

# **Springboard: Data Science Career Track Program**

## **Project 2 - Campsite Availability**

### **Proposal By Priyanka Panhalkar**

### **[October], [2024]**

#### **Business Problem**

This project is about building a machine learning model to predict the number of people expected to visit any given campsite, associated with specific campsite authorities, in a given time period.

As an example of Business need

1. When the user is logging in the website to book a campsite, they can view the total number of spots available. If they still need some time to finalize their plan, and are unsure whether it is the right time to book now or later, our model will help them provide a good estimate of a better timeline of how long they should wait to secure their desired spot.

#### **Intended Stakeholders**

The intended stakeholders are any campsite authorities in the US and other countries. Each campsite authority oversees a set of campsites in the respective country. This model will help authorities to have a better forecast about incoming crowds and can better plan for infrastructure. In upcoming versions, this model will also help campers to find an accurate timeline for their intended booking period.

#### **Data sources**

Dataset is available from - <https://ridb.recreation.gov/download>

Recreation gov has a dataset available for the last 18 years. It's a huge data which needs to be sampled down for analysis. We will be focusing on the post pandemic data set for now.

#### **Data Science Approach**

We will be using the Supervised Regression Methodology where we will train models on labeled data, meaning that both the input features, their corresponding values, and target are provided. As an example in this case we will be using detailed data for each campsite from the past 2 years about when the tent sites were booked, for which days and for how many people. These models will help predict appropriate booking timelines for a given period of time.

**Features**-FacilityState, Startdate, Enddate, Orderdate, RegionCode, Region Description etc.

**Model Architecture** - We will explore this problem using regression models with features that will be engineered to consider the temporal nature of the dataset.

**Evaluation** - We will use the following: (a) Mean Absolute Percent Error, (b) Actual Vs Predicted scatterplots, and (c) distribution of residuals. Models will be compared with respect to these performance indicators. We will also consider using explainability

approaches such as: (a) analysis of feature impact; (b) feature importance; and (c) SHAP<sup>1</sup>

We might consider using an Unsupervised approach as well.

### **Anticipate how the the Model will be helpful to clients**

By using this Model campers will have a better time estimate of when they should target to book their suitable campsite if they want to have it available.

1. The camp authorities can use this approach to understand the incoming influx of people and will help to plan the infrastructure like bathrooms, activities etc.
2. This will also open up opportunities for several businesses around the corners (example vendors selling wood logs, charcoal, ice, food etc) which can add advantage to the campers.

### **Deliverables**

As required, as part of this project, the following will be delivered: (a) all Jupyter notebooks in my Github repository; (b) written final report; and (c ) a presentation slide deck.

---

<sup>1</sup> <https://shap.readthedocs.io/en/latest/>