Solar Energy Analysis

Scott Webster, Heng Zhang, W. P. Carey School of Business

January 2023

1. Assignment Overview

Your team will develop and analyze a simulation model of solar power generation and electricity costs. Recall that one way to view analytics is in terms of three broad categories: *preparation*, *description*/*prediction*, and *prescription*. This case involves all three categories, though the data preparation element is relatively minor.

The goal is to reinforce your skills in (1) developing predictive models from data, (2) designing and implementing a relatively complex simulation model, (3) conducting analysis of your model to identify prescriptions (i.e., guidance for a decision maker). The case assignment can be divided into three main steps.

1. *Predictive models* of kWh demand and production.

    a. *Demand model*. In general, the demand model (to be estimated from the data) specifies the probability distribution of kWh demand for each hour in a year. But it is not reasonable to define a different probability distribution for each of the 8760 hours in a year. Instead, your team will conduct two-level hierarchical clustering: clustering days within a year, then clustering hours within a day. Your predictive demand model will specify a probability distribution of demand for each cluster.[1]

    In order keep time requirements of this case assignment reasonable, the top-level clusters are specified for you: each cluster is a month and each month has a demand factor to be estimated/predicted. The demand factor for a month (aka, seasonal index) is the predicted average 24-hour demand in a month as a percent of the "average" month. For example, a month with a demand factor of 1.2 means that daily demand in the month averages 20% more than demand in the "average" month. To be correct, the average of the 12 monthly demand factors should equal 1.0.

    The determination of the low-level clusters is more open ended. However, before this clustering takes place, the hourly demand data will need to be deseasonalized to strip out the top-level cluster effects (more detail on this in Section 3). Once the data are deseasonalized to reflect kWh demands in an "average" month, your team will identify <u>up to three clusters</u> within a 24-hour weekend/holiday period and up to three clusters within a 24-hour weekday period.[2] Each cluster has its own probability distribution of demand in an hour. In other words, the probability distribution of kWh demand within a cluster is stable and distinct from other clusters (e.g., time interval of low demand, time interval of moderate demand, time interval of high demand). For each cluster, your team will identify characteristics of kWh demand (e.g., mean, standard deviation, probability distribution type). Unless you have very good justification to deviate, only the probability distribution parameters will differ across clusters; the probability distribution type will be the same across all clusters. Furthermore, use a theoretical probability distribution of kWh

---

[1] By identifying probability distributions of kWh usage (e.g., as opposed to using historical actual usages), you will be able to (relatively easily) investigate the effects of changing usage patterns (e.g., by changing parameters in your probability distribution of usage).

[2] This does not mean that no more than 3 clusters is best. The reason for the constraint on the number of clusters is to help limit the time that you spend on clustering.

demand rather than a histogram constructed from data. As noted above, the main reason is that this allows you to easily study the effects of changes in character of random demand (e.g., changes in the means or standard deviations of demand).

b. *Production model*. In general, the production model (to be estimated from the data) specifies the probability distribution of solar panel kWh production for each hour in a year. kWh production is a straight-forward function of number of solar panels, capacity per panel under full sunlight, and fraction of full sunlight in an hour. The capacity is provided by the solar panel vendor and the number of panels is a value that you may specify and experiment with. The only predictive element of the production model pertains to fraction of full sunlight in an hour. For example, suppose the solar panel capacity is 0.2 kWh. There is no solar production during nighttime hours, an hour of full sun generates 0.2 kWh, and to keep things simple, you may assume that the panel generates 0.08 kWh under 60% cloud cover (i.e., $0.40 \times 0.20 = 0.08$) (and no power is generated under 100% cloud cover).

   Similar to the demand model, you will identify the probability distribution of fraction of full sunlight in an hour for different clusters, though in this case, the clusters are fully specified so that you may spend time on other aspects of the project (12 clusters corresponding to months). Data on daylight hours and Phoenix area cloud cover over a year are available. Additional detail on production model clusters are provided in Section 3 below.

2. *Measure of performance*. Your team will develop a cost model. This model that maps kWh demand, number of solar panels, kWh solar production, and electricity rate structure into cost. You will use this model to estimate the probability distribution of annual electricity cost under different scenarios (e.g., number of solar panels, rate structure), which serves as the foundation for prescriptive analysis described below.

   The electricity rate structure includes both charges and credits. A credit accrues when kWh production is greater than demand (i.e., known as net metering). A charge occurs when kWh demand is greater than production, which is offset by any accumulated credit. Your model is designed to accommodate a multi-rate structure that arises in practice. In particular, your model allows up to five charge rates (i.e., on-peak-winter, on-peak summer, off-peak winter, off-peak summer, super off-peak winter). By designing your model to include parameters that define when each rate category applies, you will have the flexibility cover fewer than five rate categories as well as different time intervals when a particular rate category applies. (APS is known to periodically change the rate structures, and you can easily examine the effect of a change.)

3. *Prescriptive analysis* of your model. One technical challenge with analysis is that solar panels and associated equipment have long (and uncertain) lives, extending 25 years or more. This leads to scalability problems, especially for a class assignment, e.g., 25 years corresponds to about 220,000 hours leading to a model requiring about 200 MB of space that is likely impractical for simulating 25-year behavior over many trials. In addition, the size and complexity of a 220,000-hour (or more) model make it difficult to understand the behavior of the system for the purposes of decision making.

   For the above reasons, you will limit your model to a one-year time frame. The overall *goal of your analysis* is to gain an understanding of the relationship between various inputs (e.g., number of solar panels, how the credit rate compares to the charge rates, how the clusters of low, mid, and high demand compare to the time intervals that different charge rates are in effect) and a key output:

probability distribution of the *percentage reduction in annual electricity bill from solar panels*, and especially its associated summary statistics – mean and standard deviation.[3]

What are elements under control of a decision maker considering an investment in solar power? There are at least three, and possibly more.

- How does the value of adding a solar panel (i.e., reduction in the electricity bill) depend the number of solar panels and other inputs?
  - For example, do you find a nonlinear relationship between the number of solar panels and the average percentage reduction in the electricity bill? If the relationship is sometimes linear, sometimes convex (i.e., increasing at an increasing rate), and sometimes concave (i.e., increasing at a decreasing rate), can you identify conditions under which a particular form is likely to be the case and intuition that may explain such behavior? How does the standard deviation in the percentage reduction in electricity bill change with the number of solar panels, and what may be the reason for observed behavior? How might answers to these questions be relevant to a decision maker trying to decide how many, if any, solar panels to install?[4]
- How does the value of changing the demand profile depend on the number of solar panels and other inputs?
  - For example, the electricity bill will clearly decrease if investments and actions are taken to shift some kWh demand from periods when rates are high to periods when rates are low. However, is the percentage reduction in the electricity bill from solar relatively stable, or are there conditions under which shifting demand to lower rate intervals are likely to either increase or decrease the percent savings? In other words, are there cases of a synergistic relationship between investment in solar and demand shifting, and if so, what is the intuition that may explain such phenomena?
- Which rate schedule should be selected?
  - A decision maker may also select among multiple electricity rate programs. For this case assignment, you will consider two possible rate programs. Your model and analysis will provide some insight into whether there is a meaningful difference in electricity cost between the two programs, and how the choice of program may influence the value of solar panels.

[3] Why is this goal meaningful? An understanding of this relationship allows a decision maker to gain a sense of hw the reduction in the electricity bill will change over the years as rate, demand, and production parameters change (e.g., charge rates change annually by uncertain amounts, the credit rate is subject to change each year, demand may grow or shrink over the years, and production efficiency of a solar panel decays over time). Furthermore, an accurate understanding can be used to identify a simple predictive model that generates the random reduction in the electricity bill by year, which in turn can be implemented in a multi-year simulation model, thereby getting around the issues of scalability discussed above. The development of such a predictive model is not part of this case assignment, but is mentioned here to highlight another reason why an understanding of the relationship between inputs and percent reduction in electric bill in a given year adds value.

[4] Some hints or suggestions related to this question: It is reasonable to assume that the cost of an additional solar panel is a constant, and it is reasonable to assume that the amortized cost of a solar panel (e.g., average cost in a year of its useful life) is constant. It is also reasonable to assume that a decision maker prefers less uncertainty to more uncertainty (though different decision makers may have different degrees of risk aversion).

You may identify factors related to the installation of solar panels (you may use the factors above, or you may argue that some other factors are relevant). Then focus your analysis on understanding the relationships between inputs and outputs that are relevant for the factors you identified. Make sure to clearly show how you arrived at your conclusions (e.g., providing relevant plots, etc.) and summarize how your findings may be used by a decision maker.

2. Data files

1. *Data.xlsx*. Contains data on historical kWh demand by hour (end of Sept 2015 to end of Oct 2022, over 62,200 observations though with some missing data), sunrise and sunset by day and month, cloud-cover history. There may be other public sources for historical cloud-cover data for the Phoenix area. With the exception of demand data, you are welcome to supplement data in the file with other sources (though it is not necessary to do so).

2. Rate schedules: *RateSchedule.pdf* and *RateScheduleDemandCharge.pdf*. Each customer is allowed to select the rate schedule that will be used to determine the customer's monthly charges. You analysis should clarify whether one schedule offers any meaningful cost advantage over the other. These two APS electricity rate sheets contain five rate categories that apply at different times (listed under 'Bundled Charges'): on-peak-winter, on-peak summer, off-peak winter, off-peak summer, super off-peak winter. These rates for kWh usage appear at the top of page 2. For example, as shown in the tables below (excerpts from the rate schedules), *RateSchedule.pdf* lists the on-peak-winter energy charge rate at \$0.28185/kWh whereas the *RateScheduleDemandCharge.pdf* lists the on-peak-winter energy charge rate at \$0.08711/kWh. Under the *RateScheduleDemandCharge.pdf* tariff there is also a demand charge, e.g., \$11.845/kWh for on-peak winter and \$16.875/kWh for on-peak summer. This rate is charged against the single highest on-peak hour in a month. As an example, suppose the highest kWh in an on-peak time window (4pm to 7pm) in the month of June is 14.5 kWh. Since June is a summer month, the demand charge in June is 16.875 × 14.5 = \$244.6875. This is in addition to all the energy charges in the month of June.

	Summer	Winter	
On-Peak Energy Charge	\$ 0.29780	\$ 0.28185	per kWh
Off-Peak Energy Charge	\$ 0.10789	\$ 0.10790	per kWh
Super Off-Peak Energy Charge		\$ 0.03166	per kWh

	Summer	Winter	
On-Peak Demand Charge:	\$ 16.875	\$ 11.845	per kW

	Summer	Winter	
On-Peak Demand Charge:	\$ 16.875	\$ 11.845	per kW
On-Peak Energy Charge:	\$ 0.12414	\$ 0.08711	per kWh
Off-Peak Energy Charge:	\$ 0.05276	\$ 0.05267	per kWh
Super Off-Peak Energy Charge:		\$ 0.03166	per kWh

By designing your model to include parameters that define the hours that each rate category applies, you will have the flexibility cover fewer than five rate categories as well as different times when a category is active. In addition, you may set up your model to identify the highest on-peak kWh usage in a month that is multiplied by the on-peak demand charge (which would be set to zero if using *RateSchedule* tariff, i.e., this tariff has no on-peak charge). Note that there is also a basic service charge ($0.40) on the rate sheets (the first table on page 2); you may ignore this charge in your analysis because this charge is not affected by solar panel installation (i.e., assume connection to the grid remains, so that the daily service charge remains). You will also see the unbundled components of various charges that you may also ignore (i.e., no need to split out the bundled rates).

3. *RCPSolarRider.pdf*. An APS electricity rate sheet that contains the net metering rate (the rate at which excess power from solar panels is credited for use against future charges). For example, the latest credit rate as shown on page 2 is $0.08465 per kWh (though this credit rate may be reduced each year by up to 10%). The historical rates by tranche provide a way to estimate how the credit rate may change in the future.

Your model will include a solar panel production capacity parameter, i.e., kWh generated by a panel under full sun. Different panels and suppliers have different production capacities. The value of this parameter does not have a meaningful effect on your analysis – it only affects the step-size of capacity as you increase the number of panels. For consistency, it makes sense for all teams to use the same parameter value; please use 0.2 kWh as the production capacity of a single solar panel.

3. Model considerations and suggestions

3.1. Predictive model of demand

The objective is to "predict" or estimate parameters and probability distributions for generating random demand. You have about five years of hourly demand data (Data.xlsx). Keep in mind that an outcome is a model of future (and uncertain/random) demand, so you may want to be on the lookout for apparent shifts in the demand profile (e.g., if you find what appear to be fundamental changes over time, you may wish to emphasize more recent, and presumably more representative data, in your analysis).[5] The parameters that need to be estimated for your demand model are described below in §3.3.

You may wish to begin with some simple plots of demand over time. This may help to assess whether there have been shifts in the demand profile, and to identify data to exclude if there is evidence of a shift.

Top-level clusters

Once you have selected the "representative portion of the demand data that you want to use for analysis, the next step is to identify monthly demand factors (discussed in more detail in §3.3). For example, you might compute average daily demand in a month, say d_i for month i for the time frame you identify as "representative" of the future. The following ratio

$$12d_i/(d_1 + \ldots + d_{12})$$

is an estimate of the demand factor for month i. You might modify this estimate if you have good reasons to do so. Regardless of how you decide to estimate the monthly demand factors, make sure to clearly explain your method and your rationale.

[5] Do not incorporate a trend factor in your model (e.g., demand increasing or decreasing over the year). In this setting, it is highly unlikely that any trend would be sustained over a long term; attempting to include trend adds to complexity and likely will reduce accuracy over the long life of a solar panel system.

You are working with a model based on two-level hierarchical clustering. I've specified the top-level clusters for you; each month is a cluster that captures the "temperature effect" of demand. For example, the probability distribution of the effect of temperature on demand in an hour is assumed to be the same for each hour in the month, but can be different from hours in other months. While the data may suggest different clusters than months, use monthly clusters in your analysis. You will be identifying clusters at the next level (described later).

Use your monthly demand factors to deseasonalize your historical demand data before the next step of analysis, i.e., divide the demand in each hour of a month by its demand factor.[6]

Low-level clusters – clusters of days within a week

The day clusters are specified in this assignment: 24-hour periods will be divided into two clusters: (1) weekday, (2) weekend/holiday. These daily clusters make sense because they align with the electricity rate structures (e.g., APS recognizes that patterns of total kWh demand are distinctly different between weekdays and weekends/holidays).

This restriction on how you will cluster days introduces an additional data preparation step. You will need write code to identify whether a particular date is a weekday or a weekend/holiday (e.g., date functions in Excel can be used, along with the list of holidays provided in 'SaverChoiceRates.pdf' that can be placed in an Excel table). With your logic that identifies whether a day belongs to a weekday cluster or a weekend/holiday cluster, you may then separate your deseasonalized demand data into two datasets: one with deseasonalized weekday demand by hour, and one with deseasonalized weekend/holiday demand by hour.

Low-level clusters – identifying clusters of hours within a day (e.g., unsupervised learning)

This is the part of the assignment where your team will identify clusters (i.e., clustering hours within a 24-hour period). Some simple plots of data in each cluster (weekday, weekend/holiday) may be useful. Your team will identify up to three clusters for each dataset (i.e., deseasonalized weekday dataset and deseasonalized weekend/holiday dataset). To restate, you will identify no more than three clusters of hours within a 24-hour period for your weekday dataset and your weekend/holiday dataset (e.g., no more than six clusters in total).

Recall that a cluster is a time interval in a 24-hour period during which the probability distribution of demand is stable and distinct from other clusters. It is not necessary to conduct sophisticated cluster analysis for this step (though you are certainly welcome to apply methods you have learned).[7] For example, you may find that cluster boundaries are relatively clear from simply examining your plots. However, each cluster you identify should be a group of consecutive hours. For example, hours 2:00-3:00 and 15:00-16:00 should not be in the same cluster if hours 0:00-1:00 and 3:00-4:00 are in different clusters from 2:00-3:00. Forming clusters that violate the "consecutive hours" rule are susceptible to the "over fitting" problem because hourly demand in reality has a relatively high degree of autocorrelation. Regardless of what clustering method you use, clearly explain your approach and rationale.

[6] Your demand generation model involves two steps: generate demand in an hour for the "average" month, then introduce the monthly pattern effect by multiplying by the demand factor. This 2-step sequence means that you have to follow the reverse sequence when estimating model parameters, i.e., estimate seasonality first (demand factors) and use to remove seasonality from the data, then identify clusters with associated parameter estimates.

[7] Avoid falling into a trap of spending many hours on cluster analysis. Apply some quick and simple methods, then move on to other aspects of the assignment.

After you have identified hourly clusters (e.g., up to three clusters each for weekend/holiday dates and weekday dates), the next step is to estimate the parameters for each cluster, e.g., mean, standard deviation, and probability distribution type. If you decide to use a uniform distribution for demand in an hour, then mean and standard deviation parameters can be replaced by min and max parameters. As noted above, use a theoretical probability distribution of kWh demand rather than a histogram constructed from data.

3.2. Predictive model of production

kWh production in an hour depends on the number of solar panels, the capacity per solar panel (e.g., 0.2 kWh), and the amount of sunlight in an hour. You may assume that all randomness in production stems from the probability distribution of cloud cover. Parameters related to the production model are discussed in more detail below.

NOTE ON PYTHON USAGE

Your team is welcome to use Python for the analysis related to Sections 3.1 and 3.2, though it is not required. Python is a natural tool for visualization and clustering, and usage can be useful to reinforce your Python skills; for example,

- describing and summarizing the data
- identifying monthly demand factors
- deseasonalize the data and divide into two data sets: (1) deseasonalized weekday data, (2) deseasonalized weekend/holiday data
- cluster analysis (distinct intervals within the 24-hour cycle) for each dataset – this can be done using visualization (e.g., identify what seem to be reasonable clusters by looking at various plots), can be done using a clustering algorithm, or a combination of both[8]
- estimation of relevant prediction parameters for each cluster – can be done using simple methods such as tables or plots of summary statistics on usage over time (e.g., mean, standard deviation, histograms or plots of the distribution of usage within a cluster, etc.).

Relevant analysis and results, whether via Python or Excel, are summarized in the document(s) that you upload prior to the assignment due data.

3.3. Model of electricity costs (with and without solar) over one year

Here a list of inputs that your model should include.

- Parameters for electricity rate data
  - Five kWh energy charge rates (on-peak-winter, on-peak summer, off-peak winter, off-peak summer, super off-peak winter) and two on-peak demand charge rates (summer, winter).
  - For on-peak, off-peak, and super off-peak: the first and last hour in a 24-hour cycle that the rate category applies.
  - For each month, an indicator identifying whether the month is a *winter* month or a *summer* month (i.e., avoid hard-coding the logic that defines what is a winter month or a summer month; in other words, include an indicator for whether a month is winter or summer with logic that refers to this

[8] To reinforce an earlier point, <u>avoid falling into a trap of spending many hours on cluster analysis</u>. Apply some quick and simple methods, then move on to other aspects of the assignment.

indicator when applying rates). This allows your model to be easily updated when APS reclassifies summer and winter months in their rate structure (due to climate change).

  - The kWh credit rate (e.g., latest is 0.08465/kWh according to the APS rate sheet 'RCPSolarRider.pdf').

  - You may assume that the rates do not change during the year (i.e., no need to add parameters for amount and timing of rate increases).

  - Table of holidays: A holiday has a different rate than a weekday. If a holiday falls on a weekend, then the holiday rate goes into effect on the closest weekday.[9]

- Parameters for demand data

  - Monthly pattern of daily demand captured by 12 demand factors, one for each month. The factor is the ratio of average daily demand in a month to the overall average daily demand. These factors are meant to capture the fact that daily demand in August tends to be different than daily demand in February (e.g., due to temperature differences)

  - Hourly pattern of demand during a weekday. As explained earlier, limit the pattern to no more than three intervals (or clusters) during which the probability distribution of demand in an hour is stable (e.g., same men and standard deviation). Each time interval has three types of parameters: (1) first hour of the time interval, (2) last hour of the time interval, (3) parameters for the probability distribution (e.g., depending on the distribution type you select, these parameters could be mean and standard deviation, or minimum and maximum, etc.).

  - Hourly pattern of demand on a weekend/holiday. The points above for weekday demand parameters apply here as well.

- Parameters for production data

  - Daylight hours. Data on the times of sunrise and sunset by day for the Phoenix are available (Data.xlsx). These times do not change much from year to year. There are a total of 24 parameters, 2 for each month.

  - Percent sunlight during a daylight hour. Data on cloud cover in the Phoenix area are available (Data.xlsx). You are welcome to use a different data source (e.g., if you find one that is more convenient or accurate for your analysis). You will likely have 25 parameters: mean and standard deviation of the percent sunlight during a daytime hour (alternatively the min and max percent sunlight during a daytime hour if you use a uniform distribution). This makes 24 parameters. You should include an additional parameter that gives percent of sunlight available during the hour of sunrise and sunset compared to an hour in the middle of the day. This captures, in a very simple way, the effect of the angle of the sun on production. In reality, even with no cloud cover, production varies in a more-or-less continuous manner as the angle of the sun changes during the day. For the purposes of this case, it is fine to approximate this effect with a single parameter for fraction of energy lost in the first and last hour of the day (compared to the rest of the daylight

[9] For example, if a holiday falls on a Saturday, then the holiday rate (same as the weekend rate) applies on the Friday before; if a holiday falls on a Sunday, then the holiday rate applies on the Monday after. Since one weekday is no different from another weekday in your model, it is fine to simplify your logic by always assigning the holiday rate to Monday, regardless of whether it falls on a Saturday or Sunday.

hours that are all treated the same). In other words, with the exception of the first and last hour of sunlight in a day, the effects of variations in sunlight angle are ignored.

  - Hourly production capacity of a solar panel under full sun (as noted above, you may use 0.2 kWh for this parameter).

- On selecting probability distribution types: As noted above, your model will include a probability distribution for demand in an hour and a probability distribution for production in an hour (you may assume random demand and production are independent). It is not necessary to conduct time-consuming analysis on this, e.g., select a simple and/or common distribution that seems reasonably representative, keeping in mind that you will have to use Excel functions (not @Risk functions, for reasons explained below) to generate your random variable values. The chi-square goodness of fit test or nonparametric alternatives *are not needed*. For example, you may examine a frequency plot of deseasonalized demand (either in aggregate or by cluster) to help guide your choice of probability distribution. Random production is tied to random cloud cover. Historical cloud-cover data and opportunities for frequency plots related to production are more limited. In sum, use a simple and quick approach to select your probability distribution types and clearly explain your approach or reasoning. Finally, <u>do not use single parameter distributions such as exponential for demand or production</u>; your model should include the flexibility to examine the effects of reductions or increases in uncertainty without changes in the mean.

- Decision variable: number of solar panels (you may experiment with different values)

A few words on the setup of your model.

- You will need a column of electricity cost in each hour of the year assuming no solar panels (i.e., to compute the percentage reduction in annual electricity cost with solar panels).

- You will need a column that tracks the accumulated credit that will be applied to cover the electricity charge in an hour when demand is more than production.[10] Here is an example to illustrate the mechanics. Suppose the accumulated credit at the end of hour 5 is \$0.50, and the credit rate is \$0.10/kWh. In hour 6, production is 20kWh and demand is 15kWh, so the accumulated credit increases by 5 kWh × \$0.1 = \$0.50 to \$1.00. In hour 7, production is 15kWh and demand is 20kWh. Suppose that the cost rate in effect during hour 7 is \$0.25/kWh. The cost for the hour is the net kWh charged at \$0.25, or 5 kWh × \$0.25 = \$1.25. The accumulated credit is applied to the charge; the cost of hour 7 is \$1.25 – \$1.00 = \$0.25, and the accumulated credit at the end of hour 7 is \$0.

- Recall that the student version of @Risk is limited to no more than 100 @Risk random variables. The model you will develop will have 17,520 random variables (if the year your model is not a leap year; otherwise 17,544 random variables): 2 variables (demand and supply) × 24 hours × 365 days. This means that all of your random variables should be Excel functions, i.e., functions that use RAND().

- @Risk also has limitations on the size of your model (i.e., fit within 300 rows). For this reason, you will likely need to use Excel data tables to simulate performance over multiple trials. If you use a data table to simulate over multiple trials, it is fine to limit the number of trials to 100 (due to the large size of your model).

[10] kWh surpluses and shortages are actually tracked instantaneously over time, e.g., a charge is assessed if demand is more than production at any time instant, and a credit is accrued if production is more than demand at any time instant. Your model that uses one hour as a time period is an approximation of the real system.

- Your model will generate random electricity cost for the year with no solar panels, and the random electricity cost for the year with the number of solar panels specified in your inputs. From these two costs, you can compute the random percentage reduction in electricity cost.[11] You may simulate annual performance over many trials (each trial is one year) to illuminate how the mean and standard deviation of percent reduction in annual electricity cost is affected by the number of solar panels and other inputs.

- The electricity charge rate that applies on a day depends on whether the day belongs to one of two categories: non-holiday weekday or weekend/holiday. Your model should contain logic that identifies the category for a particular day and applies the appropriate rate.[12]

- Make sure that your model is flexible enough to properly compute outputs when all uncertainty is removed (e.g., your model works when all standard deviation parameters are set to zero, or in the case of a uniform distribution, the min and max parameters are equal).

4. Deliverables

There are four main elements of the deliverable:

1. Support for your choice of parameters and distribution types for your demand model, including your identified hourly clusters. We will evaluate this element according to the completeness of your analysis and the clarity in which your modeling choices and your analysis are explained and justified.

2. Support for your choice of parameters and distribution types for your production model. We will evaluate this element according to the completeness of your analysis and the clarity in which your modeling choices and your analysis are explained and justified.

3. Implementation of your cost model in Excel. We will evaluate this element according to accuracy / correctness of your model and the degree to which your model conforms to the spreadsheet engineering guidelines listed below. One point to emphasize (also noted below under spreadsheet engineering guidelines) – avoid long formulas . . . break up a long formula into multiple smaller formulas (e.g., model elements). This helps to improve transparency of logic, makes a model easier to debug, and reduces the chance of model errors.

4. Analysis of your cost model. We will evaluate this element according to the extent to which your analysis leads to usable (and nonobvious, therefore helpful) insights for a decision maker, and the extent to which your analysis and conclusions are clearly supported and presented. In general, insights relate how the number of solar panels and the choice of rate schedule (*RateSchedule.pdf*, *RateScheduleDemandCharge.pdf*) affect energy costs (or savings). It is not necessary to develop many insights – you may strive to expose two or three phenomena on the relationships between inputs and outputs that may have value to a decision maker.

Your entire deliverable can potentially be contained in a single Excel file. For example, your workbook may be comprised of multiple worksheets, such as "Executive Summary", "Demand Model" (e.g., summarizes analysis supporting choices of demand model parameters and distribution type), "Production Model" (e.g., summarizes analysis supporting choices of production model parameters and distribution

[11] For modeling purposes, it is reasonable to subtract any accumulated credit at the end of the year from the total electricity cost.
[12] It is not necessary to include logic that allows for more than these two categories.

type), “Cost Model”, “Lessons from Analysis” (e.g., you may wish to include some tables or charts in support of your discoveries). If you used Python for some of your analysis (e.g., plotting, clustering, etc.), you could include a tab containing your code and plots/tables (or two tabs, one for code, one for plots/tables). If you prefer, you may submit your Excel file along with other files. For example, you may choose to submit a pdf document that contains your explanations with any supporting tables or plots.

Use the following naming convention for your Excel file (and Word or pdf or Python file if applicable): Team#x where # = last 3 digits of your team number and x = cohort identifier (class days and start time). For example, if you are team # 23193 and attend a class that begins at 10:45 on Tuesday/Thursday, then your Excel file name is “Team193TTh1045.xlsx” (if you also upload a pdf, the name is the same except for the extension, “Team193TTh1045.pdf”); all teams start with “23”, so this is dropped from your team # in the file name. If your class meets Monday and Wednesday, you would replace “TTh” with “MW”.

Do **not** lock cells or protect your worksheets, or hide rows or columns.

Your file (or files) must be uploaded to Canvas prior to 7:00 am Wednesday March 1. Only one of your team members should upload the file for your team. The penalty for turning the assignment in late is 10% if within 24 hours, 20% if within 48 hours, etc.

5. Spreadsheet engineering guidelines

Your implementation of the model should abide by spreadsheet engineering guidelines that promote transparency and ease of use (see ‘Excel format example.xlsx’ on Canvas for an example):

- separate parameters, variables, outcome measures, detailed calculations (i.e., organize into modules)
- parameters are grouped together and away from calculations, and formulas should only contain cell references, not numerical values
- easy for a user to change parameters, and easy to find key outputs
- formulas are short, decomposing complex calculations into intermediate steps
- design for communication and ease of understanding, visual cues that reinforce model’s logic
- important data and formulas are well documented, e.g.,
    - list source for important parameters
    - explain important and complex formulas
    - use cell comments or notes where appropriate.

6 Grading

We will consider the following when grading:

- Is the support for your demand model clear, complete, and free of errors?
- Is the support for your production model clear, complete, and free of errors?
- Is your Excel implementation clear and easy to follow, free of errors, and are the principles of spreadsheet engineering followed (see above)?
- Is your analysis clear and coherent, and are your conclusions clearly supported and presented and offer value to a decision maker?
- Is your executive summary clear, well-written, and complete?

There is clearly no single “correct answer” for this case assignment. A submission that we judge as meeting a minimum standard for answering ‘yes’ to above questions will receive a 92. We will make adjustments upward for elements that stand out in a positive way and downward in the case of problems.