Analyzing the Safest Vacation Destinations for Women Traveling Solo

Team 16 | Shemaine Benson, Lynnetta Bonsu, Daisy Okpa, Shantel Williams

Business Problem

Given the unfortunate fact that gender-based violence is a reality for women who travel alone, we want to find the best and safest destinations for them to go. We will achieve this by researching various factors of safety (ex. crime rates, access to embassies, inclusiveness, etc.) in specific travel destinations in order to come up with a safety index. Women should be able to travel with a peace of mind. Our vision is to provide women with a platform where they can find the best destinations for solo traveling based on this safety index. This will take the guesswork out of potential threats and precautions. We also want to provide a space for women to report their experiences and recommendations as well.

Business Impact

- This will impact our target because we want women to feel safe while traveling abroad and we are providing a service that allows them to assess the safety of a variety of areas based on country level reports and user-submitted data.
- It will also provide an economic increase to tourist companies inside of that region because it would highlight that area for safer travel.
- It will help women become more aware of their surroundings and potential danger.

Data

Crime Index by Country dataset
- `Country`: Countries around the world
- `Crime Index`: The higher the number, the more crimes reported in the country
- `Safety Index`: The higher the number, the more safe the country
- `Year`: This column will be created as the site has data that goes back to 2012, but before we combine we would need to make sure the years are included

The main disadvantage of this dataset is that the indices are not able to be split into groups of interest (women, minority, LGBTQ+). Advantages of the dataset (once combined) is being able to see a change in crime and safety over time, and having available values for 129 countries.

Social Progress Index: xls file

- `Country (string):` Country name
- `Death from interpersonal violence (float):` Age-standardized deaths rate (per 100,000 people) from interpersonal violence. Interpersonal violence is defined as death or disability from intentional use of physical force or power, threatened or actual, from another person or group not including military or police forces.
- `Discrimination and violence against minorities (float):` An assessment of the level of domestic security and the degree to which other citizens can be trusted.
- `Acceptance of gays and lesbians(float):` Is the city or area a good place or not to live for gay or lesbian people
- `Freedom of expression (float):` To what extent does government respect press & media freedom, the freedom of ordinary people to discuss political matters at home and in the public sphere, as well as the freedom of academic and cultural expression
- `Perceived criminality (float):` An assessment of the level of domestic security and the degree to which other citizens can be trusted.
- `Women with advanced education (float):` Proportion of females (age-standardized) with 12–18 years of education.

This dataset consists of 107,000 pieces of data collected and vetted from the past 11 years for 168 countries. One advantage of this dataset is that it contains an abundant amount of data points to measure social progress for most nations of the world. A disadvantage of this dataset is that it does not contain data on gender- specific population rates within each country. Another disadvantage is that the data is not city specific, which can hinder the ultimate goal of the project, which is to provide information for various destinations, not just country specific.

Legatum Prosperity Index:

- `Physical security of women (int):` A composite measure of the physical security of women within a country, encompassing (a) the degree to which women are protected from domestic violence and rape, (b) the degree to which there are taboos or norms preventing these crimes being reported, and (c) the level to which honour killings and femicide occur.
- `Safety walking alone at night (int):` Feel safe walking alone at night in the city or area where you live?
- `Equal treatment and absence of discrimination (int):` A composite measure of whether individuals are likely to be discriminated against in court, at jobs, by police or other institutions based upon their socio-economic status, ethnicity, sexuality, or resident status.

- `Non-discriminatory civil justice (int):` A composite measure of whether the civil justice system discriminates in practice based on socio-economic status, gender, ethnicity, religion, national origin, sexual orientation, or gender identity

This dataset contains approximately 50,000 rows of data. An advantage of this dataset is that it contains a very rich amount of data to measure inclusive societies around the world, which is beneficial to determining safety for women traveling solo. A disadvantage is that the data is not city-specific.

Methods

In order to obtain a more accurate assessment of the data, it will be beneficial to use exploratory data analysis to summarize our datasets. Below are some methods that we will use:
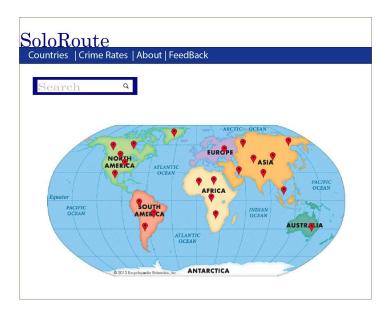
Visualizations

- Heatmap of crime index in each travel destination
- Histograms of discrimination and violence against minorities and physical security of women(univariate)
- Time series graph showing how crime rates fluctuate over time (univariate)
- Scatter plot to show the relationship between crime index and safety walking alone at night

Models

Based on the current scope of our project, we plan on using a Linear Regression Model to determine a safety index for each travel destination. The factors that we will use to determine this will be things like: crime rating, travel distance to embassy, census information (amount of women vs. men), wealth, culture, quality of life, amount of travel insurance claims, etc. Each factor will have a weight based on what we feel is important for women that contributes to our overall safety index.

We also want to incorporate Factor Analysis to help strengthen our linear model. We will gather some of the smaller factors and assess their ability to be used as a group. If time permits, we also thought it would be neat to use our linear regression model as a machine learning tool to predict what our safety index could be for an area in the future.

Interface

This is a simple mock up of our website idea. We want the user to be able to search for a location and be presented with the safety index. We also want the user to be able to click a map marker and receive statistics, charts or tables with information based on ths factors that went into our safety index. This mock up is subject to change. But this is a general idea we came up with. Of course the real website would display actual text and our objective.

Milestones

Version 1: An interactive dashboard where users can see a safety index for each county

Version 2: A website that allows you to search different destinations and their crime rate.

Version 3: An app that allows for users to submit their reviews on destinations, to view the overall ratings of countries for safety

Timeline

Date	Deliverable Details
Week 1	Program Introduction
Week 2	Team Formation

Week 3	Work on idea formation
Week 4	Idea should be finalized & Start on Scoping
Week 5	Finishing up scope and researching Datasets
Week 6	Team Project Submission: Dataset Details
Week 7	Team Project Submission: Dataset Cleaning
Week 8	Team Project Submission: Preliminary EDA
Week 9	Team Project Submission: In -Depth EDA
Week 10	Team Project Submission: Analysis/Modelling
Week 11	Create Ideas or Drafts for Front- End Mockup
Week 12	Team Project Submission: Front-End Mockup
Week 13	Team Project Submission: Front-End/DashBoard Design
Week 14	Start working on project coding/data filtering
Week 15	Continue working on project coding and design
Week 16	Finishing up last minute Bugs or Code
Week 17	Team Project Submission: Application/Dashboard Live

Concerns

Our main concern is finding datasets representative of women for our topic. Another one is finding data on all the factors based on specific destinations (i.e. cities). Determining different variables to generate our linear model is an area we are not too familiar with, but we know it would strengthen our analysis/modeling. Luckily, these concerns can be addressed through further research and learning. Overall, we want to add our own spin on how we interpret the metrics presented to us in order to truly make this a unique and special service for solo women travelers.