Critical Analysis of Ireland Tourism

Ayush Kumar
MSc Data Analytics
National College of Ireland
Dublin
x21208590

Fathima Bi Rafi
MSc Data Analytics
National College of Ireland
Dublin
x21217084

Pavithra Ashokan
MSc Data Analytics
National College of Ireland
Dublin
x22133992

***Abstract*—This article is about a tourism significant activity that represents billions of visitors each year, portrayed by the country's natural beauty, deep history, ethnic heritage, and generosity. Here, the goal is to collect various tourism datasets which suggest tourists the better option to visit Ireland. The locality information is the link between all these dataset servers which performs initially the pre-processing, loading those data into server, then performing ETL and finally visualizing to obtain best tourist place as Ireland and people have better option to visit various places in Ireland and review their opinion. Overall, our study shows that Ireland has much to offer tourists in terms of natural beauty, history, and hospitality. This study provides a snapshot of Ireland's broader tourism industry and its potential to attract visitors from around the world by highlighting the businesses' strengths and potential.**

***Index Terms*—Locality, Accommodation, Database, Tourism, Insights, ETL, Mango platform, Folium, Bokeh, (key words)**

I. Introduction

A. Project Motivation

Our project oversees marketing Ireland as a tourist destination. Our primary goal is to attract foreign visitors to the island of Ireland in order to increase tourism revenue and create jobs in the tourism industry. Up to date so far, 30% of tourism had more of domestic visitors than in 'normal' pre-Covid year and 40% have been decreased currently. Simultaneously, at the same time, two-thirds (67%) trust that foreign visitor has been reduced this year compared to last two years. Up to date so far, 30% of tourism had more of domestic visitors than in 'normal' pre-Covid year and 40% have been decreased currently. Simultaneously, at the same time, two-thirds (67%) trust that foreign visitor has been reduced this year compared to last two years. Viewing ahead, 31% of accommodation operatives have more domestic bookings than usual this year. In summer, 60% of accommodation operators have reduced their abroad bookings.

B. Objective

Creating and implementing marketing campaigns in key overseas markets to promote Ireland as a vacation destination. Our project collaborates with the tourism dataset to support the promotion and elevation of tourism areas and accommodation services. This includes collaborating with hotels and restaurants for accommodation type, develop new routes for travel destination. Exploring Ireland as a varied and appealing destination with a rich cultural heritage, breath taking panoramas, and friendly welcome. This results in highlighting key attractions, experiences, and events that tourists can enjoy on vacation. This also helps to monitors and evaluates the efficiency of accommodation marketing.[2]

***Index Terms*—Accommodation, Information, Tourist attractions, Transportation, about each place.**

C. Significance

Tourism is the commercial activity of representing and supporting vacation and business travel.[1] Travel for business or desire is the fact and practice of exploring which submerges entertaining tourists, tourism attraction, accommodation, and commercial tours. [2] Ireland is rapidly turning into one of the most significant travel and tourism destinations in Europe. Ireland is a treasure trove with miraculous scenery and endless places to visit made up of the Republic of Ireland and Northern Ireland (part of the United Kingdom). Ireland is also famous for its rich panorama and magnificent primitive castles. Endlessly, it is not surprising that millions of people visit Ireland every year. In addition, we are increasing the abroad visitors and domestic visitors to Ireland through our project.

The tourism industry is becoming a progressively prominent part of the Irish economy.[3] To do this analysis here total 3 tourism datasets had been collected where the tourist spots has been labelled with its nature type like walking, sightseeing, trekking, landscape, public sculpture, island and many more. On other hand, address locality is marked along with its longitude and latitude.

D. Tools & Libraries Used

TABLE I
Libraries Used in API Datasets

Dataset	Libraries Used
API	pymongo,pandas,MongoClient,time,requests, folium,psycopg2,matplotlib.pyplot,bokeh.io, bokeh.plotting, bokeh.models,seaborn

We have created **Docker Desktop** containers to create separate containers for **mongodb_container**, **pgadmin4_container** and **postgres_container**. We have used **MongoDB Compass** for displaying the loaded datasets. The process follows ETL method where we can Extract the data, Transform and Load the data in server. Initially, ETL does

the extraction of tourism. data in this project and related dataset like accommodation and activity datasets has been extracted from certain dataset sources. Extraction is nothing but collection of datasets. Usually, these datasets are sourced from various websites or from various API data. Here the data will collect as per the requirements. In these datasets either the data can be with less columns or limited information, in such cases the next phase of extraction validation takes place. In next phase validation takes place where the data is checked initially. It checks whether the data is suitable for current problem, and if there is any possibility to extract the insights from extracted data. Moving on to next step Transformation phase takes place. In this phase the data is pre-processed. Pre-processing follows steps of checking the null values present in the data, checking normality of the data, removing or dropping out the duplicate rows which will be problematic in visualization, removing unwanted or irrelevant columns which are not relevant for visualizations. Furthermore, various data cleaning processes are done here like data screening/masking, data merging, data union joiner and filtering the data corresponding to its requirements.

Among these, data masking process is done for managing sensitive segment where deletion and duplication should be treated safely as any kind of mishandling data would result in error when visualizing the data. The transformation of data can be done through the help of python code in Jupyter Notebook, and the data pre-processing is done or mongo data base. Usually for data warehousing the data transformation is executed in server and from there the data is further loaded into it.[4] The final method in the ETL process is Loading data where the fully processed data is loaded into a data repository.

Finally, the visualization process is done using Bokeh. Bokeh is where python module run in web browser. Here, Bokeh is used to create dynamic visualizations of processed data after it has been loaded into a target destination. Bokeh's plot formation boundary is high-level, granting users to easily indicate the type of plot, source of data, and visual elements such as size, colour, and shape. Also, bokeh depicts these plots in web browser using web tools for inter activity such as zooming, panning (viewing in horizontal and vertical view) and hover tooltip. This visualization helps in obtaining subject rather than distracting the elements. Through this visualization we can straighten up tourists about the better accommodation, attraction, and activities of Ireland tourist spot. As a conclusion, our visualization attracts tourists to visit Ireland and for the ones who are planning to visit and enjoy rather than choosing other.

II. RELATED WORK

Min-Yuan Zheng et al. Our discoveries show that the higher the degree of contribution in the travel industry programs, the higher the desire and inspiration to travel, and the more the viewers see the vacationer region as an enchanting city. As a result, tourism programs can be used to help visitors have a unique experience. In order to create distinct thematic areas that cater to various types of visitors, the local community can make use of the resources, culture, and cuisine that are distinctive to it. Nearby state run administrations can use their own social and gastronomic and provincial the travel industry explicit assets to make enchanting urban communities with extraordinary highlights to address the issues of various sightseers [5] Shah Jahan Miah et al. The study aims to design and evaluate a "big data analytics" method to support strategic decision-making in tourism destination management using a design science methodology.[6] Nonetheless, these investigations are subject to certain restrictions such as the inability to account for additional contextual variables that may impact the success, inadequate number of participants, and lack of an all-encompassing assessment of diverse machine learning methodologies. Additionally, the inferences of these findings may differ based on the specific dataset, highlighting the requirement for additional research in this domain.

III. METHODOLOGY

A. Dataset 1: Activities dataset

Fig. 1. Activity dataset

Here the data is about activities dataset [5] from the source of data gov website. This dataset has the information of tourism activities from a particular data website. The given dataset seems to be a list of businesses in Ireland, including their name, website URL, phone number, geographic coordinates (longitude and latitude), address details (region, locality, and country), and tags related to their offerings or activities. The dataset most probably focused on businesses in the food and beverage and outdoor adventure sectors. Analysts, researchers, and businesses regularly use such datasets for a variety of purposes, including market study, customer division, and location-based analysis. They can support businesses in understanding the competitive landscape and detecting growth prospects, while also allowing clients to easily find and connect with relevant businesses in their area.

Labels or tags can provide additional information about the businesses and their offerings, for making it easier to sort and search the dataset. Geographic coordinates enable mapping analysis, which can expose patterns and insights about business distribution and their relationships with other factors such as population extent or tourist hotspots. On whole, the datasets like these can be useful for a change of purposes and can help us recognize the local business environment even better.

B. Dataset 2: Attraction dataset

The next dataset is attraction dataset. This dataset appears to be a catalogue of tourist destinations in Ireland, such as historical and natural landmarks. Each data contains a name, website URL, phone number, longitude and latitude coordinates, regions, locality of address, country, and a set of labels explaining the amenities and attractive places available at a particular destination. Activities included in tags are of all kinds of attraction like dining, shopping, bird watching, nature and wildlife observation, climbing, cycling, fishing, gardening, horse riding, kayaking, kitesurfing, windsurfing, photography, learning, sailing, swimming, visiting historic houses and castles, exploring islands and natural landscapes, and enjoying local food and drink at cafes, pubs, and restaurants.

Fig. 2. Attraction dataset

This dataset could be useful for tourists interested in a variety of activities and experiences who are planning a trip to Ireland. It may also be useful for tourism companies or researchers researching tourism trends and preferences in Ireland. It should be noted, however, that this dataset only includes a limited number of destinations and may not represent the full range of attractions available in Ireland.

C. Dataset 3: Accommodation dataset

This dataset contains a list of accommodations in Ireland, such as guesthouses, camping grounds, and caravan parks. Each entry contains a name, website URL (if available), phone number, longitude and latitude coordinates, address region, address locality, address country, and a set of tags describing the type of accommodation. The tags indicate whether the

Fig. 3. Accommodation dataset

accommodation is a guesthouse, camping site, or caravan park. This information may be useful for tourists looking for a specific type of lodging while vIt should be noted, however, that the dataset is limited in scope and may not represent all accommodations available in these regions. Furthermore, some entries lack a website URL, making it difficult for users to learn more about the accommodation. Overall, this dataset may be beneficial to tourists planning a trip to Ireland who are looking for specific types of lodging in specific regions. It could also be useful for travel companies or researchers looking into Irish accommodation trends. visiting Ireland. The dataset includes Cork and Kerry.

IV. Technologies Used And Reason for Usage

PYTHON: Python plays a great support with json data because it is open choice for data interchange between web services and python based functions. SQL : Further SQL has been used for transforming the data. This design is mainly to handle huge datasets and provide various optimization techniques and SQL will be helpful to handle in this case too to handle many datasets in ETL . MONGO DB: Here we use MONGODB because our dataset is unstructured API. Plotly: This visualization has been used to create interactive data visuals including charts and dashboards, in python and programming languages.

V. ETL process

A. Extraction

Before initiating the extraction process, some pre processing is done in python to perform transformation with no hassle. The scrapped data had lot of unnecessary data which can be removed before extraction. The existing data has been altered, dropped columns so that the scrapped data can be loaded into the MongoDB. The datasets are checked for null values, as there are no null values, we are proceeding for the ETL process of the datasets. It collects data from the API in batches and stores them in a list called Accommodation1, Accommodation2 and Accommodation3. Information on accommodations in Ireland that can be further processed or analyzed will be included in the resulting list. The API requests have been made and the data has been collected for the second and third datasets as well. The collected data have been stored in lists. In the extraction process of these datasets , [4] the scrapped data, and the links data has been extracted from the Mongo Db. The above meta data explains the renaming done for meta data for the necessary columns this is done as the dataset field names was irrelevant to the records and thus considered to rename the column values

B. Transformation

The transformation [6] phase of the ETL is where the main part happens and that is where the data cleaning, data pre-processing is done. Transformation phase will make the data clean ready to visualize, model to create dashboards or to do analytics as well. The first transformation is done by removing three columns from the dataframe '_id', '@context', 'image'. The remaining columns includes @type, name, url, geo, telephone, address and tags. For the extraction process, 'address' is extracted into City, County and Country. The 'geo'

coordinates are extracted into Longitudes and Latitudes. For the 'Type of venue', we extracted the second element from the tags list, such as 'Hotel', 'Hostel', 'B&B', etc, All the renamed new columns are stored in the pandas dataframe.

C. Loading

The new database is then loaded into PostgreSQL database, with the psycopg2 library. With the help of SQL queries, the table is loaded and stored in the PostgreSQL database. The three datasets have been loaded into Mongo and extracted as dataframe.

VI. RESULTS AND EVALUATION

The table from PostgreSQL is read using the cursor connection, for the visualization processes. The data is cleaned by dropping Country columns and checking for the null values. The required data is then obtained from all the three dataset tables.

To establish the relationship between the three acquired datasets, a pivot table is created containing name of county and number of accommodations, attraction points and activities present within County. This table is used for the further representations.

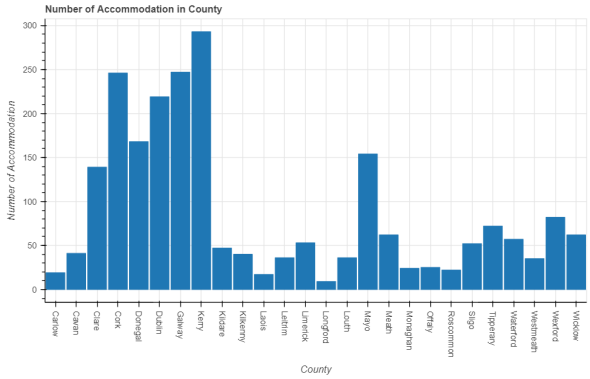

Fig. 4. Bar chart.

A bar chart is created with Number of accommodations in the County. The number of accommodations in each county of a particular region is visually represented. Data-driven decisions can be made based on the insights gained from this type of visualization. From this chart we can analyse that Kerry has highest number of accommodations present in it.

Using complete data of pivot table, we have plotted a bar graph which represent the number of accommodations, activities, and attractions in a county. As shown in figure we have plotted all three attributes together for better representations. The colours used in bars of graph represent the colours of the flag of Ireland, the background is kept blue to symbolise the Irish flag on a blue sky.

Analysing the figure 5, we got the information that although Kerry County has highest number of accommodations present in it, but Dublin County have more activities and attraction point in it.

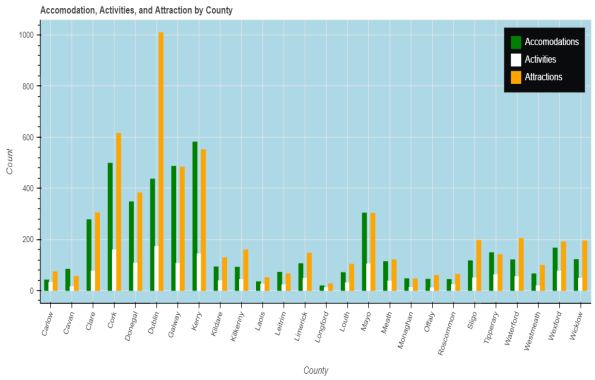

Fig. 5. Bar plot.

For exploratory data analysis we have plotted a heatmap to find correlations between the three attributes of table.as shown in Fig. 6, can interpreted that all the attributes are highly correlated to each other. That the counties with more attraction

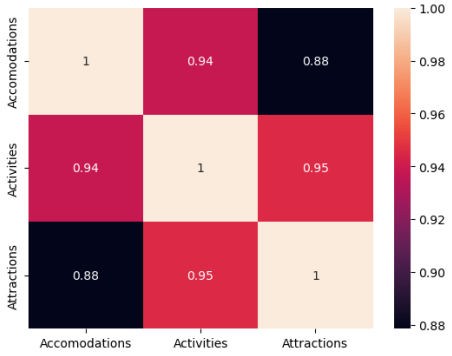

Fig. 6. Heat Map

point have more activities to do, and if there are more activities for tourists that means that there are more accommodations for them. Using this analysis, we are presenting a stacked bar chart of top 5 counties with all three attributes of table to suggest best county to visit in Ireland.

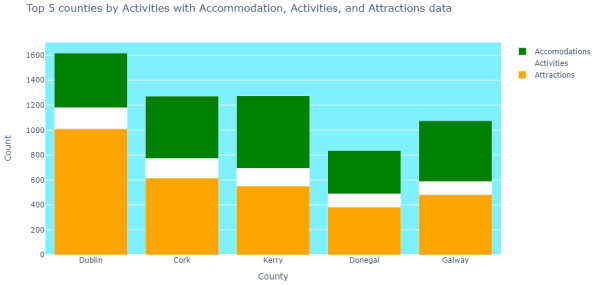

Fig. 7. Stacked Bar chart

An illustration of bars can exhibit diverse categories of tourism in Ireland, for instance, cultural tourism, thrill-seeking tourism, eco-friendly tourism, and others. The graph has the

ability to showcase the proportion of tourists intrigued by each category of tourism or the count of travelers who participate in definite activities or voyage to certain places. Furthermore, the graph might offer an understanding of the population of tourists who are fascinated by distinct categories of tourism, like age, sex, or origin.

In the figure above, we plotted a stacked bar chart between major counties such as Dublin, Cork, Kerry, Donegal and Galway. It can be seen that the best county to visit is Dublin, as all together it has a greater number of activities, accommodations, and attractions in it.

VII. CONCLUSION AND FUTURE WORK

It is evident that county Dublin is best,to visit according to data. Dublin has more historical and contemporary places to visit, it also has mouth-watering food with a wide variety of options from traditional Irish food to international cuisine. This attracts more tourists from around the world and in turn it significantly increases the number of accommodation sites. To help tourists, we have plotted food establishment in Dublin city centre which serve tasty food on map using the geo-coordinates present in data.The significance of real-time analytics is on the rise in the tourism sector since it empowers stakeholders to track and react to real-time variations in visitor behavior and other operational metrics. Precise and timely information regarding traveler's behavior, trends, and operational efficiency can be furnished to stakeholders by devising real-time analytics solutions using the datasets from Ireland's tourism industry.

Fig. 8. City map

As new data information is obtainable, the algorithms and tailored suggestions can be revised to guarantee that they are consistently enhancing and furnishing the most precise and pertinent recommendations for users. The tourism industry relies heavily on data analytics to create customized suggestions for tourists. Analyzing Ireland's tourism industry datasets can reveal patterns in visitor behavior and preferences, which can be utilized to generate personalized recommendations that cater to their interests and preferences.

REFERENCES

[1] Tourism (2023) Wikipedia. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Tourism

[2] Outline of tourism (2023) Wikipedia. Wikimedia Foundation.

[3] 50+ interesting Ireland Travel & Tourism Statistics (2023) (no date) RSS.

[4] K. S. a. Y. Lan, "SETL: A scalable and high performance ETL system," in System Science, Engineering Design and Manufacturing Informatization, 2012. https://www.dreambigtravelfarblog.com/blog/ireland-travel-tourism-statistics#: :text=10.,andaverageoverseas

[5] M.-Y. Zheng, C.-C. Chen, H.-H. Lin, C.-H. Tseng, and C.-H. Hsu, "Research on the impact of popular tourism program involvement on rural tourism image, familiarity, motivation and willingness," MDPI, 27-Apr-2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/9/4906.

[6] Author links open overlay panelShah Jahan Miah a, a, b, c, Highlights Available: https://www.sciencedirect.com/science/ article/abs/pii/S0378720616303573?via

[7] "Tourism," Wikipedia, 30-Apr-2023. [Online]. Available: https://en.wikipedia.org/wiki/Tourism.

[8] "Fáilte Ireland: National Tourism Development Authority," Failte Ireland - Accommodation Facts and Figures — Tourism Research & Reports. [Online]. Available: https://www.failteireland.ie/Research-Insights/Accommodation-Statistics-and-Reports.aspx.