

Introduction:

We conceived the primary challenge offered by the task at hand as deducing what the key considerations that individuals looking to travel make. The immediately obvious ones are the desired temperature and setting, whether the desire be a beach vacation, to find mountains to hike, etc. Through our own discussions, though, we realized just how important the cuisine is when travelling, and found this to be a market that hadn't been capitalized on. Thus, we elected to make a travel bot tailored towards foodies, meaning individuals could input some vacation parameters they desire, and be given matching cities, alongside notable restaurants. This way potential tourists can explore the dining scene in a city as an equally important factor in their planning.

How we Chose our Datasets:

We scoured potential datasets available online, but unfortunately found that the offerings tended to vary significantly from what we were trying to find. The datasets we found often had a piece of what we're looking for, but – somewhat obviously – didn't offer the streamlined data about hundreds of cities across the world. We were able to find other chatbots that were tailor made for potential travellers, but unfortunately they haven't published the datasets they pull on publicly. Thus, we decided to construct our own dataset for the project consisting of a few different metrics: average temperature in Fahrenheit during each season (fall, winter, spring, summer) and the setting (mountainous, beach, tropical, urban, rural).

Challenging Aspects:

The abovementioned challenge of locating a suitable dataset was frustrating as it resulted in tedious work crafting our own dataset, but we were fortunate that the data we needed was readily available. Another challenge we faced that briefly discussed above was identifying where in the product-market fit there was a gap. The initial conception we had for our final data project was using an API that either gave relevant information about hotel or flight pricing, but we ultimately realized that there were already many other services that provided this information. It was our goal to create something singular that would at least establish a framework upon which more features could be built. Thus, we decided to target a market that isn't being served: foodies. Therefore, most of the challenges we encountered while creating Botty were conceptual or in the data collection phase.

Key Learnings and Discoveries:

This project taught us a few things that I didn't know before. First, connecting an API and CSV file into one output was more challenging than expected. We tried a few different methods, but it's crucial that data points matched up (such as our output from the CSV files being the exact city names that are found in the API). It was also interesting for us to learn about how we best

coordinate this type of work as a team. There's an added component of communicating and exchanging information that is required when a project like this is broken up amongst three people, and the longest parts of our meetings were often just sharing the work we've done and figuring out how we can combine each part of our work together. For example, I was working on finding a suitable API that locates restaurants in different cities. After finding and creating the code for the API call, I had to then work with my other team members on explaining the code, and then implementing the code with our larger bot. Ensuring that our syntax matched up and that the combination of our code didn't interfere with how the individual codes worked was challenging, and led us to optimizing how we interacted with each other: we explained our code, highlighted key syntax, and found where we could best connect our code together (while also removing extraneous parts of it).

Potential Future Enhancements:

If we had more time to continue improving our travel chatbot, there are definitely aspects we could have added which would have enhanced the robustness and utility of our bot. As it stands now, Botty is able to provide users with some recommendations on where they might enjoy traveling based on some of their preferences, but given the dataset it has been fed, a user could quickly be limited by its capabilities. Providing it with more parameters to draw from, as well as giving it the ability to intuit, as opposed to simply using boolean search, would facilitate exponential improvement to Botty.

Something that would have been interesting to play with is adapting it to adjust its recommendations based on feedback on its first attempt. Right now, it only takes the one user input and gives recommendations, but if the user wants to ask again or adjust their parameters, they have to go back to the start. It would be more useful as an actual travel planning tool if it could remember what a user has told it and then continue to make more and more precise recommendations. This could be bolstered by adding a like/dislike feature to the stream of information, which the bot would be able to factor into future recommendations. On a related note, building out a feature that enables users to save interesting destinations would be an excellent addition to the offerings. This would allow users to come back at a later time and review cities they were previously interested in, as opposed to having to save it on some other platform.

Challenges:

A couple things that presented challenges for our team during this project were deciding what topic to focus our energy on at the beginning of the project and creating a streamlined outline for how we would use our time efficiently to get to the desired end product. Since this project was very open-ended, it left us with a lot of choices for the data we could explore; as long as we were able to find a relevant API and another accessible data source on a topic, it was an option. This kind of caused us to spend a little bit too much time exploring many different paths instead of narrowing it down to one topic early so we could start working on the actual ETL data processing. This challenge went hand in hand with creating a plan for how the project would ideally proceed. It is almost always easier to work on something with an end goal in mind, and we could have done a slightly better job of that. While these did not prove to be fatal errors, it would be a good idea to focus on using time efficiently for future projects.

Aspects That Were Easier Than Expected:

This project presented several aspects that were easier than anticipated. Combining the data into a single table and dataframe was more straightforward than expected, and integrating data from an API was already a relatively simple process. Initially, we anticipated challenges in merging data from two separate sources due to differences in column names and data points. However, renaming columns proved to be a simple task, and once we established a standardized naming convention, we were able to merge the tables with just a few lines of code. Additionally, prior to this project, we had never retrieved data from an API, and the process initially seemed daunting. However, fetching the data without errors turned out to be relatively easy, and the API data was integrated into our project seamlessly.

Aspects More Difficult than Expected:

In terms of some concrete elements of the project that proved more difficult than expected, we encountered a couple including: switching between the desired file formats (JSON, CSV, and SQL database) and making sure our API language and commands were correct for our API./

How a Utility Like This Could be Useful in Future Data Projects

The Extract, Transform, Load (ETL) process can be easily extrapolated to a wide variety of data projects because of how inherently necessary it is to derive actionable insights from raw data. In professional environments, particularly those that don't have robust data pipelines, it's very easy to imagine that significant swaths of data are never being analyzed, and thus opportunities to improve the quality of operations aren't being exploited. For instance, if in the future we're working in an industrial environment, there are certainly sectors of the business that would benefit from some of the principles we touched upon in this project. A notional example would be improving efficiencies in the fabrication process by predicting the necessitated lead-times during production. These simplistic applications of the data constantly being produced could have significant positive impacts on the synergies of virtually any business.