10 Questions an audience would ask you

1. Q: With the results of the RMSE error for the XGBoost model, wouldn't this mean that the model performed poorly?

   A: No, for this method of evaluating accuracy you need to consider the range of values. We're dealing with large range of values, so a higher RMSE is acceptable. This is why we used the Mean Absolute Percentage as well. This showed that model is still performing well with a higher RMSE.

2. Q: For the forecast of the XGBoost model with solar power, the forecast said the usage will increase but the actual data decreases. What could cause this difference?

   A: The factors that could be causing the decrease could be due to reporting from the City of Mesa website. For the specific month in question, it's possible that there are bills not recorded on the website by the time the data was pulled.

3. Q: Was dropping the rows with missing values the best way to handle them?

   A: Yes, since the data that was missing was the kWh usage it was better not to make up values for what was being used. This was prior to the mean being taken for the month, so it should have minimal impact on the data.

4. Q: Can the models account for changes in patterns of the data?

   A: The model is currently accounting for changes in patterns. However, this is not taking anomalies into consideration. For the model to account for an anomaly, it needs to be a reoccurring pattern. As for the increase in energy usage with solar power over time, the model does appear to be mapping this.

5. Q: Wouldn't the Profit model create better forecasts than the XGBoost model?

   A: The accuracy for the Profit model isn't outperforming the XGBoost model. The Profit model is not a better model. Also, with the way the Profit model is built, I do not feel it provides full transparency on how it makes its predictions.

6. Q: What could cause the cyclical behavior at the start of the time series plots?

   A: Many factors can contribute to this behavior including weather, population changes, and new businesses coming to the area. Arizona does receive severe weather that can cause

power outages which could cause the drop offs in the data. However, this could be as simple as the data was never fully reported to the website.

7. Q: Since the data had potential personal identifying information in it, was it pulled from a trustworthy cite?

   A: The Website itself is a trustworthy source. While I did see account numbers for residential buildings, these buildings could be used for business. If all the data was for business locations, then the data would have been already accessible to the public and not PII.

8. Q: What are the main features being used in the model currently to forecast energy usage?

   A: While we did keep a few columns from the data the only feature that was used was the "kWh" used. The "Month Date" column was converted into an index for the models x-values.

9. Q: Are there any models that you have not tried that may provide more accurate results and that are more transparent than the Facebook Profit models?

   A: There are other models that weren't attempted that could provide better results than both models. However, once SARIMA failed to capture the seasonality, I felt it better to use the models that could capture this aspect. Some models were overpredicting energy usage and XGBoost and Profit were two that provided results close to the actual usage.

10. Q: How would these models help stake holders decide on opening a new business in this area?

    A: For the stake holders to consider a location for the business, we need to show a growing demand for our product. We can see that non-solar energy usages are declining, and solar energy usage is increasing. This will show the demand for solar is growing, which means that the area and surrounding locations are potentially profitable.