Allison Biediger
DSC 680
Project 3-Milestone3

Cultural Tourism Data Analysis

Business Problem

In the tourism and travel planning industry creating personalized experiences is essential in keeping a competitive advantage. There are many factors that can go into the satisfaction level of clients on trips, and as a travel planner it is necessary to understand these factors to help guide data driven decisions. If the decisions made are not utilizing personal data and preferences, instead decided based on general trends the business runs a risk of not meeting the clients needs and expectations as well as potentially losing their loyalty. This analysis focuses on this problem by utilizing tourism data to gain understanding around customer preference trends, and to develop a predictive model that can recommend destinations that are best suited to client preferences. In addition this will find the most important factors that contribute to client satisfaction through a feature analysis. The findings of this analysis can be used by those in the tourism industry such as travel agencies to improve their customer targeting and to make data driven decisions. These decisions will help to enhance their clients travel experience and will lead to higher satisfaction rates and customer loyalty.

Data Explanation

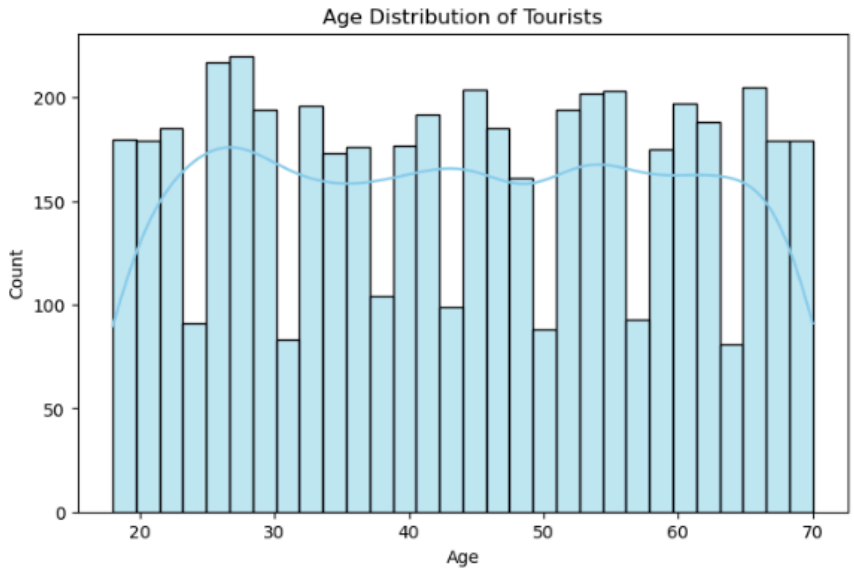
For this analysis I utilized a cultural tourism dataset Cultural Tourism Dataset from kaggle. This includes variables such as traveler demographics, previous travel history, interests, satisfaction ratings and other influencing factors for travel decisions.

Methods

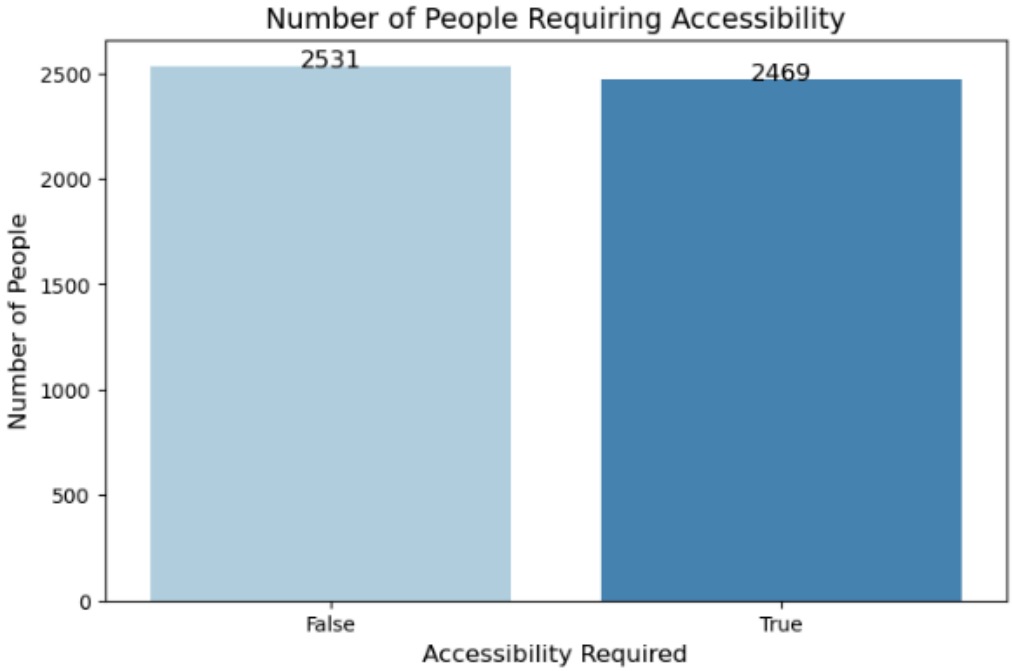
Starting this analysis I began by focusing on cleaning the data and understanding the different variables. I was then able to focus on exploratory data analysis where I was able to create visualizations to identify patterns and correlations. Then I was able to begin focusing on model creation, to begin I focused on features that would affect their potential satisfaction and used them to create a linear regression model. Following this I assessed the models evaluation metrics such as accuracy and precision. Through this analysis I was also able to focus on ethical considerations and the potential practical elements of the suggestions.

Analysis

Through this analysis of the tourism data there were many important findings. To begin gaining an understanding of the current clientele I focused on understanding their ages and interests.

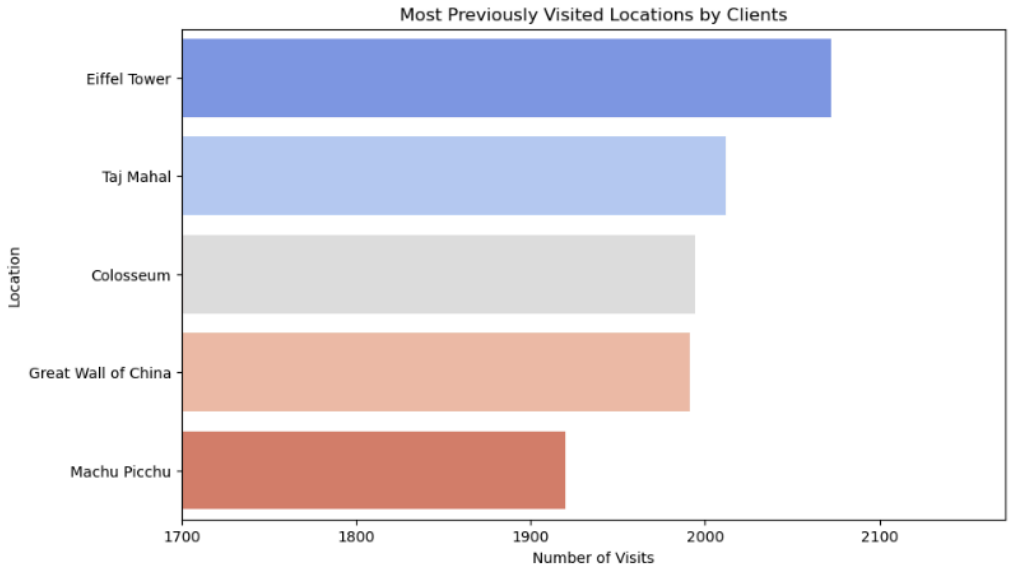


While focusing on the ages of current clients I found no clear trend, overall there is a good spread of clients across the ages 18 to 70 years old. The largest group was found to be within the 25-28 years old however this was not significantly more than the peaks around other ages.

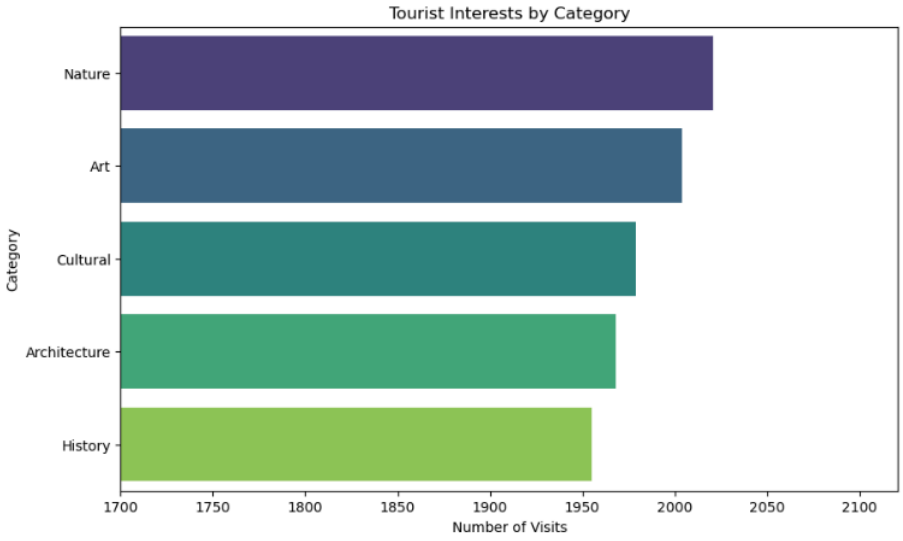


While looking closer at the need for accessible options I was able to focus on comparing if accessibility is required. I was surprised to see that there are nearly 50% of clients that have accessibility needs when they are traveling. This showcases an important factor when choosing

locations the company supports trips, making sure that areas are accessible for clients of all abilities is essential.



By looking deeper into the locations that were previously visited by clients I found that the Eiffel Tower was significantly visited more often than the other locations, with Machu Picchu visited the least. This is interesting to see the locations that have been previously highly visited to help guide decisions of locations that are supported by the travel company.



Next by looking closer at the interests of the current clientele I found that Nature is the most common interest while history is the least common. This is also a great finding to keep in mind when determining future destination locations that will be supported by the company.

Finally I created a prediction model that would predict client satisfaction based on the collected data. This included tourist rating, previously visited locations, interests, as well as tour

duration data. The model was found to have a Mean Absolute Error (MAE) of 0.33 and an R2 score of 0.74. The MAE, which indicates the absolute difference between the predicted and actual satisfaction scores, of 0.33 shows that on average the model predictions are off by about 0.33 satisfaction points. The R2 value, which measures how well the model can explain the variance in the target variable, of 0.74 shows that about 74% of the variation is explainable by the model's features.

Following the model creation I analyzed the feature importance and found that the tourist rating was the highest importance with 0.62. This shows that previous tourist ratings have a significant influence on the prediction of satisfaction. Following this was previous locations visited with 0.041 and client interest with 0.032. These being the next highest in importance shows an opportunity area for this model. Even though the model is able to predict correctly around 74% there are likely more important factors at play that are not currently monitored.

Conclusion

In conclusion there are many essential findings from this analysis. It was found that there is a large need for accessible options for travel with nearly 50% of the current clientele requesting accessible options. It was also found that the most common interest was nature, this finding can help to drive forward decisions about future supported travel locations. In addition the most previously visited location was the Eiffel Tower, this iconic location has been a previous favorite amongst the clientele knowing this when choosing future travel locations. Through the analysis a prediction model was created with a 74% accuracy of prediction that is able to predict the satisfaction of the clientele. This can be used to help when planning locations for guests to visit in the future, by using this data driven approach the company can gain a competitive edge and help increase customer loyalty. Through the factor analysis it was found that the most influential factor was tourist rating, this showcased that the previous tourist ratings often accurately reflected the locations. Utilizing these findings in combination with continuing to collect relevant travel preference data the company can keep growing their business and improve their recommendations.

Assumptions

In this analysis I had many assumptions at play, including the data accuracy and completeness that accurately represents the clients. Another major assumption made is the satisfaction variable being an accurately measurable outcome. I also assumed that there is a level of consistency for travel preferences, meaning that past travel data can effectively inform future decisions. By utilizing a linear regression I assumed a linear relationship between the different variables, this method may not showcase more complex interactions. Finally I assumed that the data was collected ethically and without bias and it accurately represents the clientele.

Limitations

There are a few limitations of this analysis including the simplification of satisfaction factors, while the model does utilize key features that influence client satisfaction, it may not showcase all nuanced or other subjective factors such as unexpected travel delays, cultural differences, and personal emotions. While I did follow the assumption that travel preferences are stable over time this could limit the analysis as other external factors may change travel preferences. A few examples that may cause this influence are economic conditions, global events or even personal circumstances.The model created is currently only utilizing historical data therefore faces a limitation when it comes to real time adaptations, not allowing for the model to update dynamically in response to changing or evolving preferences.

Challenges

Through this analysis I came across many challenges including the layout of the data. During the cleaning section for specific columns there were listed locations or interests that were categorical and that had to be broken down. Choosing the correct model also showed some challenges, the linear regression model was selected as it yielded the most accurate predictions however there are likely other factors at play that there was no collected data on.

Future Uses

There are many potential future uses for this analysis, including utilizing this to create personalized travel recommendations. Collecting more data from clients such as preferences for weather, budget, and experiences would help the model give more pointed and insightful recommendations. Using these findings would also allow the travel agency to focus on areas that best represent their demographics. For example with a large amount of clients looking for accessible options when choosing other destination locations this can help lead that decision. In addition the most common interest for the current clientele is nature, combining these findings when selecting new locations for clients would be a great way to continue growing their clientele.

Recommendations

From this analysis I recommend utilizing the findings when selecting new locations that the company supports travel to. For example knowing that nearly half of the current clientele has accessibility needs and that the most common interest is nature can help when deciding new locations to focus on. In addition I recommend obtaining more information from clients beginning their travel process, including more preferences that may affect satisfaction. These could include variables such as weather preferences, budget requirements, and experience type preferences. By also collecting these further data points the model can be updated and give more accurate predictions that can help drive decisions to personalize travel experiences.

Implementation Plan

To implement these recommendations the travel industry should aim to use data driven approaches to crafting travel experiences for their clientele. By using the model that is created they will be able to better cater to their clients needs and preferences. In addition, continuing forward the business should work on collecting more preference data and using these data points to update the model. This will help provide a more accurate model as well as give insight into trends amongst the clients. In combination with the preference data the company should try to encourage deeper member feedback, gathering further data in what aspects were positive vs negative in the clients trip will also help feed future recommendations.

Ethical Assessment

Through this analysis there are many ethical considerations at play including data privacy and security. Protecting the privacy of clients is essential, this dataset did not contain any personally identifiable information. In addition bias is something that was considered through the analysis, by providing important context through the analysis as well as clear transparency this concern was able to be managed.Continuous evaluation of the models performance is needed to ensure that it can adapt to changing ethical standards and public expectations.aining clear and informed consent is necessary before both data collection and data analysis.

Potential Questions:

1) *How did you decide what features to focus on in this analysis?*
    a) I wanted to focus on the features the travel company would have before a client goes on their tour as well as the satisfaction of the trip they took. This allowed us to see how the collected data could be used to predict the satisfaction rating the clients reported.
2) *Are you able to track seasonal trends with this data?*
    a) I could not track seasonal trends with this dataset since there was no date/time data provided in this dataset. A further collection of data regarding these factors would be essential for a seasonal trend analysis that could provide valuable insights.
3) *What are some factors that need to be considered with accessibility needs?*
    a) There are many factors to consider for accessibility, this could mean accessible hotels in the area, accessible locations that have ramps as well as destinations that can be enjoyed by people of all ability levels.
4) *How can the company collect more satisfaction data from the client after a trip?*
    a) The company can send out surveys both before and after the planned trips. The survey before can be used as a good opportunity to understand the client, gain further insights in their expectations. Some additional variables for this could include budget and weather preference. After the trip, collecting further data into what difficulties they had and specifics in what aspects they liked can also help gain a better insight into guest outcomes.

5) *Can other travel companies utilize this analysis?*
    a) Potentially yes other companies can utilize these findings, though if their supported locations are not the same the importance of the findings may not be applicable. Companies should consider utilizing these methods with their own clientele data.
6) *What other data points would help deepen this analysis?*
    a) Date and time information would help provide a better understanding into if there are any seasonal trends for the locations. Further data in preferences with data around budget and weather preferences would help guide recommendations as well as give further data for the model potentially helping increase the accuracy.
7) *When choosing a new travel location to support what should the company keep in mind?*
    a) The company should keep in mind the high demand for accessibility among their clients as well as the popular interests with the most common interest being nature. Utilizing these findings when determining the next supported location will help match the needs of the current clientele.
8) *Is there a specific demographic this company should be marketing towards?*
    a) The analysis showed a well balanced age distribution from 18 years old to 30 years old. However the accessibility needs were found to be necessary for nearly 50% of their clients. The company can use this information to make marketing campaigns that focus on the accessible options available.
9) *How can this be used to give the company a competitive edge?*
    a) The ability to use a model to give personalized recommendations will help give a competitive edge over companies that only follow overall trends. This will help give personalized experiences that will help retain loyal clients and provide higher satisfaction scores.
10) *What are some specific data points that should be collected to strengthen the model?*
    a) Budget, weather preferences, as well as date and time can be used to help strengthen the model and help gain a better understanding of client preferences.