DATA 2010: Group Project
Full Written Report

Arjun Sarkar, Sachin Bhatt, Sahil Sharma, Gustavo Nunez

April 4th 2025

1. Introduction

As a group we collectively decided to focus on the **Travel Reviews** dataset from the UCI Machine Learning Repository for our term project.

2. Tentative Analysis Questions

Here are the following questions we will be answering in our analysis.

- Is it possible to determine distinct traveler preference groups by looking at how they rate various travel categories?

- Are there various traveler groups who share similar preferences?

- Which traveler preferences are most common across all travel categories?

3. Dataset Selection

It essentially consists of many reviews of East Asian places in 10 categories which are in this collection. Our goal is to derive valuable insights from the dataset by analyzing it using statistical methods and data visualization.

3.1 Dataset Context

The dataset contains 10 Features(excluding **User ID**) and 980 instances. Also, it is important to note that this dataset supports classification and clustering tasks.

The 10 different types of travel destinations the travelers gave ratings on are **Art Galleries, Dance Clubs, Juice Bars, Restaurants, Museums, Resorts, Parks/Picnic Spots, Beaches, Theaters, and Religious Institutions**

The following is a mapping of each traveler rating:

- Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0)

4. **Cleaning Up Dataset**

5. **Data Summarization/Aggregation**

6. **Data Visualization**

7. **Key Findings**

8. **Source**

Dataset Link: https://archive.ics.uci.edu/dataset/484/travel+reviews