

# Effect of International Tourism on London Economy

## (COMP3125 Individual Project)

Ethan Rombalski  
Wentworth Institute of Technology

This report explores the relationship between tourism and the economic performance of London by analyzing data from 2011 to 2023. Using linear regression, the study finds a strong positive logarithmic connection between the number of tourists and the overall spending. Additionally, visualizations such as heatmaps of tourist attractions and spending trends by travel purpose highlight key patterns in tourist behavior. While limitations on the data and external disruptions like COVID-19 are acknowledged, the results offer valuable insight for policymakers and urban planners aiming to strengthen London's tourism-driven economy

### I. INTRODUCTION

In this report we will analyze the effect of tourism on the economy of London. We will do so by observing several key factors of London's economy, including number of travelers, purpose of travel, and how many tourists are attracted to the major monuments of the city. Key findings will cover the differences between the number of visitors and the correlation (if one exists) between total visitors and overall spending while in London. The data, pulled primarily from Statista and Data Press, represents the tourism information of the last 20 years.

### II. DATASETS

#### A. Source of dataset

The datasets come from two sources: Statista and the London Datastore. The Statista dataset tracks the number of overseas visitors to London from 2011 to 2023, using official travel and tourism statistics. They also provide a ranking of London's most-visited attractions, compiled from records and surveys by the local tourism board. Meanwhile, the London Datastore offers a detailed dataset on international visitor numbers, collected from government and statistical authorities—both sets being collected by the Office of National Statistics (UK).

#### B. Character of the datasets

The datasets being used were all produced by separate publishers, and as such, all had different formatting that needed to be cleaned. The set from Statista does not focus on raw data and instead only shows conclusive answer columns, as a physical representation of a graph. This data can still be used however, when combined with the other datasets in the project. Before the Statista sets can be used, however, the other datasets must be cleaned and uniform. To accomplish this, the data dictating tourism information was scaled 1 : 1000, to keep this information uniform across all datasets. In addition, the dataset from London Datastore had to be sorted in accordance with country of origin, so as to match the rest of the data. Once this cleaning is complete, we can merge our data based on year visited and region traveling from, something all datasets have in common.

### III. METHODOLOGY

For the modeling in this project, linear regression was used on the data. Linear regression is a method used to model the relationship between a dependent variable and one or more independent variables. This model is made by fitting a linear equation to observed data. The assumption that must be satisfied to use this model is that the relationship between variables must be linear, the error (residuals) must follow a normal distribution, and the variance must be independent. An advantage of this model is its simplicity, as it provides clear and easy to follow insights into the strength and direction of the relationship between the explanatory (independent) and response (dependent) variables. For this analysis, linear regression was chosen as it was hypothesized that there was a linear relationship between the number of tourists present in London and the amount being spent.

To implement this model using python, the libraries pandas and sklearn were used. Pandas allowed for data manipulation in a Data Frame format. Sklearn provided the linear regression packages necessary to complete the model. To address skewness in the model, the data was preprocessed, and the log values were used in the model. By using the log values (instead of the normal observed values) the data is transformed and can more closely follow a normal distribution.

### IV. RESULTS

#### A. Using Linear Regression to analyze correlations between number of US visitors and amount spent in London

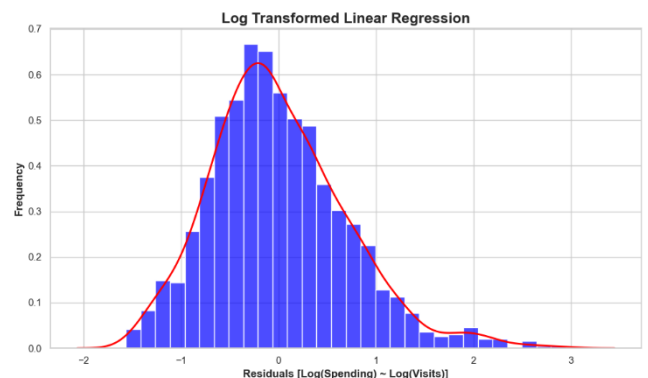


Fig. 1a. Histogram of Residuals

Before we can use our regression model, we must first ensure that the error of our data follows a normal distribution. As we can see from the histogram above, after taking the log values of our data, it satisfies the assumption of normality. Our regression model is shown below:

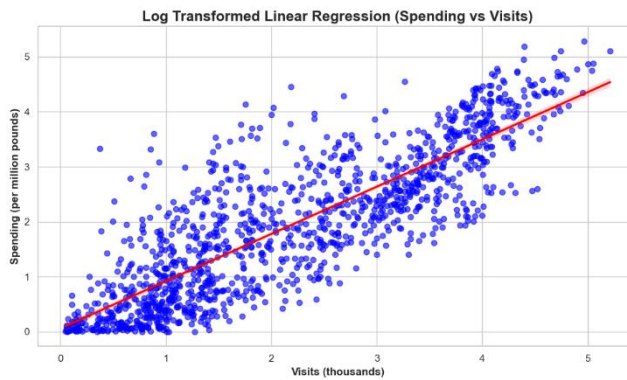


Fig. 1b. Scatterplot of Regression Model (US visitors)

The regression model shown above has the equation:

$$\text{Log}(\text{Spend}) = 0.0589 + 0.8594 * \text{Log}(\text{Visits})$$

This equation indicates a positive logarithmic relationship between the two variables. For every unit increase in the number of visitors, there is a predicted increase in spending of about 0.86%. The intercept of 0.0589 represents the expected spending when there are almost no visitors.

### B. Heatmap of Visitors to Popular British Attractions

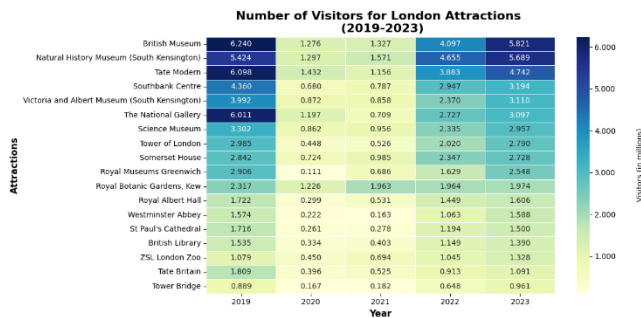


Fig. 2. Heatmap of Tourist Attractions

These results compare the number of visitors to each popular attraction with the year being observed. The darker values represent higher attendance. The British and Natural History Museum were the most attended for all four years. The max visit count was in 2019 at the British Museum, with 6,240,000 observed visitors. The minimum visitor count was observed at the Royal Museums in Greenwich with only 110,000 visitors during 2020.

### C. Line Chart of Spending During Travel by Purpose

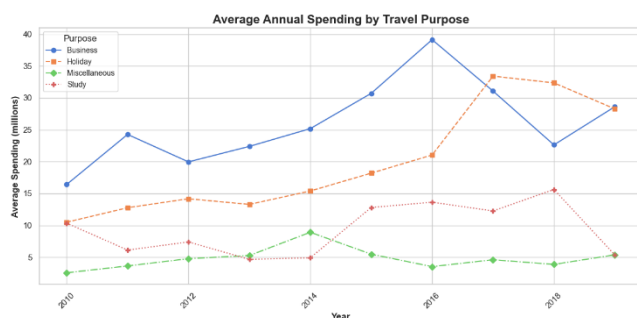


Fig. 3. Line chart of Spending

This line chart shows the differences in spending by purpose of travel. Most methods showed an increase in spending during the years of 2012 to 2016 before dropping again in 2019. The maximum spending observed was visitors traveling to London for business purposes and they spent just under 40 million pounds. The minimum spending was observed in 2010, with travelers under the 'miscellaneous' category only spending around 2.5 million pounds.

### D. Line Chart of Total Tourists Observed (2011-2023)

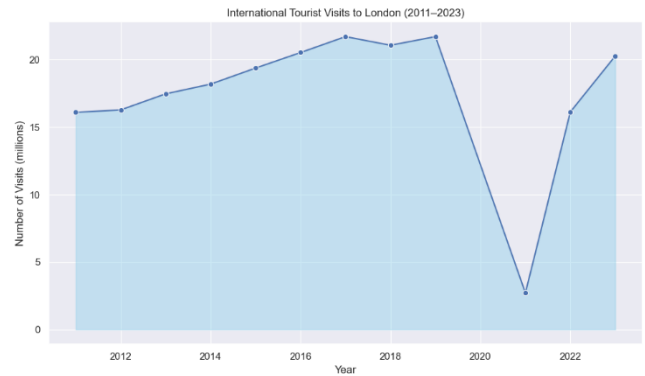


Fig. 4. Line Chart of Total Tourists

This line chart compares the total observed number of tourists over time. From 2012 to 2019 there is a steady increase in tourism from just over 15 million to a max value of about 23 million, before a steep decline in tourism during 2020 and 2021. The data point from 2020 is missing from the dataset, as this was during the COVID lockdown and tourism was not permitted in London.

## V. DISCUSSION

While the findings of the regression model show a clear correlation between the number of United States tourists visiting London and overall spending, there are several limitations to this model. Linear Regression (and further logarithmic regression) may oversimplify the relationship between tourism and economic activity. The model does not consider things like a pandemic, which the other visualizations demonstrated.

In addition, the data gathered from Statista could be more in depth with their observations. Due to the lack of daily or monthly data, the model cannot detect short-term spikes or seasonal trends that are common in the tourism sector.

In the future, using a different model like multiple linear regression or a time-series forecast could improve the accuracy and breadth of results. Other factors could be included such as exchange rates, inflation, or employment. In addition to the regression model, more specific data could be used to provide a more holistic visualization of the tourism industry in London.

## VI. CONCLUSION

This report demonstrates a strong logarithmic correlation between the number of tourists from the United States and their associated spending. The regression model shows that a 1% increase in tourist numbers is associated with an approximate 0.86% increase in spending, which highlights tourism's role in London's economy.

Additionally, the visualizations provided demonstrate the importance of culture and business in attracting visitors. Despite the disruptions in 2020 and 2021 (likely caused by the COVID-19 pandemic) the overall increasing trend from 2012 to 2019 reflects a global tourism interest in the city of London. This insight could be crucial for city planners looking to increase the number of travelers in the city. Boosting media for popular places like the British Museum or the Natural History Museum might gain more tourists, and marketing towards business travelers (the highest spenders) could only help to improve the economy. Future research, using more

data and improved modeling, could also serve to better represent how tourism impacts London's economy overall.

#### REFERENCES

- [1] Number of visits to leading tourist attractions in London, United Kingdom (UK) from 2019 to 2023 (in 1,000s) [Graph], Association of Leading Visitor Attractions, March 18, 2024. [Online]. Available: <https://www.statista.com/statistics/630146/leading-london-visitor-attractions-uk/>. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [2] O. for N. Statistics, "Number of International Visitors to London," UK Government, London Datastore, 2024. Accessed: Mar. 30, 2025. [ExcelWorksheet]. Available: <https://data.london.gov.uk/dataset/number-international-visitors-london>