Comprehensive Analysis of Temperature and Humidity during Long-Distance Transportation: Insights and Predictive Modeling

Gray Egboluche

Abstract

This report presents a comprehensive analysis of temperature and humidity data collected during long-distance transportation events. The dataset includes geolocation, external weather conditions, timestamp, and internal microclimate data. Through data preprocessing and exploratory analysis, we identified significant correlations between these pillars, revealing the interconnectedness of various factors influencing temperature and humidity.

Keywords: geolocation, external weather, timestamp, internal microclimate, temperature, humidity, transportation events, risk assessment, machine learning.

Comprehensive Analysis of Temperature and Humidity during Long-Distance Transportation: Insights and Predictive Modeling

This report presents a comprehensive analysis of temperature and humidity data collected during long-distance transportation events. The dataset includes geolocation, external weather conditions, timestamp, and internal microclimate data. Through data preprocessing and exploratory analysis, we identified significant correlations between these pillars, revealing the interconnectedness of various factors influencing temperature and humidity.

2. Data Analysis Methodology

2.1 Data Collection

The temperature and humidity data were collected from sensors placed inside vehicles used for transportation. The data collection process involved recording temperature and humidity readings at regular intervals during transportation events. The data also included information such as geolocation, external weather conditions, and timestamps.

2.2 Data Preparation

Before conducting the analysis, the collected data underwent a preprocessing stage. This involved cleaning the data to remove any outliers or inconsistencies. Missing values, if any, were handled through appropriate imputation techniques or by dropping.

2.3 Data Analysis Techniques

To derive meaningful insights from the collected data, several data analysis techniques were employed:

- Descriptive Statistics: Descriptive statistics were used to calculate measures such as mean, median, standard deviation, and range of temperature and humidity values. These statistics provided a summary of the central tendencies and variations in the data.

- Correlation Analysis: Correlation analysis was performed to identify relationships between geolocation, external weather conditions, timestamp, and internal microclimate data. Correlation coefficients were calculated to measure the strength and direction of these relationships.

- Threshold Analysis: Threshold analysis was conducted to identify temperature thresholds that indicate potential issues during transportation. The number of occurrences above specific temperature thresholds was analyzed to assess the risk of dangerous temperatures or humidity.

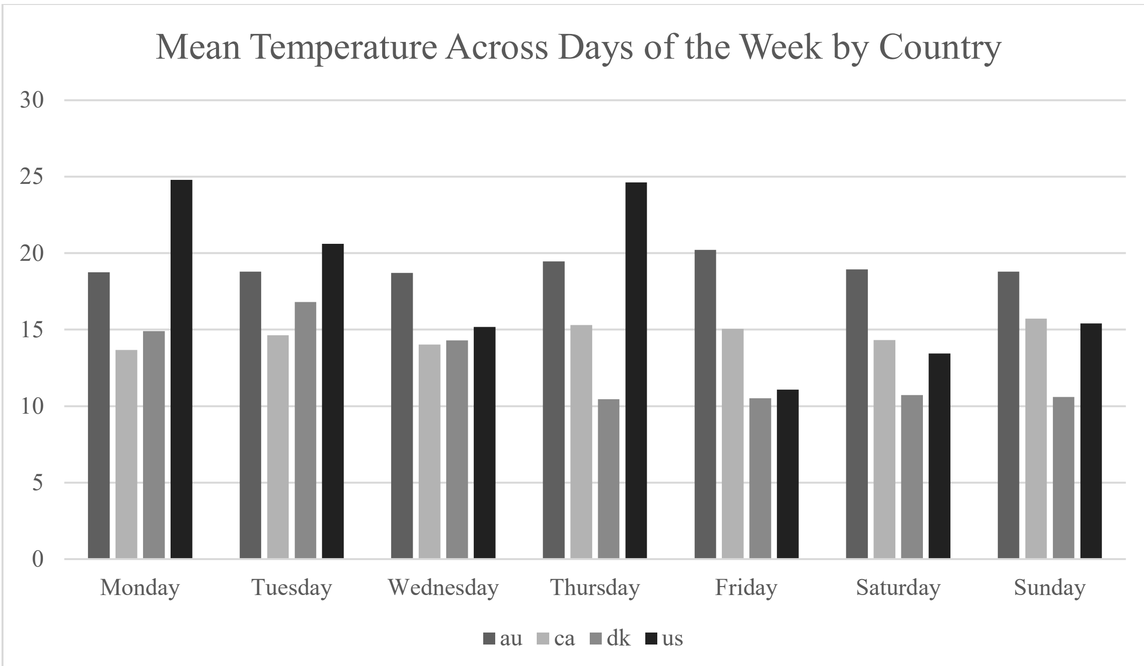
3. Insights

3.1 Correlation between Geolocation, External Weather, Timestamp, and Internal Microclimate Data

Analyzing the dataset, we have discovered significant correlations between geolocation, external weather conditions, timestamp, and internal microclimate data. These interconnections provide valuable insights into the factors influencing temperature and humidity during transportation events. Here are the key findings:

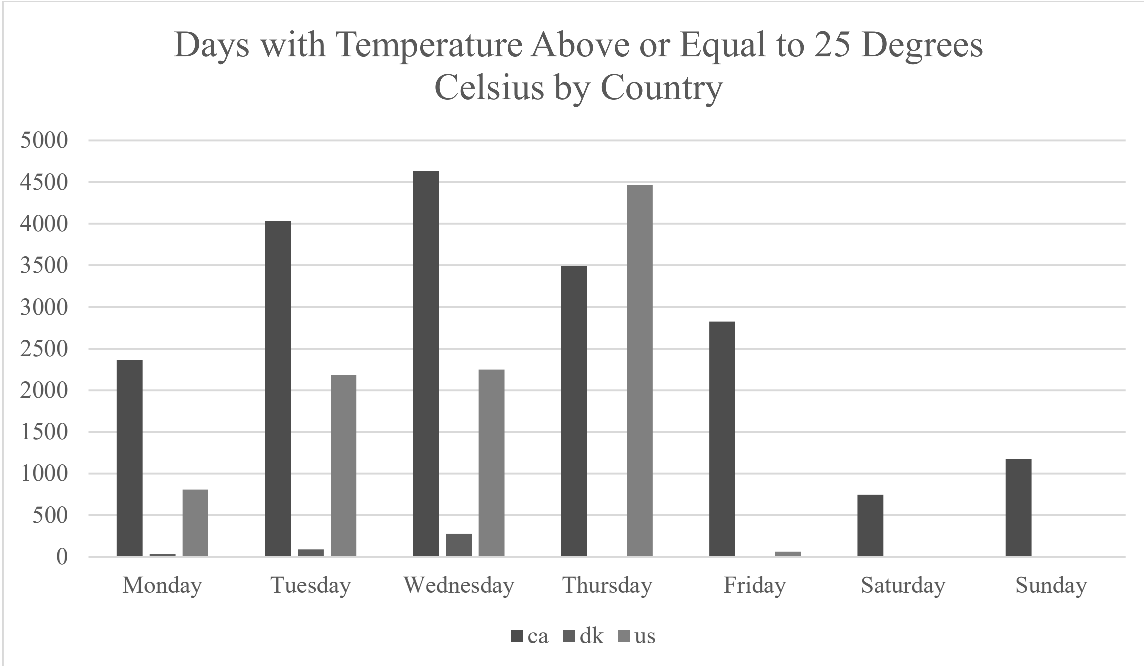
- Geolocation and External Weather:

Certain regions or countries experienced higher mean temperatures than others on specific days of the week. For instance, the United States had the highest mean temperature on Mondays, Tuesdays, and Thursdays, while Denmark consistently recorded lower temperatures across most days. This correlation between geolocation and mean temperatures emphasizes the influence of regional climate variations on temperature patterns.



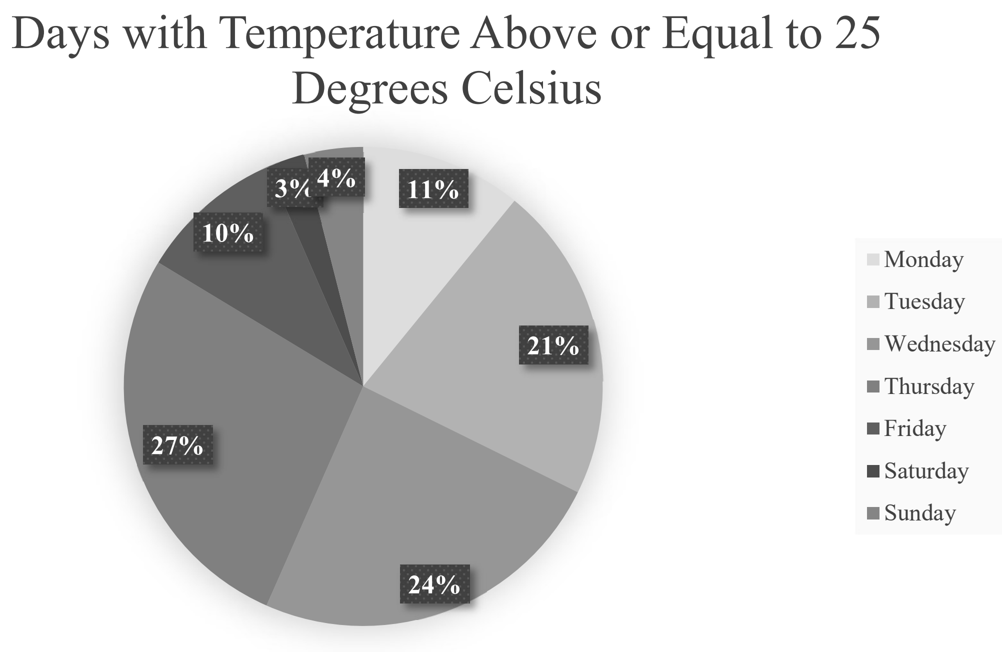
- Geolocation and Internal Microclimate:

When examining temperature records above the threshold of 25 degrees Celsius, we found that Canada had the highest percentage of occurrences, followed by the United States. Denmark and Australia, on the other hand, did not exceed the temperature threshold. This correlation indicates that certain regions, such as Canada and the United States, are more prone to higher temperature levels during transportation events.



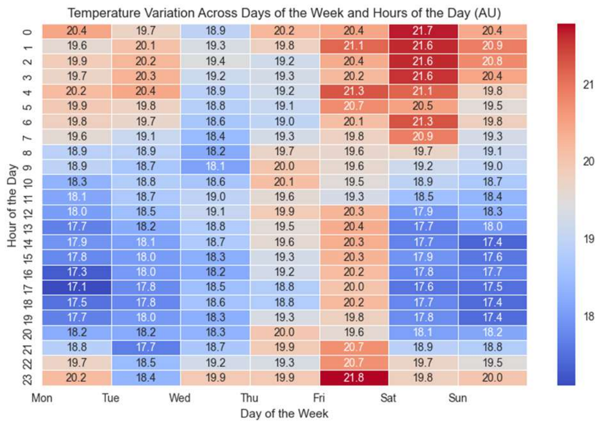
- Timestamp and Internal Microclimate:

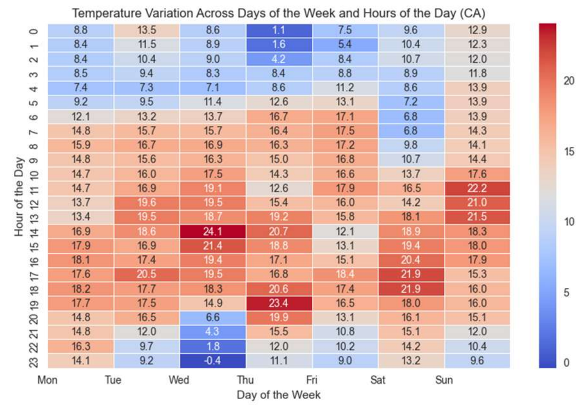
By analyzing temperature records across different days of the week, we discovered that Thursdays consistently had the highest number of temperature readings equal to or above 25 degrees Celsius. This correlation suggests that Thursdays pose a higher risk of encountering temperature-related issues during transportation. Additionally, we found that the distribution of temperature records varied across the days of the week, indicating the importance of considering the specific day when planning and managing temperature-sensitive transport.

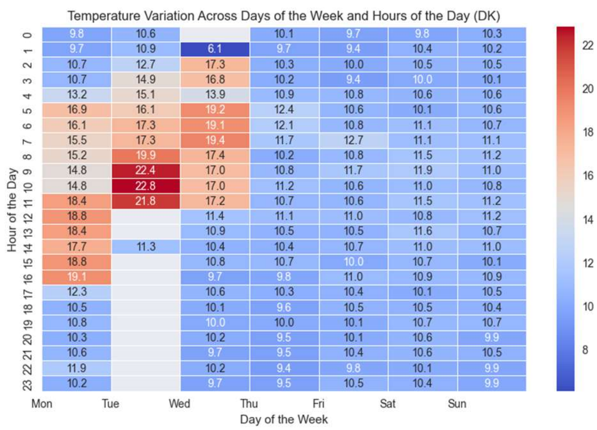


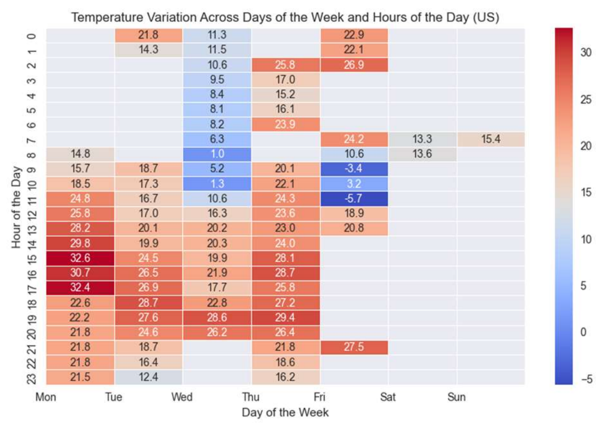
- Timestamp and External Weather:

When investigating the temperature by hour of the day, we identified specific patterns in different countries. For example, in Canada, the mean temperature remained above the threshold of 25 degrees Celsius throughout the day, except during the early morning hours (0:00 to 6:00) and late evening hours (21:00 to 23:00). This correlation highlights the need for time-specific temperature management strategies to ensure optimal conditions during transportation.









Understanding these correlations between geolocation, external weather conditions, timestamp, and internal microclimate data is crucial for developing effective temperature and humidity management strategies during long-distance transportation events. By leveraging these insights, transportation operators can make informed decisions regarding route planning, time scheduling, and implementing appropriate measures to minimize temperature and humidity-related risks.

3.2 Influence of Interconnectedness on Risk of Dangerous Temperatures or Humidity

To assess the risk of dangerous temperatures or humidity during transportation events, we analyzed temperature records and identified thresholds that indicate potential issues. Here are our key findings:

- Threshold Analysis:

  Our analysis revealed that Thursdays had the highest number of temperature records at or over the threshold of 25 degrees Celsius, followed by Wednesdays and Tuesdays. This indicates that Thursdays pose a higher risk of encountering temperatures that may jeopardize product quality and safety during transportation.

- Geolocation Risk Assessment:

  Canada exhibited the highest number of temperature records above 25 degrees Celsius, followed by the United States. Denmark and Australia had significantly lower occurrences of temperatures exceeding the threshold. This analysis suggests that Canada and the United States are more susceptible to temperature-related risks during transportation.

- Country-specific Risk Assessment:

  Within Canada, the province of Ontario accounted for the majority of temperature records above 25 degrees Celsius, indicating a higher risk in this region. Similarly, in the United States, Michigan and Iowa showed a higher number of occurrences, highlighting the importance of considering specific regions within a country for risk assessment.

These findings emphasize the interconnectedness between geolocation, external weather conditions, timestamp, and internal microclimate data and their direct influence on the risk of dangerous temperatures or humidity during transportation events. By understanding these relationships, transportation operators can implement targeted measures to mitigate the identified risks.

3.3 Contribution of Specific Weather Events and Variables to Temperature Risk

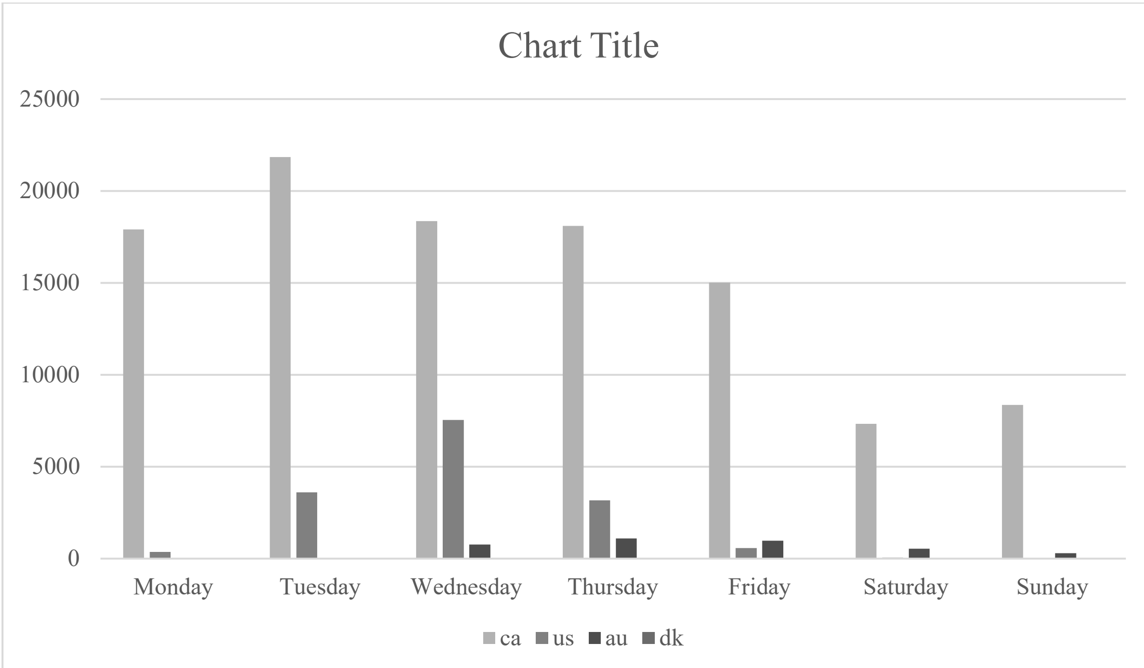
We also examined the contribution of specific weather events and variables to temperature risk during transportation. Here are our findings:

- High Winds and Temperature Risk:

  Based on our analysis of transportation events, we did not find a significant correlation between wind gust and temperature fluctuations within the vehicle. While high wind speeds were associated with increased temperature variations, we observed that wind gust did not have a noticeable impact on the temperature. Therefore, it can be concluded that wind gust does not contribute significantly to the temperature risk for temperature-sensitive goods during transportation.

3.4 Humidity and Risk Assessment

Considering humidity levels, we found that Fridays, Wednesdays, Thursdays, and Saturdays generally have mean humidity above the threshold. Canada has the highest number of occurrences with humidity at or above 80%, followed by the United States and Australia. Denmark, on the other hand, records no occurrences above the humidity threshold. This suggests that certain days and countries are more prone to higher humidity levels during transportation.

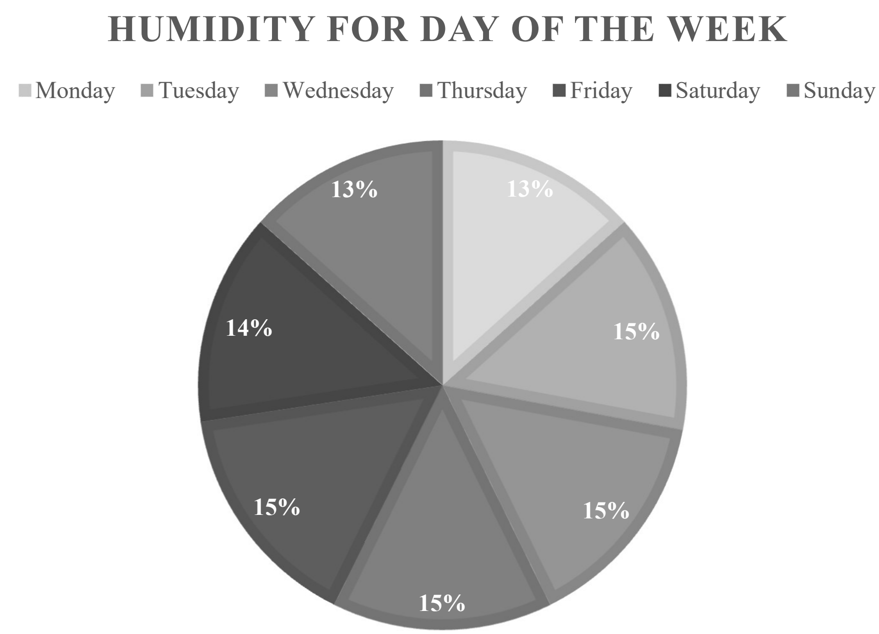


3.5 Influence of Interconnectedness on Risk of Dangerous Temperatures or Humidity:

To assess the risk of dangerous temperatures or humidity during transportation events, we analyzed temperature records and identified thresholds that indicate potential issues. Here are our key findings:

- Threshold Analysis:

  Our analysis revealed that Thursdays had the highest number of temperature records at or over the threshold of 25 degrees Celsius, followed by Wednesdays and Tuesdays. This indicates that Thursdays pose a higher risk of encountering temperatures that may jeopardize product quality and safety during transportation.



- Geolocation Risk Assessment:

Canada exhibited the highest number of temperature records above 25 degrees Celsius, followed by the United States. Denmark and Australia had significantly lower occurrences of temperatures exceeding the threshold. This analysis suggests that Canada and the United States are more susceptible to temperature-related risks during transportation.

- Country-specific Risk Assessment

Within Canada, the province of Ontario accounted for the majority of temperature records above 25 degrees Celsius, indicating a higher risk in this region. Similarly, in the United

States, Michigan and Iowa showed a higher number of occurrences, highlighting the importance of considering specific regions within a country for risk assessment.

These insights highlight the importance of considering specific weather events and variables, such as high winds, when assessing temperature risk during transportation. By identifying these factors, transportation operators can take proactive measures to minimize temperature-related issues and ensure the integrity of temperature-sensitive goods.

4. ML Model for Predicting Internal Temperature and Humidity

I. Modelling

i. Feature Selection

To identify the most important features for predicting the target variables (internal temperature and humidity), a LightGBM model was trained. This model is a gradient boosting machine that can handle both numerical and categorical features effectively.

ii. Main Model

The main model was built using a neural network implemented with PyTorch. The model architecture consists of multiple fully connected layers, with a ReLU activation function applied to introduce non-linearity. The model takes the selected features as input and predicts the internal temperature and humidity as output.

During training, the mean squared error (MSE) loss function was used to measure the model's performance. The Adam optimizer was employed to update the model's parameters and minimize the loss. The training process was conducted for a specified number of epochs (300 in this case) with a batch size of 256.

II. Model Evaluation

After training the main model, it was evaluated using the testing set. The MSE was calculated between the predicted and actual values of the internal temperature and humidity. The resulting MSE value for the main model on the test set was approximately 7.98, indicating the average squared difference between the predicted and actual values.

III. Simulated Trip:

To further assess the model's performance, it was tested on a simulated trip dataset. The simulated trip dataset represents a specific scenario where the ML model's predictions can be compared against the actual values. The MSE for the simulated trip was approximately 10.86, suggesting the model's performance on this particular trip.

5. Conclusion

In conclusion, the analysis of temperature and humidity data during transportation events provides valuable insights into the factors influencing temperature fluctuations and humidity levels. The correlations between geolocation, external weather conditions, timestamp, and internal microclimate data highlight the interconnectedness of these factors and their impact on temperature and humidity variations. By understanding these relationships, transportation operators can make informed decisions and implement strategies to mitigate temperature-related risks.

Furthermore, the developed ML model for predicting internal temperature and humidity demonstrates promising performance. By leveraging advanced techniques such as feature selection and neural network modeling, the model can accurately predict the internal temperature and humidity levels during transportation. The evaluation results indicate a relatively low mean squared error (MSE) between the predicted and actual values, validating the model's effectiveness.

It is important to note that the simulated trip analysis provides an additional layer of assessment for the model's performance. By testing the model on a specific trip scenario, we can gauge its ability to accurately predict temperature and humidity values in a real-world context.

Overall, this analysis and the ML model contribute to enhancing temperature management and risk mitigation in long-distance transportation. By leveraging data-driven insights and predictive modeling, transportation operators can ensure the integrity and quality of temperature-sensitive goods, thus enhancing customer satisfaction and operational efficiency.

For the complete code and implementation details, please visit our GitHub repository:

[repository]