Attempting Random Forest Model

In this section, we will attempt to use a Random Forest model to predict various aviation-related variables based on the number of new COVID-19 cases. The variables we will analyze include:

Aircraft movements/Local flights (number)

Aircraft movements/Cross-country flights (number)

Commercial air traffic/Passengers/Total passengers/Total passengers (number)

Commercial air traffic/Cargo/Total cargo/Total cargo (ton)

Commercial air traffic/Mail/Total mail/Total mail (ton)

We will split the data into training and testing sets, fit the Random Forest model, and evaluate its performance using metrics such as Mean Squared Error (MSE) and R-squared (R²). Additionally, we will visualize the actual vs. predicted values for each variable.

Result：

Test Mean Squared Error: 215993945235.1363

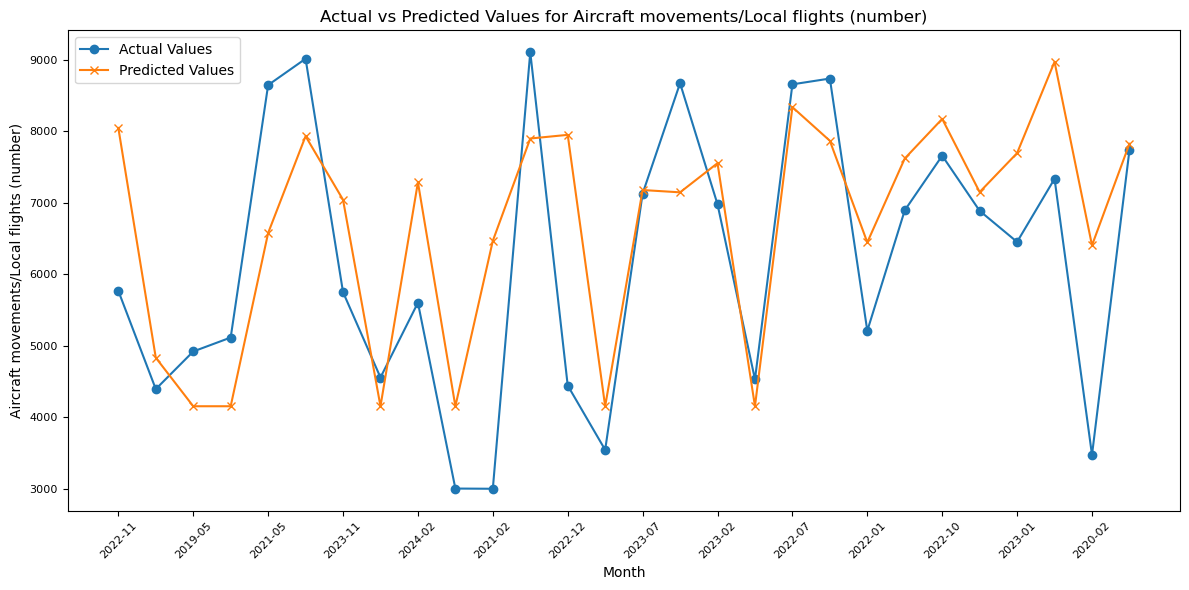
Test R^2 Score: 0.6337518052163386

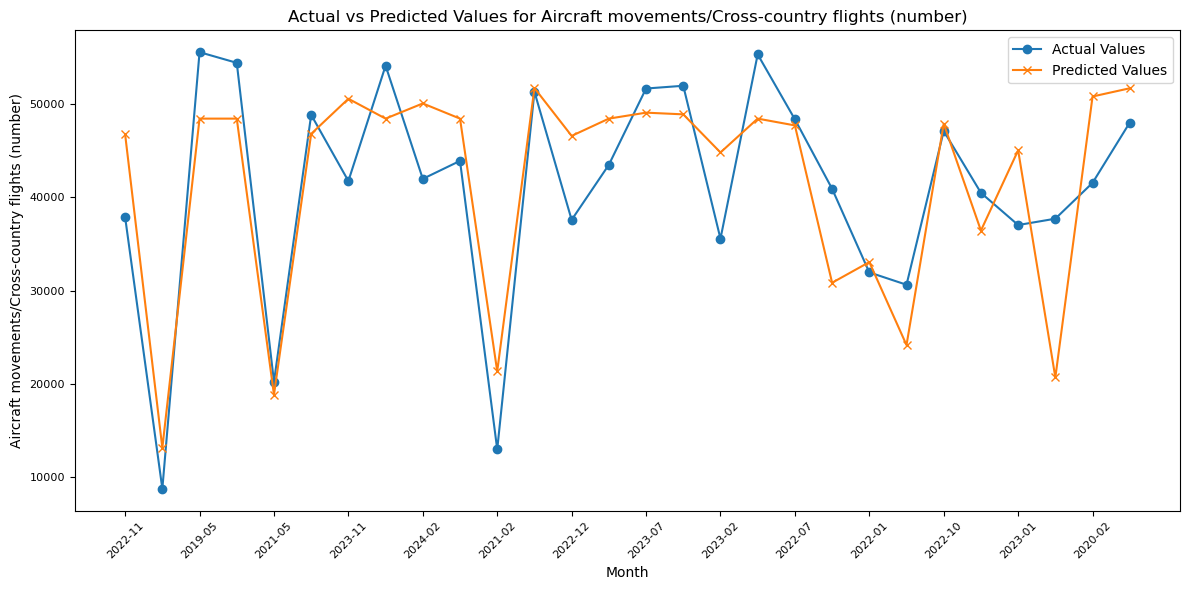
Feature Importance

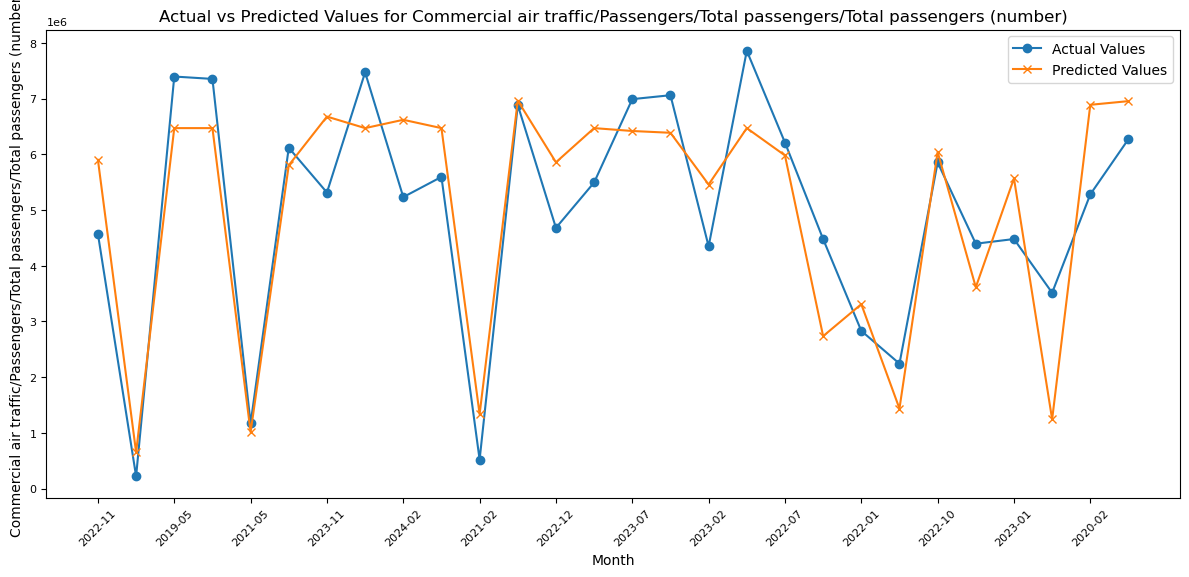
0 new\_cases 0.028951

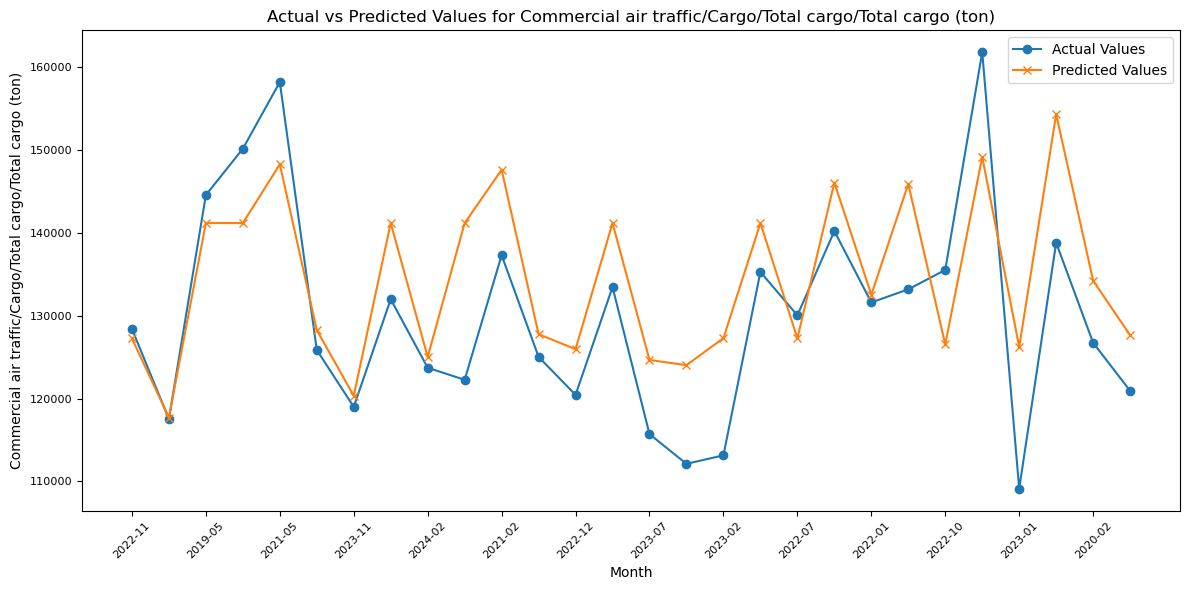
1 monthly\_deaths 0.839350

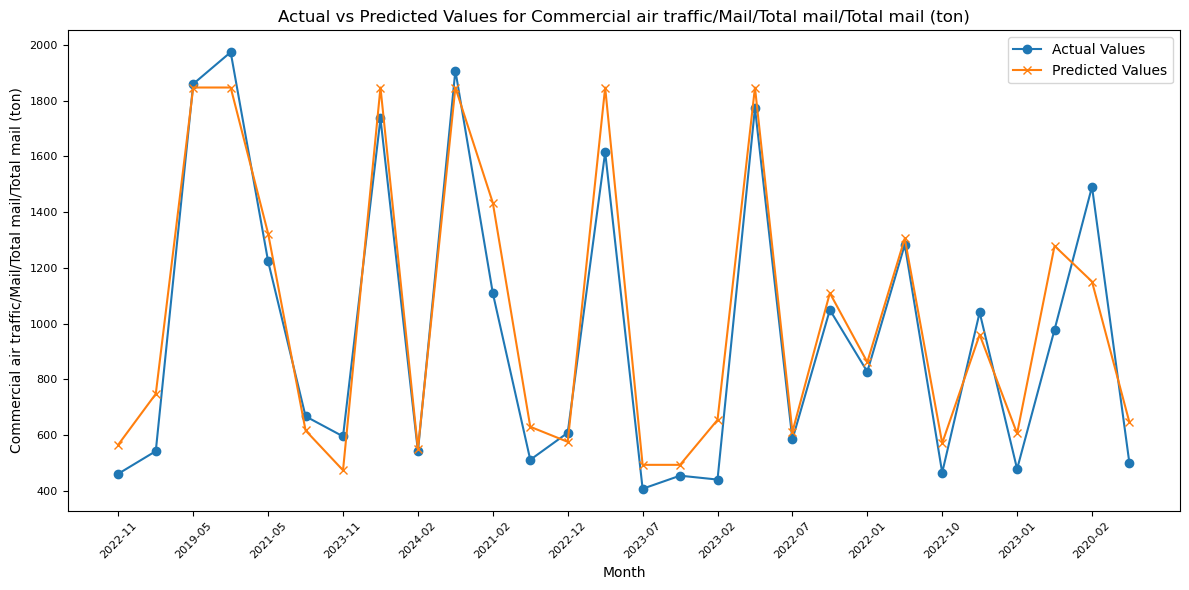
2 monthly\_vaccinations 0.131699











在上述随机森林模型中，R square=0.6337，这表示模型能够解释约63.37%的变异性，有一定的预测能力，但并不是非常理想。而模型的MSE = 215993945235.1363，MSE值非常大，表明模型在测试集上的预测误差较高。这些数据说明模型能够捕捉到一些规律，但是这三个疫情变量并不足以完全解释航空业的波动。从模型的特征重要性输出结果来看，monthly\_deaths（每月死亡人数）的重要性非常高，达到了0.8393。这说明死亡人数是对航空相关变量影响最大的疫情因素。而 new\_cases（新增病例数）的影响较小，仅为0.0289。通过图像对比可以帮助发现模型在部分时间段表现较好，其它时间段存在较大偏差。对此，我们可能需要引入更多的变量（例如封锁政策指数、疫苗接种覆盖率、人口流动性数据等）来改善模型性能。

The Random Forest model shows an R² value of 0.6337, indicating that the model explains about 63.37% of the variance. While it demonstrates some predictive ability, it is not highly accurate. The model's MSE is 215,993,945,235.1363, which is quite large, signifying a high prediction error on the test set. These results suggest the model captures some patterns, but the three pandemic variables used are insufficient to fully explain the fluctuations in the aviation industry.

Feature importance analysis shows that `monthly\_deaths` has the highest influence, with an importance score of 0.8393, indicating that death counts significantly affect aviation variables. In contrast, `new\_cases` has a much smaller impact, with an importance of only 0.0289.

Visual comparisons reveal that the model performs well during certain periods, but shows larger discrepancies in others. To improve performance, it may be necessary to include additional variables, such as lockdown policy indices, vaccination coverage, and mobility data, to better capture the factors influencing aviation trends.