“I Love It When You Call Me Big Data”

Anton Khoutiev, David Canji, Yasha Farhan, Jeson Ugale

***Overview*:**

Observing Toronto specific data, there exists signs of possible inefficiencies in the Expedia Business Model. There exists a significant amount of people who visit the Expedia site frequently but book a relatively low percentage of the time. This user usage indicates customer interest in the various travel packages, however this interest is not fit with corresponding booking rate. Additionally, there is a large portion of the population who have infrequent site usage and low booking proportions. The problem arises: How can we increase booking proportions for both population subsets through either decreasing the amount of sessions or increasing the amount of bookings?

***Methods:***

First, we cleaned the data from the first data set by removing any unwanted columns, and assigning proper class types to the variables for our purposes. We then subset the entire Toronto population through user region and user province. We then grouped the data by user ids after which we aggregated the frequency of user ID appearance to obtain the number of sessions and bookings per user. We plotted these results in terms of sessions and booking proportions (bookings per sessions). Then we were able to translate srch\_destination\_ids into the corresponding cities by creating a function we called extractcity from the 2nd data set. We then merged a simplified version of the 2nd data set with just cities and search\_destination\_ids to match the cities searched for by each user. Afterwards, we discovered the top six most popular booked and searched destinations from Toronto respectively, and plotted using a circular bar graph technique.

***Conclusions:***

The possible inefficiencies in the Expedia model could partially explain the mismatch between the most booked and most searched destinations from the city of Toronto. A potential explanation for these inefficiencies could be the limited user profile customization on the Expedia site; resulting in a lack of targeted advertisement. This classification will also help lend insight into the clusters of who we deem as the “Non-Committals” (i.e. Users with high site-frequency and low booking-frequency). Alternatively, some form of characteristic traits on the user IDs could help incentivize these individuals to book more often. This will help create a more succinct user experience.