Project Name

National Parks Spending Prediction

Contact Information

Hayden Howard, Email: how21014@byui.edu, Phone: 9072551314

Abstract

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 temperature? I want to see if we will be able to predict the annual spending amount at National Parks from the weather throughout the year.

Background

I Have found two data set relating to the project. They have done an analysis on just the annual spending of National Parks but from what I have found they did not try to find a cause for the increase in spending.

I have had some practice working with data and training a prediction model, but I have no prior experience working with this subject so I will need to explore more to gain a proper understanding of what the data means and other factors I will need to consider.

Description

Do Years with a higher average temperature have a higher annual spending then previous year. This project will take the spending and weather data from previous years to see if there is a trend. After this I will take the weather data so far for this year to predict what the annual spending will be. This project will be for those in charge of national parks so they could start planning their possible budget plans earlier. For this project to be successful I will need to gather enough data for the parks. I am planning on choosing ten different parks all ranging in size, population in the area, and popularity of the park. I am planning on getting data from these parks from the last ten years. This model will take the weather from that point in the year and can be used to predict the annual spending of National Parks. This prediction model and analysis will be complete once there is a model made that can closely predict the annual spending. The analysis will be complete when the data is thoroughly studied, and other patterns and trends are found and displayed.

Significance

This project involves gathering data. As far as I know there is not one data set that currently exists the provides the spending at National Parks and the temperature and weather of the area. Gathering the data and being able to wrangle it are great skills to have. Putting the data to see if we can predict the outcome in annual spending and analyzing the data to find trends will be an attractive addition to my resume. While I don’t believe there is a demand for this project to be completed, I believe it has an interesting question that may lead to further study.

New Computer Science Concepts

I believe I may be able to put a web scraper to use to aid me in gathering my data. I have yet to use one before so it may take some time to understand how it works. Another new concept that I have not put into full practice would be putting my analysis in a format that can be displayed to anyone. I have looked into some possible options which include Power BI, simple website to host my findings, or an in-depth mark down file.

Interestingness

I want to have my future career involved with machine learning and other AI. This project shows that I understand the importance of collecting and analyzing data to be able to properly use it in a machine learning environment. While the machine learning may end up not being too complex, it will be my own project built from the ground up. The subject of the project is the spending at national parks based on the weather. While I currently haven’t visited many of the national parks it is something I hope to be able to do when I have more time available. I love the outdoors, and knowing if the weather has an overall impact on the money the parks make, can affect how they are run in the future. As the temperatures slowly increase it could have an impact on the parks. Knowing that not only the weather can have an impact, but the rise in social media promoting the parks and increasing size of populations are factors that needed to be considered.

Tasks and Schedule

There are three main parts to this project, gathering data, analyzing data, and then the prediction model. These three parts can be split into smaller tasks, with some retrospective time as well to possibly gather more data. After the main parts I need to put everything together ready to present my findings. I have not done a project of this scale before so my time estimations may be inaccurate.

* **Gathering Data** 
  + **Web Scrapper: Time-10 hours**
  + **Manual Entry: Time-20 hours**
  + **Putting Data Together: Time-10 hours**
* **Analyzing Data**
  + **Finding Trend: Time-10 hours**
  + **Graphing and Tabeling-20 hours**
* **Retrospective**
  + **Data is Good: Time – Possible 1 hour.**
  + **Gathering More Data: Time – Possible 10 hours**
* **Building Model**
  + **Research Best Model: Time-1 hour**
  + **Make Model: Time- 2 hours**
  + **Train Model: Time-2 to 10 hours**
* **Presentation**
  + **Make The Site to Host: Time-2 hours**
  + **Put Findings on The Site: Time-2 hours**

Required Resources with Costs

I currently plan to collect the data from the National Parks Government website. Their data is free, and some are organized while others are not. I may email them asking if they have data files already of what I am looking for. This is a page that has the data I am hoping to find for previous years. <https://www.nps.gov/orgs/1778/vse2021.htm>

To gather the data for the weather, currently I am planning to use the NOAA site and put it in the location of the national parks to collect the data throughout the years. I may be able to find other API’s or data sets that have this information as well. This is a link to the NOAA site for how I am going to get the weather data as of now, <https://www.ncei.noaa.gov/access/search/dataset-search?dataTypes=CLDD&bbox=46.562,-108.899,45.312,-107.649&pageNum=1&startDate=2022-01-02T00:00:00&endDate=2023-01-02T23:59:59>

Depending on the capabilities of my laptop, and size of the dataset I may be able to train the model on my laptop. If my laptop does not have the capability to, I will try to train the model on Goggle Collab. If I can do web scraping to gather my data, I plan on using the python package, beautiful soup. For data wrangling and analysis, I am going to use the python package, Pandas. For graphing the data, I may use python packages, seaborn or plottly. I plan on using the scikitlearn library to train my prediction model. If I decide to host a site, I may need to buy a domain which will cost between 10 to 50 dollars. I don’t believe I will have any other costs to expect from this project.

References