Annual Spending at National Parks Based on Weather

Hayden Howard

Predicting the annual spending at various national parks across the United States based on the weather in the area.

Introduction

There is a lot of things that go into how much money is spent at National Parks in the United States. Is a big factor that increases the amount spent the weather? I want to see if we will be able to predict the annual spending amount at National Parks from the monthly average of various weather patterns throughout the year.

Data collected from the last decade going back to 2012 on the annual average spending at national parks and the weather there, will be utilized to try to predict what the outcome for annual spending will be for 2023. There may be other factors in data that can be collected to improve the predictability of annual spending.

Requirements

* Gather Data Set
  + Find a data set containing the national parks annual spending and monthly weather data for about the last ten years.
* Clean Data Set
  + Clean the columns so that they can be used to join other datasets and predict the target feature of annual spending.
* Answer Questions from Data
  + A written report covering the data. It will attempt to answer the following questions.
    - What parks have seen an increase in temperature?
    - Is there a visible correlation between the temperature and annual spending?
    - What notable parks have the most annual spending?
* Create Visualization of Data
  + Create charts to support the questions that are asked about the data. Minimum of one visualization per question.
* Prediction Model
  + Using machine learning to look at the features in the data to try and predict the annual spending at national parks.
* Validate Model
  + A holdout set of data will be used to test the accuracy of the model. A stage to refine and make changes to improve the model.

Stretch Requirements

* Make Popularity Metric
  + A metric built upon the average visitors and population of the main state that hosts the national park. Additional data may be needed for this.
* Simple Web Site
  + Make a site to display data analysis and machine learning models.
* Include Elevation and Size of Park
  + Incorporate the elevation and the size of the national park into the data set and model.
* Interactive Visualizations
  + An additional visualization to the already required ones. It will be interactive so the viewer can explore the chart to gain additional insight.

Design Overview of the Product

This product doesn’t have the typical user end. The user interaction will just be a report of the project. The project will contain a written report and visualization leading to the prediction model.

Resources

* Data to get the annual spending. https://irma.nps.gov/DataStore/DownloadFile/616300
* Data to get additional information about national parks. https://www.nps.gov/subjects/digital/nps-data-api.htm
* Data to get information about weather for various locations. <https://power.larc.nasa.gov/data-access-viewer/>

Verification

To verify that this works as designed the model will need to predict the annual spending of the national parks. To test that the requirements are completed I will have peers with experience with data science, evaluate each requirement.