1. **Write-Up on ML Models. (Point: 2)**

Linear regression is among the simplest machine learning models, plotting a line to predict where similar data points would appear in a dataset. However, the model also expects data to be linear. Temperature data, however, will fluctuate as the seasons change, creating a nonlinear wave over years. This model is unlikely to be the correct choice.

Random forest instead uses a collection of decision trees, each trained on random subsets of the data. These trees vote on each data point and classify it with a majority, intending to create a set of classifiers that can eliminate errors through quantity. The randomness of the subsets are intended to vary their errors, which could make a good fit for this project.

Gradient boosting also uses a collection of models, but for a different purpose: Each model is sequentially trained, using weight updates based on the errors of previously misclassified instances. This causes each model to “fix” the last one’s error, continuing until all points are correctly classified (or a maximum epoch count is reached). Outliers could greatly mess with this algorithm because of this, so it may not be the best choice due to unusually hot or cold days.

With the above information in mind, a random forest classifier will likely be the best choice.

**Retail Question**: **Customer Lifetime Value (CLV)**

* + - How can we predict long-term revenue potential to prioritize high-value customers?

1. Web Server Setup. (Point: 2)