

Week 6 Signature Assignment



**Submitted By:** Rama Niteesh Mallavarapu

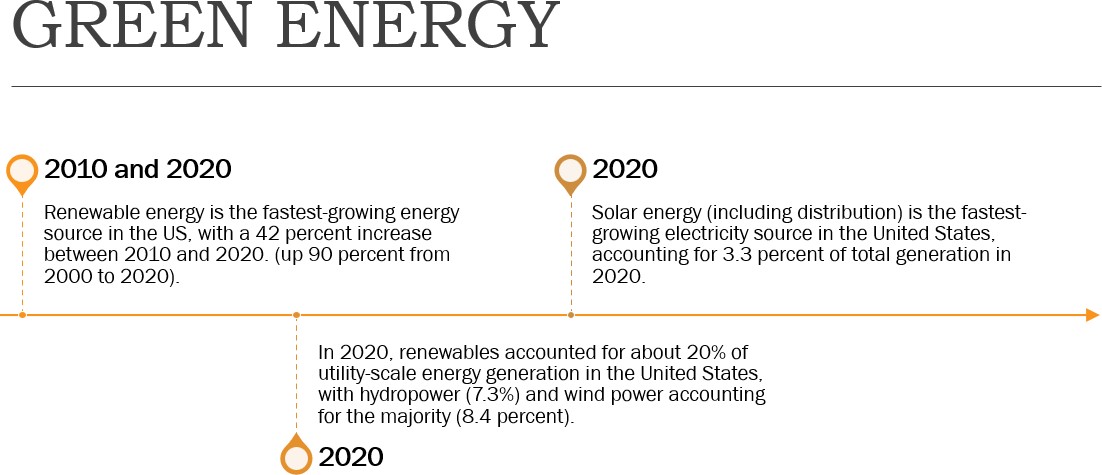
**Course Code:** ALY 6000 – Introduction to Analytics

**Instructor:** Kristen Drobnis

**Submitted on:** February 19th, 2022

Green Energy

Introduction:



Company: SunWorks.Inc

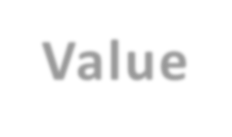
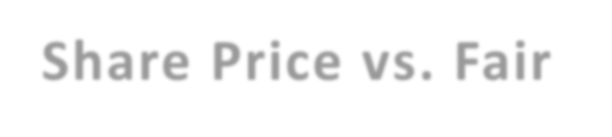
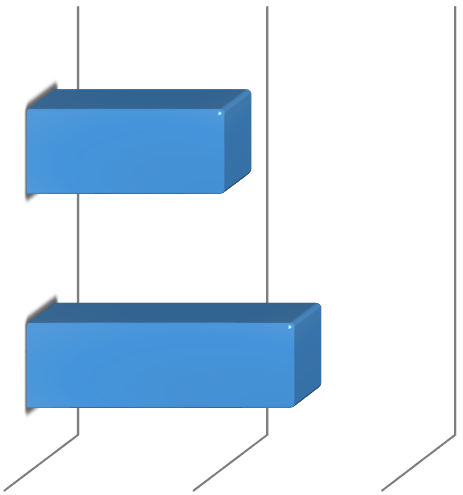
Sunworks, Inc. (NASDAQ: SUNW) is a premier provider of high-performance solar power systems. We are committed to quality business practices that exceed industry standards and uphold our ideals of ethics and safety.

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.



**Share Price vs. Fair**

**Value**

$0.00

$2.00

$4.00

Current Price

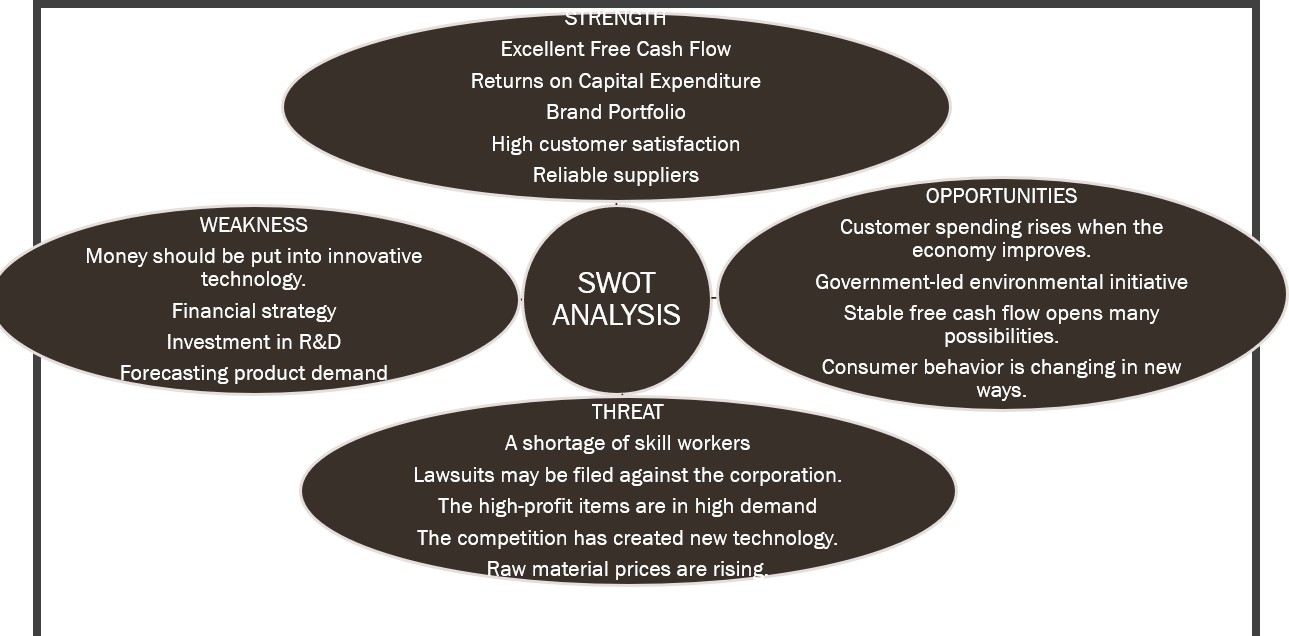
$2.08

Fair Value

$2.82

Share Price vs. Fair Value

Company with respect to industry:

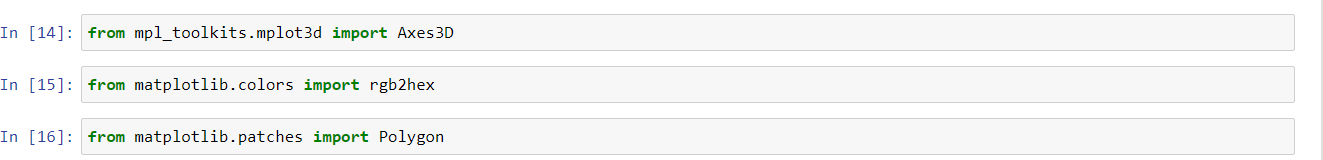


(SWOT Analysis is discussed in the observation section)

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.

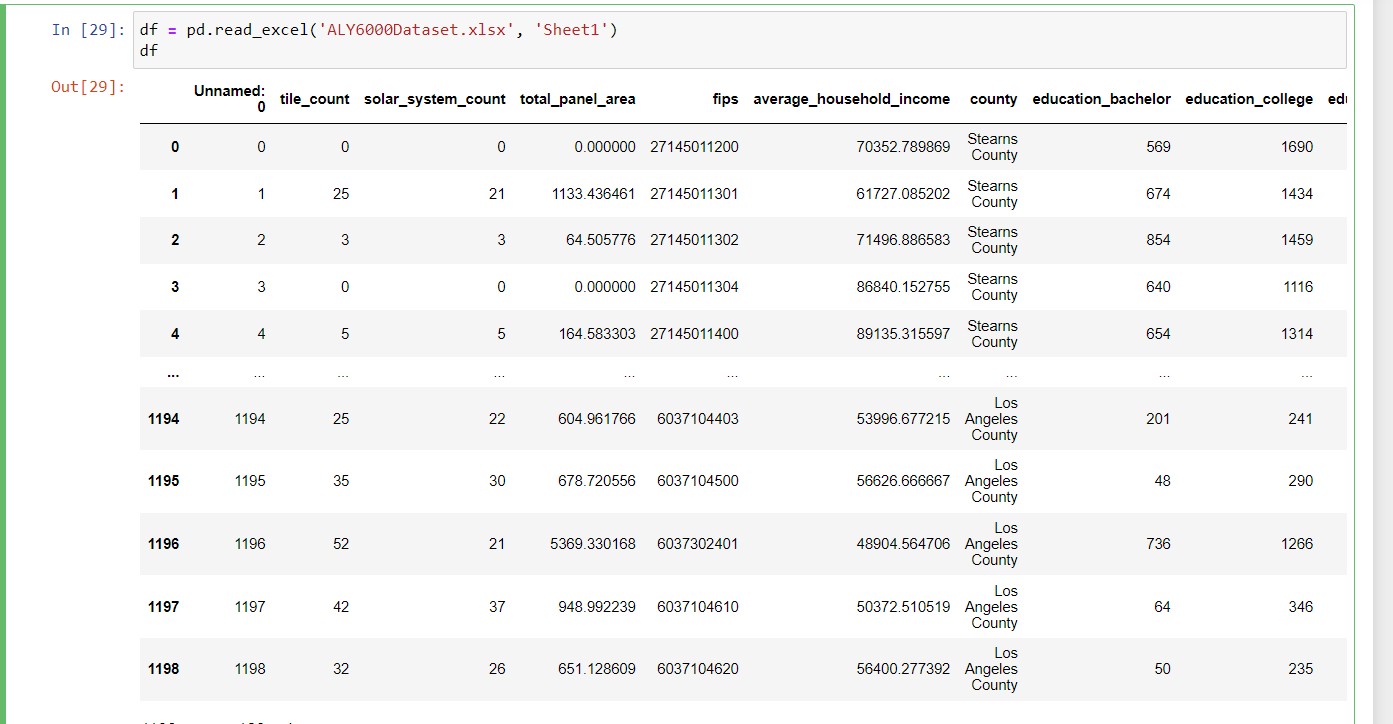
1. Importing Libraries



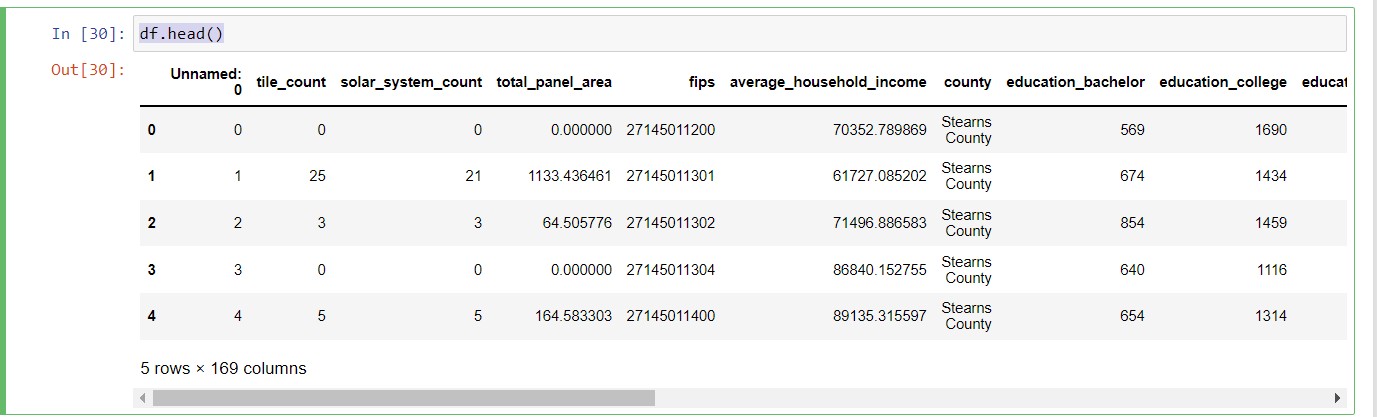
Hence, the libraries are imported.

1. Loading Data into Jupyter Notebook

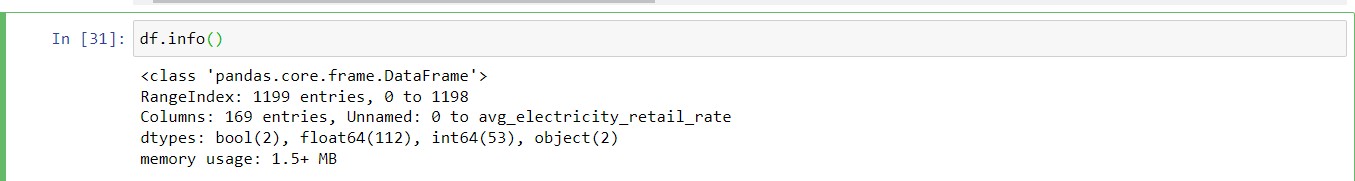
df = pd.read\_excel('ALY6000Dataset.xlsx', 'Sheet1') df



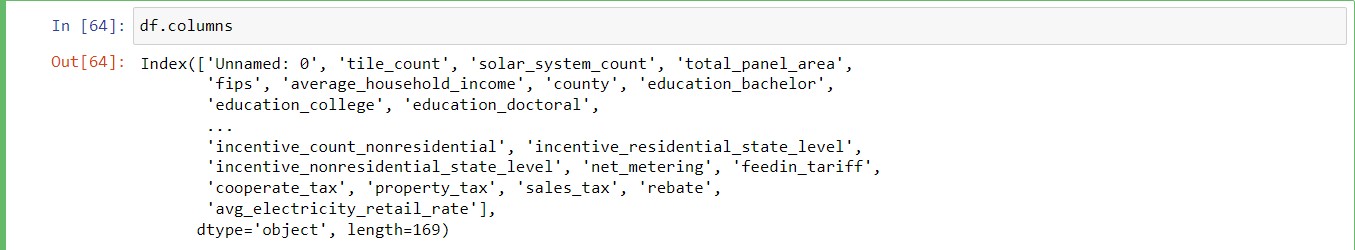
df.head()



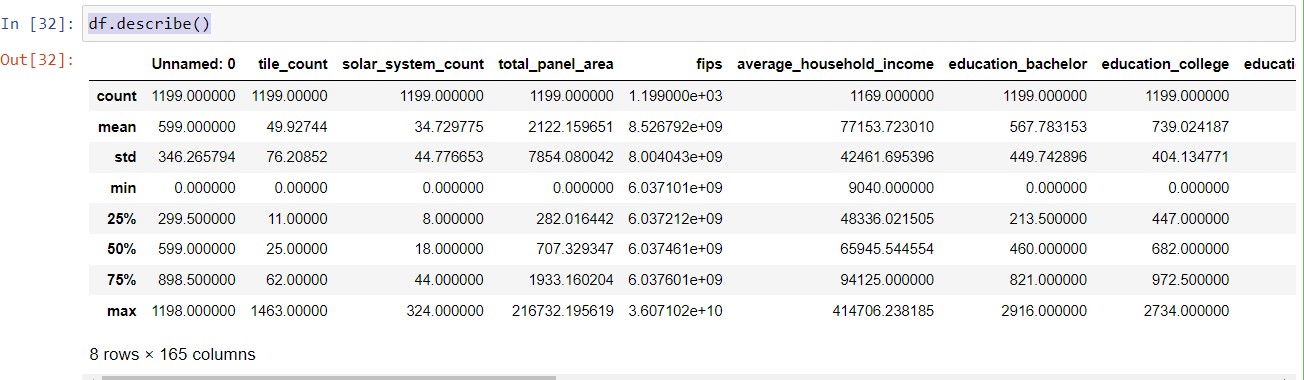
df.info ()



df.columns



df.describe()



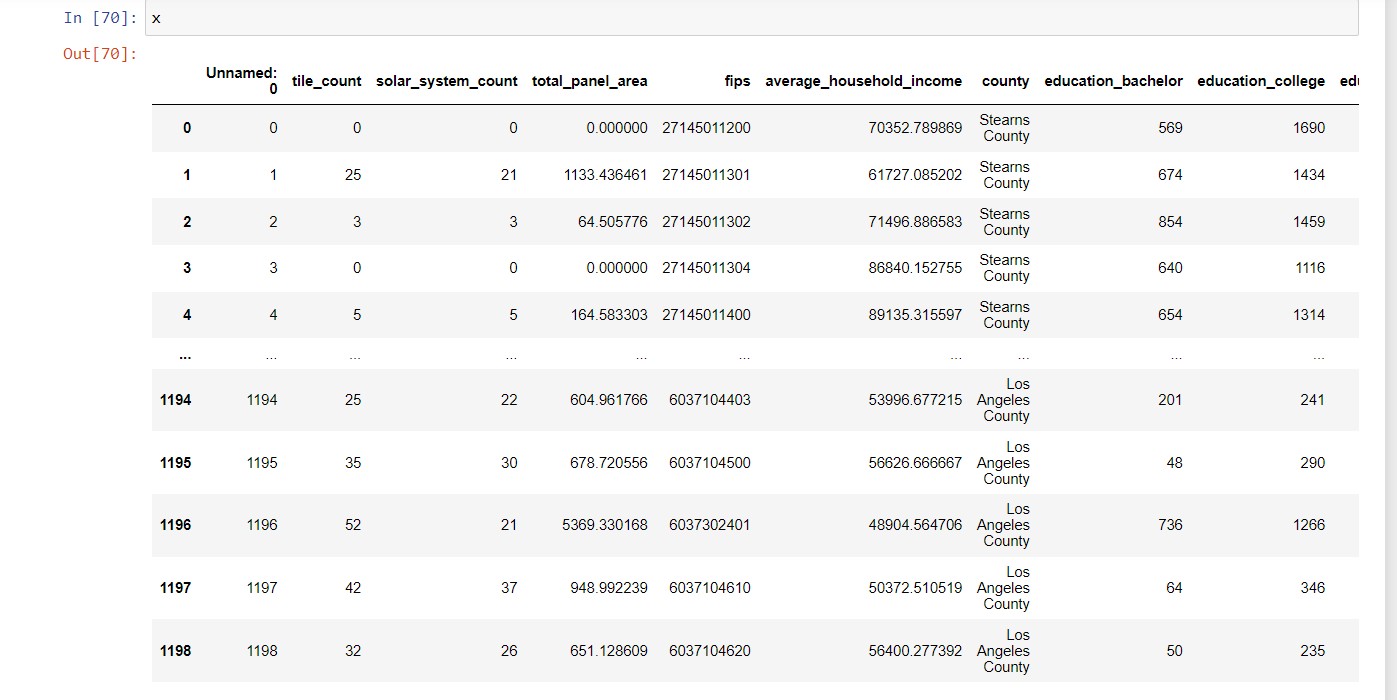
Data Cleaning:

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','educat ion\_high\_school\_graduate','education\_less\_than\_high\_school','education\_master','education\_p opulation','education\_professional\_school','employed'])

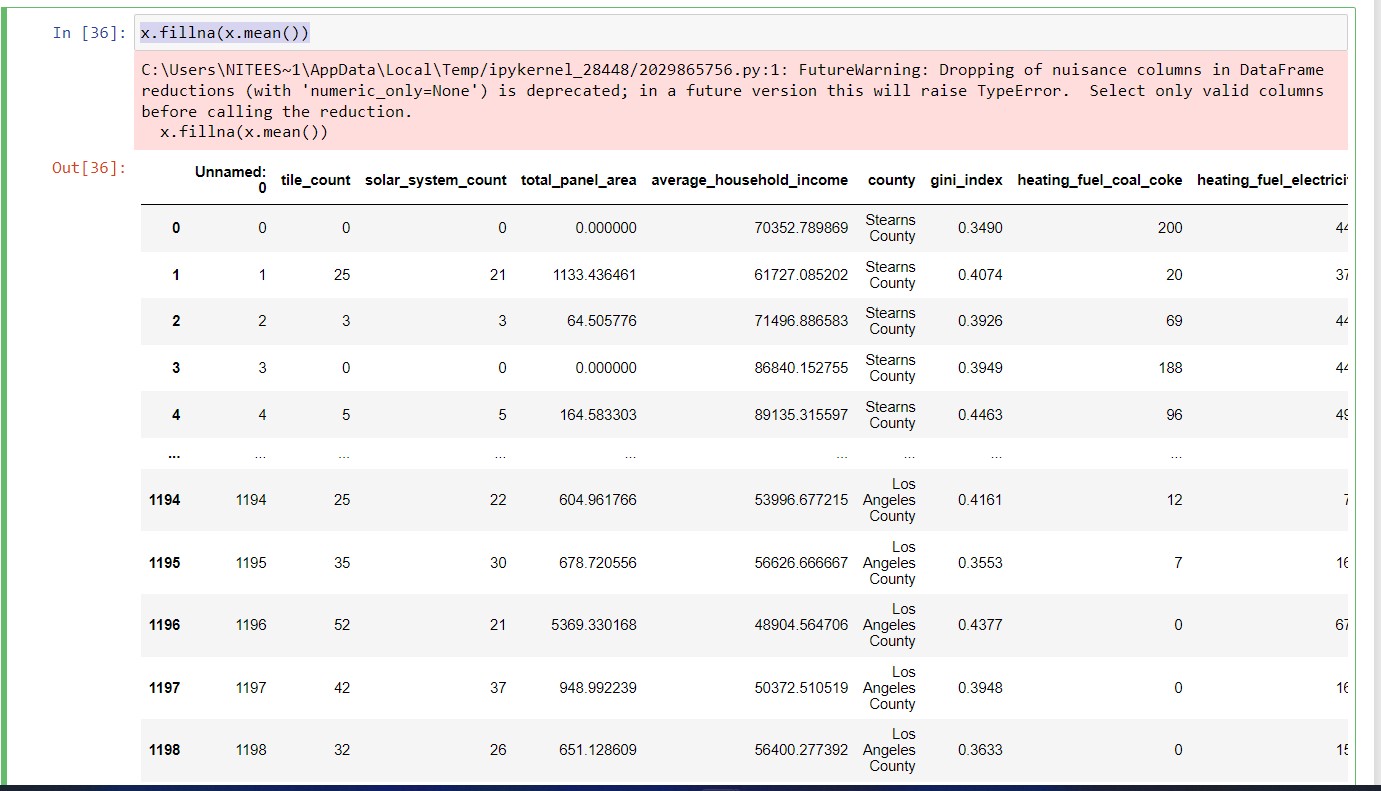
x=df.drop(columns=['poverty\_family\_below\_poverty\_level','race\_asian','race\_black\_africa','race

\_indian\_alaska','race\_islander','race\_other','race\_two\_more','race\_white','race\_white\_rate','rac e\_black\_africa\_rate','race\_indian\_alaska\_rate','race\_asian\_rate','race\_other\_rate'])

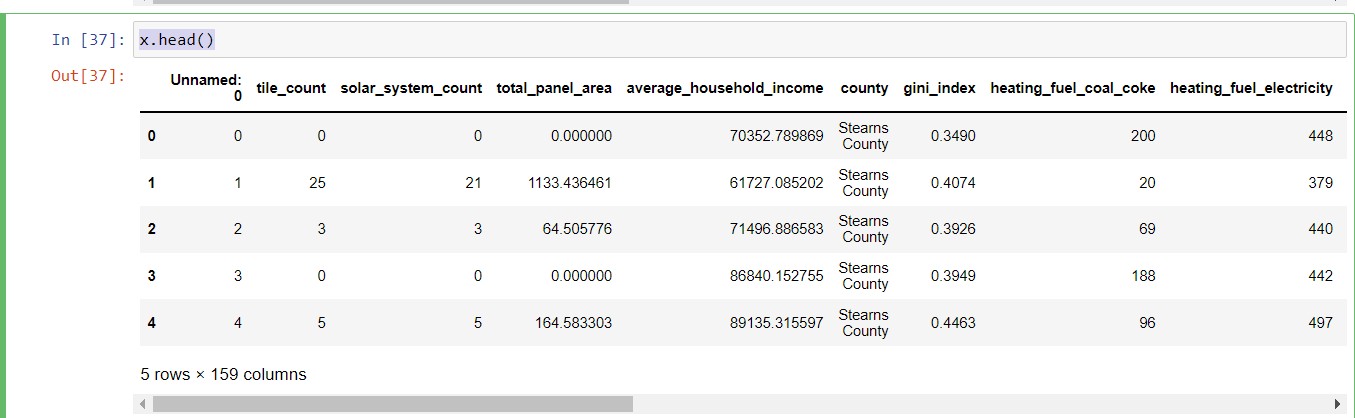


x.fillna(x.mean())

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



x.head()



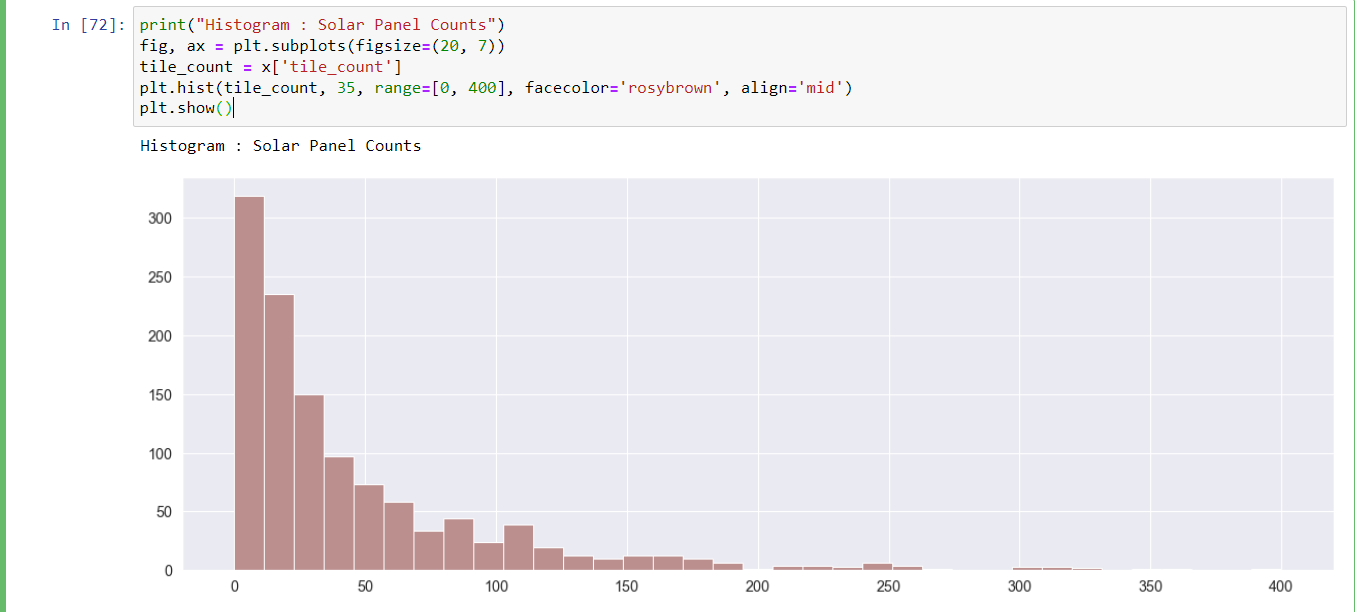
1. Data Visualization

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()



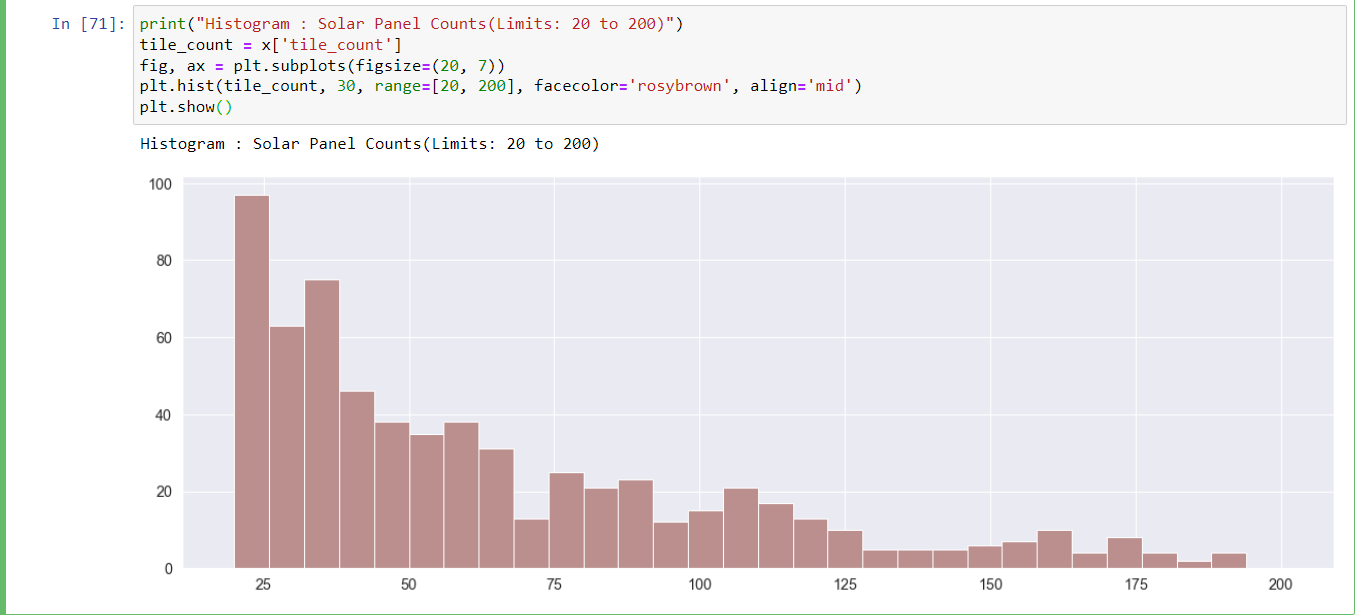
Most of the locations have no solar panels, as seen by this histogram of solar panel counts. I generated a second histogram of solar panel counts for places with at least 20 solar panels but fewer than 100 to see a more statistically rich chunk of this data.

\*) 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()

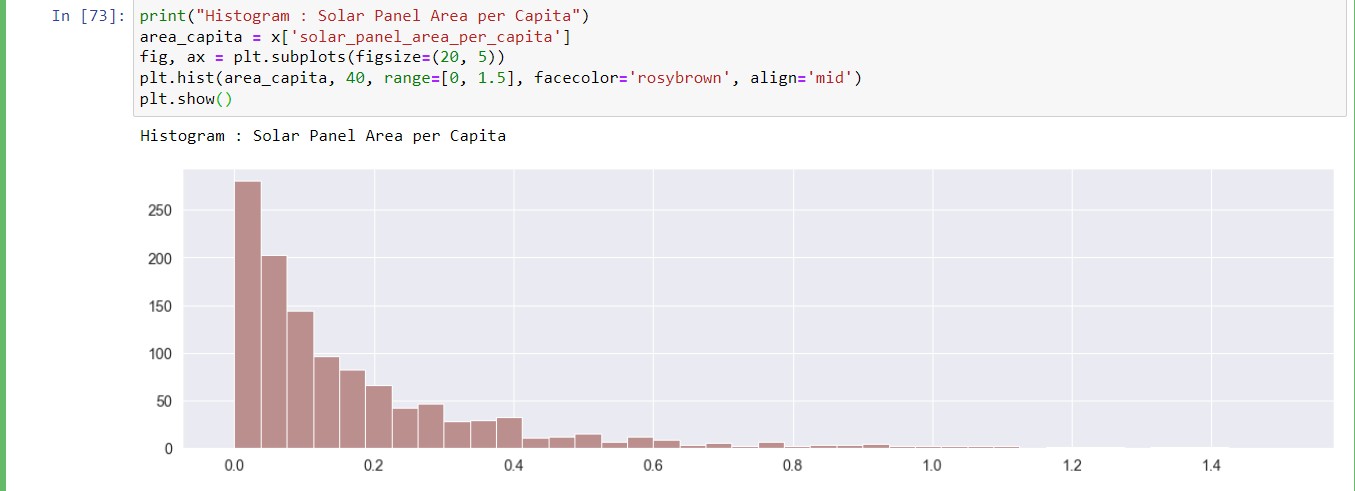


The trend is even clearer when set limits from 20 to 200. Most residential and non-residential houses have less than 60 solar panels. The frequency count is clearly visible before and after 60. Right after the count of 60, the overall number has been decreasing but spike was observed at 80 and 110. The 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()

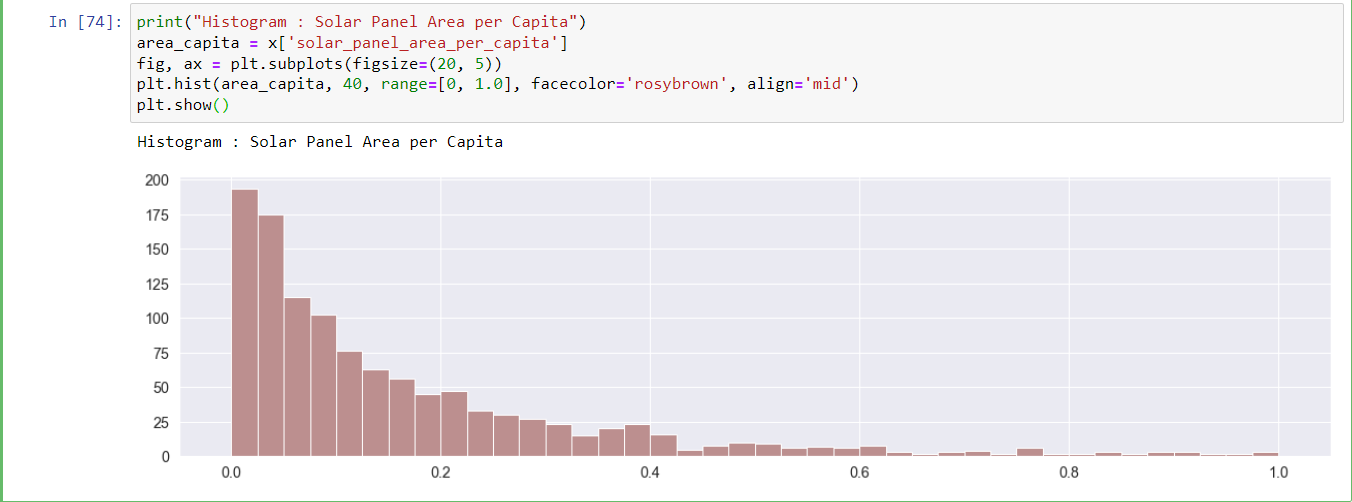


Another set of histograms, this time spanning from 0 to 1, depicts the frequency of solar panel area per population. Most of the data points are zero, as predicted from the first histogram.

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()



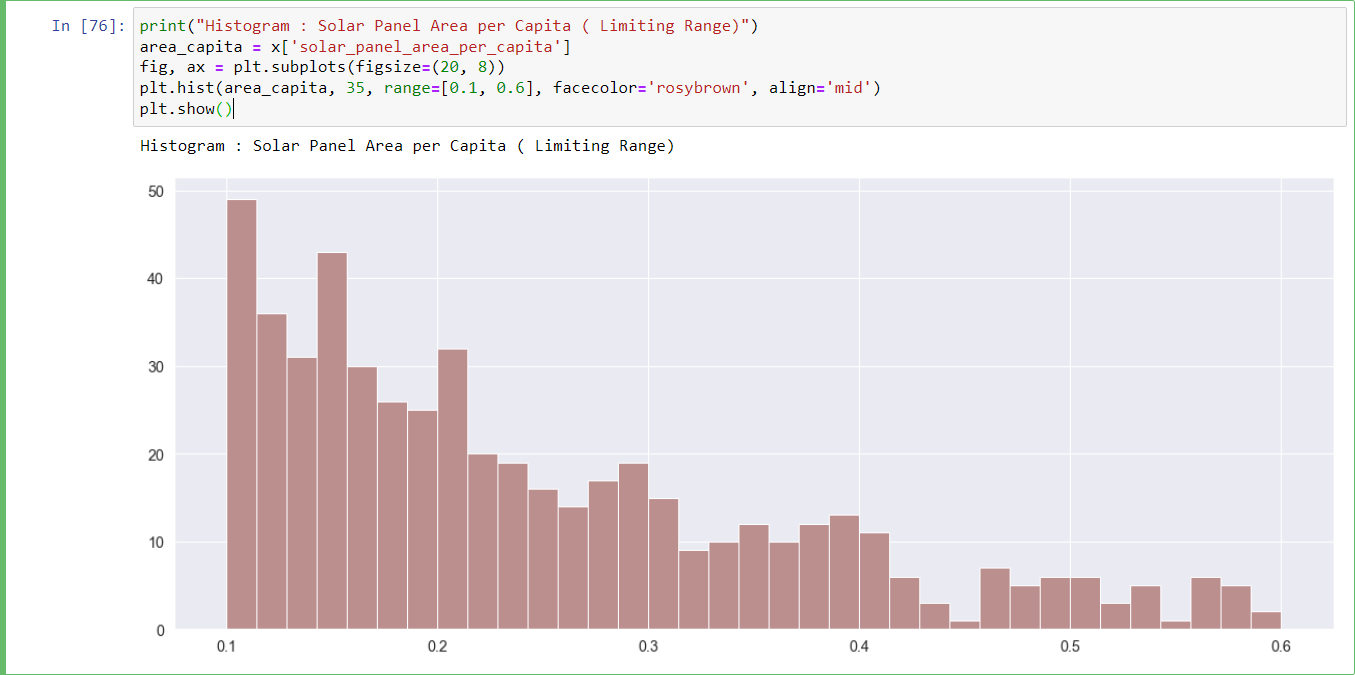
Majority of the data points are 0. Again, the number has been gradually decreasing from 0 to 1. Percapita has been bought into picture to ensure that the data is not biased by population size or density in an area. From business perspective, it is always beneficial for orders to be booked for multiple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



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]

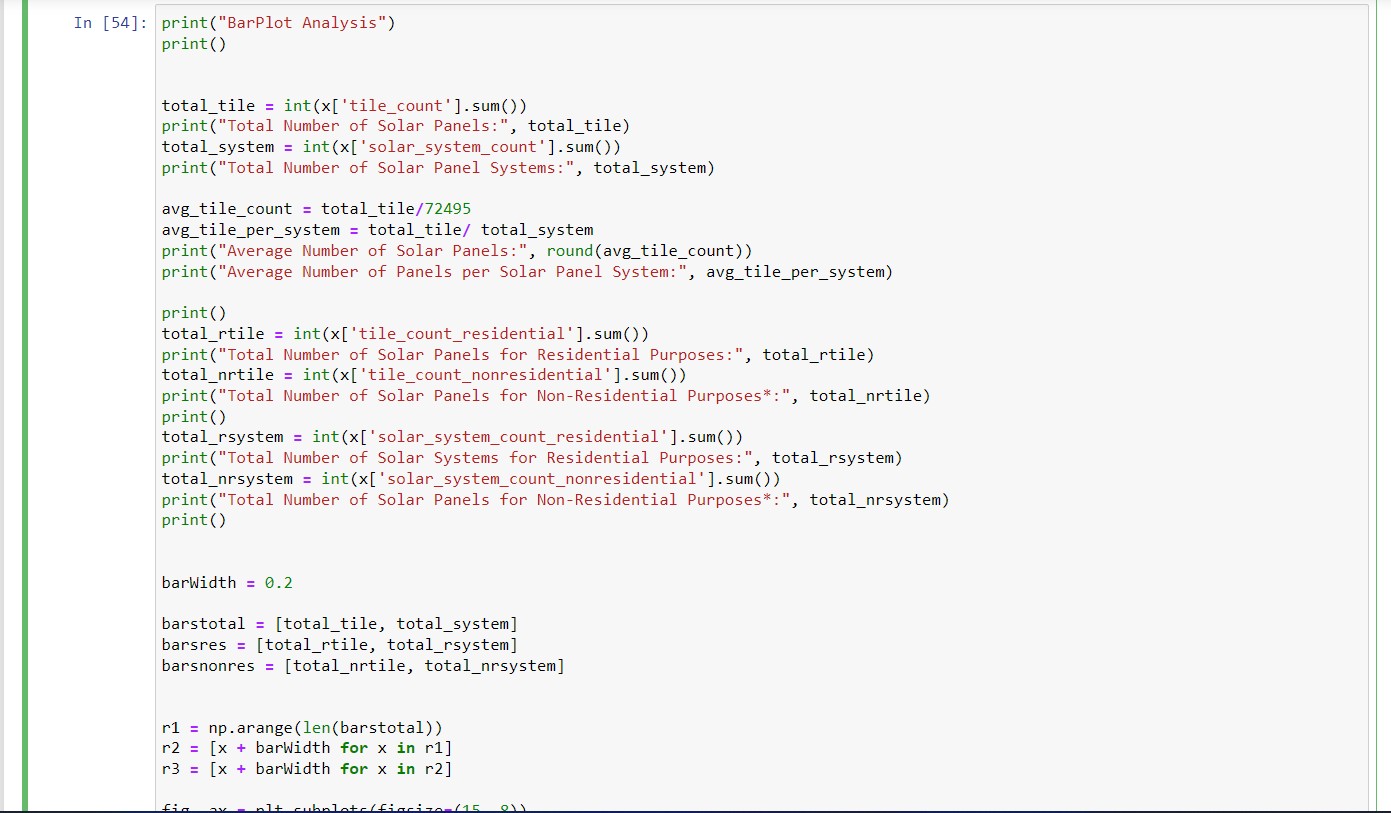
r1 = np.arange(len(barstotal)) r2 = [x + barWidth for x in r1] r3 = [x + barWidth for x in r2]

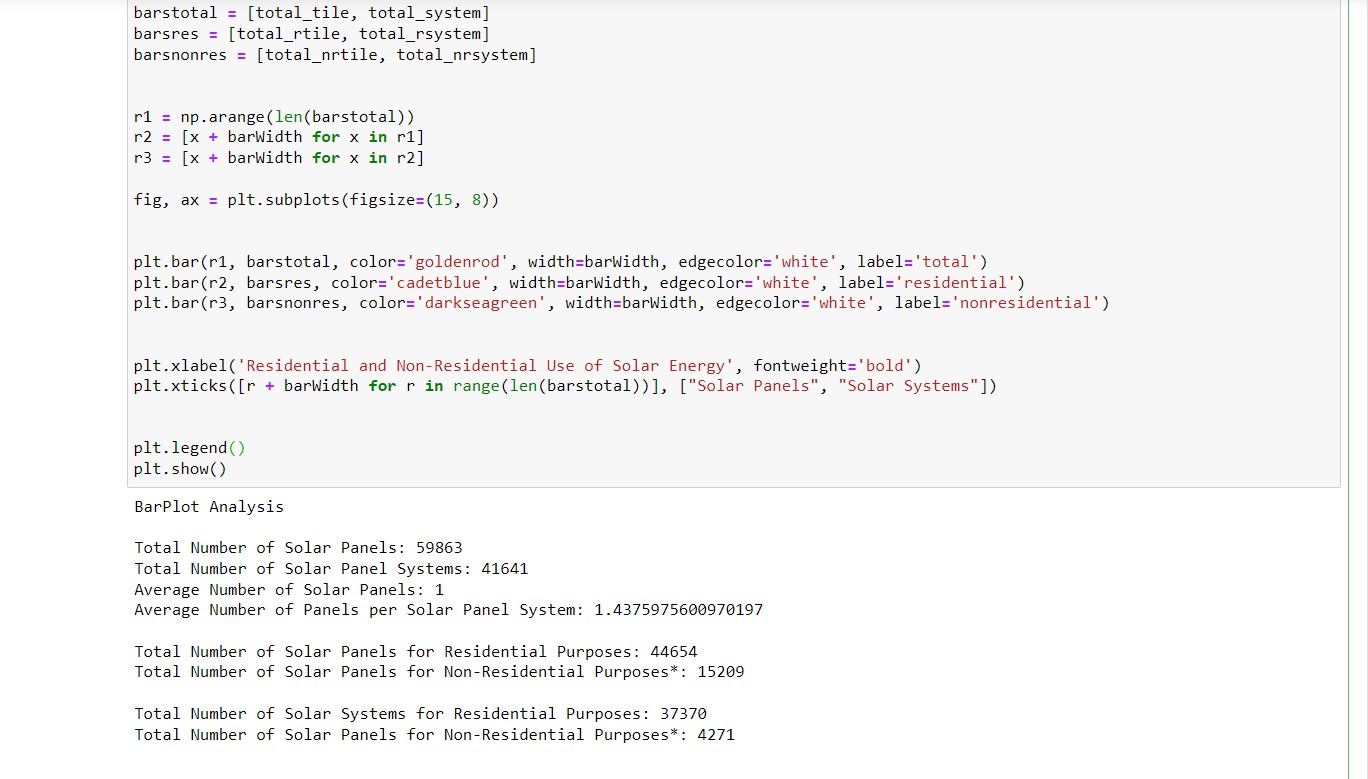
fig, ax = plt.subplots(figsize=(15, 8))

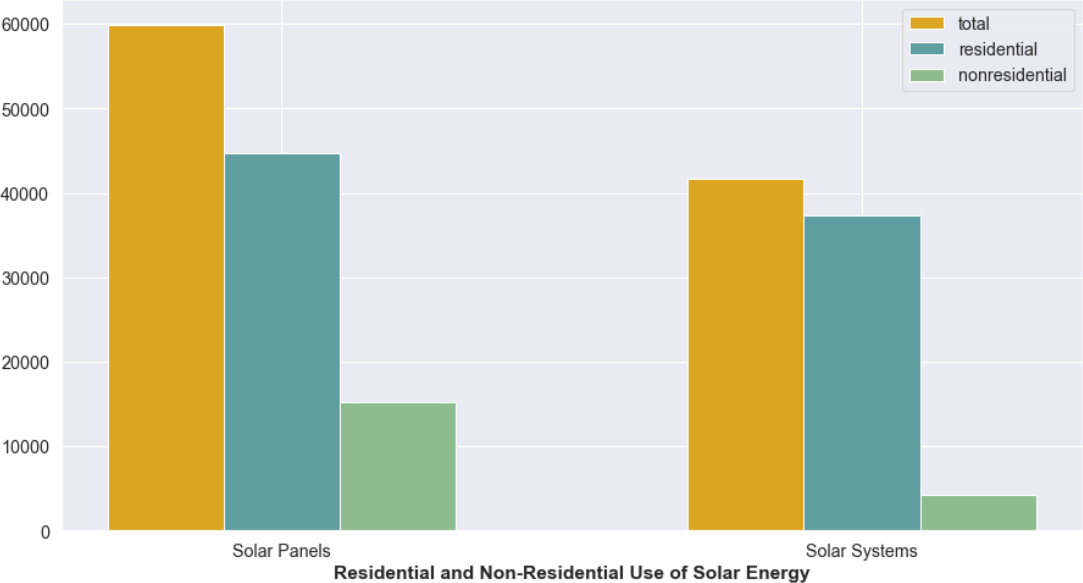
plt.bar(r1, barstotal, color='goldenrod', width=barWidth, edgecolor='white', label='total') plt.bar(r2, barsres, color='cadetblue', width=barWidth, edgecolor='white', label='residential') plt.bar(r3, barsnonres, color='darkseagreen', width=barWidth, edgecolor='white', label='nonresidential')

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()







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.

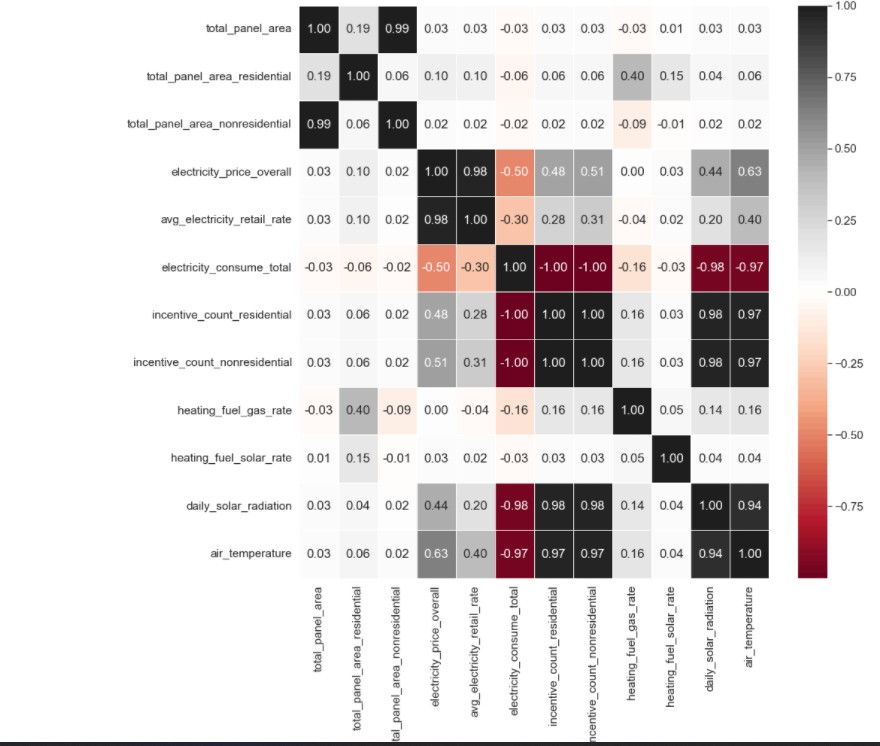
energyenv\_quant\_data = ["total\_panel\_area", "total\_panel\_area\_residential", "total\_panel\_area\_nonresidential","electricity\_price\_overall", "avg\_electricity\_retail\_rate", "electricity\_consume\_total", "incentive\_count\_residential", "incentive\_count\_nonresidential", "heating\_fuel\_gas\_rate", "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)





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.

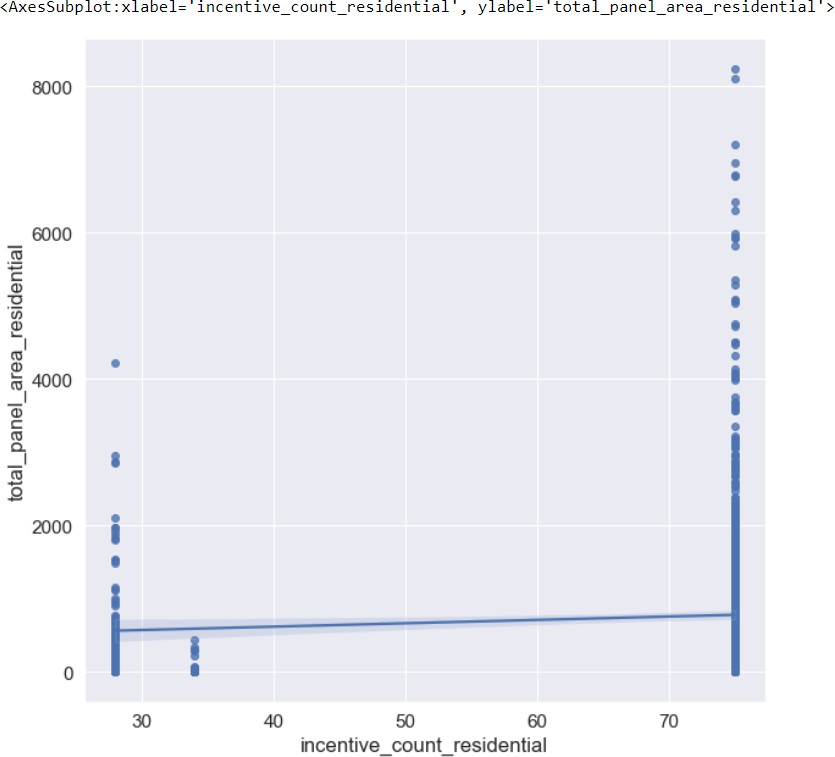
Despite that, it is interesting to see that electricity price overall is vice versa, and analyst should make note of this. This can be further improved as controlling electricity prices are still in a developing phase for green energy projects. Talking more about electricity, there is also a relation 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)



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.

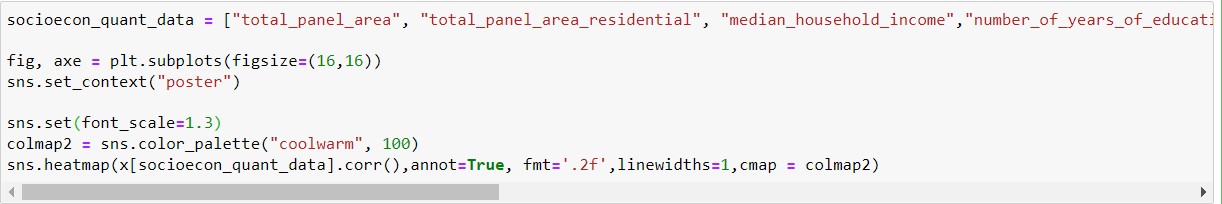
\*) Correlations between socioeconomic factors and the implementation of solar energy.

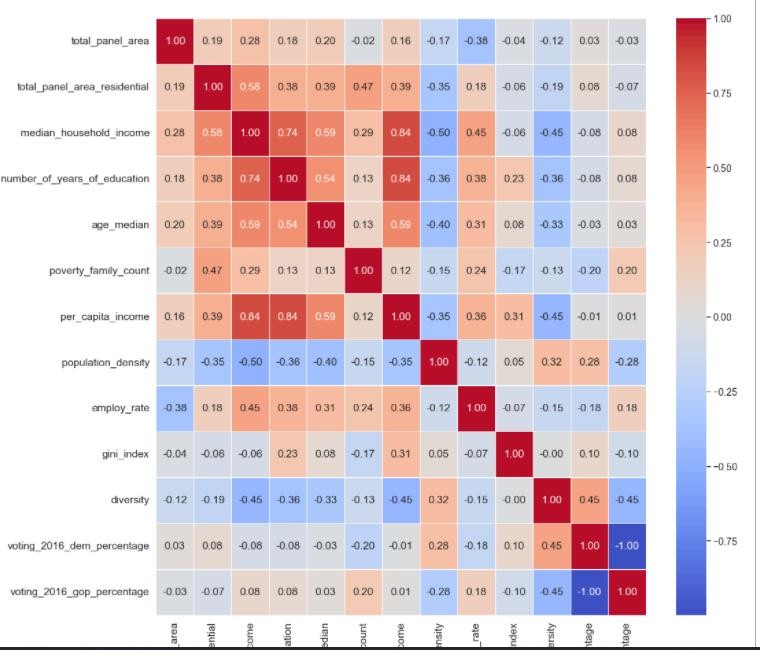
socioecon\_quant\_data = ["total\_panel\_area", "total\_panel\_area\_residential", "median\_household\_income","number\_of\_years\_of\_education", "age\_median", "poverty\_family\_count", "per\_capita\_income", "population\_density", "employ\_rate", "gini\_index", "diversity", "voting\_2016\_dem\_percentage", '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)





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.

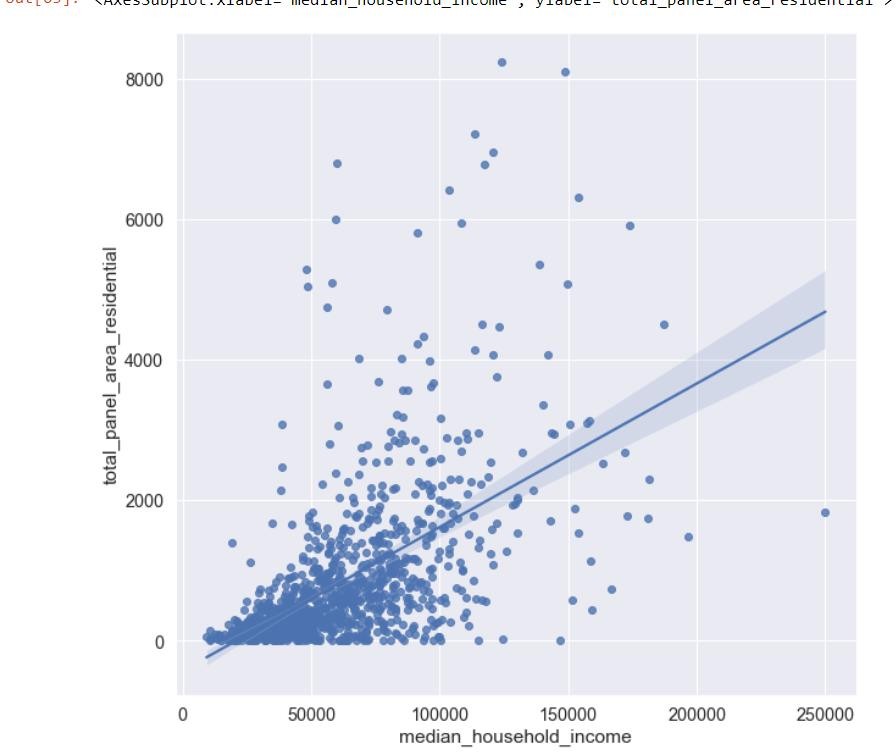
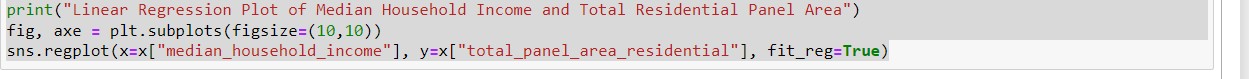
Discussion:

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)



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.

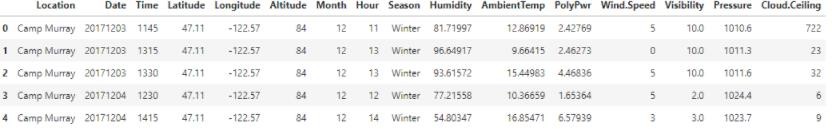
1. Conclusion

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:

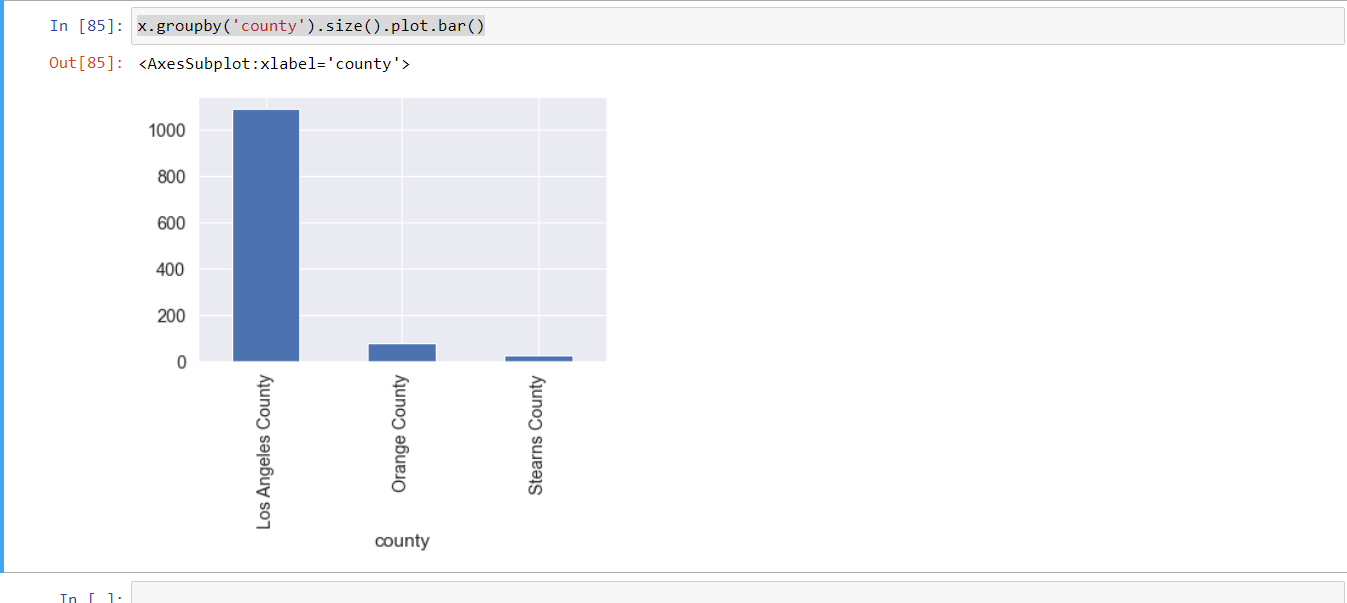


Olaoye, A. (2022, January 8). Predicting solar power output using machine learning techniques. Medium. [https://towardsdatascience.com/predicting-solar-power-output-using-machine-learning-techniques-](https://towardsdatascience.com/predicting-solar-power-output-using-machine-learning-techniques-56e7959acb1f)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

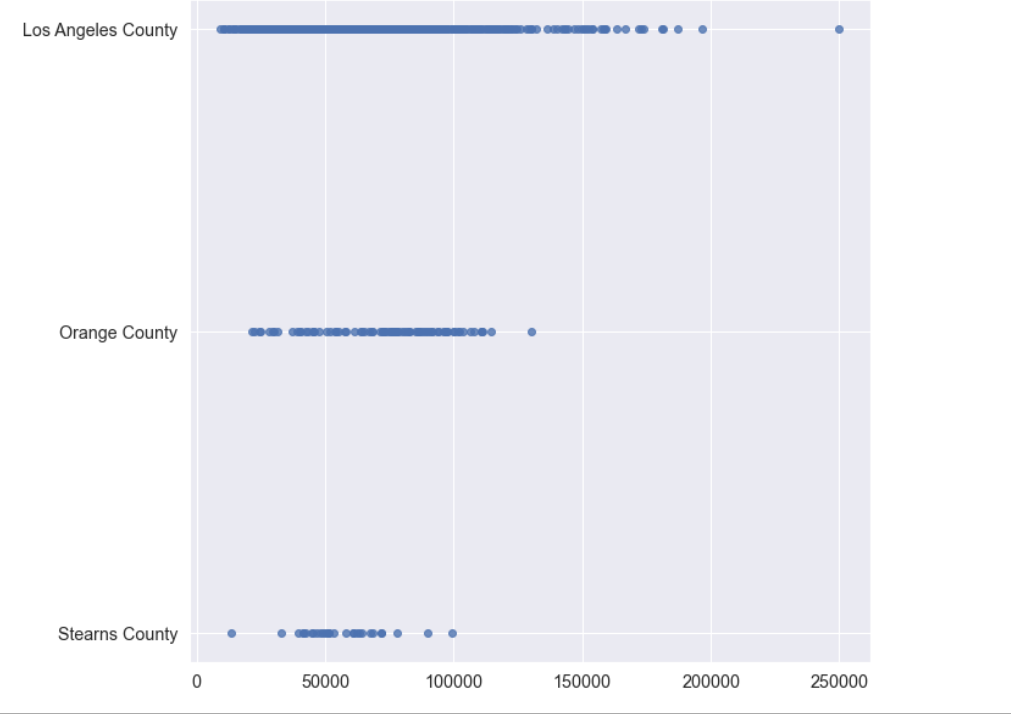
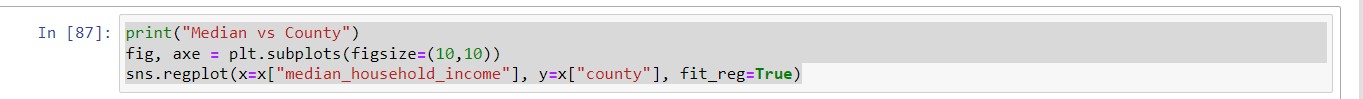


It is evident that population of users in Los Angeles County are far greater than Orange County and Stearns County. It is interesting to note that, the population in Orange County is almost negligible it

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)

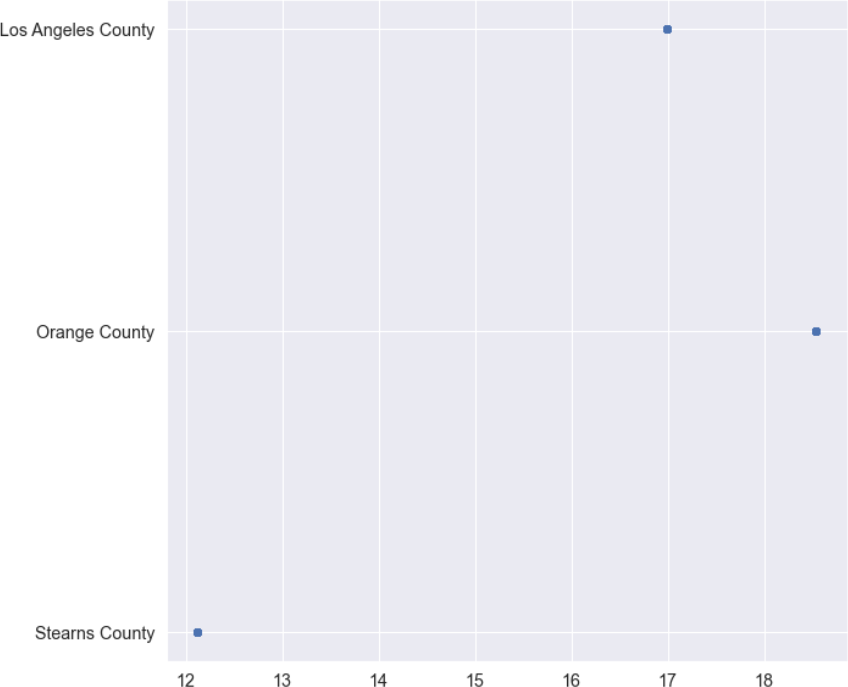


The outlier in Orange County is half the value of the outlier in Los Angeles County, though the average mean of Los Angeles County is little greater than Orange County. This explains Orange County is not an ideal location for Solar energy implementations.

Electricity vs County print("Electricity vs County")

fig, axe = plt.subplots(figsize=(10,10))

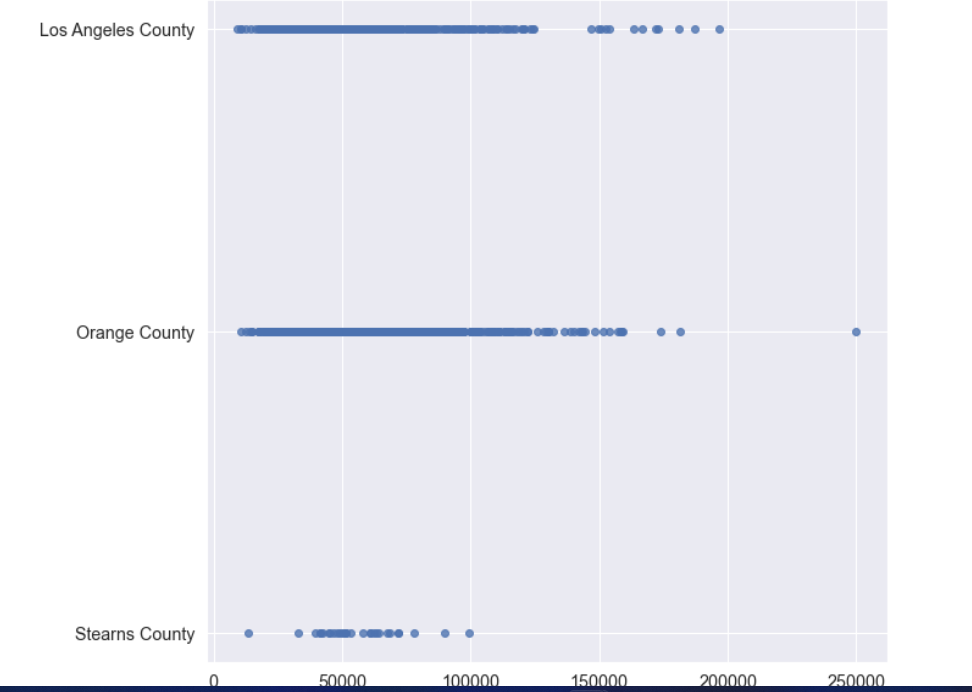
sns.regplot(x=x["electricity\_price\_residential"], y=x["county"], fit\_reg=True)



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")

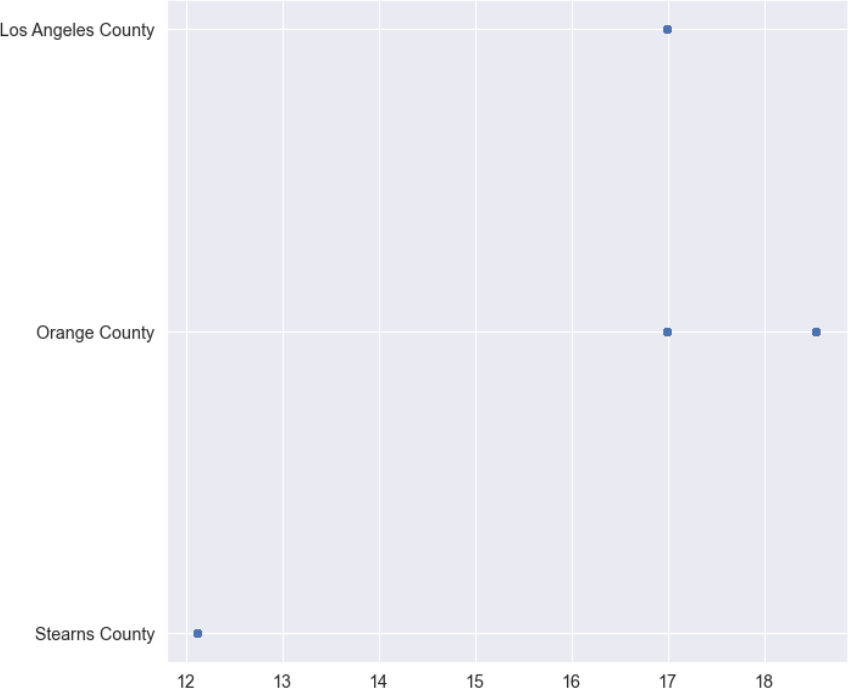
fig, axe = plt.subplots(figsize=(10,10)) sns.regplot(x=y["median\_household\_income"], y=y["county"], fit\_reg=True)



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. With about 162 megawatts deployed across the United States, the industry's and company's futures are bright since the rate of adaptation is swift and rapid. Nearly 400 megawatts of solar energy have already been deployed by major corporations. With supply chain issues and inflation uncertainty, Sunworks can improve their cash flows by implementing the above factors and see a good success rate, with a cash flow improvement of 28% and an increase of 80% in installation capacity, as it has the potential to generate enough electricity to power 1.6 million US homes each year.

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-](https://www.energy.gov/eere/solar/solar-energy-cost-and-data-analysis) [energy-cost-and-data-analysis](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-](https://towardsdatascience.com/predicting-solar-power-output-using-machine-learning-techniques-56e7959acb1f) [techniques-56e7959acb1f](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-](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/) [russias-uneasy-energy-relationship-with-the-european-union-%EF%BF%BC/](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](https://simplywall.st/stocks/us/capital-goods/nasdaq-sunw/sunworks)

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-](https://www.datapine.com/blog/financial-graphs-and-charts-examples/) [examples/](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-](http://fernfortuniversity.com/term-papers/swot/nyse/3948-sunworks--inc-.php) [papers/swot/nyse/3948-sunworks--inc-.php](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-](https://www.investopedia.com/articles/active-trading/022315/stock-analysis-forecasting-revenue-and-growth.asp) [growth.asp](https://www.investopedia.com/articles/active-trading/022315/stock-analysis-forecasting-revenue-and-growth.asp)