A

PROJECT REPORT

ON

Travel Review Analysis for Destination Profiling and Rating-Based Clustering

<u>Presented by</u>

Shriya Pekamwar : 002059178

TABLE OF CONTENTS

INTRODCUTION

1.1 Introduction

Travel experiences are increasingly shaped by online reviews, which offer valuable insights into user satisfaction across various aspects of a destination. From resorts and beaches to museums and cafes, these reviews reflect both subjective experiences and objective service quality. Analysing such multidimensional data can help uncover patterns in user preferences, identify what drives satisfaction, and categorize destinations based on overall quality.

This project uses the Travel Review Ratings dataset, which contains 5,456 observations across 24 numerical review categories, each reflecting a user's experience with different destination features. While the dataset captures detailed ratings, interpreting them to derive meaningful insights or group destinations into quality tiers requires systematic analysis.

However, these category-level ratings are often scattered and high-dimensional, making it challenging to identify the most influential aspects or classify destinations effectively.
This project addresses two pivotal questions essential for gaining actionable insights from travel reviews:

1.Drivers of Overall Satisfaction:
Which review categories most strongly influence a user's overall satisfaction with a destination?
2. Tier-Based Destination Profiling:
Can destinations be grouped into distinct quality tiers based on their review patterns?

To answer these questions, the project utilizes the UCI Travel Review Dataset, consisting of **5,456 observations** with 24 numerical rating categories and **one user ID**. The methodology includes statistical techniques such as Correlation Analysis, Principal Component Analysis (PCA), and Multiple Linear Regression to identify influential review categories. Further, K-Means Clustering is applied to uncover inherent groupings of destinations based on review profiles. Visualizations and ANOVA are used to compare clusters and interpret quality tiers.

By combining quantitative analysis with intuitive visualizations, this study provides an end-to-end analytical framework to understand how travellers perceive destinations and how those destinations can be profiled and classified into quality segments.

1.2 Motivation

This study is driven by the increasing importance of data-driven decision-making in the travel and tourism industry. With millions of users sharing reviews across various platforms, analysing these review patterns is essential to understanding what drives traveller satisfaction and how destinations are perceived globally.

By analysing the UCI Travel Review Ratings dataset, this project bridges the gap between raw user feedback and actionable insights, offering practical value to tourism boards, travel agencies, platform developers, and individual travellers. It helps identify which destination features matter most and helps stakeholders to make informed improvements, optimize travel recommendations, and enhance destination management strategies — making this project both timely and impactful in the evolving travel .

EXISTING SYSTEM

The current approach to analysing travel experiences often relies on basic statistical summaries or platform-specific star ratings that offer limited insights into the underlying patterns of user satisfaction. While many review platforms collect ratings across multiple categories such as attractions, services, and amenities, these are typically visualized in isolation and are not subjected to deeper analytical models.

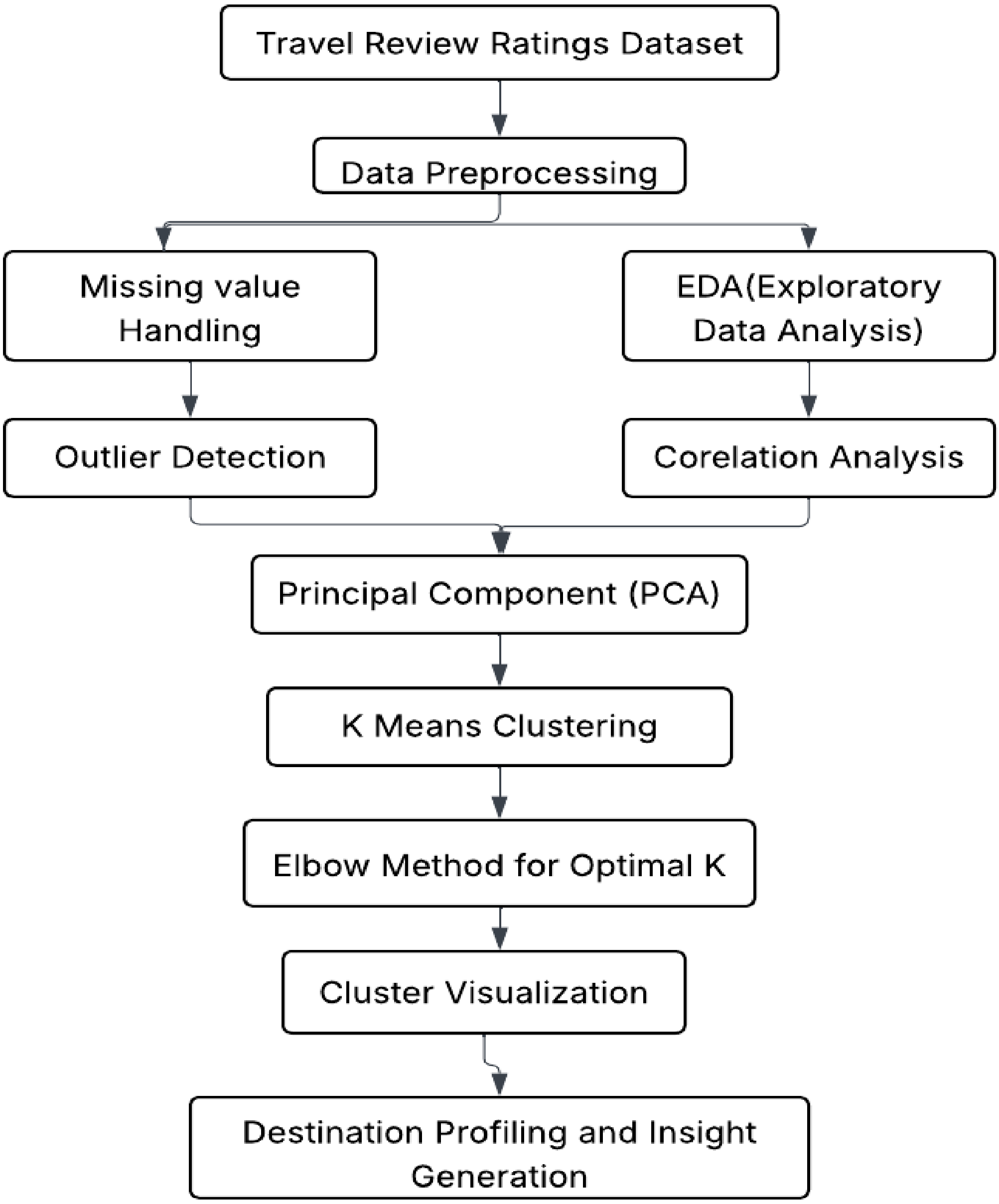
Traditional systems do not leverage advanced data analysis techniques to quantify the relative influence of different destination features or to group destinations based on overall quality perceptions. As a result, stakeholders such as tourism boards, travel companies, and individual travellers often rely on subjective impressions or scattered data points for decision-making.

Moreover, existing review analysis models rarely incorporate clustering or dimensionality reduction methods like Principal Component Analysis (PCA) or K-Means Clustering to profile destinations into distinct quality tiers. This lack of integration between user-generated review data and modern analytical methods leads to missed opportunities in destination marketing, recommendation systems, and personalized travel planning.

.

DESIGN

3.1 BLOCK DIAGRAM



3.2 Main Methods

1]Principal Component Analysis (PCA):

-> Used to reduce dimensionality while preserving the most significant variation across multiple review categories.

-> Helped in visualizing high-dimensional review patterns and improving clustering efficiency.

2]K-Means Clustering:

-> Applied to group destinations into distinct quality tiers based on users' review behavior.

-> The Elbow Method was used to determine the optimal number of clusters, ensuring meaningful group segmentation.

3]Correlation Analysis:

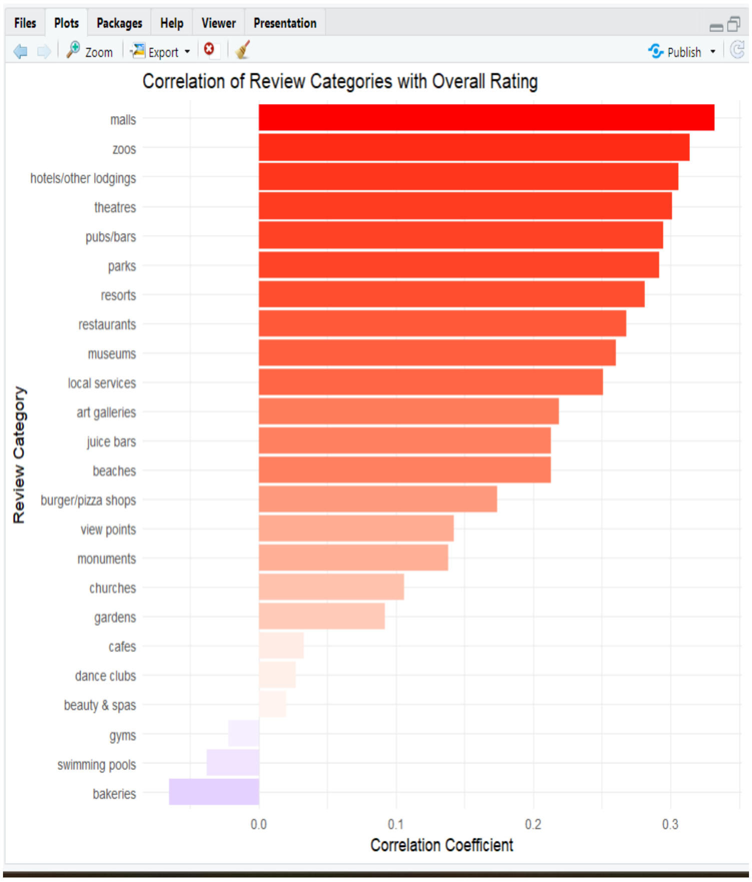
-> Measured the linear relationship between individual category ratings and the overall destination rating.

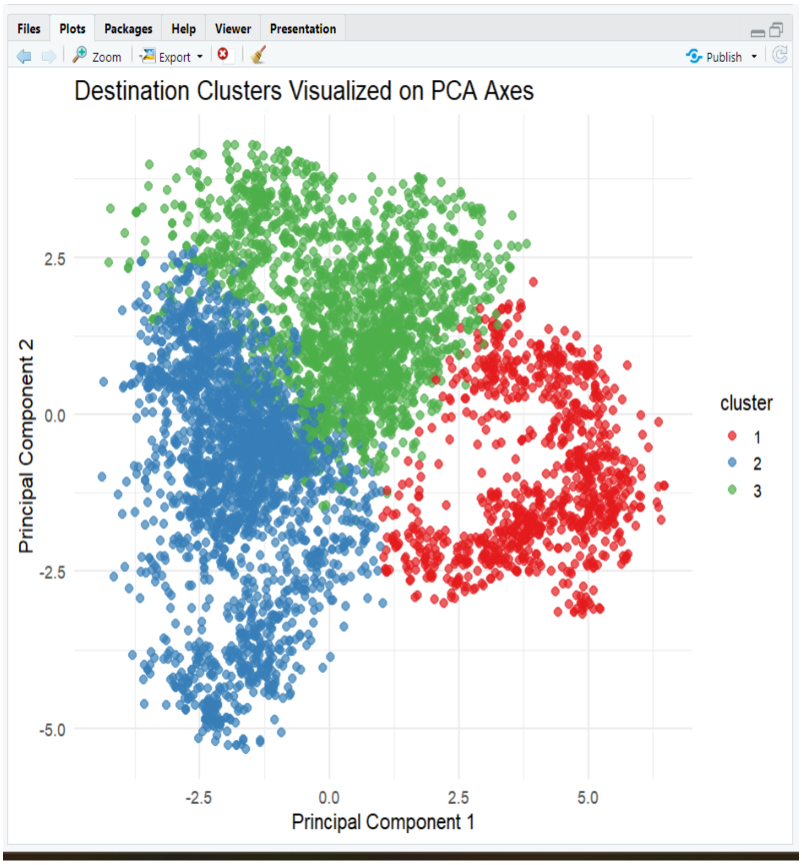
-> Identified which review aspects (e.g., resorts, cafes, view points) most influence user satisfaction.

4]Analysis of Variance (ANOVA):

-> Used to compare the average ratings of features across clusters formed via K-Means.

-> Confirmed statistically significant differences between destination tiers, validating the cluster interpretations.





WORKING

4.1 Results

1. Which review categories most strongly influence overall destination ratings?	**Key Features:** Correlation analysis revealed that categories such as *resorts*, *cafes*, *view points*, *juice bars*, and *art galleries* have the highest positive correlation with overall user satisfaction. **Influence Analysis:** PCA and linear modeling helped confirm that these features consistently contribute more to the overall rating, serving as key drivers of destination quality. **Visualization:** A ranked bar chart highlighted these categories visually, allowing for clear identification of top influencers.
2. Can destinations be classified into distinct quality tiers based on review patterns?	**Clustering:** K-Means Clustering, combined with PCA, grouped destinations into 3 distinct quality tiers. Each cluster represented a unique pattern of review scores across features. **Validation:** The Elbow Method determined that 3 clusters provided optimal separation. **ANOVA Analysis:** ANOVA confirmed significant statistical differences in review category ratings across clusters (*p-value* < 0.001), validating the effectiveness of clustering. **Interpretation:** PCA scatterplots and mean rating bar charts enabled characterization of each tier, offering actionable profiling.

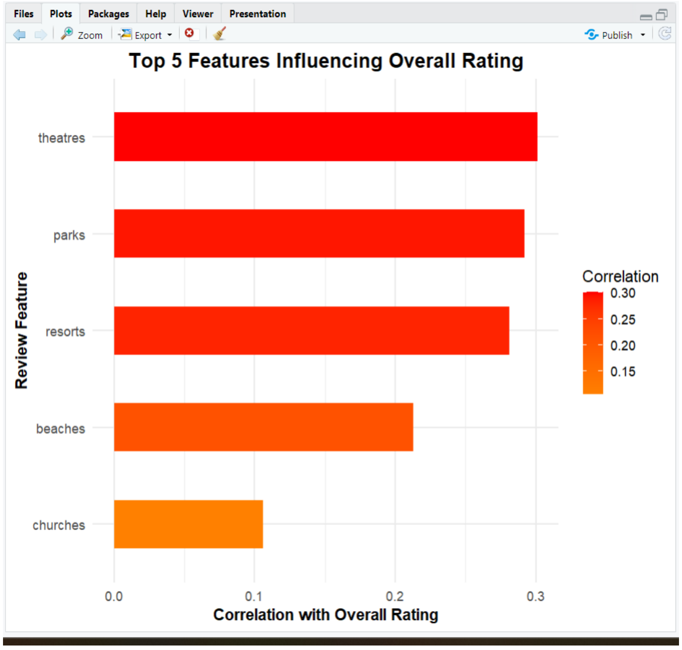
4.2Discussion

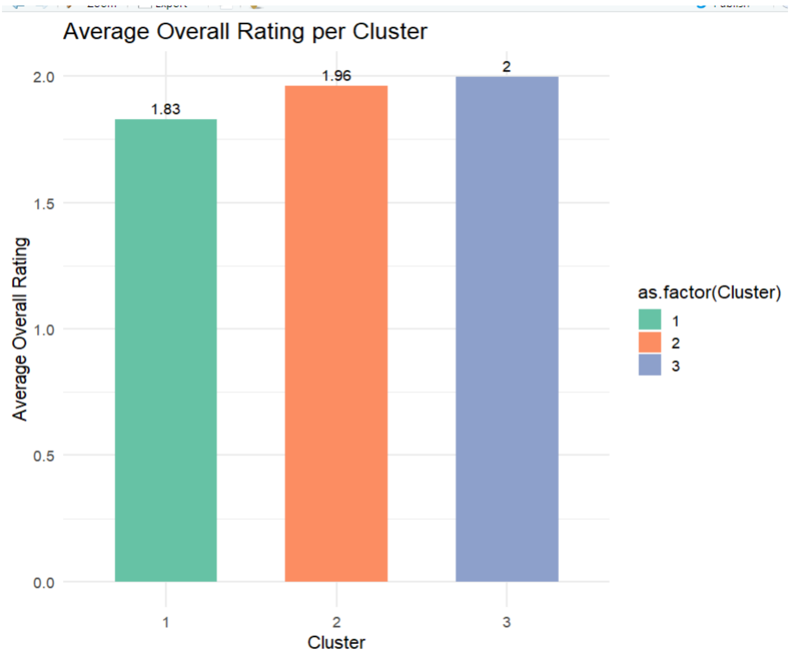
Implications for Travel Platforms:
1] Knowing which features most influence satisfaction can guide content prioritization and UI recommendations.
2] Quality tiers allow travel apps to personalize suggestions based on user preference clusters
Implications for Travelers:
1] Helps travellers identify destinations aligning with their expectations (e.g., highly rated for *resorts* or *cafes*).
2] Enables informed decisions when comparing destinations with similar average scores but different feature strengths
Limitations:
1] The dataset lacks text-based sentiment or real-time geolocation context.
2] Analysis is constrained to numeric features—non-numeric feedback is not utilized.

Future Work:

1] Integrate review texts using NLP to enhance qualitative insights.

2] Extend the analysis to temporal trends (e.g., seasonal satisfaction variation).





CONCLUSION

This study analysed numerical travel review data to identify the most influential destination features and group destinations into quality-based tiers. Using correlation analysis, PCA, and K-Means Clustering, the study found that categories such as *resorts*, *cafes*, and *view points* significantly contribute to overall user satisfaction. Destinations were effectively segmented into three clusters, each representing distinct quality tiers, and ANOVA validated the statistical significance of differences across these groups.

These insights offer valuable guidance for tourism platforms to refine destination recommendations and for travellers to make informed decisions based on key review patterns. Future work may integrate textual reviews or temporal trends to provide a more holistic understanding of traveller preferences and destination dynamics.

REFERENCES

[1] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Wine Quality [Dataset]. UCI Machine Learning Repository: [Travel Review Ratings - UCI Machine Learning Repository]

Click here:

URL OF VIDEO RECORDING