🌍 Tourism Experience Analysis – Project Report

1️⃣ Project Objective

The aim of this project was to leverage tourism datasets to build predictive and analytical models that:

- 📊 **Predict Ratings (Regression Model)**
- 🏷 **Classify Tourist Satisfaction Levels (Classification Model)**
- 🎯 **Recommend Attractions (Recommendation System)**

💡 **All models were deployed into an interactive Streamlit web application for real-time access and usage.**

2️⃣ Data Collection & Merging

📂 **Datasets Used:**

- 🌍 **Continent, Region, Country, City – *Geographic hierarchy***
- 🧍 **User – *Tourist demographic & location data***
- 🧾 **Transaction – *Tourist visit records***
- 🚇 **Mode – *Travel mode information***
- 🏛 **Item & Updated_Item – *Attraction details***
- 🏷 **Type – *Attraction type mapping***

🔄 **Key Processing & Merging Steps:**

- **Hierarchical Merges:**
  Continent → Region → Country → City → User → Transaction
- **Merged attraction details, removed duplicates (AttractionId)**
- **Cleaned categorical fields (trim spaces, standardize case)**
- **Removed invalid ratings (outside 1–5 range)**
- **Capped outliers in ratings**
- **Added time-based features: VisitSeason, VisitQuarter**
- **Engineered user-level features: average past rating, visit count, rating trend**
- **Engineered attraction-level features: average past ratings, previous visits, visitor counts**
- **Added city-level popularity metrics**

- Created interaction features:
    - user_continent
    - user_attraction_type
    - attraction_type_season

📁 Final Output:
Merged dataset saved as Final_dataset.csv for model training.

3️⃣ Regression Model – Predicting Ratings

- Algorithm: 🌲 XGB Regressor, Gradient boost regressor and ensemble model
- Features: Mix of numerical, categorical, and engineered features
- Metrics Used:
    - 📈 R^2 Score
    - 📉 Mean Absolute Error (MAE)
    - 📉 Root Mean Square Error (RMSE)

📊 Key Insights:

- Achieved high R^2 → strong predictive accuracy
- Top Predictive Features:
    1. user_avg_rating_before
    2. city_popularity
    3. VisitSeason

💡 Visualization:

- Feature Importance Graph showing historical user behavior & attraction popularity as the biggest influences.

4️⃣ Classification Model – Tourist Satisfaction

- Objective: Predict if a tourist is Satisfied or Not Satisfied
- Label Rule:
    - Rating ≥ 4 → ✅ Satisfied
    - Rating < 4 → ❌ Not Satisfied
- Algorithm: 🌲 XGB Classifier
- Metric: Accuracy

📊 **Key Insights:**

- **High accuracy achieved**
- **Top Predictors:**
    1. **VisitSeason**
    2. **attraction_avg_rating_before**
    3. **user_continent**

💡 **Visualization:**

- **Feature Importance Graph highlighting seasonal trends & attraction type influence.**

5️⃣ **Recommendation System**

- **Approach: 🤝 Collaborative Filtering using Cosine Similarity**
- **Functionality:**
    - **Input an attraction name**
    - **Output Top N similar attractions based on user-item interaction patterns**

6️⃣ **Streamlit Web Application**

📌 **Features:**

- **Sidebar Navigation: Choose Regression, Classification, or Recommendation**
- **UI Enhancements: Color themes, headers, responsive layout**
- **Interactive Inputs:**
    - **Dropdowns for categorical features**
    - **Sliders for numerical inputs**
- **Visual Outputs:**
    - **Feature importance bar charts**
    - **Metrics:**
        - **R^2 for regression**
        - **Accuracy for classification**
    - **Recommendation results in table format**

⚡ **Error Handling:**

- **Unique keys for widgets (avoiding duplicate errors)**

- Dataset load validation

7️⃣ Key Insights

- 🗓️ Seasonality heavily affects satisfaction & ratings
- 📜 User & attraction history predict future ratings well
- 🏙️ City popularity is a strong influencing factor
- 🤝 Recommendation system aligns with tourist preferences

8️⃣ Actionable Insights

- 🗓️ Promote attractions during off-peak seasons to spread demand
- 🎯 Create personalized marketing campaigns based on user history
- 📦 Bundle popular cities & attractions for premium tour packages