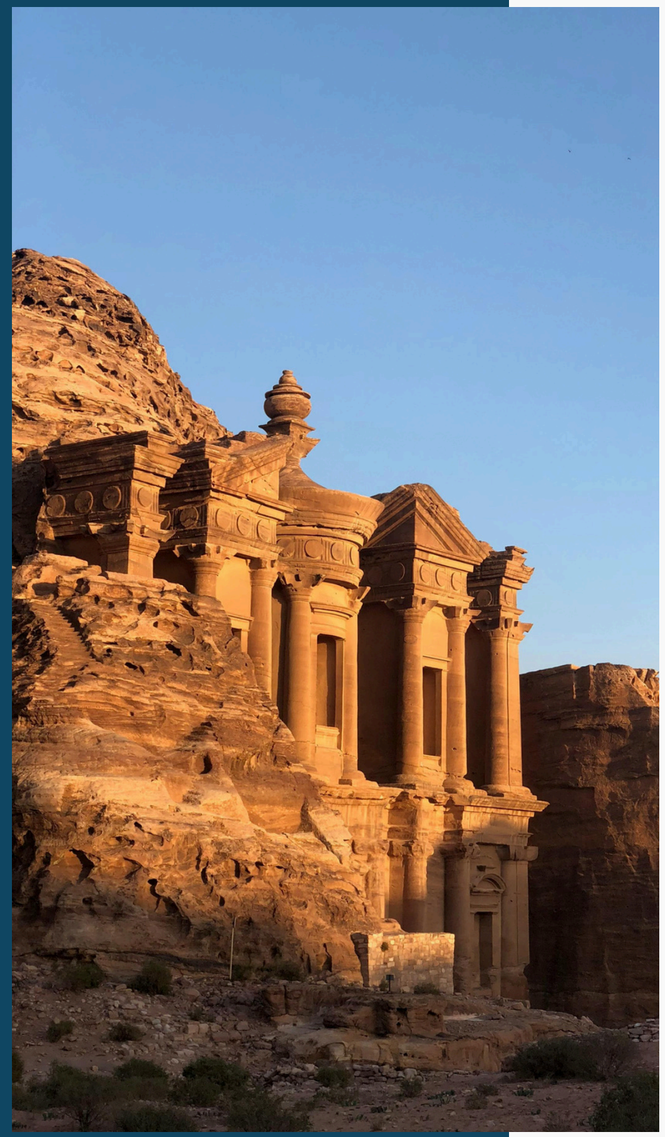


UNESCO Heritage Sites and Tourism Arrival Analysis

by Katya Kraft

Introduction

The aim of this project is to examine the correlation between **UNESCO World Heritage sites and tourism arrivals** in different countries over ten years.

By looking at tourism data from 2008 to 2018, the **goal is to find out if there's a clear link between the number of UNESCO Heritage Sites** a country has and its international tourism arrivals.

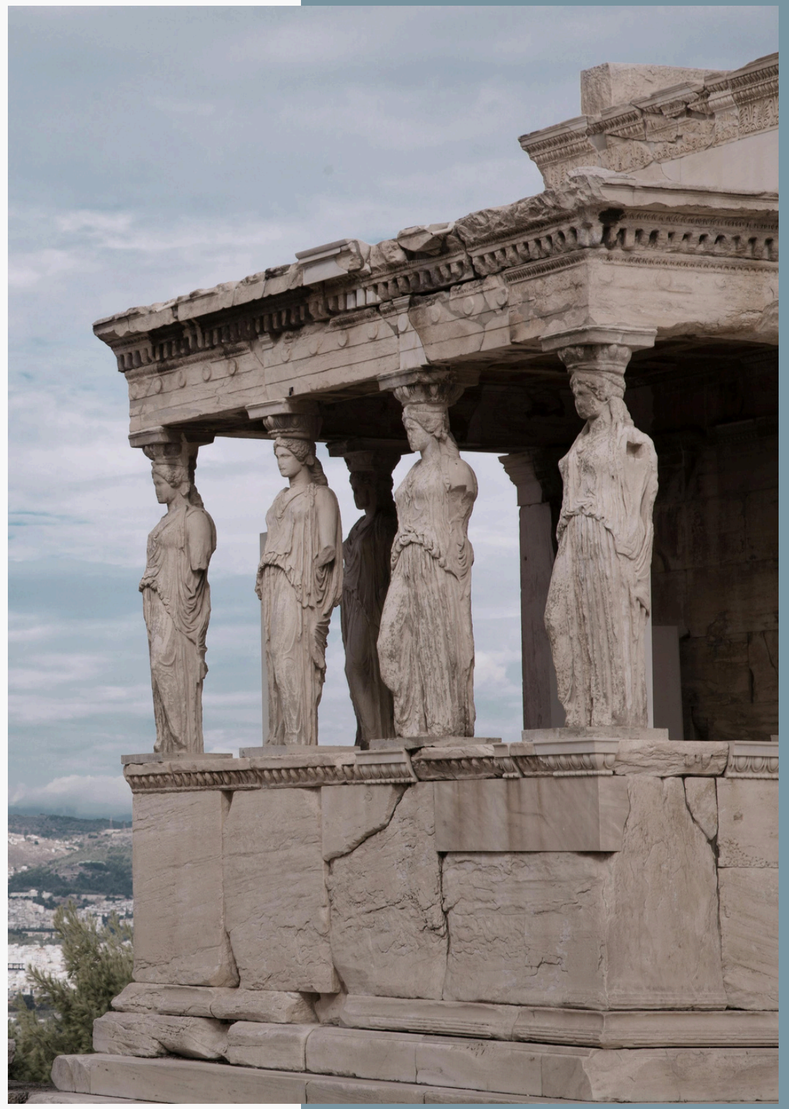
Data Overview

- **World Bank Tourism Arrivals Data:**

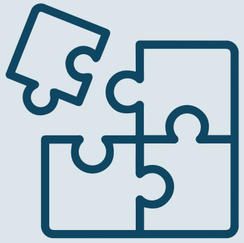
  Annual international arrivals data per country for the period of 2008–2018.

- **UNESCO World Heritage Sites (via Kaggle API):**

  The total number of UNESCO heritage site counts per country and inscription dates.



Data Merging and Cleaning



- **Dropped irrelevant columns** and removed rows with unknown or missing country data.
- **Standardized country names** to ensure consistency across sources.
- Merged data on **country name** and performed initial **EDA** with Pandas, Seaborn and Matplotlib.

Hypothesis 1: The Relationship Between UNESCO World Heritage Sites and Tourism Arrivals.

Examining the correlation between the number of UNESCO sites and tourism demand.

SQL Insights:

country	site_count	total_arrivals
France	52	986510000
Spain	50	768100000
China	57	680718000
Italy	59	584936200
United Kingdom	33	383987000
Germany	52	370587000
Mexico	35	338252000
Austria	12	298459000
Malaysia	4	298405000

country	site_count	total_arrivals
Italy	59	584936200
China	57	680718000
France	52	986510000
Germany	52	370587000
Spain	50	768100000
India	42	115091000
Mexico	35	338252000
United Kingdom	33	383987000
Iran, Islamic Rep.	27	48573000
Japan	25	174113000

Hypothesis 1: The Relationship Between UNESCO World Heritage Sites and Tourism Arrivals.

H_0: There is **no significant correlation** between the number of UNESCO sites and the number of tourist arrivals.

H_1: There is a **significant positive correlation** between the number of UNESCO sites and the number of tourist arrivals.

HYPOTHESIS TEST	RESULTS	OUTCOME
• Pearson's Correlation • Alpha = 0.05	• Pearson Correlation Coefficient: 0.8089 • P-value: 0.0000	**Reject the null hypothesis.** There is a significant positive correlation between the number of UNESCO sites and tourism arrivals.

Hypothesis 2: UNESCO Site Counts in Countries with Lower Tourism Arrivals.

Exploring the relationship between **lower tourism arrivals** and the number of UNESCO World Heritage sites.

SQL Insights:

country	site_count	total_arrivals
Kiribati	1	60800
Marshall Islands	1	61800
Comoros	0	257200
American Samoa	0	263200
Guinea-Bissau	0	304600
Micronesia, Fed. Sts.	1	370200
Djibouti	0	382000
Mali	4	387500
Guinea	1	579400
Liechtenstein	0	817200

Hypothesis 2: UNESCO Site Counts in Countries with Lower Tourism Arrivals.

H_0: There is **no association** between lower tourism arrivals and a lower number of UNESCO World Heritage sites.

H_1: Countries with lower tourism arrivals tend to have fewer UNESCO World Heritage sites.

HYPOTHESIS TEST	RESULTS	OUTCOME
• Spearman's Correlation Test • Alpha = 0.05	• Spearman Correlation Coefficient: 0.7030 • P-value: 0.0000	**Reject the null hypothesis.** There is a significant association between lower tourism arrivals and fewer UNESCO sites.

Hypothesis 3: The Impact of UNESCO Site Counts on Tourism Arrivals.

Comparing the difference in tourism arrivals between countries that have **10 or more UNESCO** World Heritage sites and those with fewer.

SQL Insights:

site_count_category	avg_annual_arrivals
Fewer than 10 UNESCO sites	24444755
10 or more UNESCO sites	249053315

Hypothesis 3: The Impact of UNESCO Site Counts on Tourism Arrivals.

H_0: There is **no significant difference** in tourism arrivals between countries with **10 or more UNESCO sites** and those with fewer than 10 sites.

H_1: Countries with 10 or more UNESCO sites **attract significantly higher tourism** arrivals than those with fewer than 10 sites.

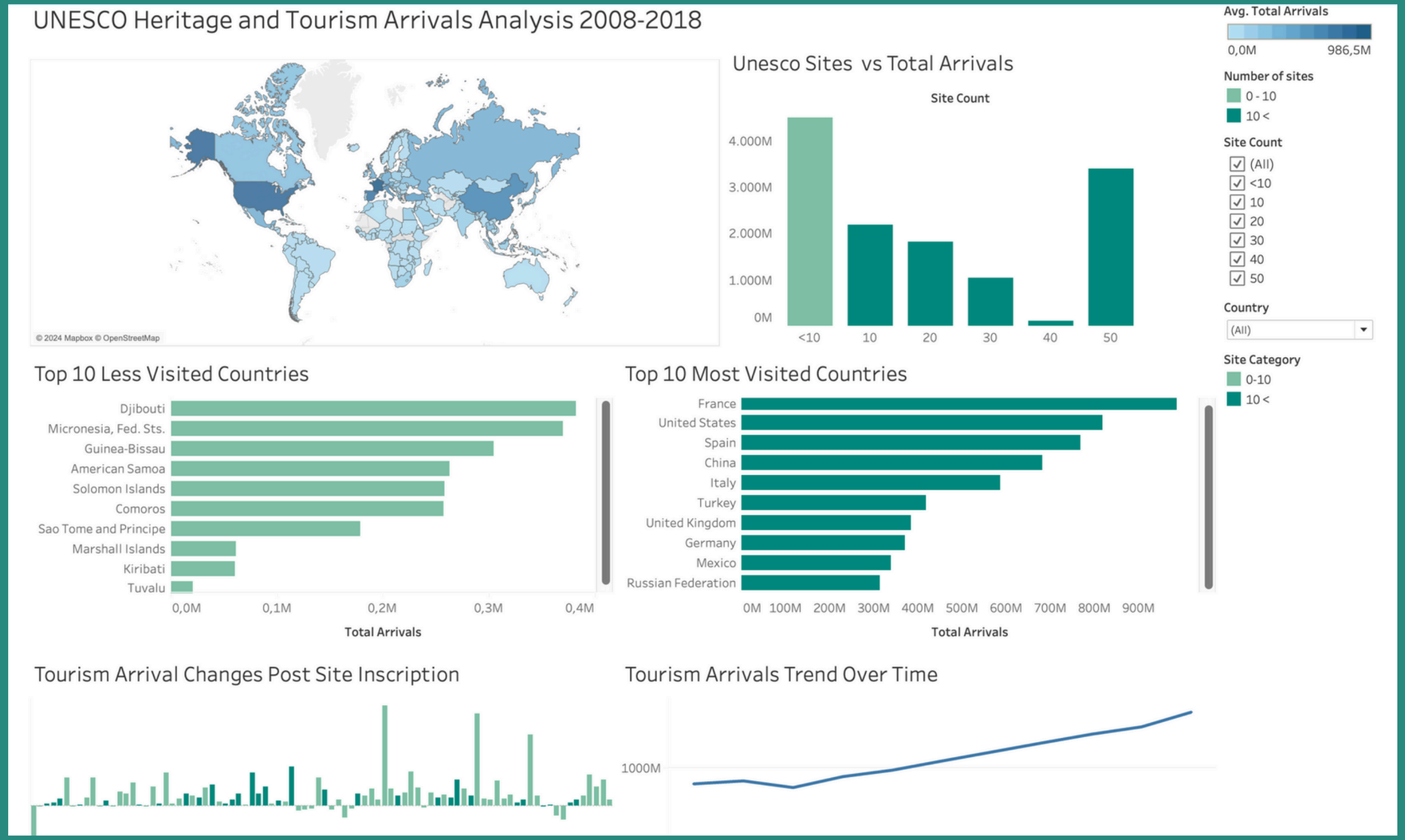
HYPOTHESIS TEST	RESULTS	OUTCOME
• Two-sample t-test • Alpha = 0.05	• T-statistic: 5.2577 • P-value: 0.0000	**Reject the null hypothesis.** Countries with 10 or more UNESCO sites attract significantly higher tourism arrivals.

Hypothesis 4: The Immediate Impact of UNESCO Site Announcements on Tourism Arrivals.

H_0: There is **no significant increase** in tourism arrivals in the year **following the announcement** of a new UNESCO site.
H_1: Tourism arrivals **significantly increase** in the year following the announcement of a new UNESCO site.

HYPOTHESIS TEST	RESULTS	OUTCOME
• Paired T-Test • Alpha = 0.05	• T-statistic: -7.1748 • P-value: 0.0000	**Reject the null hypothesis:** Tourism arrivals significantly increase in the year following the announcement of a new UNESCO site.

Dashboard


UNESCO Heritage and Tourism Arrivals Analysis 2008-2018
© 2024 Mapbox © OpenStreetMap
Unesco Sites vs Total Arrivals
Site Count
4.000M
3.000M
2.000M
1.000M
0M
<10
10
20
30
40
50
Avg. Total Arrivals
0,0M
986,5M
Number of sites
0 - 10
10 <
Site Count
(All)
<10
10
20
30
40
50
Country
(All)
Site Category
0-10
10 <
Top 10 Less Visited Countries
Djibouti
Micronesia, Fed. Sts.
Guinea-Bissau
American Samoa
Solomon Islands
Comoros
Sao Tome and Principe
Marshall Islands
Kiribati
Tuvalu
0,0M
0,1M
0,2M
0,3M
0,4M
Total Arrivals
Top 10 Most Visited Countries
France
United States
Spain
China
Italy
Turkey
United Kingdom
Germany
Mexico
Russian Federation
0M
100M
200M
300M
400M
500M
600M
700M
800M
900M
Total Arrivals
Tourism Arrival Changes Post Site Inscription
Tourism Arrivals Trend Over Time
1000M

Forecasting Tourism Arrivals: Linear Regression Model

Data Preparation:

- Train test split: (X_train, y_train).

Prediction vs Actual Sample:

	Predicted	Actual
0	3697927	3551000
1	1645576	1711000
2	25601	7100
3	27318177	30123000
4	41555951	41313000
5	61132093	62900000

Evaluation Metrics:

- **Root Mean Squared Error** (RMSE): 861,401
- **R^2 Score**: 0.9961 (Model accuracy)

Outcome:

- The model demonstrates **high accuracy** with a low RMSE of 861,401, considering the very high scale of the tourism arrivals data.

- The R^2 score of 0.996 indicates that the model explains 99.6% of the variance in the data, showing a **strong prediction performance** for tourism arrivals.

Conlcusions

Analysis of UNESCO Sites & Tourism Arrivals

- Significant positive correlation between the number of UNESCO sites and tourism arrivals.
- Countries with more UNESCO sites tend to attract more tourists.

Impact of UNESCO Inscription on Arrivals

- Tourism arrivals significantly increase in the year following the announcement of a new UNESCO site.

Forecasting Model Performance

- Linear Regression model: provides highly reliable predictions for future tourism arrivals.

Thank you!

UNESCO Heritage Sites and Tourism Arrival Analysis

by Katya Kraft