MM957 - Data Analytics in R

Part 4: Summary of Results

Univesrity of Strathclyde

Introduction

As the demand for tailored travel experiences continutes to rise, it is crucial to understand the climatic patterns fo potential cities. Tourists today make decision based on what they believe the climatic conditions of a destination are (Becken, 2010). This necessities the need for a predictive model that can estimate average tempeartures based on various geographic and meteorological variables.

This report aims to guide the agency in selecteing the most impactful varaibles for temperature prediction, ensuring confidence in the model's accuracy. Previous research in this area has included a range of models, (Chinchwad, 2019) investigated the use of a Time Series ARIMA model which was contary to prevopuus machine learning classification models.

This report focuses ona. Linear Regression model using R, enchanced with variable transformation and optimisation. This could offer improvements in predicitive accuracy and model robustness over previous models, as demonstracted by (James, Witten, Hastie, & Tibshirani, 2021).

Methodology

This section outlines the methods used to assist a European travel agency in selecting US cities for package holidays based on average temperature predictions.

1. **Variable Selection for Temperature Prediction:** The impact of each selected variable is visualised using coefficient plots created with ggplot2.
2. **Model Confidence Validation:** Employed 10-fold cross-validation via the caret package to assess model reliability.
3. **Coastal Proximity Analysis:** Analysed the influence of distance from the coast on temperature by calculating their coefficient.
4. **Temperature Prediction for Specific Locations:** Predicted the temperature for a specified city in Florida and compared it to the state's average.
5. **Evaluation of Potential New Destinations:** Assessed Springfield, Ohio as a potential destination by predicting its average temperature and it meets the agency's minimum temperature criterion of 55 degrees Fahrenheit using confidence intervals.

Analysis

Based on the boxplot analysis (Figure 1), the travel agency should prioritize Average Annual Precipitation, Latitude, and Elevation as key variables for predicting average temperature. Wind speed,

with its wide confidence interval crossing zero, shows uncertain impact on temperature, while Distance to Coast has a minimal negative coefficient, suggesting a less significant effect.

The model demonstrates strong predictive accuracy and stability, evidenced by an R-squared of 0.897, indicating it explains nearly 90% of the variance in temperature, and low error metrics (RMSE of 3.056 and MAE of 2.189).

The graph (Figure 2) and statistical outputs indicate a significant negative relationship between distance to the coast and average temperature, showing that cities farther from the coast tend to have lower temperatures, as evidenced by the negative coefficient and the low p-value (0.0015).

The predicted average temperature for a city in Florida is 61.89°F, which is 5.46°F cooler than the state average of 67.34°F, indicating that this city may be slightly cooler than the typical Florida location but still within a reasonable margin of error. In contrast, Springfield, Ohio, with a predicted average temperature of 46.61°C and a confidence interval not encompassing 55°C, does not meet the travel agency's minimum temperature criterion, thus it should not be considered as a potential travel destination.

Conclusion

This report provides key insights for selecting travel destinations based on climate. It recommends prioritizing Average Annual Precipitation, Latitude, and Elevation, as these variables significantly impact and enhance the model's accuracy, evidenced by a high R-squared of 0.897 and low error metrics (RMSE of 3.056 and MAE of 2.189). Analysis also shows that cities farther from the coast are cooler, supported by data indicating a strong negative correlation between distance to the coast and temperature. The model's high accuracy is highlighted by the example of a city in Florida, which was only 5.46°F cooler than the state's average temperature. However, Springfield, Ohio, does not meet the agency's minimum temperature requirement of 55°F, making it unsuitable as a destination.

Given these results, the agency can confidently use the model to evaluate potential destinations. It is advised that the agency continuously refines data collection and adjusts the model to ensure precise predictions and to accommodate local microclimates, catering to specific customer preferences. Regular updates and strategic use of climatic data will further enhance the agency's offerings and customer satisfaction.

References

Becken, S. (2010). *The Importance of Climate and Weather for Tourism.* Brisbane: Land Environment & People.

Chinchwad, P. (2019). *Weather Prediction for Tourism Application using ARIMA.* Dehli: International Research Journal of Engineering and Technology (IRJET) .

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* . Los Angeles: Springer.

Appendices

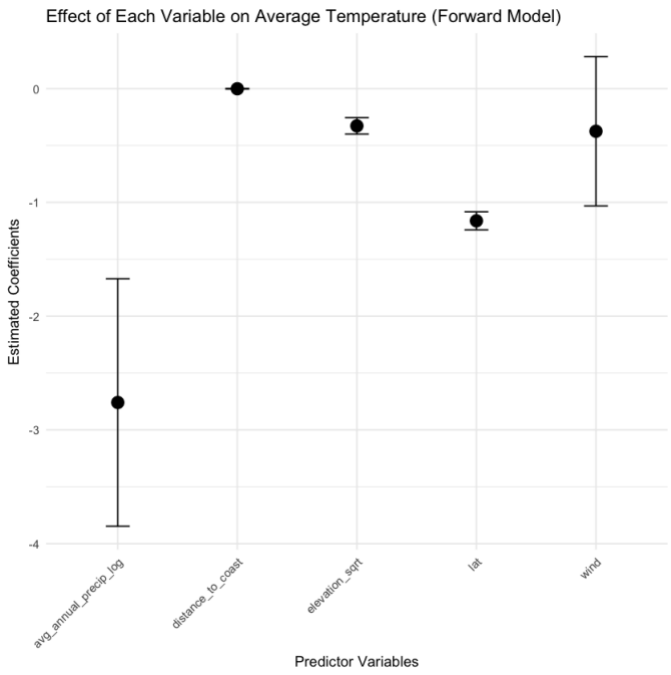

Figure 1 - Effect of Each Variable on Average Temperature

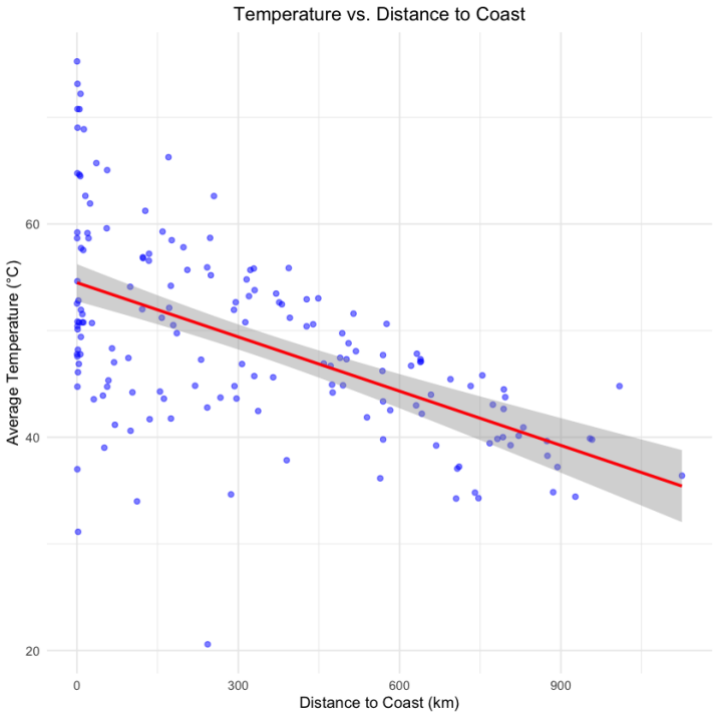

Figure 2 - Temperature vs Distance to Coast Graph