

❖ Tourism Experience Analytics – Project Workflow



Project Overview

This project demonstrates an end-to-end **Data Science workflow** implemented in a **Python Jupyter Notebook**, covering data preparation, exploration, visualization, and machine learning modeling.

The goal is to transform raw data into meaningful insights and build predictive and recommendation models.



Data Cleaning & Preprocessing

The dataset was prepared to ensure quality and consistency before analysis. Steps included:

- Handling missing values using imputation and filtering techniques
- Removing duplicate records
- Correcting data types and formatting issues
- Treating outliers where necessary
- Encoding categorical variables for modelling
- Feature scaling/normalization for machine learning algorithms

This step ensured the dataset was reliable and suitable for further analysis.



Exploratory Data Analysis (EDA)

EDA was performed to understand patterns, trends, and relationships within the tourism dataset.

Key analysis included:

- Distribution of user ratings
- Popular attractions and visit patterns
- Seasonal trends in visits
- Relationship between users, attractions, and ratings
- Correlation analysis between numerical features

EDA helped identify useful features and guided model selection.



Data Visualization

Visualizations were created to make insights easier to interpret:

- Rating distribution plots
- Attraction popularity charts
- Visit trends over time
- Heatmaps for correlation analysis
- User behavior visualizations

Libraries used: **Pandas, Matplotlib, Seaborn**



Machine Learning Models

◆ Regression Task

A regression model was trained to **predict user ratings** based on:

- User features
- Attraction characteristics
- Transaction/visit information

Models were evaluated using:

- R² Score
 - Mean Squared Error (MSE)
 - Root Mean Squared Error (RMSE)
-

◆ Classification Task

A classification model was trained to **predict VisitMode** using user and transaction features.

Algorithms explored include:

- Random Forest
- LightGBM
- XGBoost

Evaluation metrics used:

- Accuracy
- Precision
- Recall
- F1-Score

◆ Recommendation System

Two approaches were implemented:

✓ *Collaborative Filtering*

- Built using a **user-item interaction matrix**
- Recommends attractions based on similar user preferences and ratings

✓ *Content-Based Filtering*

- Uses attraction attributes such as
 - Location
 - Category / type
 - Other metadata
- Suggests attractions similar to those previously liked by the user

Evaluation metrics:

- Mean Average Precision (MAP)
 - Root Mean Squared Error (RMSE)
-

🛠️ Tools & Libraries

- Python
 - Jupyter Notebook
 - Pandas & NumPy
 - Matplotlib & Seaborn
 - Scikit-learn
-

🚀 Outcome

This project demonstrates the full lifecycle of a data science solution — from raw data cleaning to predictive modelling and recommendation generation — showcasing practical skills in data analysis, visualization, and machine learning.
