DSC412

Milestone 2: Project Proposal

Minyoung Suh

Problem Statement

What cities are the best place to travel in 2025 summer?

Proposed Solution

Decision Trees or Random Forest algorithm for comprehensive decision making with the consideration of 3 or more important criteria (*e.g.* safety, cost of living, airfare from/to RDU, and weather conditions)

Potential Stakeholder

- Airlines
- Travel agents
- Travel lovers (like me!) – as this is a very first attempt of mine to build a machine learning model, I want to stay with the topic that interests me and benefit me the most.

Potential Obstacles

- Multiple data sets from different sources might have limited compatibility with each other (*e.g.* not having many cities and countries shared, etc.)
- Publicly available data sets might not have the most recent updates for 2025

DSC412

Milestone 2: Project Plan

Minyoung Suh

Data Resource

www.kaggle.com and more

Data Collection Plan

There are multiple publicly available indexes representing the attributes that I am considering for input measures:

- Regarding Safety measures:
  Numbeo Safety index (https://www.numbeo.com)
  Safe Cities Index (https://safecities.economist.com)
  Global Peace index (https://www.visionofhumanity.org)
  OECD Better Life Index (https://www.oecdbetterlifeindex.org)
  UN Office on Drugs and Crime (https://www.unodc.org)
  Gallup Global Law and Order Report (https://news.gallup.com)

- Regarding Living Cost measures:
  Big Mac index (https://www.economist.com/big-mac-index)
  Global Petrol Prices (https://www.globalpetrolprices.com/)
  Numbeo Cost of Living Index (https://www.numbeo.com)

Data Organization

I will have to compile the relevant input and output data sets for a list of global cities.

Data Analysis

Based on the statistics on the total number of tourists for each city, I will have to find the weight for each attribute. For example, it is highly possible that safety and living cost are negatively correlated with each other. There needs to be a process to balance them out.

Machine Learning Model

Decision Trees or Random Forest model

Model Validation

Not sure yet, but I would like to explore with K-Fold Cross-Validation.

Literature on Similar Topic

Dai, F., Wang, D., & Kirillova, K. (2022). Travel inspiration in tourist decision making. *Tourism Management*, *90*, 104484. https://doi.org/10.1016/j.tourman.2021.104484

Shin, H., Nicolau, J. L., Kang, J., Sharma, A., & Lee, H. (2022). Travel decision determinants during and after COVID-19: The role of tourist trust, travel constraints, and attitudinal factors. *Tourism Management*, *88*, 104428. https://doi.org/10.1016/j.tourman.2021.104428

Karl, M., Bauer, A., Ritchie, W. B., & Passauer, M. (2020). The impact of travel constraints on travel decision-making: A comparative approach of travel frequencies and intended travel participation. *Journal of Destination Marketing & Management*, *18*, 100471. https://doi.org/10.1016/j.jdmm.2020.100471

Karl, M. (2018). Risk and Uncertainty in Travel Decision-Making: Tourist and Destination Perspective. *Journal of Travel Research, 57(1),* 129-146. https://doi.org/10.1177/0047287516678337

Bowen, J. and Whalen, E. (2017), "Trends that are changing travel and tourism", *Worldwide Hospitality and Tourism Themes, 9(6),* 592-602. https://doi.org/10.1108/WHATT-09-2017-0045

Garg, A. (2015). Travel Risks vs Tourist Decision Making: A Tourist Perspective. *International Journal of Hospitality & Tourism Systems, 8(1)*, 1-9.