

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

# **APPLIED DATA SCIENCE CAPSTONE PROJECT**

PART I

AUGUST 16, 2021

## **THE BATTLE OF NEIGHBORHOODS**

RISHI SAXENA

# INTRODUCTION

Travel between the United States and Canada has commenced once again with the borders officially opening. In the year 2019, approximately 36 million travelers crossed the U.S.-Canadian border. It is expected that these numbers would compound in the coming years to overcome the 2020 bottleneck. With the market finally finding a footing, experts predict that new businesses would flourish and travel between the two countries would soar higher than ever. As a consequence of business, international travel for employment and immigration should increase as well.

One of the key challenging elements of this travel is identifying neighborhoods that suit one's preferences. Finding suitable accommodation can be an up-hill battle for most travelers. With the rising costs of real estate and stark contrast in the structure of available amenities – making a wrong decision in accommodation planning has the potential to be catastrophic.

This project aims to provide a high-level solution to this problem. By employing the use of Machine Learning Classification models, this project endeavors to compare neighborhoods between two prominent cities and identify ideal locations for given specifications. By providing a way to statistically evaluate these similarities and identify similar neighborhoods between these two cities, this venture aims to simplify relocation efforts between the United States and Canada.

In the future, this project has the potential to expand and support relocators between multiple different cities in North America. This expansion may one day even support relocators between different continents. Additionally, this could also be a handy tool for small and large businesses alike – to assess whether a neighborhood or locality is suitable for their business. The possibilities are endless.

It should be noted that this program only offers guidelines based on publicly available data and should not be solely used for making decisions. Consequently, the accuracy of this program is a direct function of the accuracy of the source data.