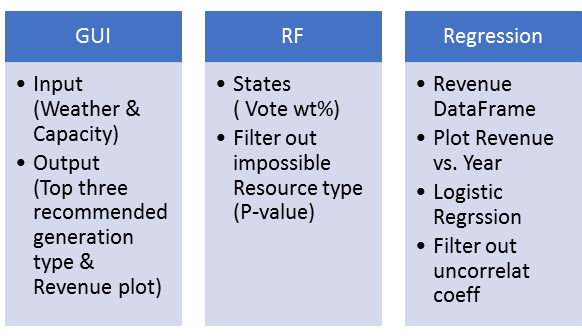
EASE Workflow

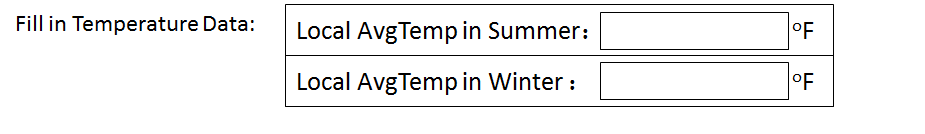
**Target Users:**

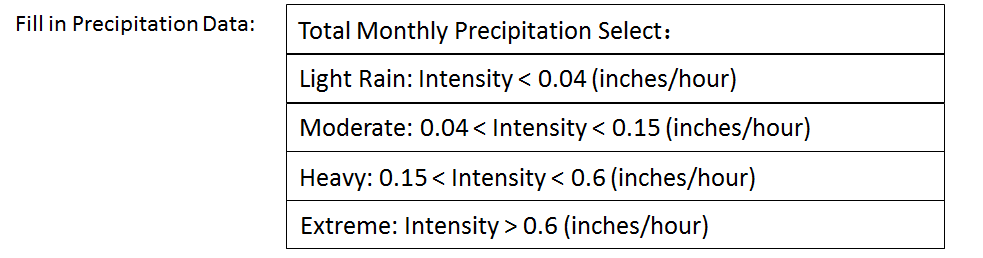
Industry users who want to build a self-sustained electricity power plant for their own factory to save money



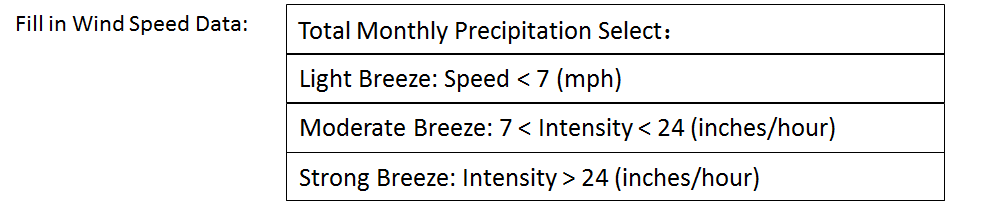
**GUI:**

**Input weather data**



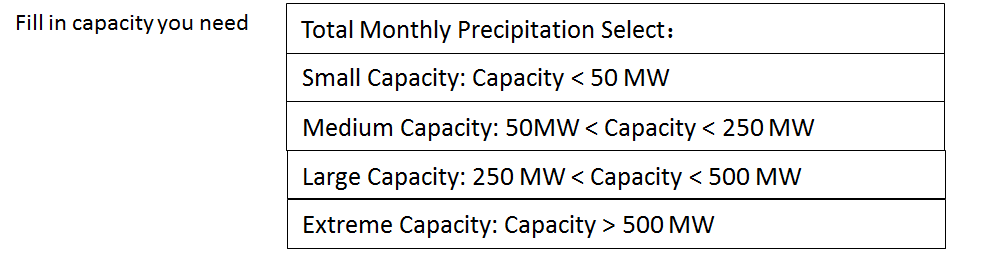


Source: https://water.usgs.gov/edu/watercycleprecipitation.html



Source:https://www.google.com/search?q=what+range+fits+into+large+precipitation&espv=2&biw=1422&bih=684&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiYu7el36zSAhVD22MKHYAMBdIQ\_AUIBygC&dpr=1.35#tbm=isch&q=wind+speed+types&\*&imgdii=VwUAii8SbtYXSM:&imgrc=PDU8nXkd9BGMpM

**Input electricity generation capacity**:



**Output:**

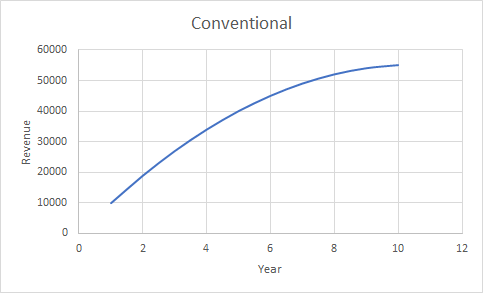
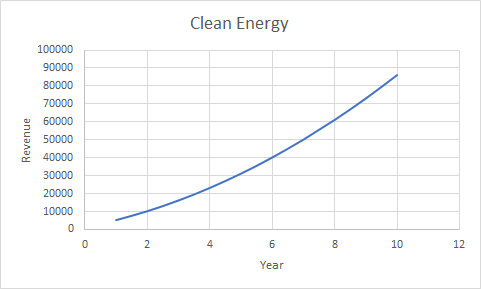
Top Three Recommend Generation Resource and their Plot (Accumulative Revenue vs. Year)

**Eg.**

1. Solar

2. Natural Gas

3. Petroleum



**RF**

**Goal:** According to the input weather data, filter out the impossible clean energy resources.

**Method:** Random Forest (Classification)

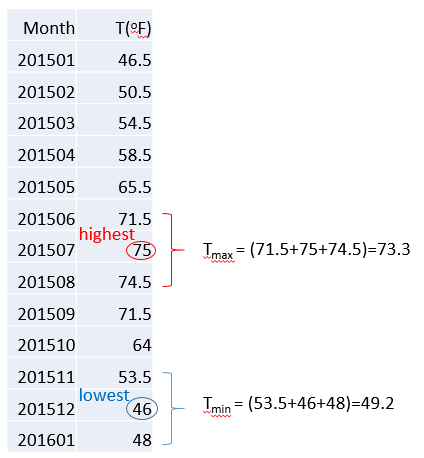
**Procedure:**

1. Temperature data management:

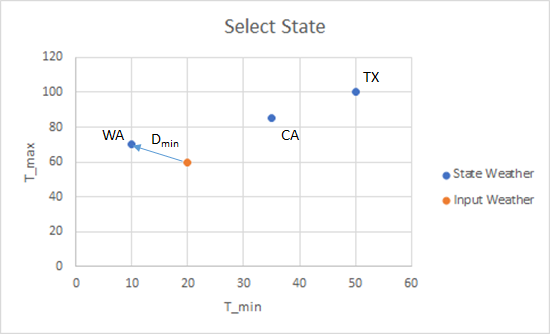
* Calculate Tmax and Tmin for each state

Find the highest and lowest temperature and their month, use the temperature data of nearby 3 months to calculate the average temperature as the final Tmax and Tmin.

Eg.For 2015,CA:



* Set Tmin as x, Tmax as y, make a scatter plot.



* Calculate the distances from input data to all the weather data of different states and find the shortest distance. Choose the corresponding state to represent the input data.

2. Random Forest (Vote for several possible states)

Results: get all the possible states that fit the weather input data

Eg. 4 votes for WA and 2 votes for CA

|  |  |  |  |
| --- | --- | --- | --- |
|  | AvgTemp | TotalMonthlyPrecip | AvgWindSpeed |
| WA |  |  |  |
| CA |  |  |  |

3. According to the states results, find the percentage of all the generation resource in these states, use the voting weight(4 votes for WA and 2 votes for CA, so WA\*66.7%, CA\*33.3%) to calculate the average percentage for each resource

Eg.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| P | Coal | NG | P | Hydro | Solar | Wind |
| WA | 4.6% | 11.9% | 0.02% | 67.2% | 0.0007% | 6.5% |
| CA | 1.6% | 44.7% | 3.3% | 14.3% | 0.2% | 1.6% |
| AvgP |  |  |  |  |  |  |

4. Use the P-Value to filter the impossible resource

**Regression**

**Goal:** In all the possible resources, select top three recommend resource according to the revenue

**Method:** Regression

**Procedure:**

1. Make the cost and revenue dataframe

Calculated average cost of certain resource by the voting weight of states

Eg. 4 votes for WA and 2 votes for CA,;The cost of NG in CA is 5.4, in WA is 5.09

so avgcost\_NG = 5.4\*33.3%+5.09\*66.7% = 5.19

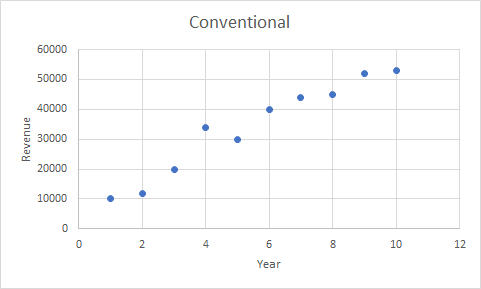
|  |  |  |  |
| --- | --- | --- | --- |
| All the Possible Resources | Cost(cents/kWh) | CO2 Tax  (Cents/MetricTon) | Sales Price  (GovernPrice)  (cents/kWh)  (industrial? total?) |
| Solar |  |  |  |
| Wind |  |  |  |
| Coal |  |  |  |
| NG |  |  |  |
| Petroleum |  |  |  |

2. Calculate the revenue of different resource by different year:

Cost of self-generated electricity = self generation Cost \* Capacity + Carbon tax

Cost of Buying electricity from Government = Sales Price (GovernPrice)\*Capacity

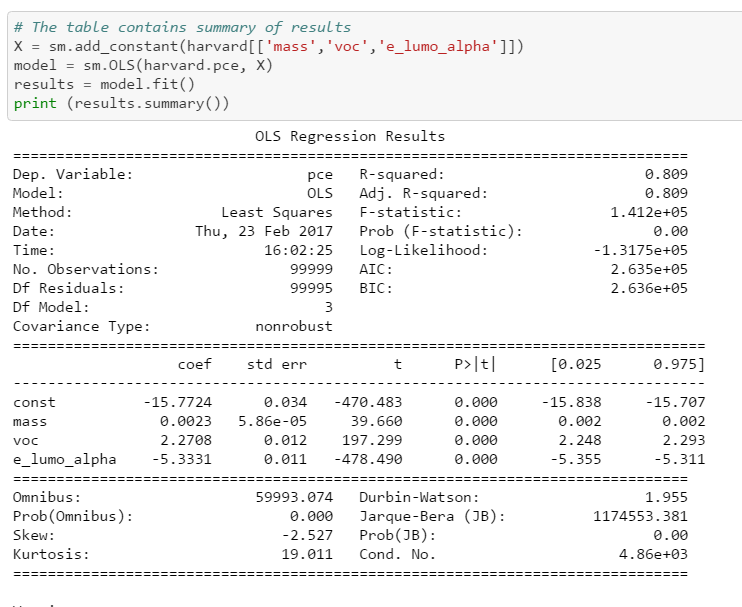
Revenue = Cost of Buying electricity from Government - cost of self generating electricity



3. Logistc Regression

Calculate P value for each coefficient

E.g HCEPDB



Drop the uncorrelated beta values

Replot Revenue Vs. Year

4. Optimized Revenue Vs. Year (By Type)

