National Parks Data Analysis Project

Project Overview

This project will be based on visitation data to the National Park System in the United States. The project will examine the number of visitors to national parks, seasonality of their visitations and examine why people choose to visit national parks.

Motivations

I have chosen to pursue this project because of my appreciation for the National Park Service and their contributions to fight climate change. The National Park Service is an often-overlooked function of the US government; however, their preservation of the natural landscape helps offset carbon emissions in the US, provide habitats for the flora and fauna of the US, and create research opportunities for scientists and historians alike.

Using this data, I hope to discover how the National Park System has been doing over the years and either highlight their value or find areas and methods for improvement in the system.

Data Source

The data for this project was provided by the United States National Park Service. Specifically, it was found on the Integrated Resource Management Application portal at www.irma.nps.gov.

The data is external data, collected by the National Park Social Science Program and owned by the National Park Service and the United States Department of the Interior. The data is usage data, that is collected by a combination of automated counting machines, counting by NPS personnel, and estimations based off vehicle and average person-per-vehicle statistics in larger parks. More information about how the data is collected is available at https://www.nps.gov/subjects/socialscience/index.htm.

The data is relevant and timely. It is collected monthly by each individual park, and it is reported to the NPS. The National Park Service is a trustworthy source of information. They have published their methodology as well as changed to their methodology over the past 100 years to create transparency between themselves and data users. This data set is classified as usage data. The NPS collects this data monthly to share with the public, as well as allocate resources within the National Park system.

Data Limitations and Ethics

Although this data is great and collected by the trustworthy and dedicated personnel of the National Parks Service, it does have certain limitations. As the NPS states in their collection methodology, in their larger parks some visitor numbers are estimates and therefore may not be completely accurate. Additionally, according to the NPS methodology it is collected by humans, and it is therefore subject to human error. However, I believe these limitations are minor and will not interfere with the results of my analysis.

This data is provided by the National Park Service for use by the public. It does not contain any personal or private information and exists as a form of transparency between the government and taxpayers. I do not foresee Ethical issues in using this data.

Questions to Investigate

1. Which National Parks receive the Highest visitation?
2. Which National Parks receive the most non-recreational visitation?
3. What classification of park has the highest visitation?
4. When do National Parks receive the highest visitation?
    a. Is this the same across regions? States? Park Style?
5. Which National Parks are the most popular per season?
6. Has National Park Attendance increased or decreased over 10 years?
    a. Are there any major increases or decreases? What could have possibly caused them?

Data Profile

Variable Name	Description
Data Grain	ParkCode-Year-Month (to create unique codes for each data point)
Park	Name of National Park
Unit Code	Four Letter Abbreviation for Park
Park Type	Style of park (list of park types)
Region	Region of the United States park is located in.
State	State park is located in.
Year	Year data was collected
Month	Month data was collected

	Rec Visits	Non Rec Visits	Rec Hours	Non Rec Hour	Con Lodge	Con Camp	Tent Camp	RV Camp	Back Country Camp	Non Rec Overnight	Misc. Overnight
Count	49,552	49,552	49,552	49,552	49,552	49,552	49,552	49,552	49,552	49,552	49,552
Mean	6.613549e+04	3.615488e+04	2.889350e+05	3.615488e+04	665.161971	289.609017	727.653576	493.340975	416.778959	23.542057	403.576667
Max	2.238812e+06	4.736394e+06	2.021239e+07	4.736394e+06	151190.000000	164885.000000	145536.000000	78060.000000	62848.000000	19918.000000	454405.000000
Min	0	0	0	0	0	0	0	0	0	0	0