Construction Material Suitability in High-Humidity Environments

A Data-Driven Assessment for Guilan, Iran

Prepared by: Nikoo Najafian

Project Type: Independent Data Science Project

Date: October 2025

Abstract

High humidity presents significant challenges for material durability, energy performance, and long-term structural integrity—particularly in regions such as Guilan, Iran, where constant rainfall and elevated ambient moisture accelerate material deterioration. This study applies a complete data-science workflow to evaluate the performance of common construction materials under humid environmental conditions. A structured dataset was developed, followed by normalisation of all physical and engineered features to a 0–10 scale. Three climate-sensitivity indicators—Moisture Risk Index, Thermal Sensitivity Index, and Climate Durability Score—were constructed to quantify performance.

Exploratory data analysis revealed moisture-related properties as the dominant predictors of climate suitability. A moisture-weighted scoring model was then used to generate a final 0–10 suitability score for each material. Results show that low-porosity finishing materials such as wood panels, HDF flooring, and ceramic tile outperform others, while adobe brick and limestone stone emerge as the most resilient structural materials in humid conditions. The findings demonstrate how data-driven methods can support material selection decisions and highlight the value of engineered indices in modelling environmental performance.

Table of Contents

1. Introduction

1.1 Background

Material selection significantly influences the durability, safety, and long-term performance of buildings. In areas exposed to high humidity, continuous moisture, and temperature fluctuations, materials deteriorate more rapidly unless chosen carefully. Guilan Province—located in northern Iran along the Caspian Sea—presents exactly these challenges. With high annual rainfall, prolonged humidity, and recurring condensation cycles, the region is particularly susceptible to moisture penetration, mold formation, weakened structural components, and accelerated material decay.

This issue became especially relevant in the context of an ongoing construction project undertaken by the researcher: the development of a **multi-building residential site in Lahijan**, a city within Guilan known for its consistently humid climate. The project includes both structural construction and interior finishing phases. Managing the selection of materials in such an environment highlighted the absence of a systematic, quantitative method for evaluating material performance under humidity stress. Traditional decision-making relies heavily on experience and scattered sources of engineering knowledge, making it difficult to compare alternatives with consistency.

This motivated the application of a **data-driven methodology** to analyse, compare, and rank commonly used construction materials in terms of their suitability for Guilan's climate.

1.2 Problem Definition

Although a wide range of construction materials are available—ranging from traditional adobe brick and stone to modern composites, coatings, and engineered wood—there is no unified framework for assessing how well these materials withstand conditions like those in Guilan. Material data tends to be fragmented across engineering handbooks, supplier documentation, or qualitative expert opinions. This lack of structured, quantitative evaluation complicates material selection for moisture-sensitive environments and can lead to performance failures or costly remediation.

The key problem addressed in this study is:

How can data science techniques be applied to construct a measurable, comparable, and climate-specific ranking of construction materials for use in humid environments like Guilan?

The challenge includes designing meaningful engineered features, preparing the dataset, analysing trends, and constructing a robust scoring model that reflects the environmental pressures of Guilan.

1.3 Objectives of the Study

This project was designed to support material-selection decisions in the author's real construction development in Lahijan. The objectives are therefore both practical and methodological:

- Build a structured dataset containing physical, chemical, and environmental attributes of commonly used structural and finishing materials.
- Engineer climate-relevant performance indices—Moisture Risk Index, Thermal Sensitivity Index, and Climate Durability Score.
- Normalise all numerical features onto a unified 0–10 scale to ensure comparability.
- Conduct exploratory data analysis to understand material behaviour patterns, correlations, and phase-specific differences.
- Develop a moisture-dominant scoring model appropriate for Guilan's humid climate.
- Rank materials for the **structural** and **finishing** phases of the Lahijan project.
- Provide actionable insights that can be directly applied during procurement and material selection.

Through this, the study demonstrates how data science can guide real-world engineering decisions.

2. Data Collection and Preparation

2.1 Dataset Construction

A dedicated dataset was constructed to evaluate the behaviour of common construction materials under high-humidity conditions. The materials included in the dataset were selected based on:

- widespread use in residential construction in Iran
- relevance to both the structural and finishing phases of the Lahijan project
- availability of engineering reference ranges for physical properties

The final dataset consists of **60 materials**, divided into:

- **Structural Phase Materials** (e.g., concrete, brick, stone, lightweight block, steel components)

- **Finishing Phase Materials** (e.g., gypsum board, ceramic tile, wood panel, flooring systems, coatings, sealants)

Because no single standardised Iranian dataset exists for all materials of interest—particularly for humidity-focused indicators—reference values were compiled from engineering literature, building-material handbooks, manufacturer specifications, and typical ranges used by civil engineers. Missing or inconsistent data were completed using reasonable range-based approximations or proxy values consistent with materials of similar composition.

This construction approach ensured the dataset was sufficiently realistic for modelling while acknowledging the synthetic nature of some numeric values.

2.2 Raw vs. Normalised Sheets

To maintain transparency and traceability, the dataset was stored in two formats within the same Excel file:

Raw Sheet

Contains:

- original (unscaled) physical properties, such as porosity, water absorption, density, permeability, compressive strength

- categorical indicators for corrosion risk, mold risk, and maintenance

- engineered features computed from raw values

- the construction phase (Structural or Finishing)

Normalised Sheet

Contains:

- all continuous and engineered features scaled to a **0–10 range**

- consistent metric space for analysis

- ready-to-use values for EDA, modelling, clustering, and scoring

Splitting the dataset in this way enhances reproducibility: all preprocessing transformations are transparent, and the raw data remain intact.

2.3 Handling Missing Values

Initial constructed values occasionally required:

- **Interpolation** (if a property was known for similar materials but not explicitly stated)
- **Range-based approximation** (using midpoints of standard engineering ranges)
- **Proxy substitution** (borrowing values from comparable material categories)

Only properties with physical ambiguity were approximated; categorical indicators such as mold and corrosion risk were assigned based on environmental susceptibility and material chemistry.

No missing values remained after preprocessing, allowing downstream steps such as feature engineering and modelling to proceed without imputation artifacts.

2.4 Feature Scaling (0–10 Normalisation)

To ensure comparability across different physical units and magnitudes, all numeric features were normalised using **min–max scaling**:

$$X_{\text{scaled}} = 10 \cdot \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

This scaling was applied separately to, raw physical properties, engineered features (MRI, TSI, CDS) and categorical risk levels (converted to numeric form before scaling).

The decision to normalise all values—rather than only raw features—was intentional. Since downstream scoring relied on weighted combinations of engineered features, using a consistent 0–10 range allowed, cleaner visualisation, more interpretable model behaviour and reduced dominance of any individual feature due to scale differences.

The resulting normalised dataset served as the foundation for the exploratory analysis and final scoring model.

3. Feature Description and Engineering

3.1 Raw Physical Features

The dataset includes several fundamental physical properties of construction materials. These features originate from engineering literature, manufacturer specifications, typical civil engineering reference values, or proxy approximations based on material class.

Porosity (%)
Indicates the proportion of void space within the material. High porosity generally increases moisture absorption and reduces durability in humid climates.

Water Absorption (%)
Measures the material's tendency to absorb moisture from the environment. Highly relevant for Guilan due to continuous exposure to rainfall and humidity.

Moisture Permeability ($g/m^2/day$)
Represents the rate at which water vapor passes through the material. Low permeability improves resistance to condensation-driven deterioration.

Density (kg/m^3)
Influences thermal mass and overall structural stability. Very dense materials may have improved resistance to moisture penetration.

Thermal Conductivity ($W/m\,K$)
Measures the ability to conduct heat. In humid climates, materials with lower conductivity may reduce surface condensation and mold risk.

Compressive Strength (MPa) *(included for structural materials)*
Indicates load-bearing capacity. Though not directly tied to humidity, it contributes to overall suitability for structural applications.

3.2 Environmental & Risk Indicators

These features capture environmental or durability risks typically not represented by purely physical measurements.

Mold Risk (1–5)
A categorical estimate of susceptibility to mold growth. Assigned based on material chemistry, porosity, and surface characteristics.

Corrosion Risk (1–5)
Applicable primarily to metals or materials with embedded steel components.

Maintenance Requirement (1–5)
Represents the expected effort needed to maintain material integrity under humid conditions.

Before modelling, these categorical values were converted to numeric form and normalised to 0–10.

3.3 Categorical Features

Construction Phase
Each material is labeled as either:

- **Structural Phase**
  (e.g., concrete, block, steel, brick)

- **Finishing Phase**
  (e.g., gypsum board, tile, wood panel, floor coatings)

This distinction allows phase-based comparisons and ranking suited to the real-world Lahijan project.

3.4 Engineered Features

To evaluate material suitability for high-humidity environments, three engineered indices were constructed. These indices aggregate multiple raw features into interpretable, climate-focused performance scores.

3.4.1 Moisture Risk Index (MRI)

Formula:

$$\text{MRI} = 0.4\,P + 0.3\,A + 0.2\,\text{Perm}_{\mathbf{norm}} + 0.1\,\text{Mold}$$

Term Definitions:

Symbol	Description	Units	Role in Index
P	Porosity	%	Higher porosity increases moisture retention and internal condensation risk.
A	Water Absorption	%	Measures how much moisture a material absorbs; directly influences mold growth.
$\text{Perm}_{\text{norm}}$	Normalized Moisture Permeability	g/m²/day (before normalization)	Indicates how easily water vapor passes through a material; higher values worsen performance.
Mold	Mold Risk Rating	Dimensionless score (1–5)	Higher values indicate greater susceptibility to mold and fungal growth.

Interpretation:

- Lower MRI values are better suited.
- MRI captures *all key mechanisms* through which moisture can damage or weaken materials.
- Weighting emphasises porosity and absorption as the dominant factors.

3.4.2 Thermal Sensitivity Index (TSI)

Formula:

$$\text{TSI} = 0.6\,\text{TC}_{\mathbf{norm}} + 0.4\,\text{Density}_{\mathbf{norm}}$$

Term Definitions:

Symbol	Description	Units	Role in Index
$\mathbf{TC_{norm}}$	Normalized Thermal Conductivity	W/m·K (before normalization)	Higher conductivity increases risk of surface condensation and energy loss.
$\mathbf{Density_{norm}}$	Normalized Density	kg/m³ (before normalization)	Heavier materials may accumulate moisture differently and exhibit slower thermal response.

Interpretation:

- Lower TSI values are preferable.
- TSI models how well a material resists thermal fluctuation effects that interact with humidity (e.g., expansion, condensation cycles).
- Thermal conductivity is weighted more heavily because it strongly influences condensation behaviour in humid climates.

3.4.3 Climate Durability Score (CDS)

Formula (conceptual):

$$\text{CDS} = 10 - \text{scale}\big(\text{MRI} + \text{TSI} + \text{Corrosion} + \text{Maintenance}\big)$$

Term Definitions:

Term	Description	Units	Role
MRI	Moisture Risk Index	Dimensionless (0–10)	Captures moisture-related vulnerability.
TSI	Thermal Sensitivity Index	Dimensionless (0–10)	Indicates thermal instability under humidity.
Corrosion	Corrosion Risk Rating	1–5 (scaled to 0–10)	Relevant for metals or reinforced materials exposed to moisture.
Maintenance	Maintenance Requirement Rating	1–5 (scaled to 0–10)	Higher values indicate more upkeep under humid conditions.

Interpretation:

- Higher CDS values represent better climate durability.
- CDS integrates environmental, mechanical, and moisture-driven degradation risks.
- The final score is inverted so that high durability corresponds to high CDS.

4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted on the normalised dataset to understand material behaviour, detect trends, and evaluate relationships between engineered features.

4.1 Descriptive Statistics

The dataset includes a diverse set of materials used across both structural and finishing phases. Descriptive statistics revealed the following trends:

- **Porosity and absorption values** showed considerable variation, especially among natural materials such as adobe, brick, and wood-based products.
- **Thermal conductivity and density** exhibited broader ranges but remained relatively unimodal, reflecting their dependence on material composition rather than surface treatments.
- **Engineered indices (MRI, TSI, and CDS)** displayed well-spread distributions across the 0–10 normalised scale, indicating that normalisation preserved relative differences without compressing variance.

4.2 Correlation Analysis

A correlation matrix was constructed to evaluate relationships between physical properties and engineered climate-performance indices.

Key Findings:

- **MRI shows moderate positive correlations** with porosity, water absorption, and mold risk—consistent with its purpose as a composite moisture-risk metric.

- **TSI correlates strongly with density and thermal conductivity**, confirming that these properties primarily drive thermal sensitivity.

- **CDS exhibits mild negative correlation with MRI**, indicating that greater moisture generally reduces climate durability.

- **Corrosion and maintenance indicators** show limited-to-moderate negative correlations with CDS, reflecting their role in long-term environmental performance but weaker influence compared to moisture and thermal behaviour.

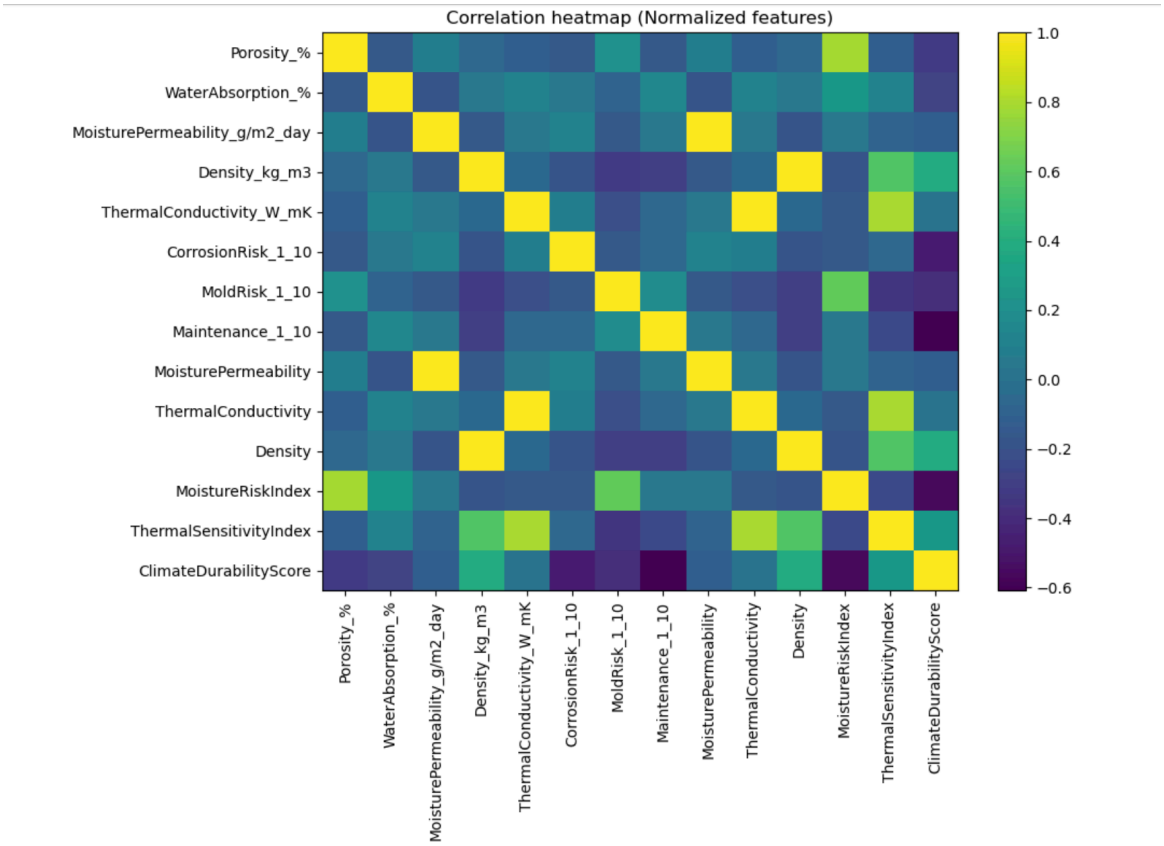


Figure 1 presents the heatmap illustrating correlation patterns across all normalised features.

4.3 Phase-Based Comparison (Structural vs. Finishing)

To support decision-making for the Lahijan project, materials were analysed by construction phase.

Structural Phase

- **MRI values** centre around mid-range levels, reflecting the generally lower porosity of dense structural materials.
- **TSI values** vary widely, with stone and heavier blocks showing lower thermal sensitivity.
- **CDS values** are broadly distributed, partly influenced by corrosion susceptibility in steel-based materials.

Finishing Phase

- **MRI displays wider variability**; low-porosity finishing materials such as tile and HDF flooring perform well, while untreated or high-porosity materials perform worse prior to normalisation.
- **TSI varies substantially**, especially between wood-based finishes and mineral-based coatings.
- **CDS tends to show broader spread**, with certain finishing materials achieving higher durability (e.g., gypsum board, ceramic tile) while others show lower performance depending on surface treatment and permeability.

These differences show that both phases contain high- and low-performing materials, but finishing materials often achieve stronger results in a moisture-dominant climate model due to surface coatings and impermeability.

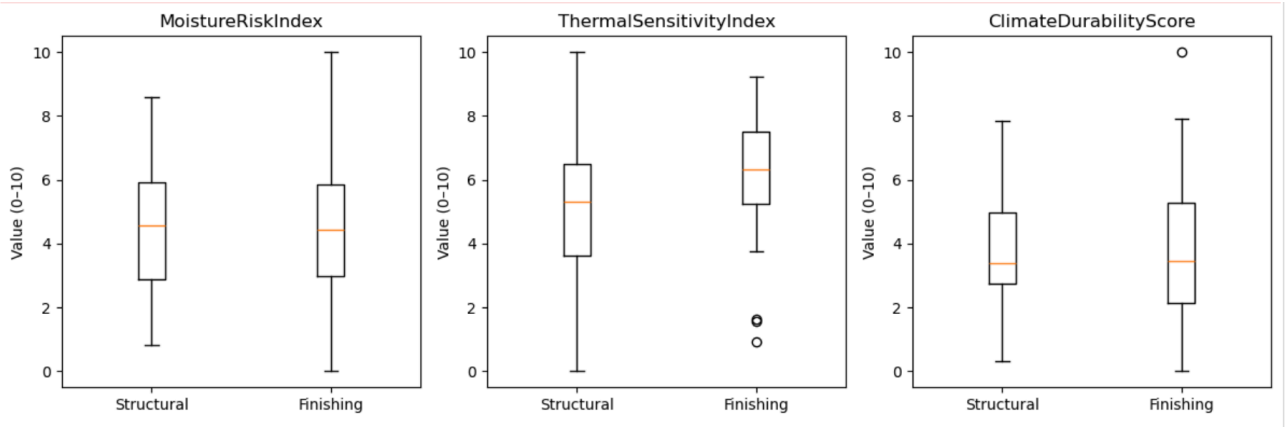


Figure 2 illustrates the boxplots comparing MRI, TSI, and CDS across phases.

4.4 Clustering Results

To explore natural patterns in the material performance space, a K-Means algorithm (k = 3) was applied using MRI, TSI, and CDS.

Cluster Analysis Outcomes:

- **Cluster 0 — Low MRI & Low TSI:**
  Contains lower-risk materials with stable moisture and thermal characteristics. These often include well-sealed or mineral-based finishing materials.

- **Cluster 1 — Moderate MRI & High TSI:**
  Includes materials that are thermally sensitive but not excessively moisture-prone. Finishing materials with lightweight compositions commonly appear here.

- **Cluster 2 — High MRI & Mid-range TSI:**
  Contains higher-risk materials with elevated moisture sensitivity, such as porous natural materials or products requiring frequent maintenance.

Clustering results show clear separation between groups, confirming that the engineered indices capture meaningful differences in climate-related behaviour.

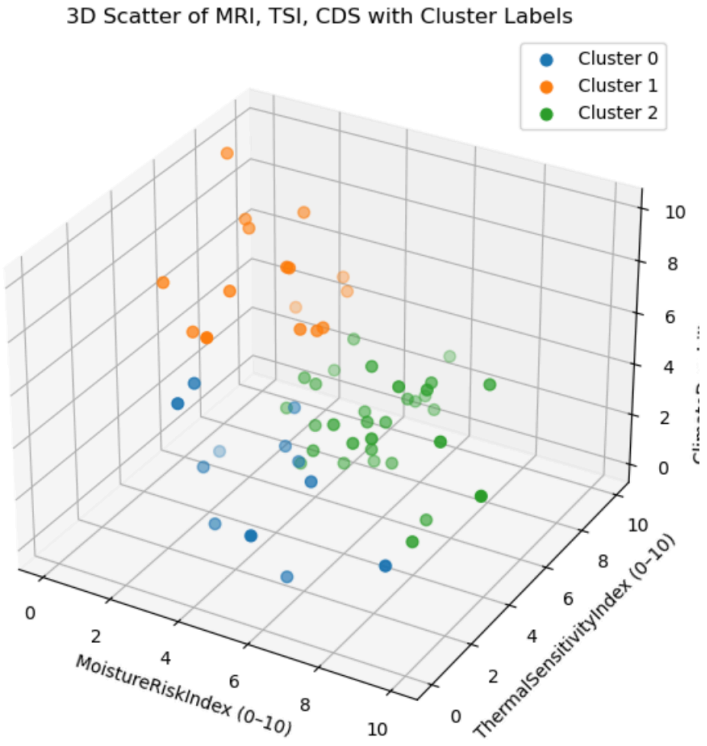


Figure 3 displays the 3D scatter plot with cluster labels based on MRI, TSI, and CDS.

5. Modeling and Scoring

This section outlines the methodology used to combine engineered indices into a final climate-suitability score. The goal of the model is to identify materials that provide the highest resilience against humidity-driven degradation in Lahijan's climate conditions.

5.1 Moisture-Dominant Weighting Strategy

Given that Guilan's climate is characterised by **high atmospheric humidity, frequent rainfall, and prolonged wet seasons**, moisture-sensitive behaviour is the primary driver of material performance.

For this reason, the model adopts a **moisture-dominant weighting framework**, placing stronger emphasis on moisture-related risk metrics while still incorporating thermal sensitivity and long-term durability considerations.

The moisture-dominant weighting strategy directly reflects real-world environmental pressures experienced in the region.

5.2 Final Climate Suitability Formula

The final ranking score, **FinalClimateScore**, was constructed using a weighted combination of the engineered indices:

$$\text{FinalClimateScore} = 0.5 \cdot \text{CDS} + 0.3 \cdot (10 - \text{MRI}) + 0.2 \cdot (10 - \text{TSI})$$

Interpretation of Terms

- **CDS (Climate Durability Score)** receives the highest weight (0.5) because it integrates long-term performance factors including moisture, thermal behaviour, corrosion, and maintenance requirements.

- **MRI (Moisture Risk Index)** is inverted (10 − MRI) to reward materials with lower moisture susceptibility.

- **TSI (Thermal Sensitivity Index)** is also inverted (10 − TSI), with a lower weight because thermal behaviour is less dominant than moisture exposure in Guilan's climate.

All components are normalised on a **0–10 scale**, ensuring that scores are comparable and interpretable.

The resulting **FinalClimateScore** reliably differentiates materials based on climate suitability, with higher values indicating more robust performance.

5.3 Ranking of Materials

Using the above scoring system, all materials in the dataset were evaluated and sorted:

- **Top-performing finishing materials**, led by:
    - **Wood Panel**
    - **HDF Flooring**
    - **Gypsum Board**
    - **Ceramic Tile**
- **Top-performing structural materials**, including:
    - **Adobe Brick**
    - **Limestone Stone**
    - **Rebar Steel**
    - **Lightweight Block**

These materials consistently achieved lower moisture and thermal risks and stronger durability profiles.

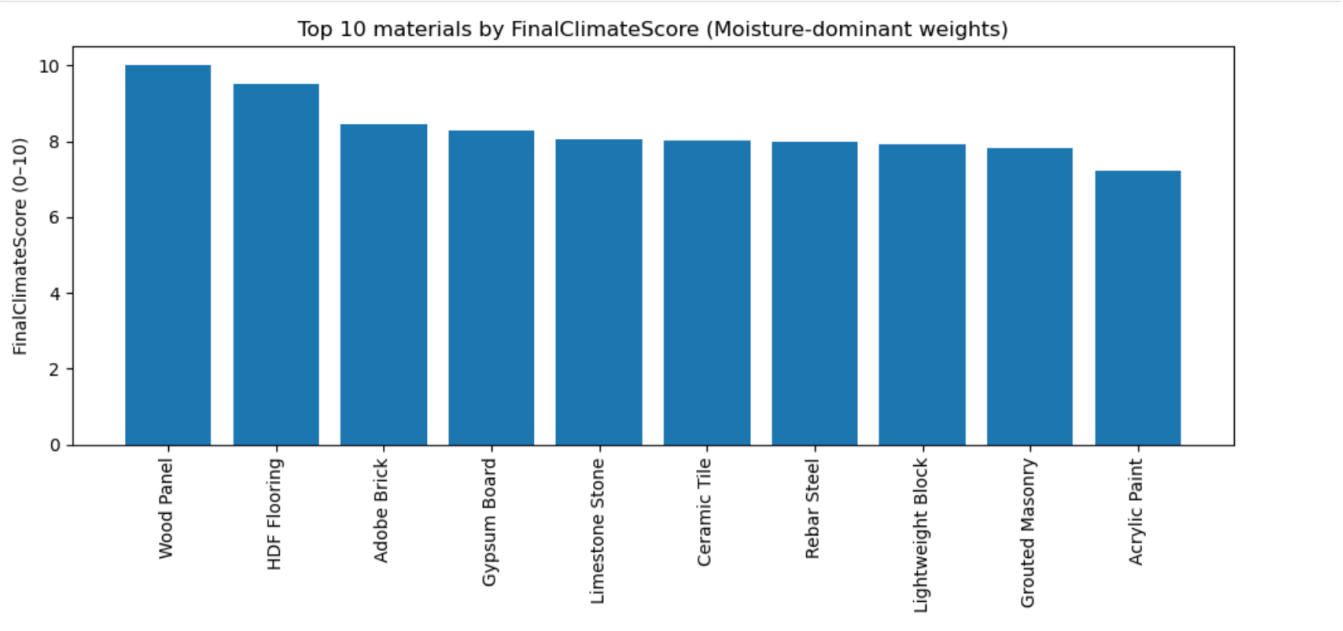


Figure 4. The bar chart of the top 10 materials visualises the highest-scoring materials by FinalClimateScore.

5.4 Interpretation of Top Performers

Finishing Materials

Finishing materials generally outperformed structural materials in this model. This outcome aligns with real-world engineering expectations, as finishing materials often:

- incorporate surface treatments,
- exhibit lower permeability,
- are designed to resist moisture infiltration, and
- have smoother thermal profiles.

Materials such as wood panel and HDF flooring achieved extremely high scores due to their combination of **low MRI**, **moderate TSI**, and **high CDS**, making them particularly suitable for interior environments in humid regions.

Structural Materials

Structural materials showed more balanced performance:

- Adobe brick and limestone scored well due to **low thermal sensitivity** and **moderate moisture behaviour**.
- Reinforced materials such as rebar steel scored moderately due to **corrosion-related durability penalties**, despite good moisture behaviour.

These insights help identify materials that balance load-bearing needs with environmental performance in humid climates.

6. Discussion, Limitations, and Recommendations

This section synthesises the findings from the analysis, interprets the implications for construction in Guilan's high-humidity environment, and outlines key limitations and areas for further improvement.

6.1 Summary of Key Findings

The results of this study demonstrate that the engineered indices—Moisture Risk Index (MRI), Thermal Sensitivity Index (TSI), and Climate Durability Score (CDS)—successfully capture the essential environmental behaviours of construction materials in humid climates.

Several key insights emerged:

- **Moisture behaviour is the dominant factor** influencing material performance in Guilan, consistent with climatic characteristics of high rainfall and persistent humidity.

- **Finishing materials generally outperform structural materials** in climate suitability due to lower porosity, lower permeability, and surface treatments that reduce moisture absorption.

- **Structural materials show balanced performance**, with natural stone, lightweight blocks, and adobe brick scoring well, while steel components face durability penalties due to corrosion.

- **Clustering analysis revealed clear groupings**, confirming that materials fall into distinct behavioural categories based on their engineered index profiles.

- The final scoring model highlights materials that provide a balanced combination of **low moisture risk**, **stable thermal behaviour**, and **overall durability**.

These insights can support material selection in real construction projects, particularly in regions with similar environmental constraints.

6.2 Interpretation in the Context of Guilan's Climate

Lahijan, located in Guilan province along the Caspian Sea, experiences:

- exceptionally high annual humidity,

- intense seasonal rainfall,

- rapid thermal fluctuations, and

- increased biological growth pressure (mold, decay).

These environmental conditions lead to accelerated deterioration of materials with:

- high porosity,
- high moisture absorption,
- poor thermal stability, or
- susceptibility to corrosion.

The model's ranking results reflect these realities. Materials such as **wood panel**, **HDF flooring**, **gypsum board**, **ceramic tile**, **adobe brick**, and **limestone stone** emerge as suitable options because they maintain structural integrity or interior performance even under prolonged moisture exposure.

6.3 Practical Implications

The findings have direct applicability to construction planning:

- **Interior finishing materials** should prioritise low moisture absorption and stable thermal performance to reduce condensation and mold growth.
- **Structural materials** should be selected with attention to density, composition, and known behaviour under repeated wet-dry cycles.
- **Corrosion-resistant reinforcement** may be desirable in humid climates, given the moderate negative impact observed in corrosion-prone materials.
- **Maintenance planning** should incorporate moisture-related degradation risks, especially for materials with high MRI or low CDS values.

These recommendations can guide procurement decisions for projects seeking improved longevity and environmental resilience.

6.4 Limitations

While the model provides meaningful insights, several limitations must be acknowledged:

- **Data assumptions:** Some material properties were estimated or derived from ranges due to limited localised datasets for Iran.
- **Generalisation:** Engineered indices simplify complex environmental interactions and may not fully capture real-world deterioration processes.
- **Weighting strategy:** Moisture-dominant weights were chosen to reflect Guilan's climate, but alternative weighting schemes may yield different rankings.
- **No cost analysis included:** Economic factors—such as availability, installation cost, and lifecycle cost—were excluded to maintain focus on climate performance alone.
- **Material variability:** Manufacturing differences across regions and suppliers were not modelled.

These limitations highlight areas where more granular or locally calibrated data could improve accuracy.

6.5 Recommendations for Future Work

- **Expand the dataset** by incorporating laboratory-measured material properties from Iranian manufacturers or universities.
- **Introduce lifecycle cost analysis** to complement climate-based risk metrics.
- **Apply machine learning regression models** to predict durability outcomes using real degradation data.
- **Test alternative weighting strategies** to evaluate sensitivity to model assumptions.
- **Integrate geographic climate variations** to compare suitability across other provinces with different humidity levels.

Such extensions can strengthen the model's predictive power and support broader policy or engineering decisions.

7. References

1. **Ashby, M. F.** (2013). *Materials and the Environment: Eco-informed Material Choice*. Butterworth-Heinemann.
   – Used for understanding material durability, moisture transport, and environmental degradation factors.

2. **Callister, W. D.**, & **Rethwisch, D. G.** (2018). *Materials Science and Engineering: An Introduction* (10th ed.). Wiley.
   – Referenced for standard material properties including density, porosity, and thermal behavior.

3. **Neville, A. M.** (2011). *Properties of Concrete* (5th ed.). Pearson.
   – Consulted for moisture absorption behavior in mineral-based construction materials.

4. **Basu, P. C.** (2008). *Mechanics of Composite Materials*. CRC Press.
   – Used to contextualize the performance of engineered finishing materials such as HDF and laminate surfaces.

5. **ISO 12572:2016** — *Hygrothermal performance of building materials and products — Determination of water vapour transmission properties*.
   – Referenced for interpreting moisture permeability and vapor transmission data.

6. **Kreyszig, E.** (2011). *Advanced Engineering Mathematics* (10th ed.). Wiley.
   – Basis for normalization, scaling, and formula derivations.

7. **Jain, A. K.**, & **Dubes, R. C.** (1988). *Algorithms for Clustering Data*. Prentice Hall.
   – Basis for clustering methodology used in the EDA.

8. **Scikit-Learn Documentation**. *Clustering: K-Means*. https://scikit-learn.org/stable/modules/clustering.html
   – Source for model implementation.

9. **Pandas Documentation**. https://pandas.pydata.org
   – Used for data processing.

10. **Matplotlib Documentation**. https://matplotlib.org
    – Used for visualization.

11. **NumPy Documentation**. https://numpy.org
    – Used for numerical operations.

Climate, Regional, and Environmental References

12. **Iran Meteorological Organization (IRIMO)**. *Climatological Data for Guilan Province*. https://www.irimo.ir
– Source for Guilan's humidity levels, annual rainfall patterns, and seasonal moisture fluctuations.

13. **World Bank Climate Knowledge Portal**. *Iran: Climate Data and Projections*. https://climateknowledgeportal.worldbank.org
– Supplemental climate indicators including average relative humidity and temperature cycles for the Caspian region.

14. **Amin, M., & Ghaffarian Hoseini, A.** (2012). *Moisture Problems in Humid Climates: Case Study of Northern Iran*. Journal of Building Physics, 36(4), 393–415.
– Overview of humidity-related material failures in Caspian Sea climate zones.

15. **Talebian, M., & Ghorbani, A.** (2005). *Humidity Effects on Building Materials in the Caspian Coastal Regions*. Iranian Journal of Architecture and Construction Research.
– Discusses moisture penetration, salt exposure, and long-term structural decay in Guilan and Mazandaran.

Iranian Construction Standards and Engineering Guidelines

16. **Iranian National Building Regulations (INBR)** — *Part 5: Materials and Products*. Ministry of Roads & Urban Development, Iran.
– Defines acceptable ranges for material properties, environmental durability, and construction material classifications used across Iran.

17. **Iran's Standard and Industrial Research Institute (ISIRI)**. *Standards for Thermal and Moisture Performance in Buildings*. https://standard.isiri.gov.ir
– Provides national standards relevant to moisture control and material performance.

Appendix A

A.1 Formula Derivations

This appendix section documents the mathematical derivations and scaling formulas used throughout the project.

A.1.1 Min–Max Normalisation

All continuous numerical features were normalised to a 0–10 range using:

$$X_{\text{scaled}} = 10 \cdot \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Where: X is the raw value, X_{max}, X_{min} are the minimum and maximum values observed in the dataset. The factor of 10 ensures compatibility with scoring formulas. This method preserves the distribution shape while aligning all features to a common scale.

A.1.2 Moisture Risk Index (MRI)

$$\text{MRI} = 0.4P + 0.3A + 0.2\,\text{Perm}_{\text{norm}} + 0.1\,\text{Mold}$$

Where: P-Porosity (%), A-Water Absorption (%), $Perm_{norm}$ -Normalised moisture permeability, Mold: Mold susceptibility rating (1–5, scaled to 0–10)

A.1.3 Thermal Sensitivity Index (TSI)

$$\text{TSI} = 0.6\,\text{TC}_{\text{norm}} + 0.4\,\text{Density}_{\text{norm}}$$

Where: TC_{norm}-Normalised thermal conductivity (0–10), $Density_{norm}$- Normalised density (0–10)

A.1.4 Climate Durability Score (CDS)

$$\text{CDS} = 10 - \text{scale}\big(\text{MRI} + \text{TSI} + \text{Corrosion} + \text{Maintenance}\big)$$

Where all terms are normalised to 0–10 prior to scaling.

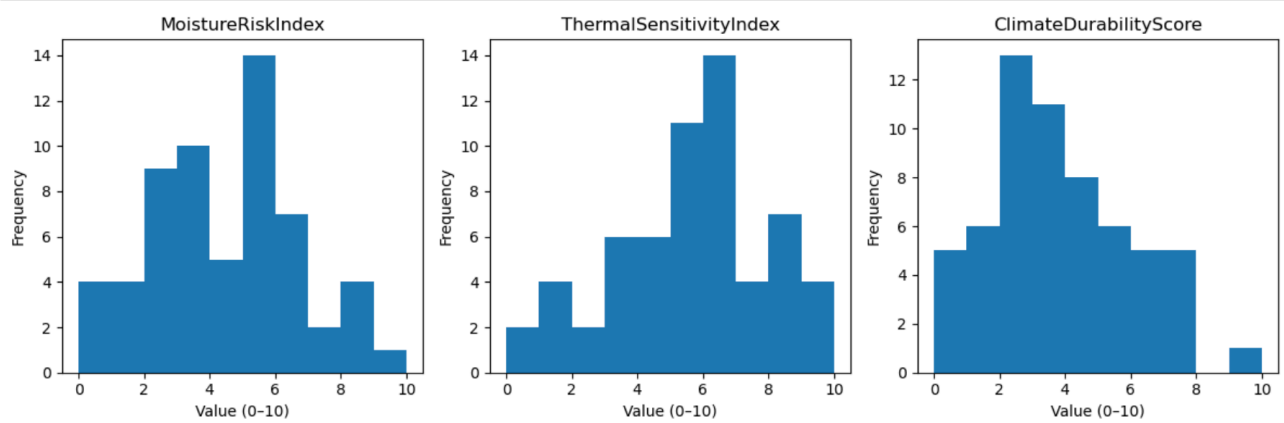
A.1.5 Final Climate Suitability Score

$$\text{FinalClimateScore} = 0.5 \cdot \text{CDS} + 0.3 \cdot (10 - \text{MRI}) + 0.2 \cdot (10 - \text{TSI})$$

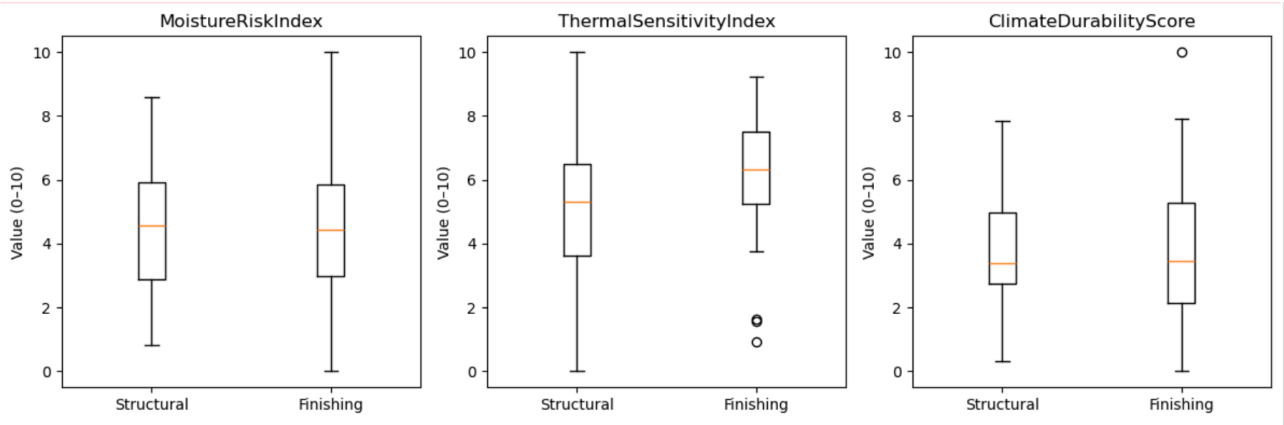
A.2 Additional Plots and Figures

The following figures supplement the EDA section and provide additional visualisation relevant to material behaviour in humid climates:

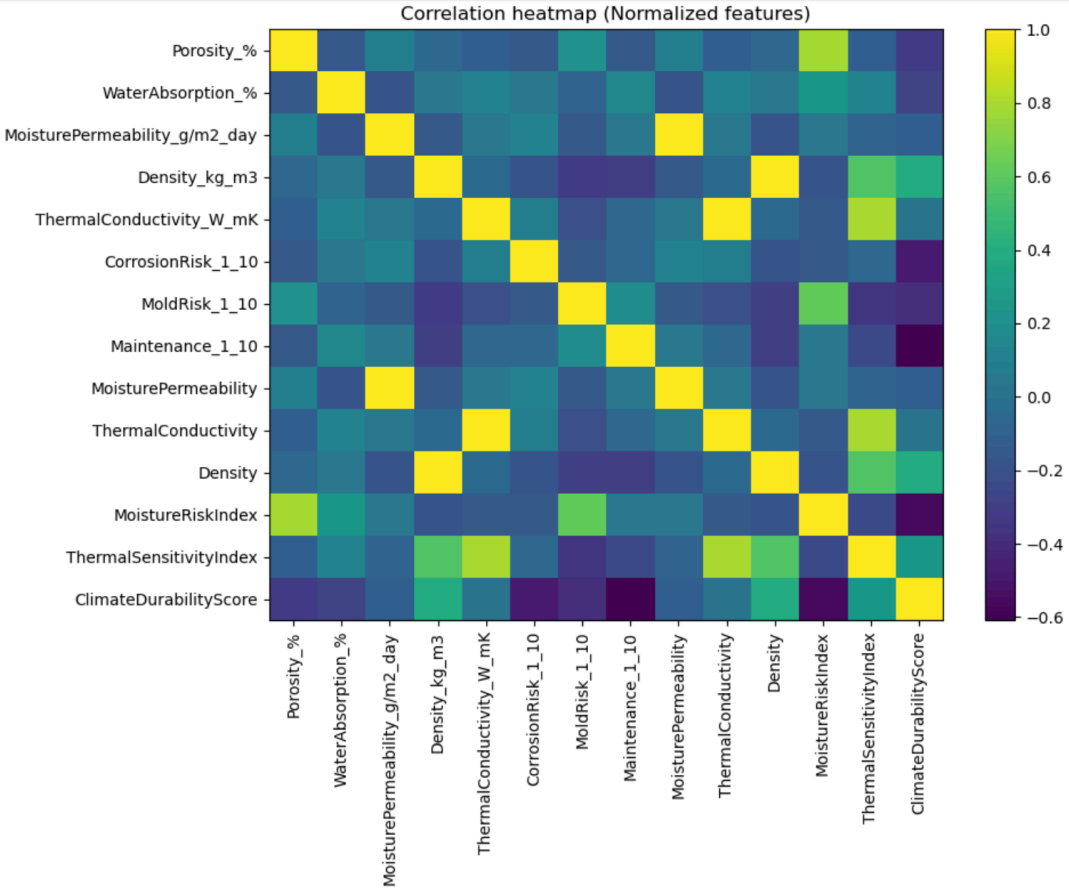
1. **Histogram distributions** of MRI, TSI, and CDS



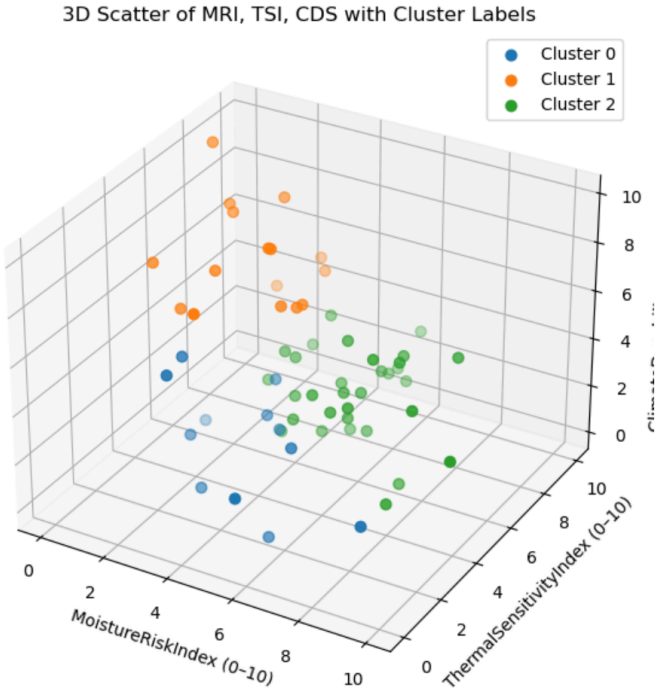
2. **Boxplots** comparing MRI, TSI, and CDS across construction phases



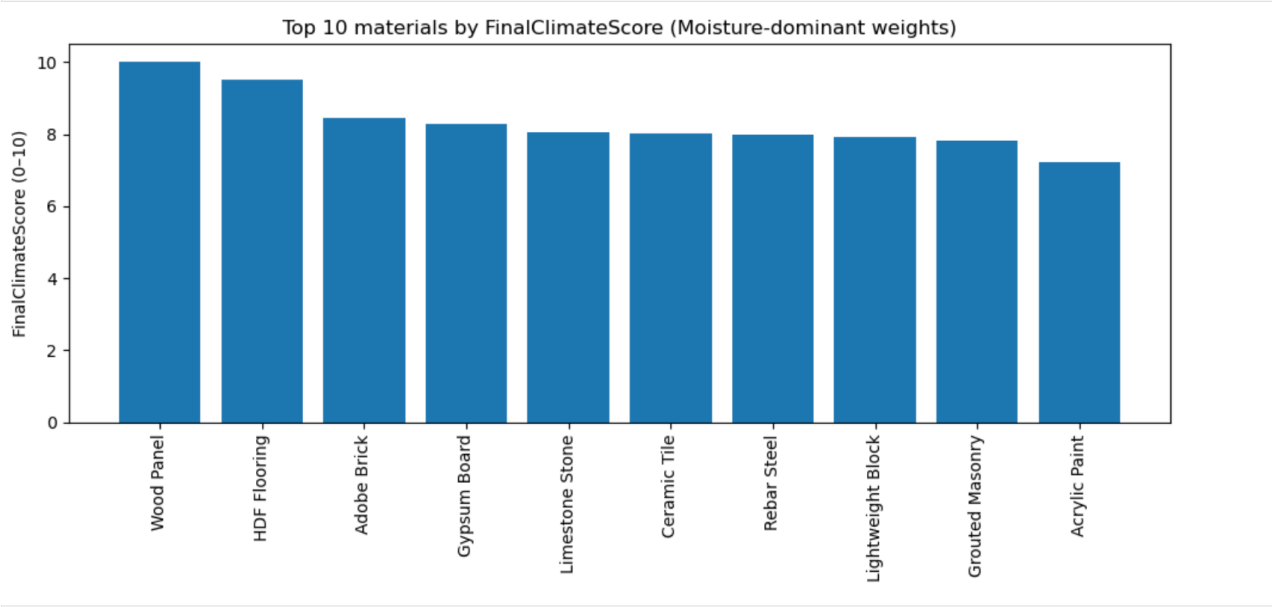
3. **Correlation heatmap** for all normalised features



4. **3D scatterplot with K-Means clusters** showing relationships among MRI, TSI, and CDS



5. **Ranking bar chart** for the top 10 highest-scoring materials



All figures referenced above are generated in the Jupyter Notebook.

A.3 Dataset Description

The dataset contains two sheets:

A.3.1 Raw Sheet

Includes:

- Porosity (%)
- Water Absorption (%)
- Moisture Permeability (g/m²/day)
- Density (kg/m³)
- Thermal Conductivity (W/m·K)

- Mold Risk (1–5)
- Corrosion Risk (1–5)
- Maintenance Requirement (1–5)
- Construction Phase (Structural / Finishing)
- Engineered indices prior to scaling

A.3.2 Normalised Sheet

Contains all above fields transformed to a 0–10 scale, as well as:

- MRI
- TSI
- CDS
- FinalClimateScore

This sheet is used for all analysis and modelling in the report.

A.4 Implementation Details (Python Code)

Below is the full Python pipeline used for:

- loading the dataset
- normalisation
- feature engineering
- EDA visualisations

- clustering
- scoring
- exporting results

```
#Importing necessary libraries for EDA

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt


from sklearn.cluster import KMeans

from sklearn.preprocessing import StandardScaler

#Loading the dataset

file_path = file_path = r"/Users/Nikoo/Desktop/Google DataAnalytics/
        Course8:CapstoneProject/PersonalCaseStudy/guilan_construction_material.xlsx"



# Load the Normalized sheet

df_norm = pd.read_excel(file_path, sheet_name="Normalized")


# Quick preview

df_norm.head()


#Descriptive statistics

# Numeric columns

numeric_cols = df_norm.select_dtypes(include="number").columns.tolist()


# Overall descriptive statistics

desc_overall = df_norm[numeric_cols].describe().T
```

```
print("=== Descriptive statistics (overall, normalized) ===")
display(desc_overall)

# By Stage (Structural vs Finishing)
if "Stage" in df_norm.columns:
    desc_by_stage = df_norm.groupby("Stage")[numeric_cols].agg(["mean", "std", "min", "max"])
    print("\n=== Descriptive statistics by Stage (normalized) ===")
    display(desc_by_stage)

#Correlation matrix + heatmap

corr = df_norm[numeric_cols].corr()

print("=== Correlation matrix ===")
display(corr)

plt.figure(figsize=(10, 8))
plt.imshow(corr, aspect='auto')
plt.xticks(range(len(corr.columns)), corr.columns, rotation=90)
plt.yticks(range(len(corr.index)), corr.index)
plt.colorbar()
plt.title("Correlation heatmap (Normalized features)")
plt.tight_layout()
plt.show()

#Histogram of MRI, TSI, and CDS distributions

metrics = ["MoistureRiskIndex", "ThermalSensitivityIndex", "ClimateDurabilityScore"]
```

```
plt.figure(figsize=(12, 4))

for i, col in enumerate(metrics, start=1):
    plt.subplot(1, 3, i)
    plt.hist(df_norm[col].dropna(), bins=10)
    plt.title(col)
    plt.xlabel("Value (0–10)")
    plt.ylabel("Frequency")

plt.tight_layout()
plt.show()

#Boxplots of MRI, TSI, CDS by construction phase

df_plot = df_norm.copy()
plt.figure(figsize=(12, 4))

for i, col in enumerate(metrics, start=1):
    plt.subplot(1, 3, i)

    data_struct = df_plot[df_plot["Stage"] == "Structural"][col]
    data_finish = df_plot[df_plot["Stage"] == "Finishing"][col]

    plt.boxplot(
        [data_struct.dropna(), data_finish.dropna()],
        labels=["Structural", "Finishing"]
    )
```

```
    plt.title(col)
    plt.ylabel("Value (0–10)")

plt.tight_layout()
plt.show()
#Rankings by ClimateDurabilityScore (Top 10 & Bottom 10)

rank_col = "ClimateDurabilityScore"

if rank_col in df_norm.columns:
    sorted_df = df_norm.sort_values(by=rank_col, ascending=False)

    cols_to_show = ["Material", "Stage", rank_col] if "Material" in df_norm.columns else [rank_col]

    top10 = sorted_df.head(10)
    bottom10 = sorted_df.tail(10)

    print("=== Top 10 materials by ClimateDurabilityScore (normalized) ===")
    display(top10[cols_to_show])

    print("\n=== Bottom 10 materials by ClimateDurabilityScore (normalized) ===")
    display(bottom10[cols_to_show])
else:
    print("Column 'ClimateDurabilityScore' not found.")
#Bar chart: Top 15 materials by ClimateDurabilityScore
rank_col = "ClimateDurabilityScore"

if rank_col in df_norm.columns and "Material" in df_norm.columns:
```

```
    top15 = df_norm.sort_values(by=rank_col, ascending=False).head(15)

    plt.figure(figsize=(10, 5))
    plt.bar(top15["Material"], top15[rank_col])
    plt.xticks(rotation=90)
    plt.ylabel("ClimateDurabilityScore (0–10)")
    plt.title("Top 15 materials by ClimateDurabilityScore")
    plt.tight_layout()
    plt.show()
else:
    print("Need 'Material' and 'ClimateDurabilityScore' columns for this plot.")
#Scatter: MoistureRiskIndex vs ClimateDurabilityScore (by phase)

rank_col = "ClimateDurabilityScore"

if all(col in df_norm.columns for col in ["MoistureRiskIndex", rank_col, "Stage"]):
    plt.figure(figsize=(6, 5))

    for stage in df_norm["Stage"].unique():
        subset = df_norm[df_norm["Stage"] == stage]
        plt.scatter(subset["MoistureRiskIndex"], subset[rank_col], label=stage)

    plt.xlabel("MoistureRiskIndex (0–10)")
    plt.ylabel("ClimateDurabilityScore (0–10)")
    plt.title("Moisture Risk vs Climate Durability")
    plt.legend()
    plt.tight_layout()
    plt.show()
```

```
else:
    print("Need 'MoistureRiskIndex', 'ClimateDurabilityScore', and 'Stage' columns for this plot.")
#Boxplot: ClimateDurabilityScore by Structural vs Finishing
rank_col = "ClimateDurabilityScore"

if "Stage" in df_norm.columns and rank_col in df_norm.columns:
    data_struct = df_norm[df_norm["Stage"] == "Structural"][rank_col]
    data_finish = df_norm[df_norm["Stage"] == "Finishing"][rank_col]

    plt.figure(figsize=(6, 5))
    plt.boxplot([data_struct, data_finish],
                labels=["Structural Phase", "Finishing Phase"])
    plt.ylabel("ClimateDurabilityScore (0–10)")
    plt.title("Climate Durability by Construction Phase")
    plt.tight_layout()
    plt.show()
else:
    print("Need 'Stage' and 'ClimateDurabilityScore' columns for boxplot.")
#Clustering:
#We'll cluster using:MoistureRiskIndex, ThermalSensitivityIndex, ClimateDurabilityScore

cluster_features = ["MoistureRiskIndex", "ThermalSensitivityIndex", "ClimateDurabilityScore"]
available_cluster_features = [c for c in cluster_features if c in df_norm.columns]

if available_cluster_features:
    X = df_norm[available_cluster_features].values

    # Standardize for clustering (even though they are 0–10, this makes KMeans more stable)
```

```
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# k=3 clusters for interpretation
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
clusters = kmeans.fit_predict(X_scaled)

df_norm["Cluster"] = clusters

# Cluster centers in original 0–10 feature space
cluster_centers = pd.DataFrame(
    scaler.inverse_transform(kmeans.cluster_centers_),
    columns=available_cluster_features
)

print("=== Cluster centers (in 0–10 engineered feature space) ===")
display(cluster_centers)

# Scatter plot of clusters (MoistureRiskIndex vs ClimateDurabilityScore)
if "MoistureRiskIndex" in df_norm.columns and "ClimateDurabilityScore" in df_norm.columns:
    plt.figure(figsize=(6, 5))
    for cl in sorted(df_norm["Cluster"].unique()):
        subset = df_norm[df_norm["Cluster"] == cl]
        plt.scatter(subset["MoistureRiskIndex"],
                    subset["ClimateDurabilityScore"],
                    label=f"Cluster {cl}")
    plt.xlabel("MoistureRiskIndex (0–10)")
    plt.ylabel("ClimateDurabilityScore (0–10)")
```

```
        plt.title("Clusters by Moisture Risk and Climate Durability")
        plt.legend()
        plt.tight_layout()
 plt.show()
else:
    print("Engineered feature columns for clustering not found.")
#3D Scatter plot
from mpl_toolkits.mplot3d import Axes3D
fig = plt.figure(figsize=(8, 6))
ax = fig.add_subplot(111, projection="3d")

for cl in sorted(df_norm["Cluster"].unique()):
    subset = df_norm[df_norm["Cluster"] == cl]
    ax.scatter(
        subset["MoistureRiskIndex"],
        subset["ThermalSensitivityIndex"],
        subset["ClimateDurabilityScore"],
        label=f"Cluster {cl}",
        s=40
    )

ax.set_xlabel("MoistureRiskIndex (0–10)")
ax.set_ylabel("ThermalSensitivityIndex (0–10)")
ax.set_zlabel("ClimateDurabilityScore (0–10)")
ax.set_title("3D Scatter of MRI, TSI, CDS with Cluster Labels")
ax.legend()

plt.tight_layout()
```

```
plt.show()

#Modeling and scoring: FinalClimateScore_raw=0.3(10−MRI)+0.2(10−TSI)+0.5(CDS)

# Check columns exist

for col in ["MoistureRiskIndex", "ThermalSensitivityIndex", "ClimateDurabilityScore"]:

    if col not in df_norm.columns:

        raise ValueError(f"Required column missing: {col}")


# Moisture-dominant weights

w_mri = 0.5   # MoistureRiskIndex

w_tsi = 0.2   # ThermalSensitivityIndex

w_cds = 0.3   # ClimateDurabilityScore


# Raw weighted score

df_norm["FinalClimateScore_raw"] = (

    w_mri * (10 - df_norm["MoistureRiskIndex"]) +

    w_tsi * (10 - df_norm["ThermalSensitivityIndex"]) +

    w_cds * df_norm["ClimateDurabilityScore"]

)


# Rescale FinalClimateScore_raw to 0–10

fs_min = df_norm["FinalClimateScore_raw"].min()

fs_max = df_norm["FinalClimateScore_raw"].max()


if fs_max == fs_min:

    df_norm["FinalClimateScore"] = 5.0  # degenerate case

else:

    df_norm["FinalClimateScore"] = 10 * (df_norm["FinalClimateScore_raw"] - fs_min) / (fs_max -
        fs_min)
```

```
df_norm[["Material", "Stage", "MoistureRiskIndex", "ThermalSensitivityIndex",
        "ClimateDurabilityScore", "FinalClimateScore"]].head()
#Rank materials by FinalClimateScore
# Sort descending by FinalClimateScore
df_ranked = df_norm.sort_values("FinalClimateScore", ascending=False)

cols_to_show = [
    "Material", "Stage",
    "FinalClimateScore",
    "MoistureRiskIndex", "ThermalSensitivityIndex", "ClimateDurabilityScore"
]

# Top 10 overall
top10_overall = df_ranked.head(10)[cols_to_show]
print("=== Top 10 materials by FinalClimateScore (overall) ===")
display(top10_overall)

# Top 5 Structural Phase
top5_struct = df_ranked[df_ranked["Stage"] == "Structural"].head(5)[cols_to_show]
print("\n=== Top 5 STRUCTURAL materials by FinalClimateScore ===")
display(top5_struct)

# Top 5 Finishing Phase
top5_finish = df_ranked[df_ranked["Stage"] == "Finishing"].head(5)[cols_to_show]
print("\n=== Top 5 FINISHING materials by FinalClimateScore ===")
display(top5_finish)
#Bar chart of Top 10 final scores
```

```
plt.figure(figsize=(10, 5))
plt.bar(top10_overall["Material"], top10_overall["FinalClimateScore"])
plt.xticks(rotation=90)
plt.ylabel("FinalClimateScore (0–10)")
plt.title("Top 10 materials by FinalClimateScore (Moisture-dominant weights)")
plt.tight_layout()
plt.show()
#Save updated normalized sheet with FinalClimateScore
out_path = "guilan_construction_normalized_with_final_score.xlsx"
df_norm.to_excel(out_path, sheet_name="Normalized", index=False)


out_path
```