Sample Research Questions and Datasets for Battery Technology & Clustering

---

1. Clustering Battery Cells by State of Health (SOH)

- **Research Question**:
  How can we group lithium-ion battery cells based on their SOH degradation patterns, and what does this imply for battery pack uniformity?

- **Dataset Recommendation**:

  1. **NASA Prognostics Data Repository**
     [**Link**: NASA Battery Data]
- **Programming Exercise**:

  1. **Load and preprocess** the NASA battery dataset (handle missing values, scale features).
  2. **Extract key features** such as capacity fade, voltage dips, or internal resistance.
  3. **Apply clustering** (e.g., K-Means or DBSCAN) to group cells with similar SOH trajectories.
  4. **Visualize clusters** and discuss how consistent they are with battery cycle life expectations.
  5. **Evaluate cluster quality** using silhouette score or domain knowledge.

---

2. Identifying Fast vs. Slow Degradation Paths

- **Research Question**:
  Can we cluster battery cells into groups of **fast degrading** and **slow degrading** patterns based on capacity fade rate and voltage curves?

- **Dataset Recommendation**:

  1. **MIT Battery Degradation Dataset**
     [**Link**: MIT Battery Data]
- **Programming Exercise**:

1. **Import data** containing cycling profiles over time (voltage, current, capacity).
2. Compute **degradation rate** features, e.g., drop in capacity per cycle.
3. **Cluster** cells using hierarchical clustering (or K-Means) to find distinct degradation paths.
4. **Plot degradation curves** by cluster to see how quickly each group fades.
5. Relate clustering results to **usage conditions** (like temperature or charge rate).

---

3. RUL Prediction via Cluster-Based Approaches

- **Research Question**:
  How can clustering be used as a **preprocessing step** to support remaining useful life (RUL) estimation for different battery cells?

- **Dataset Recommendation**:

  1. **CALCE Battery Data**
     [**Link**: CALCE Battery Data]
- **Programming Exercise**:

  1. **Load** the CALCE dataset focusing on cycle life and capacity.
  2. **Cluster** cells based on their early-cycle behavior (voltage curves, capacity retention).
  3. For each cluster, **train a regression model** (e.g., linear or neural network) to predict RUL.
  4. Compare RUL predictions **with** and **without** clustering-based preprocessing.
  5. Present findings on whether clustering helps in **improving RUL accuracy**.

---

4. Temperature & Rate Clustering for Thermal Management

- **Research Question**:
  Can we identify battery clusters that respond differently to **temperature variations** and **charge/discharge rates**, and how can this inform thermal management strategies?

- **Dataset Recommendation**:

  1. **Harvard Dataverse A123 LiFePO4 Dataset**
     [**Link**: Harvard Battery Dataset]
  2. **A123's Lithium Iron Phosphate (ANR26650M1-B) Battery Cell Data** [Link]

- **Programming Exercise**:

    1. **Load** LiFePO4 data with different temperature and charge rate conditions.
    2. **Create features** such as [max temperature rise, average charge time, capacity fade under each condition].
    3. **Cluster** the battery cells to see which ones are most/least sensitive to temperature or rate changes.
    4. **Analyze** how cluster membership correlates with overall performance or safety margins.
    5. Make recommendations for **thermal management** or **charging protocols**.

---

5. Clustering-Based Fault Detection in EV Batteries

- **Research Question**:
  Can clustering algorithms detect **faulty or abnormal** EV battery cells before they fail?

- **Dataset Recommendation**:

    1. **Kaggle EV Battery Datasets**
       [**Link**: Kaggle] (search for EV-specific data)

- **Programming Exercise**:

    1. **Gather** EV battery data (voltage, current, temperature, SoC over time).
    2. **Feature engineering**: compute voltage drop under load, internal resistance trends, temperature spikes.
    3. **Use an outlier detection** approach or a cluster-based technique (e.g., DBSCAN) to **flag anomalies**.
    4. Investigate which points are **far from clusters** (potential early failures).
    5. **Evaluate** your model by how accurately it detects known faulty samples (if labeled data is partially available).

---

6. Battery Pack Configuration via Cell Clustering

- **Research Question**:
  How do we **select homogeneous cells** from a large pool to build a battery pack with **minimal performance variation**?

- **Dataset Recommendation**:

1. **Oxford Battery Degradation Dataset**
   [**Link**: Oxford Battery Dataset]

- **Programming Exercise**:

  1. **Load** multiple cells tested under the same conditions.
  2. **Cluster** cells based on their key performance metrics (voltage, capacity, temperature rise, SoH).
  3. **Select** one cluster representing the **most stable and uniform** cells.
  4. Demonstrate how forming a pack from that cluster can **reduce mismatch** and extend pack life.
  5. **Discuss** potential trade-offs (like cost vs. uniformity).

General Guidance for All Exercises

- **Preprocessing**: Clean the data, handle missing values, and normalize or standardize features.
- **Clustering Methods**: Experiment with **K-Means**, and **DBSCAN**, clustering to compare results.
- **Evaluation**:
  - **Internal metrics**: Silhouette score.
  - **Domain knowledge**: Validate whether clusters make sense in terms of battery engineering.
- **Visualization**:
  - 2D/3D plots (using **PCA** or t-SNE) to see how clusters separate.
  - Bar charts or box plots to compare cluster statistics (e.g., average capacity).

These **research questions** and **datasets** will give students hands-on experience in **data preprocessing, feature extraction, clustering, visualization, and interpretation** for **battery technology**.