**Applied Data Science Term Project Week 1-4**

Paul Galvez

Department of Data Science, Bellevue University

Summer DSC 680 T301: Applied Data Science

Dr. Catherine Williams

9/17/23

**Applied Data Science Term Project Week 1-4**

**Business Problem:**

Hawaii’s primary source of income for the state is the tourism and hospitality industry. While the state has made efforts over the past twenty years to diversify the income streams that fuel the state’s ability to generate revenue, the sheer number of tourists from all over the world remains the most reliable source of income for Hawaii. However, tourism is heavily impacted by the domestic and world economies. Because travel and leisure can be expensive for most people, a luxury vacation to the islands might not always be a possibility or even a reality. The unpredictable economic fluctuations can result in a lack of or excess numbers of visitors to the Hawaiian Islands, which can make running a hotel, restaurant, or other business that relies on tourists to survive. So, the business problem for hotels is the knowledge and business insight to predict better the number of tourists and the number of nights they stay. If there was access to or a way to predict the number of nights tourists will most likely stay, better hotels can plan their workloads and resource planning.

**Background/History:**

The history that connects Hawaii and tourism is well known, and the reliance on the vacation industry can be traced back to the early 1940s when the ability for people to travel across the Pacific Ocean was becoming mainstream amongst the American public. What has yet to be available to businesses like hotels, where the number of tourists staying fluctuates throughout the year and the number of nights they stay changes due to several factors can be challenging to plan for from a resource requirement perspective. Historically, hotels have mainly utilized educated best guesses based on historical and common knowledge data. For example, it is common knowledge that more people will visit Hawaii in the winter months because tourists want to escape the cold temperatures on the mainland or in a foreign country where winter can be lengthy and unfriendly. However, the guessing game can be just that, a sheer shot in the dark based on nothing more than a feeling or instinct that people will come and stay in high or low numbers as the year moves from season to season. There is also the factor of not being able to predict the number of nights tourists stay, which is also a factor in planning for resources across the visitor industry in Hawaii. The more nights guests stay, the more requirements hotels are going to need to meet the demands of both human and physical.

**Data Explanation (Data Prep/Data Dictionary/etc):**

The dataset is from Kaggle and is based on the average number of night stays tourists to Hawaii booked from multiple sources. The years that comprise the dataset start in 1999 and run through 2021, which means the data shown covers multiple world and national events that could have negatively and positively impacted the tourism industry in Hawaii. For example, the global economic downturn in 2007 affected people’s financial ability to spend on vacations and travel. The dataset also shows several critical categories of travelers to Hawaii labeled by group and indicator. For the group indicator, the following categories identify the type of traveler; they include all visitors by air, hotel-only visitors, first-time visitors, and honeymoon visitors, to name a few. The second variable is the indicator, which categorizes where those visitors stayed in Hawaii. For example, LOS on Oahu is the length of stay on Oahu specifically. There are also duplicate data points for LOS on Maui, Kauai, Molokai, and the Island of Hawaii, or the Big Island. In terms of data prep, NAN values needed to be removed because the dataset needed numerical points of information that focused on predicting the number of night stays, or LOS, in Hawaii and on which island. Also, some duplicate values needed to be dropped to avoid confusion and possible overcounting of the night stays that would have doubled the data points for those categories with that duplicated data. Below is a picture of the prepared dataset ready for further analysis and visualization.

A table of numbers and symbols

Description automatically generated

**Methods:**

The critical method for predictive analysis is regression, specifically linear regression, and creating some data visualizations that support the overall analysis. The visualizations focus on the year of the data range and the lengths of stay by year. Some indicators in the visuals show an overall increase in night stays in the last years of the dataset, 2019- 2021, prior to the COVID-19 pandemic shutting down the global economy, which devastated the travel and tourism industry. For linear regression, the project needed to show the prediction of the dependent variable, in this case, the number of night stays, LOS, and the independent variable, in this case, the year ranging from 1999-2021. The strength of the two variables can be shown in the results of the project, and we can use those scores in predicting the number of nights to plan for throughout the year; deploying those resources is critical to managing the business of hotels, which can be difficult and costly with no insight into the best way to predict those numbers. The visual below shows the length of stay in Hawaii. The x axis shows the length of stay in days and the y axis shows the year those length of stay averages occurred.

A graph of blue bars

Description automatically generated

**Analysis:**

The models yielded exciting results, which can be deployed to predict the length of stays in Hawaii. Overall, the analysis showed that the number of nights remained steady, with some drop in the length of stays during more economically challenging years. However, a critical point in the data was not necessarily the number of tourists. However, the number of stays was shorter during slower or more difficult economic periods. The linear regression scores were as follows: The year used was 2018.

A screenshot of a computer

Description automatically generated

For the purposes of modeling another set of dependent and independent variables, the project also looked at the following model and scores for 2004:

A screenshot of a computer code

Description automatically generated

**Conclusion:**

The project overall helped show the number of night stays as being steady throughout the year ranges that comprise the dataset from a booking perspective. However, the number of night stays does fluctuate with the national and global economy. The project's original goal was to find the number of night stays for resource planning and deployment, both physical and human. With that goal in mind, the communication from the project results would be to plan for a similar number of tourists to the islands based on historical data and plan for shorter stays should there be negative economic periods. On the other hand, the project also shows to plan for longer lengths of stay when the economy is strong, and people can travel to Hawaii and stay longer, which is a valuable insight for hotels to manage their businesses.

**Assumptions:**

One of the key assumptions is that the dataset presents the total night stays based on the guests staying their entire booking length rather than cutting their time in Hawaii short. In other words, the data assumes those lengths of stays are tourists completing their entire booking. In most cases, hotels do not refund or credit their guests who check out early. However, in some cases, the hotel can offer refunds for nights not stayed based on their reason for leaving early. For example, a family emergency out of the guest’s control could prompt a refund based on the number of nights not completed from the intended lengths of stays. Another assumption is that the dataset is from all sources of guests, which means that those lengths of stay are from guests who are not just American but also international travelers from other countries who favor Hawaii as a vacation destination.

**Limitations:**

The dataset was complete except for knowing where the country-of-origin visitors were coming from, which can significantly impact business planning and resource procurement. For example, Japanese tourists comprise a sizable total number of Hawaii guests and are historically known to be the steadiest stream of international tourists. Hotels and other businesses that support Japanese tourists would benefit from being able to predict the lengths of stays for that group because Japanese-speaking people in Hawaii is a finite resource, and not having enough bilingual staff could have significant negative impacts on the ability of businesses to deliver the same quality of stay and experience to those guests that are not able to communicate in English.

**Challenges:**

A challenge with the project was linking the information to the intended goal of the predictive analysis aspect of the endeavor. Because there were lengths of stray across multiple islands, the data needed to connect the islands to reflect a model that any hotel or business that relies on those night stays can apply to manage their resource needs. Also, the challenge of not knowing the guests' country of origin was a challenging aspect of the dataset because human resources could be challenging to plan for and acquire without that knowledge.

**Future Uses/Additional Applications:**

The project’s findings can be applied to future predictions for hotels and businesses that rely heavily on tourism to run their companies more efficiently and cost-effectively. Because Hawaii’s reliance on tourism will never change, the project’s modeling and findings can be applied to any hotel on any island. Hotels in Waikiki will continue to be the epicenter of tourism in Hawaii. However, hotels on Maui, Kauai, and Hawaii Island can also apply the same predictive models for their businesses because the lengths of stay on a more remote island can face more complexities than Waikiki, which is located on Oahu, the most advanced and modern of the all the Hawaiian Islands and is home to primary harbors and ports for goods and services to the entire state.

**Recommendations:**

Recommendations are to deploy the model and the findings to apply necessary resource allocation for future night stay planning. Also, the findings can be applied to other hotels in more remote areas of Hawaii. For example, hotels located more than twenty miles from Waikiki are typically less populated tourist-wise on the island of Oahu. However, they can still sell out of rooms because of their smaller capacity. With that in mind, the project’s findings can support the insights needed to run smaller hotels that cater to their needs and requirements to host their guests in a way that delivers a high-quality experience that guests expect when visiting Hawaii. The increase in overall value for the guests can lead to repeat business and a positive perception amongst the traveling public.

**Implementation Plan:**

The plan to implement the model and findings will be sent out to those leaders who can make the most effective change and impact on their lines of business. Departments such as finance, project management, and human resources will be the first departments to be presented with the findings and recommendations resulting from the project.

**Ethical Assessment:**

Ethically, the project and the data rely on honesty from the reporting locations. Hotels and Air B and Bs reported these numbers based on their bookings from that specific year. Those businesses reported those based on bookings rather than actual check-ins for those locations. The project ethically relies on the honesty and transparency of those hotels and businesses. Another ethical concern is the ability of hotels to be the dominant option for housing when tourists visit Hawaii. The state has placed stringent regulations on other housing options, such as Air B and Bs, because of hotels' reliance on a steady stream of international and national visitors.

**Ten Questions from an Audience:**

1. How does the project asses’ areas of the tourism industry that are not necessarily hotels?
2. What are the implications for the state in terms of overall planning and development?
3. Given Hawaii’s limited resources both human and physical, are there possible predictions that could overtax the Hawaii population and ecosystem?
4. When we have an influx of people staying longer, how does the number of night stays affect the overall natural environment and resources of Hawaii?
5. If we know there are going to be visitors staying longer, should hotels and other businesses look at an increase in amenity offerings?
6. From a government perspective, does the longer number of stays in fact increase revenue?
7. Are there too many tourists staying longer?
8. How do such a model and predictions apply to other industries?
9. Should hotels look at increasing the cost of per night stay? Or would that affect effect the ability of guests to spend ancillary dollars?
10. What should hotels do if the number of night stays does not meet the predictive number?

**References**

[Hawaii Travel Length of Trip 1999-2021]. (n.d.). Retrieved September 17, 2023, from https://www.kaggle.com/datasets/csafrit2/hawaii-travel-length-of-trip

**Appendix A**

Subject of the Appendix

The appendix covers the visualizations from the project that show the outcomes of the analysis from the dataset. Those visualizations were crucial to the understanding of the information and allows the audience to better comprehend the narrative the data is attempting to show.

A table of numbers and symbols

Description automatically generated

A graph of blue bars

Description automatically generated

The image below shows a scatter plot for dates ranging from the start and end of the dataset:A graph with numbers and dots

Description automatically generated