DSCI 560 Lab 2

Part 1.
Team name: For Resume

Team details:
Saavani Vaidya 9385579920
Yuxuan Liu 4780355176
Shubham Darekar 1641138809

Part 2.
Saavani Vaidya
Financial Education Chatbot

- Regional Cost of Living Analysis:
  - https://www.kaggle.com/datasets/heidarmirhajisadati/regional-cost-of-living-analysis
  - This dataset provides insights into the cost of living and average monthly income across various countries and regions worldwide from 2000 to 2023.
  - Reasoning: This data can help the chatbot offer practical advice rooted in real world financial conditions. The chatbot can give users insights on how to adjust their finances based on regional expenses, such as rent, utilities, and daily needs, as well as help with budgeting.
- Credit Card Eligibility Data
  - https://www.kaggle.com/datasets/rohit265/credit-card-eligibility-data-determining-factors
  - This dataset provides a variety of attributes that can be used for analysis and modeling to understand the factors influencing credit card eligibility.
  - Reasoning: By understanding patterns from the dataset, the chatbot can offer suggestions to users on how to improve their creditworthiness, such as increasing income and paying off debts.

Course Textbook Chatbot

- Course to Textbook Mapping
  - https://www.kaggle.com/datasets/polartech/us-college-textbooks-and-courses-dataset
  - The datasets contain over 200,000 courses from 40+ American universities and maps them to their textbooks and information about them.
  - Reasoning: With the mapping and information on textbook content, chapters, and topics, the chatbot can assist students in quickly locating specific sections or topics within a textbook, identify overlap between courses that use the same textbook, and give better content summaries depending on the course.

Shubham Darekar
Medical Datasets:

- First Aid Help chatbot:

    - First Aid Intents Dataset, First Aid Recommendations Intents, Websites like:First Aid Instructions for 10 Medical Emergencies
    - First Aid Data is publicly available and can be helpful to chat with a bot in situations where quick help is required rather than skimming through the texts and available First Aid brochures
-

Itinerary planner for a Tourist Town:

- Websites like City of Sedona | Home and its outlinks as well as data from A Jam-Packed Sedona Itinerary! - My Perfect Itinerary
- As most of the tourist spots have documented important information, as well as there are websites which list down itineraries according to the availability of time, a chatbot can help in planning the day

Customer Service for Online stores:

- eCommerce Customer Service Satisfaction, Telecom Customers
- With the transcribed information from customer service calls, as well as the written communication, a chatbot can be trained and used to answer primary questions of the end user.

Yuxuan Liu

Programming Resources:

- Stack Overflow Data:
    - https://www.kaggle.com/datasets/stackoverflow/stackoverflow
    - This dataset includes real world programming questions, answers, and discussions.
    - Enables the chatbot to address common programming issues and provides detailed troubleshooting and debugging assistance.

FAQ Dataset:

- FAQ Dataset:
    - https://www.kaggle.com/datasets/umairnasir14/all-kaggle-questions-on-qoura-dataset
    - This dataset contains frequently asked questions across various domains, such as customer support, education, and finance. It includes question-answer pairs, which can be used to train the chatbot to handle routine inquiries.
    - Enables the chatbot to provide quick and accurate responses to common questions and improves user experience by automating repetitive task

Part 3.

We selected the travel itinerary planner idea from above to use in this part because we thought it would have more data available to collect for this assignment than the other topics.

Saavani

https://www.kaggle.com/datasets/fuarresvij/bali-popular-destination-for-tourist-2022
https://www.kaggle.com/datasets/vitaliymalcev/russian-touris-attractions
https://www.kaggle.com/datasets/faizadani/european-tour-destinations-dataset

Shubham
Tourist Chatbot for Hill Track Areas Bangladesh
Home - My Perfect Itinerary
Travel Dataset: Guide to India's Must See Places
Tour-itinerary.pdf
https://latourist.com/documents/LA_Tourist_Itinerary.pdf

Yuxuan
https://www.kaggle.com/datasets/vitaliymalcev/russian-touris-attractions
NZJYBG1.pdf
https://www.mytouragent.com/6D_PEK_-_XIA_PVT.pdf

Part 4.

Data Selection:

i. CSV or Excel

   https://www.kaggle.com/datasets/faizadani/european-tour-destinations-dataset

ii. ASCII Texts like Forum Postings and HTML

   Home - My Perfect Itinerary

iii. PDF and Word Documents that require conversion and OCR

   https://latourist.com/documents/LA_Tourist_Itinerary.pdf

We picked these out of the options in the previous part because they had the most diverse and descriptive data.

Chatbot Improvements:

1. Kayak Chatbot

One example of an existing tourism related chatbot is the Ask Kayak Chatbot. It helps users search for flights, hotels, and car rentals, provides pricing details and booking assistance, and can answer simple travel-related queries. Some limitations are that it has limited itinerary customization for activities beyond booking, it cannot plan multi-destination trips efficiently, and it is focused mainly on transactions instead of detailed travel planning. Our dataset focuses

more on the activity and detailed planning side of tourism which will improve the performance in these aspects.

2. Copilot2trip Chatbot

Copilot2trip, like many AI travel assistants, faces challenges in emotional intelligence, complex query handling, contextual understanding, and creative problem-solving. However, by enhancing our dataset, we can significantly improve its performance. Incorporating a diverse range of real-world travel planning conversations will help the AI better understand and respond to users' emotional needs and complex requests. Expanding the database with detailed, up-to-date information on various travel destinations will improve contextual understanding, allowing Copilot2trip to provide more nuanced and personalized recommendations. Including case studies of complex travel issues and their resolutions will enhance the AI's problem-solving capabilities, enabling it to offer more tailored and creative solutions. These improvements will help Copilot2trip evolve from a basic information provider to a more empathetic, adaptable, and innovative travel planning assistant, capable of handling the diverse and often unpredictable needs of modern travelers.

3. Wanderboat Chatbot

Wanderboat faces limitations such as the lack of accommodation recommendations and flight booking support, as well as a dependency on existing datasets that might be outdated or inaccurate, leading to discrepancies between suggestions and real-world conditions. In contrast, our dataset offers rich contextual data with detailed multi-turn conversations and annotations, enabling the chatbot to maintain context throughout complex interactions. Additionally, our dataset includes domain-specific knowledge, such as examples of hotel bookings, visa requirements, and budget constraints, allowing the chatbot to address these gaps and provide more accurate and comprehensive travel planning assistance.

Script:

In the report, describe what the script does (conversion tasks and tools to keep only the relevant data) to create a clean single dataset.