BAX 422 DDR Final Project

**Developing a Go-to-Market Strategy for a Travel E-Commerce Platform in the United States**

Group 13: Avani Sharma, Jodie Chen, Winnie Chen

Date: Mar 19, 2024

Course: BAX 422 Data Design & Representation

Instructor: Jörn Boehnke

## 1. Executive Summary

This project endeavors to **facilitate the strategic expansion of a travel product e-commerce company into the United States market.** Utilizing a meticulously compiled dataset of 884 travel products from "[Get Your Guide](https://www.getyourguide.com/)," this report offers a deep dive into essential business queries related to **competitive positioning, product focus, geographical strategy, supplier partnerships, and pricing mechanisms**.

The analysis is centered around structured data encompassing various travel product categories, locations within the U.S., and comprehensive pricing and supplier information. By dissecting the competitive landscape, the report **identifies key competitors** and their operational platforms, enabling a strategic approach to market entry that minimizes direct confrontation. A focused examination of product categories reveals **the most popular listings**, guiding the prioritization of product development efforts. Geographical analysis pinpoints high-activity regions, suggesting **optimal locations for establishing a local corporate presence**. Furthermore, insights into activity providers unveil **potential supplier partnerships**, critical for diversifying and enhancing the product portfolio. The dataset's exploration also facilitates a nuanced understanding of **pricing strategies**, balancing competitiveness with profitability.

This project underlines the value of a cohesive data strategy, leveraging a relational database for its robust querying capabilities, which supports the company's informed decision-making processes. The strategic insights derived from this comprehensive data analysis are poised to significantly influence the company's approach to capturing the U.S. travel product market, setting a foundation for sustained growth and competitive advantage.

## 2. Background, Context, and Domain Knowledge

The travel and tourism industry is not only a significant contributor to the global economy but also a highly competitive and constantly evolving sector. Within this dynamic ecosystem, travel product e-commerce platforms have emerged as pivotal facilitators of travel experiences, connecting consumers with a wide array of travel-related products and services. These platforms cater to a diverse range of consumer needs, from booking guided tours and securing entry tickets to popular tourist destinations to arranging water activities and adventure experiences. **The U.S., with its vast and varied landscape, offers an abundant array of tourist attractions** ranging from natural wonders and outdoor activities to cultural and historical sites, making it an attractive market for expansion.

Entering the U.S. travel product market requires a thorough understanding of the business scenario, including the identification of target customer segments, the differentiation of products and services offered, and the formulation of a competitive strategy that leverages unique value propositions. The e-commerce platform must navigate a landscape marked **by established competitors, emerging trends in consumer preferences, and the challenges of logistical and operational integration across diverse geographical locations.**

The proposed expansion into the U.S. market by our travel product e-commerce company is driven by the ambition to **tap into this lucrative sector by offering a carefully curated selection of travel experiences**. Our product and service offerings will encompass a broad spectrum, including but not limited to guided tours, entry tickets to must-see attractions, water-based activities, and adventure experiences tailored to diverse consumer interests and preferences. The strategic focus will be on **identifying and developing the most popular listings first**, ensuring a strong market entry with offerings that resonate with target customer segments.

To ensure competitiveness and market relevance, our approach emphasizes the importance of understanding the competitive landscape, consumer behavior, and preference trends, as well as the operational intricacies of the travel and tourism industry in the U.S. **Success in this endeavor will hinge on our ability to source, aggregate, and leverage data effectively to inform strategic decisions, from product development to market positioning and pricing strategies.** The ultimate goal is to establish a strong foothold in the U.S. market, creating a distinctive and compelling value proposition that differentiates our platform from competitors, drives customer engagement, and fosters sustainable growth.

## 3. Data Sources, Web-Scraping Routines, and Databases

The primary source of data for this project is "[Get Your Guide](https://www.getyourguide.com/)," specifically the 58 pages of activities within the United States. Each activity page lists the following information:

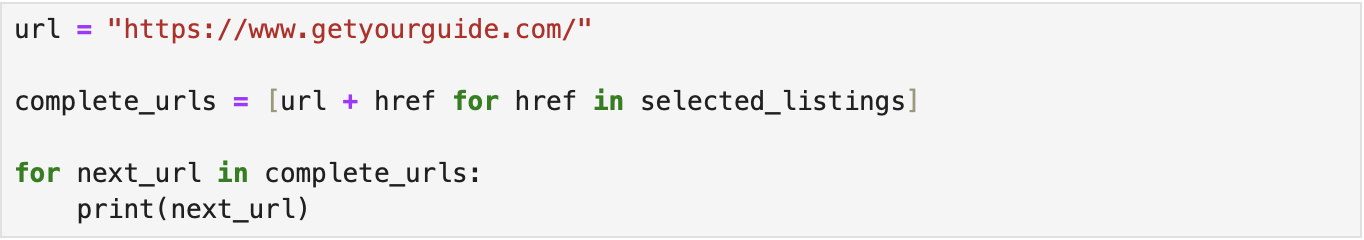
| **Variable** | **Description** |
| --- | --- |
| Country | Country |
| Location 1 | State |
| Location 2 | City |
| Location 3 | Specific location (NULL allowed) |
| Product Category | The category of the activity, such as “Water activity”, “Guided tour”, "Entry ticket", "Adventure” etc. |
| Product Name | Name or title |
| Review Score | Rating, ranging from 0 to 5, with one decimal. (NULL allowed) |
| Review Count | Count of reviews (NULL allowed) |
| Activity Provider | Provider of activity |
| Cancellation Policy | Cancellation policy indicating either a specific cancellation policy or free cancellation |
| Cancellation Policy Details | Additional details about the cancellation policy |
| Promotion Tag | Any promotional tags such as ‘New Activity’ or ‘ Likely to sell out’ (NULL allowed) |
| Price | Price |

As a part of the web scraping process, Selenium’s WebDriver was used to access the GetYourGuide website. The site’s design required users to click a “show more” button to reveal all of the 800+ activities available. By examining the HTML structure, we identified the CSS selector for the “show more” button as “div#load-more-page-{page}”. This allowed us to implement a loop to navigate through each page, effectively clicking all “show more” buttons using ChromeDriver and expanding the entire website.



After the additional listings loaded, we stored the HTML content of the entire website to our hard drive and used BeautifulSoup to search the content for the CSS selector for each individual activity block (“div.reloadable-block”). This method allowed us to locate and extract the “href” attribute of each individual listing from the activity blocks. By combining the “href” attribute with the original website structure, we were able to create a list of all the URLs of the individual listings. These URLs were then used to extract the relevant information about the activities from GetYourGuide.





We decided to store all the HTML files for each listing locally, ensuring easy access to the data in scenarios where the website structure changes or if there are issues with the web scraping process. However, during this process, we encountered an issue when attempting to save and open the HTML files using GET requests. Instead of the expected content, all of the saved HTML files displayed error pages. We believe that GetYourGuide has such measures in place to block web scraping efforts. To overcome this challenge, we used Selenium to open and save each HTML file individually instead. This approach was more time consuming, but it allowed us to successfully save and utilize the HTML files for extracting the relevant variables.

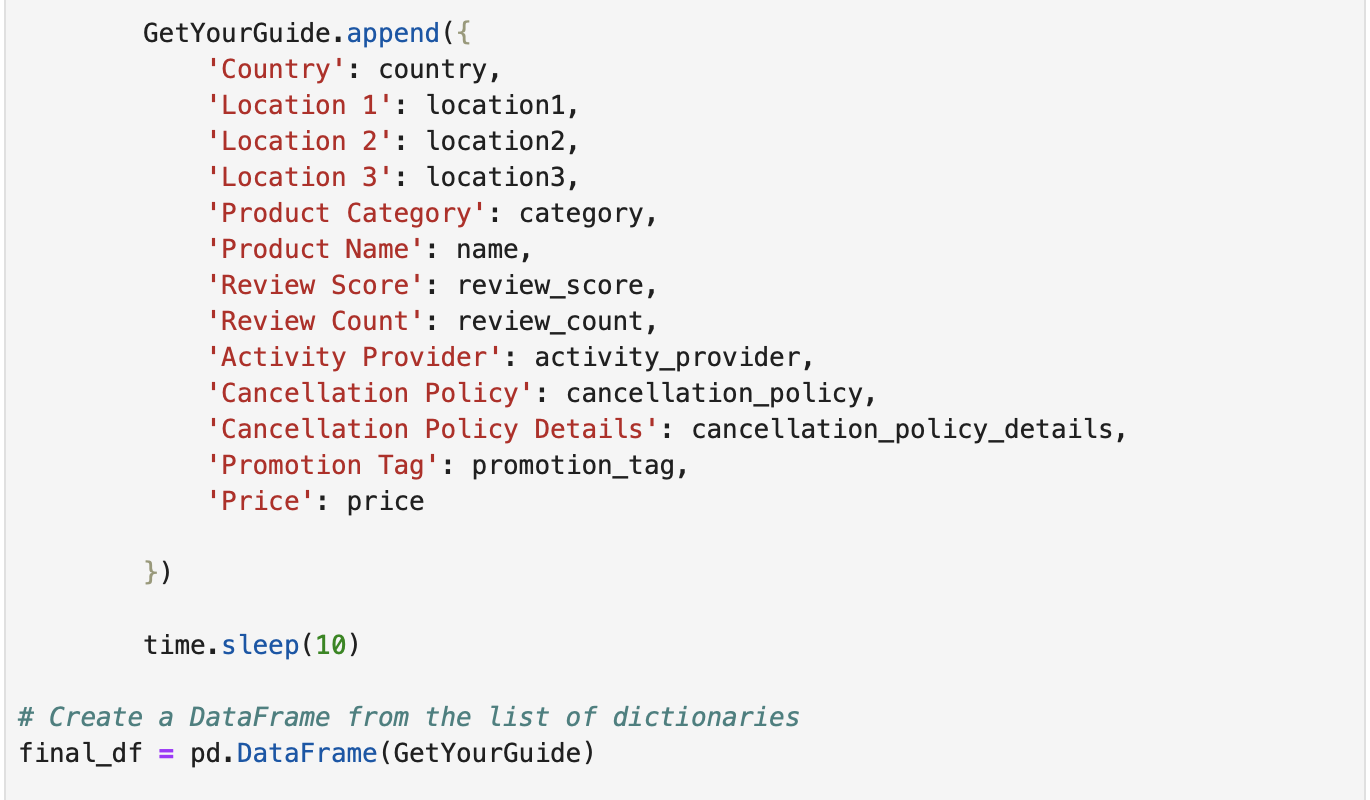


Once all of the HTML files were stored locally, we looped through each file and used BeautifulSoup to search for the CSS selectors regex patterns associated with the variables of interest. Some variables such as “Location 3”, “Review Score”, “Review Count” and “Promotion Tag” included NULL values. For example, if the activity was new, then it did not have a review score or count.

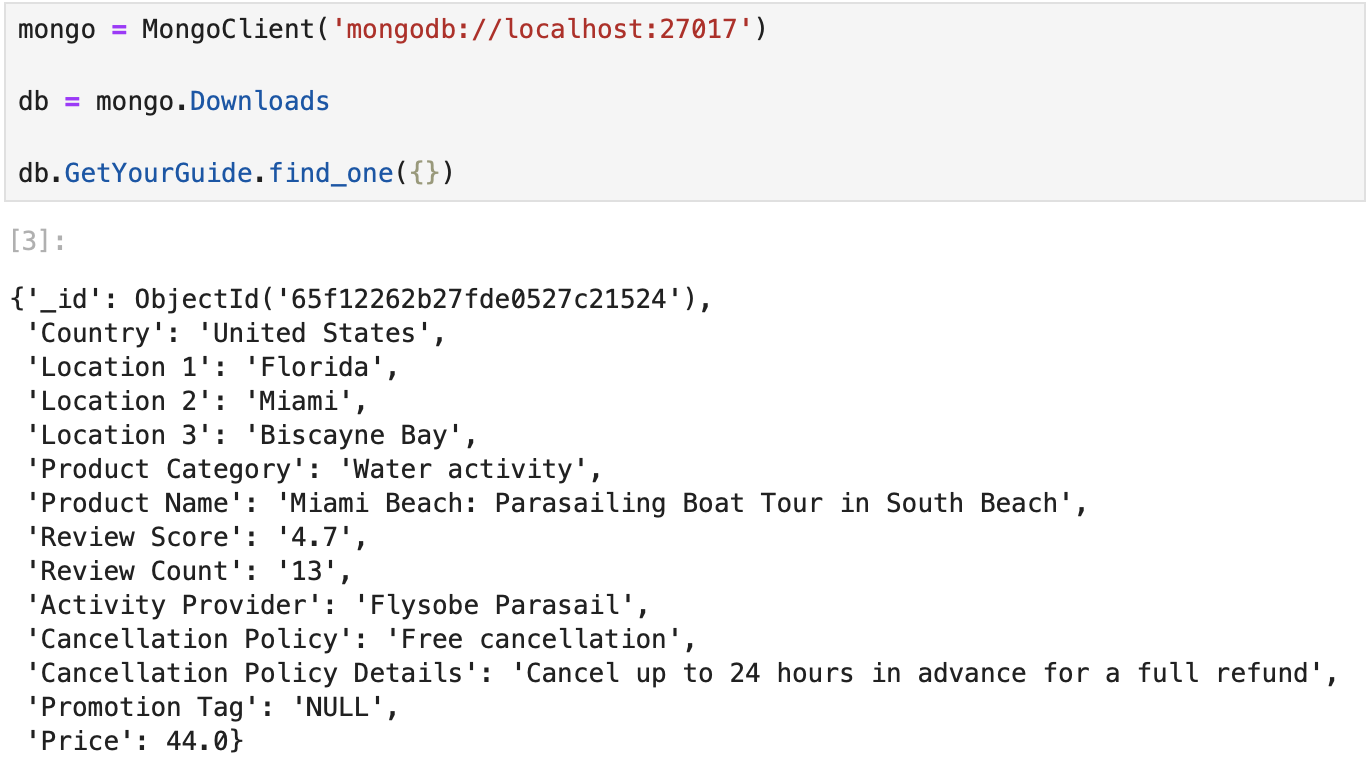




The extracted information was organized into a list of dictionaries, with each dictionary representing a single activity. Then, we used Pandas to create a data frame from the list of dictionaries.



Lastly, the data frame was exported as a CSV file and imported into MongoDB where further analysis can occur.



## 4. Discussion of the Strategic Use of Data

The dataset from "Get Your Guide" features insights from 884 diverse travel experiences used to guide strategic business decisions. Here’s how the dataset addresses our key business questions and the strategic advantages it offers:

* **Identifying Key Competitors and Their Platforms:**
  + The dataset's breadth across multiple product categories and locations allows us to map out the competitive landscape, identifying both direct and indirect competitors.
  + Advantage: By understanding where and how competitors are concentrated, we can identify market gaps or areas of less intense competition for strategic entry points.
* **Determining Popular Product Categories for Initial Development:**
  + Analysis of product categories and their associated review scores and counts reveals consumer preferences and popularity trends.
  + Advantage: This insight guides our product development strategy, ensuring we focus on offerings that are most likely to attract and satisfy customers.
* **Geographical Strategy for Product Listings and Local Office Placement:**
  + Detailed location data enables a deep dive into geographic demand patterns, highlighting regions with high activity concentrations.
  + Advantage: This geographical insight informs not only our local office placement for logistical and partnership advantages but also guides regional marketing strategies.
* **Strategic Supplier Partnerships:**
  + Information on activity providers uncovers potential suppliers for popular listings, offering a path to forge beneficial partnerships.
  + Advantage: Identifying key suppliers allows us to curate a diverse and appealing product portfolio, directly impacting our competitive edge and customer satisfaction.
* **Pricing Strategy and Market Positioning:**
  + The dataset provides crucial insights into pricing across different categories and locations, alongside the impact of promotional tags.
  + Advantage: Understanding market pricing dynamics enables us to set competitive prices that attract customers while ensuring profitability. The analysis of promotion tags offers additional strategies for enhancing product visibility and demand.

The strategic decision to manage this dataset within a relational database system presents numerous advantages, pivotal among them being the ability to perform complex queries and analyses efficiently. This capability is crucial for dynamically exploring the dataset to answer the above business questions, allowing for a level of flexibility and depth in analysis that would be challenging to achieve otherwise. The relational database's structured approach ensures that as our dataset grows and evolves, our ability to extract valuable insights and make informed decisions remains robust and scalable. This strategic use of data not only empowers our market entry strategy but also lays the groundwork for ongoing optimization and growth in the competitive U.S. travel market.

## 5. Summary and Conclusions

This comprehensive project, aimed at guiding a travel product e-commerce company's strategic entry into the U.S. market, leverages an extensive dataset of 884 products from "Get Your Guide" to address critical business questions. The analysis has provided invaluable insights into the competitive landscape, product popularity, geographical trends, supplier partnerships, and pricing strategies. These findings form the cornerstone of our go-to-market strategy, ensuring a data-driven approach to product development, market positioning, and operational planning.

Key conclusions from this project underline the importance of a strategic focus on popular product categories and the identification of high-demand geographical areas. These insights not only guide our initial product offerings but also inform the optimal placement of local operational bases to maximize market impact and logistical efficiency. Furthermore, the analysis of supplier data and pricing information has illuminated pathways to forge strategic partnerships and adopt competitive pricing strategies that promise both market attractiveness and profitability.

The decision to employ a relational database for managing and querying the dataset has proven instrumental, offering the flexibility and depth of analysis required for informed decision-making. As we move forward, these data-driven insights will continue to inform our strategic choices, ensuring that our market entry is not only successful but also positions us for sustainable growth and competitiveness in the U.S. travel product market. This project exemplifies the power of data in shaping strategic business decisions and highlights the potential for data-driven insights to drive success in the dynamic and competitive landscape of travel e-commerce.