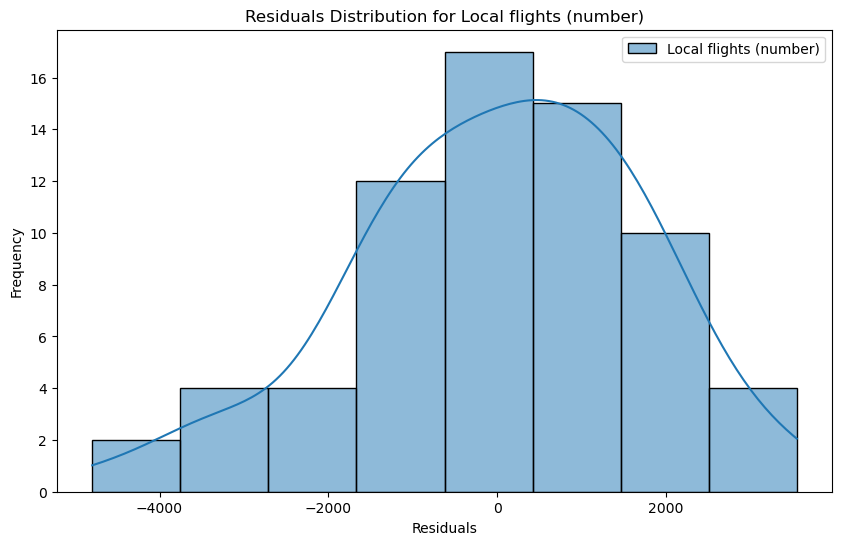
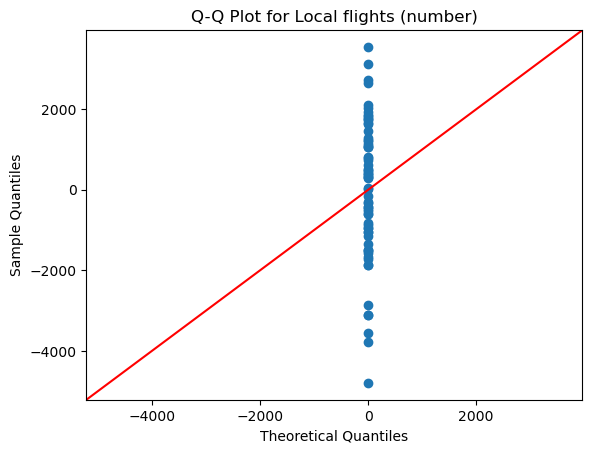
Multi-linear Regression

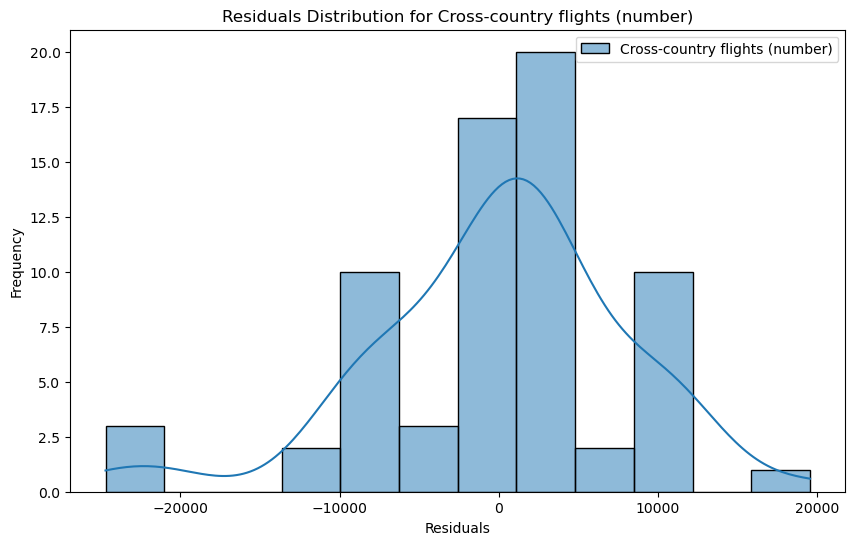


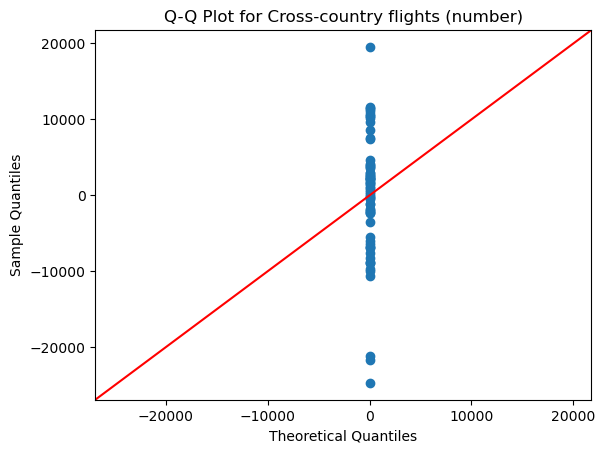
The first two graphs show the residuals for the number of local flights and the quantiles of the residuals versus a theoretical normal distribution. Despite the residuals clustering around zero, the spread and presence of outliers suggest that the model may not capture some of the variability in local flights data. The low R² (R²=0.324) also supports the insignificant relationship between local flights number and the variables.



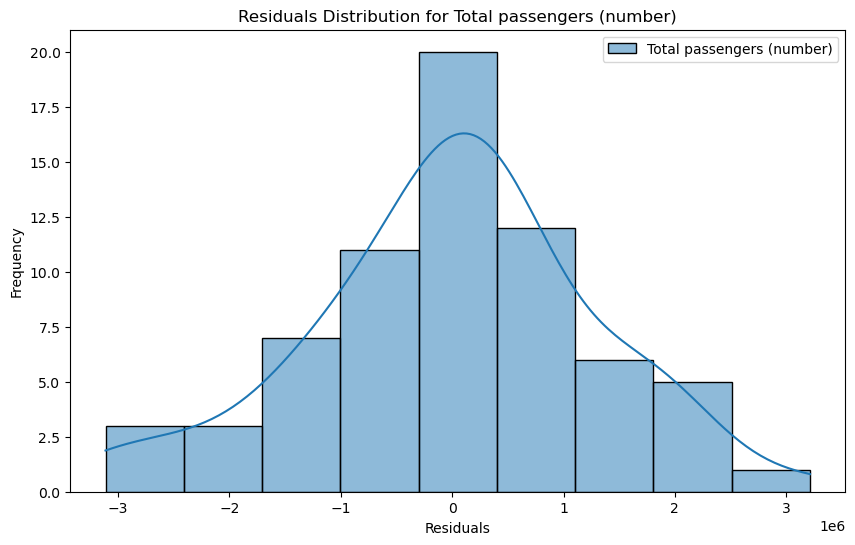


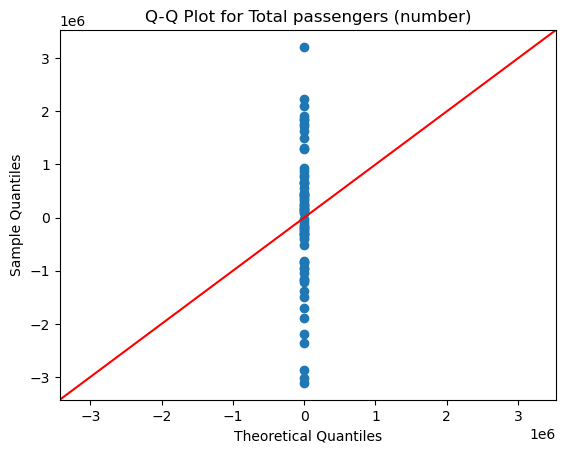
The following graphs, along with the R² value of 0.621, indicate a stronger relationship between cross-country flights number and COVID-related variables compared to domestic flights. This result is logical considering the Netherlands' significant role in international air traffic.



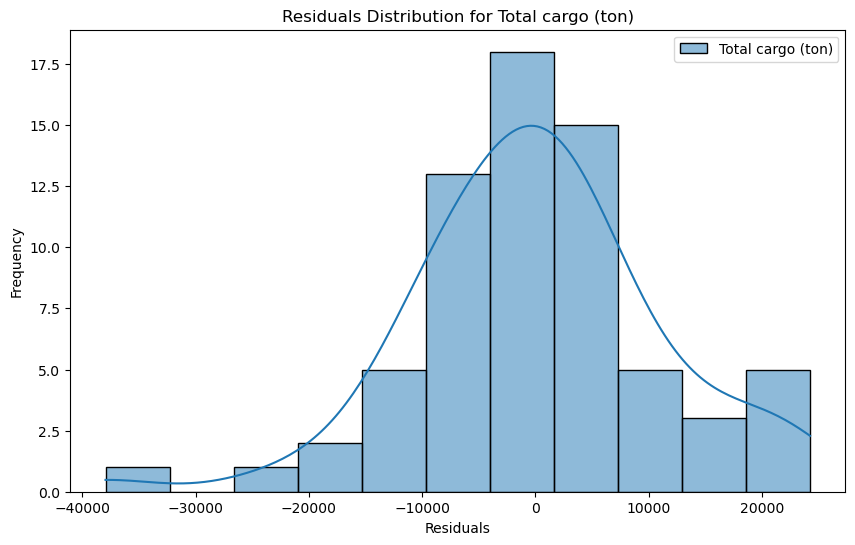


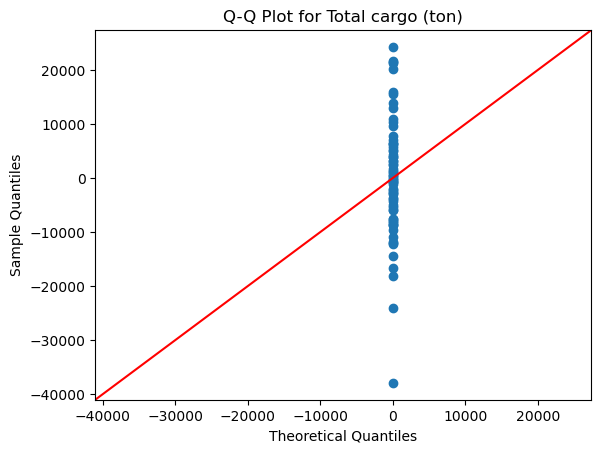
The following graphs illustrate that total passenger numbers are closely related to COVID-related variables, as reflected by an R² value of 0.689. This relatively strong correlation suggests that the pandemic has had a notable impact on passenger aviation, with fluctuations in COVID-19 cases, deaths, and vaccinations likely influencing travel demand.



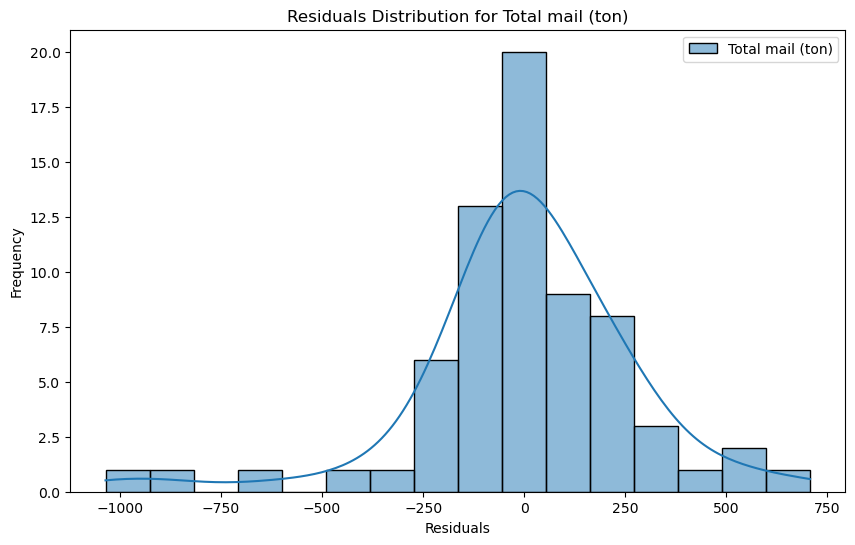
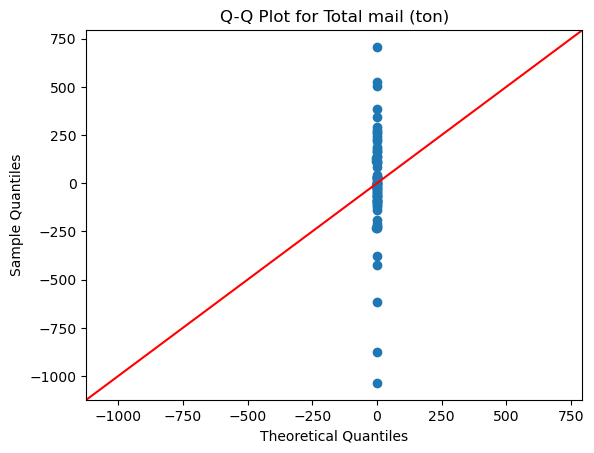


In comparison, the following figures show that total cargo has a much weaker relationship with COVID-related indicators, also indicated by an R² value of 0.349. This suggests that cargo transport has been less affected by pandemic factors such as case numbers, deaths, and vaccinations. The lower correlation may be due to the essential nature of cargo transport, which likely continued relatively unaffected during the pandemic to meet global supply chain demands, regardless of fluctuations in COVID-19 variables.





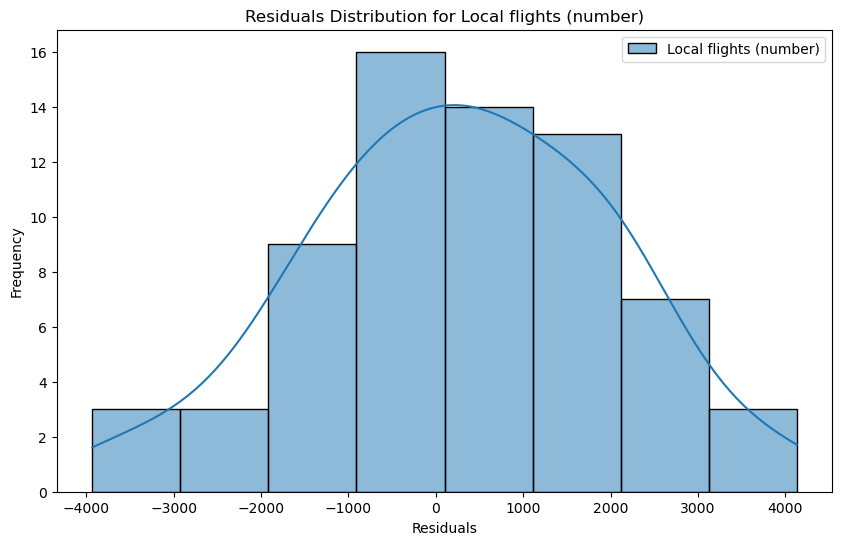
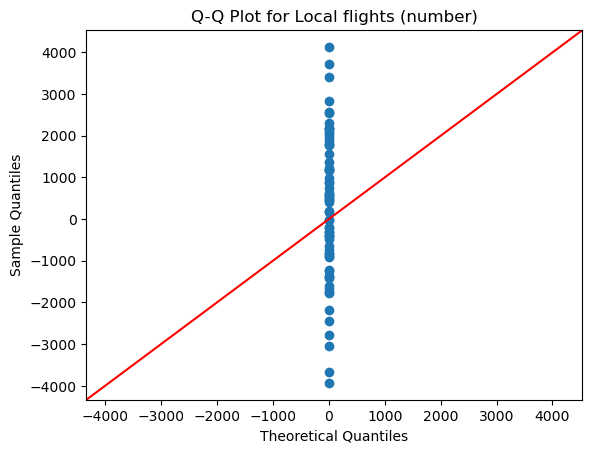
The following graphs reveal a significant relationship between total mail and COVID-related variables, demonstrated by an R² value of 0.743. This correlation indicates that the pandemic has an influence on mail volumes, likely due to shifts in consumer behavior and increased reliance on delivery services during lockdowns.

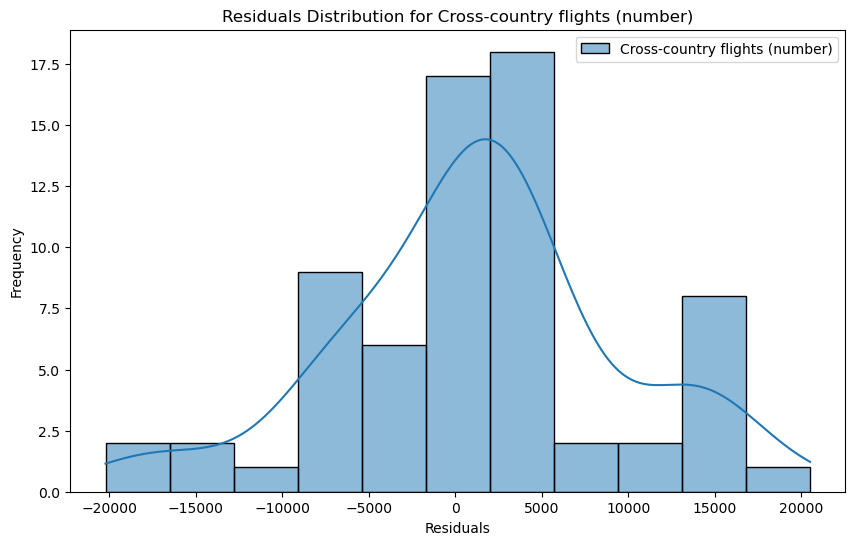
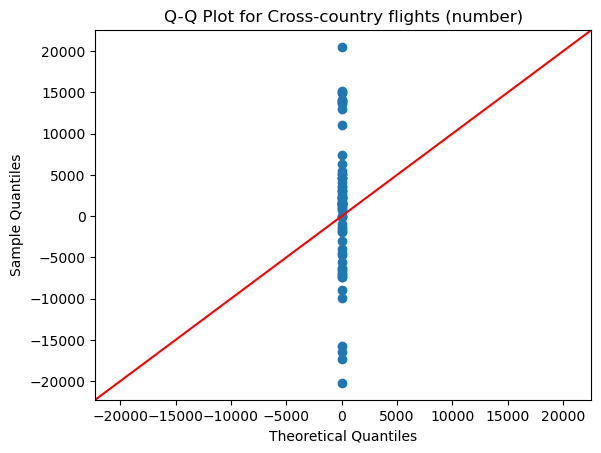
To summarize this section, while the multiple linear regression model performs reasonably well for predicting some variables, it falls short in capturing the relationships for others. This suggests that a direct linear approach may not fully reflect the complexities within the data. To address this, we plan to apply some transformation to the variables and investigate whether this approach can reveal stronger or more linear relationships.

In this section, we apply a log transformation to the variables to see whether this approach can reveal stronger or more linear relationships. By transforming the data, we aim to enhance the model's accuracy and better account for potential nonlinearities, leading to more robust predictions and deeper insights.

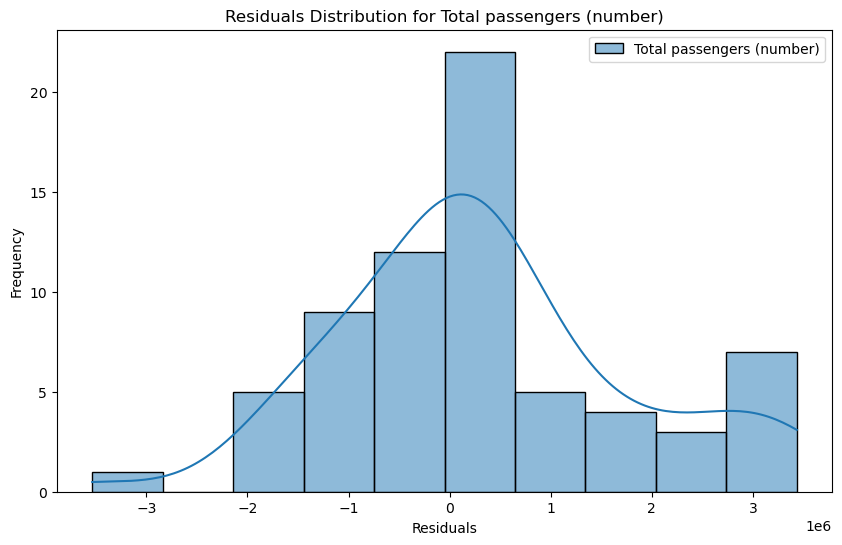
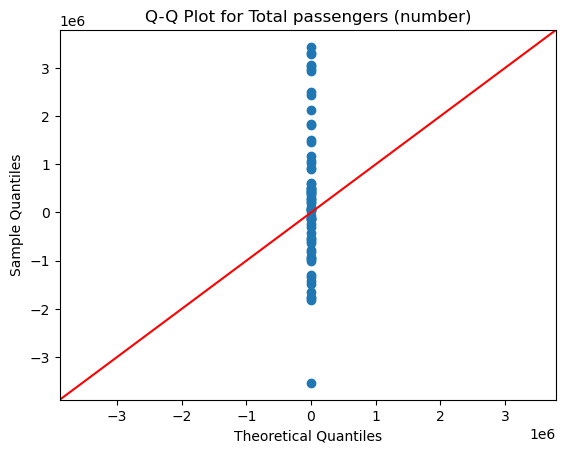
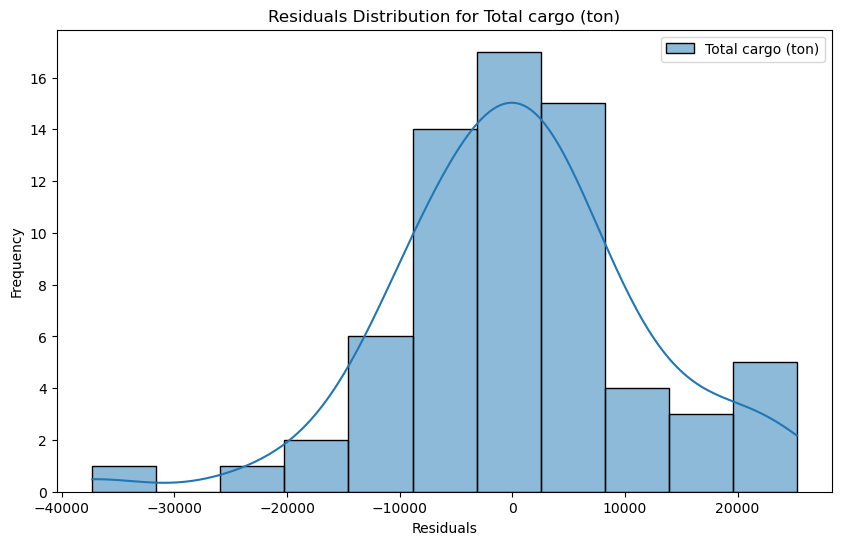
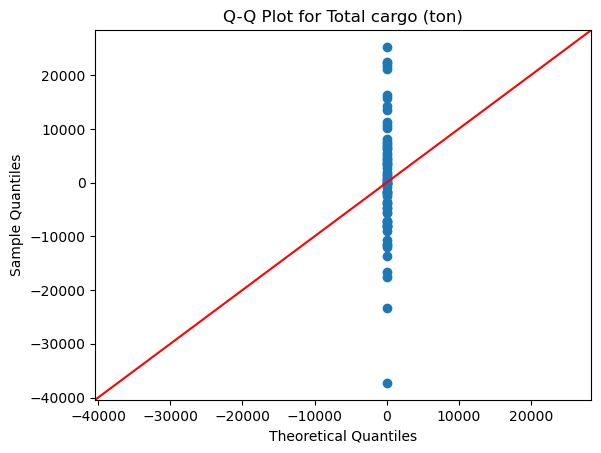
In the following two graphs, the log transformation does not appear to improve the relationship between local flights and the variables. This conclusion is supported by a low R² value of 0.305, indicating that the log transformation fails to uncover a stronger connection. The result suggests a weak or nonexistent relationship between local flights and the COVID-related factors.

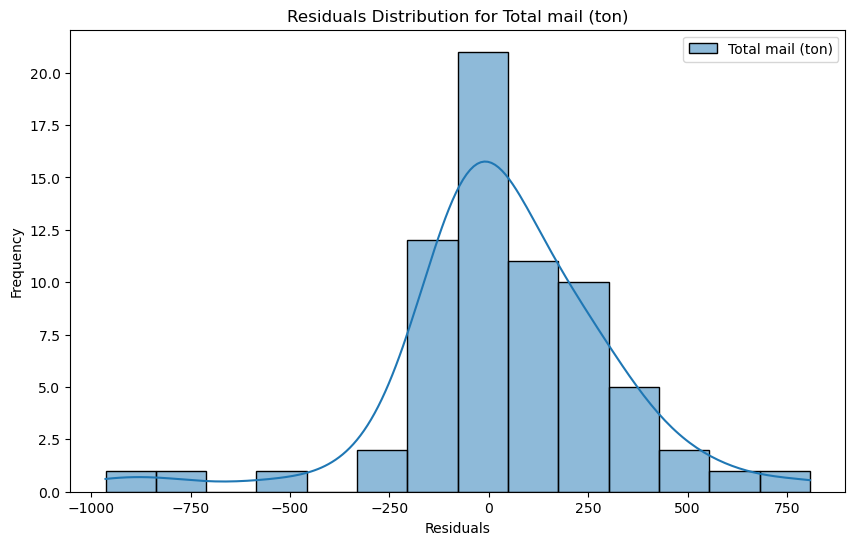
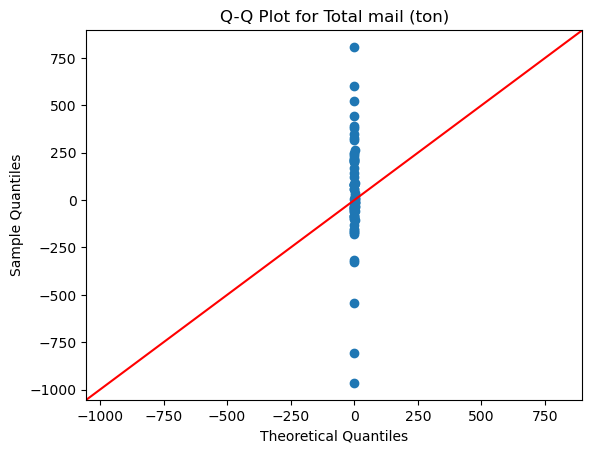
Similarly, the log transformation does not appear to enhance the relationship between cross-country flights and the variables. This is indicated by the relatively low R² value of 0.528, which is lower than that of the original model. This finding suggests that the transformation failed to improve the model's ability to capture the underlying relationship, highlighting the limitations of both the original and transformed models in this context.

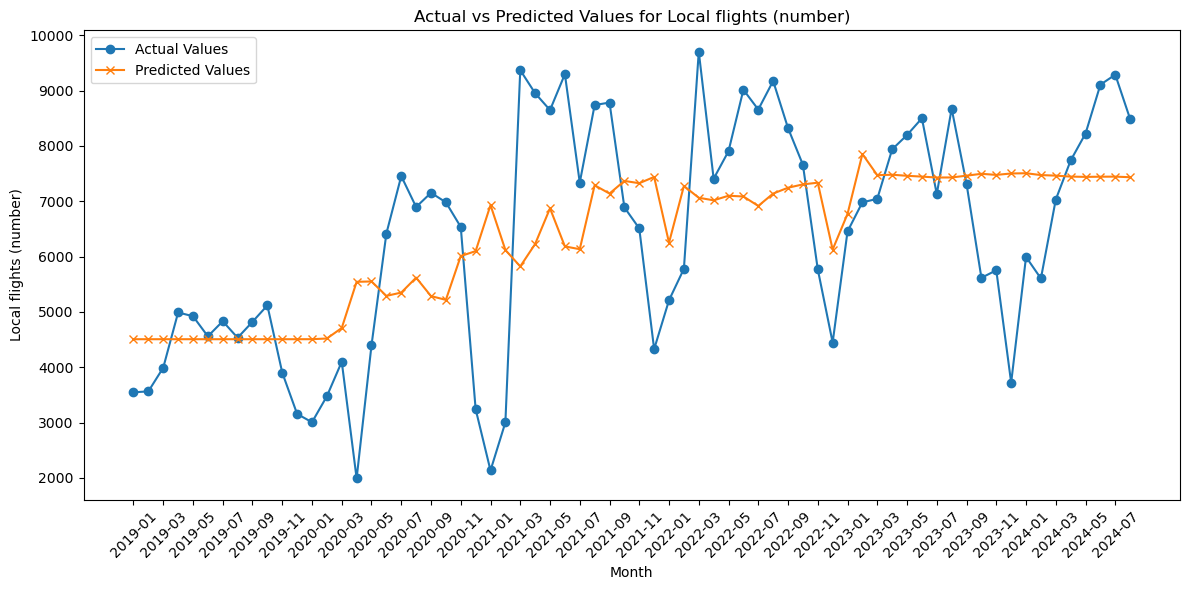
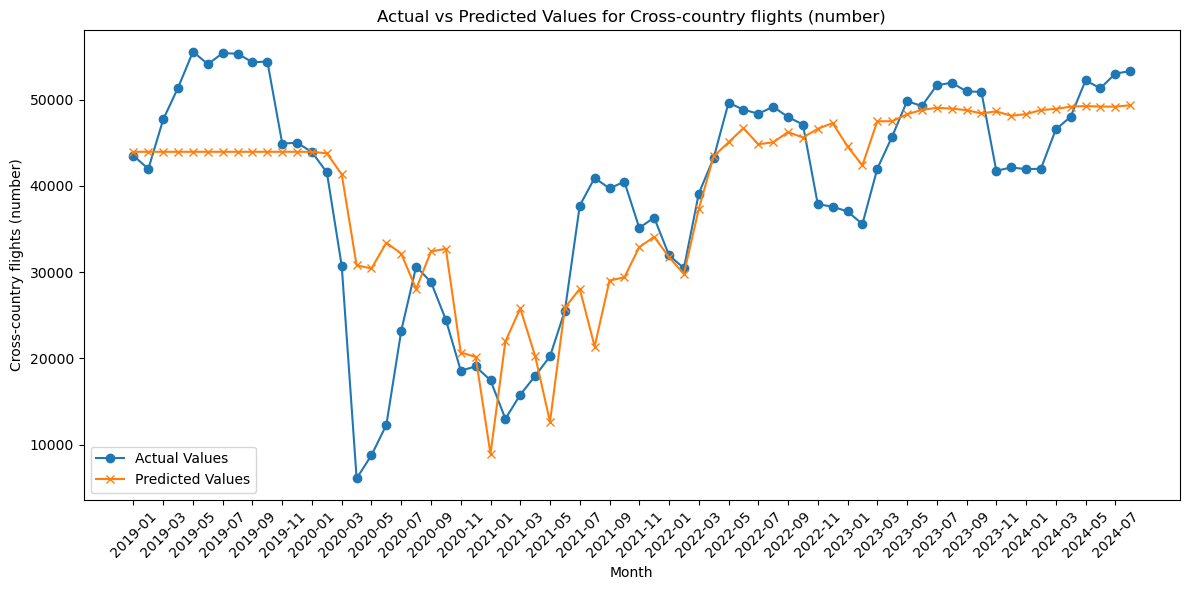
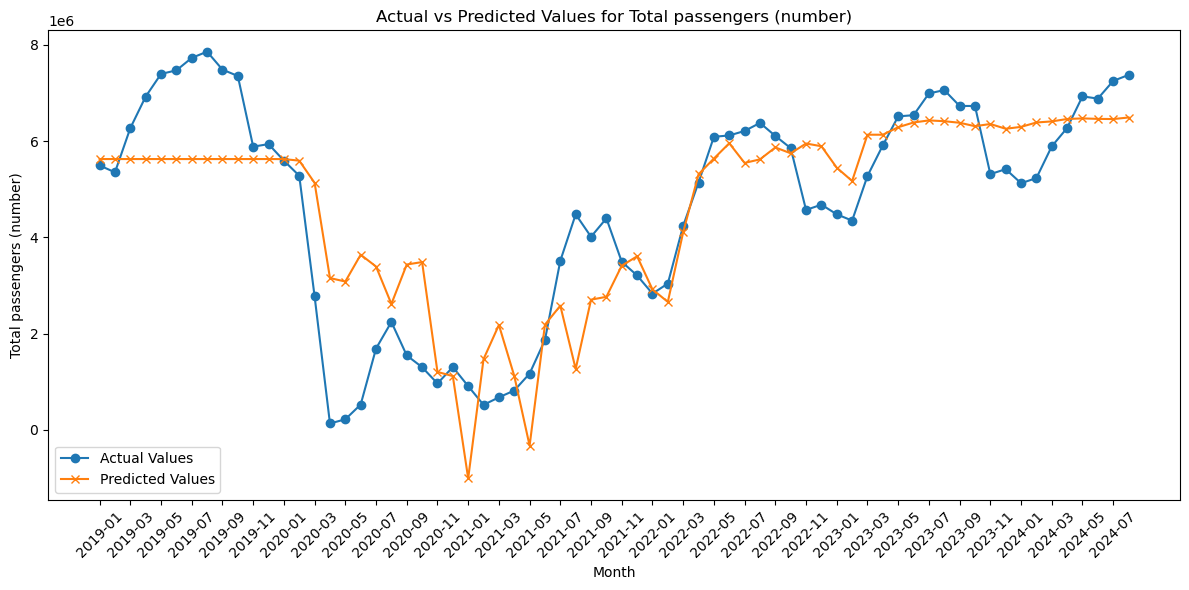
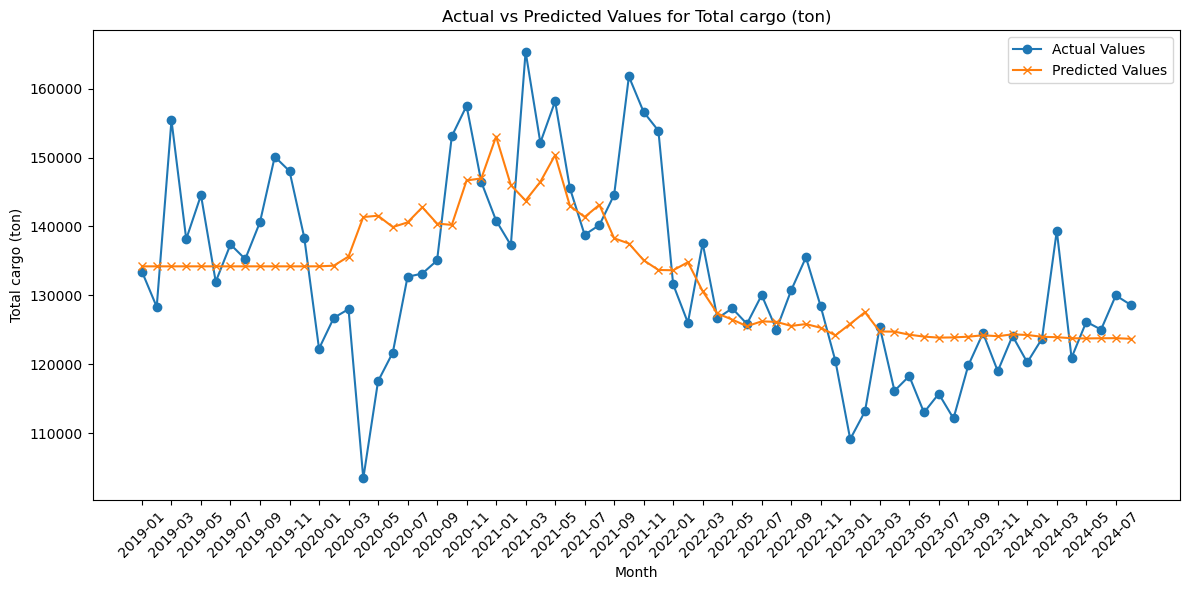
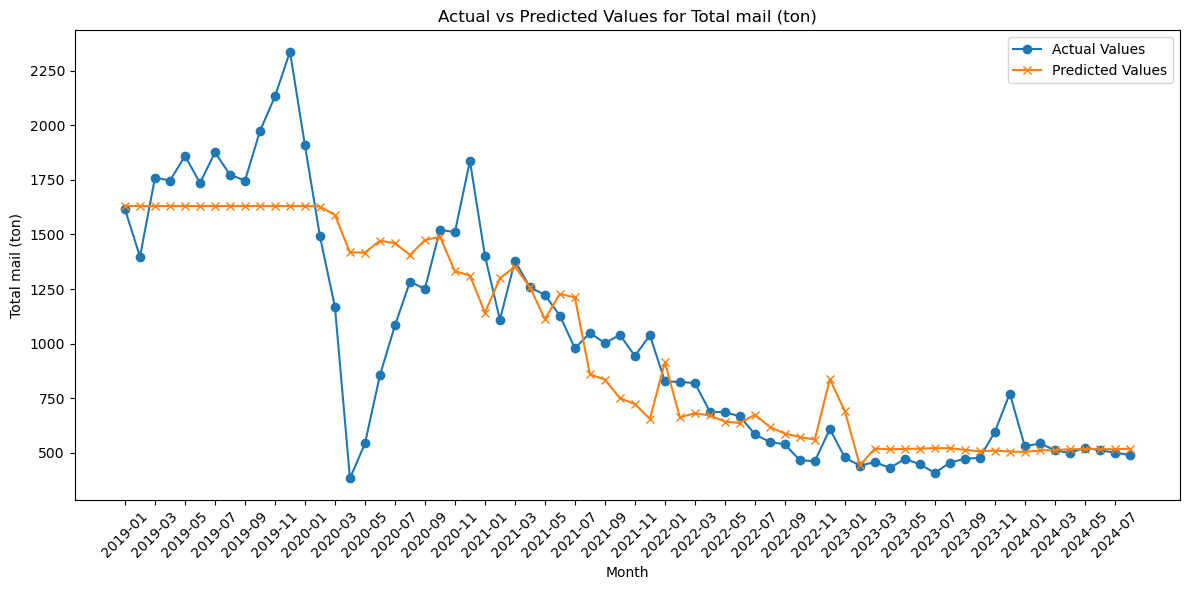
The same situation applies to the relationships between total passengers and total cargo with the variables. The R² values are 0.561 and 0.339, respectively, both of which are lower than those of the original models. This indicates that the log transformation did not improve the model's ability to capture the underlying relationships between total passengers, total cargo, and the COVID-related factors.

The relationship between total mail and the variables stands as the only exception. With an R² value of 0.750, which is slightly higher than that of the original model, this improvement may be attributed to the close relationship between total mail and the COVID-related factors.

Based on these results, the log transformation does not lead to any improvement in the model's performance. As a result, we have decided to discontinue its use. Finally, we present the actual vs. predicted value graphs for the five indicators, illustrating the model's predictions compared to the observed data for each variable.

The findings in this section reveal that while multiple linear regression offers some insights to predict aviation indicators, it does not accurately capture the relationship with COVID-related variables. To address this limitation, we plan to apply alternative predictive models, such as random forest model, to further investigate the connection between aviation performance and the pandemic. This change in methodology aims to provide a more comprehensive understanding of the relationships and improve prediction accuracy.