**Subject:** Investigating Customer Churn – Initial Data Requirements and Approach

**Dear Associate Director,**

Estelle and I have reviewed PowerCo’s case, and we are outlining our initial plan to investigate the churn issue in line with our 5-step data science methodology. Below are the details of our problem framing, proposed data needs, and analytical approach.

**🧠 1. Business Understanding & Problem Framing**

**Problem Statement:**  
PowerCo is experiencing customer churn due to increased competition in the energy market. The goal is to determine the key factors influencing customer churn — particularly the hypothesis that price sensitivity is a major driver — and to develop actionable insights to help PowerCo reduce churn.

**📊 2. Exploratory Data Analysis & Data Cleaning**

**Key Data Required from PowerCo:**

* **Customer Profile Data:**
  + Customer ID
  + Business size (e.g. SME classification)
  + Industry type
  + Location
* **Usage & Billing Data:**
  + Monthly energy consumption
  + Billing history
  + Plan types and pricing history
* **Customer Interaction Data:**
  + Customer service contact history
  + Complaint logs
  + Survey results (if available)
* **Churn Label:**
  + Indicator if customer has churned or not
  + Churn date (if applicable)
* **External Data (optional but useful):**
  + Competitor pricing data (if available)
  + Energy market trends

**🛠️ 3. Feature Engineering**

We will derive features such as:

* Monthly cost fluctuations
* Tenure with PowerCo
* Frequency of support interactions
* Average response time from customer service
* Competitiveness of pricing compared to regional average
* Historical consumption patterns (seasonal trends, spikes, drops)

**🤖 4. Modeling and Evaluation**

We plan to use classification models (e.g., Logistic Regression, Random Forest, or XGBoost) to predict churn probability and identify feature importance.  
Evaluation metrics will include:

* Accuracy
* Precision/Recall
* ROC-AUC

**📈 5. Insights & Recommendations**

We will visualize insights using:

* **Churn distribution by price plans and regions**
* **Correlation heatmaps** for variable relationships
* **Bar plots** of churn rate by customer service score, tenure, and usage tier
* **SHAP or feature importance plots** to explain the model results

Based on this, we will recommend:

* Strategies for retention (e.g., re-pricing, loyalty offers)
* Target segments for intervention
* Operational improvements in customer support