**TRAVEL DESTINATION RECOMMENDATION SYSTEM**

#### Authors[¶](http://localhost:8889/notebooks/Downloads/CAPSTONE%20PROJECT/travel-destination-recommendation-sys/notebook.ipynb#Authors)

1. Dennis Mwanzia
2. Pamela Awino
3. Joshua Rwanda
4. Pauline Wambui
5. Kane Mwendo
6. Ian Macharia
7. Nelson Kemboi

**BUSINESS OVERVIEW**.

AfricuraAI is a business dedicated to revolutionizing the way tourists explore Africa by leveraging the power of machine learning and recommendation systems. With the understanding that tourists often face challenges in finding suitable travel destinations that meet their preferences, budget, and time constraints, AfricuraAI aims to provide personalized recommendations for the best tourist destinations in Africa.

Our core objective is to develop a robust machine learning model that analyzes various factors, including customer reviews, budget limitations, specific locations, available amenities, and residence types, to suggest the ideal tourist destination that aligns with each user's unique preferences. By considering these factors and analyzing text reviews, our model seeks to accurately predict the preferred tourist destinations for both new and existing users.

The AfricuraAI project brings immense value by addressing a real-world problem faced by tourists, namely the lack of personalized recommendations when planning their trips in Africa. By employing machine learning algorithms and examining a wide range of factors, our model aims to enhance the travel experience by suggesting destinations that not only meet the users' preferences but also fulfill their specific requirements.

Through the AfricuraAI platform, tourists will have access to a comprehensive and tailored travel planning experience. They can input their budget limitations, desired location preferences, preferred amenities, physical attractions of interest, and desired residence type. Our machine learning model will then analyze these inputs along with customer reviews and available data to provide the most suitable and personalized recommendations for the best tourist destinations in Africa.

With AfricuraAI, tourists will no longer struggle to find suitable travel destinations within their budget and time constraints. They can confidently plan and embark on a satisfying trip, knowing that their preferences and requirements have been taken into account. By harnessing the power of data analysis and machine learning, AfricuraAI is poised to revolutionize the way tourists explore Africa and create unforgettable travel experiences.

**PROBLEM STATEMENT.**

Tourists visiting Africa often struggle to find suitable travel destinations that align with their preferences, making it challenging to plan a satisfying trip within their budget and time constraints.

Our main aim as AfricuraAI is to develop a machine learning model i.e. recommendation system that provides personalized recommendations for the best tourist destinations in Africa. By considering customer reviews, budget constraints, specific locations, available amenities, and residence type, the model aims to suggest the ideal tourist destination that aligns with the user's preferences.

By analyzing the text reviews along with the additional factors available in the dataset, the project seeks to build a model that accurately predicts the preferred tourist destination for both new and existing users. The model will take into account factors such as customer reviews, budget limitations, location preferences, amenities, physical attractions, and residence type to recommend the most suitable tourist destination in Africa.

The value of this project lies in its ability to address the real-world problem of providing personalized recommendations for tourists in their search for the best tourist destinations in Africa. By utilizing machine learning algorithms and analyzing a range of factors, the project aims to enhance the travel experience for users by suggesting destinations that meet their preferences and requirements.

**OBJECTIVES.**

**MAIN OBJECTIVE**

1. The goal is to build a machine learning model that can accurately predict hotel ratings based on customer reviews, budget constraints, specific locations, and the type of residence. By analyzing the text reviews and incorporating additional factors, such as sentiment analysis, budget range, geographical distance, and residence type, the objective is to develop a robust model that can make accurate predictions for new, unseen hotels. The model will help users make informed decisions when selecting hotels by considering their preferences and constraints.

**SPECIFIC OBJECTIVES**

1. To create a Recommendation system that can be used to suggest travel destinations based on budget.
2. To identify top destinations in Africa
3. To foster customer loyalty and repeat customers, building long term relationships between customers and travel destinations.
4. To increase customer engagement and satisfaction by providing comprehensive and accurate information about tourist destinations.
5. To generate revenue by engaging travel destinations to have them listed on the site.
6. To collect user information and feedback to improve the recommendations in the long run.

**THE DATA**

We sourced data by scraping destination review data from **TripAdvisor. Each column contained the following**

• id: Unique identifier for each destination.

• type: Type of destination, i.e. whether an attraction, a hotel etc.

• category: Category of the destination. It describes the same features as type and hence a repletion of the type column.

• subcategories: Subcategories associated with each type. i.e. the kind of attraction or hotel type.

• name: Name of the tourist destination.

• locationString: String representation of the location of the tourist destinations in each country.

• description: Description or details about the tourist destination.

• image: Image associated with each tourist destination.

• photoCount: Number of photos available for each destination.

• awards: Awards received by each tourist destination.

• rankingPosition: Ranking position of each destination within its location based on trip advisors trip ranking.

• rating: Rating of each tourist destination.

• rawRanking: Raw ranking of the destination.

• phone: Phone number associated with each destination.

• address: Address of the destination.

• addressObj: Address information in object format.

• localName: Local name of the destination.

• localAddress: Local address of the destination.

• localLangCode: Language code for the locals where each destination is located.

• email: Email address associated with each destination.

• latitude: Latitude coordinate of the destination's location.

• longitude: Longitude coordinate of the destination's location.

• webUrl: URL associated with the destination.

• website: Website URL of the destination.

• rankingString: Ranking information of the destination in string format.

• ranking Denominator: Denominator for the ranking.

• neighborhood Locations: Other destinations with the target destination neighborhood.

• nearestMetroStations: Nearest metro stations around the destination.

• ancestorLocations: Ancestor locations of the destination.

• ratingHistogram: Histogram data for the destination's ratings.

• numberOfReviews: Number of reviews for the destination.

• reviewTags: Tags associated with the reviews.

• reviews: Reviews of the destination.

• booking: Booking information for the destination.

• offerGroup: Group offers associated with the destination.

• subtype: Subtype or specific type of the destination.

• hotelClass: Class or rating of a hotel item.

• hotelClassAttribution: Attribution information for the hotel's class.

• amenities: Amenities available at each destination.

• numberOfRooms: Number of rooms available (for hotels).

• priceLevel: Price level or range of the destination.

• priceRange: Price range of the destination.

• roomTips: Tips or recommendations for rooms (for hotels).

• checkInDate: Date for check-in (for hotels).

• checkOutDate: Date for check-out (for hotels).

• offers: Offers associated with the destination.

• guideFeaturedInCopy: Information about guides featuring the destination.

• isClosed: Indicates if the destination is closed.

• isLongClosed: Indicates if the destination has been closed for a long time.

• openNowText: Text indicating if the item is currently open.

• cuisines: Cuisines offered (for restaurants).

• mealTypes: Types of meals available (for restaurants).

• dishes: Dishes served (for restaurants).

• features: Features or highlights of the destination.

• dietaryRestrictions: Dietary restrictions or considerations.

• hours: Operating hours of the destination.

• menuWebUrl: URL for the menu (for restaurants).

• establishmentTypes: Types of establishments.

• ownersTopReasons: Top reasons provided by owners.

• rentalDescriptions: Descriptions related to rentals.

• photos: Photos associated with the item.

• bedroomInfo: Information about bedrooms (for accommodations).

• bathroomInfo: Information about bathrooms (for accommodations).

• bathCount: Number of bathrooms (for accommodations).

• baseDailyRate: Base daily rate (for accommodations).

**PROJECT GOALS**

* Build a machine learning model that can accurately predict hotel ratings based on customer reviews, budget constraints, specific locations, and the type of residence.
* Establish evaluation metrics to assess the performance of the recommendation system.
* Deploy and Implement a real-time recommendation feature that can adapt to users' changing preferences and provide up-to-date tourist destination suggestions.

**SUCCESS METRIC**

Root Mean Squared Error of close to 0 to evaluate model efficiency.

**METHODS USED**

* Descriptive Statistics
* Data Visualization
* Machine Learning

**DATA UNDERSTANDING**

* Data Scraping: AfricuraAI employed web scraping techniques to collect relevant data from TripAdvisor. Web scraping involved extracting information from web pages in an automated manner. Tools like BeautifulSoup or Selenium were used by AfricuraAI to scrape data from TripAdvisor's website.
* Data Types: The data scraped from TripAdvisor included a variety of information that was useful for understanding tourist preferences and destinations in Africa. This included:
* Reviews and Ratings: AfricuraAI collected textual reviews and ratings given by users who had visited different tourist destinations in Africa. These reviews provided insights into the experiences, opinions, and satisfaction levels of the tourists.
* Destination Information: The system extracted details about various destinations in Africa, such as their names, locations, attractions, amenities, accommodation options, and more. This information helped in building a comprehensive database of African tourist destinations.
* User Profiles: AfricuraAI also collected information about the users who provided reviews on TripAdvisor. This included demographic data, travel preferences, previous destinations visited, and other relevant details. User profiles helped in understanding the preferences and characteristics of different types of tourists.

**DATA CLEANING**

During the data preprocessing phase, we performed data cleaning to ensure the collected data from TripAdvisor was in a suitable format for analysis. The data cleaning process involved several steps:

* 1. **Removing Irrelevant Data**: Any irrelevant or unnecessary data that was not relevant to the recommendation system, such as advertisements or unrelated information, was removed from the dataset.
  2. **Handling Missing Values**: If the scraped data contained missing values, AfricuraAI applied strategies to handle them. We removed some rows with missing values, filling others with appropriate values, such as using mean imputation for numerical attributes or mode imputation for categorical attributes.
  3. **Deduplicating Data**: Duplicate entries in the dataset were identified and removed to ensure that each record was unique. This helped in preventing bias and errors during analysis.
  4. **Normalizing Text**: Textual data, such as reviews and destination names, may contain variations in formatting, punctuation, or case sensitivity. We applied techniques like text normalization, removing special characters, and converting text to lowercase to ensure consistency and improve analysis accuracy.
  5. **Filtering Outliers**: Outliers, which are extreme values that deviate significantly from the majority of the data, were identified and handled appropriately.
  6. **Standardizing Data Formats**: Data collected from different sources or in different formats were standardized to a common format. This included ensuring consistent date formats, numerical representations, and categorical labels for better analysis.
  7. **Addressing Data Integrity Issues**: In some cases, the scraped data might have inconsistencies or integrity issues. We performed checks and corrections to address any inconsistencies in the data to maintain data quality and reliability.

By applying these data cleaning techniques, we ensured that the dataset used for analysis was accurate, consistent, and suitable for building the recommendation system. This enhanced the reliability and effectiveness of the subsequent data analysis and modeling stages

**EXPLORATORY DATA ANALYSIS**

Once the data was preprocessed, we analyzed it to gain insights and identify patterns. This involved various techniques such as:

Sentiment Analysis: The system used natural language processing techniques to analyze the sentiment expressed in the reviews. This helped in understanding the positive or negative experiences shared by tourists.

Topic Modeling: We employed topic modeling algorithms, such as Latent Dirichlet Allocation (LDA), to identify the main topics or themes discussed in the reviews. This helped in categorizing the destinations based on common attributes or experiences.

Statistical Analysis: The scraped data was analyzed statistically to identify correlations, trends, and patterns. This analysis helped in understanding the relationships between different variables and features, aiding in the recommendation process.

**MODELING**

In this section, we built classification models using

## NMF

Memory Based

SVD

## KNNWithMean

## Baseline Model

## KNNBasic

**CHALLENGES.**

While developing the recommendation system for tourist destinations in Africa, we encountered several challenges. These challenges were addressed to ensure the accuracy and reliability of the system. Some of the key challenges faced include:

**Data Quality and Reliability:** The quality and reliability of the data scraped from TripAdvisor posed a significant challenge. The data collected from user reviews and ratings could be subjective and biased. AfricuraAI had to implement techniques to filter out unreliable or fake reviews and ensure that the dataset used for analysis was of high quality.

**Data Volume and Scalability**: TripAdvisor contains a vast amount of data, including reviews, ratings, and destination information. Handling and processing such a large volume of data required efficient storage, computational resources, and scalable algorithms. We had to optimize data processing techniques to handle the large dataset effectively.

**Data Privacy and Legal Compliance**: Scraping data from online platforms like TripAdvisor raises concerns regarding data privacy and legal compliance. We had to ensure that the data scraping process adhered to legal guidelines and respected user privacy rights. This involved obtaining necessary permissions and anonymizing personal information.

**Variability in User Preferences**: Tourists have diverse preferences when it comes to selecting travel destinations. We had to develop a recommendation system that could handle the variability and subjectivity of user preferences. This required implementing advanced machine learning techniques, such as collaborative filtering or content-based filtering, to capture and accommodate diverse user preferences.

**Data Noise and Inconsistencies**: The data scraped from TripAdvisor might contain noise, inconsistencies, or inaccuracies. This could include misspelled words, ambiguous language, or incomplete information. We had to employ data cleaning techniques and perform thorough preprocessing to address these issues and ensure data accuracy.

**Dynamic Nature of the Data**: TripAdvisor's data, including reviews and ratings, can change over time as new reviews are added or existing ones are updated. We needed to account for the dynamic nature of the data by regularly updating and retraining the recommendation system to incorporate the latest information.

**Handling Cold Start Problem**: The recommendation system faced the challenge of providing accurate recommendations for new users who had limited or no historical data. We had to develop strategies to address the cold start problem by leveraging information such as location preferences, budget constraints, and user demographics to provide initial recommendations.

**RECOMMENDATIONS**

In summary, our recommendation system

1. Offers personalized destination recommendations.
2. Identifies top tourist destinations.
3. Fosters customer loyalty and engagement, and continually improves through user feedback.
4. By leveraging advanced algorithms and user data, we aim to provide the best travel recommendations tailored to each user's preferences and budget constraints, ultimately enhancing their overall travel experience in Africa.

**CONCLUSIONS**

Overall, the recommendation system has brought significant benefits to both users and the business.

It has streamlined the travel planning process, enhanced personalization in travel experiences, facilitated informed decision-making, and opened up business opportunities within the travel industry.

The system has successfully addressed the challenges in finding suitable travel destinations and has the potential to continue improving and expanding its services in the future.

**APPENDIX**

By scraping data from TripAdvisor and analyzing it, we gained valuable insights into tourist preferences, experiences, and destination characteristics. This enabled the development of a powerful recommendation system that suggested personalized tourist destinations in Africa based on the scraped data.