

My Vacation Planner: Personalised, data-driven vacation recommendations

Skills City – Data Analytics Bootcamp (DA – APR24)

|  |  |
| --- | --- |
| **Project Group** | |
| Roberta Johnson (Project Team Leader)  David Sheffield | [*Bertajohnson15@gmail.com*](mailto:Bertajohnson15@gmail.com)  [*Davesheff@live.co.uk*](mailto:Davesheff@live.co.uk) |
| Swati Patil | [*swatideshmukh84@gmail.com*](mailto:swatideshmukh84@gmail.com) |
| Evante Oduro | [*evanteoduro@gmail.com*](mailto:evanteoduro@gmail.com) |
| Yafet Wolde | [*Japhetwolde@gmail.com*](mailto:Japhetwolde@gmail.com) |
| Mohammed Hamza Ali | [*Mohammed.hamza.ali@outlook.com*](mailto:Mohammed.hamza.ali@outlook.com) |

|  |  |
| --- | --- |
| SUPERVISED BY | *Victoria Mellor* |
| CLIENT’S DELEGATE FULL NAME | *Laetitia Sterling* |
| CLIENT’S DELEGATE EMAIL ADDRESS | [*Laetitia\_sterling@uk.ibm.com*](mailto:Laetitia_sterling@uk.ibm.com) |

|  |  |
| --- | --- |
| **Document Submission Date**  *23/6/24* | ***Final Document Wordcount***  4,668 |

Contents

[Abstract 3](#_Toc2039507906)

[Introduction 4](#_Toc8511459)

[Methodology 4](#_Toc1331685404)

[Initial Research, the Team’s Approach to the User Interface and Establishing Comparable Data Measures for Possible Destination Countries 6](#_Toc503522024)

[Sustainability and Emissions Data Reporting Measures 8](#_Toc1216981201)

[Sourcing and Presentation of Cost of Living, Weather, Hotel, Restaurant and Attractions Data. 11](#_Toc1978955416)

[Key findings and Trends in Dashboard. 13](#_Toc1882444336)

[Conclusion 16](#_Toc1234979301)

[Recommendations 16](#_Toc1749663660)

[Bibliography 17](#_Toc964504320)

[Appendix 22](#_Toc1301857971)

# Abstract

This project report will guide you through the development of a comprehensive PowerBI (Power Business Intelligence) dashboard that our client, IBM, requested from the group via a project brief in April 2024. The brief requested the creation of a platform that provides personalised vacation recommendations for users based on open data sources that encourages environmentally responsible and sustainable travel choices with a priority of limiting their impact on the environment. The project found that rail, ferry and transport modes that avoid using flights are the most sustainable with European countries also ranking the best destinations for sustainability.

# Introduction

IBM has been dedicated to every client's success and creating innovations that matter to the world. (1) They bring together technologies and services to help their clients solve the most pressing problems. The goal is to offer users insights into travel destinations, accommodations, and activities to enhance their vacation experience, all with a priority of limiting their impact on the environment. (1)

This report will detail how the project team attempted to meet the brief set out by IBM in line with the above goal and the client’s values. The report will, firstly, summarise the methodology used to create the dashboard including how the team decided to layout the dashboard, what data was necessary to find to achieve this and where all the data was sourced from. The report will then detail in the substantive chapters how the team worked together to accomplish the sourcing and research of data and creation of the dashboard broken down sequentially in the order in which each of the steps were completed to arrive at the final product before explaining how the data was analysed and the limitations that were faced.

Next the report will detail the project team’s conclusions and findings from the creation of the dashboard and provide the team’s overall views on sustainable transport planning and the insights that can be gained from the dashboard in its current form.

The report will then detail conclusions that were drawn from the process of creating the report as well as from the anaylsis of the data taken from the dashboard in its current form and provide recommendations to the client and the user based on the dashboard.

Finally, the report will make recommendations to the client company about their next steps to further develop the product including what the project team would have done differently/additionally given further time and what the team would prioritise when implementing the findings of the project as well as how what the team learned could help the client company.

# Methodology

The research question that was issued to the project team within the brief was - ***“How can personalized, data-driven algorithms enhance vacation planning by providing tailored recommendations based on individual user preferences?”***

The team were also provided with a purpose and set of aims and objectives in the brief as follows:

**PURPOSE:**

* To employ predictive analytics by analysing historical data and real-time market trends. This enables accurate travel demand forecasting, helping businesses prepare for peak periods, optimize resource allocation, and minimize revenue loss.
* This vacation planning project with open data leverages travel-related datasets to provide users with personalised and data-driven vacation recommendations, making their trip planning experience more enjoyable and efficient

**AIMS AND OBJECTIVES:**

* IBM is aiming to provide personalised vacation recommendations for users based on open data sources and encouraging environmentally responsible and sustainable travel choices.
* The goal is to offer users insights into travel destinations, accommodations, and activities to enhance their vacation experience, all with a priority of limiting their impact on the environment.
* It helps companies see trends and what customers like or don’t like. It also helps them understand how customers feel. This is useful for making business plans and finding new ways to sell.
* BI is adaptable and beneficial for businesses of all sizes. Smaller travel businesses can leverage BI tools to enhance efficiency, understand customer preferences, and compete effectively. The scalability of BI solutions makes them suitable for enterprises of varying sizes.

In order to ensure the purpose and aims of the project were met, the team used Discord - an instant messaging and VoIP social platform - to communicate via message thread and online audio/video meetings. The team held a series of initial meetings to go through the project brief and brainstorm early thoughts on how to approach the project and assign the responsibilities of Project Team Leader, note-taker etc. Following this, the team met at least once per week to update and review progress, assign tasks to team members and review the project plan for the remaining time to ensure the project was completed on time.

During the first round of initial research and to get a flavour of the amount and types of available data each team member picked a country they have a potential interest in visiting and did some fledgling quantitative research to find available and ideally recent or live data in the following areas which may be useful for the project - Weather (live data), Prices For Accommodations, travel options, restaurants, events, sustainability, recommendations. The team created a shared spreadsheet to hold all of the data found and then reviewed the data before refining our approach to data sourcing in order to ensure that the data for each country was comparable.

Next, each team member was assigned to qualitatively research different elements of the client’s suggested activities from the brief - Data Collection and storage, Data Cleaning and Pre-processing, Data Analysis, User Preferences and Profiling, Interactive Travel Itinerary, Recommendations and Suggestions, Real-time Weather Updates, Carbon Footprint. This initial research was discussed in team meetings so as to agree and refine the team’s approach to completing the task.

Once it was agreed between the members the comparable data for destinations that the team would use and the approach the team would take to transforming this into a PowerBI dashboard, it was then time to progress with collecting the data, cleansing it and reiteratively loading it into the PowerBI dashboard before working as a team to create the most impactful and user friendly visualisations and interface for a prospective user. After several phases of the Extract, Transform, Load (ETL) data integration process (2) as well as numerous of full-team reviews of the layout and format of the visualisations and controls in PowerBI, the team agreed that the dashboard was ready to present to the client and had progressed sufficiently in the time available such that further improvement and development would require presenting it in its current form to the client and receiving feedback.

Having provided a summary of the methodology used to create the dashboard, this report will, now in the following chapters, describe in detail the main methodological decisions and processes used to create the dashboard before detailing the team’s conclusions and recommendations.

# Initial Research, the Team’s Approach to the User Interface and Establishing Comparable Data Measures for Possible Destination Countries

This section of the report will detail how the team conducted their initial qualitative and quantitative research as well as how the final decisions were made about what data to use and how to present it in the dashboard as well as what factors were considered in making the decisions and the conclusions from these decisions.

Tourism is a social, cultural and economic phenomenon (3) which entails the movement of people to countries or places outside their unusual environment for person or business/ professional purpose. This project aims to innovate vacation planning by utilizing advanced data analytics.

Traditional travel planning methods, often reliant on manual research and generic suggestions, are increasingly complemented or replaced by digital tools. However, many existing systems offer recommendations based on broad categories, which may not adequately reflect individual traveller needs and preferences. This project highlights our ability to use data visualization and analytics tools to identify key insights and provide actionable recommendations, thereby facilitating data-driven decision-making at all levels.

As noted in the summary methodology section about, our initial approach as a project team was to each take a country, we were familiar with and then look to see what freely available data sets were available on the internet to an early sense of the types and quality of data we may be able to use. The team felt it was important at this stage to prioritise using open source data that was not behind a paywall or required a subscription as this may limit the usability of the product since this cost may have to be passed on to the user. Following several team meetings at which any data found by the team members, which had been stored in a newly created shared spreadsheet, was reviewed, the team came to a very important realisation and decision which affected the development of the project going forward. In order for the dashboard to have the functionality to compare different destinations against the same criteria, particularly sustainability criteria, the data for each potential destination must be comparable i.e. measure the same factor in the same format. It was this realisation that led us to make the decision to focus our efforts on finding data on the capital cities of countries the user may wish to visit. Capital cities, by their nature are the centre and focus of study and measurement in all manner of different ways and our initial research indicated that the quantity, quality and comparability of data available for capital cities was significantly higher which we anticipated would best support producing the optimum user experience.

After making this decision to focus on capital cities, the team engaged in a further round of data research focusing specifically on finding data sets under these categories that contained data for as many capital cities as possible:

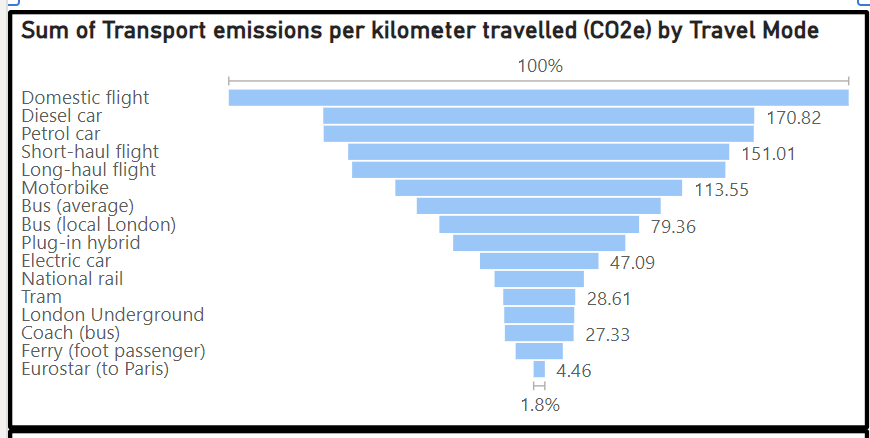
* Weather (live data)
* Prices For Accommodations
* Travel options
* Restaurants
* Events
* Sustainability
* Recommendations

Whilst searching for these data sets under the revised brief, the team were also mindfully beginning to consider how the data would need to be collected, stored, cleaned, processed, and analysed going forward in order to present it in the most useful way to the user.

Once the team had initial useable datasets for each topic area and the data was downloaded and stored in our shared spreadsheet, we began initial discussions about how to layout the dashboard. After importing the data into PowerBI Desktop and drafting various initial layout ideas, the team made the decision to prioritise offering the user high-level sustainability and CO2 emissions data by country and continent on the opening tab as well as cost of living and comparable pricing data by country on the next tab before allowing more detailed and granular data by country and city on the remaining tabs. The team felt this offered a smooth user experience by naturally starting with high-level data allowing the user to get an idea of continents, then countries they may wish to visit, perhaps those which are more sustainable, produce lower emissions and have a reasonable cost index, before then, on the remaining tabs, finding useful hotel, weather, restaurant and activities data for cities in those countries. In the next section we’ll detail how the team progressed with developing the dashboard with this approach in mind.

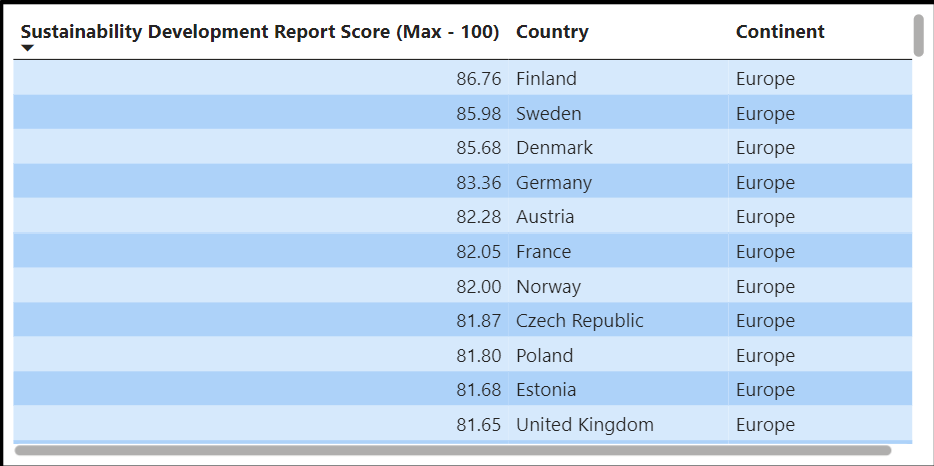
# Sustainability and Emissions Data Reporting Measures

This section of the report details how the sustainability and emissions data as well as the opening/introductory board of the dashboard were prepared. The first element of data that was present on the board was which travel modes emitted the most CO2 emissions. It was fell this was a key piece of data that would significantly affect which destinations the user my go on to select if sustainability is a priority. The team found the latest report by the UK government published in 2022 and widely cited set of accompanying data that provided a clear picture of this data and so this was prominently place on the opening tab. It indicated that



The data, whilst only data for the UK, gives a good indication that transport modes that avoid flights such as rail, ferry, Eurostar and electric cars have the lowest emissions.

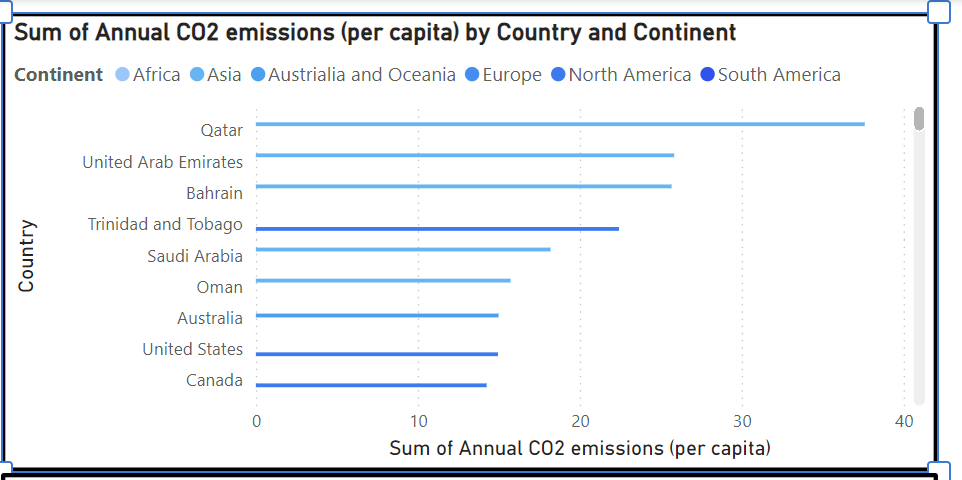
The next piece of top-level data added to the dashboard was a measure of sustainability and, after conducting research, the team settled on using the 2024 [Sustainable Development Report 2024 (sdgindex.org)](https://dashboards.sdgindex.org/) . This report has been published each year since 2016 incorporating global data coverage and all of the data from the report is available to download and use. It measures sustainability of country across 17 key areas combined to provide an overall score.



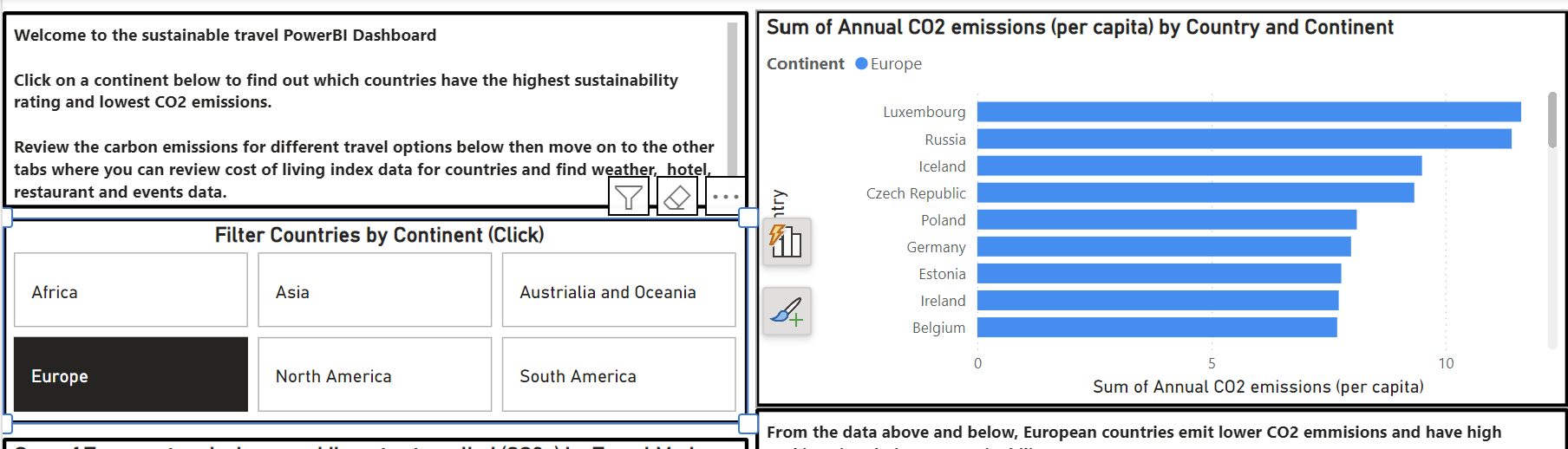
The data was filtered such that those countries ranked highest for sustainability were at the top and clearly indicated that European countries currently lead the table in meeting sustainability progress.

Finally on the first dashboard tab, CO2 emissions data by country for all countries was sourced from a 2022 European Commission report

<https://edgar.jrc.ec.europa.eu/report_2022> and loaded into PowerBI to produce a horizontal bar chart:



The final element of the opening tab of the dashboard relates back to our earlier decision to focus on comparable data from capital cities. Each of the emissions sets of data discussed above were cleaned and processed in such a way that a continent was logged for each country. This allowed a slicer mechanism to be used in PowerBI enabling the user to filter the results by continent:



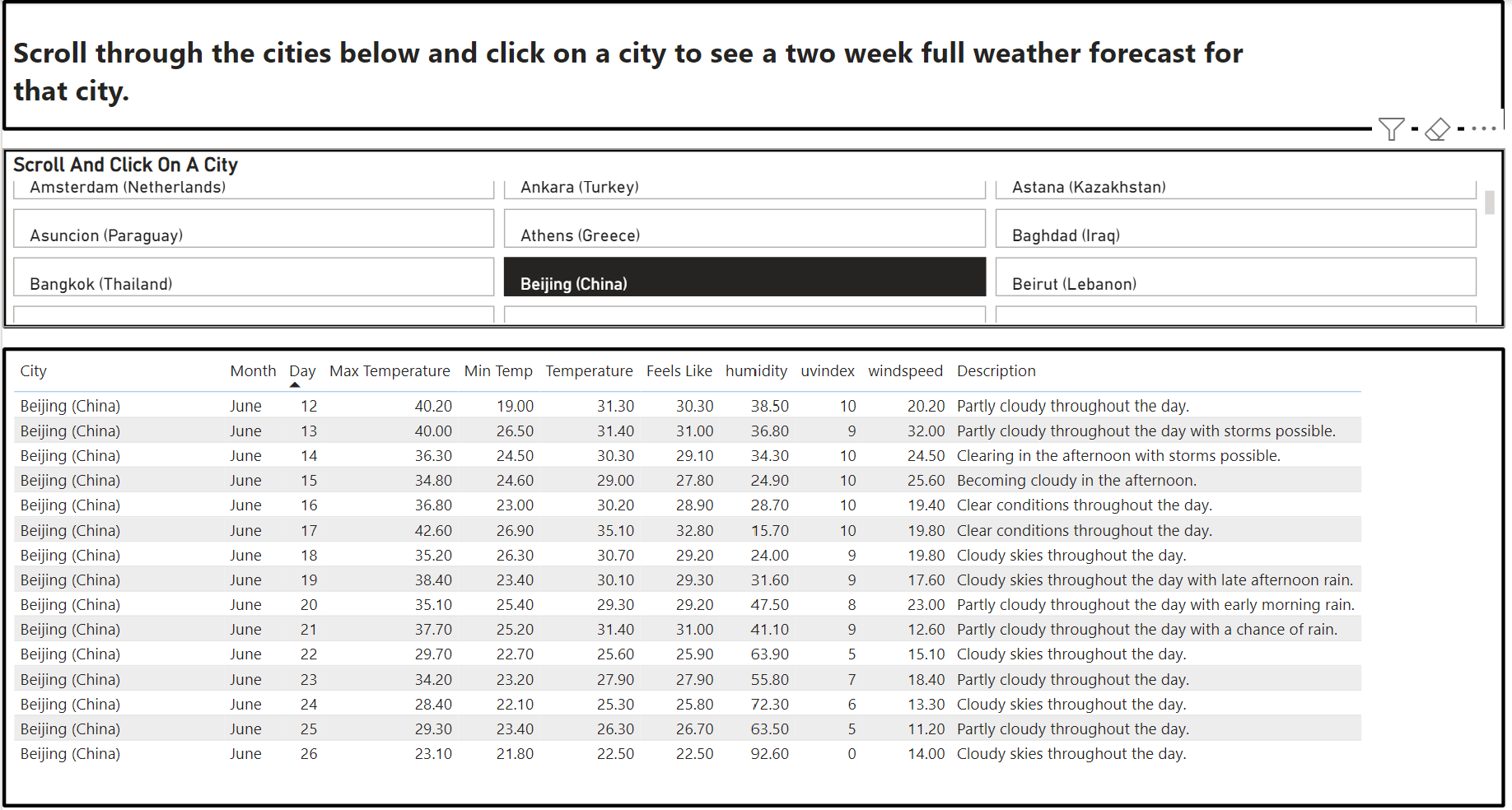
This filter by continent ability helps guide the user to the clear conclusion that European countries are much more sustainable and offer lower emissions overall when factoring in the transport modes required to visit them which will inform the user experience and choices on the proceeding tabs of the dashboard. Having discussed the opening tab of the dashboard, this report will now detail how the remaining tabs were created using the approach of progressing from top-level data by continent and country to granular-level data country and city as mentioned earlier.

# Sourcing and Presentation of Cost of Living, Weather, Hotel, Restaurant and Attractions Data.

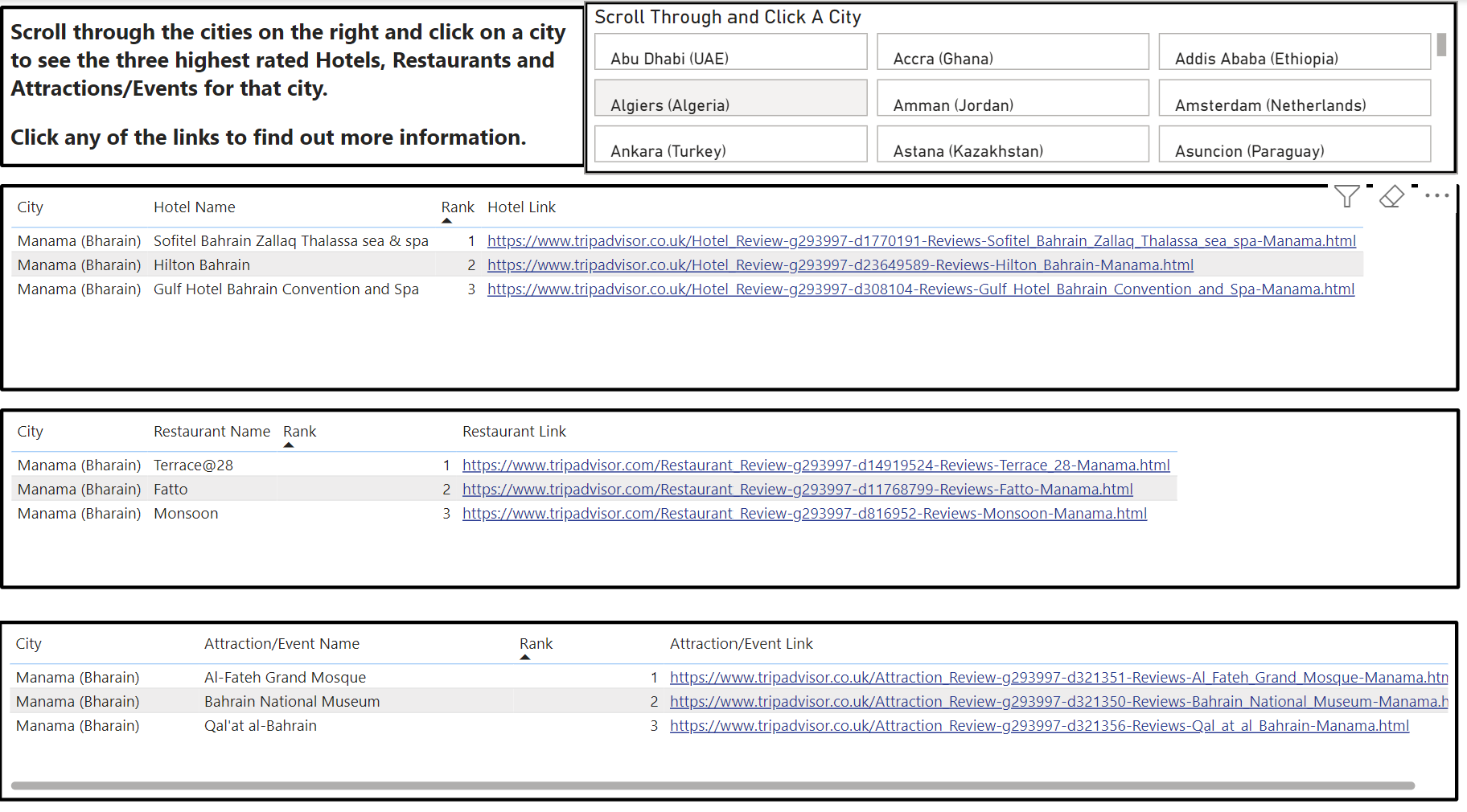
The next tab that the team created on the dashboard used recent data freely available on the Kaggle website (6) that details four different cost of living indexes for each city to give the user a varied summary of relatively expensive the city they choose to visit will be split in to costs for groceries, rent, restaurants and a combined cost of living index score. Each of these tables and be scrolled through and ordered highest to lowest or lowest to highest. There is also a world map chart on this tab so the user can zoom in on a continent and pick a country. This tab, in a similar way to the first tab has a slicer that allows for all of the data to be filtered by continent to help streamline the results for the user based on their preference.



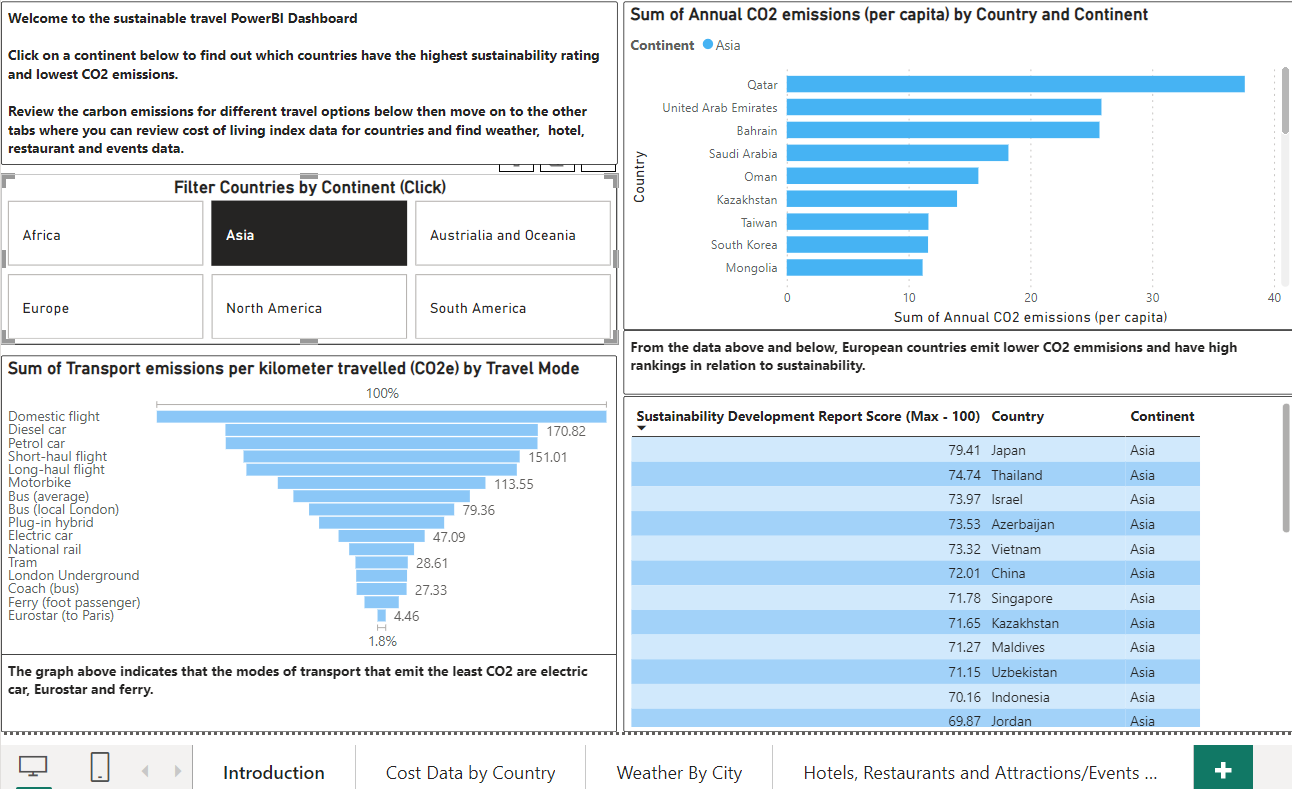
The third tab in the dashboard helps the user start to home in on particular cities they may wish to visit by providing a 14-day weather forecast for each of the 100 cities visualised in the dashboard. Detailed and accurate weather data was collected from the [Virtual Crossing website](https://www.visualcrossing.com/weather-api?ga_api20&key=weather%20api&loc=9046643&gad_source=1&gclid=CjwKCAjwydSzBhBOEiwAj0XN4GdomQRHZSgE8rylb2PZHud_FT9Bo_3dCC5SXQPqpJotyRpiYTPk9RoCds4QAvD_BwE) (7) that was then downloaded, cleaned to remove unwanted columns of data and then loaded in to the dashboard before being visualised in the form of a table with a slicer so that the user can scroll through all of the cities and countries listed and click on a city to filter the results.



The final tab on the dashboard is designed to be the final step in the process of the user determining which city they would like to visit based on the emissions, sustainability, cost of living and weather data they have been able to review so far. For this tab, the details of the top three hotels, restaurants and attractions for every city listed were manually gathered from the Trip Advisor website and processed into a large spreadsheet forming a significant portion of the nearly 22,000 data points in the dashboard overall. In order continue to make the user experience as smooth as possible, the data is presented in a similar format to the previous tab i.e. a filterable table. The tab contains a scrollable list of cities which, when clicked display only the information for that city with web links for every hotel, restaurant and attraction such that the user can, having determined which city they would like to visit, can now progress to a page with details of hotel and being the booking process should they wish.



# Key findings and Trends in Dashboard.

Figure 1.1 - Introduction page of Dashboard with Asia selected.

In figure 1.1, Asia the continent has been selected and it shows that Asia has a high level of sum of annual CO2 emissions. Qatar has the highest CO2 emissions, and this could be due to the country being a hot spot for travellers. In research, it showed that in Q1 of 2024 they welcomed 1.6M people with 47% of people travelling by air, by land 39% and by sea with 14%. A high contributor to CO2 emissions is air travelling. This is a probable cause for the high CO2 emissions in Qatar.

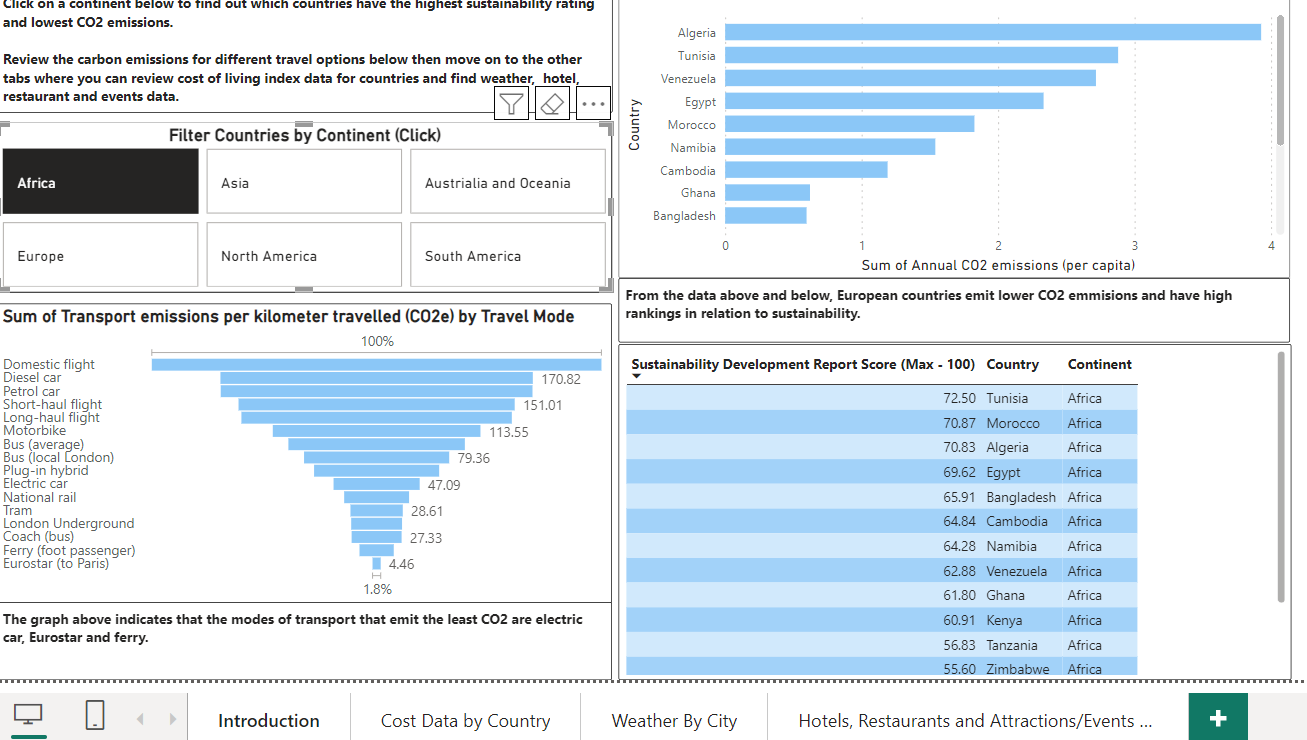
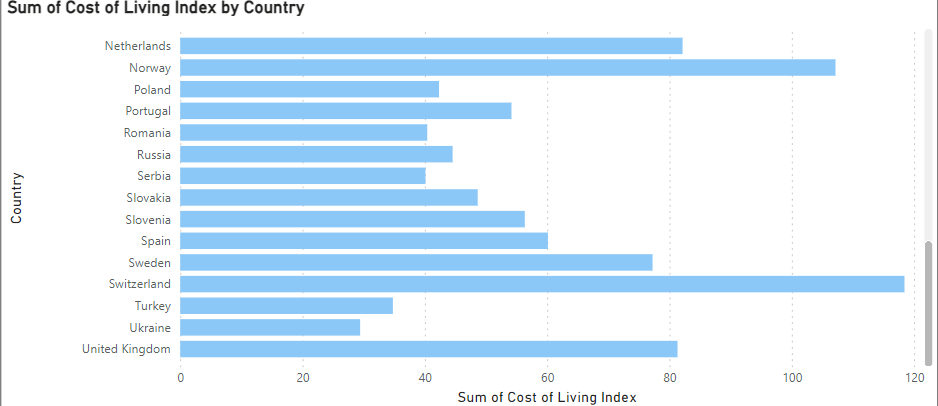
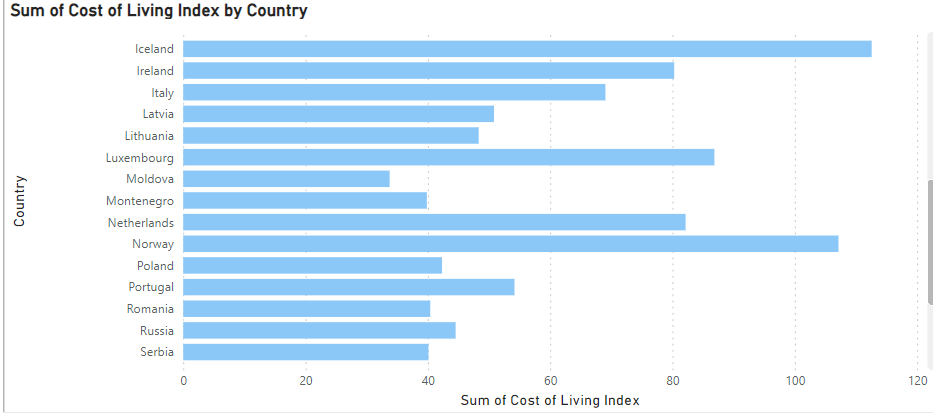


Figure 1.2 - Introduction page of Dashboard with Africa selected.

In figure 1.2, that continent Africa has been selected and it has the lowest amount of CO2 emissions annually. It's said that Africa only contributes to 4% of global carbon emissions with 1.45 billion. They have fairly high numbers of sustainability development but as you go down the numbers decrease, and this could be due to countries which are underdeveloped will not have as many resources to create methods/plans for sustainable development.

Figure 1.3 - Cost data by country: Europe

In figure 1.3, the country selected is Europe and it shows that to have a holiday in Europe is expensive. The prices for restaurants are expensive and also grocery shopping. This shows to stay in Europe for a holiday would require the traveller to bring more money with them to be able to have a comfortable holiday stay. In the image, it shows that Iceland, Switzerland & Norway are quite expensive holiday destination as all three countries are at the top of the groceries index and restaurant index. The higher the index the more expensive the prices are. In the cost-of-living index in these three countries is also high, as shown in the images below.



# Conclusion

This project report details the creation of a PowerBI dashboard for IBM, aimed at delivering personalised and sustainable vacation recommendations. We analysed a broad spectrum of data on destinations, transport modes, accommodations, and activities, emphasising sustainability. Our findings revealed that travel by rail and ferry, particularly to European destinations, significantly reduces environmental impact compared to air travel. We underscored the importance of integrating real-time, comparable data and a user-friendly interface to enhance the vacation planning experience. The dashboard combines data-driven insights with a commitment to eco-friendly travel. Future steps should focus on expanding data sources and optimising algorithms for even more precise recommendations.

# Recommendations

An essential element of any project is to determine what can be delivered within the constraints: Time, Money, People, Resources. Many of the activities and suggestions may not be possible within the timescales available.

Here are some key recommendations to enhance the user experience and drive continuous improvement.

* In current project we failed to complete activities like User Preferences and Profiling, Interactive Travel Itinerary and utilisation of data-driven algorithms to recommend travel destinations based on user preferences, travel trends, and current events.

To overcome this, we can expand data sources for more personalized recommendations.

**Expand Data Integration:**

* Integrate additional APIs and data sources for more comprehensive and up-to-date information.
* Ensure real-time updates for dynamic travel conditions and deals.
* Incorporate a wider range of data sources, such as local events, seasonal activities, and user reviews, to provide more comprehensive and relevant destination recommendations.
* **Visual Itinerary Builder:** Develop an interactive, drag-and-drop itinerary builder that allows users to easily customize and visualize their vacation plans.
* **Budgeting and Cost Estimates:** Integrate tools that provide real-time cost estimates and budgeting features to help users plan their vacations within their financial constraints.
* **Focus on refining and optimizing the recommendation algorithms** using the latest machine learning techniques. Emphasize the hybrid model combining collaborative filtering and content-based filtering for better personalization.
* Up-to-date information on travel deals, weather conditions, and local events, making the planner more reliable and attractive. This can set the company apart from competitors who may not offer real-time updates.
* Continuously refine algorithms using the latest machine learning techniques to improve recommendation accuracy.
* Establish a process to regularly incorporate customer feedback into product roadmaps and implement updates to continuously improve the vacation planning experience.

# Bibliography

**References – Where we have got data from**

1. Skills City - (2024) IBM Project Brief, accessed May 2024 - June 2024.
2. IBM (2024) - <https://www.ibm.com/topics/etl,> accessed May 2024 - June 2024.
3. UN Tourism (2024), <https://www.unwto.org/glossary-tourism-terms#:~:text=Tourism%20is%20a%20social%2C%20cultural,personal%20or%20business%2Fprofessional%20purposes.>
4. Sustainable Development Report (2024) - <https://dashboards.sdgindex.org/>
5. European Commission - Emissions Database for Global Atmospheric Research (2022) <https://edgar.jrc.ec.europa.eu/report_2022>
6. Kaggle (2024) [Cost of Living Index by Cities (kaggle.com)](https://www.kaggle.com/datasets/debdutta/cost-of-living-index-by-country)
7. Virtual Crossing (2024) <https://www.visualcrossing.com/weather-api?ga_api20&key=weather%20api&loc=9046643&gad_source=1&gclid=CjwKCAjwydSzBhBOEiwAj0XN4GdomQRHZSgE8rylb2PZHud_FT9Bo_3dCC5SXQPqpJotyRpiYTPk9RoCds4QAvD_BwE>
8. Kaggle (n.d.), *Best Restaurants in Germany*, <https://www.kaggle.com/datasets/mahnoorshahidshakir/best-restaurants-in-germany?rvi=1>, accessed May 2024 - June 2024.
9. OECD (n.d.), *Portugal Data*, <https://data.oecd.org/portugal.htm>, accessed May 2024 - June 2024.
10. Electricity Maps (n.d.), *Electricity Maps - Portugal*, <https://app.electricitymaps.com/zone/PT>, accessed May 2024 - June 2024.
11. Turismo de Portugal (n.d.), *Open Data*, <https://dadosabertos.turismodeportugal.pt/>, accessed May 2024 - June 2024.
12. EU Data (n.d.), *Data Europa*, <https://data.europa.eu/en>, accessed May 2024 - June 2024.
13. DataHub (n.d.), *REN DataHub - Portugal*, <https://datahub.ren.pt/en/>, accessed May 2024 - June 2024.
14. Meteostat (n.d.), *Weather Data for Lisbon*, <https://meteostat.net/en/place/pt/lisbon?s=08535&t=2024-05-08/2024-05-15>, accessed May 2024 - June 2024.
15. OECD (n.d.), *Domestic, Inbound, and Outbound Tourism - Greece*, <https://stat.link/qehykj>, accessed May 2024 - June 2024.
16. Thomson, T. (2023), *A Guide to Data Analytics in the Travel Industry*, <https://www.alation.com/blog/data-analytics-in-travel-industry/>, accessed May 2024 - June 2024.
17. Giatamedia (n.d.), *My Hotel - Hotel Directory*, <https://myhotel.giatamedia.com/hotel-directory?page=1&country=DE>, accessed May 2024 - June 2024.
18. UNWTO (n.d.), *Sustainable Development - Biodiversity*, <https://www.unwto.org/sustainable-development/biodiversity>, accessed May 2024 - June 2024.
19. Future Market Insights (n.d.), *Greece Tourism Market*, <https://www.futuremarketinsights.com/reports/greece-tourism-market>, accessed May 2024 - June 2024.
20. Original Travel (n.d.), *Things to Know Before Traveling to Greece*, <https://www.originaltravel.co.uk/travel-guide/greece/things-to-know-before-travelling>, accessed May 2024 - June 2024.
21. Hellenic Statistical Authority (n.d.), *FSS 2023*, <https://www.statistics.gr/en/fss-2023>, accessed May 2024 - June 2024.
22. Kaggle (n.d.), *BI Professional Certificate - Create Charts*, <https://www.kaggle.com/code/fekihmea/bi-professional-certificate-create-a-charts>, accessed May 2024 - June 2024.
23. SETE (n.d.), *Sustainability and Greek Tourism*, <https://sete.gr/viosimotita-kai-ellinikos-tourismos/>, accessed May 2024 - June 2024.
24. INSETE (n.d.), *Air Data Tracker*, <https://insete.gr/bi/airdata-tracker/?lang=en>, accessed May 2024 - June 2024.
25. Hikersbay (n.d.), *Climate Data for Greece*, <https://hikersbay.com/climate/june/greece?lang=en>, accessed May 2024 - June 2024.
26. OECD (n.d.), *Air and GHG Emissions Data*, <https://data.oecd.org/air/air-and-ghg-emissions.htm>, accessed May 2024 - June 2024.
27. Office for National Statistics (n.d.), *TravelPAC*, <https://www.ons.gov.uk/peoplepopulationandcommunity/leisureandtourism/datasets/travelpac>, accessed May 2024 - June 2024.
28. NOAA (n.d.), *Daily Summaries for Switzerland*, <https://www.ncei.noaa.gov/access/search/data-search/daily-summaries?bbox=47.450,8.467,47.304,8.613&startDate=2023-04-01T00:00:00&endDate=2024-04-01T23:59:59&pageNum=1>, accessed May 2024 - June 2024.
29. Inside Airbnb (n.d.), *Get the Data*, <https://insideairbnb.com/get-the-data/>, accessed May 2024 - June 2024.
30. Zurich.com (n.d.), *Culture in Zurich*, <https://www.zuerich.com/en/api/v2/data?id=96>, accessed May 2024 - June 2024.
31. World Weather (n.d.), *Zurich Live Weather*, <https://worldweather.wmo.int/en/city.html?cityId=312>, accessed May 2024 - June 2024.
32. Hikersbay (n.d.), *Accommodation Prices in Switzerland*, <https://hikersbay.com/prices/switzerland?lang=en#hotel-prices>, accessed May 2024 - June 2024.
33. Zurich.com (n.d.), *Sustainable Hotels in Zurich*, <https://www.zuerich.com/en/hotels-bbs-apartments-and-hostels-your-accommodation-in-zurich?f=sustainability%3A391&page=1>, accessed May 2024 - June 2024.
34. EU Data (n.d.), *Zurich Tourism Data*, <https://data.europa.eu/data/datasets?query=zurich%20tourism&locale=en>, accessed May 2024 - June 2024.
35. World Weather (n.d.), *Weather in Italy*, <https://worldweather.wmo.int/en/country.html?countryCode=176>, accessed May 2024 - June 2024.
36. WeatherSpark (n.d.), *Historical Weather Data for Rome, Italy*, <https://weatherspark.com/h/y/71779/2020/Historical-Weather-during-2020-in-Rome-Italy#google_vignette>, accessed May 2024 - June 2024.
37. Statista (n.d.), *CO2 Emissions Data for Italy (1970-2022)*, <https://www.statista.com/statistics/449779/co2-emissions-italy/>, accessed May 2024 – June 2024
38. · Lonely Planet (n.d.), *Best Places to Visit in Italy*, <https://www.lonelyplanet.com/articles/best-places-to-visit-in-italy>, accessed May 2024 - June 2024.
39. · TripAdvisor (n.d.), *Luxury Beach Resorts in Italy*, <https://www.tripadvisor.co.uk/HotelsList-Italy-Luxury-Beach-Resorts-zfp10281610.html>, accessed May 2024 - June 2024.
40. · Sabino, C. (2018), *7 Fabulous and Affordable Places in Italy You Need to Check Out Now*, <https://www.forbes.com/sites/catherinesabino/2018/05/10/7-fabulous-and-affordable-places-in-italy-you-need-to-check-out-now/?sh=7d55f488174a>, accessed May 2024 - June 2024.
41. · YouTube(n.d.), *How to Plan a Travel Itinerary*, <https://youtu.be/u71JhuYLsrE?si=UXFO6JYrNbZFgOJ2>, accessed May 2024 - June 2024.
42. · YouTube (n.d.), *How to Do Machine Learning in Power BI Desktop*, [https://www.youtube.com](https://www.youtube.com/), accessed May 2024 - June 2024.
43. · Microsoft (n.d.), *Build a Machine Learning Model in Power BI*, <https://learn.microsoft.com/en-us/power-bi/guidance/tutorial-build-machine-learning-model>, accessed May 2024 - June 2024.
44. · Microsoft (n.d.), *Data Sources in Power BI Desktop*, <https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-data-sources>, accessed May 2024 - June 2024.
45. · YouTube (n.d.), *Real Time Weather Updates*, <https://www.youtube.com/watch?v=vlnS7JlLRQU>, accessed May 2024 - June 2024.
46. · Carbon Monitor (n.d.), *Carbon Monitor Europe*, <https://eu.carbonmonitor.org/>, accessed May 2024 - June 2024.
47. · City Carbon Footprints (n.d.), *Carbon Footprint Data for Cities*, <https://citycarbonfootprints.info/>, accessed May 2024 - June 2024.
48. · Our World in Data (n.d.), *CO2 Emissions per Capita*, <https://ourworldindata.org/grapher/co-emissions-per-capita?tab=table>, accessed May 2024 - June 2024.
49. · Electricity Maps (n.d.), *Live Data on Electricity and Carbon Intensity*, <https://app.electricitymaps.com/map>, accessed May 2024 - June 2024.
50. TripAdvisor (n.d.), TripAdvisor Main Website, <https://www.tripadvisor.com/> , accessed May 2024 - June 2024.

# Appendix

**Appendix -A Raw Data**

This appendix provides the raw data collected from external sources, which was used to generate personalized vacation recommendations.

**Table 1: Destination Data**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Continent | Country | Capital City | Cost of Living index | Rent Index | Grocery Index | Restaurant Price Index |
| Africa | Egypt | Cairo | 25.68 | 5.24 | 21.31 | 23.26 |
| Africa | Venezuela | Caracas | 27.82 | 8.43 | 28.4 | 23.63 |
| Asia | India | New Delhi | 28.67 | 8.73 | 25.8 | 25.30 |
| Asia | Uzbekistan | Tashkent | 29.94 | 9.43 | 24.99 | 21.66 |
| Europe | Ukraine | Kiev | 29.42 | 14.53 | 23.17 | 23.51 |

**Table 2: CO2 Emission and Sustainability**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Continent | Country | Capital City | Annual CO2 emissions (per capita) | Sustainability Development Report Score (Max - 100) |
| Africa | Egypt | Cairo | 2.333106 | 69.62 |
| Africa | Venezuela | Caracas | 2.716869 | 62.88 |
| Asia | India | New Delhi | 1.996682 | 63.45 |
| Asia | Uzbekistan | Tashkent | 3.48306 | 71.15 |
| Europe | Ukraine | Kiev | 3.557854 | 76.52 |

**Table 3 Modes of transfer and Co2 Emission**

|  |  |
| --- | --- |
| Travel Mode | Transport emissions per kilometer travelled (CO2e) |
| Bus (average) | 96.5 |
| Bus (local London) | 79.36 |
| Coach (bus) | 27.33 |
| Diesel car | 170.82 |
| Domestic flight | 245.87 |
| Electric car | 47.09 |
| Eurostar (to Paris) | 4.46 |
| Ferry (foot passenger) | 18.74 |
| London Underground | 27.81 |
| Long-haul flight | 147.87 |
| Motorbike | 113.55 |
| National rail | 35.49 |
| Petrol car | 170.48 |
| Plug-in hybrid | 68.4 |
| Short-haul flight | 151.01 |
| Tram | 28.61 |

**Table 4 City, Hotels, Restaurants and Events Data**

|  |  |  |  |
| --- | --- | --- | --- |
| City | Rank | Hotel Name | Link |
| Helsinki (Finland) | 1 | Hotel F6 | <https://www.tripadvisor.co.uk/Hotel_Review-g189934-d10263887-Reviews-Hotel_F6-Helsinki_Uusimaa.html> |

|  |  |  |  |
| --- | --- | --- | --- |
| Restaurant Name | Link | Attraction/Event Name | Link |
| Ravintola Rioni | <https://www.tripadvisor.com/Restaurant_Review-g189934-d23087799-Reviews-Ravintola_Rioni-Helsinki_Uusimaa.html> | Sea Fortress Suomenlinna | <https://www.tripadvisor.co.uk/Attraction_Review-g189934-d199918-Reviews-Sea_Fortress_Suomenlinna-Helsinki_Uusimaa.html> |

The raw data collected forms the basis of the personalized vacation recommendations. By analyzing user preferences, destinations, Co2 emission and sustainability data, hotels, restaurant, events data, the vacation planner can match users with suitable destinations that are sustainable with their interests and budget.

**Appendix -B Algorithms for Vacation Recommendations**

This appendix provides detailed descriptions and code of the algorithms used to generate personalized vacation recommendations.

The script for code is as follows:

import pandas as pd

df = pd.read\_csv('/content/Best Restaurants in Germany - Recovered\_Sheet1.csv')

df.dropna(subset=["Quality Score", "Postal Code", "City", "Open for Lunch", "Open for Dinner", "Open for Midday", "Plant Holidays", "Price Per Person", "Courses", "Chef", "Culinary Category", "Stars"], inplace=True)

df.sort\_values("Rank")

df.drop\_duplicates(subset=None, keep='first', inplace=False)

from google.colab import drive

drive.mount('drive')

df.to\_csv('data.csv')

!cp data.csv "drive/My Drive/"

