MITA CAPSTONE PROJECT

Google Project Sunroof

Under the guidance of: Prof. Meng Qu

Done by:

Darshan Dalvi

181006386

Table of Contents

Introduction

As the price of installing solar has gotten less expensive, more homeowners are turning to it as a possible option for decreasing their energy bill. Google wants to make installing solar panels easy and understandable for anyone.

Project Sunroof puts Google's expansive data in mapping and computing resources to use, helping calculate the best solar plan for customers.

Project Sunroof computes how much sunlight hits your roof in a year. It takes into account:

- Google's database of imagery and maps
- 3D modeling of your roof
- Shadows cast by nearby structures and trees
- All possible sun positions over the course of a year
- Historical cloud and temperature patterns that might affect solar energy production

Project Sunroof recommends an installation size to generate close to 100% of your electricity use, based on roof size, the amount of sun hitting the roof, and your electricity bill.

To get more information about solar panels section visit Google Sunroof Project

Dataset

- This dataset includes data on rooftop panels, carbon offset, and yearly sunlight from Google Project Sunroof.
- This data is used for Project Sunroof's Data, Trends, and Maps database, which provides national and state specific data on carbon offset, yearly sunlight, and number of panels.
- The dataset consists of 31 data columns and 48723 entries.

```
##  [1] "carbon_offset_metric_tons"
##  [2] "count_qualified"
##  [3] "existing_installs_count"
##  [4] "install_size_kw_buckets"
##  [5] "kw_median"
##  [6] "kw_total"
##  [7] "lat_avg"
##  [8] "lat_max"
##  [9] "lat_min"
## [10] "lng_avg"
## [11] "lng_max"
## [12] "lng_min"
## [13] "number_of_panels_e"
## [14] "number_of_panels_f"
## [15] "number_of_panels_median"
## [16] "number_of_panels_n"
## [17] "number_of_panels_s"
## [18] "number_of_panels_total"
## [19] "number_of_panels_w"
## [20] "percent_covered"
## [21] "percent_qualified"
## [22] "region_name"
## [23] "state_name"
## [24] "yearly_sunlight_kwh_e"
## [25] "yearly_sunlight_kwh_f"
## [26] "yearly_sunlight_kwh_kw_threshold_avg"
## [27] "yearly_sunlight_kwh_median"
## [28] "yearly_sunlight_kwh_n"
## [29] "yearly_sunlight_kwh_s"
## [30] "yearly_sunlight_kwh_total"
## [31] "yearly_sunlight_kwh_w"
```

Problem Statement

- This dataset provides national and state specific data on carbon offset, yearly sunlight, and No. of panels.
- The problem statement deals with whether solar roofing has an impact on carbon offset.
- In the analysis, the qualified cities for solar rooftop in each state is compared and linear regression is performed to compare the carbon offset of these cities and how it can be reduced by installing solar rooftops.

Experiment

Explanatory Data Analysis

We start by creating a table showing us the Percentage of Non-Null Values.

```
GetPercentageOfNonNullValues <- function(OurDataSet) {

  PercentageOfNonNullValues = ( colSums(!is.na(OurDataSet)) / nrow(OurDataSet) ) * 100

  PercentageOfNonNullValues = as.data.frame(PercentageOfNonNullValues)
  PercentageOfNonNullValues = rownames_to_column(PercentageOfNonNullValues,"VariableName")

  PercentageOfNonNullValues =  PercentageOfNonNullValues %>% arrange(desc(PercentageOfNonNullValues))

  PercentageOfNonNullValues = PercentageOfNonNullValues %>%
    mutate(VariableName =  reorder(VariableName,PercentageOfNonNullValues)) %>% tail(20)

datatable(PercentageOfNonNullValues, style="bootstrap", class="table-condensed", options = list(dom = 'tp',scrollX
 = TRUE))

}


GetPercentageOfNonNullValues(sunroof)
```

In this table below, only the Lowest Twenty Percentage Values are shown.

	VariableName	PercentageOfNonNullValues
12	number_of_panels_n	100
13	number_of_panels_s	100
14	number_of_panels_w	100
15	percent_covered	100
16	percent_qualified	100
17	region_name	100
18	state_name	100
19	yearly_sunlight_kwh_e	100
20	yearly_sunlight_kwh_f	100
21	yearly_sunlight_kwh_n	100

Previous 1 2 Next

	VariableName	PercentageOfNonNullValues
22	yearly_sunlight_kwh_s	100
23	yearly_sunlight_kwh_w	100
24	yearly_sunlight_kwh_kw_threshold_avg	99.9979475390994
25	install_size_kw_buckets	99.880957267764
26	kw_median	99.880957267764
27	kw_total	99.880957267764
28	number_of_panels_median	99.880957267764
29	number_of_panels_total	99.880957267764
30	yearly_sunlight_kwh_median	99.880957267764
31	yearly_sunlight_kwh_total	99.880957267764

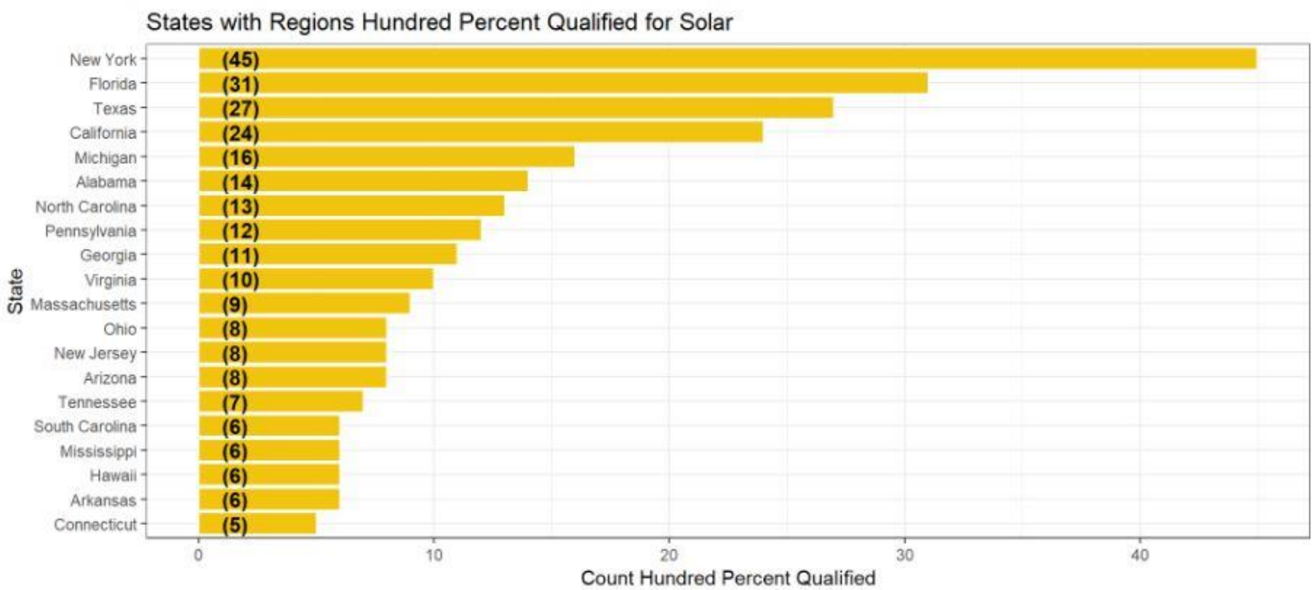
Previous 1 2 Next

We observe all the variables have more than 99% values.

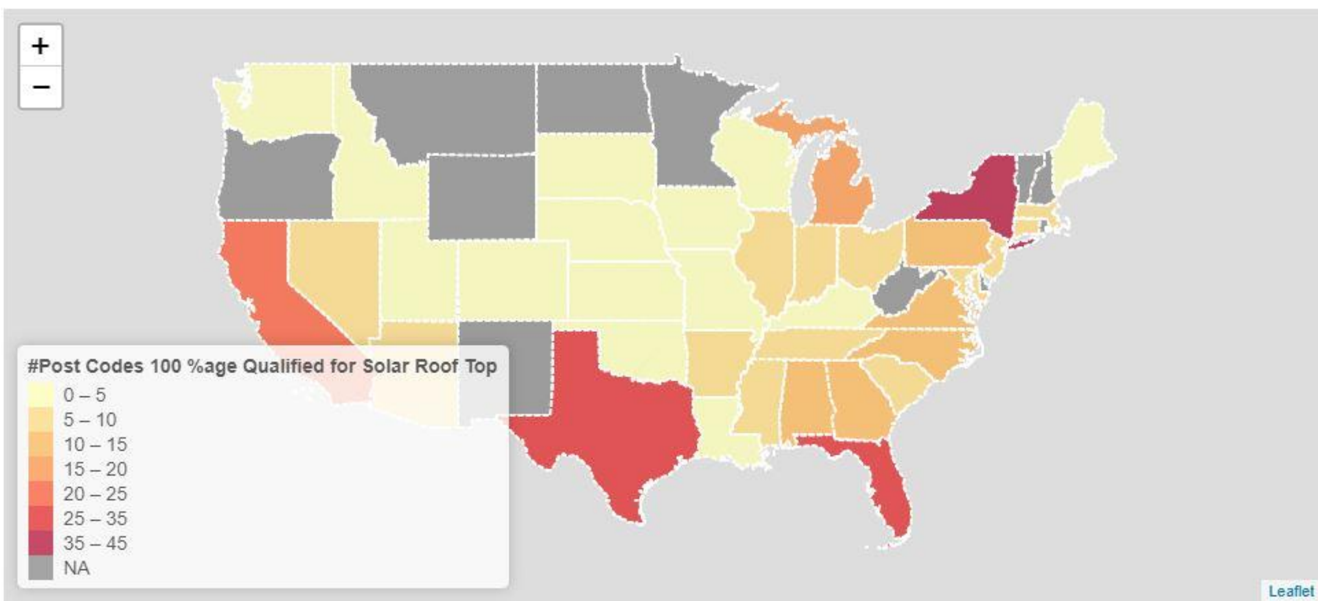
We find the states which has regions with Hundred percentage Qualification for Solar Roof Top.

In the below bar plot, we show the states which has the Maximum number of such regions.

```
sunroof$percent_qualified = as.numeric(sunroof$percent_qualified)

sunroof %>%
  filter(percent_qualified == 100) %>%
  group_by(state_name) %>%
  tally() %>%
  mutate(state_name = reorder(state_name,n)) %>%
  arrange(desc(n)) %>%
  head(20) %>%

  ggplot(aes(x = state_name,y = n)) +
  geom_bar(stat='identity',colour="white", fill =fillColor2) +
  geom_text(aes(x = state_name, y = 1, label = paste0("(",n,")",sep="")),
            hjust=0, vjust=.5, size = 4, colour = 'black',
            fontface = 'bold') +
  labs(x = 'State', y = 'Count Hundred Percent Qualified',
       title = 'States with Regions Hundred Percent Qualified for Solar') +
  coord_flip() +
  theme_bw()
```



This choropleth shows the **No of Regions** in each state which is **100 Percentage Qualified for Solar Roof Top**.



We observe that the States of **New York**, **California**, **Texas**, **Pennsylvania**, **Illinois** have very High Number of Post Codes with Hundred Percent Qualification for Solar.

We observe that New York has the Highest number of regions with **Hundred Percent** Qualifications for Solar. Therefore, we now go deeper and examine the **Counties in New York** and show in a bar plot the counties and their Percentage Qualification for Solar.

```
CountyData = read_csv("C:/Users/ddalv/Documents/Courses/Capstone/google-project-sunroof/project-sunroof-county-0908
2017.csv")

NYCountyData = CountyData %>%
  filter(state_name == "New York") %>%
  arrange(desc(percent_qualified)) %>%
  mutate(region_name =  reorder(region_name,percent_qualified))

NYCountyData %>%
  head(20) %>%

  ggplot(aes(x = region_name,y = percent_qualified)) +
  geom_bar(stat='identity',colour="white", fill =fillColor) +
  geom_text(aes(x = region_name, y = 1, label = paste0("(",round(percent_qualified),"%",")",sep="")),
            hjust=0, vjust=.5, size = 4, colour = 'black',
            fontface = 'bold') +
  labs(x = 'County Name', y = 'Percent Qualified for Solar',
       title = 'Counties in NY With Percent Qualified for Solar') +
  coord_flip() +
  theme_bw()
```

Methodology

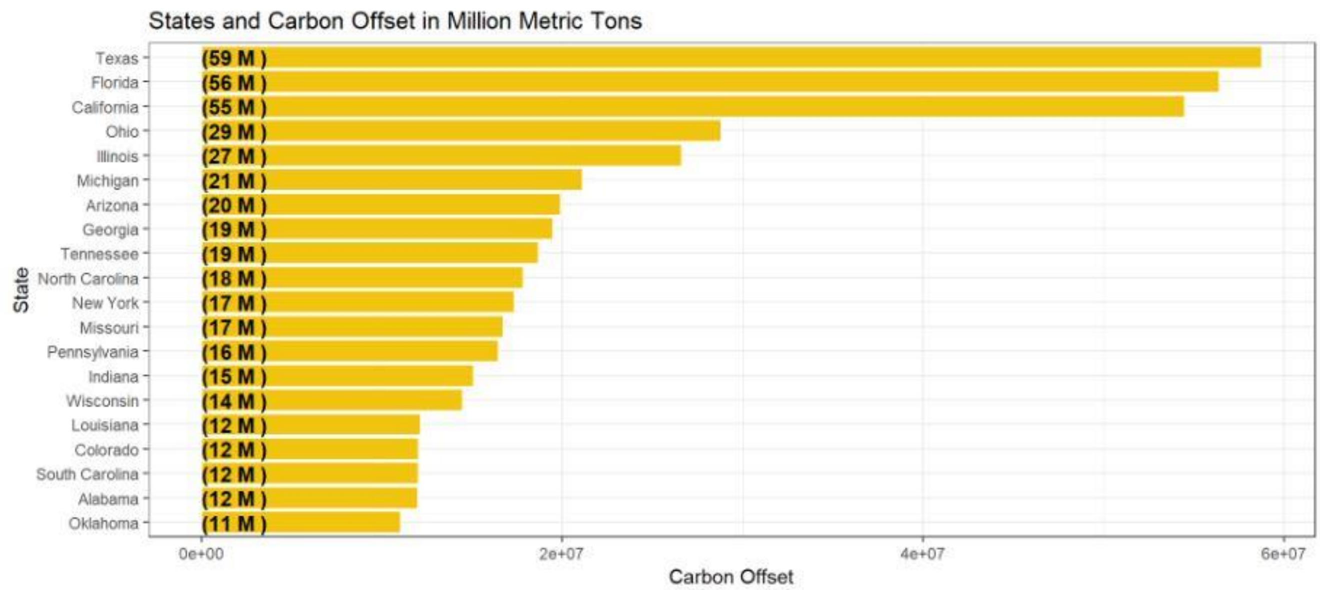
Carbon Offset v/s States

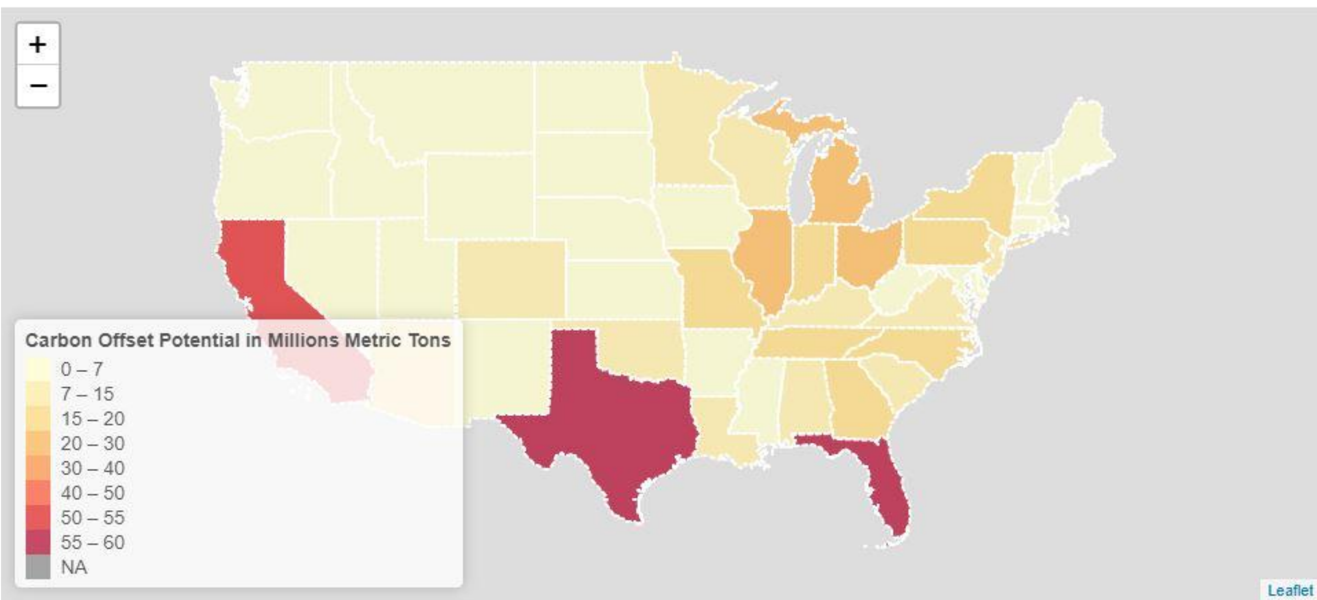
A carbon offset is a reduction in emissions of carbon dioxide or greenhouse gases made in order to compensate for or to offset an emission made elsewhere.

Carbon offsets are measured in metric tons of carbon dioxide-equivalent (CO_2e) and may represent six primary categories of greenhouse gases:[5] carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), perfluorocarbons (PFCs), hydrofluorocarbons (HFCs), and sulfur hexafluoride (SF_6).[6] One carbon offset represents the reduction of one metric ton of carbon dioxide or its equivalent in other greenhouse gases.

We measure the Carbon offset in accordance with each state and plot them on a choropleth graph.

```
CarbonOffsetState = sunroof %>%
  group_by(state_name) %>%
  summarise(CarbonOffset = sum(carbon_offset_metric_tons)) %>%
  arrange(desc(CarbonOffset)) %>%
  mutate(state_name = reorder(state_name,CarbonOffset))


CarbonOffsetState %>%

  head(20)%>%

  ggplot(aes(x = state_name,y = CarbonOffset)) +
  geom_bar(stat='identity',colour="white", fill =fillColor2) +
  geom_text(aes(x = state_name, y = 1, label = paste0("(",round(CarbonOffset/1000000)," M ",")",sep="")),
            hjust=0, vjust=.5, size = 4, colour = 'black',
            fontface = 'bold') +
  labs(x = 'State', y = 'Carbon Offset',
       title = 'States and Carbon Offset in Million Metric Tons') +
  coord_flip() +
  theme_bw()
```



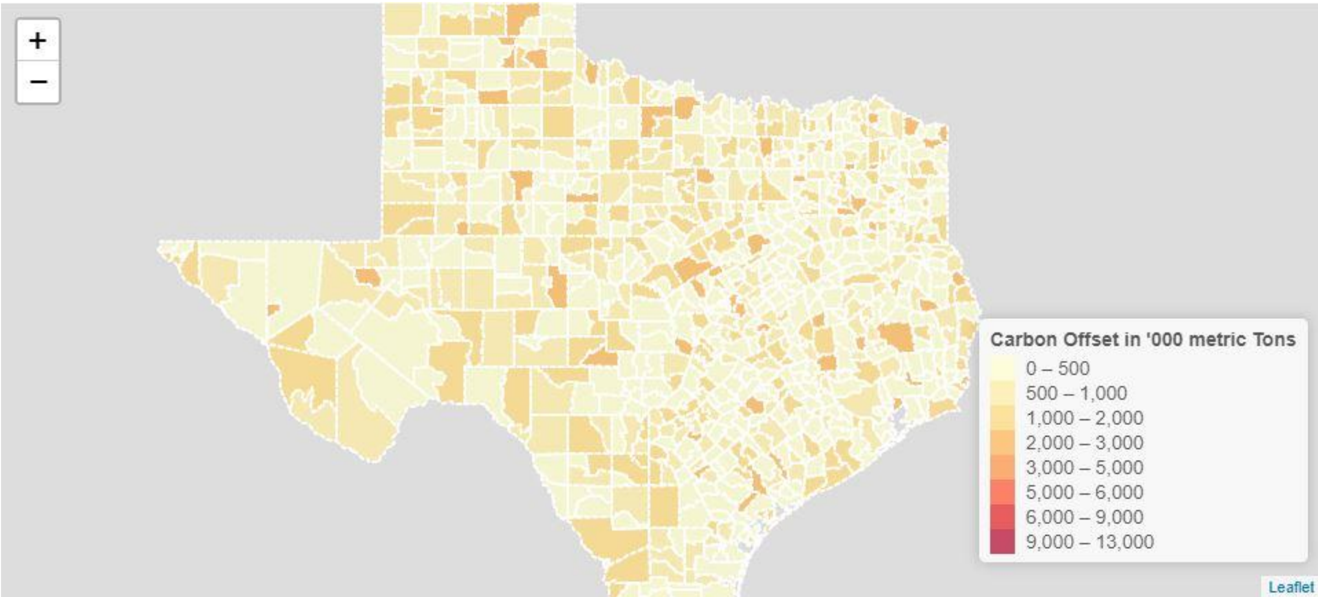


We observe that **Texas , Florida and California** have very high potential for Carbon Offset.

We show the **Choropleth** for the Texas Counties with the Carbon Offset Data.

```
CountyData = read_csv("C:/Users/ddalv/Documents/Courses/Capstone/google-project-sunroof/project-sunroof-county-0908
2017.csv")

TexasCountyData = CountyData %>%
  filter(state_name == "Texas") %>%
  group_by(region_name) %>%
  summarise(CarbonOffset = sum(carbon_offset_metric_tons)/1000) %>%
  arrange(desc(CarbonOffset)) %>%
  mutate(region_name =  reorder(region_name,CarbonOffset))


TexasCounties2 = sapply(str_split(TexasCountyData$region_name," "), head, 1)
TexasCountyData$region_name = TexasCounties2

USCounties@data$NAME = as.character(USCounties@data$NAME)

USCounties@data = inner_join(USCounties@data, TexasCountyData, by = c("NAME" = "region_name"))

bins <- c(0,500,1000, 2000, 3000, 5000, 6000,9000, 13000)

pal <- colorBin("YlOrRd", domain = USCounties@data$CarbonOffset, bins = bins)

labels <- sprintf(
  "<strong>%s</strong><br/>%g",
  USCounties@data$NAME, USCounties@data$CarbonOffset
) %>% lapply(htmltools::HTML)
```

```
leaflet(data = USCounties) %>% setView(-98.613281, 31.203405, 6)  %>%
  addPolygons(
    fillColor = ~pal(CarbonOffset),
    weight = 2,
    opacity = 1,
    color = "white",
    dashArray = "3",
    fillOpacity = 0.7,
    highlight = highlightOptions(
      weight = 5,
      color = "#666",
      dashArray = "",
      fillOpacity = 0.7,
      bringToFront = TRUE),
    label = labels,
    labelOptions = labelOptions(
      style = list("font-weight" = "normal", padding = "3px 8px"),
      textsize = "15px",
      direction = "auto")) %>%
  addLegend(pal = pal, values = ~CarbonOffset, opacity = 0.7, title = "Carbon Offset in '000 metric Tons",
            position = "bottomright")
```

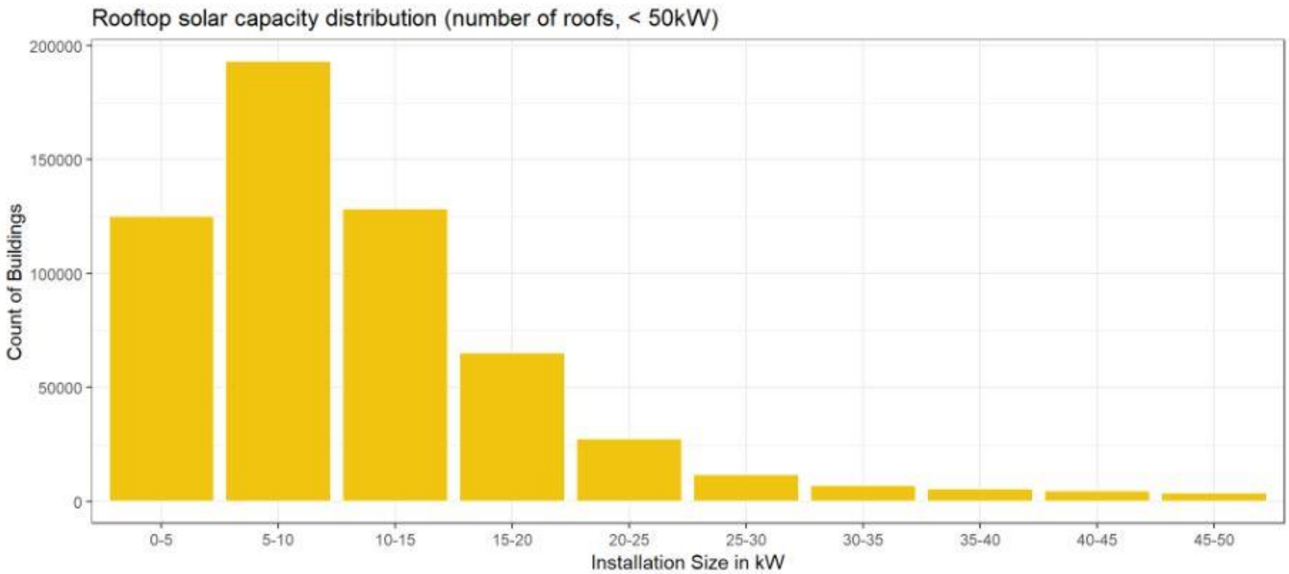


We can see that Huston has the greatest number of Counties with carbon Offset Potential.

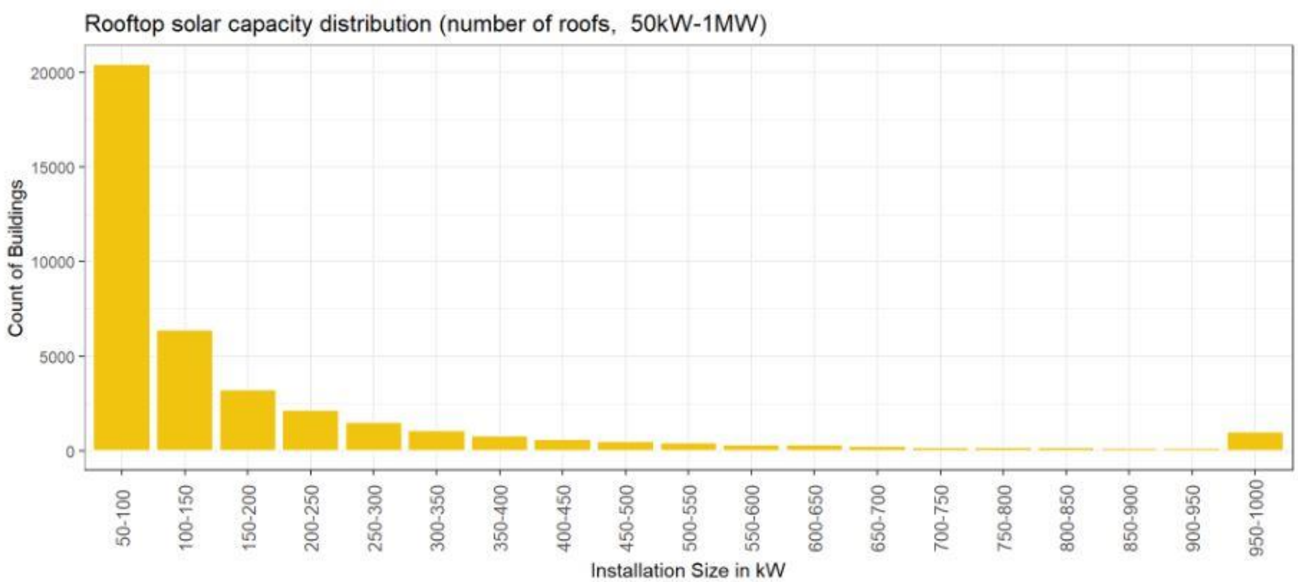
Find the number of building for various Installation Size Buckets in Huston.

- Installation Size Buckets <= 50Kw

```
CityData = read_csv("C:/Users/ddalv/Documents/Courses/Capstone/google-project-sunroof/project-sunroof-city-0908201
7.csv")

HoustonCity = CityData %>%
  filter(region_name == "Houston") %>%
  filter(state_name == "Texas")


HoustonBuckets = jsonlite::fromJSON(HoustonCity$install_size_kw_buckets_json, simplifyDataFrame = TRUE)

Buckets = data.frame(BucketName = character(), BucketKW = numeric())

for( counter in  1:10)
{
  BucketName = paste0(HoustonBuckets[counter,1],"-",HoustonBuckets[counter+1,1])

  BucketRow  = data.frame(BucketName = BucketName, BucketKW = HoustonBuckets[counter,2])

  Buckets  = rbind(Buckets,BucketRow)


}

Buckets %>%

  ggplot(aes(x = BucketName,y = BucketKW)) +
  geom_bar(stat='identity',colour="white", fill =fillColor2) +
  labs(x = 'Installation Size in kW', y = 'Count of Buildings',
       title = 'Rooftop solar capacity distribution (number of roofs, < 50kW)') +
   theme_bw()
```



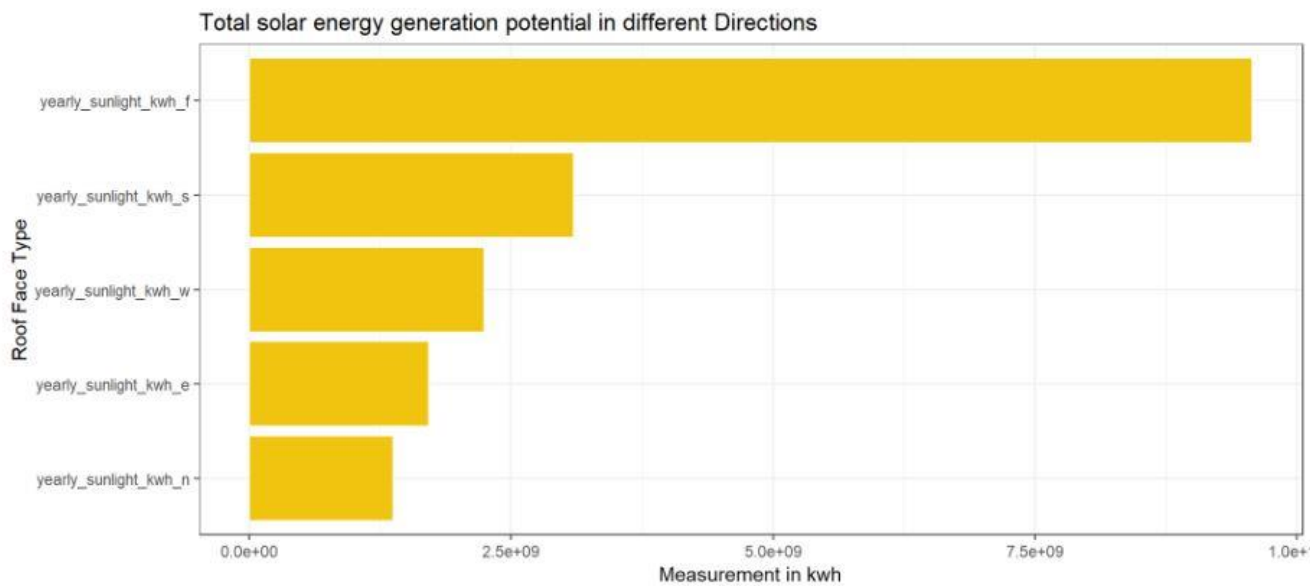
- Installation Size Buckets 50Kw – 1MW



Yearly Solar energy generation potential for different roof faces in Huston.

Description of the different roof faces and the total solar energy generation is shown below:

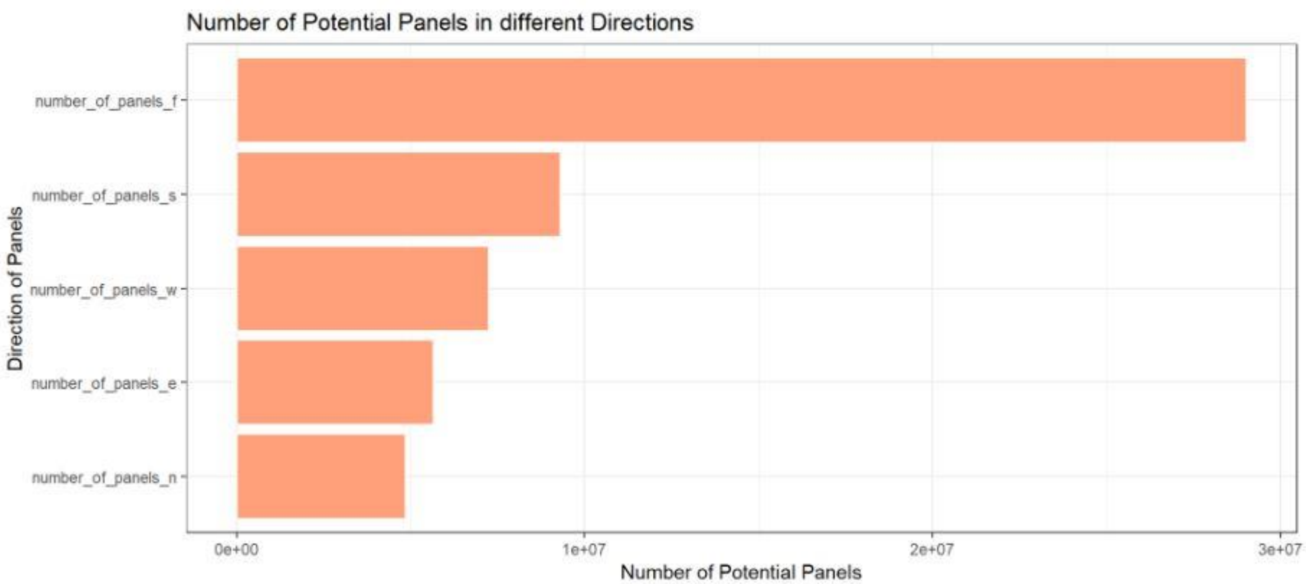
Type	Description
yearly_sunlight_kwh_f	total solar energy generation potential for flat roof space in that region
yearly_sunlight_kwh_n	total solar energy generation potential for north roof space in that region
yearly_sunlight_kwh_e	total solar energy generation potential for east roof space in that region
yearly_sunlight_kwh_s	total solar energy generation potential for south roof space in that region
yearly_sunlight_kwh_w	total solar energy generation potential for west roof space in that region



Number of Solar Panels potential for different roof faces in Houston

Description of the different roof faces and the total solar energy generation is shown below

Type	Description
number_of_panels_f	# of solar panels potential for flat roof space in that region, assuming 1.650m x 0.992m panels
number_of_panels_n	# of solar panels potential for north roof space in that region, assuming 1.650m x 0.992m panels
number_of_panels_e	# of solar panels potential for east roof space in that region, assuming 1.650m x 0.992m panels
number_of_panels_s	# of solar panels potential for south roof space in that region, assuming 1.650m x 0.992m panels
number_of_panels_w	# of solar panels potential for west roof space in that region, assuming 1.650m x 0.992m panels



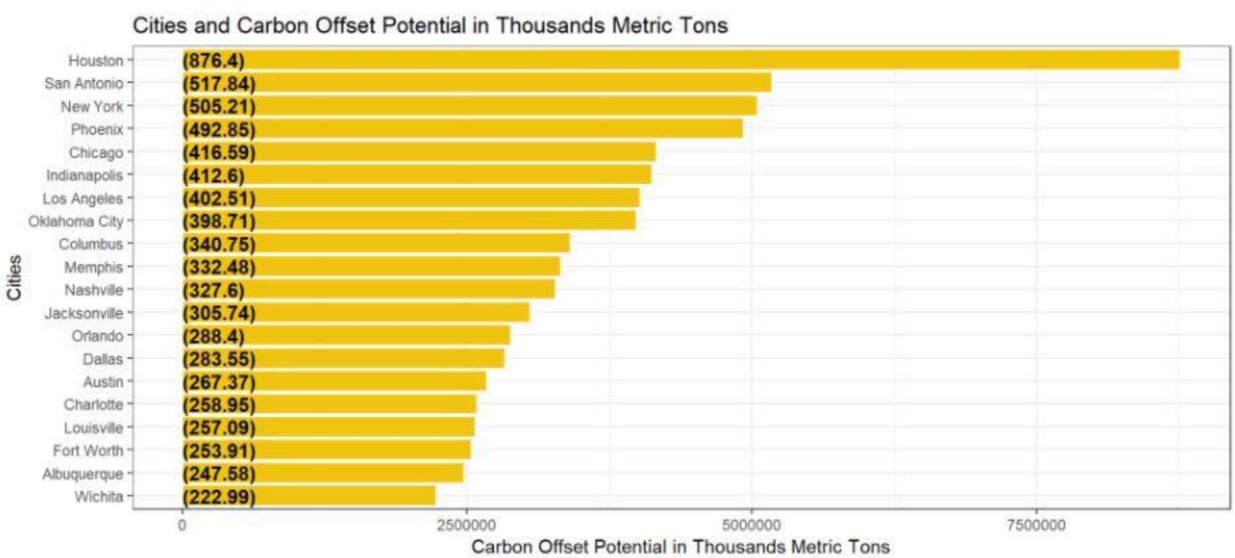
Future scope

Analysis on Sunroof data can be done on by taking state-wise data, city-wise data and following a similar procedure to effectively find whether yearly sunlight and panels installation have a significant effect on reducing carbon offset rate. Like we have identified the potential of USA in reducing carbon offset below, same can be done on a global scale.

Carbon Offset Potential of Cities

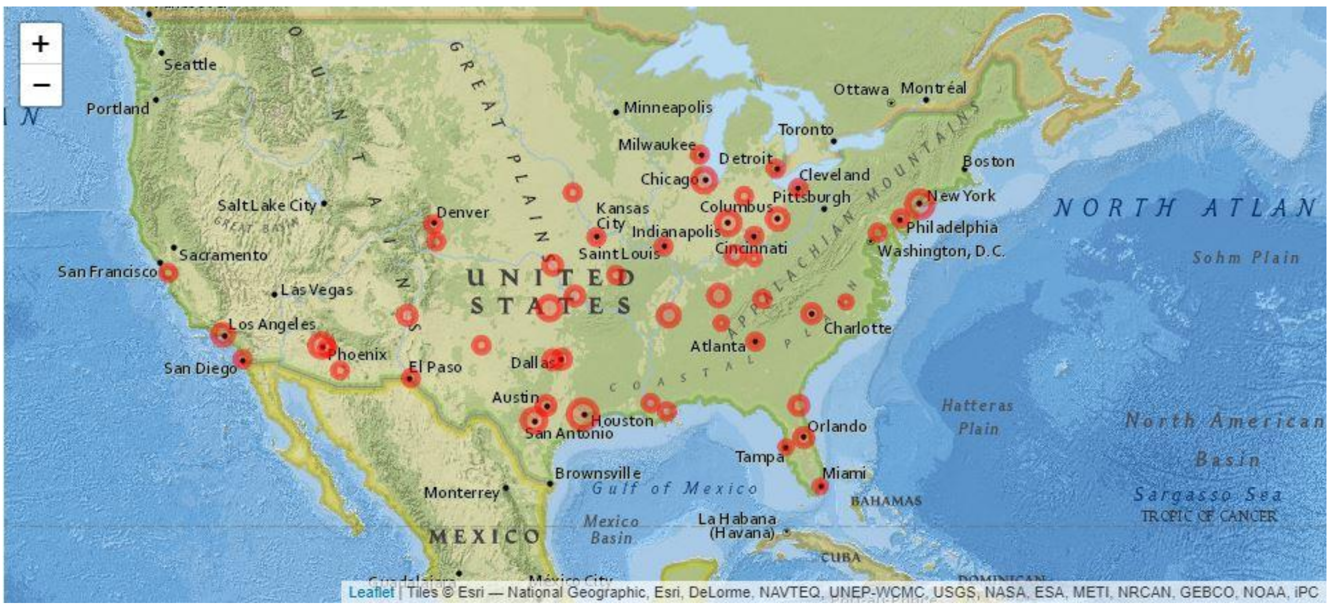
We examine the Carbon Offset Potential of the Cities and plot the Top Twenty in a bar plot.

```
CarbonOffsetCities = CityData %>%
  arrange(desc(carbon_offset_metric_tons)) %>%
  head(20)

CarbonOffsetCities %>%
  mutate(region_name = reorder(region_name,carbon_offset_metric_tons)) %>%
 ggplot(aes(x = region_name,y = carbon_offset_metric_tons)) +
  geom_bar(stat='identity',colour="white", fill =fillColor2) +
  geom_text(aes(x = region_name, y = 1,
                label = paste0("(",round( (carbon_offset_metric_tons/10e+3),2),")",sep="")),
            hjust=0, vjust=.5, size = 4, colour = 'black',
            fontface = 'bold') +
  labs(x = 'Cities', y = 'Carbon Offset Potential in Thousands Metric Tons',
       title = 'Cities and Carbon Offset Potential in Thousands Metric Tons') +
  coord_flip() +
  theme_bw()
```



We plot the map of Top Fifty Cities which have high Carbon Offset Potential.

```
CarbonOffsetCities = CityData %>%
  arrange(desc(carbon_offset_metric_tons)) %>%
  head(50)

center_lon = median(CarbonOffsetCities$lng_avg)
center_lat = median(CarbonOffsetCities$lat_avg)

leaflet(CarbonOffsetCities) %>% addProviderTiles("Esri.NatGeoWorldMap") %>%
  addCircles(lng = ~lng_avg, lat = ~lat_avg,radius = ~sqrt(carbon_offset_metric_tons)*30,
             color = c("red"))  %>%
  # controls
  setView(lng=center_lon, lat=center_lat, zoom=4)
```



Conclusion

- The mean Carbon offset rate for United States of America is 370 in thousands of metric tons.
- Project Sun Roof will have a major effect on carbon offset.
- Potential of this project is huge in USA itself, on a global scale it will solve energy crisis and help with sustainable living.