**Data Immersion Task 6.1**

1. **Summary**

* This is an open-source dataset that shows us the type of travelers and to see if there are any patterns that could help travel agencies or just people who like to travel in general determine what is popular and when.

1. **Data Source:**

* Traveler data (excel file) Citation: <https://www.kaggle.com/datasets/rkiattisak/traveler-trip-data>
* countries.geojson (spatial data) Citation: https://pkgstore.datahub.io/core/geo-countries/countries/archive/23f420f929e0e09c39d916b8aaa166fb/countries.geojson
* time series data. Citation: https://data.nasdaq.com/data/EIA-us-energy-information-administration-data?page=2

1. **Data Contents:**

* It has destination, start date, end date, duration, travelers name, travelers age, travelers’ gender, travelers nationality, accommodation type, accommodation costs, transportation type, and transportation costs.

1. **Data Cleaning:**

* Changed any abbreviations in nationality column.
* Changed any abbreviation in destination column.
* In Jupyter I checked for duplicates, missing data, and shape of the data.

1. **Limitations and Ethics:**

* An ethics issue could be that the data does provide personal information on people including their name, age, nationality, and gender.
* In destination there are a few entries that just list the country and not the city like most of the entries.

1. **Data Profile:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column title** | **Column description** | **Data type** | **Time variant/ invariant** |
| **Trip ID** | ID number for the trip | Quantitative | Invariant |
| **Destination** | Where the person is taking a trip to | Qualitative | Invariant |
| **Start date** | The start of the person’s trip | Quantitative | Variant |
| **End date** | The end of the person’s trip | Quantitative | Variant |
| **Duration** | The number of days the person spent on the trip | Quantitative | Variant |
| **Traveler name** | Name of the person traveling | Qualitative | Invariant |
| **Traveler age** | Age of the person traveling | Quantitative | Variant |
| **Traveler gender** | Gender of the person traveling | Qualitative | Invariant |
| **Traveler nationality** | Nationality of the person traveling | Qualitative | Invariant |
| **Accommodation type** | What kind of place the traveler is staying at | Qualitative | Invariant |
| **Accommodation cost** | The cost of where the traveler is staying | Quantitative | Variant |
| **Transportation type** | What kind of transportation the traveler is using to get to their destination | Qualitative | Invariant |
| **Transportation cost** | The cost of how much that transportation was to get to the destination | Quantitative | Variant |

1. **Questions to Explore:**

* Does a certain country cost more to travel to or to stay in than another?
* According to the folium map France and Canada have the darkest coloring indicating that they would be the most expensive.
* Is there an age when people tend to travel more or less?
* As per the results found so far it seems that young adults aged 20-35 are more likely to travel and stay longer on their trips.
* Does the cost of accommodation play a factor in the duration that someone takes for their trip?
* It seems like more people would be willing to have longer trips if the price of where they were staying was lower. But we can see that as the price increases the duration of stays decreases.
* Does cost affect the duration of someone’s trip?

1. **Hypothesis:**

* If a traveler is between the ages of 20 to 35, they are more likely to travel and stay longer on their trip then someone 51 and older
* If the cost of transportation is low then travelers will be more likely to stay longer for their trip