Northeastern University
College of Professional Studies

Week 6 Signature Assignment

Submitted By: Rama Niteesh Mallavarapu

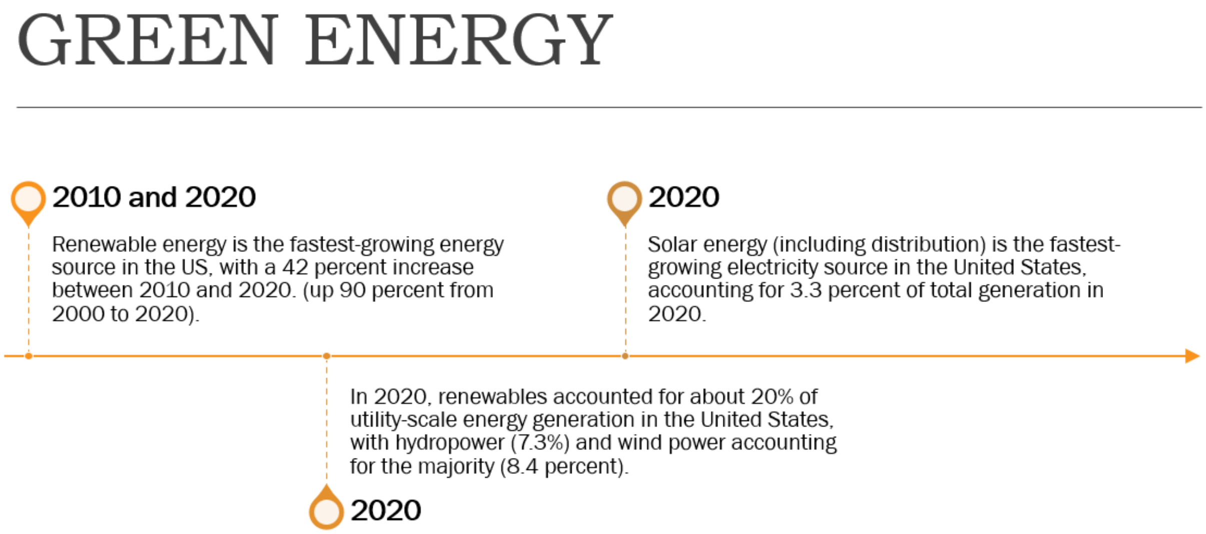
Course Code: ALY 6000 – Introduction to Analytics

Instructor: Kristen Drobnis

Submitted on: February 19th, 2022

Green Energy

Introduction:



Company: SunWorks.Inc

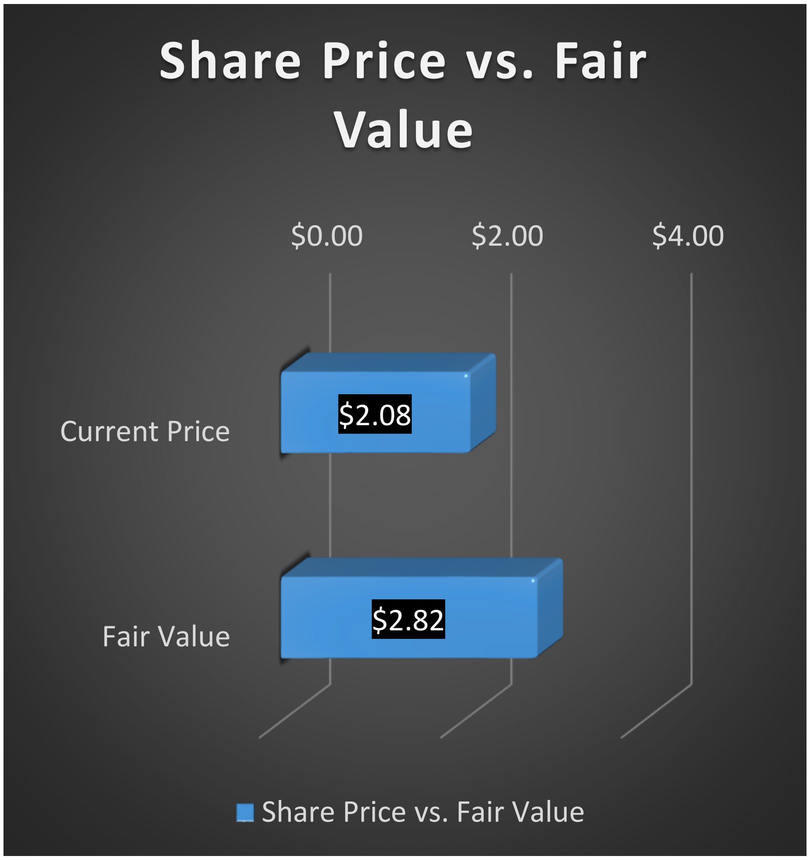
Sunworks, Inc. (NASDAQ: SUNW) is a premier provider of high-performance solar power systems. We are committed to quality business practices that exceed industry standards and uphold our ideals of ethics and safety.

We strive to consistently deliver high quality, performance-oriented solutions for customers in a wide range of industries including agricultural, commercial, and industrial, federal, public works, and residential.

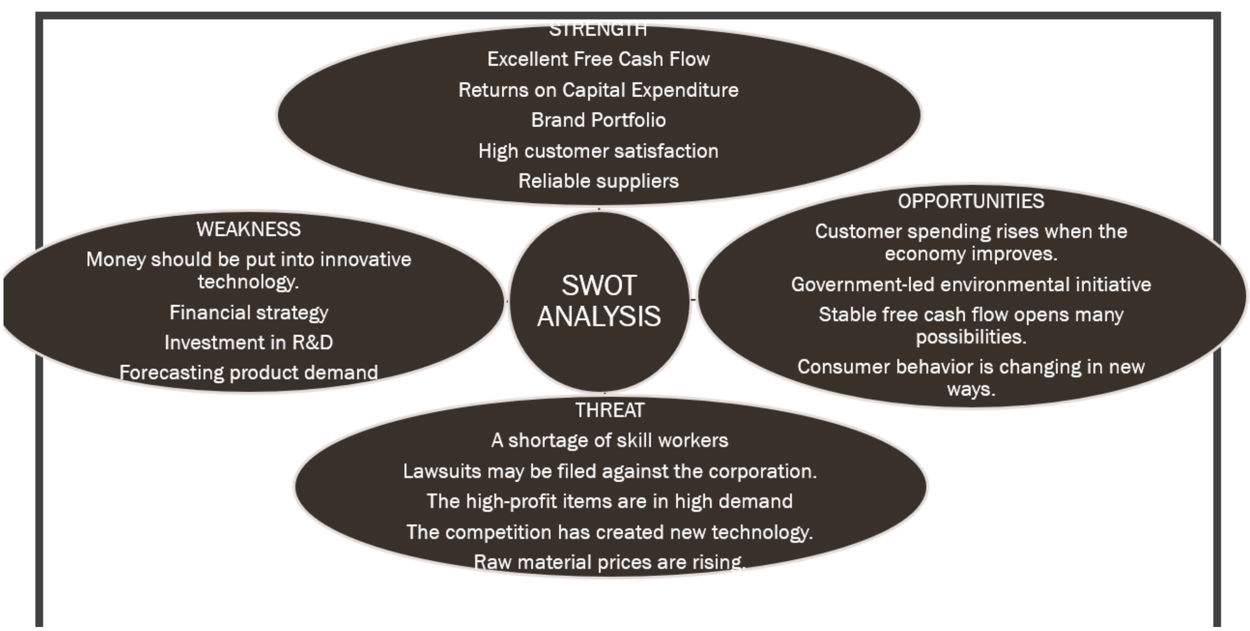
Sunworks, Inc. (2021, December 23). About SunWorks | Solar Solutions. SUNWORKS | SOLAR SOLUTIONS. https://sunworksusa.com/about-sunworks/

Financial Analysis:

Stock Price: Sunworks.Inc is currently trading at $2.08 on American Stock Exchange. Furthermore, it is trading at a discount of 26.3 percent to our estimate of its fair value. Explains it is a perfect buy opportunity.



Company with respect to industry:



(SWOT Analysis is discussed in the observation section)

Part 1: Visualizing trends in Solar Energy for SunWorks.Inc

Using Jupyter Notebook, I will investigate the deployment of solar energy in the United States (Python 3). Stanford University's Deep Solar Project provided the data. I'll illustrate the data by charting these variables against one other and against data from solar panels. Finally, I will analyze the charts and look for patterns and connections in solar energy deployment.

1) Importing Libraries

```
In [10]: import os
         import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         import matplotlib.ticker as ticker
         import seaborn as sns
         import scipy
         from mpl_toolkits.mplot3d import Axes3D
         from sklearn.preprocessing import StandardScaler
```

```
In [11]: # libraries for maps
         from mpl_toolkits.basemap import Basemap as Basemap
         from matplotlib.colors import rgb2hex
         from matplotlib.patches import Polygon
```

```
In [14]: from mpl_toolkits.mplot3d import Axes3D
```

```
In [15]: from matplotlib.colors import rgb2hex
```

```
In [16]: from matplotlib.patches import Polygon
```

Hence, the libraries are imported.

2) Loading Data into Jupyter Notebook

df = pd.read_excel('ALY6000Dataset.xlsx', 'Sheet1')
df

```
In [29]: df = pd.read_excel('ALY6000Dataset.xlsx', 'Sheet1')
         df
```

Out[29]:

	Unnamed: 0	tile_count	solar_system_count	total_panel_area	fips	average_household_income	county	education_bachelor	education_college	edi
0	0	0	0	0.000000	27145011200	70352.789869	Stearns County	569	1690	
1	1	25	21	1133.436461	27145011301	61727.085202	Stearns County	674	1434	
2	2	3	3	64.505776	27145011302	71496.886583	Stearns County	854	1459	
3	3	0	0	0.000000	27145011304	86840.152755	Stearns County	640	1116	
4	4	5	5	164.583303	27145011400	89135.315597	Stearns County	654	1314	
...	...	...	...	...	...	...	...	...	...	
1194	1194	25	22	604.961766	6037104403	53996.677215	Los Angeles County	201	241	
1195	1195	35	30	678.720556	6037104500	56626.666667	Los Angeles County	48	290	
1196	1196	52	21	5369.330168	6037302401	48904.564706	Los Angeles County	736	1266	
1197	1197	42	37	948.992239	6037104610	50372.510519	Los Angeles County	64	346	
1198	1198	32	26	651.128609	6037104620	56400.277392	Los Angeles County	50	235	

df.head()

```
In [30]: df.head()
```

Out[30]:

	Unnamed: 0	tile_count	solar_system_count	total_panel_area	fips	average_household_income	county	education_bachelor	education_college	educat
0	0	0	0	0.000000	27145011200	70352.789869	Stearns County	569	1690	
1	1	25	21	1133.436461	27145011301	61727.085202	Stearns County	674	1434	
2	2	3	3	64.505776	27145011302	71496.886583	Stearns County	854	1459	
3	3	0	0	0.000000	27145011304	86840.152755	Stearns County	640	1116	
4	4	5	5	164.583303	27145011400	89135.315597	Stearns County	654	1314	

5 rows × 169 columns

df.info ()

```
In [31]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1199 entries, 0 to 1198
Columns: 169 entries, Unnamed: 0 to avg_electricity_retail_rate
dtypes: bool(2), float64(112), int64(53), object(2)
memory usage: 1.5+ MB
```

df.columns

```
In [64]: df.columns
```

```
Out[64]: Index(['Unnamed: 0', 'tile_count', 'solar_system_count', 'total_panel_area',
               'fips', 'average_household_income', 'county', 'education_bachelor',
               'education_college', 'education_doctoral',
               ...
               'incentive_count_nonresidential', 'incentive_residential_state_level',
               'incentive_nonresidential_state_level', 'net_metering', 'feedin_tariff',
               'cooperate_tax', 'property_tax', 'sales_tax', 'rebate',
               'avg_electricity_retail_rate'],
              dtype='object', length=169)
```

df.describe()

```
In [32]: df.describe()
```

Out[32]:

	Unnamed: 0	tile_count	solar_system_count	total_panel_area	fips	average_household_income	education_bachelor	education_college	educati
count	1199.000000	1199.00000	1199.000000	1199.000000	1.199000e+03	1169.000000	1199.000000	1199.000000	
mean	599.000000	49.92744	34.729775	2122.159651	8.526792e+09	77153.723010	567.783153	739.024187	
std	346.265794	76.20852	44.776653	7854.080042	8.004043e+09	42461.695396	449.742896	404.134771	
min	0.000000	0.00000	0.000000	0.000000	6.037101e+09	9040.000000	0.000000	0.000000	
25%	299.500000	11.00000	8.000000	282.016442	6.037212e+09	48336.021505	213.500000	447.000000	
50%	599.000000	25.00000	18.000000	707.329347	6.037461e+09	65945.544554	460.000000	682.000000	
75%	898.500000	62.00000	44.000000	1933.160204	6.037601e+09	94125.000000	821.000000	972.500000	
max	1198.000000	1463.00000	324.000000	216732.195619	3.607102e+10	414706.238185	2916.000000	2734.000000	

8 rows × 165 columns

Data Cleaning:

Removing columns from the dataset that are no longer needed. Majority of the columns in this dataset are not dropped because they can act as a dependent variable. Dropping dependent variables has an impact on data analysis and causes visualization problems.

x=df.drop(columns=['fips','education_bachelor','education_college','education_doctoral','education_high_school_graduate','education_less_than_high_school','education_master','education_population','education_professional_school','employed'])

x=df.drop(columns=['poverty_family_below_poverty_level','race_asian','race_black_africa','race_indian_alaska','race_islander','race_other','race_two_more','race_white','race_white_rate','race_black_africa_rate','race_indian_alaska_rate','race_asian_rate','race_other_rate'])

In [70]:
```
x
```

Out[70]:

	Unnamed: 0	tile_count	solar_system_count	total_panel_area	fips	average_household_income	county	education_bachelor	education_college	edi
0	0	0	0	0.000000	27145011200	70352.789869	Stearns County	569	1690	
1	1	25	21	1133.436461	27145011301	61727.085202	Stearns County	674	1434	
2	2	3	3	64.505776	27145011302	71496.886583	Stearns County	854	1459	
3	3	0	0	0.000000	27145011304	86840.152755	Stearns County	640	1116	
4	4	5	5	164.583303	27145011400	89135.315597	Stearns County	654	1314	
...	...	...	...	...	...	...	...	...	...	
1194	1194	25	22	604.961766	6037104403	53996.677215	Los Angeles County	201	241	
1195	1195	35	30	678.720556	6037104500	56626.666667	Los Angeles County	48	290	
1196	1196	52	21	5369.330168	6037302401	48904.564706	Los Angeles County	736	1266	
1197	1197	42	37	948.992239	6037104610	50372.510519	Los Angeles County	64	346	
1198	1198	32	26	651.128609	6037104620	56400.277392	Los Angeles County	50	235	

x.fillna(x.mean())

Replace all NA values in each data column with the mean of the column

In [36]:
```
x.fillna(x.mean())
```

```
C:\Users\NITEES~1\AppData\Local\Temp/ipykernel_28448/2029865756.py:1: FutureWarning: Dropping of nuisance columns in DataFrame
reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns
before calling the reduction.
  x.fillna(x.mean())
```

Out[36]:

	Unnamed: 0	tile_count	solar_system_count	total_panel_area	average_household_income	county	gini_index	heating_fuel_coal_coke	heating_fuel_electrici
0	0	0	0	0.000000	70352.789869	Stearns County	0.3490	200	44
1	1	25	21	1133.436461	61727.085202	Stearns County	0.4074	20	37
2	2	3	3	64.505776	71496.886583	Stearns County	0.3926	69	44
3	3	0	0	0.000000	86840.152755	Stearns County	0.3949	188	44
4	4	5	5	164.583303	89135.315597	Stearns County	0.4463	96	49
...	...	...	...	...	...	...	...	...	
1194	1194	25	22	604.961766	53996.677215	Los Angeles County	0.4161	12	7
1195	1195	35	30	678.720556	56626.666667	Los Angeles County	0.3553	7	16
1196	1196	52	21	5369.330168	48904.564706	Los Angeles County	0.4377	0	67
1197	1197	42	37	948.992239	50372.510519	Los Angeles County	0.3948	0	16
1198	1198	32	26	651.128609	56400.277392	Los Angeles County	0.3633	0	15

x.head()

```
In [37]: x.head()
```

Out[37]:

	Unnamed: 0	tile_count	solar_system_count	total_panel_area	average_household_income	county	gini_index	heating_fuel_coal_coke	heating_fuel_electricity
0	0	0	0	0.000000	70352.789869	Stearns County	0.3490	200	448
1	1	25	21	1133.436461	61727.085202	Stearns County	0.4074	20	379
2	2	3	3	64.505776	71496.886583	Stearns County	0.3926	69	440
3	3	0	0	0.000000	86840.152755	Stearns County	0.3949	188	442
4	4	5	5	164.583303	89135.315597	Stearns County	0.4463	96	497

5 rows × 159 columns

3) Data Visualization

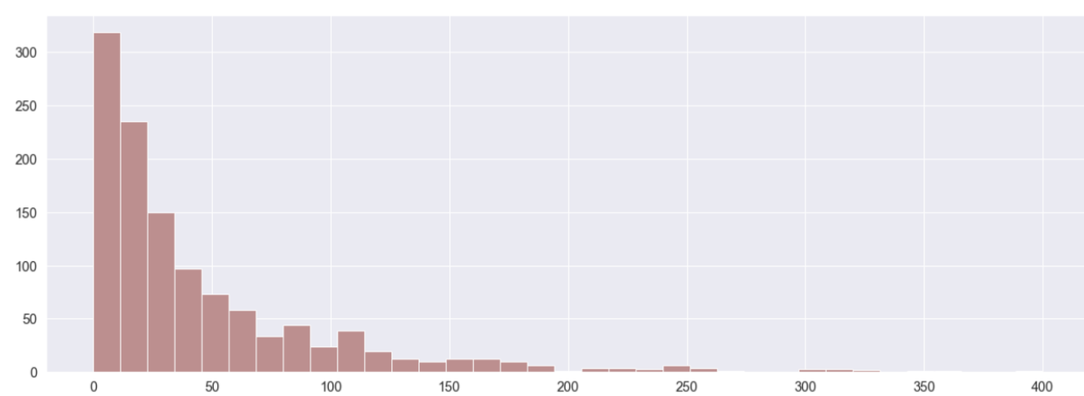
Using Matplotlib to create histograms to show the frequency of solar panel counts.

*) Histogram : Solar Panel Counts

print("Histogram : Solar Panel Counts")
fig, ax = plt.subplots(figsize=(20, 7))
tile_count = x['tile_count']
plt.hist(tile_count, 35, range=[0, 400], facecolor='rosybrown', align='mid')
plt.show()

```
In [72]: print("Histogram : Solar Panel Counts")
         fig, ax = plt.subplots(figsize=(20, 7))
         tile_count = x['tile_count']
         plt.hist(tile_count, 35, range=[0, 400], facecolor='rosybrown', align='mid')
         plt.show()
```

```
Histogram : Solar Panel Counts
```



Most of the locations have no solar panels, as seen by this histogram of solar panel counts. I generated a second histogram of solar panel counts for places with at least 20 solar panels but fewer than 100 to see a more statistically rich chunk of this data.

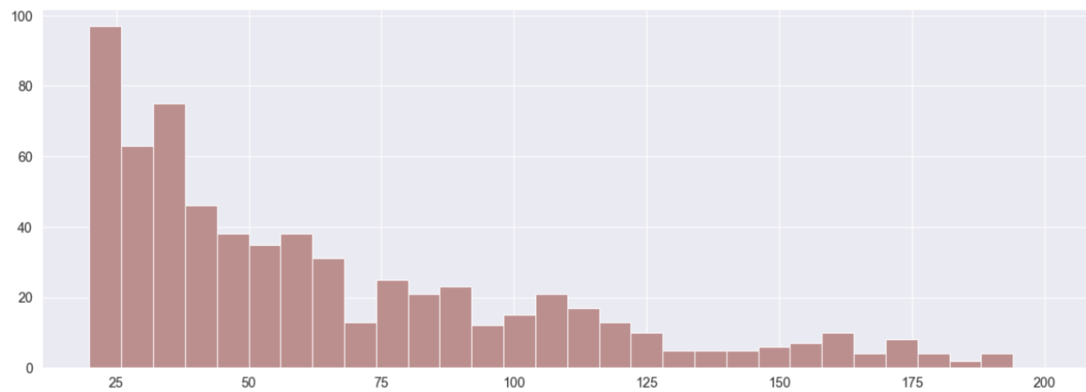
*) Histogram : Solar Panel Counts(Limits: 20 to 200)

print("Histogram : Solar Panel Counts(Limits: 20 to 200)")
tile_count = x['tile_count']

fig, ax = plt.subplots(figsize=(20, 7))
plt.hist(tile_count, 30, range=[20, 200], facecolor='rosybrown', align='mid')
plt.show()

```
In [71]: print("Histogram : Solar Panel Counts(Limits: 20 to 200)")
         tile_count = x['tile_count']
         fig, ax = plt.subplots(figsize=(20, 7))
         plt.hist(tile_count, 30, range=[20, 200], facecolor='rosybrown', align='mid')
         plt.show()
```

```
Histogram : Solar Panel Counts(Limits: 20 to 200)
```



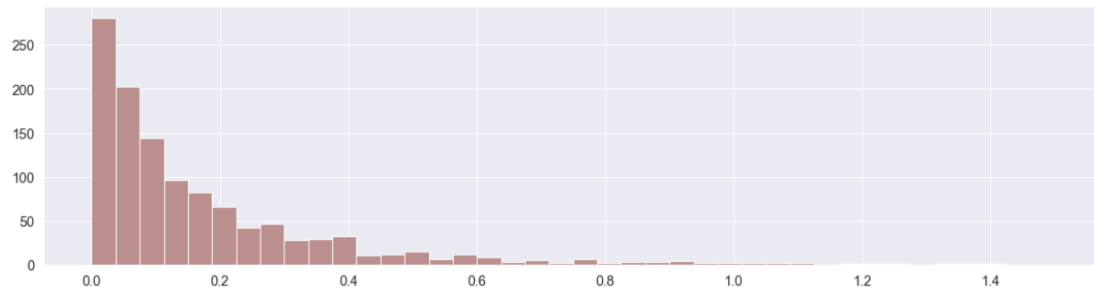
The trend is even clearer when set limits from 20 to 200. Most residential and non-residential houses have less than 60 solar panels. The frequency count is clearly visible before and after 60. Right after the count of 60, the overall number has been decreasing but spike was observed at 80 and 110. The prospect for future solar panel employment appears to be increasing at a good rate. From a business perspective, solar panel contracts of 60 or less than that are highly profitable.

*) Histogram : Solar Panel Area per Capita

print("Histogram : Solar Panel Area per Capita")
area_capita = x['solar_panel_area_per_capita']
fig, ax = plt.subplots(figsize=(20, 5))
plt.hist(area_capita, 40, range=[0, 1.5], facecolor='rosybrown', align='mid')
plt.show()

```
In [73]: print("Histogram : Solar Panel Area per Capita")
         area_capita = x['solar_panel_area_per_capita']
         fig, ax = plt.subplots(figsize=(20, 5))
         plt.hist(area_capita, 40, range=[0, 1.5], facecolor='rosybrown', align='mid')
         plt.show()
```

```
Histogram : Solar Panel Area per Capita
```



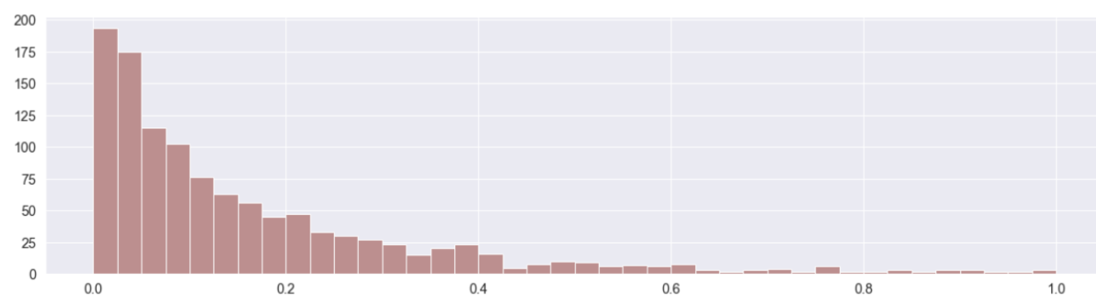
Another set of histograms, this time spanning from 0 to 1, depicts the frequency of solar panel area per population. Most of the data points are zero, as predicted from the first histogram.

Histogram : Solar Panel Area per Capita

print("Histogram : Solar Panel Area per Capita")
area_capita = x['solar_panel_area_per_capita']
fig, ax = plt.subplots(figsize=(20, 5))
plt.hist(area_capita, 40, range=[0, 1.0], facecolor='rosybrown', align='mid')
plt.show()

```
In [74]: print("Histogram : Solar Panel Area per Capita")
         area_capita = x['solar_panel_area_per_capita']
         fig, ax = plt.subplots(figsize=(20, 5))
         plt.hist(area_capita, 40, range=[0, 1.0], facecolor='rosybrown', align='mid')
         plt.show()
```

```
Histogram : Solar Panel Area per Capita
```



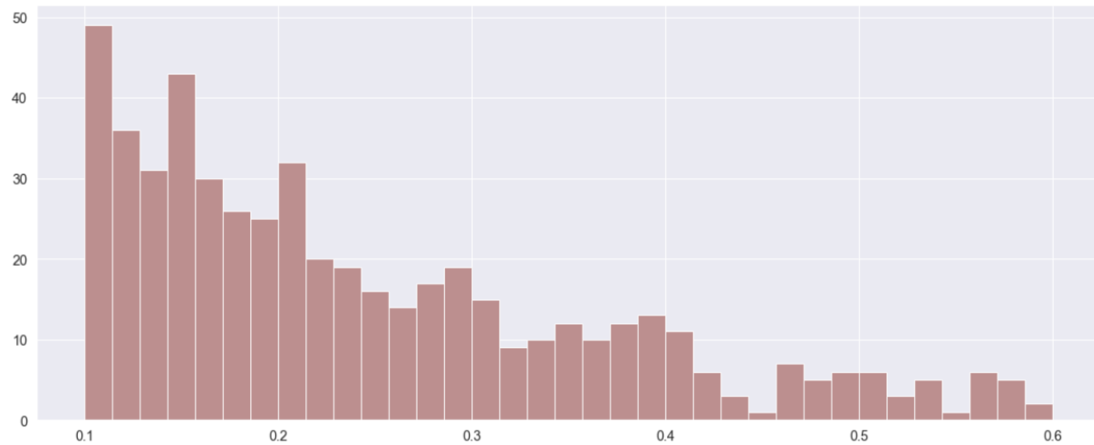
Majority of the data points are 0. Again, the number has been gradually decreasing from 0 to 1. Percapita has been bought into picture to ensure that the data is not biased by population size or density in an area. From business perspective, it is always beneficial for orders to be booked for multiple families residing in the same building.

Histogram : Solar Panel Area per Capita (Limiting Range)

print("Histogram : Solar Panel Area per Capita (Limiting Range)")
area_capita = x['solar_panel_area_per_capita']
fig, ax = plt.subplots(figsize=(20, 8))

plt.hist(area_capita, 35, range=[0.1, 0.6], facecolor='rosybrown', align='mid')
plt.show()

```
In [76]: print("Histogram : Solar Panel Area per Capita ( Limiting Range)")
         area_capita = x['solar_panel_area_per_capita']
         fig, ax = plt.subplots(figsize=(20, 8))
         plt.hist(area_capita, 35, range=[0.1, 0.6], facecolor='rosybrown', align='mid')
         plt.show()

         Histogram : Solar Panel Area per Capita ( Limiting Range)
```



A bar plot created with NumPy and matplotlib was the second technique of visualization

*) BarPlot Analysis

```
print("BarPlot Analysis")
print()


total_tile = int(x['tile_count'].sum())
print("Total Number of Solar Panels:", total_tile)
total_system = int(x['solar_system_count'].sum())
print("Total Number of Solar Panel Systems:", total_system)

avg_tile_count = total_tile/72495
avg_tile_per_system = total_tile/ total_system
print("Average Number of Solar Panels:", round(avg_tile_count))
print("Average Number of Panels per Solar Panel System:", avg_tile_per_system)

print()
total_rtile = int(x['tile_count_residential'].sum())
print("Total Number of Solar Panels for Residential Purposes:", total_rtile)
total_nrtile = int(x['tile_count_nonresidential'].sum())
print("Total Number of Solar Panels for Non-Residential Purposes*:", total_nrtile)
print()
total_rsystem = int(x['solar_system_count_residential'].sum())
print("Total Number of Solar Systems for Residential Purposes:", total_rsystem)
total_nrsystem = int(x['solar_system_count_nonresidential'].sum())
```

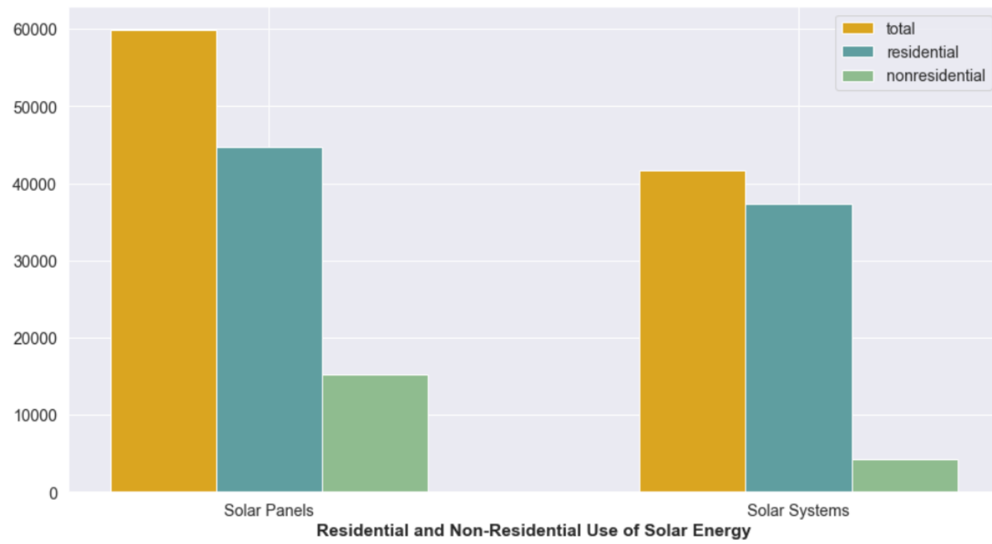
```
print("Total Number of Solar Panels for Non-Residential Purposes*:", total_nrsystem)
print()


barWidth = 0.2

barstotal = [total_tile, total_system]
barsres = [total_rtile, total_rsystem]
barsnonres = [total_nrtile, total_nrsystem]


r1 = np.arange(len(barstotal))
r2 = [x + barWidth for x in r1]
r3 = [x + barWidth for x in r2]

fig, ax = plt.subplots(figsize=(15, 8))


plt.bar(r1, barstotal, color='goldenrod', width=barWidth, edgecolor='white', label='total')
plt.bar(r2, barsres, color='cadetblue', width=barWidth, edgecolor='white', label='residential')
plt.bar(r3, barsnonres, color='darkseagreen', width=barWidth, edgecolor='white',
label='nonresidential')


plt.xlabel('Residential and Non-Residential Use of Solar Energy', fontweight='bold')
plt.xticks([r + barWidth for r in range(len(barstotal))], ["Solar Panels", "Solar Systems"])


plt.legend()
plt.show()
```

```
In [54]: print("BarPlot Analysis")
print()


total_tile = int(x['tile_count'].sum())
print("Total Number of Solar Panels:", total_tile)
total_system = int(x['solar_system_count'].sum())
print("Total Number of Solar Panel Systems:", total_system)

avg_tile_count = total_tile/72495
avg_tile_per_system = total_tile/ total_system
print("Average Number of Solar Panels:", round(avg_tile_count))
print("Average Number of Panels per Solar Panel System:", avg_tile_per_system)

print()
total_rtile = int(x['tile_count_residential'].sum())
print("Total Number of Solar Panels for Residential Purposes:", total_rtile)
total_nrtile = int(x['tile_count_nonresidential'].sum())
print("Total Number of Solar Panels for Non-Residential Purposes*:", total_nrtile)
print()
total_rsystem = int(x['solar_system_count_residential'].sum())
print("Total Number of Solar Systems for Residential Purposes:", total_rsystem)
total_nrsystem = int(x['solar_system_count_nonresidential'].sum())
print("Total Number of Solar Panels for Non-Residential Purposes*:", total_nrsystem)
print()


barWidth = 0.2

barstotal = [total_tile, total_system]
barsres = [total_rtile, total_rsystem]
barsnonres = [total_nrtile, total_nrsystem]


r1 = np.arange(len(barstotal))
r2 = [x + barWidth for x in r1]
r3 = [x + barWidth for x in r2]

fig, ax = plt.subplots(figsize=(15, 8))
```


```
barstotal = [total_tile, total_system]
barsres = [total_rtile, total_rsystem]
barsnonres = [total_nrtile, total_nrsystem]


r1 = np.arange(len(barstotal))
r2 = [x + barWidth for x in r1]
r3 = [x + barWidth for x in r2]

fig, ax = plt.subplots(figsize=(15, 8))
```


```
plt.bar(r1, barstotal, color='goldenrod', width=barWidth, edgecolor='white', label='total')
plt.bar(r2, barsres, color='cadetblue', width=barWidth, edgecolor='white', label='residential')
plt.bar(r3, barsnonres, color='darkseagreen', width=barWidth, edgecolor='white', label='nonresidential')


plt.xlabel('Residential and Non-Residential Use of Solar Energy', fontweight='bold')
plt.xticks([r + barWidth for r in range(len(barstotal))], ["Solar Panels", "Solar Systems"])


plt.legend()
plt.show()
```

```
BarPlot Analysis

Total Number of Solar Panels: 59863
Total Number of Solar Panel Systems: 41641
Average Number of Solar Panels: 1
Average Number of Panels per Solar Panel System: 1.4375975600970197

Total Number of Solar Panels for Residential Purposes: 44654
Total Number of Solar Panels for Non-Residential Purposes*: 15209

Total Number of Solar Systems for Residential Purposes: 37370
Total Number of Solar Panels for Non-Residential Purposes*: 4271
```



BarPlot Analysis

Total Number of Solar Panels: 59863
Total Number of Solar Panel Systems: 41641
Average Number of Solar Panels: 1
Average Number of Panels per Solar Panel System: 1.4375975600970197

Total Number of Solar Panels for Residential Purposes: 44654
Total Number of Solar Panels for Non-Residential Purposes*: 15209

Total Number of Solar Systems for Residential Purposes: 37370
Total Number of Solar Systems for Non-Residential Purposes*: 4271

The bar plot compares number of solar panels used in residential and non-residential areas. There is an average of one solar panel and about 1.5 solar panels per solar panel system in a certain location in United States. There are way more solar panels deployed residentially than non-residential purposes (i.e., commercial, industrial and transportation sectors combined).

Discussion:

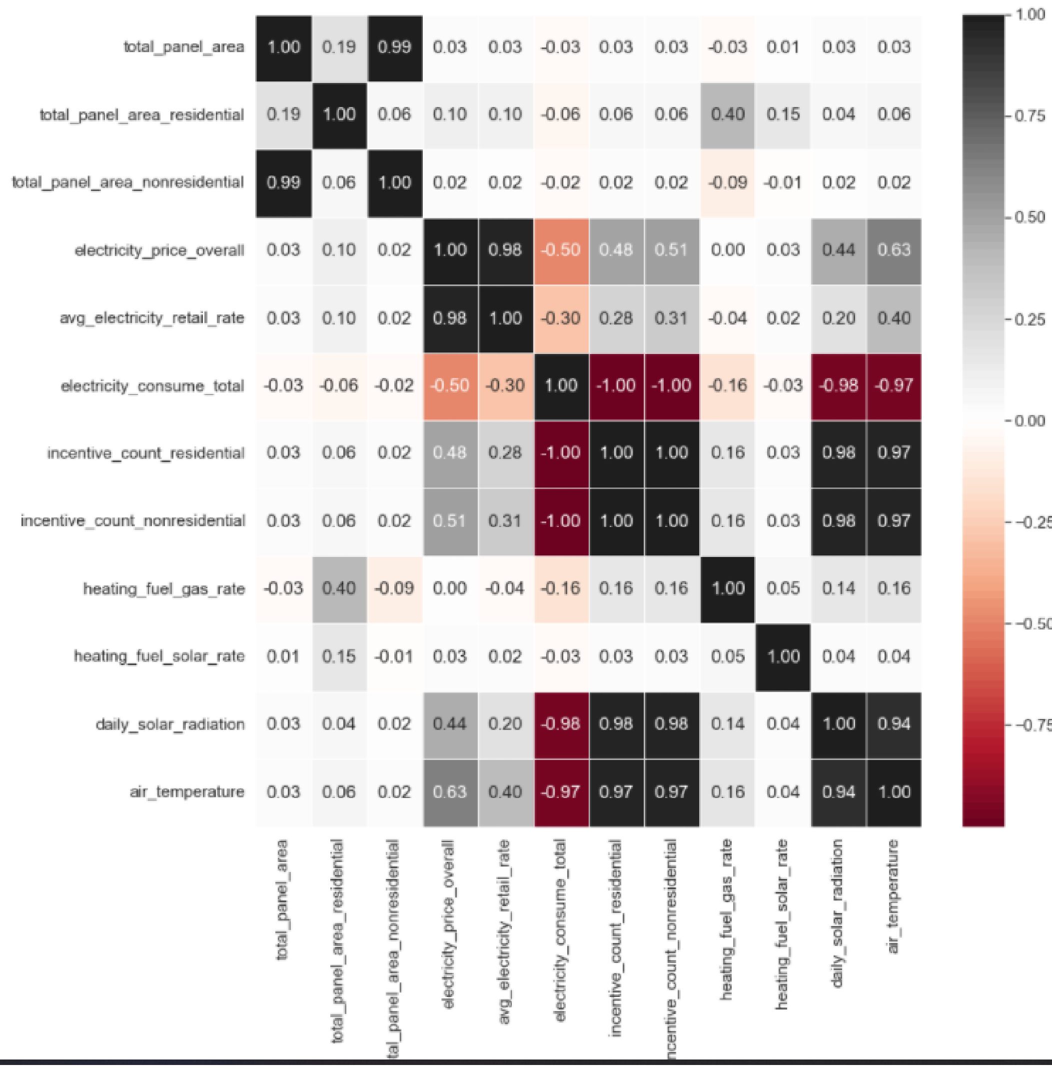
Interesting part is that there are fewer solar panel systems utilized for nonresidential uses than for residential, each nonresidential system has many more solar panels. more houses in the United States have solar panels than organizations and enterprises, although if these institutions have installed a solar panel system, it often includes a far larger number of panels than the ordinary family.

To further illustrate patterns and trends in data, we used correlation heat maps made using matplotlib and seaborn, as well as linear regression plots created with seaborn.

*) Heatmap visualization is used to understand the relationship between multiple attributes in the dataset.

```
energyenv_quant_data = ["total_panel_area", "total_panel_area_residential",
"total_panel_area_nonresidential","electricity_price_overall", "avg_electricity_retail_rate",
"electricity_consume_total", "incentive_count_residential", "incentive_count_nonresidential",
"heating_fuel_gas_rate", "heating_fuel_solar_rate","daily_solar_radiation",
"air_temperature",]

fig, axe = plt.subplots(figsize=(14,14))
sns.set_context("poster")

sns.set(font_scale=1.4)
corrmap = sns.color_palette("RdGy", 100)
sns.heatmap(x[energyenv_quant_data].corr(),annot=True, fmt='.2f',linewidths=1,cmap =
corrmap)
```

```
In [59]: energyenv_quant_data = ["total_panel_area", "total_panel_area_residential", "total_panel_area_nonresidential","electricity_price_

         fig, axe = plt.subplots(figsize=(14,14))
         sns.set_context("poster")

         sns.set(font_scale=1.4)
         corrmap = sns.color_palette("RdGy", 100)
         sns.heatmap(x[energyenv_quant_data].corr(),annot=True, fmt='.2f',linewidths=1,cmap = corrmap)
```



This heat map finds the co-relation factors between solar energy, environmental factors, and economic incentives. All these will provide an idea how soon solar energy will be implemented. In heat map, factors with strong positive correlation range from light silver to stone black. Neutral correlation ranges in white color and strong negative relation white to dark red.

Discussion:

There is a strong positive correlation between the daily solar radiation with the incentive count residential and non-residential (0.98). This is most likely due to the greater proclivity of local governments and businesses to offer incentives, and people and businesses are more likely to take advantage of these offers. Total Panel area non-residential is greater than residential, which shows that the solar panels are in more demand for non-residential than residential. Despite that, it is interesting to see that electricity price overall is vice versa, and analyst should make note of this. This can be further improved as controlling electricity prices are still in a developing phase for green energy projects. Talking more about electricity, there is also a relation between the electricity price, the average retail rate, and the incentive count.

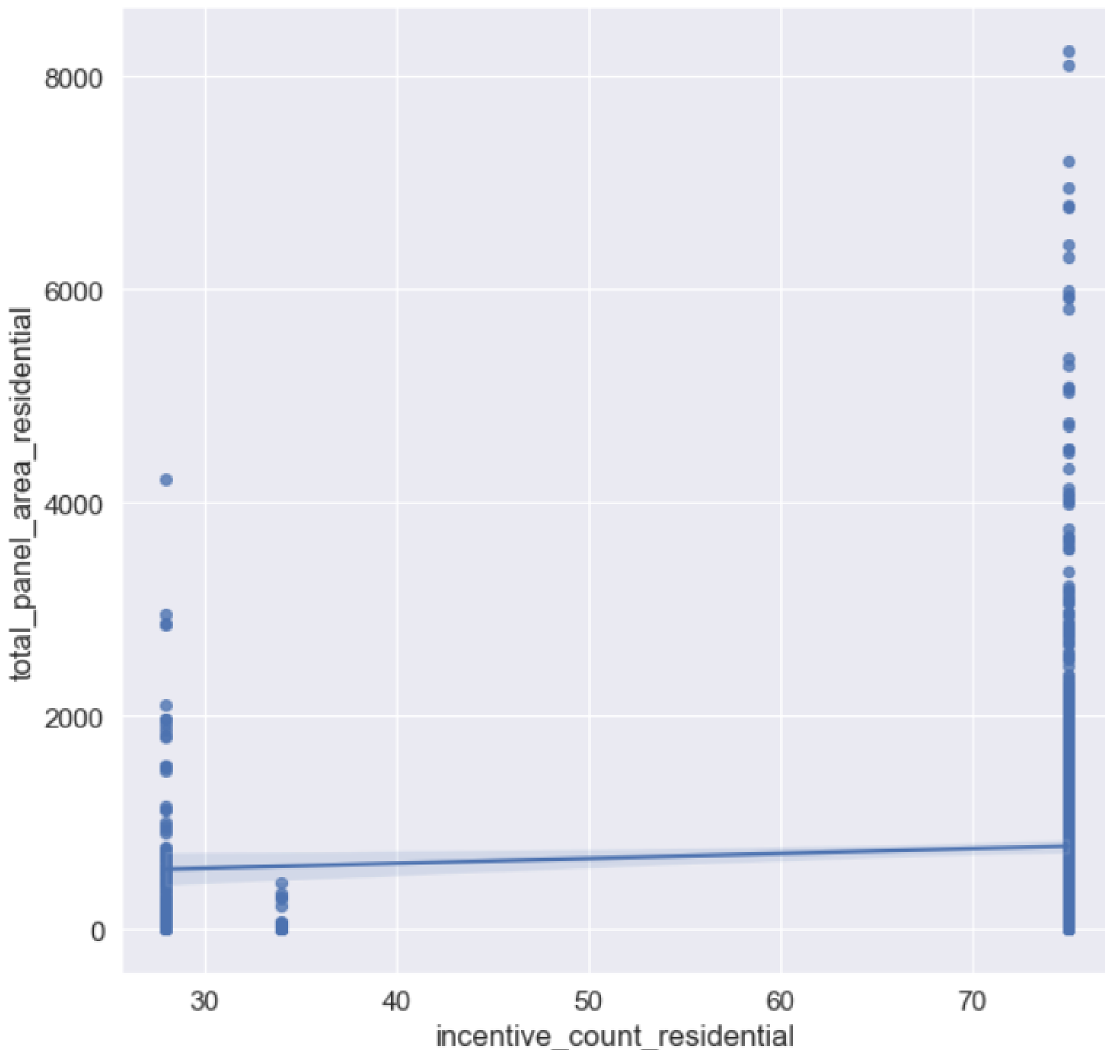
Residential incentive count was the environmental and economic element that was most strongly associated to solar energy. Overall, while performing this data visualization, zero values were kept because they are a realistic picture of how many locations do not have solar energy.

*)Relationship between total_panel_area_residential and incentive_count_residential

fig, axe = plt.subplots(figsize=(10,10))
sns.regplot(x=x["incentive_count_residential"], y=x["total_panel_area_residential"], fit_reg=True)

```
In [61]: fig, axe = plt.subplots(figsize=(10,10))
         sns.regplot(x=x["incentive_count_residential"], y=x["total_panel_area_residential"], fit_reg=True)
```

```
<AxesSubplot:xlabel='incentive_count_residential', ylabel='total_panel_area_residential'>
```

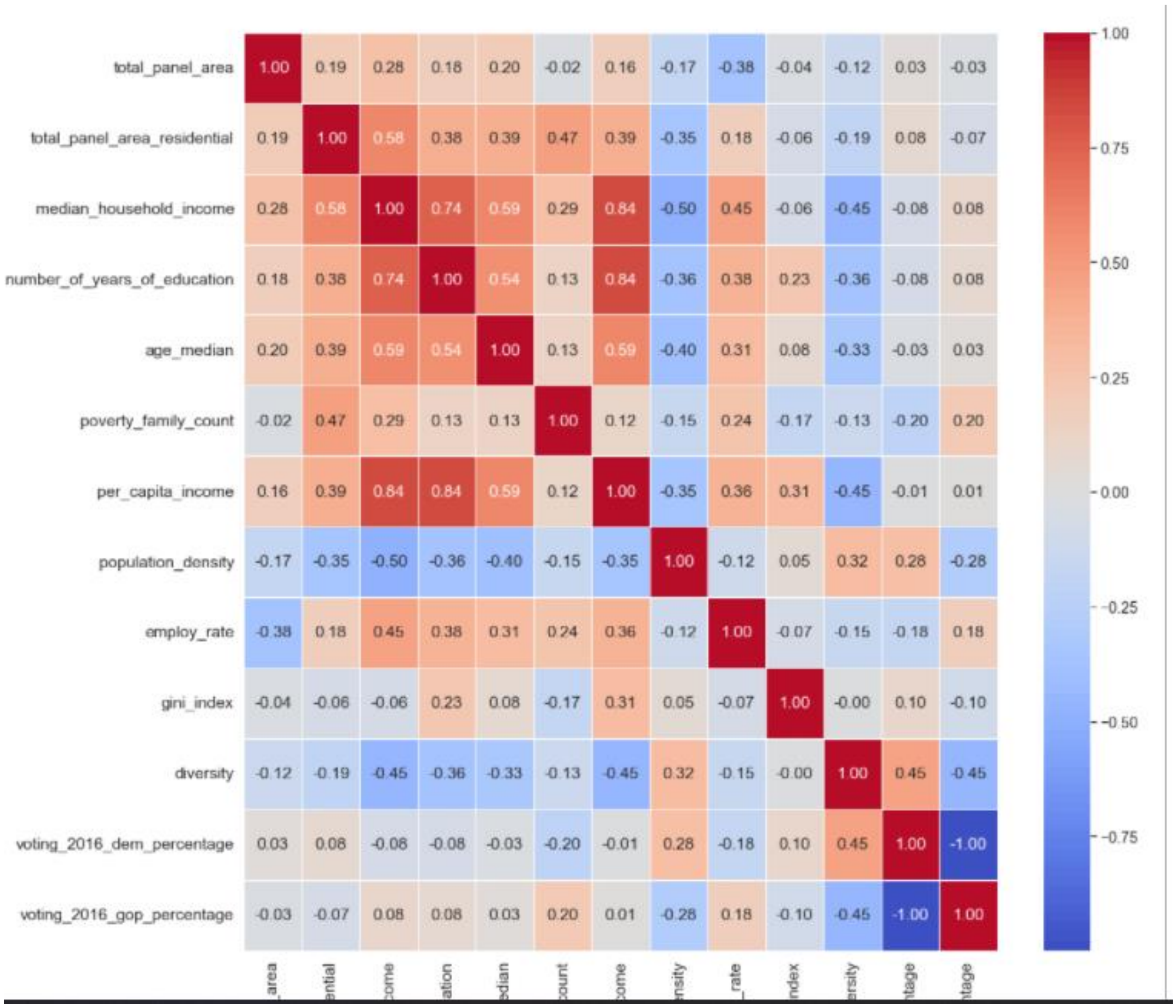


This visualization is done based on the previous heatmap where we discussed on the positive relation between total_panel_area_residential and incentive_count_residential. This is clearly clear in this plot, but there is still a positive link - individuals are more likely to put solar panels on their homes when they are offered greater financial incentives to do so.

*) Correlations between socioeconomic factors and the implementation of solar energy.

```
socioecon_quant_data = ["total_panel_area", "total_panel_area_residential",
"median_household_income","number_of_years_of_education", "age_median",
"poverty_family_count", "per_capita_income", "population_density", "employ_rate",
"gini_index", "diversity", "voting_2016_dem_percentage", 'voting_2016_gop_percentage']

fig, axe = plt.subplots(figsize=(16,16))
sns.set_context("poster")

sns.set(font_scale=1.3)
colmap2 = sns.color_palette("coolwarm", 100)
sns.heatmap(x[socioecon_quant_data].corr(),annot=True, fmt='.2f',linewidths=1,cmap =
colmap2)
```

```
socioecon_quant_data = ["total_panel_area", "total_panel_area_residential", "median_household_income","number_of_years_of_educati

fig, axe = plt.subplots(figsize=(16,16))
sns.set_context("poster")

sns.set(font_scale=1.3)
colmap2 = sns.color_palette("coolwarm", 100)
sns.heatmap(x[socioecon_quant_data].corr(),annot=True, fmt='.2f',linewidths=1,cmap = colmap2)
```



The socio-economic factors include household income, employment rate, years of education, GINI Index and voting behaviors in 2016 election. exterior variables such as a population's surroundings and receipt of economic incentives and internal factors can impact solar deployment.

Discussion:

Some factors which showed strong positive relationship are median household income and education. Furthermore, it showed a positive connection with employment rate and inverse connection with GINI Index. Interesting to see that, median household income was the most strongly connected with solar energy (total_panel_area_residential). We will investigate it closely below.

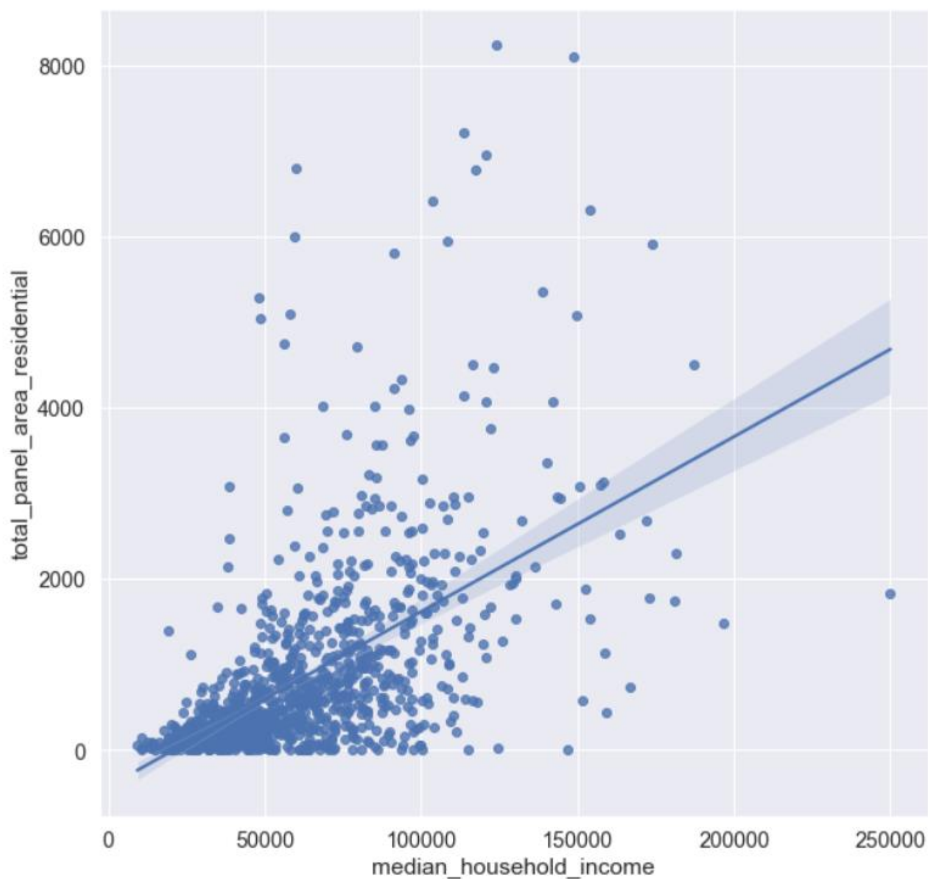
*) Linear Regression Plot of Median Household Income and Total Residential Panel Area

```
print("Linear Regression Plot of Median Household Income and Total Residential Panel Area")
fig, axe = plt.subplots(figsize=(10,10))
```

sns.regplot(x=x["median_household_income"], y=x["total_panel_area_residential"], fit_reg=True)

```
print("Linear Regression Plot of Median Household Income and Total Residential Panel Area")
fig, axe = plt.subplots(figsize=(10,10))
sns.regplot(x=x["median_household_income"], y=x["total_panel_area_residential"], fit_reg=True)
```

<AxesSubplot:xlabel='median_household_income', ylabel='total_panel_area_residential'>



The data is normally distributed exception is for the outliers. The median is 50000 to 100000. Even though they didn't continually increase after 100000, it is certain that there is a direct relationship between income and solar energy.

4) Conclusion

Members of middle- to upper-income households are more inclined to invest in solar energy if they have more discretionary money and can afford it. Including daily solar radiation and retail electricity prices, median family income and economic incentives remain the most important indicators of whether a certain area would adopt solar energy.

Part II: Creating new attributes on questions and data in Part I

The dataset which I used covered all range of attributes that is linked to the Solar Power and Energy. The main reason why did not drop all the attributes is because of existence of dependent variables.

However, I was able to draw insights which explained a lot about the company and their statistics. Based on my observations if I had a chance to include data from my side that would be “Humidity “, “Season “, “Pressure” and “Visibility “can also be a good attribute. These attributes can be added to existing dataset. These characteristics aid in understanding power output predictability, which is crucial for solar photovoltaics integration into traditional electrical grid systems. (Will discuss about this in the last section). The dataset can be further be worked on with machine learning algorithms to solve predictive analytics problems in green sector. A sample of small dataset involving these attributes looks something like this:

	Location	Date	Time	Latitude	Longitude	Altitude	Month	Hour	Season	Humidity	AmbientTemp	PolyPwr	Wind.Speed	Visibility	Pressure	Cloud.Ceiling
0	Camp Murray	20171203	1145	47.11	-122.57	84	12	11	Winter	81.71997	12.86919	2.42769	5	10.0	1010.6	722
1	Camp Murray	20171203	1315	47.11	-122.57	84	12	13	Winter	96.64917	9.66415	2.46273	0	10.0	1011.3	23
2	Camp Murray	20171203	1330	47.11	-122.57	84	12	13	Winter	93.61572	15.44983	4.46836	5	10.0	1011.6	32
3	Camp Murray	20171204	1230	47.11	-122.57	84	12	12	Winter	77.21558	10.36659	1.65364	5	2.0	1024.4	6
4	Camp Murray	20171204	1415	47.11	-122.57	84	12	14	Winter	54.80347	16.85471	6.57939	3	3.0	1023.7	9

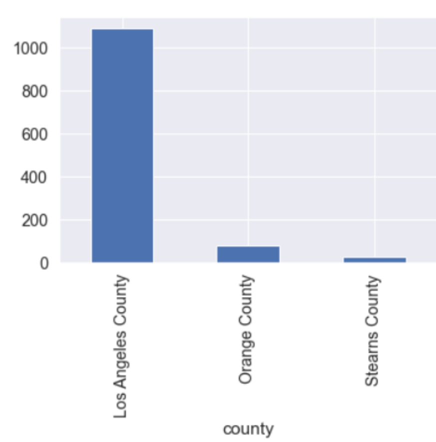
Olaoye, A. (2022, January 8). Predicting solar power output using machine learning techniques. Medium. https://towardsdatascience.com/predicting-solar-power-output-using-machine-learning-techniques-56e7959acb1f

*) For your data set, compute differences between appropriate variable values and create a new variable

I had run commands on “County “attribute to understand the number of people in each county. I would like to find out new interpretations on this like number of people who have high median of income in these counties and how much electricity is it costing.

x.groupby('county').size().plot.bar()

```
In [85]: x.groupby('county').size().plot.bar()
Out[85]: <AxesSubplot:xlabel='county'>
```

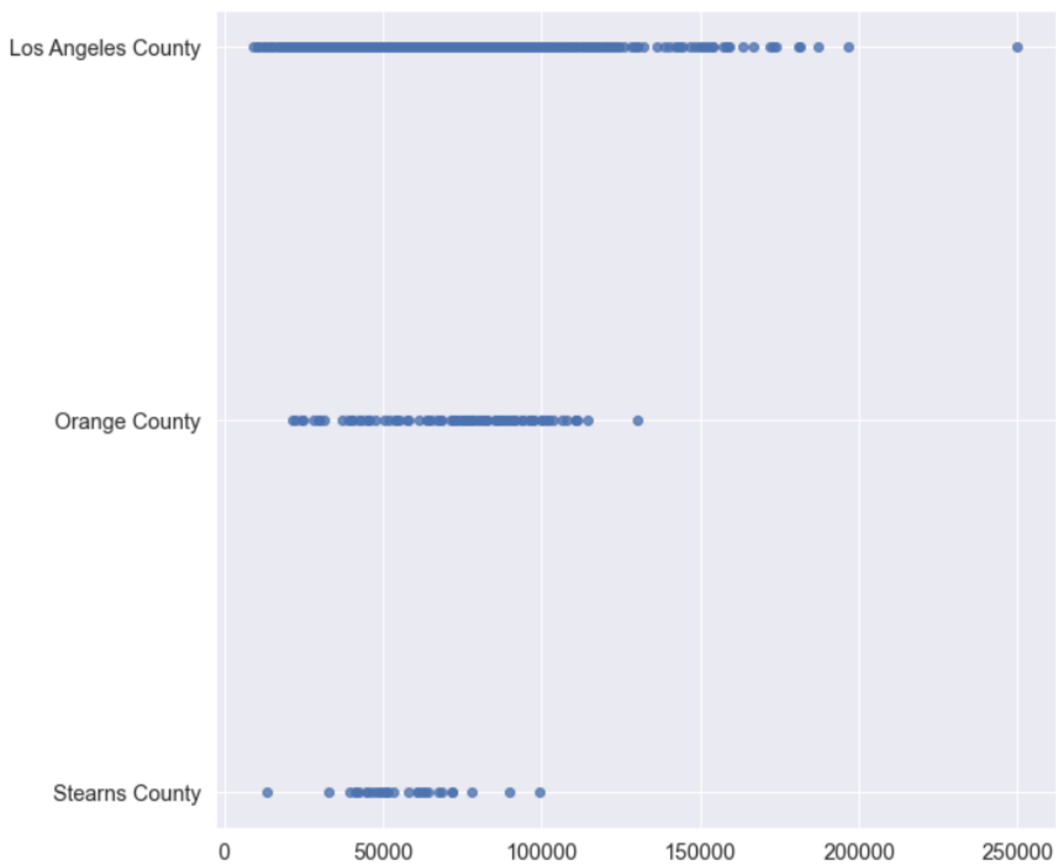



```
In [ ]:
```

It is evident that population of users in Los Angeles County are far greater than Orange County and Stearns County. It is interesting to note that, the population in Orange County is almost negligible it being the third most Populus city in California. This can be the reason that it is not an ideal location for implementing solar energy. Many factors come into picture for this for example environmental factors, cost of living, average mean income etc. We try to find out new interpretations on this by sorting it down on basis on mean income.

print("Median vs County")
fig, axe = plt.subplots(figsize=(10,10))
sns.regplot(x=x["median_household_income"], y=x["county"], fit_reg=True)

```
In [87]: print("Median vs County")
         fig, axe = plt.subplots(figsize=(10,10))
         sns.regplot(x=x["median_household_income"], y=x["county"], fit_reg=True)
```

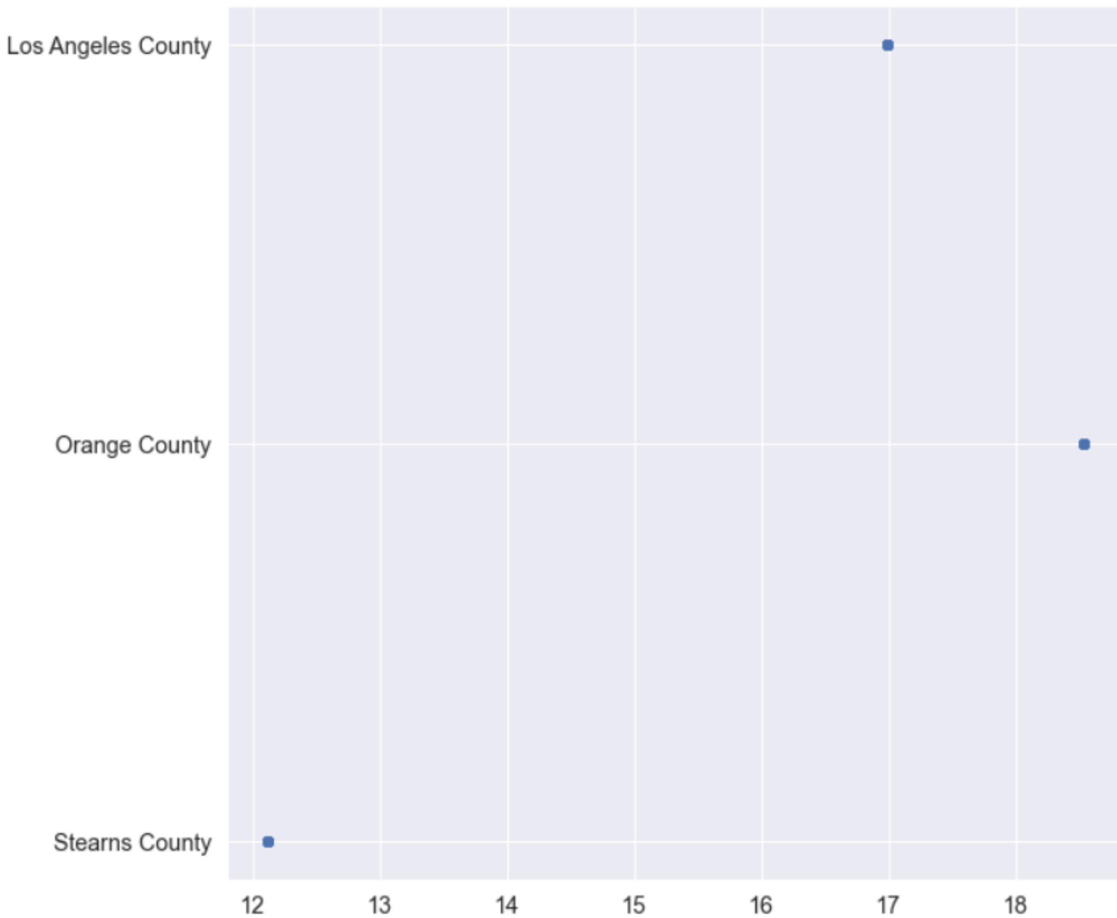


The outlier in Orange County is half the value of the outlier in Los Angeles County, though the average mean of Los Angeles County is little greater than Orange County. This explains Orange County is not an ideal location for Solar energy implementations.

Electricity vs County

print("Electricity vs County")
fig, axe = plt.subplots(figsize=(10,10))

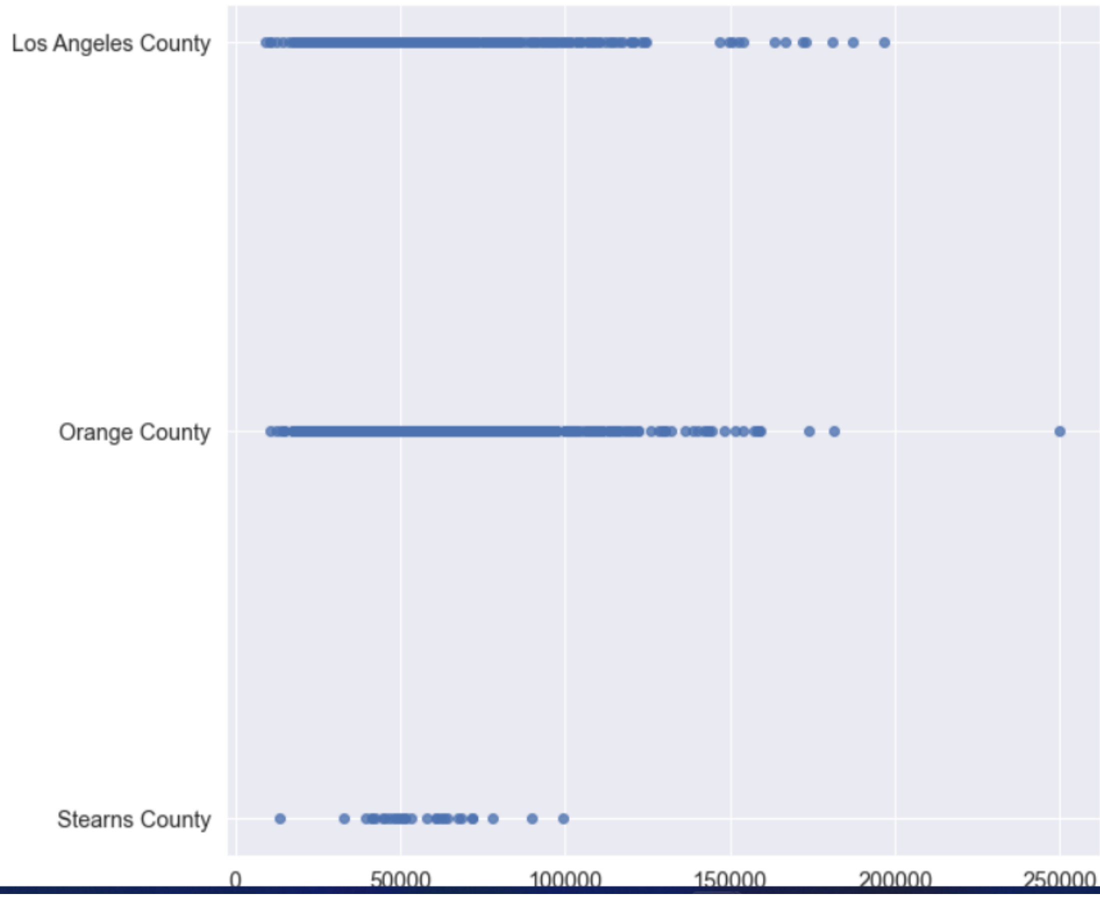
```
sns.regplot(x=x["electricity_price_residential"], y=x["county"], fit_reg=True)
```



After, changing the Attributes of “County “, again analyze the same representations and check whether the previous factors apply to these changes.

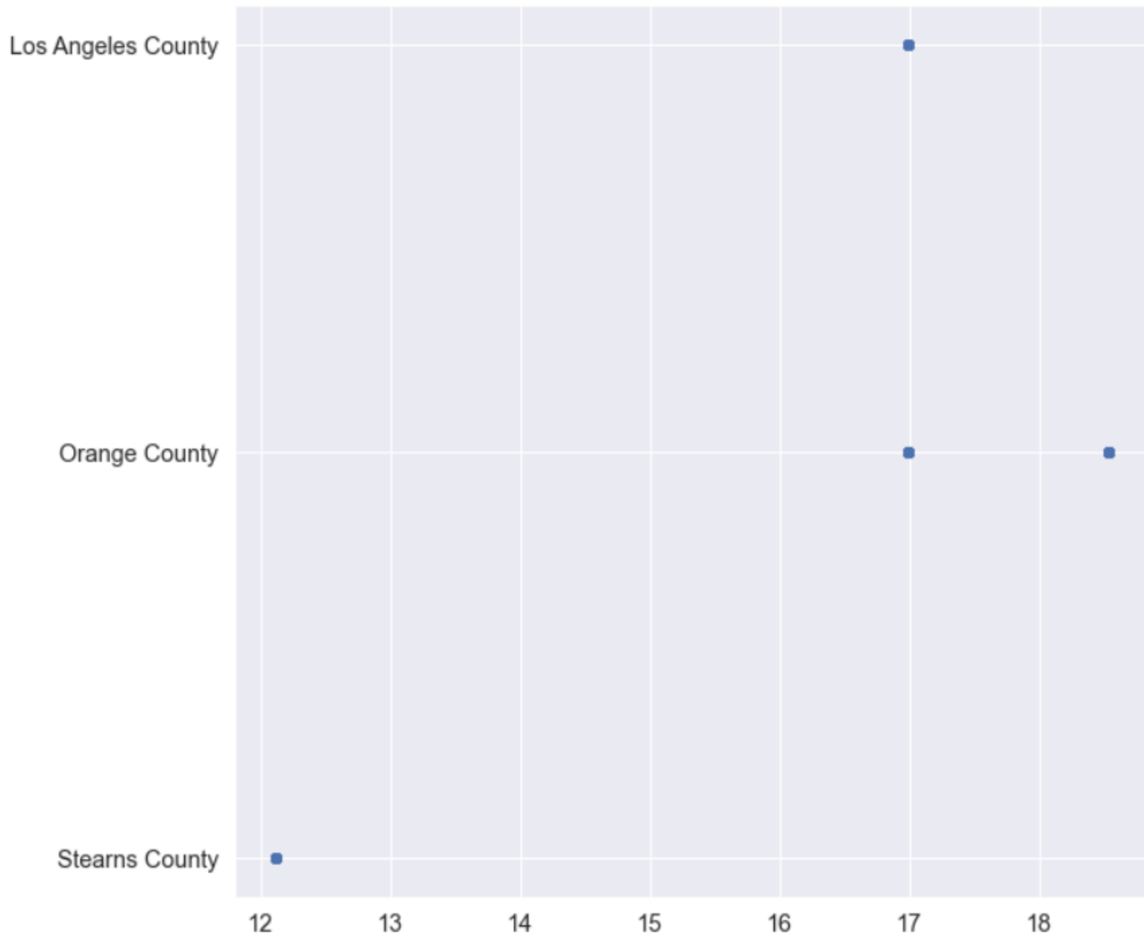
*) Median vs County

```
print("Median vs County")
fig, axe = plt.subplots(figsize=(10,10))
sns.regplot(x=y["median_household_income"], y=y["county"], fit_reg=True)
```



Electricity vs County:

```
print("Electricity vs County")
fig, axe = plt.subplots(figsize=(10,10))
sns.regplot(x=y["electricity_price_residential"], y=y["county"], fit_reg=True)
```



Despite the similarity in median income for both Los Angeles and Orange County, the electricity still dominates for orange county. Hence, by changing the variables in attributes we conclude that the previous conclusion is appropriate and correct. The same implies for the reasons behind it as well.

Part III: What is the data saying to you?

It was a great experience working on this project. I attribute provided me a lot of technical and non-technical information about Solar Energy. I was able to access both ins and outs of Solar Energy. Upon completion of this project, I was able to understand the relationship between economic, environmental, and technical incentives. I drew many valuable conclusions from the representations. I believe more analysts should work on green sector and enhance their skills on improving this. With this knowledge, they will be better able to understand what environmental, social, and economic factors influence a population's use of solar energy. Hopefully, in the future, we will reduce the inequity of solar energy distribution and be more equipped to advance, support, manage, and maintain the implementation of solar energy to power the United States.

Identify 3-5 observations or follow-up questions that you have:

I believe the efficiency, usage, factors involved, and rate of growth should not be limited to just analytics. All the data which is analyzed must be fir into latest machine learning algorithms to solve

predictive analytic problems in renewable space sector. This is the major takeaway I will take from my observation.
“This field has the potential to change the perspective of analytics is not a business aspect but also environmental aspect. “

With respect to my company, I want SunWorks.Inc to invest in technology for developing even more (As explained in SWOT Analysis). Based on my observations, the difference between the success factors for mid-cap companies and large-cap companies is their investment in technology. Technology is not limited to analyzing trends but also putting forth algorithms and eradicating problems. One such problem, which we discussed above is the uncertain cost of electricity, Analysis of power markets' production rates and soft costs, as well as solar access and environmental effect, and PV integration for grid planning and reliable operation. Sunworks.Inc must hire quality data scientists for this task and must be in the trend with this sector along with an ideal financial strategy. One such interesting research going on this sector is horizontal photovoltaics.

One such aspect Sunworks.Inc must investigate the effect of inflation in United States. They must be aware that the prices of equipment’s are rising. Financial Analyst must be aware that the revenue, net income, and cash flow have met at a meeting point on Sept 1. This is due to the risk of de-listing from the NASDAQ in revenue terms. Although, they had got an extension on Sep 16 for half a year, things are turning out well for Sunworks.Inc. An informal flotilla of ships transporting liquefied natural gas (LNG) across the Atlantic from the US Gulf Coast is the most visible sign of the Biden administration's support for its EU partners. Europe's gas supply is running low, and prices are at all-time highs, thanks in part to a slowdown in Russian gas shipments. Overall, Ukraine-Russia war tensions might affect stock prices and energy prices in Texas state.

I would like to thank Prof. Drobnis Kristen for providing me this wonderful opportunity to analyze “Green Sector and take way valuable insights from this project and coursework.

Resources and References:

(2019, February 22). Implementation of Solar Energy in the US. Kaggle. https://www.kaggle.com/andromedasagan/implementation-of-solar-energy-in-the-us/data

Sunworks, Inc. (2022, January 31). Home. SUNWORKS | SOLAR SOLUTIONS. https://sunworksusa.com/

Solar Energy Cost and Data Analysis. (2022). Energy.Gov. https://www.energy.gov/eere/solar/solar-energy-cost-and-data-analysis

Olaoye, A. (2022b, January 8). Predicting solar power output using machine learning techniques. Medium. https://towardsdatascience.com/predicting-solar-power-output-using-machine-learning-techniques-56e7959acb1f

Why the future looks bright for solar energy. (2020, November 9). World Economic Forum. https://www.weforum.org/agenda/2020/01/the-future-looks-bright-for-solar-energy/

Weisbrod, K. (2022, February 13). How the Ukraine Conflict Looms as a Turning Point in Russia’s Uneasy Energy Relationship with the European Union. Inside Climate News.

https://insideclimatenews.org/news/12022022/how-the-ukraine-conflict-looms-as-a-turning-point-in-russias-uneasy-energy-relationship-with-the-european-union-%EF%BF%BC/

10 Biggest Renewable Energy Companies in the World. (2021, December 27). Investopedia. https://www.investopedia.com/investing/top-alternative-energy-companies/

Sunworks (NasdaqCM:SUNW) - Stock price, News & Analysis. (2022, February 15). Simply Wall St. https://simplywall.st/stocks/us/capital-goods/nasdaq-sunw/sunworks#about

Calzon, B. (2021, April 30). Financial Graphs And Charts - See Here 20 Business Examples. BI Blog | Data Visualization & Analytics Blog | Datapine. https://www.datapine.com/blog/financial-graphs-and-charts-examples/

How to Write & Present an Industry Analysis. (2021, December 23). Industrial Distribution. https://www.inddist.com/operations/article/21545231/how-to-write-present-an-industry-analysis

Strategic Management Department. (2021). Sunworks, Inc. SWOT Analysis Matrix (Strengths, Weakness, Opportunities, Threats). Fern Fort University. http://fernfortuniversity.com/term-papers/swot/nyse/3948-sunworks--inc-.php

Stock Analysis: Forecasting Revenue and Growth. (2021, December 1). Investopedia. https://www.investopedia.com/articles/active-trading/022315/stock-analysis-forecasting-revenue-and-growth.asp