Classification of Tourist Towns in the United States

David Sachs

December 15, 2019

1. Introduction

1.1 Background

In 2014, 76.9 million people visited the Unites States of America. This industry represents roughly 2.8% of the entire US GDP, with a spend of $251 billion dollars. A key feature of this is that people come to visit different regions and cities within the US for very different reasons. America offers everything from historic locations to world-class skiing to warm beaches to amazing cities. Tourists must research and plan which regions they want to visit based off their interests. The US tourist industry gains value through positive tourism experiences, as this encourage repeat trips. This therefore means that tourists correctly selecting travel locations that match their desired vacation is critical to the overall longevity of US tourism. Being able to create clusters or classifications of tourist towns both enables tourists to determine what cities are interesting, and to find more like them for subsequent trips.

1.2 Problem

Data that might contribute to determining tourist town classifications include the locations and stores that make up that area, as well as the names of those locations. This project aims to classify tourist towns and create clusters to help tourists identify towns of interest and similar locations.

1.3 Interest

People planning trips to the US would be interested in this tool, as it is an easy way to understand what a city has to offer. Cities would be interested as well, as this helps them understand potential competition as well as if there is another cluster they would rather be a part of. Retail companies and restaurants would also interest, as this may help them find similar markets to open additional locations.

2. Data Acquisition and Cleaning

2.1 Data Sources

The types and names of locations in small towns can be found using the FourSquare API. The list of 25 samples towns was gathered from the US News Best Small Towns list available here. Other potential location lists were considered, but would require removing significant portions of the list such as national parks due to a lack of nearby locations. The latitude and longitude of the locations was gathered using Google's Geocoding API. Examples of the target data would be the locations in Telluride, Colorado, gained from FourSquare. The types of locations and words in the location names would then be used to classify the towns.