

# Report: Investigating User Patterns in the Nunavut eBird Dataset using DBSCAN and Linear Regression

Clara Risk, May 2018

## Executive Summary

### *Motivation & Goals*

Wood et al (2011) stated the majority of the eBird observations are submitted by a small number of users (“power users”). They asserted that about 90% of the observations come from just 10% of participants (Wood et al, 2011). A quick look through the eBird data for June 2015 for the Canadian territory of Nunavut indicates that this phenomenon may help explain the distribution of observations. Interestingly, it seems that eBird observations in Nunavut are not correlated with population. For example, given that Iqaluit, the capital of Nunavut, has the highest population in the territory, it would be expected that the majority of the eBird observations come from this area. However, most of the observations come from Bylot Island (in Sirmilik National Park). Bylot Island is remote – it is located at around 73°N and the nearest town (Pond Inlet) is a 2.5-hour flight from Iqaluit. Given the information about eBird power users from Wood et al (2011), this report will investigate the hypothesis that eBird observations in Nunavut are more correlated with the presence of power users than population, using a case study time period of June 2015.

### *Results Highlight*

Three power users (defined for this report as the top 10% of users) were found in the eBird June 2015 dataset for Nunavut located in DBSCAN Cluster 3 (Karrak Lake Area or W Karrak Lake), Cluster 9 (Akimiski island – OMNR base camp), and Cluster 10 (primarily Camps 1 & 2 on Bylot Island). Simple linear regression confirmed that the presence of these users in a cluster accounted for approximately 49% ( $R^2 = 0.49$ ) of the variation in the number of observations for June 2015. The multiple linear regression (MLR) showed that the adding the independent variables of population and income only increased the  $R^2$  to 0.52. In the MLR, the standardized regression coefficient for the presence of a power user was 0.75, greater than that of population or income. It can be concluded from this study that the presence of a power user in a cluster is a significant explanatory variable for eBird observations in Nunavut.