King Saud University
College of Computer and Information Sciences
Department of Information Technology

جامعة الملك سعود
King Saud University
كلية علوم الحاسب والمعلومات
قسم تقنية المعلومات

Phase 3 - Logbook: Detailed Overview

Date of Collection: February 1-6, 2025

Data Collection Methods Overview:

1. **TripAdvisor and Booking.com (Feb 1 and Feb 4, 2025):** Utilized Apify for web scraping, focusing on hotel reviews across Saudi Arabia. Targeted data included customer reviews, ratings, and accommodation specifics.
2. **Ministry of Tourism Open Data Portal (Feb 2, 2025):** Direct download of tourist statistics including overnight stays and spending patterns.
3. **Survey Distribution (Feb 4-6, 2025):** Implemented through Google Forms, targeting both domestic and international tourists' preferences and satisfaction levels.

King Saud University
College of Computer and Information Sciences
Department of Information Technology

جامعة الملك سعود
King Saud University
كلية علوم الحاسب والمعلومات
قسم تقنية المعلومات

Modelling Task

Q1: What are the most popular tourist destinations in Saudi Arabia, and why?

Approach:

- Descriptive analysis (Survey + TripAdvisor)
- K-Means clustering
- Agglomerative clustering

Finding: Riyadh and Jeddah were the most visited and best-rated cities. Clustering grouped destinations into performance tiers, highlighting high-visibility cities and others with potential for growth.

Q2: Which seasons experience the highest tourist inflow?

Approach:

- Seasonal decomposition (time series)
- Linear regression on month numbers

Finding: Tourist inflow peaks during summer, followed by fall and winter. Seasonal decomposition clearly showed repeating patterns and trends, helping inform event timing and marketing strategies.

Q3: How do spending patterns differ between domestic and international tourists?

Approach:

- Descriptive bar chart by spending range
- Logistic regression to predict spending category

Finding: International tourists consistently spend more, particularly in the higher budget brackets. Domestic tourists tended to stay within the lower to mid-range.

Q4: What factors contribute most to tourist satisfaction and dissatisfaction?

Approach:

- Logistic regression (baseline)
- Random forest
- Support Vector Machine (SVM)

Finding: Trip purpose, spending range, season, and duration were key predictors. Logistic regression and SVM both performed well (80% accuracy), while Random Forest was slightly weaker.

Q5: What impact have large-scale projects like Riyadh Season and NEOM had on tourism trends?

Approach:

- Boxplot comparison of satisfaction
- Logistic regression using event attendance and influence score

Finding: Tourists who attended these events were significantly more likely to report high satisfaction. Events also played an important role in influencing travel decisions.

Focus Question Conclusion (Based on All Models): What are the key trends and factors driving tourism in Saudi Arabia, and how can they be used to enhance tourism strategies?

To answer this, we broke it down into 5 sub-questions, applying various modelling approaches as per the task requirements.

Tourism in Saudi Arabia is influenced by a mix of destination appeal, seasonal timing, spending behavior, satisfaction drivers, and major event experiences.

Using a variety of models, we gained valuable insights to enhance tourism strategy through:

- Better timing of events
- Promotion of high-value cities
- Targeting high-spending international tourists
- Focusing on experiences that drive satisfaction