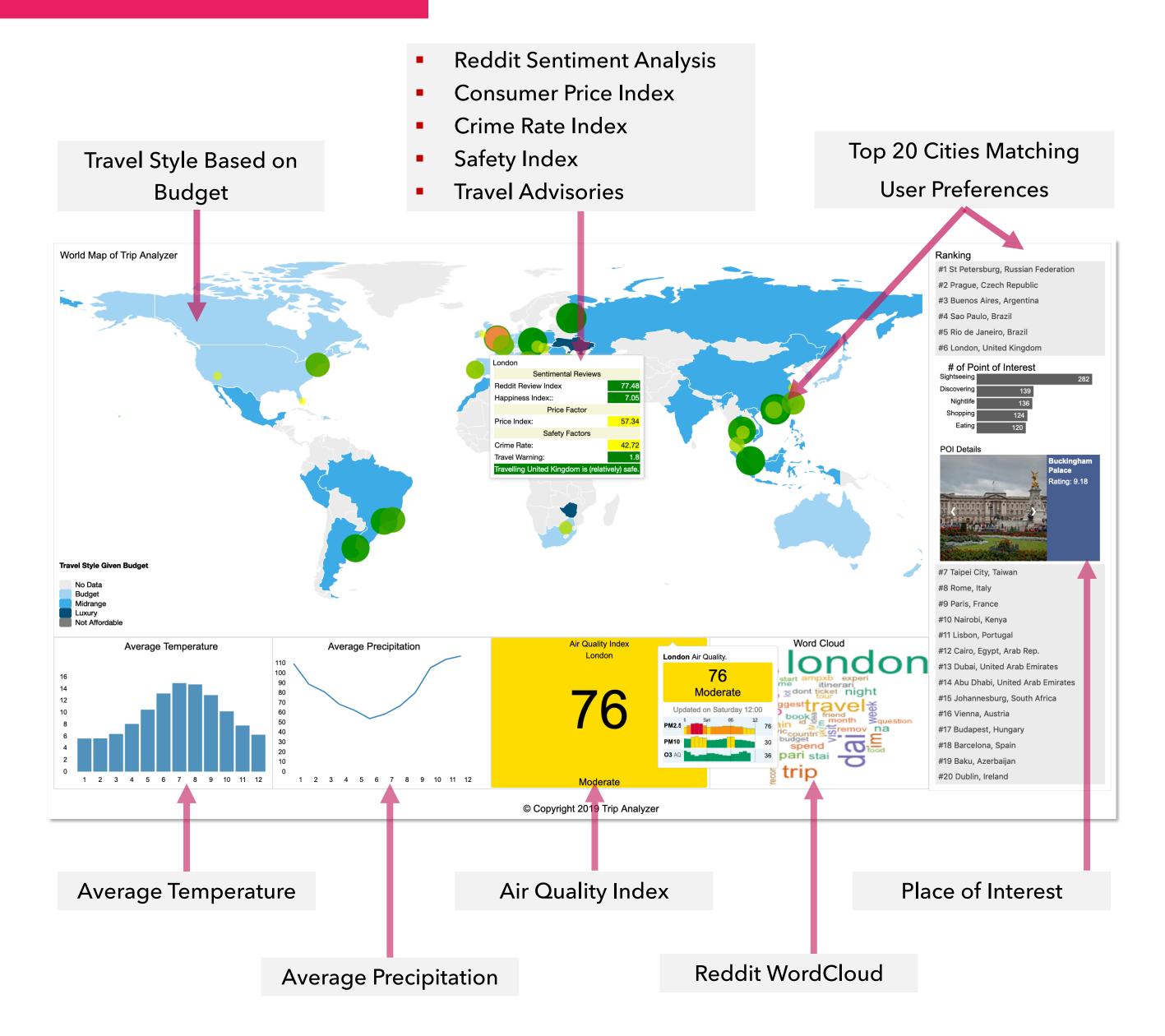
of Records

Data Realtime Air Quality Index Access API - AQICN.ORG Size Realtime Data Access # of Records ~ 200 cities (for project) Variables AQI, PM2.5, PM1.0, O3



Data Happiness Index Data Scrapping (UN SDSN) Access Size ~ 80 Mb

Index computed for countries since # of Records 26 years



Google BigQuery

contraction reddit SubReddit 'Travel' User Comments Google BigQuery Reddit Dataset

Access Size 250 Gb # of Records ~ 1.5 Million

Variables 25 (Comment, Heading etc) **Analysis** Sentiment Analysis & Word Cloud

Travel-Advisory, in

Data

Data **World Travel Alerts** API calls Access

Size Realtime Data Access

Combine travel alerts issued for over # of Records 200 countries

NUMBEO

Data Safety Index Access Dara scrapping

Size ~ 21 Mb

Safety Index scrapped for 200 # of Records

countries

The Budget Your Trip API

Data Trip Data API calls Access

Size Realtime Data Access

of Records

Variables Trip Data, Daily/Total Expense etc **Analysis** World Map showing affordability

Approach

Intuition

Our solution relies on simple premises that:

- Every tourist is different
- presented with relevant information in a simple graphical interface, without any commercial bias, tourists can intuitively decide on a travel destination that is closest to their personal preferences.

Innovation

- No other tourism related solution provides such unified view of information distributed across several public/government databases
- Unlike Wiki (Voyage), our solution provides data in an interactive graphical interface
- Multiple indexes are scaled and baselined for easier user appreciation
- Sentiment analysis of user comments on reddit is presented as positive/negative percentages
- WordCloud gives user a quick snapshot of high frequency keywords

Algorithms

- Multi source data access and processing
- Multi source data association
- Deterministic Record Linkage
- Sentiment Analysis on over 1.5 Million SubReddit 'Travel" user comments
- WordCloud on High Frequency keywords appearing in user comments

Experiments and Results

