Analysis Report

The Relationship Between Train Usage and Greenhouse Gas Emissions in EU Tourism Ecosystems

Methods of Advanced Data Engineering SS24

Presented by: Hend Abdo Hussien (23008093)

Contents

1. Introduction

Tourism is a major economic driver in the European Union countries, contributing approximately 11% to employment, encompassing direct, indirect, and induced jobs in 2019 according to the estimates of the European Commission [1]. However, the influx of tourism poses significant environmental challenges, particularly in terms of greenhouse gas (GHG) emissions which was estimated be the European Commission to be as high as 48 Tons of GHG emissions per Million Euro of Gross Value Added on average across the EU countries [2]. As global awareness of climate change increases, the need for sustainable tourism practices has become more urgent. One critical aspect of this is the mode of transportation used by tourists within their destinations. Trains as a mode of transport, are known for their lower environmental impact compared to cars and airplanes. As part of the efforts the necessity to accelerate green transition in EU Tourism, the Council of the European defined the share of trips by train as one of the green pillars for this green transition, as it measures the relative importance of sustainable means of transportation within a tourism destination, approximated by the share of trips taken by train. Higher values indicate a more widespread use of the train compared to other modes of transport with a higher environmental impact. This report aims to address the research question: How does the utilization of trains as a mode of transportation within a tourism destination influence the overall greenhouse gas intensity of the tourism sector within EU countries? By doing so, it is examining the potential of train travel to reduce GHG emissions.

2. Used Data

The dataset used for this analysis was aggregated from datasets provided by the European Commission and are subject to the European Commission Reuse and Copyright Notice. The table below describes the data used in the analysis.

Fields	Description	Usage in Analysis	Unit of Measure
TERRITORY_ID	Identifier key of the EU countries	Key	
TERRITORY_NAME	Name of the EU counties	Categorical Info	
YEAR	Year of observation, it covers the period from 2019 to 2021	Key	
GHG_VALUE	This indicator measures the amount of greenhouse gas (GHG) emissions produced by the tourism ecosystem per Million Euro of Gross Value Added (GVA) in the tourism sector. [2]	Quantitative Indicator	Ton of CO2 per Million EUR
TRAIN_TRIPS_SHARE	The indicator is obtained dividing the number of trips by train by the number of trips done using all means of transport, i.e.: Air, land, railways, buses, coahes, motor vehicles, waterway, other. [3]	Quantitative Indicator	Percentage
NIGHTS_SPENT_COUNT	This descriptor measures the total number of nights spent at tourist accommodation establishments in a destination (country or region) in a given year, from both domestic and foreign tourists.[4]	Quantitative Indicator	Night
GHG_NORMALIZED	Derived data created by dividing GHG_VALUE by NIGHTS_SPENT_COUNT in Millions to consider the variation in tourism demand across countries	Derived data. Quantitative Indicator	Ton of CO2 per Million EUR per Million Nights Spent

3. Analysis

3.1. Method

The analysis implements statistical methods to investigate the relationship between the share of trips by train and the GHG intensity of the tourism sector. First, a trend analysis of the variables of interest is performed, to understand the effects of temporal anomalies like the pandemic and to take it into consideration. Then, correlation analysis to evaluate the strength and direction of their linear relationship, with the correlation coefficient indicating whether the variables are positively, negatively, or not correlated. Also linear regression is used to model this relationship determining the slope and intercept of the regression line in order to identify the rate of change in GHG_NORMALIZED for each unit change in TRAIN_TRIPS_SHARE. Data visualization techniques such as scatter plots with regression lines further illustrate this relationship visually, and sorted bubble plot with countries as categorical key to show trends, outliers and help in the interpretation of data points.

3.2. Result and interpretation

a) Trend analysis to identify temporal anomalies and geographical variance

The grouped bar chart below (Figure 1) reflects the trend in tourism demand over the years of interest. It is measured by the total number of nights spent at tourist accommodation establishments in a destination from both domestic and foreign tourists. A significant drop in demand in 2020 and 2021 can be attributed to the COVID-19 pandemic peak period. This measure is used in the analysis normalize the GHG intensity given the variation in demand across EU countries. For example, a country with generally high demand in tourism like Spain is expected to have higher GHG intensity from tourism than a low demand country like Latvia.

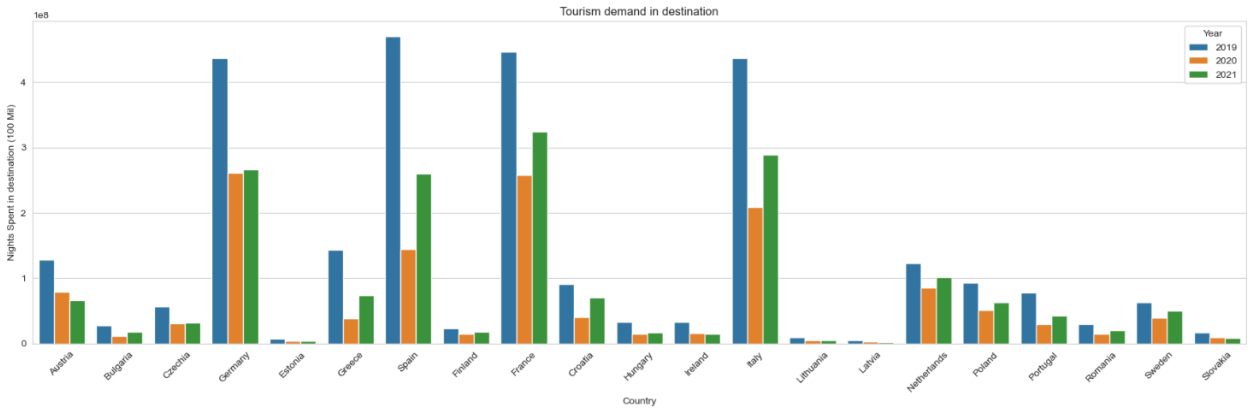

Figure 1: Tourism demand in EU countries represented by nights spent in destinations

This reduction in tourism can lead to a decrease in the GVA generated by the tourism sector, while fixed GHG emissions from maintenance of tourism infrastructure (hotels, resorts, etc.) remain relatively constant. As a result, the GHG emissions per unit of GVA would appear higher in the pandemic time as the bar char shows (Figure 2).

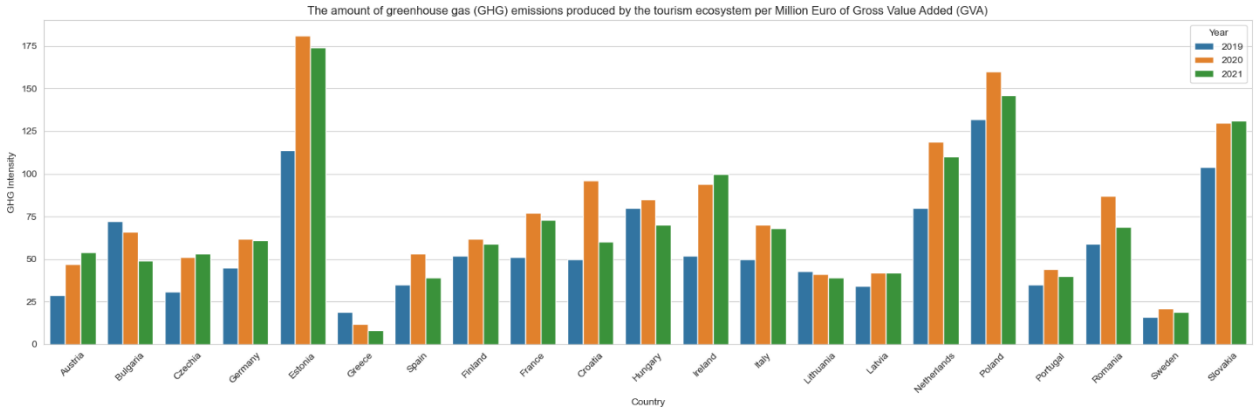

Figure 2: The Greenhouse Gas emissions intensity in tourism sector per Gross Value Added by tourism activities

The use of train trips also fluctuated during the period of interest, which can be attributed to the preference of avoiding public transport during the lock down period.

b) Correlation analysis

The correlation coefficient between the two variables of interest is approximately -0.237, indicating a weak negative correlation. This suggests that as the share of trips by train TRAIN_TRIPS_SHARE increases, the normalized greenhouse gas intensity of the tourism sector GHG_NORMALIZED tends to decrease slightly. However, the strength of this relationship is not strong enough to imply a clear relationship between these variables and the observed relationship does not necessary mean that an increase in train trips causes a reduction in greenhouse gas intensity. Therefore further analysis using regression modeling is performed to understand the nature of the relationship between these variables.

C) Regression Analysis

The visualization below (Figure 3) shows a scatter plot with regression lines for the years. It shows a negative linear relationship between TRAIN_TRIPS_SHARE and GHG_NORMALIZED.

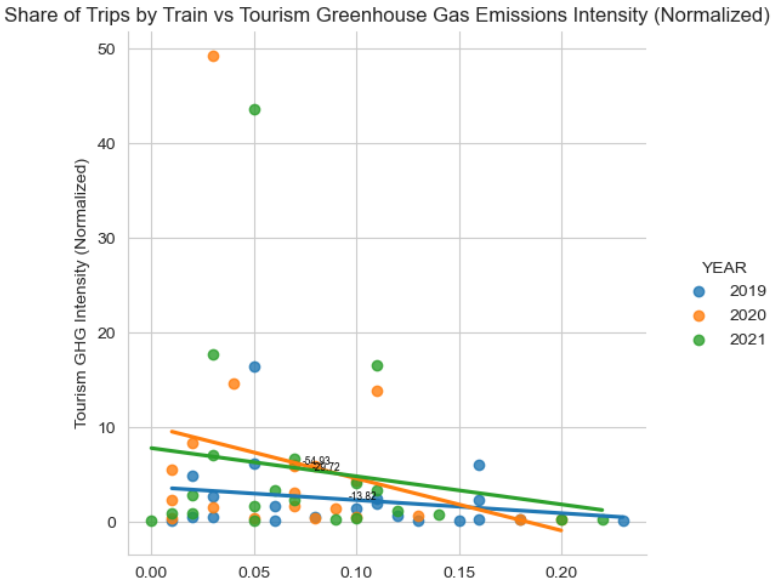

Figure 3: Regression analysis of GHG intensity and Share of trips by trains

The average slope value -34.30 is suggesting that even a small increase in train travel can lead to a significant reduction in GHG emissions intensity. However there is a noticeable difference in slop across the years that can be justified by the anomalies identified earlier.

The average intercept = 7.43 provides a baseline for emissions intensity when train travel is not considered, which is higher than the 75th Percentile Q3 = 5.59 which also provides an insight into the influence of train use. Given that the scatter plot provide an insight into the quantitative variables only. The sorted bubble chart below (Figure 4) was created to show the categorical variable of the countries, with the size of the bubble representing the GHG intensity.

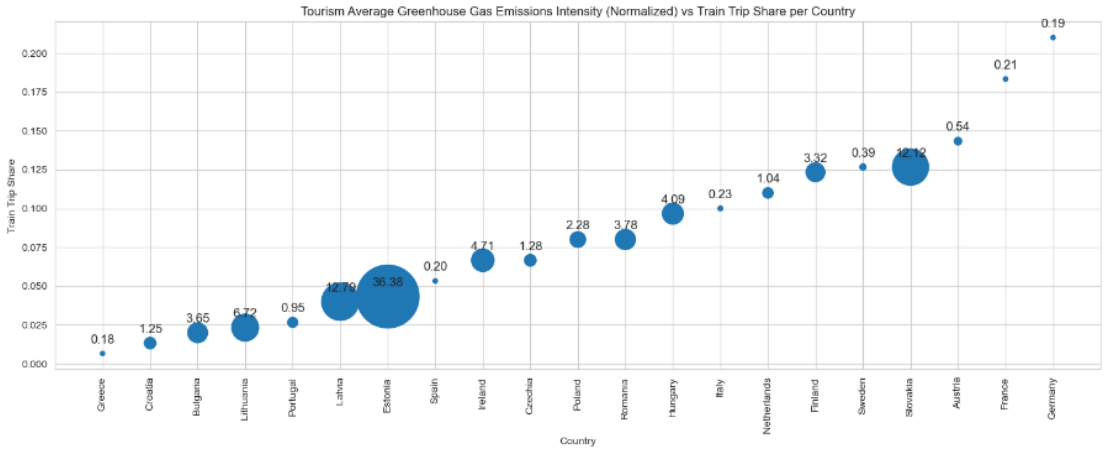

Figure 4: Average share of train trips and average GHG intensity (normalized by demand) per country

4. Conclusion

The research question posed was to understand the relationship between the share of trips by train and the greenhouse gas GHG intensity in the tourism sector. The proposed statistical analysis through correlation and regression methods suggests that an increase in train travel is generally associated with a reduction in GHG intensity. This finding indicates that promoting sustainable means of transport could potentially mitigate GHG emissions within the tourism industry. However, the weak correlation implies that other factors may also influence GHG intensity and the issue should be approached with more complex datasets and analysis perspectives. Additionally, the data used included periods of significant disruption such as the pandemic, which may have impacted the results. Further research with more stable and comprehensive data is needed to fully confirm these findings and to understand the broader context of other influencing factors. Therefore, while the question has been answered to an extent, the conclusions should be viewed while considering these limitations.

5. References

[1] EU Tourism Dashboard (2019) : https://tourism-dashboard.ec.europa.eu/background-methodology?lng=en

[2] Silvia Iodice (2022): UDP - Tourism GHG intensity. European Commission, Joint Research Centre (JRC)

[3] Silvia Iodice (2022): UDP - Share of trips by train. European Commission, Joint Research Centre

[4] Batista, Filipe (2022): UDP - Nights spent. European Commission, Joint Research Centre (JRC)