

# Travel Score – Open data for tourists

## Benchmark Results

### 1. Introduction

As discussed in the Analysis Report<sup>1</sup>, there is a lack of open<sup>2</sup> datasets in the context of travel and tourism. This project set out to explore, whether it is possible to create such a dataset based on existing open data sources. While the Analysis Report describes, how such a dataset can be created, it remained unclear, how accurate the resulting dataset actually is. The following benchmark results shall answer this question.

### 2. Methodology

The Travel Score combined dataset, which is sourced from several open data sources, was compared to the Travel & Tourism Development Index<sup>3</sup>, which serves as the ground truth here (see Table 1). The comparison was realized by analyzing the correlation between similar indicators from both datasets.

Open Data Sources	Ground Truth
Inform Risk Index <sup>4</sup>	Travel & Tourism Development Index <sup>3</sup>
International Comparison Program <sup>5</sup>	
OpenStreetMap <sup>6</sup>	
Natural Earth <sup>7</sup>	

Table 1: Data sources, which the following benchmark is based upon.

### 3. Results

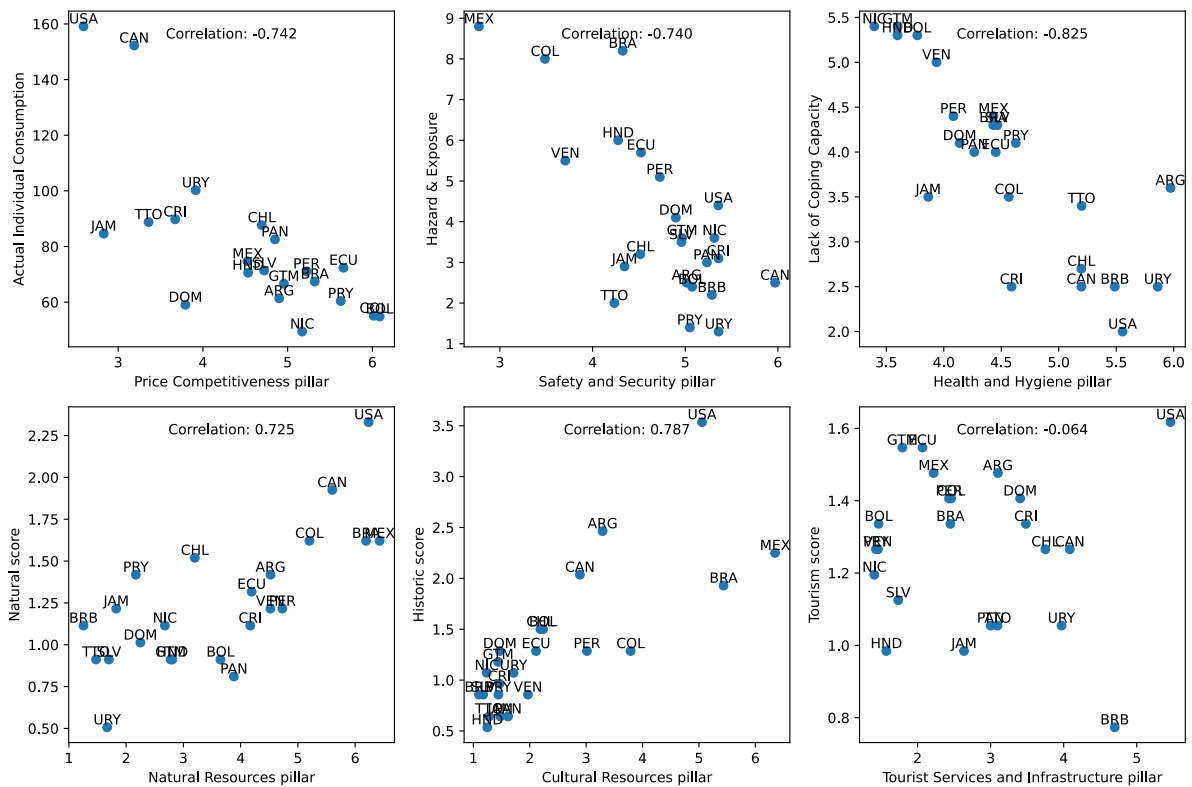


Figure 1: Open data indicators for American countries plotted in relation to their respective TTDI score.

## 4. Discussion

The benchmark results show a strong (either positive or negative) correlation for most pairs of open and TTDI indicators. This speaks for the quality of both datasets, considering that the indicators were calculated using different primary indicators and methodologies but still achieve very similar results in describing the real world situation.

The only exception is the OpenStreetMap-derived *Tourism score*, which has no correlation with the TTDI indicator *Tourist Services and Infrastructure*. In partial, this might be explained through the fact that the TTDI indicator is broader, including also investment and productivity aspects in addition to the availability of tourism resources itself. More likely however, this could indicate incomplete data on the OpenStreetMap side or a flaw in the methodology that was used to calculate the resource scores from the raw OpenStreetMap data.

As a consequence, the *Tourism score* and potentially also identically derived scores (*Natural score* and *Historic score*) should be inspected for flaws in the underlying data or methodology, for example by comparing them with additional datasets if available. Additionally, investigation of existing outliers from the otherwise well-correlated indicator pairs could lead to interesting insights.

## 5. Bibliography

1. "Travel Score - Analysis Report," <<https://raw.githubusercontent.com/tims777/travel-score/main/project/analysis-report.pdf>>
2. "The Open Definition," <<https://opendefinition.org/>>
3. "Travel & Tourism Development Index 2024," <<https://www.weforum.org/publications/travel-tourism-development-index-2024/>>
4. "INFORM Risk," <<https://drmkc.jrc.ec.europa.eu/inform-index/INFORM-Risk>>
5. "International Comparison Program," <<https://www.worldbank.org/en/programs/icp>>
6. "OpenStreetMap," <<https://www.openstreetmap.org/about>>
7. "Natural Earth," <<https://www.naturalearthdata.com/>>