

# **Final Project Proposal — TravelSafe: A Data-Driven Analysis of International Travel Safety Indicators**

## **Team Members**

Sylvan Wang (USC ID: 6176724527, [zichenw@usc.edu](mailto:zichenw@usc.edu))

Ningjun Li (USC ID: 5528989772, [ningjunl@usc.edu](mailto:ningjunl@usc.edu))

## **1. Problem Statement**

International travelers, especially students, often rely on fragmented or biased information when evaluating the safety of different countries. Travel advisories from governments and media coverage vary widely and can create misleading impressions. To provide a more objective understanding, this project aims to examine how measurable safety indicators differ across countries and identify which demographic or regional factors are most strongly associated with higher travel risk. Our goal is to construct a unified global dataset and use statistical analysis to reveal meaningful safety patterns.

## **2. Data Sources and Collection**

The project draws on both API-based data and web-scraped information. Structured country-level attributes such as names, regions, subregions, population, and geographic classifications will be collected through the REST Countries API using Python requests. This dataset forms the contextual backbone.

Safety indicators will be obtained by scraping the Wikipedia page “List of Countries by Intentional Homicide Rate,” which contains a clean HTML table with homicide rates and corresponding years. Data will be extracted with BeautifulSoup or pandas. If time allows, we will also scrape the U.S. Department of State’s travel advisories to add a categorical safety measure (Levels 1–4). After collection, all datasets will be merged using country names, with naming inconsistencies resolved during cleaning.

## **3. Data Cleaning Plan**

Data cleaning will include removing HTML tags, standardizing inconsistent country names, converting homicide rates into numeric form, addressing missing or incomplete entries, and removing duplicates. All fields will be formatted in English to meet project guidelines. The final output will be a single, consistent pandas DataFrame ready for analysis.

## **4. Analysis Plan**

We will begin by identifying the countries with the highest and lowest homicide rates and then compare safety levels across regions to determine whether geographic clusters show distinct patterns. Correlation analysis will help us understand how indicators such as population size, region, or optional economic classifications relate to safety outcomes. If time permits, we will also explore simple clustering to group countries into broader “risk tiers,” providing a multidimensional interpretation of travel risk.

## **5. Visualization Plan**

The project will produce visualizations using Matplotlib. These will include bar charts comparing homicide rates across countries, correlation heatmaps showing relationships among indicators, and, if feasible, regional or map-based graphics that illustrate geographic variation. Additional visuals such as radar charts may be used to highlight differences between selected countries.