



DATATHON CHALLENGE

The Challenge

As part of our commitment to enhancing the experience for tourists visiting Sri Lanka, we are introducing a new personalized recommendation feature. Following the visa process, we will conduct an experimental survey to gain insights into each tourist's preferences and travel desires. This allows us to suggest destinations with desired activities and experiences that align with their expectations, creating a truly unforgettable experience.

To provide personalized travel destination recommendations, which will be included in their visa confirmation email, we collect the following details from tourists:

- Name
- Email Address
- Preferred Activities
- Bucket List Destinations in Sri Lanka

You have been provided with two datasets for this purpose:

- The first dataset ([Visitors Preference Dataset](#)) consists of data of the users such as **User ID, Name, Email Address, Preferred Activities, and Bucket List Destinations in Sri Lanka**. This dataset provides information about activities that the visitor wishes to do and places they want to visit in Sri Lanka. Click [here](#) to access it.
- The second dataset ([Places Dataset](#)) contains details about various locations in Sri Lanka. This includes **name, latitudes, longitudes, formatted address, ratings, total number of ratings, and latest reviews** of the places. Click [here](#) to access it.

Your team will be responsible for devising a recommendation engine that utilizes the provided data to recommend places to users. Your solution should take a user's details as input and return the top 5 recommended places as output to that user.

Given that these recommendations are issued with visa approval, it is crucial that the suggested destinations meet a high standard of quality and relevance. The recommendations should not only align with the tourist's preferences but also reflect recent and positive attributes of the destinations. This implies that the selection process should prioritize places that are well-regarded and have maintained a high level of satisfaction among recent visitors.



Data Enrichment Guidelines

You are welcome to use open data sources to enrich the second dataset ([Places Dataset](#)). However, it is important to remain mindful of the following:

- **Licenses and Permissions:** Ensure that any external data sources you use are legally accessible under their respective licenses. Always review and comply with the terms and conditions of the data provider before incorporating the data into your project.
- **Attribution:** Proper attribution must be given to the original data source. This includes citing the source in your project or acknowledging the authors as specified in the license agreement.
- **Data Usage Restrictions:** Some open data sources may impose restrictions on how the data can be used, modified, or shared. Make sure you adhere to these limitations to avoid violating any legal agreements.
- **Respect for Privacy and Confidentiality:** If using datasets that include personal or sensitive information, be sure to handle the data in accordance with privacy laws and ethical guidelines, ensuring confidentiality where required.

Evaluation

We understand that there are multiple approaches to building an effective recommendation system. To foster creativity and innovation, you are given the flexibility to define your own model evaluation metric within the guidelines provided below. This guideline encourages you to consider multiple aspects of recommendation quality.

Model Evaluation Aspects

We strongly encourage you to evaluate your models based on the following aspects:

- **Relevance:** How well do the recommended places align with the user's preferences?
- **Diversity:** Does your recommendation list offer a variety of options? A diverse set of recommendations ensures that users are exposed to a wide range of destinations, which can enhance their overall experience.

Rationale for Selection of Model Evaluation Metric/s

As part of your submission, please include a brief explanation of the model evaluation metric/s you developed/chose. Specifically, address the following:

- **Reasoning for metric/s selection:** Describe the reasoning behind your selection of metric/s and how they fit with your approach to building the recommendation system.
- **Relevance to the evaluation aspects:** Describe how your metric/s is related to the evaluation aspects of relevance and diversity

Grading of Model Evaluation Method

Your model evaluation method and the justification you provide will be graded as part of the competition. We will assess:

- **The thoughtfulness and appropriateness of your chosen metric/s:** We will look for a clear, logical explanation behind your selection of specific model evaluation metric/s. Your justification should be well-organized, concise, and explicitly outline how each metric contributes to evaluating the effectiveness of your recommendation model. A compelling rationale will demonstrate the connection between the chosen model evaluation metric/s and the goals of your model.
- **Coverage of Recommendation Quality Aspects:** We will assess the extent to which your model evaluation method effectively addresses the key aspects of recommendation quality, with particular focus on relevance and diversity. This includes evaluating how well your chosen metrics capture the ability of the model to deliver recommendations that closely match user preferences. A strong model evaluation method will demonstrate a balanced approach that ensures recommendations are both highly relevant and sufficiently diverse.

Deliverables

- **Architecture Diagrams:** Submit diagrams that illustrate the architecture of your recommendation model and any preprocessing pipelines. These should clearly convey the flow of data and the design choices you made in the architecture of your model.
- **Documentation:** Document the approaches you have used while analyzing the data, insights derived from the analysis, data enrichment process and pre-processing methods you have used, the evaluation metric/s you have selected, and the rationale and reasoning behind your selection. (*For more info, read the 'Evaluation' slide above.*)
- **Recommendation Model:** Provide the final trained model in .h5 or .pkl format.
- **Notebooks:** Submit all the notebooks used for experimentation, data enrichment process, and data preprocessing as .ipynb files. Please mark the final notebook with your best-performing model as TeamName_FinalNotebook.ipynb.
- **Demo Video:** Upload a 3 to 5-minute demo video on YouTube. This video should provide a walkthrough of your model architecture, data preprocessing steps used, the model evaluation techniques you used, the problems you encountered, and how you overcame them.

Please add your model architecture diagrams, documentation, recommendation model, and notebooks into one folder. Afterward, compress this folder into a zip format, ensuring it retains the name **TeamName_Datathon.zip**. Upload the zip file to the submission form.

Deadline for submissions: **14th September, 2024 at 11.59 PM**

🔗 **Submission Form:** <https://forms.gle/yK7SiMDHRsXsq7b86>



Rules and Regulations

- The submission form will stop accepting responses after the deadline.
- You can't use any pre-trained model on the final recommendation model. But you can use pre-trained models in the data preprocessing stage if needed.
- If the team is found to have cheated, plagiarized, or violated the rules of the competition, they will be disqualified
- By participating in this competition, teams agree to be bound by these rules and regulations and by the decisions of the judges, which are final and binding in all respects.

Judging Criteria

- Data Wrangling - 25%
- Model and Architecture Implementation - 35%
- Model Evaluation - 20%
- Demo Video - 20%



Terms and Conditions

1. Use of Data

You are permitted to use the provided datasets solely for the purpose of this competition. Any other use, including but not limited to commercial purposes, academic research, or personal projects, is strictly prohibited.

2. Data Sharing

The dataset must not be shared, distributed, or transmitted in any form, whether publicly or privately, with any third party. This includes uploading the dataset to external websites, forums, or social media platforms.

3. Publication and Disclosure

You are not allowed to publish, disclose, or otherwise make the dataset or any derivatives of the dataset publicly available unless it's authorized by the competition organizers.

4. Data Confidentiality

By participating in the competition, you agree to maintain the confidentiality of the dataset and any sensitive information contained within it.

5. Violation of Terms

Any violation of these terms and conditions may result in disqualification from the competition and may even lead to extreme cases such as legal action if deemed necessary.

Disclaimer

The information in the second dataset (*Places Dataset*), including reviews and comments, reflects the personal opinions of users. These views do not necessarily represent the opinions or views of the organizers (i.e Rootcode Pvt Ltd). Please note that the reviews may vary in accuracy and may touch upon sensitive or controversial topics.

Wish you all the best!



TECH-TRIATHLON

A competition by rootcode