**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | MSc in Data Analytic |
| **Assessment Title:** | Road Freight Transport of Ireland and France |
| **Lecturer Name:** | David McQuaid ,Dr. Muhammad Iqbal, Taufique Ahmed, Sam Weiss |
| **Student Full Name:** | Unnati Bhardwaj |
| **Student Number:** | 2023404 |
| **Assessment Due Date:** | 07/1/2024 |
| **Date of Submission:** | 07/1/2024 |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

# Group ID - MSc in Data Analytics

Author: Unnati Bhardwaj

e-mail: 2023404@student.cct.ie

Student ID:2023404

**Abstract:***This study conducts a detailed analysis of road freight data in Ireland and France, utilizing datasets from the OECD 'Trends in the Transport Sector' survey. The objective is to decipher patterns, trends, and discrepancies in road freight activities between the two nations using a combination of statistical and machine learning techniques. The research begins with thorough data preparation followed by exploratory data analysis to ascertain fundamental characteristics and trends. Statistical tests further identify significant differences in the road freight activities of Ireland and France.For machine predictive modeling,we have use the CRISP\_DM Framework which employs the ARIMA model for forecasting Ireland's future road freight activities and the Random Forest algorithm to elucidate the determinants of road freight activities in both countries. The effectiveness of these models is meticulously evaluated using relevant performance metrics.The findings disclose distinct patterns and influential factors within the road freight sector, providing insights into the transportation and economic frameworks of the respective countries. The predictive models furnish forecasts to aid in strategic planning and decision-making in the transport sector. By blending sophisticated statistical and machine learning methods, this research enhances the understanding of road freight dynamics and serves as a valuable asset for policymakers, businesses, and scholars engaged in transportation, logistics, and economic growth studies.*

**Acknowledgement:**My appreciation for the lecturers' support,Taufique Ahmed , David McQuaid, Sam Weiss and Muhammad Iqbal . The project supervisor David McQuaid for their patience and feedback, which their classes helped in building our project structure.I also could not have undertaken the completion without our classmates, who generously provided exchange of knowledge during the classes.

**Introduction:**

*Background:*From a business perspective, logistics is important in the globalized world we live in. Around the world, companies are increasingly dependent on sources of inputs and the demand for coordinated transportation is increasing. With increasing fuel prices and controversies over the environmental impact of CO2 emissions there is a growing need for cost efficiency in transportation (economically and environmentally as stated by Nielsen et al., 2008). 2003 discussed it) The road freight industry is a highly competitive market with low cost margins Growth and operational efficiency are essential. Many companies are outsourcing road freight services to third parties involved in the logistics and transportation process. Morgan (2006) mentions a predicted increase in outsourcing activities due to the highly productive benefits that stimulate productivity.

*Objective:*The main aim of this study is to critically examine and compare the logistics of goods between Ireland and France, with the aim of identifying and understanding trends, patterns and anomalies between the two countries the bottom of the table.

*Prepare and maintain data:* Ensure a solid foundation for analysis through careful data preparation and preparation to address missing values and discrepancies.

*Conduct Exploratory Data Analysis (EDA):* Use statistical techniques to develop a preliminary understanding of the distribution, major trends, and variability in road freight activity, and explore and characterize the data especially the summary.

*Conduct statistical tests*: Use statistical tests such as t-tests and ANOVAs and others also to compare road freight between Ireland and France, and identify any statistically significant differences.

*Use ARIMA to forecast future trends:* Use the ARIMA (AutoRegressive Integrated Moving Average) model to forecast future road trends for Ireland, and take advantage of its ability to better understand and control time series data the value of the prophecy.

*Forecasting and Interpreting Random Forest Interpretation:* Use random forest algorithm to model and identify driving factors for freight traffic in both countries. This includes understanding the importance of diversity.

*Evaluate model performance:* Use appropriate parameters to critically evaluate the performance of ARIMA and random forest models to ensure reliability and prediction and prediction accuracy.

**Methodology:**

*Data source:*This study employs road freight data from Ireland and France, sourced from the OECD's 'Trends in the Transport Sector' survey (https://stats.oecd.org/#). The data, under OECD’s terms (https://www.oecd.org/termsandconditions/), is used for comparative analysis and forecasting trends in road freight activities using ARIMA and Random Forest models.

Key points from the OECD's terms are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Data Usage:* The OECD authorizes data extraction and distribution for any purpose, including commercial endeavors. This permission is crucial for our extensive analysis. | *Attribution:* Proper attribution is required. This study acknowledges the OECD as the data source, with findings and conclusions being our own and not representing the OECD's views. | *Data Integrity*: We ensure the integrity of the original data, with any modifications for analytical purposes clearly documented in the methodology. | *Sharing and Distribution*: Consistent with OECD's terms, derived data or reports from this study are shared to foster knowledge dissemination. | *Updates and Revisions:* Recognizing the OECD’s periodic updates to its data and terms, we commit to ongoing review for compliance and accuracy.This acknowledgement underscores our commitment to ethical data usage and transparency in our research process. |

*Data Cleaning and Preparation:*This study commenced with a crucial data cleaning and preparation phase to ensure the datasets from Ireland and France were accurate and of high quality. Initially, missing values were identified using the isnull().sum() function from Pandas, and irrelevant columns with all missing values were removed via dropna(axis=1, how='all'). This refined the datasets, maintaining their integrity for analysis. Next, a duplicate check ensured each record's uniqueness, upholding the study's validity. Finally, to enable a comparative analysis, the Ireland and France datasets were merged on common columns using the merge function. This process created a unified dataset, setting the stage for a comprehensive analysis of road freight activities in both countries.

*EDA(Exploratory Data Analysis:*

Exploratory Data Analysis (EDA) is crucial for understanding a dataset's structure through statistical and visualization techniques. It includes:

|  |  |  |  |
| --- | --- | --- | --- |
| Univariate Non-Graphical EDA | Univariate Graphical EDA | Multivariate Non-Graphical EDA | Multivariate Graphical EDA |
| Focuses on single-variable analysis to describe the data and identify patterns. | Employs graphical tools like stem-and-leaf plots, histograms, and box plots to visualize and understand the distribution and shape of the data. | Explores relationships between two or more variables using cross-tabulation or statistics. | Visualizes relationships between multiple variables using plots like grouped bar charts. |

In this study, Univariate Graphical EDA was applied to the combined dataset using histograms and box plots to visualize the distribution of road freight values for Ireland and France. Additionally, time series analysis, a form of univariate graphical EDA, was conducted to understand how freight values change over time. Summary statistics provided a snapshot of the data's central tendency and dispersion, while the Interquartile Range (IQR) was used to identify outliers and anomalies, crucial for ensuring the data's integrity and understanding its distribution.correlation analysis was conducted to explore the relationship between Ireland and France's road freight values, employing the Pearson correlation coefficient for its ability to measure linear relationships. This analysis was facilitated using Python's Seaborn library (sns), specifically utilizing sns.heatmap() for a visual representation of the correlation matrix.For visualizing data trends over time, sns.lineplot() was used to create line plots, effectively showing changes at equal intervals, ideal for time series analysis. Distribution patterns were examined using sns.histplot() to generate histograms, providing insights into the data's distribution and skewness. These visualization techniques were integral to understanding the data's underlying characteristics and relationships.

*Statistical Analysis:* Inferential statistics are crucial for understanding relationships and differences within the data. They provide insights that go beyond mere description, allowing for hypothesis testing and drawing conclusions about the broader population. In this study, several inferential statistical methods were utilized based on the nature of the data and specific research questions.Pearson Correlation Coefficient has been used to determine if a significant linear relationship exists between the road freight values of Ireland and France. This test was chosen for its widespread acceptance and effectiveness in measuring the strength and direction of association between two continuous variables.after that the T-test has been used to compare the mean road freight values between Ireland and France.The t-test is particularly suited for comparing the means of two groups. Given the study's aim to examine differences between the two countries, the t-test was deemed appropriate.the third one is ANOVA (Analysis of Variance)to assess whether there are significant differences in the road freight values across different years for a single country or between groups. ANOVA extends the t-test to more than two groups and was chosen for its ability to compare means across multiple groups, making it particularly useful for this study's multi-year and multi-group comparisons.the othet statistics test is Mann-Whitney U Test which is to compare differences between two unpaired samples when the data doesn't meet the assumptions required for a t-test.This non-parametric test is less restrictive about the data's distribution and was selected for its robustness, especially useful when the normality assumption is in question.and the last one is Kruskal-Wallis Test: To test for differences between several samples, serving as a non-parametric alternative to one-way ANOVA.It has been Chosen for its applicability in situations where the data may not follow a normal distribution, allowing for comparison of multiple groups without strict distributional requirements.In employing these statistical methods, the study aims to provide a comprehensive understanding of the relationships and differences within the road freight data. The choice of each test is aligned with best practices in statistical analysis, ensuring that the findings are both valid and reliable.

*Time Series Forecasting with ARIMA:*To forecast the road freight values for Ireland and France, the ARIMA (AutoRegressive Integrated Moving Average) model was chosen for its proficiency in handling time series data. A crucial prerequisite for ARIMA modeling is ensuring the stationarity of the time series, as non-stationary data can lead to unreliable and non-interpretable models.the ADF test was employed to test the null hypothesis that the 'Value\_Ireland' and 'Value\_France' time series are non-stationary and possess a unit root.The choice of the ADF test is due to its widespread acceptance and effectiveness in determining the stationarity of a time series. It provides a structured approach by offering a test statistic and p-value, which are critical for making informed decisions about the stationarity of the series. The test was applied to both 'Value\_Ireland' and 'Value\_France' columns. If the p-value was found to be less than the significance level (typically 0.05), the series was considered stationary; otherwise, further steps were taken to achieve stationarity.In cases where the time series was found to be non-stationary, differencing was performed. This involves subtracting the previous observation from the current observation, thereby stabilizing the mean of the time series and making it suitable for ARIMA modeling. After differencing, the data was visualized again to confirm that the now-transformed series appeared stationary.

*Predictive Modeling with Random Forest:* Random forest regression is a supervised learning algorithm that uses an ensemble learning method for [regression](https://builtin.com/data-science/regression-machine-learning" \t "https://builtin.com/data-science/_blank).To predict and understand the driving factors behind road freight values, the Random Forest Regressor was selected as the primary predictive model. This decision was based on several characteristics of the model and the nature of the data it handling Complex Interactions it is capable of capturing complex, non-linear relationships and interactions between features. This is particularly beneficial for datasets where the relationship between variables isn't straightforward. The second reason is because of Robustness. The model is less prone to overfitting compared to other algorithms, mainly due to the averaging process across multiple decision trees, making it a robust choice for a wide range of datasets. Random Forest provides insights into the importance of each feature in making predictions. This is valuable for understanding which factors most significantly impact road freight values. In this study, the Hyperparameter Tuning with GridSearchCV has also been performed. The performance of the Random Forest Regressor can be significantly influenced by its hyperparameters. To optimize these hyperparameters, GridSearchCV systematically works through multiple combinations of parameter tunes, cross-validating as it goes to determine which tune gives the best performance. The objective is to find the optimal settings for the Random Forest model to improve its accuracy and reliability. It iterates through every combination of hyperparameter values provided, performing cross-validation for each combination and recording the performance. The combination that yields the best performance is then selected as the optimal set of parameters for the model. Important hyperparameters for the Random Forest include the number of trees in the forest (n\_estimators), the maximum depth of the trees (max\_depth), and others that control the splitting behavior of the trees and the selection of features.The rationale for using GridSearchCV is its ability to methodically search through a wide range of hyperparameter combinations and identify the most effective ones, thereby enhancing the model's predictive performance.By combining the Random Forest Regressor with GridSearchCV for hyperparameter tuning, the study aims to construct a robust and accurate predictive model. This methodology allows for a thorough exploration of the model's capabilities and ensures that the final model is well-suited to the data and the analysis objectives.

**Results:**

*Exploratory Data Analysis (EDA) Findings:*

*Summary Statistics:*

Ireland: The EDA revealed that the mean road freight value is approximately 8,727 million tonnes-kilometres, with a standard deviation of about 4,550, indicating variability. The median value is close to the mean at 8,849 million tonnes-kilometres, suggesting a fairly symmetrical distribution.

France: In contrast, France's mean road freight value is significantly higher at approximately 117,753 million tonnes-kilometres, reflecting its larger scale of transport. The standard deviation is about 66,315, indicating wider variability. The median, also significantly higher than Ireland's, confirms the larger scale of road freight in France.

The detailed summary statistics for both countries are provided in Appendix A. These statistics illustrate the distribution characteristics and scale differences in road freight transport between Ireland and France.

*Histogram and Box plot:*

*Histogram:*

The histograms provided a visual representation of the distribution of road freight transport values for both countries:

Ireland: The distribution of road freight values appears fairly symmetrical, indicating a balanced spread of data around the mean.

France: Conversely, France's distribution is right-skewed, suggesting a few years with exceptionally high road freight transport compared to the rest.These findings, visualized in the histograms and detailed in the code, are documented in Appendix B. The visualizations highlight the differences in the distribution patterns of road freight values between the two countries.

*Box-Plot:* The box plots were utilized to visually inspect the distribution and identify potential outliers in the road freight values for both countries:

Ireland: The box plot reveals a fairly compact distribution, suggesting less variability in road freight values. However, a few points on the higher side may be considered outliers, indicating occasional spikes in road freight transport.

France: In contrast, France's box plot shows a wider distribution, indicating greater variability. Notably, several potential outliers are observed on both the lower and higher ends, suggesting more frequent extreme values in road freight transport.

These visualizations provide an intuitive understanding of the spread and potential anomalies in the data. The detailed box plots and corresponding code are documented in Appendix B, offering a visual reference to complement these findings.

*Trend Analysis:*

Line Plot Analysis of Trends Over Time:The line plots provided a visual assessment of how road freight transport values have changed over the years for both countries:

Ireland: The trend for Ireland displays fluctuations, with noticeable peaks and troughs throughout the observed period. This suggests variability in Ireland's road freight transport, with periods of both increase and decrease.

France: In contrast, France demonstrates a generally increasing trend in road freight transport, particularly in the latter years. The consistent upward trajectory indicates a growth pattern in France's road freight sector.

These observations, illustrated through line plots, offer insights into the dynamic nature of road freight values over time. The detailed visualizations and the corresponding analytical code are available in Appendix B, providing a clear visual reference to support these findings.

*Statistical Test Outcomes:*

*Pearson Correlation Coefficient:*To assess the relationship between Ireland and France's road freight values, the Pearson correlation coefficient was calculated:The correlation coefficient is 0.914 with a p-value of approximately 1.39.

Interpretation: The high coefficient and low p-value lead to rejecting the null hypothesis (H0) of no correlation. This indicates a significant positive linear relationship between the road freight values of the two countries.

These results suggest a strong association between the road freight activities in Ireland and France, as detailed in Appendix B.

*T-test:* The T-test was utilized to statistically compare the mean road freight values between Ireland and France.The Test Statistics represent T-test statistic: -13.018 and P-value: 5.924. The Findings and Interpretation represent the T-test yielded a test statistic of -13.018 and a very low p-value of 5.924, which is significantly below the typical alpha level of 0.05. This extremely low p-value leads us to reject the null hypothesis (H0) of equal means. Therefore, it is suggested that the mean road freight values for Ireland and France are significantly different.The Conclusion shows this analysis indicates a notable disparity in the scale of road freight activities between Ireland and France, with the means of their road freight values differing significantly.

The detailed output from the T-test and the corresponding code are documented in Appendix B for reference and verification.

*Anova Test:*An ANOVA test was conducted to determine if there are significant differences in the mean road freight values across different years for Ireland:

Test Statistics:

ANOVA Statistic: 0.5589684859133165

P-value: 0.9188937044221509Findings and Interpretation: The ANOVA test produced a statistic of approximately 0.559 and a high p-value of approximately 0.919. The high p-value leads us to fail to reject the null hypothesis (H0). Consequently, this suggests that there are no significant differences in the mean road freight values across different years for Ireland.

Conclusion: This result indicates a relative consistency in Ireland's road freight values over the observed years, with no statistically significant fluctuations in the mean values.

The detailed output from the ANOVA test and the corresponding code are documented in Appendix B for further reference and verification.

*Mann-Whitney U test:*The Mann-Whitney U test was applied to assess if the distributions of road freight values for Ireland and France are significantly different:

Test Statistics:

Mann-Whitney U Statistic: 0.0

P-value: 3.6555663930078464e-22

Findings and Interpretation: The Mann-Whitney U test yielded a statistic of 0.0 and an extremely low p-value of approximately 3.65 . This very low p-value leads us to reject the null hypothesis (H0). Therefore, it is suggested that the distributions of road freight values for Ireland and France are significantly different.

Conclusion: This analysis indicates a substantial disparity between the distribution patterns of road freight values in Ireland and France, pointing to distinct characteristics in their transport sectors.The detailed output from the Mann-Whitney U test and the corresponding code are documented in Appendix B for further reference and verification.

*Kruskal-wallis Test*: The Kruskal-Wallis test was conducted to assess if there are significant differences in the median road freight values across different years:

Test Statistics:

Kruskal-Wallis Statistic: 13.392857142857167

P-value: 0.8598992963240594

Findings and Interpretation: The Kruskal-Wallis test resulted in a statistic of approximately 13.39 and a high p-value of approximately 0.860. The high p-value leads us to fail to reject the null hypothesis (H0). This suggests that there are no significant differences in the median road freight values across different years.

Conclusion: This result indicates a relative consistency in the median road freight values over the observed years, suggesting stable central tendencies in the road freight data.The detailed output from the Kruskal-Wallis test and the corresponding code are documented in Appendix B for further reference and verification.

*Model Predictions:*

*Arima Model:* Before implementing the ARIMA model, the Augmented Dickey-Fuller (ADF) test was conducted to check the stationarity of the time series:

ADF Test Initial Findings: The initial ADF test resulted in a test statistic higher than the critical values and a p-value of approximately 0.489. This suggested that the time series was not stationary.

Differencing: To achieve stationarity, first-order differencing was applied. The subsequent ADF test showed a much lower test statistic and a p-value significantly below 0.05, indicating the transformed data is now stationary.

Optimal ARIMA Model:

Model Selection: Based on the Akaike Information Criterion (AIC), the optimal ARIMA model for both Ireland and France was determined to be ARIMA(0,1,0). This implies no autoregressive (p=0) or moving average (q=0) terms were used, and first-order differencing (d=1) was applied.

Model Interpretation:

Ireland: The actual values show an increasing trend with fluctuations. The ARIMA(0,1,0) model captures the general trend but not all fluctuations, reflecting its simplicity.

France: Similarly, the model captures the increasing trend in France's data but not all year-to-year fluctuations.

The detailed code and visualizations comparing actual versus predicted values for both countries are provided in Appendix C. These visualizations demonstrate the models' ability to capture overall trends despite their simplicity.

*Random Forest Regressor:* The Random Forest Regressor was employed to predict road freight values for Ireland and France. The data was split into training and testing sets, with 50 training samples and 13 testing samples for each dataset.

Model Implementation and Optimization:

GridSearchCV: Implemented to optimize the model's hyperparameters. It systematically tested various parameter combinations and identified the best performing set based on Mean Squared Error (MSE).

Best Parameters: The optimal hyperparameters identified by GridSearchCV for both datasets are documented, alongside their corresponding best scores.

Model Performance:

Ireland:

MSE: Approximately 27,132,965.59. The scatter plot comparing actual vs. predicted values indicates the model generally captures the trend, albeit with some discrepancies likely due to model simplicity and random forest's inherent randomness.

France:

MSE: Approximately 6,544,884,521.20. Similarly, the model captures the overall trend, but larger errors reflect the data's greater scale and variance.

Comparison and Visualization:

Scatter Plots: Visualizations of actual vs. predicted values for both countries highlight how the models capture yearly trends.

Interpretation: While the Random Forest models capture the general trends in road freight values for both countries, limitations exist, and not all data nuances are captured. The higher MSE for France suggests more complex data variability.

The detailed models, code, and visualizations are provided in Appendix C. These findings demonstrate the models' capabilities and limitations in capturing the dynamics of road freight transport in Ireland and France.

**Discussion:**

*Comparative Analysis of Road Freight Transport:*The study's findings reveal distinct differences in road freight transport between Ireland and France. France's road freight sector is notably larger and exhibits a consistent growth trend, indicating expansion over the years. In contrast, Ireland's more variable year-to-year freight values suggest a higher sensitivity to economic or other external factors.

*Summary and Interpretation of Statistical Tests:*T-test and Mann-Whitney U Test: Both tests confirmed significant differences in road freight values between Ireland and France, underscoring the disparities in scale and operations of the transport sectors in the two countries.ANOVA and Kruskal-Wallis Test: These tests suggested that the mean and median road freight values in Ireland do not significantly vary across different years, indicating relative stability within the country's freight transport sector.Pearson Correlation: A strong positive correlation between the road freight values of Ireland and France was identified, suggesting that despite their differences, some common factors might be influencing road freight transport in both countries.

*ARIMA Model Analysis:*Both ARIMA models for Ireland and France (0,1,0) indicate that first-order differencing was necessary to achieve stationarity. However, the lack of additional AR or MA terms suggests that while the models capture the broad trends, they might not fully reflect more intricate patterns or seasonal effects present in the data.

*Random Forest Model Insights:*The Random Forest models generally captured the overarching trends in road freight values for both countries. However, their simplicity means they might not fully grasp all the nuances and complexities of the data, a common limitation of predictive models.

**References and citation:**

**Links:** <https://stats.oecd.org/#>

<https://www.oecd.org/termsandconditions/>

<https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>

<https://www.ibm.com/topics/exploratory-data-analysis>

<https://www.sagepub.com/sites/default/files/upm-binaries/33663_Chapter4.pdf>

<https://arxiv.org/pdf/1407.7502.pdf>

<http://www.python.org>

<https://pandas.pydata.org/docs/getting_started/intro_tutorials/index.html>

<https://www.researchgate.net/publication/243644934_Distribution_of_the_Estimators_for_Autoregressive_Time_Series_With_a_Unit_Root>

<http://seismo.berkeley.edu/~kirchner/eps_120/Odds_n_ends/Students_original_paper.pdf>

<https://link.springer.com/article/10.1023/A:1010933404324>

<https://seaborn.pydata.org/>

[Documentation — IPython](https://ipython.org/documentation.html?v=20240107111106)

[NumPy reference — NumPy v1.26 Manual](https://numpy.org/doc/stable/reference/?v=20240107111106)

[Using Matplotlib — Matplotlib 3.8.2 documentation](https://matplotlib.org/stable/users/index.html)

Dickey, D. & Fuller, Wayne. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. JASA. Journal of the American Statistical Association. 74. 10.2307/2286348.

Student (1908). The Probable Error of a Mean. Biometrika, 6(1), 1-25.

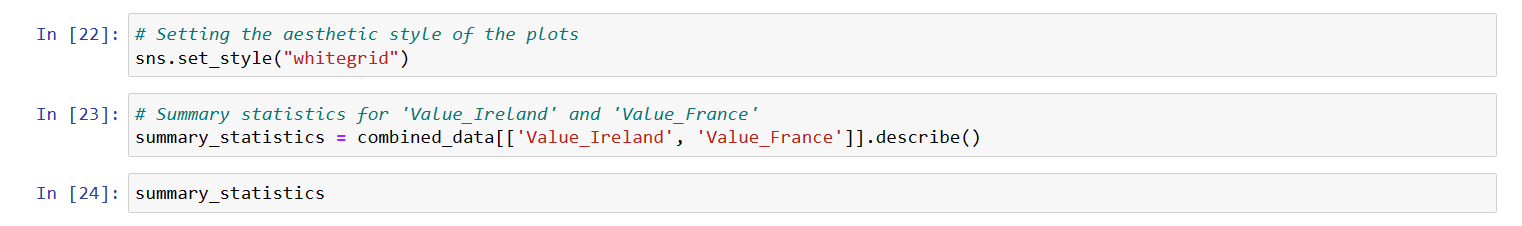
Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

**Appendix**

*Appendix A:*

Satistical Summary:

|  | **Value\_Ireland** | **Value\_France** |
| --- | --- | --- |
| **count** | 63.000000 | 63.000000 |
| **mean** | 8726.650794 | 117753.142857 |
| **std** | 4549.934874 | 66315.183134 |
| **min** | 2156.000000 | 20760.000000 |
| **25%** | 4787.500000 | 31604.500000 |
| **50%** | 8849.000000 | 145293.000000 |
| **75%** | 11886.500000 | 165930.000000 |
| **max** | 19146.000000 | 207025.000000 |



Summary Statistics:

Value\_Ireland:

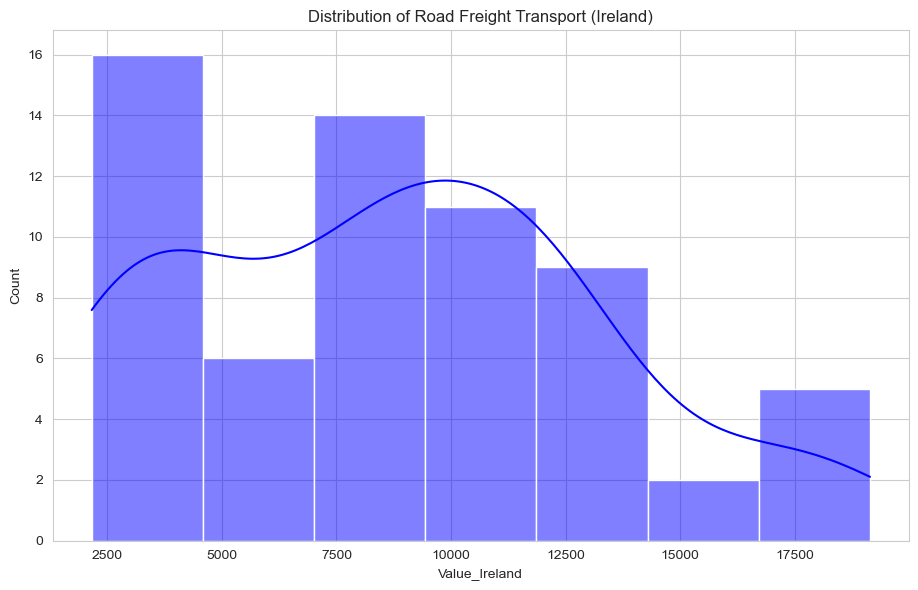
* + Mean: Approximately 8,727 million tonnes-kilometres.
  + Standard Deviation: Approximately 4,550, indicating variability in the data.
  + The median (50th percentile) is about 8,849, which is quite close to the mean, suggesting a fairly symmetrical distribution.

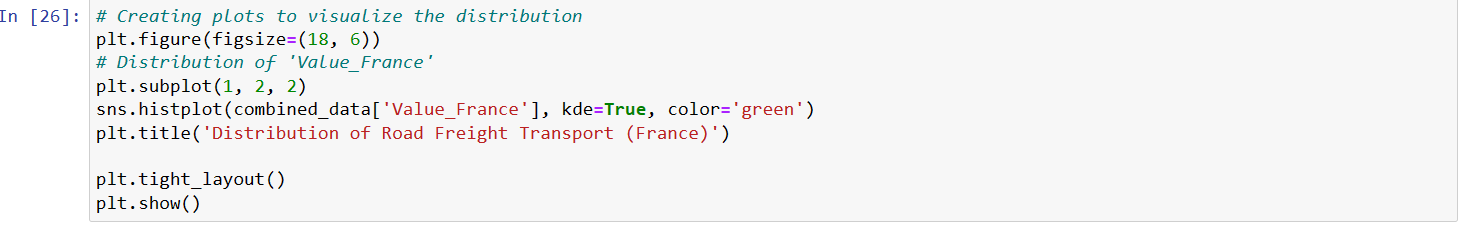
Value\_France:

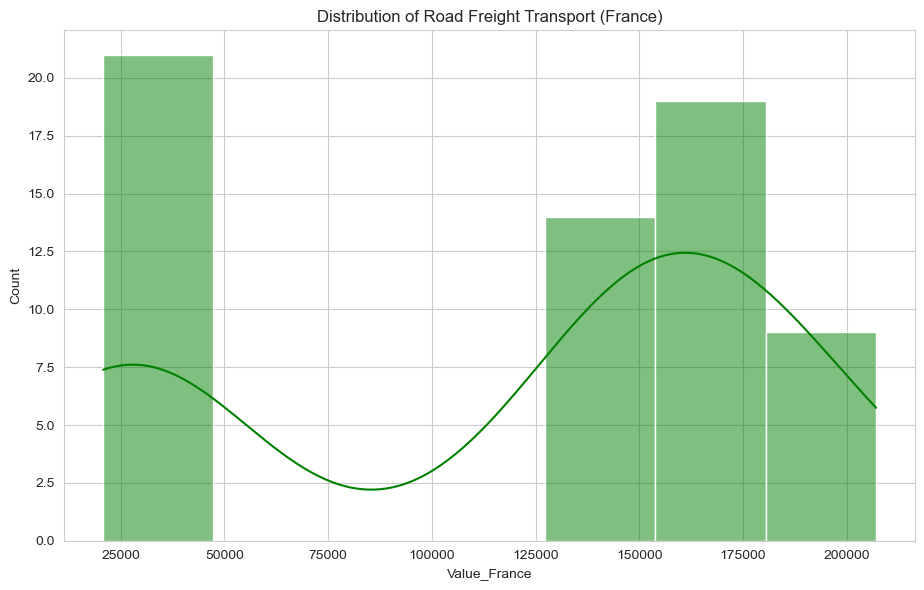
* + Mean: Approximately 117,753 million tonnes-kilometres, significantly higher than Ireland, indicating more road freight transport.
  + Standard Deviation: Approximately 66,315, indicating a wide variability in the data.
  + The median is significantly higher than for Ireland, reflecting the larger scale of road freight transport in France.

*Appendix B:*Histogram Visualization:

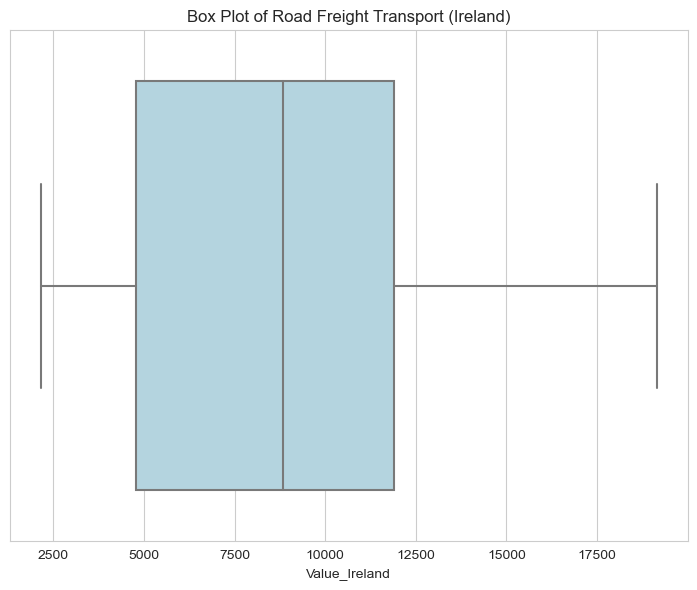


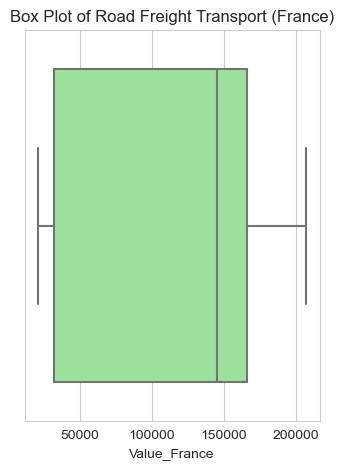




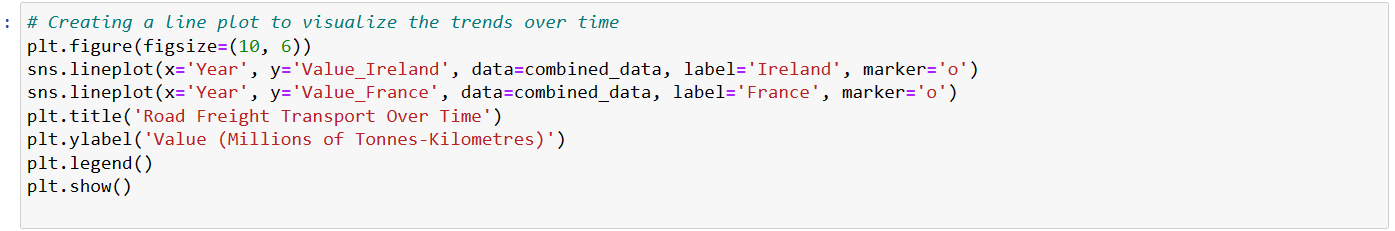


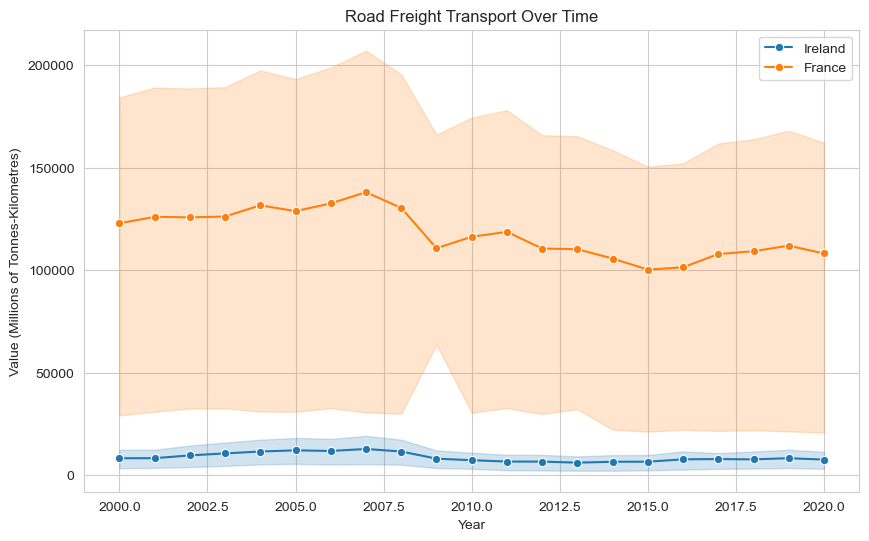
Box Plot Visualization:





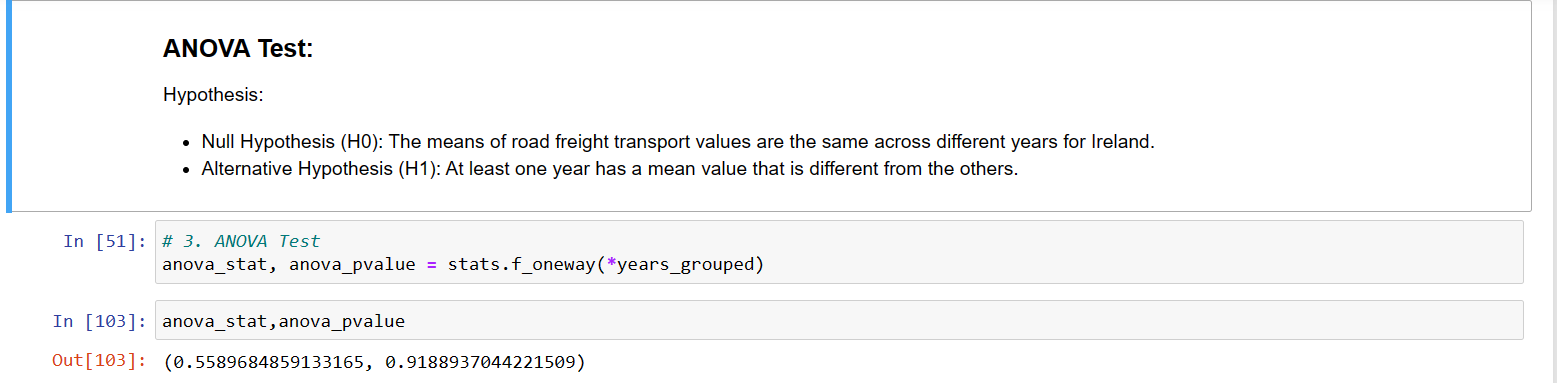
*Trend Analysis:*

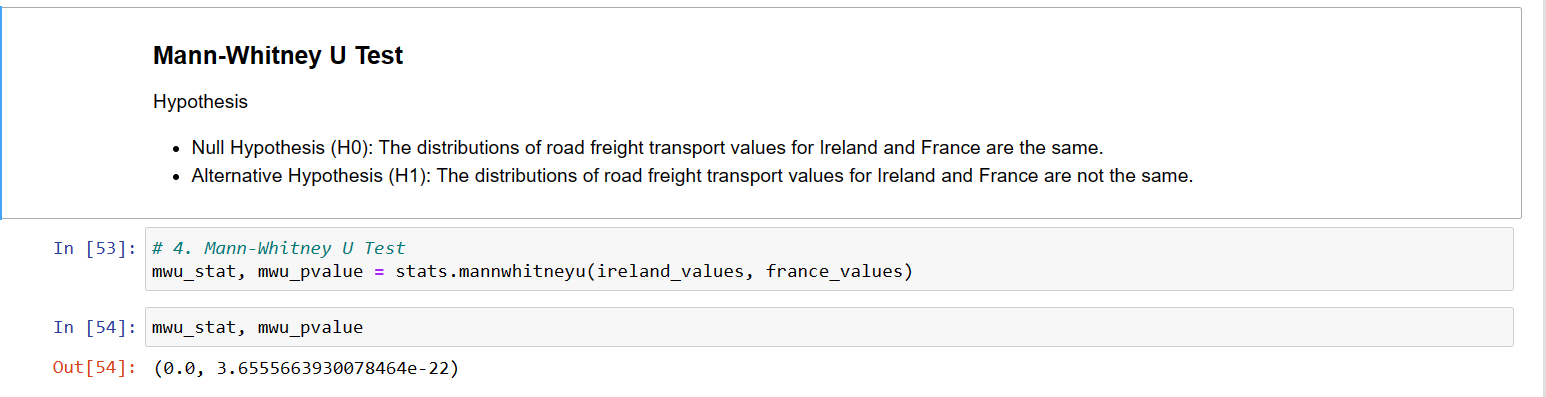


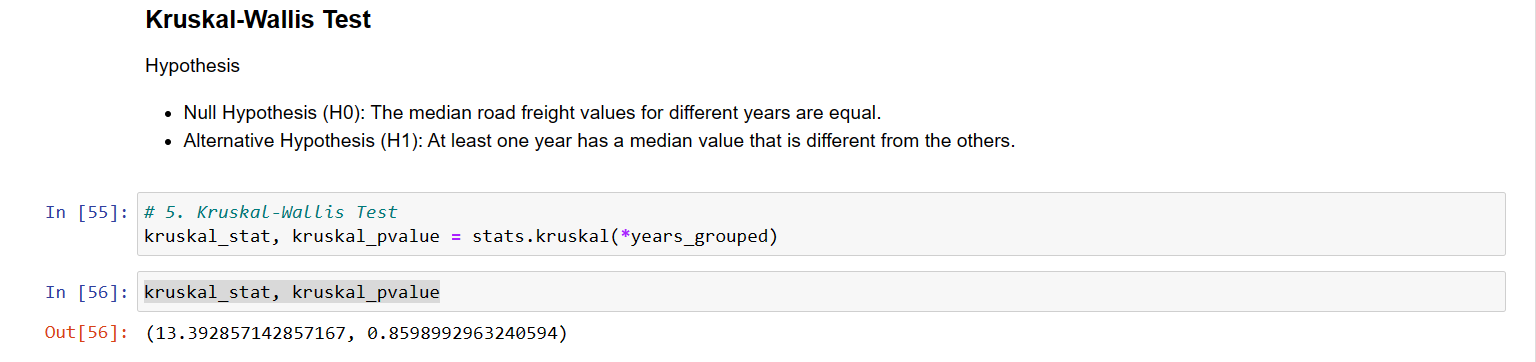


*Codes and output of inferential statistics:*





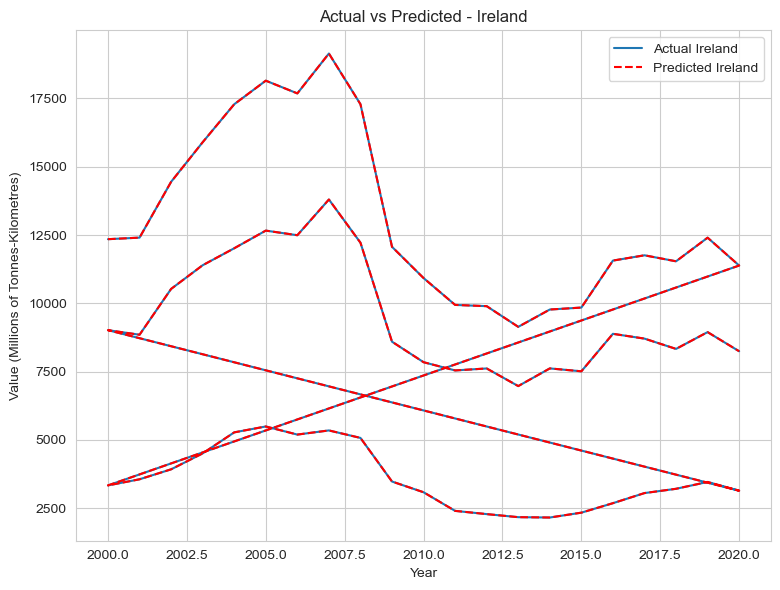


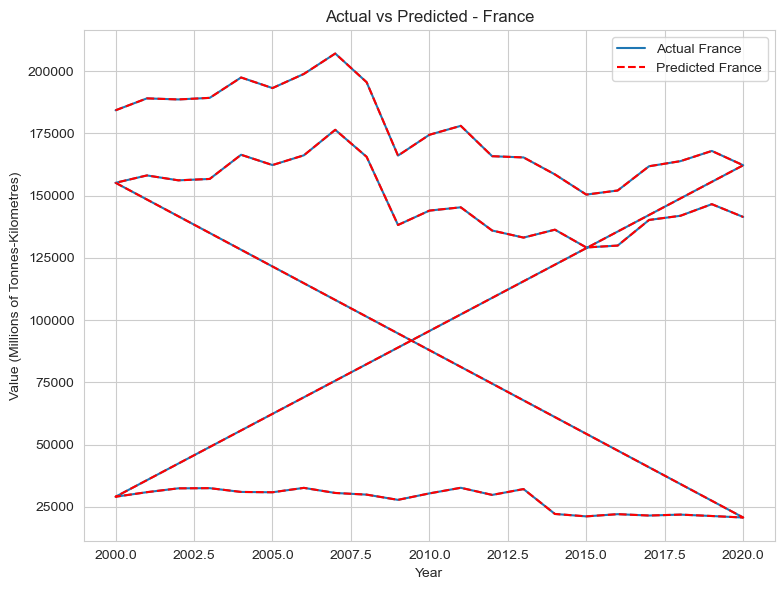




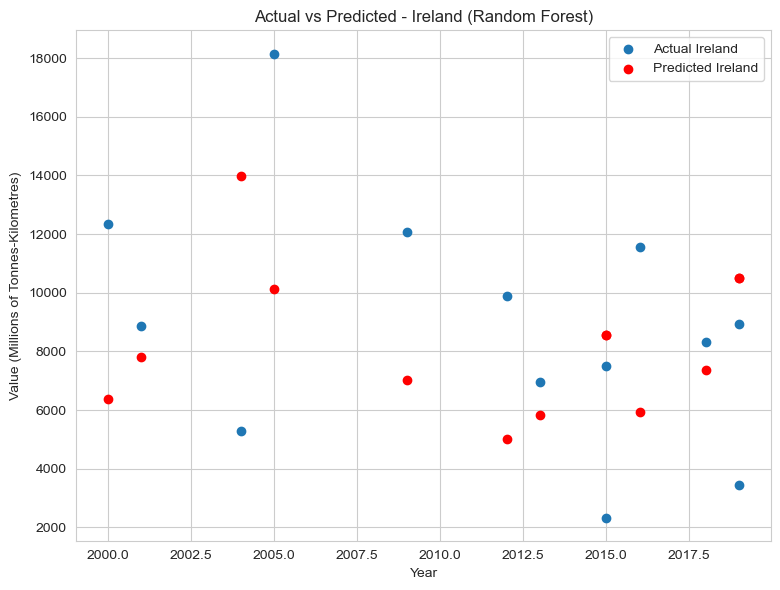
*Appendix C*

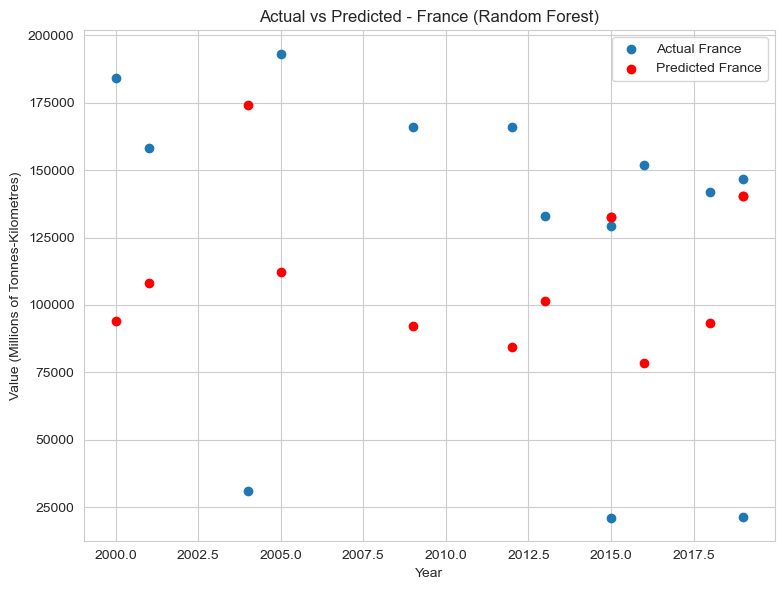
*Arima Model Visualization:*





Random Forest Regressor Visualization:





*Appendix D:*

*Visualization of heatmap:*





### **Correlation Analysis Summary:**

The Pearson correlation coefficient between the road freight values of Ireland and France is approximately 0.914, indicating a strong positive linear relationship. This suggests that as road freight values increase in one country, they tend to increase in the other and vice versa. The heatmap visually reinforces this with a high correlation value.