Description of the Use Case

Casey Ellis, Meenu Ravi, Kentaro Hama

Hypothesis (Research Questions)

As our final project, we will analyze the patterns of the number of visitors to national parks and predict the number of visitors in the future. Accurate forecasting of the number of visitors is an important issue for park managers because if the number of visitors exceeds the expectation, there will be a shortage of park guides and facilities such as restrooms and camping sites. This would be detrimental to the visitors and result in poor visitor experience. Conversely, if expectations are off and the number of visitors is smaller, excess staff and facilities can create financial burden. Thus, accurate prediction of visitors is an issue that directly affects the optimal distribution of management resources.

To understand the pattern of visitor numbers, we first apply the decomposition methods. Presumably, the series has a seasonality, with a peak during the summer vacation season. There may also be an upward trend due to lower moving costs. Next, we will apply the ARIMA models to predict the number of visitors to the park. As the class progresses, if there is a more appropriate method of analysis, we will try that as well. For example, it may be informative to see how variation in visitation can be explained by other variables such as environmental factors, social policy and economic conditions.

Dataset Description (data source: https://irma.nps.gov/STATS/)

The data used in our analysis is the number of visitors by park and by month from 1979 to 2020. The data will be obtained from Visitor Use Statistics provided by the U.S. National Park Service. We will focus our analysis on Yosemite National Park. The number of visitors is recorded separately for recreational and non-recreational visits, and we will focus on recreational visitors.

Work plan for Group Members

Casey	**Meenu**	**Kentaro**	**Notes**
EDA	EDA	EDA	Since we each may have different approaches to EDA, we plan to all do this to get familiarized with the dataset and combine unique features into our final presentation
	data cleaning and preparation diff		Running stationarity tests
Model 1	Model 2	Model 3	We can each choose a model a) ARIMA model b) another model c) another model From this, we can choose the one that gives us the best results for our presentation
			Model evaluation - RMSE, MSE, MAE, SMAPE, MAPE
Deepen the research questions	Deepen the research questions	Deepen the research questions	- Conclusions and next steps - Search for covariates that could be related to the variation in visitation and merge them with the visitation data set.