Inbound Tourism to Ireland: Using Machine Learning to Predict Demand

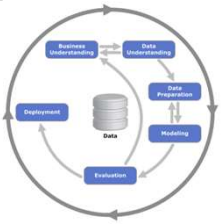
Parhad Keyim Idikut, CCT College Dublin, May 2025

INTRODUCTION

- **Project aim**: produce Irish inbound tourism demand forecasting models.
- **Research hypothesis**: Irish inbound tourism, as a part of international tourism, will continue to grow overall, but will experience temporary fluctuations.

STRATEGIC OVERVIEW OF THE BUSINESS PROBLEM: IRISH INBOUND TOURISM

- **General Goal**: The prediction could supplement the Irish tourism sustainable development strategy by equally considering the environmental and socio-economic impacts of the industry.
- **Ethical Considerations**: It is important to consider all potential forecasting biases which might be caused by inadequate data in terms of size and content.
- Project Plan and **Methodology**: Project is implemented by utilising the **CRISP-DM** methodology. It is a **cyclical** process that encourages continuous improvement. Thus, any problems that arose from the stage of this project are resolved by looping back to an earlier stage, until a desirable result is produced.



Source: Shearer, C., 2000. The CRISP-DM model: the new blueprint for data mining. Journal of Data Warehousing, 5(4), pp.13-22.

BUSINESS UNDERSTANDING: IRISH INBOUND TOURISM

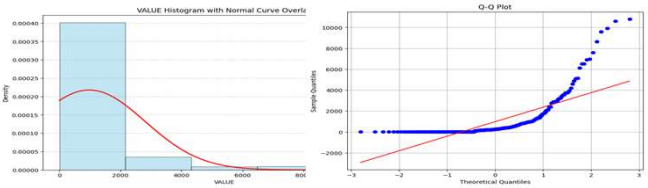
- Tourism is the country's largest indigenous industry and biggest regional employer.
- The Irish government aims to implement a sustainable development strategy.
- The inbound tourism demand prediction is relevant to the Irish government's strategy of prioritising the economic value per visitor over the volume of tourists visiting.

TECHNOLOGIES USED

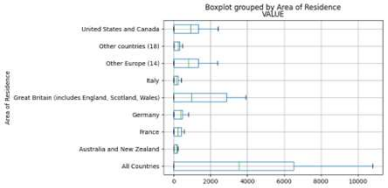
- **Libraries and Models:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Sklearn, Sktime, Shapiro-Wilk Test, QQ-plot, ANOVA, Box plot, Adfuller, ACF, PACF, ARIMA, SARIMAX, and Theta Forecaster.
- **Model Selection and Appropriateness**: focused on **time series analysis** because forecasting tourism demand is a prediction problem.

DATA UNDERSTANDING

- **Data Sources/Acknowledgements**: Two datasets, covering the years 1983 – 2008 and 2009 – 2019, were obtained from the Central Statistics Office of Ireland. The datasets are open source for non-commercial use.
- **Descriptive Statistics and Findings:** It needs to perform a transformation, e.g., **differencing**, to **normalise the time series data to deal with non-stationarity**. Since stationary data **could improve the prediction accuracy**.

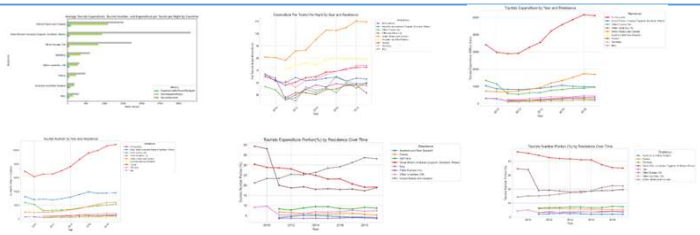


- **Boxplot** of nationality of the inbound tourists to Ireland: the most "VALUE" of inbound tourists to Ireland come from Great Britain, the United States and Canada (combined), Germany and France.



DATA PREPARATION

- **Data Cleansing, Formatting, EDA, and Identification of Relevant Variables and Features of Prediction:** Overall, the data preparation results, and visualization suggests, **the United States, Canada and Great Britain** are the main contributors to Irish inbound tourism.

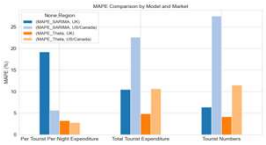


FORECASTING AND EVALUATING IRISH INBOUND TOURISM: ARIMA, SARIMA, AND THETA MODELS

- Data **stationarity** is checked by using the **ADF test**. Performed **ACF and PACF plots** to estimate the time-series model-building parameters such as **p, d, and q**.
- Key **forecasting indicators**: Total Tourist Expenditure, Per Tourist Per Night Expenditure, and Tourist Numbers.
- **Forecasting for the UK:**

- **Forecasting for the US/Canada:**

- **Model Evaluation and Comparison**: the Mean Absolute Percentage Error (**MAPE**); lower values reflect better model performance.
- The Theta model consistently outperforms the SARIMA model across all tourism indicators examined.



CHALLENGES ENCOUNTERED AND SOLUTIONS

- Main **challenges**: data size and its contents.
- **Solutions**: Combined data to produce a bigger dataset that covers the period from 1983 to 2019. Produced 10 years of future yearly dates from 2020 to 2029.

FINDINGS, RECOMMENDATIONS, AND CONCLUSION

- The UK remains the highest-volume source market, i.e. in terms of **tourists' number**.
- The US/Canada tourists represent a **higher per-visitor expenditure**, and the **forecasting** indicates that this trend is a **gradual growth**.
- **Recommendations**:
- The **Theta model** should be considered for use in the Irish inbound tourism forecasting.
- The Irish Tourism Authority should conduct dedicated **tourist behaviour surveys and statistical reporting on key markets** of Great Britain and the United States and Canada. These enriched datasets would help to conduct of more effective forecasting.
- Ireland's tourism **strategy should shift from mass tourism to value-driven tourism** products and services.
- In **summary**, there is a shift from British tourists' **dominance towards the United States & Canadian tourists**, in both spending and visitor numbers.
- The prediction models **should be continuously improved** in future projects.

REFERENCES

- Arunraj, N.S., Ahrens, D. and Fernandes, M., 2016. Application of the SARIMAX model to forecast daily sales in the food retail industry. International Journal of Operations Research and Information Systems (IJORIS), 7(2), pp.1-21.
- Central Statistics Office, 2020a. TMA09 - Overseas Trips (incl. Expenditure) to Ireland - data.gov.ie (Date updated 2020-09-15). Available at: https://data.gov.ie/dataset/tma09-overseas-trips-incl-expenditure-to-ireland (Accessed: 20 October 2024).
- Central Statistics Office, 2020b. TMA06 - Overseas Visits to Ireland (Date updated 2020-05-14). Available at: https://data.cso.ie/ (Accessed: 13 April 2025).
- Cheer, J.M. and Lew, A.A., 2017. Understanding tourism resilience: Adapting to social, political, and economic change. In Tourism, resilience and sustainability (pp. 3-17). Routledge.
- Government of Ireland, 2024. Tourism Policy Framework 2025–2030. Available at: https://www.gov.ie/pdf/?file=https://assets.gov.ie/31416/2b0fbb57-62df-4b6c-b90f-4af38ff2590e.pdf#page=null (Accessed: 24 November 2024).
- Gunter, U. and Onder, I., 2015. 'Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data', Tourism Management, 46, pp. 123–135. Available at: https://doi.org/10.1016/j.tourman.2014.06.017.
- Irish Tourism Industry Confederation, 2024. 'VISION 2030 – An Irish Tourism Strategy for Growth', Irish Tourism Industry Confederation - ITIC. Available at: https://www.itic.ie/vision-2030/ (Accessed: 24 November 2024).
- Isik, M., 2025. 'Airline Passenger Forecasting'. Available at: https://kaggle.com/code/mehmetisik/airline-passenger-forecasting (Accessed: 19 January 2025).
- McKinney, W., 2012. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.
- Perktold, , et al., 2024. The Theta Model. 'statsmodels/statsmodels: Release 0.14.2'. Zenodo. Available at: https://doi.org/10.5281/ZENODO.593847.
- Shearer, C., 2000. The CRISP-DM model: the new blueprint for data mining. Journal of Data Warehousing, 5(4), pp.13-22.
- Song, H. and Li, G., 2008. 'Tourism demand modelling and forecasting—A review of recent research', Tourism Management, 29(2), pp. 203–220. Available at: https://doi.org/10.1016/j.tourman.2007.07.016.
- Song, H., Qiu, R.T.R. and Park, J., 2019. 'A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting', Annals of Tourism Research, 75, pp. 338–362. Available at: https://doi.org/10.1016/j.annals.2018.12.001.
- Weiss, N., 2016. Introductory Statistics, MyLab Revision, Global Edition. 10th edition. Boston; Columbus; Indianapolis, New York: Pearson.