Team Insight

Vidit, Tejus, Gilmo, Andrew

23/09/2021

# Prediction of Economic Growth from Electrical Power Consumption

Nowadays we have become energy driven a lot. Everything from our basic household equipment to the hugh factories to our day to day transport require energy. And the amount of electric consumption could relate to how developed the state is. So it wouldn’t be wrong to say that there is a relation between electric consumption and the GDP of a state.

If you analyse consumption in a specific area, you can find out which area has the greatest potential and through this, we will be able to open a new business in the optimal location.

Even power consumption of a particular company can tell you lots about them, and can help you in your stock investment.

## What we need to establish with this report.

1. *How does the state’s power consumption relate to the GDP ?*
2. *How well does each state’s power consumption predict their GDP ?*

## Data Sources Used

1. Annual power usage by state of USA [<https://www.eia.gov/electricity/data/state/sales_annual.xlsx>]

* We got this data from the Energy Information Administration website . It is an annual energy consumption data for the states in US for the year 1997 to 2018. Here the data was available as years in columns so we had to transpose that data .

1. GDP by state of USA [<https://github.com/vidit-sharma/Data_GDP_US/raw/main/GDP%20by%20state%20(1).xls>]

* We got this data from the Bureau of Economic Analysis. It is the annual GDP data for the states of US for the year from 1990 to 2018. Here we only use the column that has the total energy consumption for our analysis.

Libraries Used in this analysis:

library(readxl)

## Warning: package 'readxl' was built under R version 4.1.1

library(tidyr)  
library(dplyr)  
library(stringr)  
library(ggplot2)  
library(patchwork)

## Warning: package 'patchwork' was built under R version 4.1.1

## Loading and Cleansing the Datasets

Lets load the electric power consumption data into R.

url <- "https://www.eia.gov/electricity/data/state/sales\_annual.xlsx"  
destfile <- "sales\_annual.xlsx"  
curl::curl\_download(url, destfile)  
elec\_usage <- read\_excel(destfile)

## New names:  
## \* `` -> ...2  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5  
## \* `` -> ...6  
## \* ...

Looking at the dataset, we need to cleanse it and transform the data as per our requirement .

colnames(elec\_usage) <- elec\_usage[1,]  
elec\_usage <- elec\_usage[-1,]  
elec\_usage <- elec\_usage[,-c(4:8)]  
elec\_usage <- elec\_usage %>% group\_by(Year) %>%  
 filter(`Industry Sector Category` == "Total Electric Industry")

We see that the data available to us for the state name is in the form of State Abbreviation. So we have to use the state abbreviation to get full State name.

state\_abbreviation <- read.csv("https://worldpopulationreview.com/static/states/abbr-name.csv", header = FALSE)  
colnames(state\_abbreviation) <- c("State", "State\_full")  
elec\_usage <- merge(elec\_usage, state\_abbreviation, by = "State")  
elec\_usage[,1] <- elec\_usage[,5]  
elec\_usage <- elec\_usage[-5]  
table(is.na(elec\_usage))

##   
## FALSE   
## 6120

Now we need to load and cleanse the GDP data for the state of US so as to transform the data as per our requirement.

url1 <-   
 "https://github.com/vidit-sharma/Data\_GDP\_US/raw/main/GDP%20by%20state%20(1).xls"  
destfile1 <- "GDP%20by%20state%20(1).xls"  
curl::curl\_download(url1, destfile1)  
GDP\_by\_state <- read\_excel(destfile1)

## New names:  
## \* `` -> ...2  
## \* `` -> ...3  
## \* `` -> ...4  
## \* `` -> ...5  
## \* `` -> ...6  
## \* ...

colnames(GDP\_by\_state) <- GDP\_by\_state[5,]  
GDP\_by\_state <- GDP\_by\_state[-c(1:5),-1]  
GDP\_by\_state <- pivot\_longer(GDP\_by\_state, -GeoName, names\_to = "Year", values\_to = "GDP")  
colnames(GDP\_by\_state) <- c("State", "Year", "GDP")

## Merging the Datasets

Now as both the data are cleansed and ready, we will merge both the datasets on the key as State and Year so as to get the GDP data of the state and Electric consumption of the state in the same data frame.

elec\_usage[,"state-year"] <- data.frame(paste(elec\_usage$State, elec\_usage$Year, sep = "-"))  
GDP\_by\_state[,"state-year"] <- data.frame(paste(GDP\_by\_state$State, GDP\_by\_state$Year, sep = "-"))  
GDP\_by\_state <- GDP\_by\_state[,-c(1:2)]  
elec\_GDP <- merge(elec\_usage, GDP\_by\_state, by = "state-year")  
elec\_GDP <- elec\_GDP[,-c(2:4)]  
colnames(elec\_GDP) <- c("State-Year", "Electric Usage", "GDP")  
table(is.na(elec\_GDP))

##   
## FALSE   
## 3450

elec\_GDP <- cbind((t(data.frame(elec\_GDP$`State-Year` %>% str\_split("-")))),  
 elec\_GDP)  
elec\_GDP <- elec\_GDP[,-3]   
colnames(elec\_GDP) <- c("State", "Year", "Electric Usage", "GDP")  
rownames(elec\_GDP) <- NULL

Lets compare the GDP of the state with the GDP on the next year of the same state.

GDP <- elec\_GDP[4]  
GDP[1151,] <- 9999999  
GDP <- GDP[-1,]  
elec\_GDP\_plus1 <- cbind(elec\_GDP,GDP)  
colnames(elec\_GDP\_plus1) <- c("State", "Year", "Electric Usage", "GDP", "GDP+1year")  
elec\_GDP\_plus1 <- elec\_GDP\_plus1 %>% filter(Year != 2019) # state change correction  
  
elec\_GDP\_plus1[,"GDP"] <- round(as.numeric(elec\_GDP\_plus1$GDP))  
elec\_GDP\_plus1[,"GDP+1year"] <- round(as.numeric(elec\_GDP\_plus1$`GDP+1year`))

## Visualization

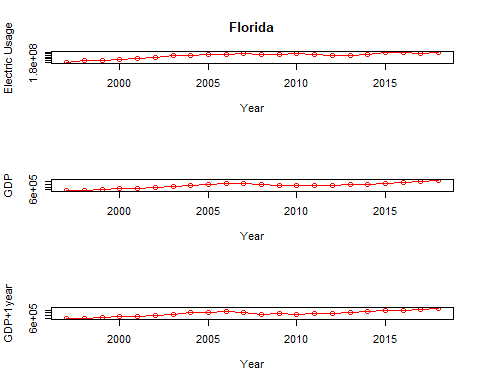
Now as we will segregate the data based on the state.

state\_abbreviation <- state\_abbreviation[,-1]  
  
for(i in state\_abbreviation){  
 assign(paste0("elec\_GDP\_plus1\_",i), elec\_GDP\_plus1 %>% filter(State == i))  
}

Lets take a look at the graphs for Energy consumption vs year and GDP vs Year so as to analyse the trend on change in both the graphs and to find a correlation between them. We also compare the data for GDP of the next year vs year. We will look for a couple of states so as to generalize the result.

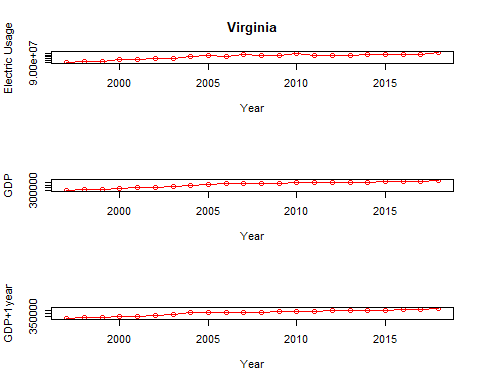
1. For the state of Florida

par(mfrow = c(3,1))  
plot(elec\_GDP\_plus1\_Florida$Year, elec\_GDP\_plus1\_Florida$`Electric Usage`,   
 main = "Florida", xlab = "Year", ylab = "Electric Usage",   
 type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Florida$Year, elec\_GDP\_plus1\_Florida$`GDP`,   
 xlab = "Year", ylab = "GDP", type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Florida$Year, elec\_GDP\_plus1\_Florida$`GDP+1year`,  
 xlab = "Year", ylab = "GDP+1year", type = "o", col = "red")



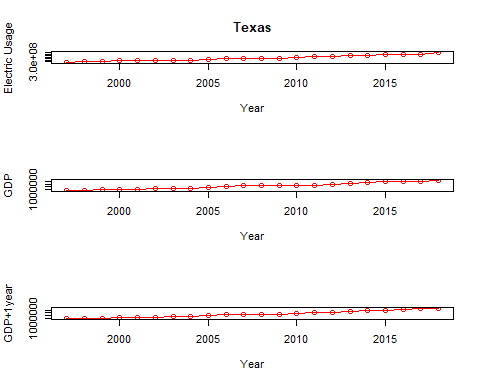
1. For the state of Virginia

par(mfrow = c(3,1))  
plot(elec\_GDP\_plus1\_Virginia$Year, elec\_GDP\_plus1\_Virginia$`Electric Usage`,  
 main = "Virginia", xlab = "Year", ylab = "Electric Usage",   
 type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Virginia$Year, elec\_GDP\_plus1\_Virginia$`GDP`,  
 xlab = "Year", ylab = "GDP", type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Virginia$Year, elec\_GDP\_plus1\_Virginia$`GDP+1year`,  
 xlab = "Year", ylab = "GDP+1year", type = "o", col = "red")



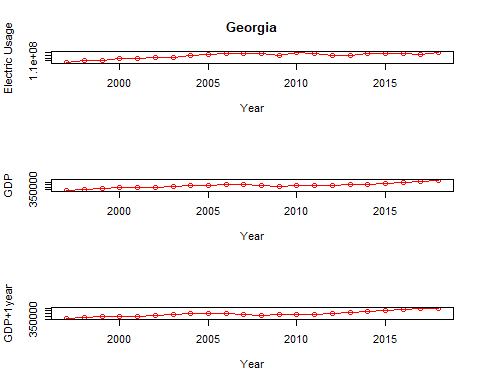
1. For the state of Texas

par(mfrow = c(3,1))  
plot(elec\_GDP\_plus1\_Texas$Year, elec\_GDP\_plus1\_Texas$`Electric Usage`,  
 main = "Texas", xlab = "Year", ylab = "Electric Usage",  
 type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Texas$Year, elec\_GDP\_plus1\_Texas$`GDP`,  
 xlab = "Year", ylab = "GDP", type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Texas$Year, elec\_GDP\_plus1\_Texas$`GDP+1year`,  
 xlab = "Year", ylab = "GDP+1year", type = "o", col = "red")



1. For the state of Georgia

par(mfrow = c(3,1))  
plot(elec\_GDP\_plus1\_Georgia$Year, elec\_GDP\_plus1\_Georgia$`Electric Usage`,   
 main = "Georgia", xlab = "Year", ylab = "Electric Usage",  
 type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Georgia$Year, elec\_GDP\_plus1\_Georgia$`GDP`,  
 xlab = "Year", ylab = "GDP", type = "o", col = "red")  
plot(elec\_GDP\_plus1\_Georgia$Year, elec\_GDP\_plus1\_Georgia$`GDP+1year`,  
 xlab = "Year", ylab = "GDP+1year", type = "o", col = "red")



From the above plots, we found that electricity consumption and GDP have a positive correlation. And whenever there is a dip in electrical consumption , there is a dip in GDP too. Thus this establishes the answer to the first question of our report.

## Models

Now let us use the above data to predict GDP for further years based on Electric consumption. For this we will need two models. First will help us predict values of electric consumption for the years 2019 to 2023.And the second will use the predicted data of electric usage to predict GDP data for the year 2019 to 2023.

Colorado\_rearrage <- elec\_GDP\_plus1\_Colorado[, c(1, 4, 5, 3, 2)]  
Colorado\_rearrage<-transform(Colorado\_rearrage, `Electric Usage` = as.numeric(`Electric Usage`),Year=as.numeric(Year))

Model 1: Electric Usage based on previous years electric data

mod1<- lm(Electric.Usage ~ Year,data=Colorado\_rearrage)  
summary(mod1)

##   
## Call:  
## lm(formula = Electric.Usage ~ Year, data = Colorado\_rearrage)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2766221 -1265788 218084 1078845 2454216   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.566e+09 1.058e+08 -14.81 3.05e-12 \*\*\*  
## Year 8.048e+05 5.269e+04 15.28 1.72e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1568000 on 20 degrees of freedom  
## Multiple R-squared: 0.921, Adjusted R-squared: 0.9171   
## F-statistic: 233.3 on 1 and 20 DF, p-value: 1.721e-12

Model 2: GDP based on the Electric consumption data

mod <- lm( GDP ~ Electric.Usage , data = Colorado\_rearrage)  
summary(mod)

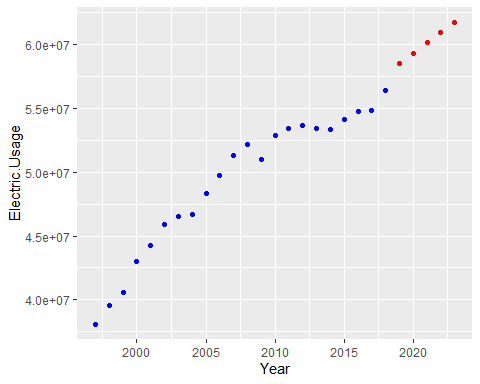
##   
## Call:  
## lm(formula = GDP ~ Electric.Usage, data = Colorado\_rearrage)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21840 -10701 -3400 11311 32347   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.919e+04 3.045e+04 -2.272 0.0343 \*   
## Electric.Usage 6.726e-03 6.143e-04 10.950 6.74e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15330 on 20 degrees of freedom  
## Multiple R-squared: 0.857, Adjusted R-squared: 0.8499   
## F-statistic: 119.9 on 1 and 20 DF, p-value: 6.737e-10

Lets calculate the prediction data for GDP based on the predicted data of electric consumption.

Colorado\_rearrage<-Colorado\_rearrage %>% mutate(pred = predict(mod))  
  
Colorado\_add<- data.frame(State = c("Colorado","Colorado", "Colorado","Colorado", "Colorado"),   
 GDP = c(0,0,0,0,0),  
 GDP.1year=c(0,0,0,0,0),  
 Electric.Usage=c(predict(mod1, data.frame(Year = 2019)),  
 predict(mod1, data.frame(Year = 2020)),  
 predict(mod1, data.frame(Year = 2021)),  
 predict(mod1, data.frame(Year = 2022)),  
 predict(mod1, data.frame(Year = 2023))),  
 Year=c(2019,2020,2021,2022,2023),  
 pred=c(predict(mod, data.frame(Electric.Usage =  
 predict(mod1, data.frame(Year = 2019)))),  
 predict(mod, data.frame(Electric.Usage =   
 predict(mod1, data.frame(Year = 2020)))),  
 predict(mod, data.frame(Electric.Usage =  
 predict(mod1, data.frame(Year = 2021)))),  
 predict(mod, data.frame(Electric.Usage =  
 predict(mod1, data.frame(Year = 2022)))),  
 predict(mod, data.frame(Electric.Usage =   
 predict(mod1, data.frame(Year = 2023))))))  
  
Colorado\_rearrage <- rbind(Colorado\_rearrage,Colorado\_add )

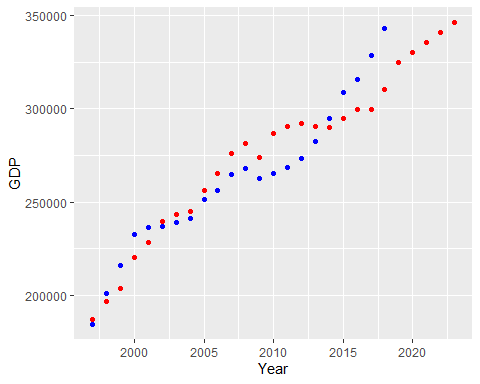
### Model Visualizations for the state of Colorado

ggplot() +  
 geom\_point(data=Colorado\_rearrage[Colorado\_rearrage$Year<=2018,],  
 aes(x = Year, y = Electric.Usage), color = "blue") +  
 geom\_point(data=Colorado\_rearrage[Colorado\_rearrage$Year>2018,],  
 aes(x = Year, y = Electric.Usage), color = "red")



The above graph shows energy usage prediction based on previous year data for energy usage for the state of Colorado. Here Blue points represents the value from our data and Red points represent the prediction data.

ggplot() +  
 geom\_point(data=Colorado\_rearrage[Colorado\_rearrage$Year<=2018,],  
 aes(x = Year, y = GDP), color = "blue") +  
 geom\_point(data=Colorado\_rearrage,aes(x = Year, y = pred), color = "red")



The above graph shows GDP data prediction based on energy usage data predicted previously on year for the state of Colorado. Here Blue points represents the actual GDP data from our data and Red points represent the GDP predictions based on the electric usage data.

## Github Link

We have uploaded our report in rmd and knitted format at the provided link :[<https://github.com/vidit-sharma/Predict-GDP-based-on-Electric-Consumption.git>]

## Conclusion

As a result of data analysis, it was found that electricity consumption and GDP had a positive correlation.

Daily and monthly correlation analysis between electricity consumption and GDP can be used as a leading indicator of economic trends.

## Possible sources of Bias

One of the possible sources of bias is the source data, regarding how it was collected and reported.

There is a possibility of selection bias, as the power consumption records could be of only the major cities and not the countryside.