Describe analytics models and data that could be used to make good recommendations to the power company.

Here are some questions to consider:

- **The bottom-line question is which shutoffs should be done each month, given the capacity constraints. One consideration is that some of the capacity – the workers' time – is taken up by travel, so maybe the shutoffs can be scheduled in a way that increases the number of them that can be done.**

- **Not every shutoff is equal. Some shutoffs shouldn't be done at all, because if the power is left on, those people are likely to pay the bill eventually. How can you identify which shutoffs should or shouldn't be done? And among the ones to shut off, how should they be prioritized?**

Think about the problem and your approach. Then talk about it with other learners, and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

This is how I would approach this problem:

PART 1: Finding which customers will most likely not pay their electricity bill.

Given: { address/zipcode, customer id, amount of electricity used per month, months that they have/ haven't paid bills on time, recent history of payment, income of household, amount paid/due, days since the last payment)

The goal here would be too classify customers into two categories which include customers that never pay the bill and customers that will pay the bill.

Use: {SVM (Support Vector Machines) } to fit the data. A Support Vector Machine (SVM) is well-suited for the power company's objective of identifying homes at risk of not paying electricity bills on time. SVMs are effective in handling non-linear relationships, making them suitable for scenarios with complex payment histories. The model excels in situations of imbalanced data, focusing on instances that are more challenging to classify. By aiming to maximize the margin between classes, SVMs can identify a clear boundary between homes with a consistent payment history and those that may default. Additionally, SVMs are robust to outliers and can handle higher-dimensional data, making them adaptable to various features influencing the decision to shut off power. Customizable kernel functions, such as radial basis function (RBF) or polynomial kernels, enable the model to capture nuanced relationships in the data. Overall, SVMs offer a powerful tool for binary classification tasks in this context, effectively prioritizing homes with a higher likelihood of payment issues.

After customers have been identified for the shutoffs, we will be able to prioritize them based on how many days it's been since they last paid their bill, and/or the amount they need too pay

Result : {Now we have a generalized picture of houses that haven't paid their bills, and we have ranked them on both the amount of money they need to pay and the length of time that has passed since they have paid their last bill. }

PART 2: Determining order of how shutoffs need to be completed.

Given: { customer id of the individuals who we determined are not going to pay the bill from the SVM model, amount due, location of the customer(zip code or address)

Use: { clustering algorithm(K-means) in addition to a Simulation Model} . By integrating the SVM model's predictions with customer data, including customer ID, amount due, and location information, clustering algorithms like K-Means or hierarchical clustering can be applied to group customers based on shared characteristics. This process enables the identification of clusters with a higher likelihood of non-payment. Analyzing these clusters provides valuable insights, such as specific neighborhoods or customer groups consistently appearing in clusters with a higher risk of bill non-payment. By prioritizing efforts based on these clusters, businesses can strategically allocate resources and focus interventions on areas or groups with a greater potential impact. Continuous monitoring ensures that the prioritization remains dynamic, allowing for adaptive responses to changing customer behavior and payment patterns.

After identifying clusters of customers at risk of non-payment through clustering, we can use a simulation model employed to optimize shutoff scheduling and minimize workers' travel time. The simulation dynamically adjusts routes and worker allocations based on real-time data. Scenario analyses within the model allow for the exploration of various scheduling strategies, optimizing the order and allocation of shutoffs. By defining and monitoring key performance metrics, businesses can continuously improve scheduling strategies, ensuring maximum impact on shutoffs with minimal impact on workers' time spent traveling.

Result: {By using a combination of K-means and simulation models we can hopefully determine the shortest route that workers can take so they can minimize the amount of time they spend on the road.}