**Datathon 2019 -- Predicting Rare Events**

Github link: https://github.com/Yuhan-Liu-Heidi/Datathon-2019

**Background**

In digital advertising, a “conversion” refers to the event when the shopper clicks on the ad and performs a valuable action such as signup, registration, or makes a purchase. Since “conversion” is a measurable event, it represents a reasonable proxy for the number of customers acquired during the ad campaign. Increasingly, brands and agencies looking to put a value on the Return on Advertising Spend (ROAS) require marketers such as us to optimize the ad spend such that customer acquisition is maximized.

In order to wisely spend the limited marketing dollars, we need to identify the shoppers who are more likely to respond to our ad and convert. While the number of devices to target is nearly one billion, the number of conversion events range from just a few hundreds to few thousands during the period of the ad campaign. In other words, these conversion events are extremely rare.

Data: Provided by Valassis, a leader in marketing technology and consumer engagement.

**Methods**

**1. Convert with false alarm**

Introduction

**File**:

**Question**: Will this shopper convert with minimal false alarm?

**Importance**: With a given shopper and their interest profile, this machine learning algorithm will be able to tell whether they are likely to convert, thus advise the marketers on whether to send this customer more digital advertisements.

Process: How did you clean and prepare the data, and what data did you use?

Data used:

Clean and prep:

Analysis: What analytical techniques did you use, and why?

Findings: What did you discover (include visualizations)?

Conclusion: What can a layperson at Valassis conclude from your team’s work?

**2. Convert rate within category**

Introduction

**File**: [customer\_analysis.py]()

**Question**: Which categories of shoppers are more likely to convert?

**Importance**: With the given data set of shoppers and their interest profiles, this program finds the interest category with the highest shopper conversion rate. This will help the marketer decide which category of customers to gear their advertisements towards.

Process: How did you clean and prepare the data, and what data did you use?

Data used:

Clean and prep:

Analysis: What analytical techniques did you use, and why?

Findings: What did you discover (include visualizations)?

Conclusion: What can a layperson at Valassis conclude from your team’s work?

