Report about the analysis about electricity

cost balancing in Corolado

Sutong Zhang

# Abstract

This white paper examines the challenges Colorado faces in balancing the energy demands of its rapidly growing population and economy with the need for conservation and sustainability. This paper provides the analysis electricity usage and efficiency in Colorado in past 20 years and discuss about the reasons that may cause the increasing trends in electricity cost besides the increasing of the population. Also, it will provide some machine learning process with different model that have some prediction on the trends of electricity cost in the future.

# 1. Introduction

Based on the latest report, the population of Colorado increase about 3 million in the past 20 years. Although the overall rank is 21st most, the growing population also bring Colorado some challenges on the energy usage. After research, I find out there are two ways that wo can dig in: the renewable recourses and the efficiency of usage of energy. I choose to do the analysis on the electricity usage and cost base for the past 20 years. Because I think no matter which energy resources we choose, we can use a more efficient way to save the cost.

After the research on bic dataset, I choose the Electricity Revenue in Colorado dataset which contains all the electricity consumption revenue, amount, unit price and number of customers from 1990 to 2020(monthly). Also, it also separates the data to five section of electricity usage: residential, commercial, industrial, transportation and others. So, I try to find the relationship between total revenue each month and other variables.

# 2 Data Processing

To begin with, I start with data cleaning. As this is a large dataset, I separate it into different section that I need. First, I plot graphs about the change of each revenue and sales by years.

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Base on the first plot about the number of megawatt hours sold, we can find that all section of sales is in the slowly increasing trend besides in 2020. Baes on the second graph about revenue from electricity consumption, we can find that the trend is increasing and drop in 2020 as the last graph. However, it is interesting to notice that the the revenue of residential revenue is more than commercial revenue in the past few years even if the sales amount is opposite. It may cause by different price and customer number.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

For the next step, I plot the percentage proportional graph about the two data sets. Based on the graph above, we can find that commercial usage of electricity is the largest part of the whole period in both sales and revenue and residential usage of electricity comes after. On the other hand, the transportation and other sections are the least. Now, we understand the largest part of cost of electricity usage is on commercial and residential parts. So, we should focus on how to let commercial and residential usage more efficient.

Chart, scatter chart

Description automatically generated

The price of different type of electricity usage is valid base on the law and the price the changing every month. Base on the graph I plot about electricity unit price against month, we can find some interesting facts. On one hand, we can see that the price on winter is higher than other months, especially in other and residential section. On the other hand, the price of residential, transportation and commercial is above the average(total).

#3 Model Fitting

Base on the information we get above, I decide to fit some models and use machine learning to predict the future revenue base on the previous data. To achieve the goal, I choose month, number of customers and unit price as the predictor. My aim is to build up model to find the relationship between predictors and the total revenue.

Chart, shape, bubble chart

Description automatically generated

I do the correlation analysis between the few predictors. From the graph above, we can find that the Price have the most correlation with the total revenue of electricity usage and the month has the least correlation with total revenue.

Then I split the train and test dataset randomly by 80% and 20 %.

Chart, scatter chart

Description automatically generated

First, I use the glm model base on lm(totalRevenue ~ month + totalCustomers + totalPrice, data = train\_data). The graph above is the result between prediction and actual data by glm model. The mse is 676394947.

Chart, scatter chart

Description automatically generated

Second, I use the random forest model base on lm(totalRevenue ~ month + totalCustomers + totalPrice, data = train\_data). The graph above is the result between prediction and actual data by glm model. The mse is 661775168. Comparing these two models, the random forest has the less mse than the glm model. However, the difference is not that large.

Chart

Description automatically generated with medium confidence

For the last model, I try to use the time series forecasting by month and try to predict the next 15 months' total revenue after the 2020. The plot above can see that the revenue is kind of increase in a linear slope.

#4 Conclusion

Base on the result I got above, I build a simple shiny website (http://127.0.0.1:4924/) to show the result. From the analysis about the cost of electricity in Colorado from 1990 to 2020, we can find that the commercial and residential takes the most proportion in the total revenue. The government may want to make more advertising about how to use electricity more efficient. Also, the price of electricity has the most influence on the total revenue besides consumer number and time(weather). In my opinion, the prediction is keep climbing for the next 15 years, however, it may be change due to lack of the data after 2020. All in all, we should focus on the research about how to control price to make a more efficient electricity cost balancing on residential & commercial usage for the future.

# 5. References

Data sourse: https://data.colorado.gov/Business/Electricity-Revenue-in-Colorado/q6sk-tjm9

Other article: https://www.simplilearn.com/tutorials/data-science-tutorial/time-series-forecasting-in-r