Project Title	Tourism Experience Analytics: Classification, Prediction, and Recommendation System
Skills take away From This Project	Data Cleaning and Preprocessing, Exploratory Data Analysis (EDA), Data Visualization ,SQL, Streamlit, Machine Learning (Regression, Classification & Recommendation).
Domain	Tourism

Problem Statement:

Tourism agencies and travel platforms aim to enhance user experiences by leveraging data to provide personalized recommendations, predict user satisfaction, and classify potential user behavior. This project involves analyzing user preferences, travel patterns, and attraction features to achieve three primary objectives: regression, classification, and recommendation.

Business Use Cases:

1. **Personalized Recommendations:** Suggest attractions based on users' past visits, preferences, and demographic data, improving user experience.
2. **Tourism Analytics:** Provide insights into popular attractions and regions, enabling tourism businesses to adjust their offerings accordingly.
3. **Customer Segmentation:** Classify users into segments based on their travel behavior, allowing for targeted promotions.
4. **Increasing Customer Retention:** By offering personalized recommendations, businesses can boost customer loyalty and retention.

Objective:

1. Regression: Predicting Attraction Ratings

Aim:

Develop a regression model to predict the rating a user might give to a tourist attraction based on historical data, user demographics, and attraction features.

Use Case:

- Travel platforms can use this model to estimate the satisfaction level of users visiting specific attractions. By identifying attractions likely to receive lower ratings, agencies can take corrective actions, such as improving services or better setting user expectations.
- Personal travel guides can provide users with attractions most aligned with their preferences to enhance overall satisfaction.

Possible Inputs (Features):

- User demographics: Continent, region, country, city.
- Visit details: Year, month, mode of visit (e.g., business, family, friends).
- Attraction attributes: Type (e.g., beaches, ruins), location, and previous average ratings.

Target:

- Predicted rating (on a scale, e.g., 1-5).

2. Classification: User Visit Mode Prediction

Aim:

Create a classification model to predict the mode of visit (e.g., business, family, couples, friends) based on user and attraction data.

Use Case:

- Travel platforms can use this model to tailor marketing campaigns. For instance, if a user is predicted to travel with family, family-friendly packages can be promoted.
- Hotels and attraction organizers can plan resources (e.g., amenities) better based on predicted visitor types.

Inputs (Features):

- User demographics: Continent, region, country, city.
- Attraction characteristics: Type, popularity, previous visitor demographics.
- Historical visit data: Month, year, previous visit modes.

Target:

- Visit mode (categories: Business, Family, Couples, Friends, etc.).

3. Recommendations: Personalized Attraction Suggestions

Objective:

Develop a recommendation system to suggest tourist attractions based on a user's historical preferences and similar users' preferences.

Use Case:

- Travel platforms can implement this system to guide users toward attractions they are most likely to enjoy, increasing user engagement and satisfaction.
- Destination management organizations can identify emerging trends and promote attractions that align with user preferences.

Types of Recommendation Approaches:

1. **Collaborative Filtering:**
   - Recommend attractions based on similar users' ratings and preferences.
2. **Content-Based Filtering:**
   - Suggest attractions similar to those already visited by the user based on features like attraction type, location, and amenities.
3. **Hybrid Systems: (Optional)**
   - Combine collaborative and content-based methods for enhanced accuracy.

Inputs (Features):

- User visit history: Attractions visited, ratings given.
- Attraction features: Type, location, popularity.
- Similar user data: Travel patterns and preferences.

Output:

- Ranked list of recommended attractions.

Approach:

1. **Data Cleaning:**
   - Handle missing values in the transaction, user, and city datasets.
   - Resolve discrepancies in city names or other categorical variables like VisitMode, AttractionTypeId, etc.
   - Standardize date and time format, ensuring consistency across data.
   - Handle outliers or any incorrect entries in rating or other columns.
2. **Preprocessing:**
   - **Feature Engineering:**
     - Encode categorical variables such as `VisitMode`, `Continent`, `Country`, and `AttractionTypeId`.
     - Aggregate user-level features to represent each user's profile (e.g., average ratings per visit mode).
     - Join relevant data from transaction, user, city, and attraction tables to create a consolidated dataset.
   - **Normalization:** Scale numerical features such as `Rating` for better model convergence.
3. **Exploratory Data Analysis (EDA):**
   - Visualize user distribution across continents, countries, and regions.
   - Explore attraction types and their popularity based on user ratings.
   - Investigate correlation between `VisitMode` and user demographics to identify patterns.
   - Analyze distribution of ratings across different attractions and regions.
4. **Model Training:**
   - **Regression Task:**
     - Train a model to predict `ratings` based on user, attractions, transaction features, etc.
   - **Classification Task:**
     - Train a classifier (e.g., Random Forest, LightGBM, or XGBoost) to predict `VisitMode` based on user and transaction features.
   - **Recommendation Task:**
     - Implement collaborative filtering (using user-item matrix) to recommend attractions based on user ratings and preferences.
     - Alternatively, use content-based filtering based on attractions' attributes (e.g., location, type).
5. **Model Evaluation:**
   - Evaluate classification model performance using accuracy, precision, recall, and F1-score.
   - Evaluate regression model using R2, MSE, etc.

- - Assess recommendation system accuracy using metrics like Mean Average Precision (MAP) or Root Mean Squared Error (RMSE).
  6. **Deployment:**
     - Build a Streamlit app that allows users to input their details (location, preferred visit mode) and receive:
       - A prediction of their visit mode (Business, Family, etc.).
       - Recommended attractions based on their profile and transaction history.
     - Display visualizations of popular attractions, top regions, and user segments in the app.

End Output:
A user-friendly Streamlit application where tourists can input their data and receive personalized recommendations for attractions, as well as get predictions on their likely visit mode.

Expected Results:

1. **Analyze Trends:** Explore trends in tourism experiences by user demographics, visit modes, and popular attractions across different regions and cities.
2. **Predictive Modeling:** Develop models to predict the most likely visit mode (e.g., Business, Family, Solo) for users based on their profiles and past interactions.
3. **Actionable Insights:** Provide recommendations for personalized attraction suggestions and targeted marketing strategies to enhance customer experience.
4. **Streamlit Application:** Develop a Streamlit application to showcase predicted visit modes and personalized attraction recommendations based on user inputs.tion to showcase the predicted value based on selected features.

Project Evaluation metrics:

- Data Preparation
  - Completeness and accuracy of data cleaning.
  - Clear documentation of preprocessing steps.
- Exploratory Data Analysis
  - Depth and clarity of insights derived.
  - Use of relevant statistical methods.
- Data Visualization
  - Relevance and quality of visualizations.

- ○ Effective use of charts, maps, and interactivity.
- ● Model Performance
  - ○ Metrics involving accuracy, f1, recall, etc..
  - ○ Give comparison of different models.
- ● Business Insights
  - ○ Actionable and stakeholder-focused findings.
  - ○ Addressing key use cases like hotspots and trends.
- ● Presentation
  - ○ Logical structure and clarity.
  - ○ Effective communication of results.

Technical Tags:

- ● Data Cleaning
- ● Data Preprocessing
- ● Exploratory Data Analysis (EDA)
- ● Machine Learning
- ● Streamlit

Data Set:

Dataset: 🖻 Tourism Dataset

Data Set Explanation:

1. **Domain Code & Domain:**

1. Transaction Data:

- ● **Purpose:** Contains information on user visits to various attractions, including ratings and visit details.
- ● **Columns:**
  - ○ `TransactionId`: Unique identifier for each transaction.
  - ○ `UserId`: Identifier for the user making the transaction.
  - ○ `VisitYear`: The year the visit occurred.
  - ○ `VisitMonth`: The month the visit occurred.
  - ○ `VisitMode`: The mode of visit (Business, Couples, Family, etc.).
  - ○ `AttractionId`: Unique identifier for the visited attraction.
  - ○ `Rating`: The user's rating for the attraction.

- **Usage:** This data will be used to understand user behavior, predict visit modes, and recommend attractions based on user preferences and ratings.

2. User Data:

- **Purpose:** Contains information about the users, such as their geographical location (continent, region, country, city).
- **Columns:**
  - `UserId`: Unique identifier for each user.
  - `ContinentId`: The continent to which the user belongs.
  - `RegionId`: The region within the continent.
  - `CountryId`: The country in which the user resides.
  - `CityId`: The city of residence for the user.
- **Usage:** This data will help in analyzing the user demographics, which can be used to predict their behavior, visit modes, and attractions they might prefer.

3. City Data:

- **Purpose:** Contains information about different cities, used to link users and attractions to specific geographical locations.
- **Columns:**
  - `CityId`: Unique identifier for each city.
  - `CityName`: Name of the city.
  - `CountryId`: The country associated with the city.
- **Usage:** This data helps to relate the `CityId` in the user and attraction data to a specific city name and country.

4. Type:

- **Purpose:** Contains details about type of tourist attractions.
- **Columns:**
  - `AttractionTypeId`: The type of the attraction (e.g., Beach, Museum, Park).
  - `AttractionType`: Type of the attraction.

5. Visit Mode Data:

- **Purpose:** Contains information about the types of visit modes users may have, such as business trips, family vacations, or solo travel.
- **Columns:**
  - `VisitModeId`: Unique identifier for the visit mode.
  - `VisitMode`: Name or description of the visit mode (e.g., Business, Couples, Family).
- **Usage:** This data will be used to predict the user's likely visit mode based on their transaction history and demographic information.

6. Continent Data:

- **Purpose:** Contains information about the continents, helping to link users to their respective continents.
- **Columns:**
  - `ContinentId`: Unique identifier for each continent.
  - `Continent`: Name of the continent (e.g., Africa, Asia, Europe).
- **Usage:** Helps associate users with their continent, which may influence their travel behavior and preferences.

7. Country Data:

- **Purpose:** Contains information about countries, linking users and attractions to specific countries.
- **Columns:**
  - `CountryId`: Unique identifier for each country.
  - `Country`: Name of the country.
  - `RegionId`: The region within the country.
- **Usage:** Helps in understanding user preferences and travel behavior in relation to different countries.

8. Region Data:

- **Purpose:** Contains information about regions within countries.
- **Columns:**
  - `RegionId`: Unique identifier for each region.
  - `Region`: Name or description of the region (e.g., North America, East Africa).
  - `ContinentId`: Unique identifier for each continent.
- **Usage:** This data helps classify users and attractions based on regional preferences and trends.

9. Item Data:

- **Purpose:** Contains information about regions within countries.
- **Columns:**
  - `AttractionId`: Unique identifier for each attraction, linking it to other datasets.
  - `AttractionCityId`: City where the attraction is located, connecting it to the City Data.
  - `AttractionTypeId`: Category of the attraction (e.g., beach, historical site, park) for personalized recommendations.
  - `Attraction`: Name of the attraction used to identify and suggest places.
  - `AttractionAddress`: Physical address, useful for mapping, distance calculations, and travel planning.

- **Usage:** This data will be used to recommend attractions based on user preferences, location, and visit modes.

Project Deliverables:

Cleaned Dataset:

- Final preprocessed dataset used for analysis.
- Description of cleaning steps (e.g., handling missing values, formatting).

Source Code:

- Python or other scripts used for data cleaning, analysis, model building and visualization.

Application:

- Streamlit Application for recommendation/ classification

Documentation:

- A concise report explaining the approach, key findings, and actionable insights.
- Include explanations for visualizations, trends, and analyses performed.

Project Guidelines:

Follow the steps mentioned:

1) **Data Preparation**: Cleaning → Transforming → Feature Engineering → Encoding.
2) **Exploratory Data Analysis (EDA)**: Understanding distributions → Correlations → Relationships.
3) **Visualization**: Plot trends → Identify outliers → Spot patterns.
4) **Model Building**: Train → Validate → Test models.
5) **Evaluation & Insights**: Evaluate performance → Interpret results → Derive actionable insights

Code Quality:

- Write clean, well-commented, and reusable code.

Presentation:

- Keep the final presentation clear and concise.
- Emphasize key findings and actionable recommendations.

Timeline:

The project must be completed and submitted **within 7 days from the assigned date**.

References:

Project Orientation Recording (English)	📄 ML_ Project Session Recordings(Englis…
Project Live Evaluation	📄 Project Live Evaluation
EDA Guide	📄 Exploratory Data Analysis (EDA) Guide
Streamlit Reference	https://docs.streamlit.io/get-started/fundamentals/main-concepts
Capstone Explanation Guideline	📄 Capstone Explanation Guideline
GitHub Reference	🅿 How to Use GitHub.pptx

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.
Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday to Saturday (4:00PM to 5:00PM)

Booking link :https://forms.gle/XC553oSbMJ2Gcfug9

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.
Note: This form will Open on Saturday and Sunday Only on Every Week

Timing: Monday-Saturday (5:30PM to 6:30PM)

Booking link : https://forms.gle/1m2Gsro41fLtZurRA