Data-Driven Classification of Canadian Crude Oils

Ryan Bulger

University of Colorado Boulder

DTSA 5506

ABSTRACT

This project investigates whether the current classification of crude oil grades accurately reflects their underlying chemical and physical similarities. Using unsupervised machine learning techniques, the analysis explores whether natural groupings emerge within the data that differ from existing defined categories. The objective is to determine if data-driven clustering can uncover more meaningful relationships among crude grades, potentially revealing a simpler and more representative grouping structure that better captures the true variation in crude oil characteristics.

1. INTRODUCTION

This study addresses the problem of whether hidden patterns within Canadian crude oil quality data can reveal more natural groupings of crude oils than those defined by current classification systems. Accurate grouping is important because the existing categories, often based only on density and sulfur content, may not fully capture the chemical and physical complexity of each crude. As a result, misclassification can distort pricing, complicate blending and transport decisions, and reduce overall efficiency in the supply chain.

To address these limitations, this project applies unsupervised machine learning techniques to cluster crude oils based on a comprehensive range of laboratory-measured properties. By uncovering data-driven groupings that reflect intrinsic relationships between crudes, the analysis provides a more objective framework for understanding quality variation. The findings can support more accurate market valuation, optimize pipeline and blending operations, and improve refinery yield predictability within the Canadian energy sector.

The analysis proceeds through several structured stages: data summary to understand the dataset's structure, exploratory data analysis (EDA) to identify patterns and correlations, and data cleaning to handle missing or inconsistent values. Three clustering algorithms, K-Means, Agglomerative Clustering, and Gaussian Mixture Models (GMM), are then applied and evaluated using multiple performance metrics and visualized through PCA and t-SNE. The unsupervised results are later compared with supervised learning models, Multinomial Logistic Regression, Support Vector Classifier, and Gradient Boosting, to assess alignment with true class labels. The project concludes with final discussion and takeaways, highlighting key insights and implications for applying machine learning to crude quality analysis in the energy industry.

2. RELATED WORK

Over the past two decades, several approaches have been developed to classify crude oil grades based on their physical and chemical properties. Traditional methods rely on API gravity and sulfur content to distinguish between light, medium, and heavy crudes, as well as sweet and sour categories [1]. While effective for general use, these classifications overlook finer compositional differences. Chemical engineering studies have expanded on this by incorporating indices such as the Watson characterization factor and correlation index to better capture variations in paraffinic and aromatic content [2].

More recently, researchers have introduced spectroscopic and chemometric techniques to automate and improve accuracy in crude classification. Studies using Fourier Transform Infrared (FTIR) spectroscopy, Near-Infrared (NIR) spectroscopy, and Nuclear Magnetic Resonance (NMR) have demonstrated that supervised machine learning models such as Support Vector Machines and Partial Least Squares Discriminant Analysis (PLS-DA) can predict crude oil grades and key properties with high accuracy [3][4]. These methods provide faster, non-destructive analysis compared to conventional lab assays.

Unsupervised statistical methods have also been used to discover natural groupings within crude oil datasets. Techniques like K-Means, hierarchical clustering, and self-organizing maps have revealed that many crude oils can be clustered into fewer groups than their assigned categories, often aligning with geographic origin or refining characteristics [5]. This supports the idea that data-driven clustering can yield more meaningful categorizations than fixed classification systems.

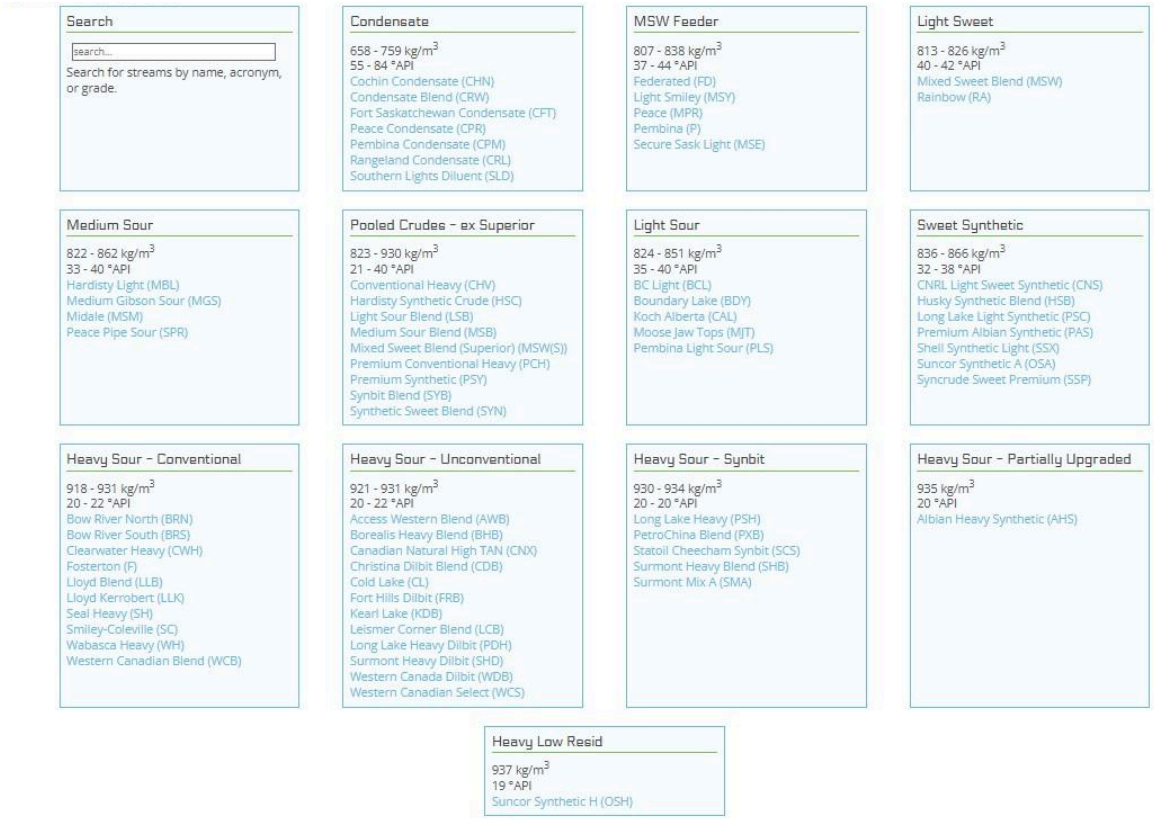
The proposed study builds upon this prior work by combining multiple clustering techniques, K-Means, Agglomerative Clustering, and Gaussian Mixture Models, and comparing them with supervised learning models. Unlike previous studies that focused on narrow datasets or specific analytical tools, this project integrates a broad range of crude quality features to evaluate whether unsupervised methods can uncover more representative groupings, offering both technical and commercial insights into crude oil classification.

3. METHODOLOGY

3.1 Data Source

The dataset for this project is derived from CrudeMonitor.ca, an established platform managed by Crude Quality Inc., which collects and publishes detailed crude oil assay data from pipelines and refineries across Western Canada [6]. Each assay provides extensive information on the chemical and physical characteristics of crude oils, including density, sulfur content, metals (nickel, vanadium), distillation yields, and hydrocarbon compositions. As shown in Figure 1, CrudeMonitor categorizes these crude oils into 12 defined groups based on quality and refining characteristics. The platform serves as the industry-standard source for Canadian crude oil quality data, used extensively by producers, refiners, marketers, research institutions, and government agencies. Because the data is both publicly accessible and quality-assured, CrudeMonitor is an ideal and trusted source for developing a transparent and reproducible dataset for this project.

Figure 1
CrudeMonitor Crude Oil Grades



The project is being developed in Python using Jupyter Notebooks within Visual Studio Code (VS Code). All scripts and supporting materials are stored and version-controlled through a public GitHub repository. The dataset itself is relatively small so it is stored in CSV format rather than a database, allowing for efficient access and portability within the repository.

Initially, CrudeMonitor defined 12 groups of Canadian crude oils; however, upon further inspection, the Condensate group was found to contain a unique set of features that are not directly comparable to the remaining crude grades. To maintain analytical consistency, the Condensate group was excluded from the study, resulting in a final dataset containing 11 crude oil grades for clustering analysis.

To construct the working dataset, a Python function was developed to interact directly with the CrudeMonitor API. This function loops through each of the 11 defined crude groups, extracts the associated crude assay data, and converts it into structured DataFrames. Once all groups are collected, the DataFrames are concatenated into a single unified dataset for analysis. This automated process ensures consistency across all groups and simplifies future data updates if new assays or crude types are introduced.

3.2 Data Summary

The first stage of analysis involved a systematic review of the dataset's structure and composition. After loading the dataset, it was found to contain 8,561 rows and 133 columns. The dataset consists primarily of numeric features, with 128 columns of type float32, one column of type datetime64[ns], and four columns of type object, which represent categorical or identifier fields.

A detailed table summarizing each column including its description, physical property, unit of measurement, and expected value range is presented in Figure 2. This summary establishes a clear understanding of the dataset's structure and provides the foundation for determining appropriate preprocessing, encoding, and modeling strategies in the subsequent stages of analysis.

Figure 2
Feature Descriptions

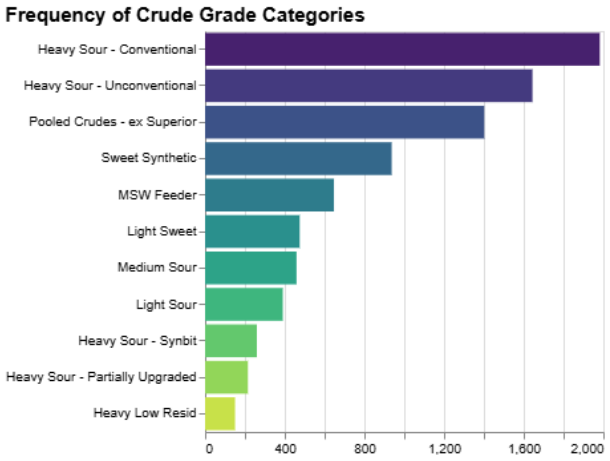
Feature	Description
Crude	Name of the crude oil grade.
Batch	Shipment or lot identifier for the sample.
Sample Date (yyyy-mm-dd)	Date the crude sample was collected.
Location	Site or facility where the sample was taken.
Density (kg/m³) [ASTM D5002]	Mass per unit volume; indicates heaviness of crude.
Gravity (°API) [ASTM D5002]	Measure of crude lightness; inverse of density.
Sulphur (wt%) [ASTM D4294]	Sulphur content; affects refining and emissions.
Micro Carbon Residue (wt%) [ASTM D4530]	Carbon left after pyrolysis; indicator of coke-forming tendency.
Sediment (ppmw) [ASTM D4807]	Solid impurities or particles in crude.
Total Acid Number (mgKOH/g) [ASTM D664]	Measure of acidity; indicates corrosive potential.
Salt (ptb) [ASTM D3230]	Salt concentration; impacts corrosion and desalting needs.
Nickel (mg/kg) [ASTM D5708A]	Metal contaminant affecting catalyst life.
Vanadium (mg/kg) [ASTM D5708A]	Metal impurity impacting refining catalysts.
C1–C10 Components (vol%) [GC/FID]	Light hydrocarbon composition (methane to decane) by gas chromatography; indicates volatility and gas content.
iC4–nC5 (vol%) [GC/FID]	Branched and normal butane/pentane fractions; measure of light-end structure.
C6–C10 (vol%) [GC/FID]	Heavier paraffins; affect vapor pressure and yield profiles.
Nitrogen – C4- (vol%) [GPA 2177M]	Nitrogen content in light gas fraction.
Carbon Dioxide – C4- (vol%) [GPA 2177M]	CO_2 concentration in gas fraction.
C1–nC4 – C4- (vol%) [GPA 2177M]	Light hydrocarbon breakdown (methane through butane) in gas stream.
IBP (°C) [ASTM D7169]	Initial boiling point; start of vaporization during distillation.
1–99 Mass% Recovered (°C) [ASTM D7169]	Distillation temperatures where 1–99% of sample mass is vaporized; describes boiling range and fractionation behavior.
target_label	CrudeMonitor crude group name

3.3 Exploratory Data Analysis

The exploratory analysis phase began with organizing the dataset into logical column groups to ensure consistency throughout the analysis process. A categorical group was created containing columns such as Sample Date, Location, and Crude. A second group of numeric columns was defined to capture features describing crude quality (e.g., density, sulfur, metals, and hydrocarbon composition). Finally, a third group of numeric columns was established for the distillation temperature measurements.

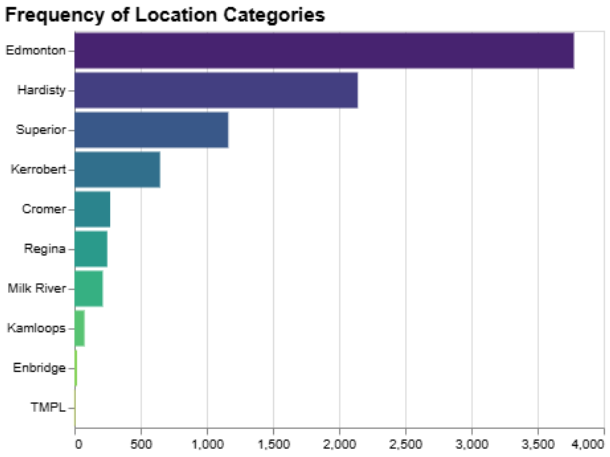
A frequency chart of Crude Grade Categories was created (Figure 3). The dataset is dominated by Heavy Sour – Conventional, Heavy Sour – Unconventional, Pooled Crudes – ex Superior, and Sweet Synthetic grades. This is expected since Western Canada primarily produces heavy crude oils derived from oil sands operations.

Figure 3
Crude Grade Categories



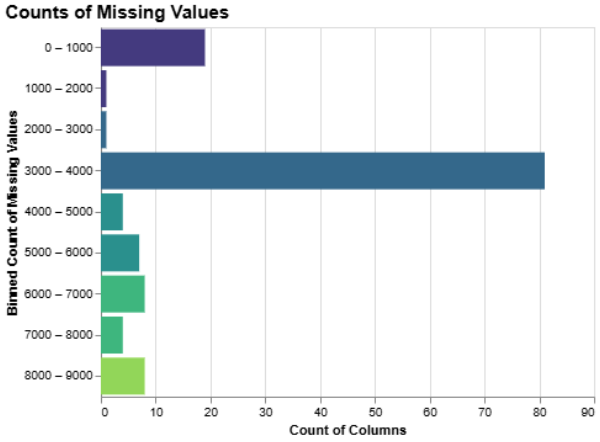
A frequency chart of Location Categories (Figure 4) shows that most samples originate from Edmonton, Hardisty, and Superior. This distribution aligns with the structure of Western Canada's pipeline system, where heavy crude and oil-sands by-products typically flow south from northern Alberta through Edmonton and then on to Hardisty, a major crude-oil hub.

Figure 4
Location Categories



A histogram comparing the number of columns to the binned count of missing values revealed that most columns have between 3,000 and 4,000 missing entries (Figure 5).

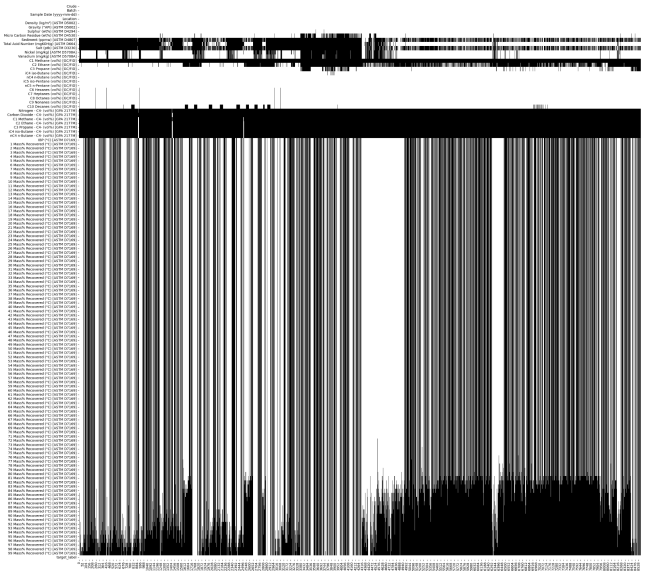
Figure 5
Missing Value Counts



To visualize missing data more comprehensively, a missing-values heatmap was generated (Figure 6). The heatmap showed several columns that are almost entirely empty. These columns will be dropped during data cleaning, as imputing values where data are scarce would add bias. The proposed columns to drop are:

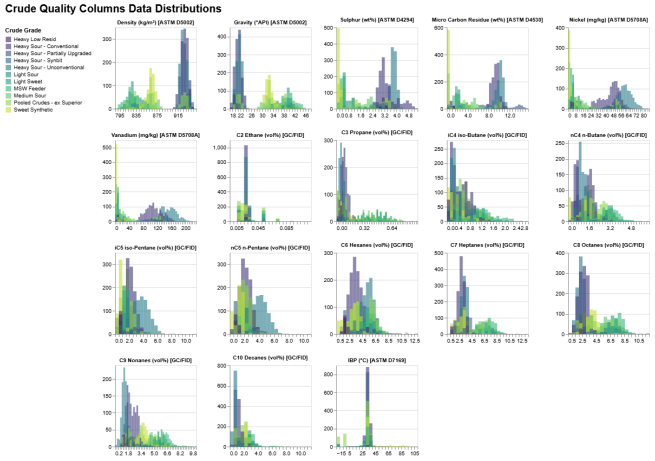
- Sediment (ppmw) [ASTM D4807],
- Total Acid Number (mgKOH/g) [ASTM D664],
- Salt (ptb) [ASTM D3230],
- C1 Methane (vol%) [GC/FID],
- Nitrogen – C4- (vol%) [GPA 2177M],
- Carbon Dioxide – C4- (vol%) [GPA 2177M],
- C1 Methane – C4- (vol%) [GPA 2177M],
- C2 Ethane – C4- (vol%) [GPA 2177M],
- C3 Propane – C4- (vol%) [GPA 2177M],
- iC4 iso-Butane – C4- (vol%) [GPA 2177M], and
- nC4 n-Butane – C4- (vol%) [GPA 2177M].

Figure 6
Missing Values Heatmap



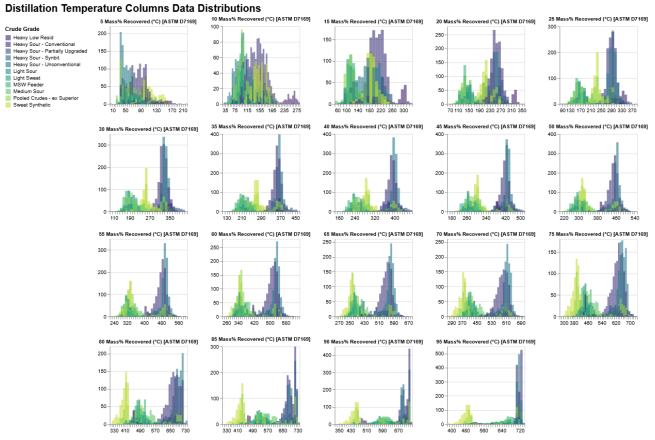
Histograms of crude quality columns were then plotted, color-coded by target group (Figure 7). These plots reveal distinct groupings among certain crude grades, suggesting that several features (e.g., density, sulfur, metals, and hydrocarbon composition) are strong indicators of grade similarity and could enable the dataset to be clustered into fewer groups.

Figure 7
Crude Quality Distributions



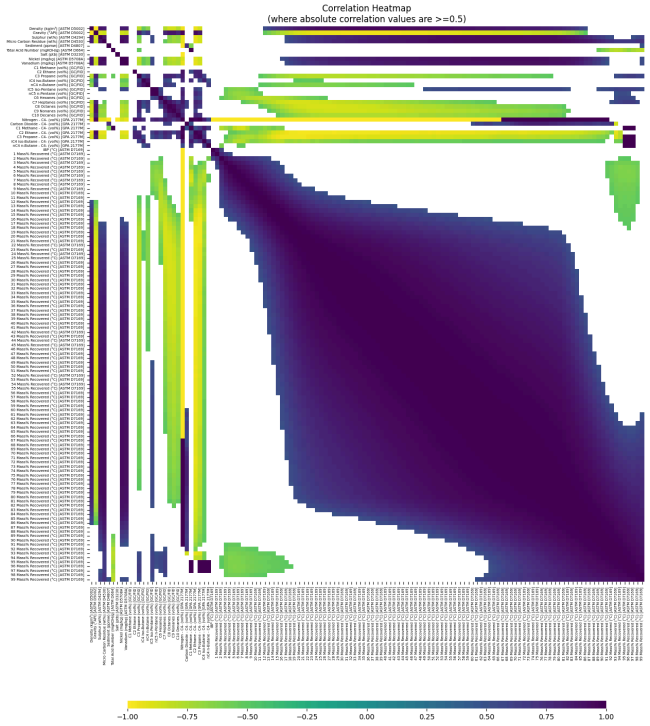
Next, histograms of the distillation temperature columns were created (Figure 8). To avoid excessive visual complexity, only every fifth column was included out of the total 99. The resulting patterns show clear separation among crude grades, reinforcing that the dataset contains identifiable chemical signatures suitable for clustering.

Figure 8
Distillation Temperature Distributions



Finally, a correlation heatmap was produced (Figure 9), limited to variables with correlation coefficients less than −0.5 or greater than 0.5 to focus on the strongest relationships. Several notable positive and negative correlations were observed among the crude-quality variables, and strong positive correlations were found among the distillation temperature columns which is an expected result since these measurements originate from the same experimental process.

Figure 9
Correlation Heatmap



3.4 Pre-Processing

The preprocessing stage prepared the dataset for modeling. After importing the data from CrudeMonitor.ca, the column types were first converted into categorical, numeric, and datetime groups to ensure consistent handling throughout analysis. The data was then split into features (X) and the target (y) variable. The feature matrix X combines the three column groups defined during the EDA stage: a categorical group containing the Sample Date, Location, and Crude columns; a set of numeric columns describing the crude quality characteristics; and a second set of numeric columns representing the distillation temperature measurements. The target labels representing the eleven crude grades were then encoded into integer values using Scikit-Learn's LabelEncoder to prepare them for modeling and external validation metrics.

A Scikit-Learn ColumnTransformer pipeline was created to standardize and automate the preprocessing workflow. The pipeline includes a FunctionTransformer that removes the columns with excessive missing values identified during EDA. A second FunctionTransformer adds two datetime-derived columns, Month and Year, where the Month value represents the count of months since the earliest sample date in the dataset. This approach ensures that the model treats time as a continuous sequence rather than resetting every calendar year, since machine learning algorithms do not inherently understand that month 1 follows month 12. A KNNImputer using ten nearest neighbors is then applied to estimate missing values in the numeric crude quality and distillation temperature columns. After imputation, a StandardScaler is applied to all numeric columns to normalize their magnitudes and prevent features with large ranges from dominating the clustering process. Finally, a OneHotEncoder is applied to the Location category to convert it into binary variables suitable for modeling.

Once the pipeline was configured, the feature matrix X was processed through the ColumnTransformer to produce a fully preprocessed dataset ready for clustering and dimensionality reduction.

3.5 K-means Clustering

The first model I will be using is the K-Means clustering algorithm, one of the most widely used unsupervised learning methods for identifying patterns within multivariate data. K-Means partitions the dataset into a predefined number of clusters by minimizing the sum of squared distances between data points and their respective cluster centroids. The algorithm begins by assigning random centroids, then iteratively reassigns data points and recalculates centroids until convergence is reached and cluster assignments stabilize. The result is a compact grouping of data points where each point is closer to its own cluster centroid than to any other.

K-Means is computationally efficient and performs well on large datasets with continuous numeric features, making it a strong initial choice for clustering high-dimensional crude assay data. However, the algorithm assumes spherical clusters of similar variance and is sensitive to feature scaling and the initial choice of cluster centroids. It may also struggle with overlapping clusters or when the true underlying structure is not well separated. Despite these limitations, its simplicity, speed, and interpretability make it a practical starting point for exploring the underlying structure in crude oil quality data.

I am choosing to include K-Means in this project because it will provide a clear and interpretable baseline for understanding the primary cluster formations within the crude dataset. Since the goal of the analysis is to determine whether natural groupings exist that differ from current classifications, K-Means offers a straightforward way to observe initial cluster boundaries and compare them against known crude oil groups. Its performance will also serve as a benchmark for evaluating more complex models.

3.6 Agglomerative Clustering

The second model used will be Agglomerative Clustering, a type of hierarchical clustering that builds nested groupings by progressively merging the most similar data points or clusters. The process begins with each observation as its own cluster and iteratively combines pairs based on a chosen linkage criterion, such as average, complete, or Ward's linkage, until all data points are merged into a single hierarchy. The resulting structure is often visualized as a dendrogram, which provides insight into how clusters form and merge at different levels of similarity.

Agglomerative Clustering has the advantage of not requiring a predefined number of clusters, allowing flexibility to explore hierarchical relationships within the dataset. It also performs well on data where the natural clusters vary in shape or size. However, it can be computationally expensive for large datasets and may be sensitive to noise and outliers. Additionally, once clusters are merged, the algorithm cannot reverse previous steps, making it somewhat less flexible than iterative partitioning methods like K-Means.

This model is being included in the project because it can reveal hierarchical relationships that K-Means may overlook. In the context of crude oil quality data, hierarchical clustering can help visualize how different grades relate to one another and whether certain groups form subclusters with shared characteristics. This can provide a richer understanding of how crude oils group together beyond simple distance-based partitioning.

3.7 Gaussian Mixture Model Clustering

The third model applied will be the Gaussian Mixture Model (GMM), a probabilistic clustering approach that assumes data points are generated from a mixture of multiple Gaussian distributions, each representing a cluster. Unlike K-Means, which assigns points to a single cluster, GMM uses a soft clustering approach where each data point has a probability of belonging to each cluster. The model uses the Expectation-Maximization (EM) algorithm to iteratively estimate the parameters of these distributions, means, covariances, and mixing weights, until the likelihood of the data under the model is maximized.

GMM is more flexible than K-Means because it can capture clusters of different shapes, sizes, and orientations due to its use of covariance matrices. It also provides probabilistic cluster memberships, which can be valuable for interpreting ambiguous cases where samples exhibit characteristics of multiple groups. However, GMM can be sensitive to initialization, prone to overfitting if too many components are chosen, and may require more computational effort than simpler models.

I am selecting GMM for this project because its probabilistic framework aligns well with the goal of exploring nuanced relationships among crude oil grades. Many crude samples exhibit overlapping physical and chemical properties, and GMM's ability to model such uncertainty provides a more realistic representation of these boundaries. This model will complement the results from K-Means and Agglomerative Clustering by offering a probabilistic perspective on how crude grades may overlap or transition between one another, thereby helping assess whether the existing classification system adequately reflects the natural structure of the data.

4. MODEL EVALUATION

4.1 Clustering Scoring Metrics

The performance of the clustering models will be evaluated using a combination of internal and external validation metrics. Internal metrics assess clustering quality using only the model's output, without reference to true labels, while external metrics require the known target labels to measure how closely the model's cluster assignments align with the actual classes. By employing both types, the evaluation will capture both the intrinsic structure of the data and the extent to which that structure reflects the existing classification system.

Three internal metrics will be applied to each clustering result. The Silhouette Score measures how well each sample fits within its assigned cluster compared to neighboring clusters, with higher values indicating better-defined separation. The Davies–Bouldin Index (DBI) calculates the average ratio of intra-cluster similarity to inter-cluster separation, where lower scores represent better clustering. The Calinski–Harabasz Index (CHI) assesses the ratio of between-cluster dispersion to within-cluster dispersion, with higher scores indicating more distinct cluster boundaries. Together, these internal metrics provide a robust assessment of cluster compactness and separation.

External validation will use three additional metrics that compare predicted clusters to the true crude group labels. The Adjusted Rand Index (ARI) measures the similarity between two clusterings while correcting for chance alignment, where higher values

suggest closer agreement. The Normalized Mutual Information (NMI) quantifies how much information is shared between the predicted and true labels, with values closer to 1 indicating greater mutual dependence. Lastly, the Fowlkes–Mallows Index (FMI) calculates the geometric mean of precision and recall between predicted and actual labels, balancing false positives and false negatives. These external metrics provide insight into how accurately the unsupervised models capture known classifications.

4.2 Clustering Plots

To visually interpret the clustering performance, a set of three plots will be created for every model and cluster size combination. The first is a Silhouette Plot, which displays the silhouette coefficient for each sample within its assigned cluster. The average silhouette score is also plotted as a reference line. When clusters are well-defined, the silhouette coefficients will be high and relatively consistent in width across groups, indicating that points are strongly associated with their clusters and well-separated from others. In contrast, negative or unevenly distributed silhouettes suggest overlapping or poorly defined clusters. The ideal cluster configuration is typically identified where the mean silhouette score is maximized, and cluster widths are approximately uniform.

The second visualization will be a Principal Component Analysis (PCA) scatter plot. PCA reduces the multidimensional dataset into two principal components that capture the highest variance in the data, allowing the clusters to be plotted in two dimensions. While PCA does not always preserve nonlinear relationships, it provides a straightforward and interpretable way to observe how the algorithm has partitioned the data and whether distinct boundaries exist between clusters.

The third visualization will use t-distributed Stochastic Neighbor Embedding (t-SNE), a nonlinear dimensionality reduction technique designed to preserve local relationships between data points. Unlike PCA, t-SNE focuses on maintaining relative distances between nearby points, which makes it particularly effective for visualizing high-dimensional data such as crude assay features. The resulting 2D plots often reveal more nuanced groupings and substructures that may not be visible in PCA space.

Finally, to directly compare model predictions to known labels, three 11×11 heatmaps, one for each clustering model, will be created for a cluster size of 11. Each heatmap will display the relationship between predicted clusters and actual crude groups, illustrating how well each model aligns with existing classifications. This provides an intuitive summary of model performance and misclassification patterns.

4.3 Comparison to Supervised Learning

To further evaluate the clustering models, their performance will be compared to supervised machine learning algorithms trained on the same feature matrix and true target labels. The three supervised models selected are the Multinomial Logistic Classifier, Support Vector Classifier, and Gradient Boosting Classifier. Each of these algorithms represents a different approach to classification, logistic regression for linear separability, SVC for margin-based classification, and gradient boosting for nonlinear decision boundaries.

After training, the supervised models will be used to predict crude oil groups, and the results will be visualized through three additional 11×11 heatmaps, one for each model. These will mirror the layout of the unsupervised heatmaps, allowing for direct comparison between how the supervised and unsupervised approaches classify samples relative to the true labels.

This comparative evaluation will help determine how closely the data-driven, unsupervised clusters reflect the structure captured by supervised learning methods. If the unsupervised models achieve similar grouping performance without label guidance, it would suggest that the existing classification system may not fully represent the natural structure of the data, supporting the hypothesis that fewer or alternative groupings of crude oils may be more appropriate.

5. DISCUSSION

This project is currently at the midpoint of the planned two-week timeline and remains on schedule for completion within the next week. To date, the Data Source, Data Summary, Exploratory Data Analysis (EDA), Data Cleaning, and Pre-Processing stages have been completed. The next phase will involve developing and evaluating the machine learning models, which will be carried out over the coming week as originally planned.

During the course of the work so far, several important refinements have been made to improve data quality and model consistency. The number of crude oil categories was reduced from twelve to eleven after identifying that the Condensate group contained a distinct set of features not comparable to the other grades. Excluding it ensures that the remaining data represent a consistent feature space suitable for clustering. Exploratory analysis also revealed that most assays originated from Edmonton, Hardisty, and Superior, which aligns with the flow of heavy crude oil through Alberta's major pipeline hubs. Additionally, missing value analysis showed that several columns contained insufficient data to be imputed reliably, leading to their removal before modeling.

The preprocessing pipeline was successfully constructed using Scikit-Learn's ColumnTransformer to automate all feature preparation. This includes a FunctionTransformer for dropping columns with excessive missing data, another to generate time-based month and year features, a KNNImputer for numeric imputation, a StandardScaler for feature normalization, and a OneHotEncoder for categorical encoding. With this automated structure in place, the dataset is now fully prepared for clustering and dimensionality-reduction analysis.

Looking ahead, the next stage will involve applying and comparing multiple unsupervised learning algorithms such as K-Means, Gaussian Mixture Models, DBSCAN, and Hierarchical Clustering to determine whether the existing eleven crude grades can be reduced into a smaller number of meaningful groups. This modeling phase will also include validation through both internal and external metrics to assess the quality and interpretability of the clusters. The project remains on track and has established a strong analytical foundation to build the modeling component efficiently in the final week.

6. CONCLUSION

This project applies an unsupervised machine learning approach to evaluate whether the current classification of Canadian crude oil grades accurately reflects their chemical and physical relationships. The analysis focuses on clustering algorithms such as K-Means, Agglomerative Clustering, and Gaussian Mixture Models to identify natural groupings within the data that may reveal a more concise or representative structure than the existing classification system.

Throughout the project so far, significant progress has been made in developing a clean, structured, and reproducible dataset. The CrudeMonitor data has been consolidated into a unified format, refined from twelve to eleven crude groups to exclude the Condensate category due to its incompatible feature set. The dataset has undergone extensive exploratory analysis, revealing strong clustering tendencies among certain crude properties and clear correlations across distillation temperature measurements. A robust Scikit-Learn preprocessing pipeline has also been implemented to automate feature cleaning, imputation, scaling, and encoding, ensuring the data is prepared for modeling.

The upcoming modeling and evaluation phase will determine whether unsupervised learning can effectively replicate or simplify the industry's existing classification scheme. By comparing clustering results to known labels and validating performance through internal and external metrics, the project aims to provide a data-driven perspective on crude quality relationships. The findings could offer insight into how Canadian crude grades may be more efficiently grouped for refining, transportation, and market optimization.

7. REFERENCES

[1] Sancho, C., López, M., & Acha, E. (2021). *Clustering of crude oils using key quality parameters*. Energy & Fuels, 35(7), 5671–5678.

[2] Al-Dahhan, M., & Mahmood, A. (2019). *Characterization of Basra crude oils using Watson and correlation indices*. Journal of Petroleum Research and Studies, 28(1), 1–10.

[3] Mohammadi, A., Kamal, M. R., & Shahraki, F. (2020). *Rapid classification of crude oils using FTIR and chemometric methods*. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 239, 118488.

[4] Moro, M., dos Santos, L. M., & Sousa, J. (2021). *Prediction of crude oil properties using NIR and machine learning models*. Fuel, 289, 119918.

[5] Olea, R. A., Houseknecht, D. W., & Garrity, C. P. (2021). *Cluster analysis of crude oil geochemical data for source characterization*. Marine and Petroleum Geology, 133, 105292.

[6] Crude Quality Inc. (n.d.). CrudeMonitor.ca. Retrieved October 19, 2025, from https://www.crudemonitor.ca/