Evaluating Material Performance for Space Applications

A Data Science Approach Using NASA's Outgassing Data

Project Outline

01

PROJECT OBJECTIVES

Objectives & Outcome

02

THE PROCESS

Data Processing & Analysis

Model Development & Performance

03

OUTCOME & CONCLUSION

Overall Conclusion & Steps Moving Forward

Project Objectives

Project Objectives

01

Evaluate the performance of different materials based on the test values of materials they provide.

+

02

Use ML methods to create a performance evaluation framework to consistently monitor and assess material reliability and material quality.

=

OUTCOME

Simple Framework for quality assessment.

The Process

Data Processing & Analysis

Data Description

Project Data: The data is from NASA's Outgassing Database, which tests materials for space use. *(dataset "Outgassing Db" on NASA's Open Data Portal provides information on materials that have been tested for outgassing, which is the release of gas that was dissolved, trapped, frozen, or absorbed in some material.)*

Source: *https://data.nasa.gov/Applied-Science/Outgassing-Db/r588-f7pr/about_data*

Key Features (for identifying quality & reliability):

- **Total Mass Loss (TML):**
  measures how much material is lost in space conditions.
  (Lower values are better.)

- **Collected Volatile Condensable Material (CVCM):**
  measures potential contamination that can condense on a cold surface.
  (Lower the value, the better it is for keeping surfaces clean in space.)

- **Water Vapor Regained (WVR):**
  measures how much moisture a material absorbs.
  (Lower values are preferred to avoid issues like rust.)

Data Preparation

Preparation Steps:

1. **Cleaning Data**
   Cleaned the data for accuracy *(removal of unnecessary data)*.

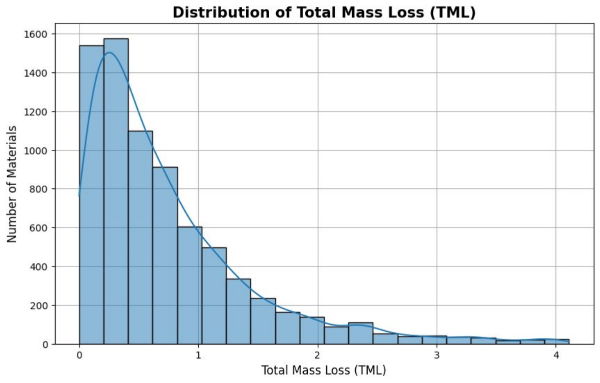
2. **Handling Duplicates & Outliers:**
   Removed duplicates and outliers.

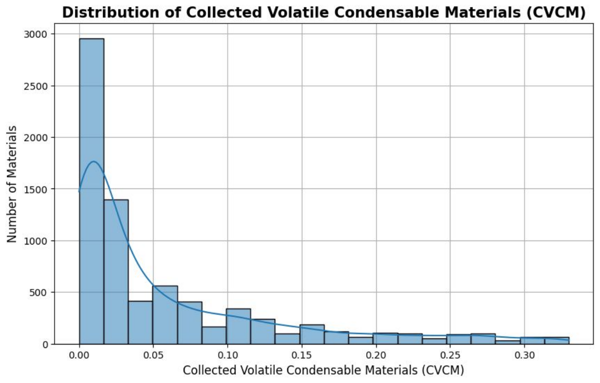
3. **Checking and Saving Updated Data**:
   Ensured data quality for EDA (Exploratory Data Analysis).

Exploratory Data Analysis

Individual Material Performance - Overall Data Distribution

Key Takeaway : Majority of the materials have low Values





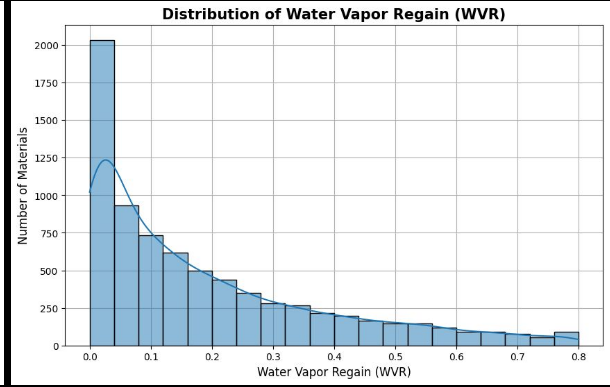


Diagram shows that the **majority of materials have a low TML, with a significant number of materials having TML values close to zero**.

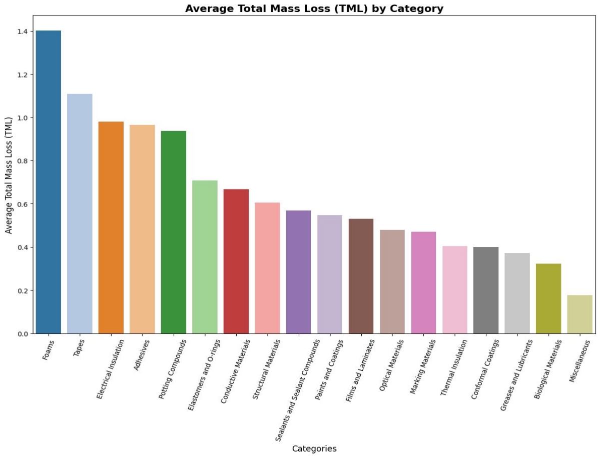
Diagram for CVCM displays a similar pattern to TML, where **most materials have very low CVCM values, clustering near zero.**

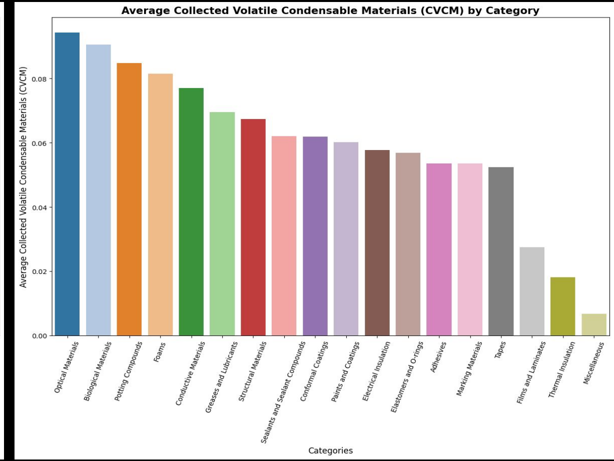
The WVR distribution shows that the **majority of materials have WVR values close to zero,** with the frequency decreasing as WVR increases.

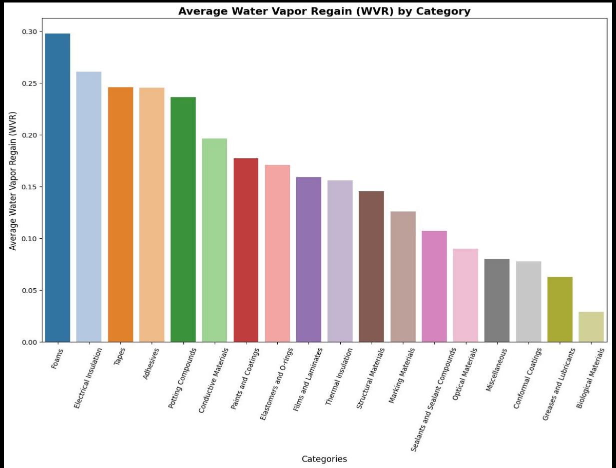
Exploratory Data Analysis

Categorical Material Performance

Key Takeaway : How different material categories perform based on key features.







- The bar chart shows that **Foams have the highest average TML, followed by Tapes and Electrical Insulation**.
- **Miscellaneous and Biological Materials have the lowest average TML**, indicating their better performance in terms of lower mass loss.

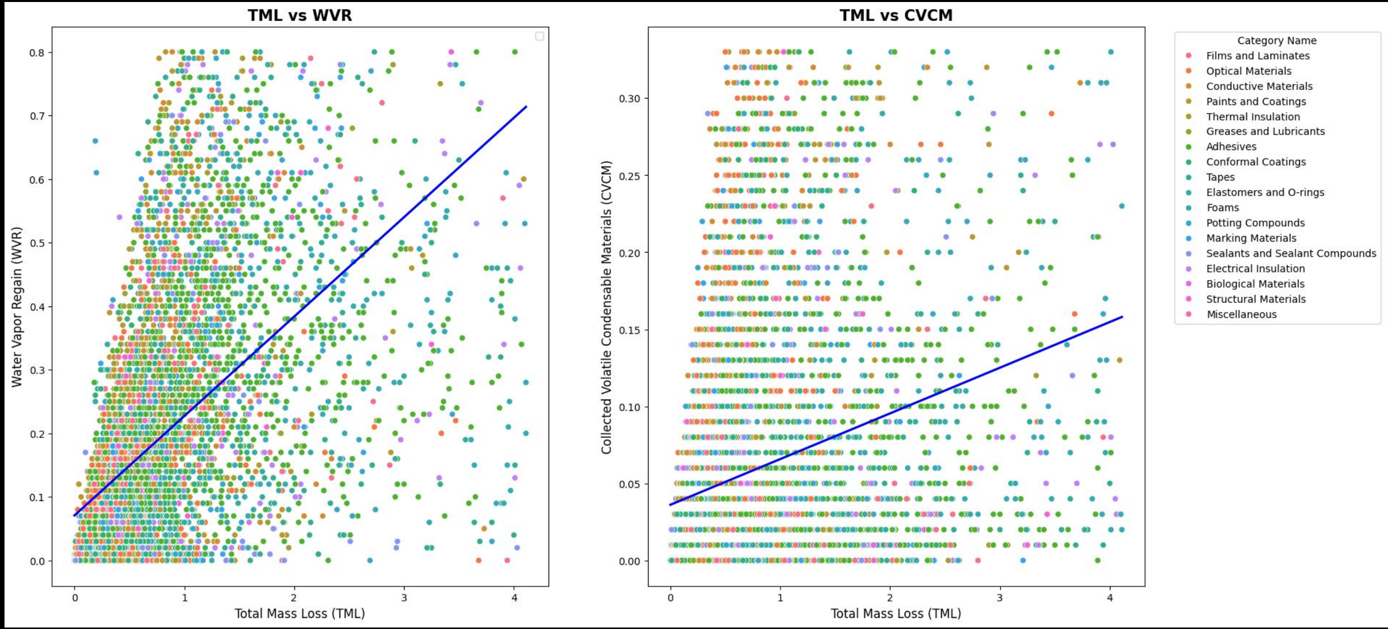
- **Biological Materials have the highest average CVCM values,** indicating higher condensable volatile material loss.
- **Thermal Insulation and Miscellaneous have the lowest average CVCM**, showing better performance in terms of lower volatile condensable material loss.

- Similar to the TML chart, **Foams have the highest average WVR, followed by Electrical Insulation and Tapes.**
- **Categories like Biological Materials and Miscellaneous have the lowest WVR,** indicating less moisture absorption.

Exploratory Data Analysis

Material Relationship Performance

Key Takeaway : Materials with high TML generally have high WVR and CVCM.



- The scatter plots show a diverse distribution of materials across different TML, WVR, and CVCM values.

- **General positive correlation between TML, WVR and CVCM**.

Findings and Insights

TML, WVR and CVCM Relationship: The positive correlation between Total Mass Loss (TML), Water Vapour Regain(WVR), and Collected Volatile Condensable Material(CVCM) indicates that materials prone to higher mass loss are also likely to absorb more water vapor and produce contamination that can condense on material surfaces.

Consistent Performance Across Categories: High-performing materials(low TML WVR, and CVCM) are generally distributed across various categories, indicating that achieving high-quality materials is not restricted to specific categorical types.

Insights from Exploratory Data Analysis suggests **Quality Control Measures** can be applied across the board to improve material selection process.

The Process

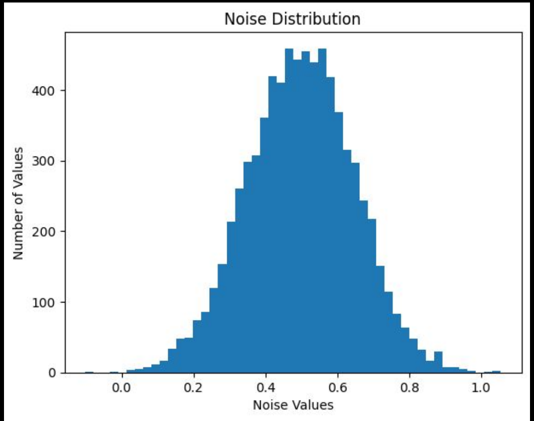
Model Development & Performance

Feature Engineering

Methods Taken:

Normalizing Features:
To ensure all features are on a common scale, TML, CVCM, and WVR are normalized using **MinMaxScaler()** function. This scales the features between 0 and 1.

Adding Noise:
Added some random noise to make the data more realistic. The noise values are generated from a normal distribution with a mean of 0.5 and a standard deviation of 0.15 *(to create value distribution between 0 and 1)*.

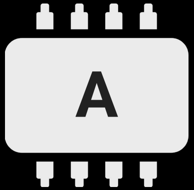


Performance Score Calculation:
The performance score is calculated using the formula

$$\text{performance\_score} = (1 - \text{TML}) + (1 - \text{CVCM}) + (1 - \text{WVR}) + \text{noise}$$

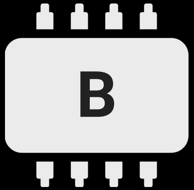
(reversed interpretation of the feature values, performance_score calculated to show higher scores for lower values)

Models



Regression Model:

Objective: Predict material performance scores based on test variables (features)



Classification Model:

Objective: Classify material performance

Based on material performance scores.

Model Performance

Regression Model:

Objective: Predict material performance.

Models:

Linear Regression: Performed with

R^2 of 0.90 = 90% indicating it could explain 90% of the variability in material performance.

Random Forest Regressor:

Performed with

R^2 of 0.87 indicating it could explain **87%** of the variability in material performance.

Model Performance

Classification Model:

Objective: Classify material performance.

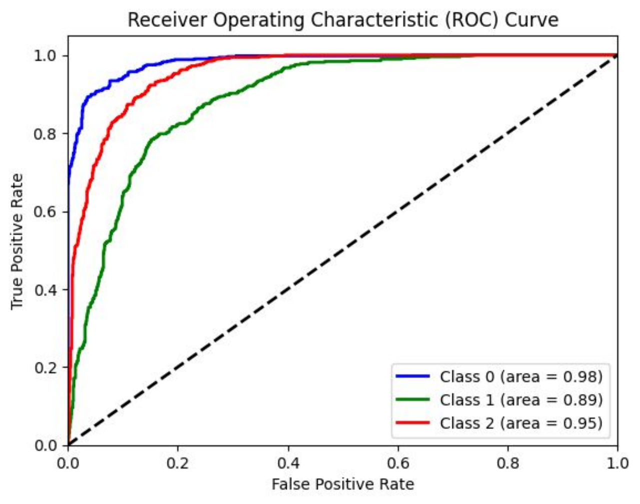
Model:

Logistic Regression: Performed with an **Overall accuracy of 82%**

Overall Performance ROC Curve:

The closer the curve follows the left-hand border the better the model is at distinguishing between the classes.

An AUC (Area Under Curve) provides an aggregate measure of performance across all possible classification thresholds.
Value close to 1 indicates a very good model performance.



Class 0 = Low Performance values

Class 1 = Medium Performance values

Class 2 = High Performance values

Conclusion

"To summarize,

The project is aimed to evaluate material performance and develop a evaluation framework.

The model insights can help in optimizing material selection, implementing proactive quality control, and help drive R&D innovations.”

Moving Forward

Further Research Recommendations:

- **Expand Dataset:** Include more materials and test values to improve the models performance and accuracy.
- **Experiment with more Classification Models and improvement in evaluations of the models.**
- **Incorporate Categorical Features:** Analyze the impact of different material categories on performance.

Potential applications of the Model:

- **Optimized Material Selection:** Leverage predictive analytics to choose the best materials for specific space applications, reducing costs and improving reliability.
- **Proactive Quality Control:** Implement continuous monitoring systems to detect material performance issues early.
- **R&D Innovations:** Apply model findings to drive research and development efforts, focusing on improving existing materials or developing new ones with better performance characteristics.

References

- **Data Source:** *https://data.nasa.gov/Applied-Science/Outgassing-Db/r588-f7pr/about_data*
- **Information about outgassing:** *https://www.kistler.com/INT/en/outgassing/C00000135*
- Outgassing additional information: *https://ntrs.nasa.gov/api/citations/20030053424/downloads/20030053424.pdf*

Image Source: *https://wallpapercrafter.com/238996-___space-rocket-launch-and-reflection-hd.html*

NOTE: *For more in depth EDA processes, further analysis, and detailed model performance & evaluations please view my notebooks.*

Thank you.