Personalized Travel Itinerary Generation with Flight & Hotel Recommendations

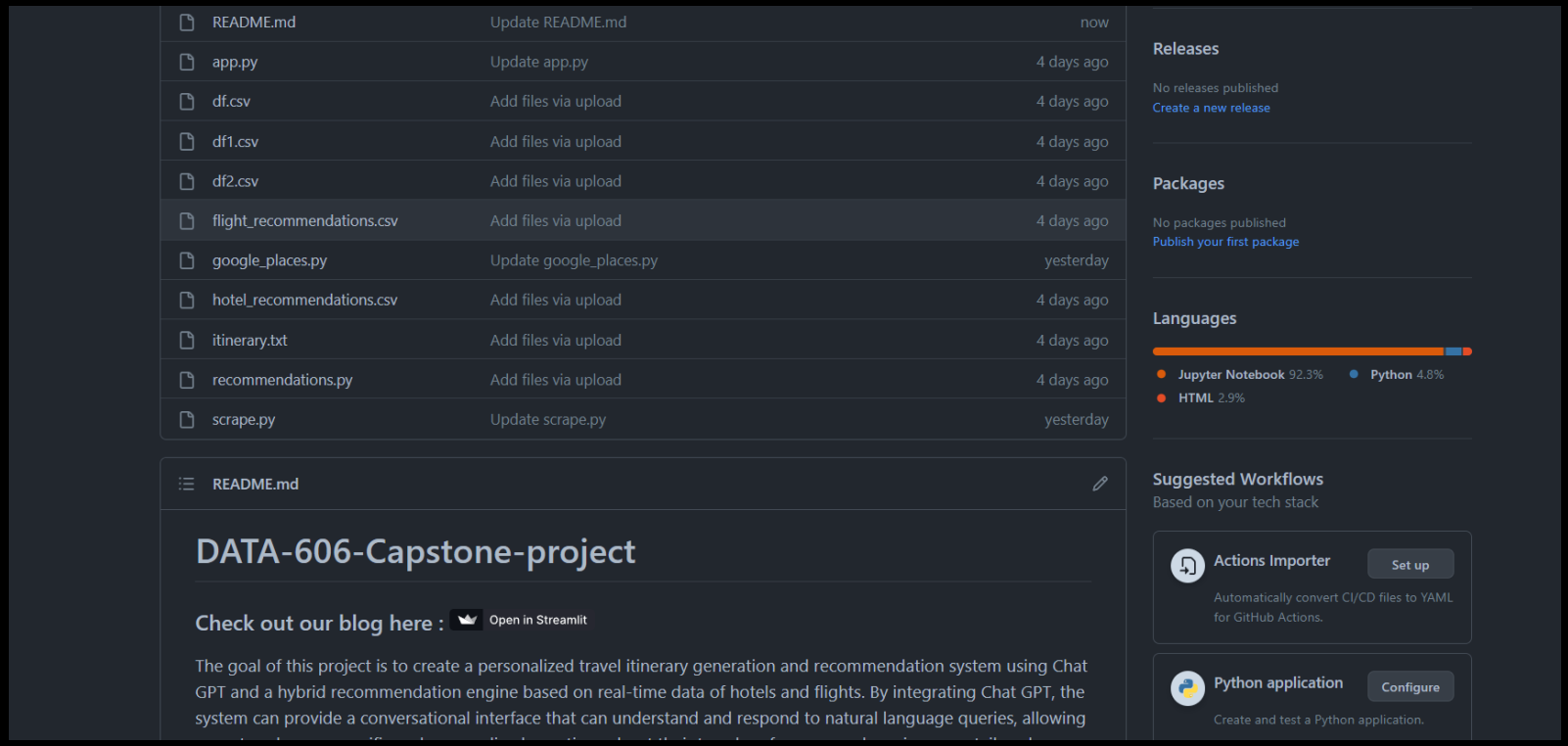
TEAM D – VIRTUAL TRAVEL ASSISTANT

- **NIKITA DHARMADHIKARI**
- **KOMAL LONDHE**
- **PARTH ZAVERI**

GITHUB REPOSITORY

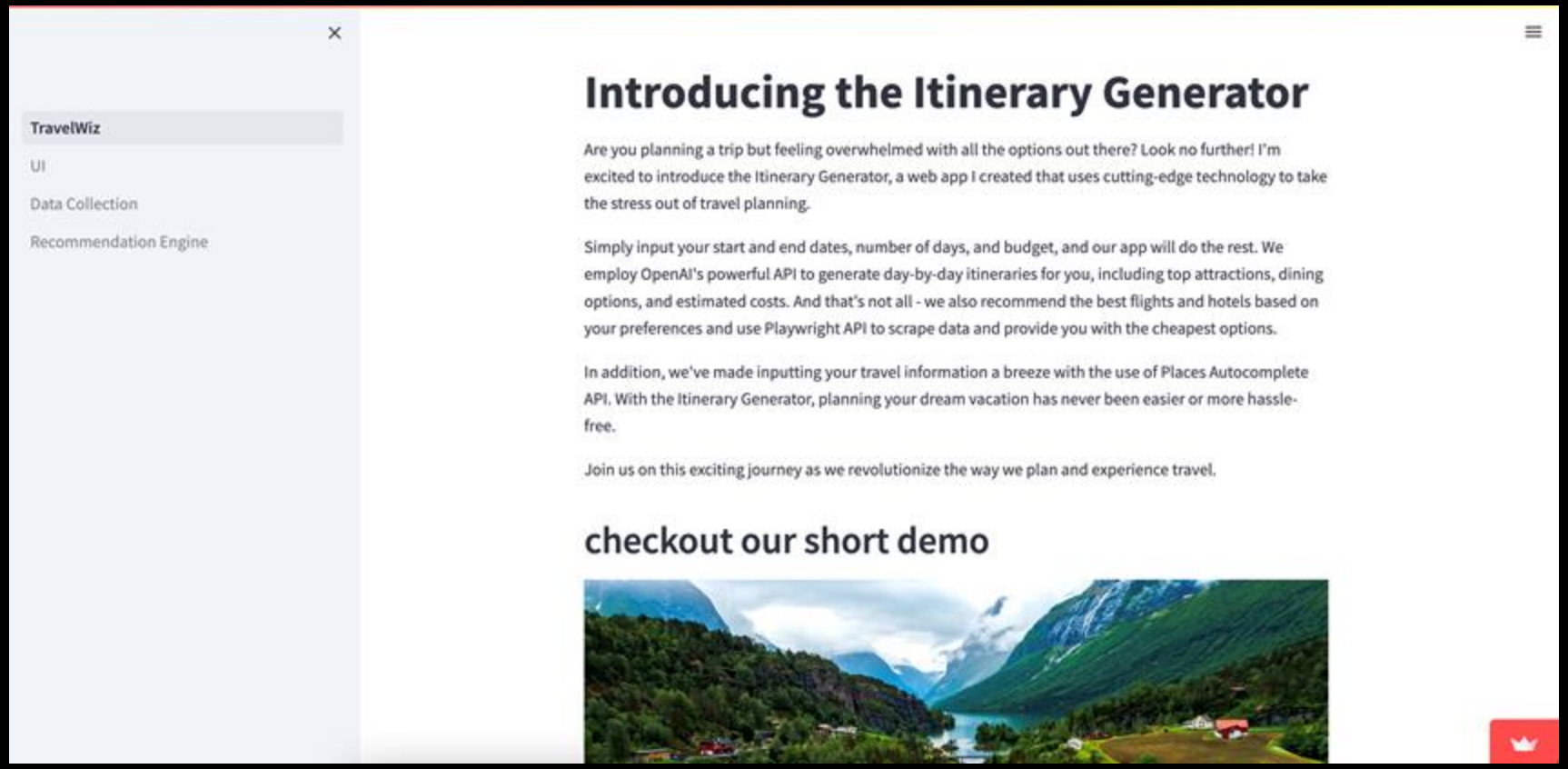
GITHUB LINK

https://github.com/epic-coder97/DATA-606-Capstone-project



SNAPSHOT – GITHUB PAGE

BLOG


TravelWiz
UI
Data Collection
Recommendation Engine
Introducing the Itinerary Generator
Are you planning a trip but feeling overwhelmed with all the options out there? Look no further! I'm excited to introduce the Itinerary Generator, a web app I created that uses cutting-edge technology to take the stress out of travel planning.
Simply input your start and end dates, number of days, and budget, and our app will do the rest. We employ OpenAI's powerful API to generate day-by-day itineraries for you, including top attractions, dining options, and estimated costs. And that's not all - we also recommend the best flights and hotels based on your preferences and use Playwright API to scrape data and provide you with the cheapest options.
In addition, we've made inputting your travel information a breeze with the use of Places Autocomplete API. With the Itinerary Generator, planning your dream vacation has never been easier or more hassle-free.
Join us on this exciting journey as we revolutionize the way we plan and experience travel.
checkout our short demo

Link

How a Frustrating Travel Planning Experience Led to an Innovative Solution for Personalized Travel Recommendations

- We've all been there: excitedly planning a trip, only to be overwhelmed by the endless options and information available online. It can be frustrating and time-consuming, leaving us feeling uncertain about our travel plans.
- That's exactly what happened to us when we were planning a trip to New York. We spent hours researching and comparing hotels, flights, and tourist attractions, only to end up with a mediocre itinerary that didn't meet our expectations.
- We knew there had to be a better way. That's when we had the idea to develop a personalized travel itinerary generator that would make travel planning easier and more enjoyable.

Product Description

Our project aims to develop a highly personalized travel itinerary generator product using ChatGPT.

The system provides recommendations for hotels, flights, and tourist attractions based on the user's budget.

Real-time data extraction using Playwright from the Travel (Google) website ensures that the recommendations are up-to-date.

The itinerary provided will have a per day budget to help users plan their trip more efficiently.

The generated itinerary includes links to recommended hotels and flights, along with the cost for the flights and hotels.

Users can directly access the links and easily book their flights and hotels according to their budget and preferences.

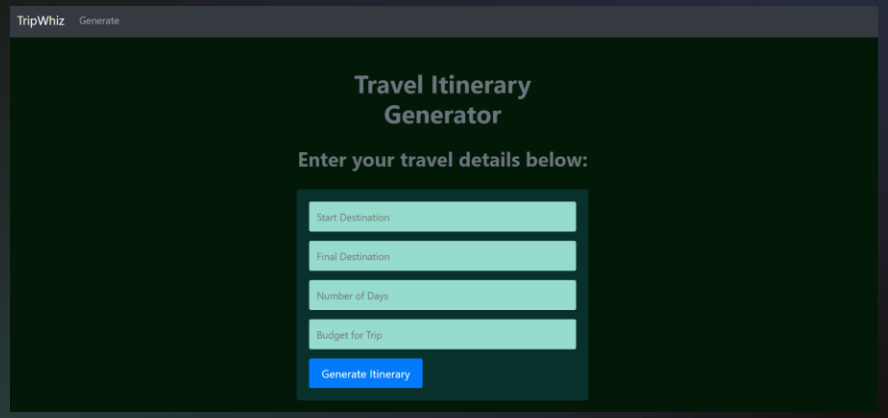
We're proud to offer a solution that makes travel planning more personalized and enjoyable. Our mission is to help travelers create unforgettable experiences, one itinerary at a time.

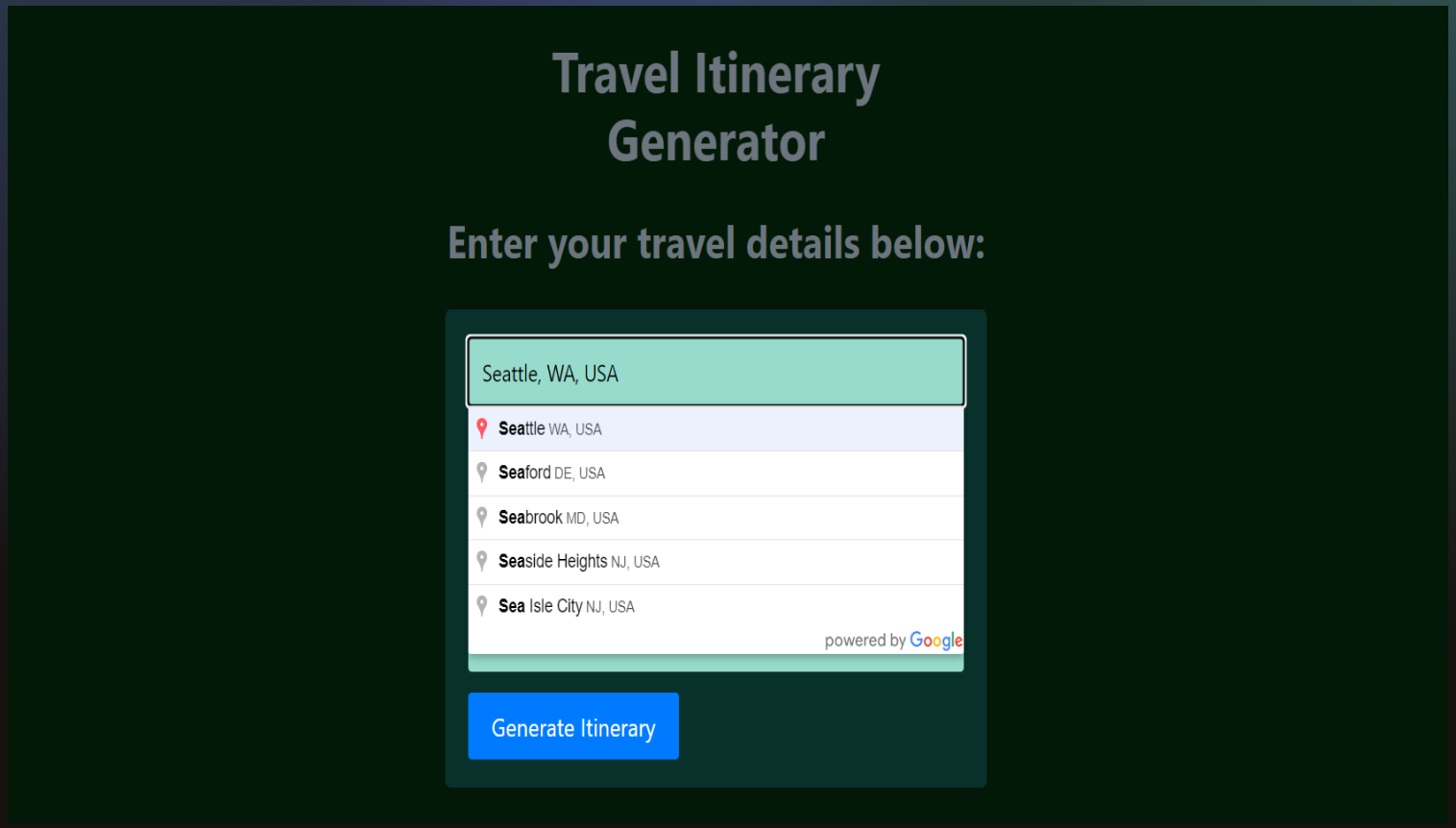
Implementation Approach

Step 1	Taking user inputs using Google Places Autocomplete API for Start and Final destination, number of days, budget.
Step 2	Using OpenAI API key, design a prompt for ChatGPT to give detailed customized itinerary based on user inputs.
Step 3	Collect real-time data for flights and hotels using Playwright from the travel website of Travel - Google. (https://www.google.com/travel/).
Step 4	Build a content-based recommendation engine for flights and hotels based on the attributes like flight and hotel prices, hotel ratings and flight duration column.
Step 5	Build a web app for the users where the user can get links to the recommended hotels, flights along with their cost.

Step 1: Taking user inputs using Google Places Autocomplete API for Start and Final destination, number of days, budget.

To begin creating a personalized travel itinerary, we need to gather some key information from the user. By collecting this information upfront, we can tailor our recommendations to their specific needs and preferences.





Tool used for taking user inputs – Google Places Autocomplete API

- Have you ever typed something into a search box and noticed that it helps you fill in what you're looking for? That's called autocomplete, and it can make it a lot easier and faster to find what you're looking for.
- Google Places Autocomplete API is a tool that helps people fill in information about places, like addresses or names of businesses, when they're filling out a form online. It uses a database of places that Google has collected and suggests options as you start typing.
- This can be helpful for a lot of reasons. It can save time because you don't have to type out the whole thing, and it can also help you make sure you're entering the right information. For example, if you're trying to enter your address to order food, the autocomplete might suggest your correct address before you even finish typing it all out.

Step 2: Using OpenAI API key, design a prompt for ChatGPT to give detailed customized itinerary based on user inputs.

To create detailed and customized recommendations for the user, we will utilize the powerful OpenAI API key. This will help ensure that our itinerary is tailored to the user's specific preferences and interests by giving the appropriate ChatGPT prompt.

```
model_engine = "text-davinci-003"
# prompt = f"Generate a detailed {num_days}-day itinerary from {start_dest} to {final_dest} and also places to vi
# prompt = f"Can you help me plan an itinerary for my upcoming trip? I'm starting in {start_dest} and I want to e
prompt = f"Act as an professional travel agent.Generate an itinerary for a trip starting at {start_dest} and expl
         f"The itinerary should include {num_days} " \
         f"days and cover the following details: recommended activities, places to eat, accommodation " \
         f"options ( suggest only one accommodation based on the budget {budget} for the entire trip )," \
         f"transportation options ( from the start destination to final Desination), " \
         f"and any other relevant information to ensure a memorable trip. " \
         f"Give the total estimate for the whole trip as well as for each day. " \
         f"Suggest only 1 accommodation for the whole trip which fits the budget {budget} at the start of the " \
         f"itinerary."
# f"Also, Display the output in JSON format."
response = openai.Completion.create(
    engine=model_engine,
    prompt=prompt,
    max_tokens=1024,
    n=1,
    stop=None,
    temperature=0.5,
```

Prompt Engineering

- The generative AI tools like ChatGPT put a lot of power in people's hands, but if everyone has access to the same technology
  - how can we use it while still expressing our individuality and creativity?
  - And how do we make sure our work stands out above that of other people and creates a competitive edge for our businesses?

  This is where prompt engineering skills come in.

❖ Prompt engineering is a comprehensive process that encompasses the entire cycle of interaction between humans and the AI.

❖ It involves deliberate and systematic design and refinement of prompts and underlying data structures to manipulate AI systems towards achieving specific and desired outputs.

TripWhiz Generate

About Us

We are a team of travel enthusiasts who love exploring new places and creating memorable experiences. Our mission is to help travelers plan their trips more efficiently and effectively with the help of our Travel Itinerary Generator tool powered by ChatGPT API.

With our tool, you can input your starting and final destinations, along with the number of days you plan to travel, and we'll generate a detailed itinerary that covers recommended activities, places to eat, accommodation options, transportation options, and any other relevant information to ensure a memorable trip.

Our team is committed to providing the best possible service to our users. We are constantly updating our tool to make it more user-friendly and efficient. We also value your feedback, so please feel free to reach out to us with any comments, questions, or suggestions.

Travel Itinerary

Day 1:

Total Estimate: $200

Accommodation: Staypineapple at Belltown ($100 for the night)

Transportation: Take a flight from Boston to Seattle ($100)

Activity: Explore the city of Seattle by taking a walking tour of the city.

Places to Eat: Check out the local favorites such as the Pike Place Market, Seattle's original farmers market, or grab a bite at one of the many restaurants in the area.

Day 2:

Total Estimate: $200

Accommodation: Staypineapple at Belltown ($100 for the night)

Transportation: Take a ferry from Seattle to Bainbridge Island ($10)

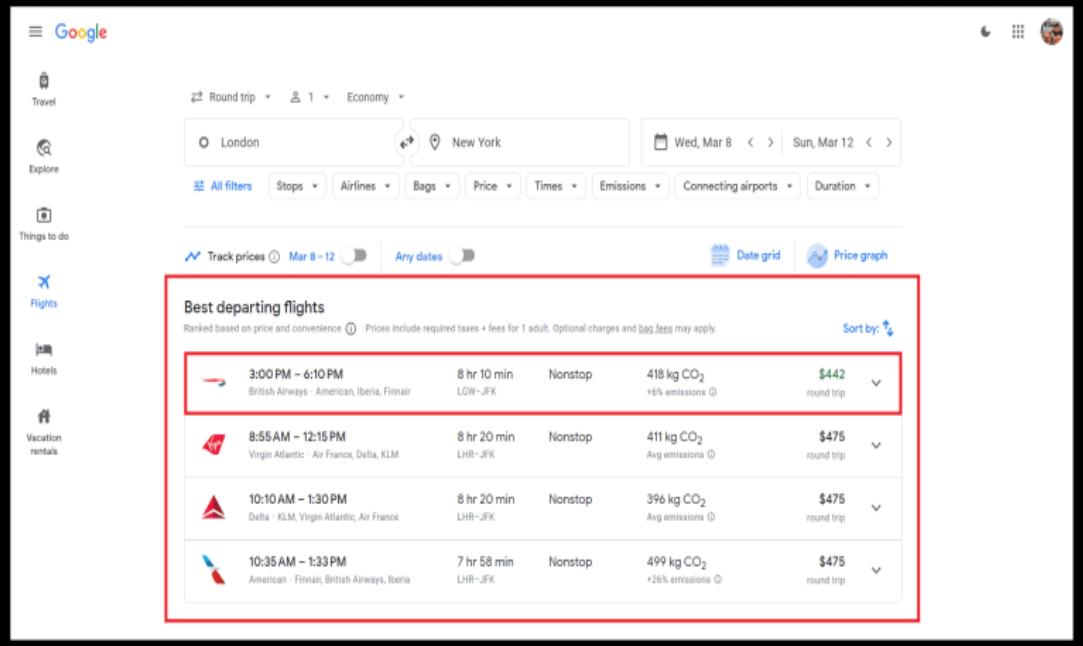
Activity: Explore the island by taking a tour of the island.

Places to Eat: Check out the local restaurants such as Bainbridge Island Brewing Company, or grab a bite at one of the many restaurants in the area.

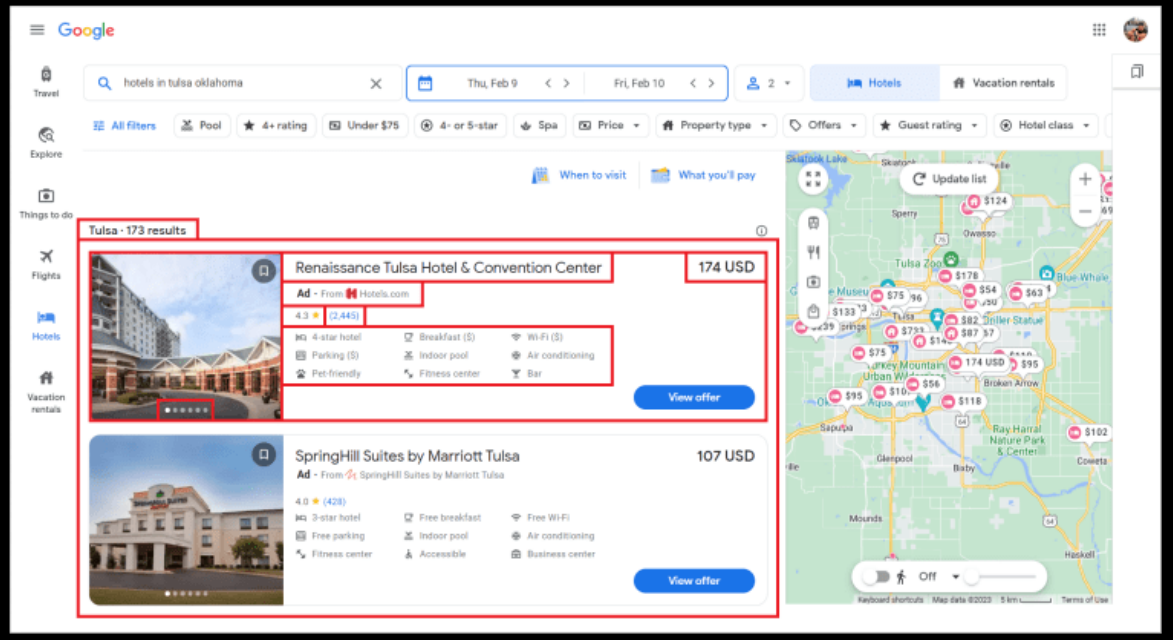
Step 3: Collect real-time data for flights and hotels using Playwright from the travel website of Travel - Google. (https://www.google.com/travel/).

By using real-time data, we can provide the most current and relevant recommendations for the user.

What flight data will be scraped



What hotel data will be scraped



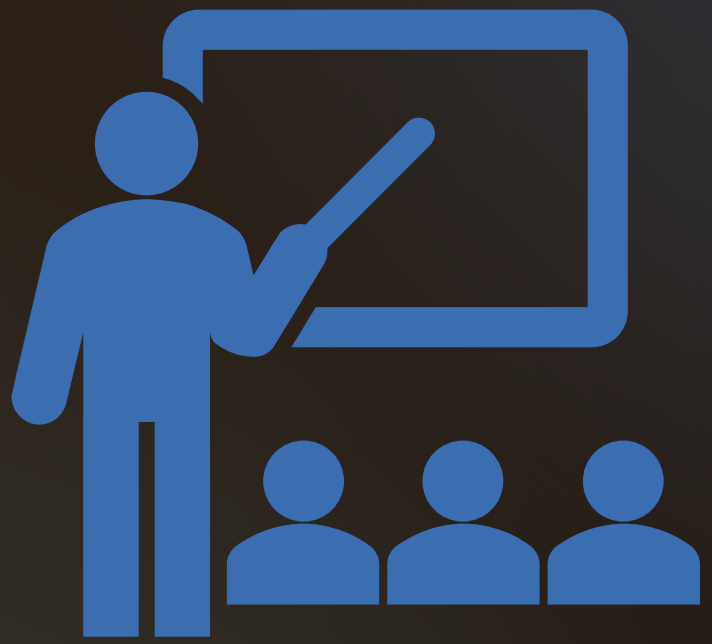
company	duration	stops	emissions	emission_comparison	price	price_type	flight_link
Frontier	12 hr 54 min	1 stop	202 kg CO2	+34% emissions	$298.00	round trip	https://www.google.com/trave
Spirit	8 hr 27 min	1 stop	185 kg CO2	+23% emissions	$367.00	round trip	https://www.google.com/trave
Spirit	6 hr 12 min	1 stop	184 kg CO2	+22% emissions	$373.00	round trip	https://www.google.com/trave
Spirit	11 hr 27 min	1 stop	165 kg CO2	+9% emissions	$377.00	round trip	https://www.google.com/trave
Spirit	4 hr 32 min	1 stop	144 kg CO2	Avg emissions	$414.00	round trip	https://www.google.com/trave
Spirit	8 hr 30 min	2 stops	218 kg CO2	+44% emissions	$436.00	round trip	https://www.google.com/trave
American	5 hr 40 min	1 stop	177 kg CO2	+17% emissions	$447.00	round trip	https://www.google.com/trave
American	5 hr 13 min	1 stop	177 kg CO2	+17% emissions	$447.00	round trip	https://www.google.com/trave
American	5 hr 2 min	1 stop	177 kg CO2	+17% emissions	$586.00	round trip	https://www.google.com/trave
Delta	7 hr 10 min	1 stop	204 kg CO2	+35% emissions	$606.00	round trip	https://www.google.com/trave
Delta	5 hr 35 min	1 stop	204 kg CO2	+35% emissions	$606.00	round trip	https://www.google.com/trave
DeltaOperated by Endeavor Air DBA Delta Connection	7 hr 30 min	1 stop	229 kg CO2	+52% emissions	$606.00	round trip	https://www.google.com/trave
Delta	5 hr 55 min	1 stop	204 kg CO2	+35% emissions	$612.00	round trip	https://www.google.com/trave
American	7 hr 7 min	1 stop	177 kg CO2	+17% emissions	$671.00	round trip	https://www.google.com/trave
United	8 hr 14 min	1 stop	261 kg CO2	+73% emissions	$685.00	round trip	https://www.google.com/trave
United	7 hr 47 min	1 stop	214 kg CO2	+42% emissions	$718.00	round trip	https://www.google.com/trave
UnitedOperated by Mesa Airlines DBA United Express	6 hr 32 min	1 stop	221 kg CO2	+46% emissions	$718.00	round trip	https://www.google.com/trave
American	8 hr 53 min	1 stop	177 kg CO2	+17% emissions	$735.00	round trip	https://www.google.com/trave
American	8 hr 23 min	2 stops	243 kg CO2	+61% emissions	$744.00	round trip	https://www.google.com/trave
American	7 hr 31 min	2 stops	243 kg CO2	+61% emissions	$771.00	round trip	https://www.google.com/trave
American	8 hr 5 min	2 stops	221 kg CO2	+46% emissions	$798.00	round trip	https://www.google.com/trave
American	10 hr 22 min	2 stops	231 kg CO2	+53% emissions	$814.00	round trip	https://www.google.com/trave
American	9 hr 22 min	2 stops	239 kg CO2	+58% emissions	$814.00	round trip	https://www.google.com/trave
American	12 hr 22 min	1 stop	177 kg CO2	+17% emissions	$821.00	round trip	https://www.google.com/trave
American	10 hr 26 min	1 stop	177 kg CO2	+17% emissions	$821.00	round trip	https://www.google.com/trave

Sample flight data

	title	link	price	rating	reviews	extensions
2	Park MGM Las Vegas	https://www.google.com/travel/search?ts=CAESC	$280.00	4.2	23570.0	['4-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
3	The LINQ Hotel + Experience	https://www.google.com/travel/search?ts=CAESC	$183.00	4.2	39016.0	['4-star hotel', 'Breakfast ($)', 'Wi-Fi ($)',
4	The STRAT Hotel, Casino & SkyPod	https://www.google.com/travel/search?ts=CAESC	$71.00	4.1	55400.0	['3-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
5	Hilton Vacation Club Cancun Resort Las Vegas	https://www.google.com/travel/search?ts=CAESC	$104.00	4.2	6916.0	['3-star hotel', 'Breakfast ($)', 'Free Wi-Fi
6	Luxor Hotel & Casino	https://www.google.com/travel/search?ts=CAESC	$370.00	4.2	85155.0	['4-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
7	MGM Grand	https://www.google.com/travel/search?ts=CAESC	$490.00	4.4	94404.0	['4-star hotel', 'Breakfast ($)', 'Wi-Fi ($)',
8	Las Vegas - 1 Week - Luxury Resort on Strip!	https://www.google.com/travel/search?ts=CAESC	$327.00			['Apartment', 'Sleeps 6', '2 bedrooms', '2
9	Rio All-Suite Hotel & Casino	https://www.google.com/travel/search?ts=CAESC	$82.00	3.9	37710.0	['4-star hotel', 'Breakfast ($)', 'Free Wi-Fi
10	Harrah's Las Vegas	https://www.google.com/travel/search?ts=CAESC	$120.00	4.1	26716.0	['3-star hotel', 'Breakfast ($)', 'Free Wi-Fi
11	Flamingo Las Vegas Hotel & Casino	https://www.google.com/travel/search?ts=CAESC	$142.00	4.1	59750.0	['3-star hotel', 'Breakfast ($)', 'Free Wi-Fi
12	Caesars Palace	https://www.google.com/travel/search?ts=CAESCgoCCAMI		4.5	114845.0	['4-star hotel', 'Breakfast ($)', 'Free Wi-Fi
13	Rio All-Suite Hotel & Casino	https://www.google.com/travel/search?ts=CAESC	$101.00	3.9	37708.0	['4-star hotel', 'Breakfast ($)', 'Free Wi-Fi
14	The Venetian Las Vegas	https://www.google.com/travel/search?ts=CAESC	$426.00	4.7	103098.0	['5-star hotel', 'Breakfast', 'Wi-Fi', 'Free
15	Excalibur Hotel & Casino	https://www.google.com/travel/search?ts=CAESC	$105.00	4.1	68763.0	['3-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
16	Four Queens Hotel & Casino	https://www.google.com/travel/search?ts=CAESC	$102.00	4.2	17122.0	['3-star hotel', 'Free Wi-Fi', 'Parking', 'Air
17	Beautiful High Rise Condo-307	https://www.google.com/travel/search?ts=CAESC	$207.00	3.9	16.0	['Sleeps 4', '1 bathroom', '615 sq ft', 'Air
18	SAHARA Las Vegas	https://www.google.com/travel/search?ts=CAESC	$90.00	4.1	22994.0	['4-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
19	ARIA Resort & Casino	https://www.google.com/travel/search?ts=CAESC	$452.00	4.5	37609.0	['5-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
20	The LINQ Hotel + Experience	https://www.google.com/travel/search?ts=CAESC	$179.00	4.2	39018.0	['4-star hotel', 'Breakfast ($)', 'Wi-Fi ($)',
21	The Cosmopolitan of Las Vegas	https://www.google.com/travel/search?ts=CAESC	$470.00	4.6	66663.0	['5-star hotel', 'Breakfast ($)', 'Wi-Fi ($)',
22	Mandalay Bay Beach	https://www.google.com/travel/search?ts=CAESC	$646.00	4.5	61867.0	['4-star hotel', 'Breakfast', 'Wi-Fi ($)', 'Pa
23	Excalibur Hotel & Casino	https://www.google.com/travel/search?ts=CAESC	$135.00	4.1	68761.0	['3-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
24	The Mirage	https://www.google.com/travel/search?ts=CAESC	$290.00	4.4	42845.0	['4-star hotel', 'Breakfast', 'Free Wi-Fi', 'I
25	Planet Hollywood Las Vegas Resort & Casino	https://www.google.com/travel/search?ts=CAESC	$179.00	4.3	54195.0	['4-star hotel', 'Breakfast ($)', 'Wi-Fi ($)',

Sample hotel data

Dataset



- Our system collects real-time data to provide up-to-date recommendations for hotels, flights, and tourist attractions based on user input.
- The size of the data collected varies depending on the user's input, ensuring that the recommendations are personalized and relevant.
- To gather data for flights, our script extracts all the flight data from Google Flights up to 200-300 rows of data.
- For hotels, we extract around 100-150 rows of data, equivalent to up to 5 pages where each page has 20 hotels.
- While we have a dynamic script to extract all hotel data, it took our extraction engine half an hour to run and extract all the data, so we decided to scrape sample data for efficiency.

Why Playwright ?

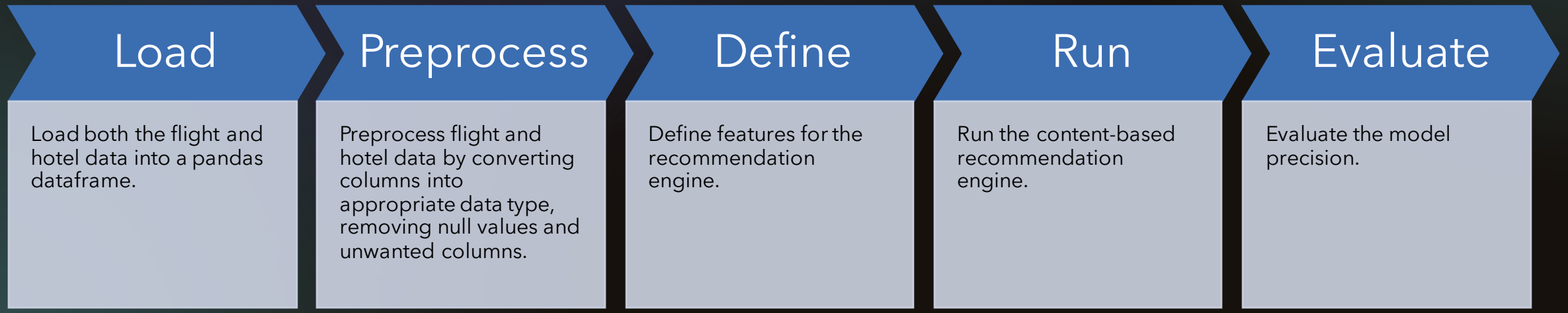
- **Multi-browser support**
- **Cross-platform support**
- **Modern APIs**
- **Robust automation capabilities**
- **High performance**
- **Headless mode**
- **Powerful debugging and tracing tools:**

Step 4: Build a content-based recommendation engine for flights and hotels based on the attributes like flight and hotel prices, hotel ratings and flight duration column.

Content – based recommendation engine

- A content-based recommendation engine is a type of recommendation system that uses the attributes or features of items to make recommendations. This type of recommendation engine analyzes the content of the items and recommends similar items with comparable features.

Model Implementation Approach



```python
# define a function to get the top n most similar flights to a given
def get_top_similar_recommendations(id, df, n=20):
    # create a TF-IDF vectorizer object to convert the features colum
    tfidf = TfidfVectorizer(stop_words='english')
    tfidf_matrix = tfidf.fit_transform(df['features'])

    # compute the cosine similarity matrix between all flights
    cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
    sim_scores = list(enumerate(cosine_sim[id]))
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
    top_similar_recommendations = [i[0] for i in sim_scores[1:n + 1]]
    return top_similar_recommendations


# define a function to recommend the cheapest flights based on a give
def cheapest_recommendations(id, df, attributes):
    top_similar_recommendations = get_top_similar_recommendations(id,
    sorted_similar_recommendations = df.iloc[top_similar_recommendati
    return sorted_similar_recommendations.head(20)


def recommendation_engine(df, features, attributes):
    df['features'] = df[features].apply(lambda x: ' '.join(x), axis=1

    recommendations = cheapest_recommendations(5, df, attributes)
    return recommendations
```

How the recommendation engine works ?

- The get_top_similar_recommendations function computes the cosine similarity matrix between all flights in the df dataframe. It uses a TfidfVectorizer object to convert the text-based features column of the dataframe into a matrix of feature vectors. Then it computes the cosine similarity between all flights based on this matrix and returns the top n most similar flights to the flight with the given id.

- The cheapest_recommendations function takes in the id of a flight, the df dataframe, and a list of attributes to sort the similar flights by. It first calls get_top_similar_recommendations to get the most similar flights to the given flight, and then sorts them by the specified attributes. It returns the top 20 cheapest flights.

- Finally, the recommendation_engine function takes in the df dataframe, a list of features to be used as text-based features, and a list of attributes to sort the flights by. It first converts the specified features into a matrix of feature vectors, and then calls the cheapest_recommendations function to get the top 20 cheapest flights that are most similar to flight 5. The function returns these flight recommendations as a dataframe.

```python
def evaluate_model(df, attributes):
    # split the data into train and test sets
    from sklearn.model_selection import train_test_split
    train_df, test_df = train_test_split(df, test_size=0.1, random_state=42
    # Define the number of recommendations to make
    k = 10
    # Initialize the precision array
    precisions = []
    # Loop over all test hotels
    for i, test_hotel in test_df.iterrows():
        # Get the top k recommendations for the test hotel
        recommended = cheapest_recommendations(i, df, attributes)[:k]
        # Get the set of recommended hotel IDs
        recommended_ids = set(recommended.index)
        # Get the set of actual hotel IDs (excluding the test hotel itself)
        actual_ids = set(train_df[train_df.index != i].index)
        # Calculate the true positives (i.e., recommended hotels that are a
        true_positives = len(recommended_ids.intersection(actual_ids))
        # Calculate the false positives (i.e., recommended hotels that are
        false_positives = len(recommended_ids.difference(actual_ids))
        # Calculate the precision
        precision = true_positives / (true_positives + false_positives)
        # Append the precision to the respective array
        precisions.append(precision)

    # Calculate the average precision across all test hotels
    avg_precision = sum(precisions) / len(precisions)
    return avg_precision
```

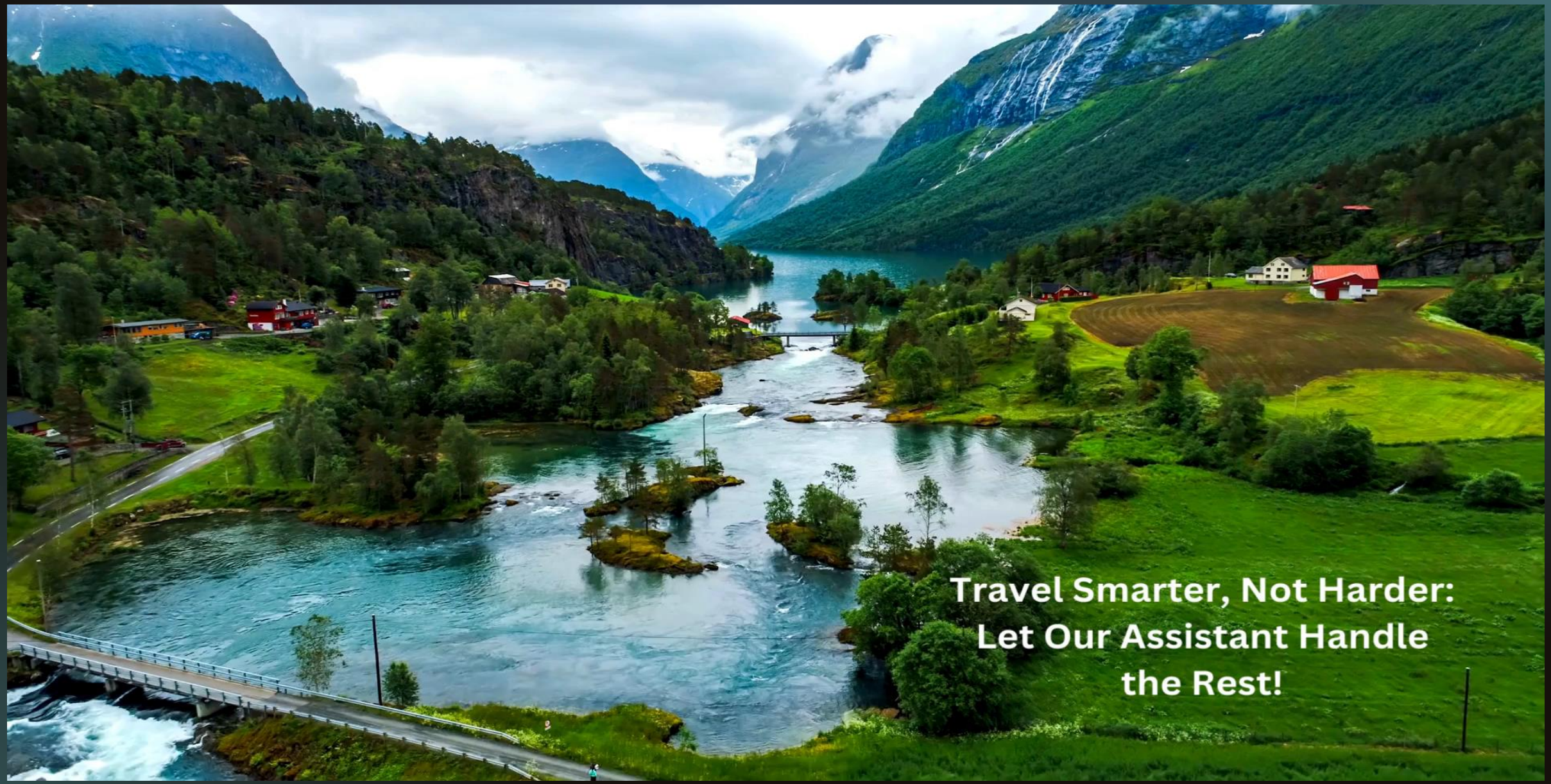
Model Evaluation

evaluate_model function evaluates the recommendation engine's performance by splitting the data into training and testing sets, getting the top k recommendations for each test hotel using cheapest_recommendations, calculating the precision for each test hotel, and returning the average precision as the evaluation metric. It uses true positives and false positives to calculate precision.

```
Model Precision for flights -
86.22222222222223
```

```
Model Precision for Hotels -
85.83333333333336
```

Step 5: **Build a web app for the users where the user can get links to the recommended hotels, flights along with their cost.**



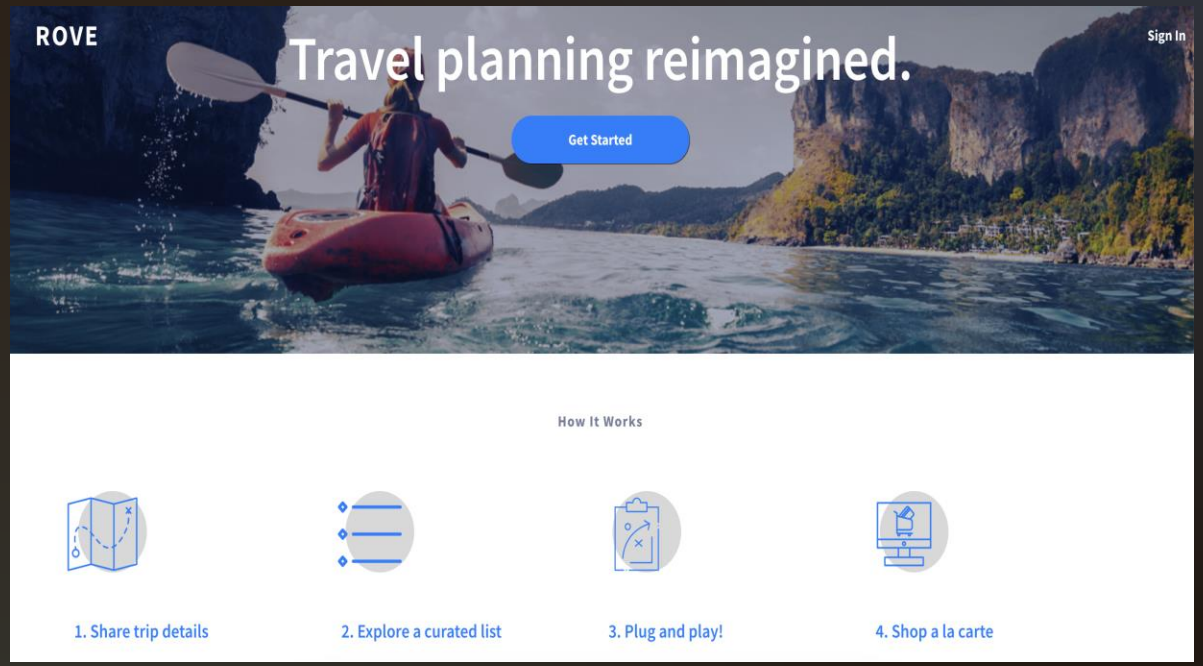
Literature Review

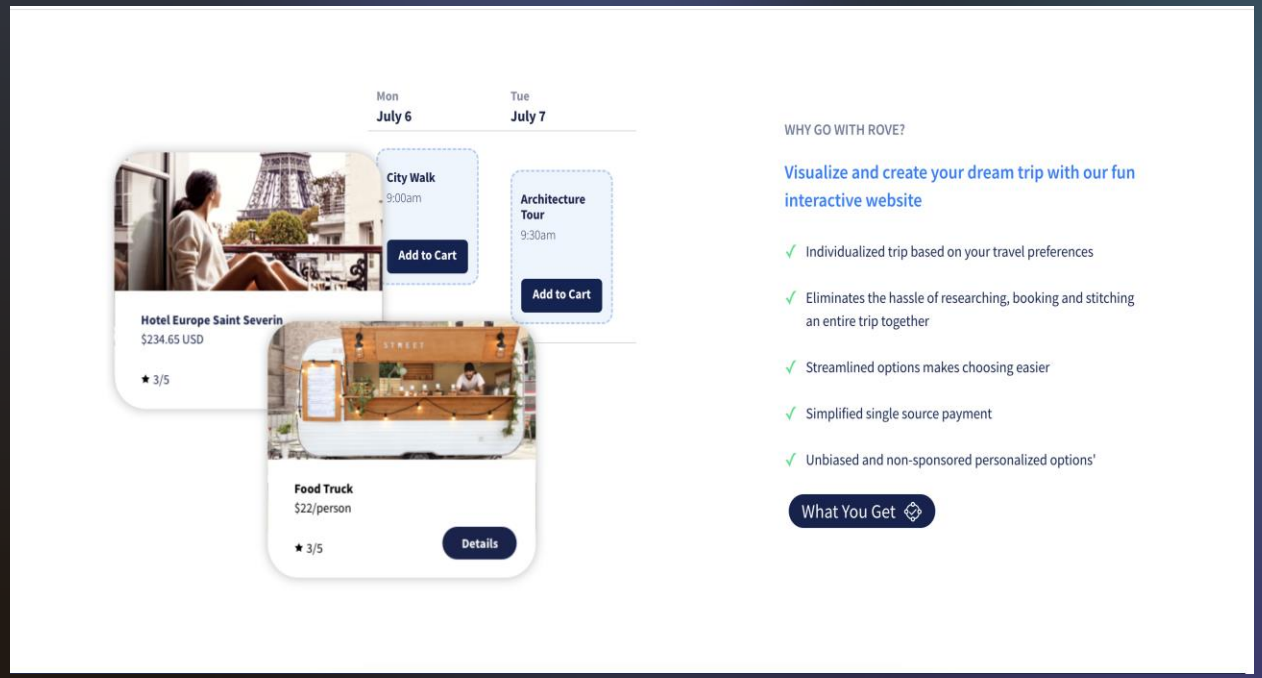
Zhai et al. (2020) present a personalized tourism recommendation system that uses data mining and collaborative filtering algorithms. The system provides personalized travel recommendations based on user preferences and behavior, such as past bookings and ratings. The paper highlights the importance of personalization in the tourism industry and the potential benefits of using data mining and collaborative filtering algorithms for recommendation systems.

Additionally, there are several other studies that explore the use of recommendation systems in the travel industry. For example, Chen et al. (2019) develop a travel recommendation system that uses machine learning algorithms and social media data to provide personalized travel recommendations. The system incorporates user preferences and behavior, as well as real-time data such as weather and event information.

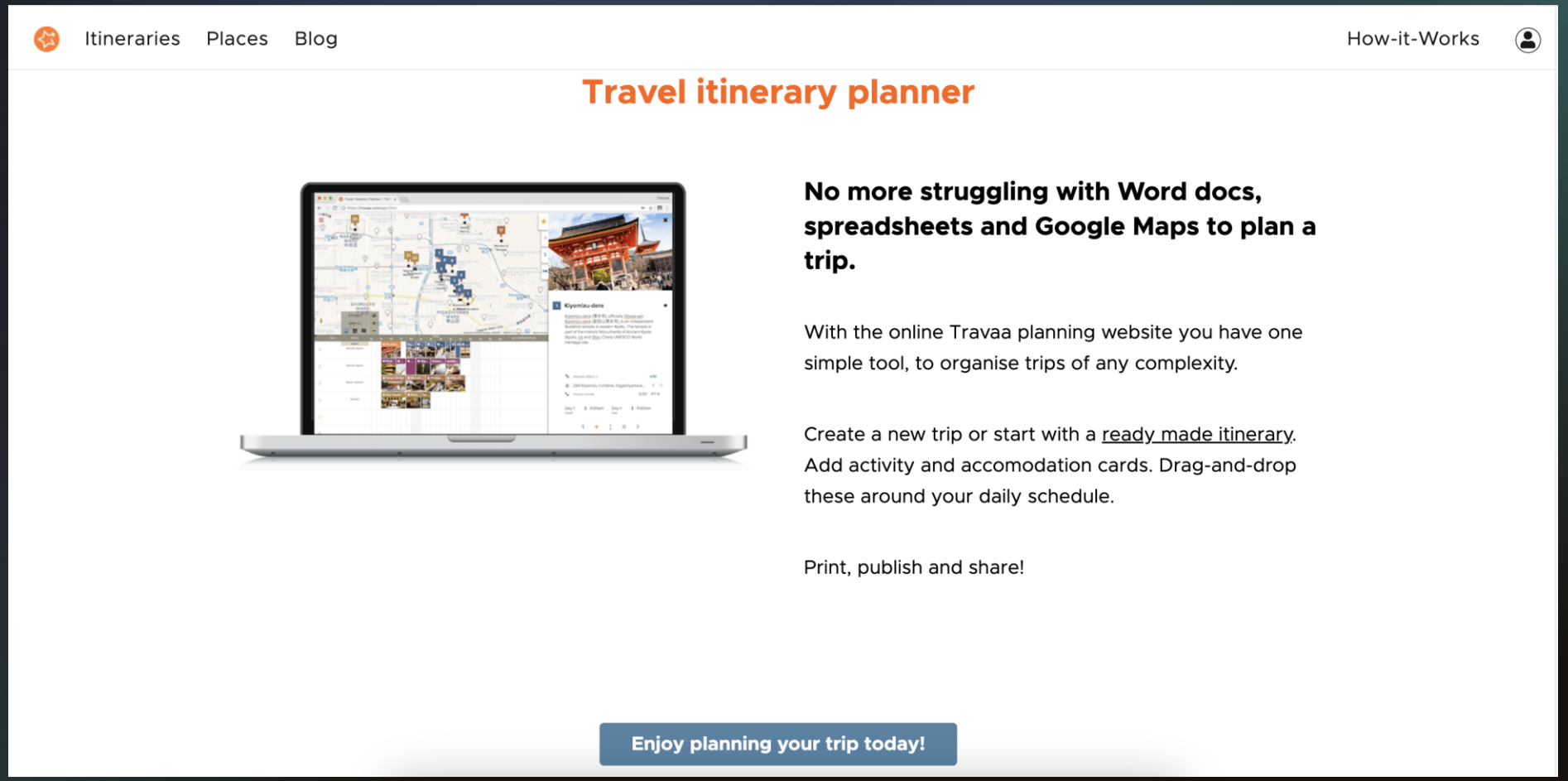
Similar Products

- https://www.travelwithrove.com/





➢ https://travaa.com/



Challenges

Prompt Engineering - Developing an appropriate prompts for ChatGPT.

Coding challenges – Getting airport name for the places mentioned in user input, collecting data using playwright with the locators.

Data extraction – Initially used Octoparse tool to get real time data (Integration was paid) but later developed own code for extraction using playwright.

Recommendation Engine – Initially recommendation engine was giving precision score 72 percent. After training on more data the score improved to 86 percent.

Future Work

1. Software Engineering :
   - The process of itinerary generation and data extraction can be run in parallel to optimize the app run time.
2. UI Features :
   - Pinpointing the places from the itinerary on google maps using the Google Maps API (Paid API).
   - Taking user information through sign-up forms.
   - Taking user reviews/ratings for the itinerary and recommendations.
   - Adding feature where user can save the itinerary and recommendations.
3. Data :
   - We have currently extracted data from. Google Travel site, but more data can be extracted from various travel sites like Expedia, Booking.com.
4. Booking System :
   - Implementing a feature on the website that enables users to schedule appointments or reserve services directly through the website without having to communicate with a human representative. This feature requires integrating a booking system, selecting a suitable provider, and customizing the booking system to meet the needs of the business and the preferences of the customers.

References

- Zhai, Y., Wei, X., & Song, J. (2020). Design and implementation of a personalized tourism recommendation system based on the data mining and collaborative filtering algorithm. Complexity, 2020, 1-13. https://www.researchgate.net/publication/363159690_Design_and_Implementation_of_a_Personalized_Tourism_Recommendation_System_Based_on_the_Data_Mining_and_Collaborative_Filtering_Algorithm
- Chen, M., Zhang, B., & Zhang, J. (2019). Personalized travel recommendation system based on social media data and machine learning algorithms. Journal of Ambient Intelligence and Humanized Computing, 10(4), 1367-1376. https://doi.org/10.1007/s12652-018-0881-1
- https://developers.google.com/apis-explorer
- https://openai.com/blog/chatgpt/
- https://www.travelwithrove.com/
- https://travaa.com/