# Data Science Capstone Project Proposal: TripAdvisor

Trevor McCormick

Tmccormick2@bellarmine.edu

January 12, 2023

# Data Science Capstone Project Proposal: TripAdvisor

## Executive Summary:

In this project, I am going to be using TripAdvisor API data in order to tell people when the best time of year to visit a location is. This will be done by moving the data through many applications for cleaning, storage, visualization, analysis, and modeling. All of these are necessary for the end goal of making a useful and working program. Most of the project will be done in Python using the Pandas package. The storage method will be MySQL due to that being what I am semi-familiar with. Then, the visualization will be done in Tableau because I am both familiar with and enjoy using it. Finally, everything will go back into Python for the analysis and modelling. While these are my ideas as of writing this, I am aware and sure that many things are likely to change through the process of trial and error.

## Project Idea:

The topic that I have decided to pursue for the data science capstone project is a dive into the API from the travel organization TripAdvisor. This project will have many steps that will all cumulate to achieve a final goal at the end. The TripAdvisor API gives users (like myself) access to numerous aspects of information from their website. Some of which include locations, reviews, ratings, images, precise latitude/longitude of destinations, and URLs for where to find these listings. As you can see, all of these options provide so many potential ideas and goals that could be used for a data science project. One of the hardest decisions is deciding what aspect to work on. After considering all of the possible options that I could pursue for my project, I had to pick the one that seemed both the most interesting to me and actually doable for a class project. The end goal for this project as of right now (it may change if the need arises) is to use the TripAdvisor API data to predict when the best time to visit a location is. Ideally, you would be able to look at a certain location and the program will tell you when the best time of year to visit is. The ways the program could decide this would be by looking at crowd levels, reviews, and activities for a given location. These are my ideas for the project at this time.

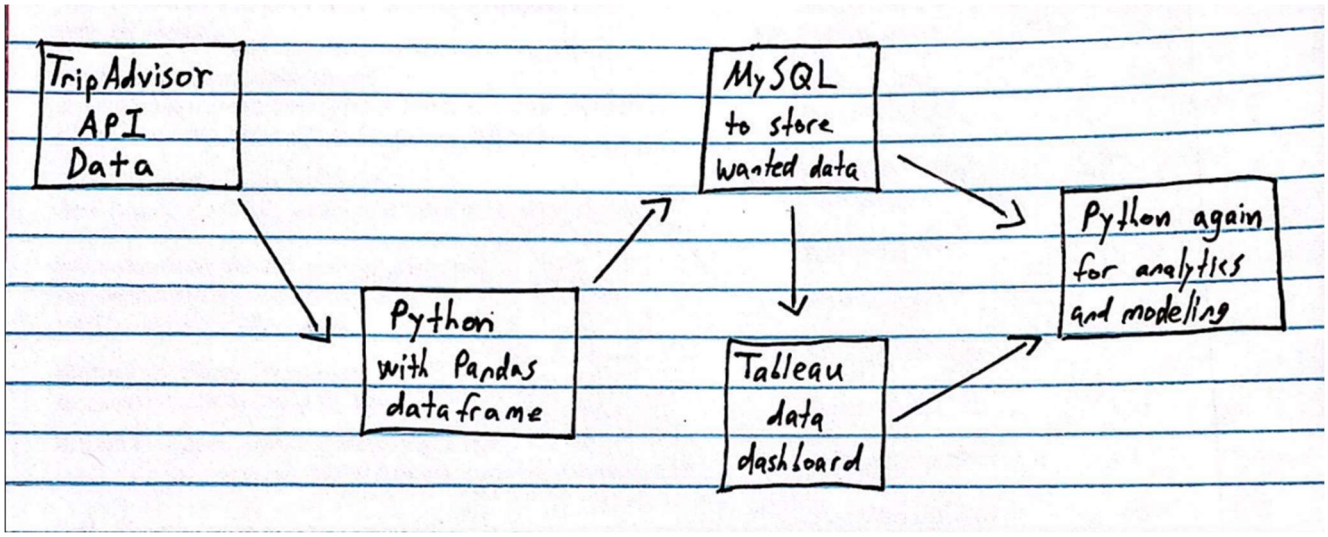
**Background:**

My main drive for picking TripAdvisor as my project area is due to always being interested in travel and exploring/finding new places. So, I figured a project that enables me to reflect on and build off of that would be a perfect fit. A project should always be focused on something that you are passionate about or at the very least, interested in. My data for this project is going to come from the TripAdvisor API that is available for free (in limited quantities) or pay in certain larger projects. The data needed for this project should fall under the free section of their API use. Due to TripAdvisor being such a popular and large website today, it has constant data flowing through it. This means that their information is constantly updating to include new locations, reviews, etc. This makes it a perfect choice for this project. New data means more current and accurate information that we can work with when using the TripAdvisor API. TripAdvisor has this description for their current API system, "Increase your user experience, engagement, and conversion with TripAdvisor's globally recognized and highly trusted content. Our partner API provides you with dynamic access to TripAdvisor content, enabling seamless integration with your website and applications. Locations are defined within this API as hotels, restaurants or attractions." (Tripadvisor.com)

**Preliminary Architecture:**

The architecture of a data science project like this includes the data, how to collect the data, where to store the data, cleaning the data, and finally, using the data. As stated many times above, the data for this project will come from the TripAdvisor API that is available for use in projects like this. As of right now, I am planning to store the data in MySQL. SQL is often used as a data store and it is one that I am at least somewhat familiar with. It should do the job for what I need it to accomplish. I will use Python (or maybe R) for the data cleaning and analysis. These are the two programs that I am the most familiar with as of now. Python and R are both excellent recourses for what I need to do with this project. For the data dashboard I am planning on using Tableau since it is something that I am familiar with due to using it in a course my junior year. It seems like it would be a good fit for what I need, bit if it doesn't end up working properly there are other options out there such as Power BI. Considering these things, Python and Tableau/Power BI will most likely be the most important programs that I will be using during

this project. However, every application is necessary to reach the final goal of completion. All of this can be seen in the graph below:




**Modeling:**

The modeling section of this project is the main area where I am not quite sure which idea to settle on. There are so many options for what to do here and so many to choose from. I do know that with what my project goal is, I am going to want to use predictive analytics. Predictive analytics uses the data to learn and make predictions based on past data. I think the main methods that I'll be using are Scikit-learn and decision trees. However, these may need to chance once I get further in the project and see how thing work and inevitably don't work.

**Conclusion:**

In conclusion, I will be using the data available from the TripAdvisor API to conduct and data science project which will let people see when the optimal time of year to visit a location is. The data which comes from the API will then be moved into Python for cleaning. Next it will go to MySQL for storage. Then to Tableau as a dashboard and back to Python for analysis and modeling. I am excited to begin working on this project and seeing how close thing go to plan. Even though things with most likely need to change.

**References**

"Overview." *Tripadvisor Content API*, https://tripadvisor-content-api.readme.io/reference/overview.