

Is North America more suitable for tourists than South America?

The aim of my MADE¹ project is to answer this question based on open data.

Data Sources

As of now, I am using two main data sources, as well as one extra data source for visualization purposes.

Data Source 1: INFORM Risk

The INFORM Risk Index² is provided by the Joint Research Centre of the European Commission. Its purpose is to evaluate the risk of humanitarian crises. While the risk of humanitarian crises is obviously not the same thing, I have chosen this data source to indicate the individual risk for tourists, because the data is of high quality and has global coverage. Additionally, the index is not only open data but also open source, meaning that it is possible (although out of scope for this one-semester project) to adapt the methodology towards the risk profile of tourists.

The INFORM Risk Index is published under the Creative Commons Attribution 4.0 (CC-BY 4.0) license³. As such, it is possible to use the data without major limitations, provided that proper attribution is given. In addition to the attribution already present here, I will add additional notices (directly or as metadata) to derivatives like the final report or the resulting dataset.

Data Source 2: International Comparison Program

The International Comparison Program⁴ is managed by the World Bank on behalf of the United Nations Statistical Commission. One of its aims is, to produce “comparable price level indexes (PLIs) for participating economies”. As such, it was a natural choice for me to use this dataset to compare the price levels of different American countries.

The results of the International Comparison Program 2021 are also licensed under the CC-BY 4.0 license⁵. Consequently, I will take the same measures as with the INFORM Risk dataset to ensure proper license adherence.

Extra Data Source: Natural Earth

The Natural Earth dataset⁶ is a public domain map of the world, which is primarily intended for visualization purposes but a closer inspection shows that there is also lots of useful metadata attached to the dataset.

As the Natural Earth dataset is in the public domain (CC0)⁷, the data can be used without limitations and no attribution is required.

Data Pipeline

My data pipeline is based on Dagster⁸. After comparison of different competing frameworks, I chose Dagster because it offers a good compromise between framework size and feature set.

Result and Limitations

License Glossary

CC-BY 4.0 – Creative Commons Attribution 4.0: <https://creativecommons.org/licenses/by/4.0/> 1

CC0 – public domain: <https://creativecommons.org/publicdomain/> 1

Bibliography

1. “Methods of Advanced Data Engineering,” <<https://oss.cs.fau.de/teaching/specific/made/>>
2. “INFORM Risk,” <<https://drmkc.jrc.ec.europa.eu/inform-index/INFORM-Risk>>
3. “INFORM Risk - Copyright,” <https://commission.europa.eu/legal-notice_en>
4. “International Comparison Program,” <<https://www.worldbank.org/en/programs/icp>>
5. “International Comparison Program - Details,” <<https://datacatalog.worldbank.org/search/dataset/0066092/International-Comparison-Program-2021>>
6. “Natural Earth,” <<https://www.naturalearthdata.com/>>
7. “Natural Earth - Terms of Use,” <<https://www.naturalearthdata.com/about/terms-of-use/>>
8. “Dagster,” <<https://dagster.io/>>