Aaron: Gasoline Prices with Economic Indicators

Objective: Build a machine learning model to predict future gasoline prices for New York State or specific cities.

Steps:

1. Data Integration:
   * Use crude oil prices, Consumer Price Index (CPI), and local/state gasoline tax data as external features.
2. Feature Engineering:
   * Include time-related features like seasonality (month, week, holidays).
   * Compute rolling averages and volatility of crude oil prices.
3. Model Building:
   * Train regression models such as Random Forest, Gradient Boosting (e.g., XGBoost), or LSTM for time-series forecasting.
   * Evaluate using metrics like RMSE, MAE, and MAPE.
4. Insights:
   * Identify key drivers of gasoline prices using feature importance analysis (SHAP or permutation importance).

Jen: Forecasting Fuel Transition (EV vs. Gasoline)

Objective: Identify the key factors influencing EV adoption and predict future adoption rates.

Techniques:

* Classification Models:
  + Use models like logistic regression, decision trees, or neural networks to classify regions or individuals as EV adopters or non-adopters based on demographic, economic, and policy factors.
* Clustering:
  + Group regions by EV adoption patterns using clustering algorithms (e.g., K-Means) to identify areas with high or low adoption potential.

Data Sources:

* EV sales data (e.g., [DOE Alternative Fuels Data Center](https://afdc.energy.gov/)).
* Demographics (e.g., [US Census Bureau](https://www.census.gov/)).

Pat: Anomaly Detection

Anomaly detection focuses on identifying unusual price fluctuations or discrepancies across regions and time.

Objective

* Detect and analyze unexpected changes in fuel prices that might be driven by supply chain disruptions, weather events, or economic policies.

Steps

1. Data Preprocessing
   * Detrend and deseasonalize time series data using moving averages or seasonal decomposition.
   * Aggregate data weekly or monthly for consistency.
2. Feature Engineering
   * Calculate price differentials:
     + Between regions (e.g., Albany vs. Buffalo).
     + With spot prices (e.g., NY Conventional Gasoline Spot Price vs. Albany Average).
   * Derive additional features:
     + Rolling averages/volatility (e.g., 7-day or 30-day).
     + External data:
       - Weather Data: Look for extreme conditions or storm events (e.g., NOAA).
       - Traffic Data: Examine congestion spikes (e.g., INRIX).
3. Anomaly Detection Models
   * Statistical Methods:
     + Z-scores for price outliers.
     + Seasonal-trend decomposition to detect abrupt deviations from seasonal norms.
   * Machine Learning Models:
     + One-Class SVM: Identify anomalies by modeling normal price distributions.
     + Autoencoders: Train neural networks to reconstruct normal price patterns; deviations signal anomalies.

Some Other External Data Sources:  
  
Traffic Data

* Source: INRIX, Google Maps API.
* Use: Understand price surges related to holiday or commuter traffic. Cluster regions based on their sensitivity to traffic patterns.

Weather Data

* Source: NOAA Climate Data Online.
* Use: Analyze weather-related disruptions (e.g., hurricanes causing price spikes) and include these patterns in anomaly detection.

Commodity Futures

* Source: CME Group.
* Use: Incorporate futures market volatility into anomaly detection to flag instances of speculative price behavior