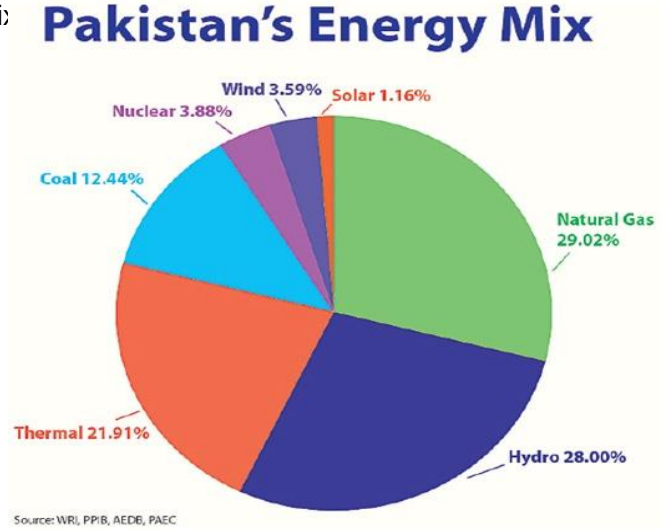
BATTLE OF THE SOLAR POWER PLANTS

-by Muhammad Ali Shamozai
Islamabad, Pakistan
Orient Energy Systems (Pvt.) Ltd.

Introduction:

Pakistan' peak electricity demand is estimated at 25,000 MW while its total power generation capacity is 28,000 MWs. Pakistan meets two-thirds of its energy requirement from fuel oil and natural gas. The share of renewables in Pakistan's energy mix

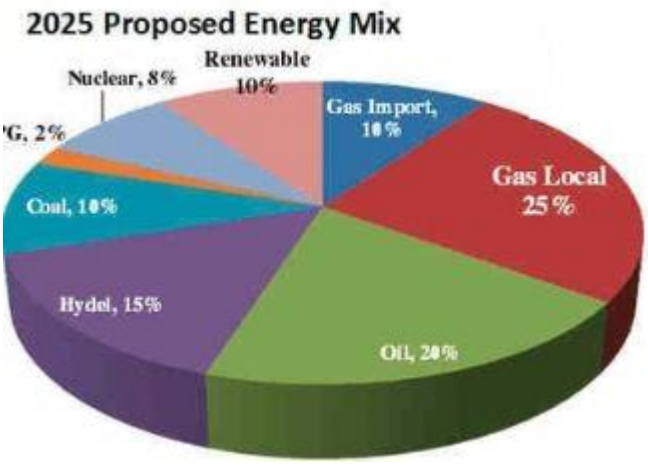


In November 2019 the Government of Pakistan unveiled its new Renewable Energy Development Policy.

The policy aims to increase the share of renewables in Pakistan's Electricity Generation by 30%.

Around 8,000MW cheap renewable clean and green energy will be added to the system by 2025 while it will be increased to 20,000MW by 2030.

This will bring the total share of renewables in Pakistan's energy mix to around 10% by 2025.



In order to meet the goals stated in this policy, Pakistan will have to invest heavily in Utility scale Solar and Wind energy plants. These plants are called Independent Power Producers or IPPs.

Investment by private investors is encouraged by offering lucrative buying tariffs per kilo watt hour, tax reliefs, lower or no import duties on import of equipment etc.

In this study we will try to determine the best places in Pakistan for setting up a solar Independent Power Plant (Solar IPP) which will provide better electricity output at optimized Capital Expense CAPEX to the investor/project owner.

Problem:

Like with all major utility scale projects, transmission cost and transmission losses are a major bottleneck in a project's feasibility. Pakistan's regions which receive very high Direct Normal Irradiation (DNI) of above 6 kWh/m2/day are very distant from the national power grid. Which means developing a solar IPP in one of those regions will have huge transmission costs. Also their remoteness, distance from seaports and nearby population centers will drive up the cost of project management and operation.

Another major factor in developing a Solar IPP is the cost of Land. The nearer the IPP site is to a city or a major population center the higher will be the cost of land and there will also be difficulties in acquiring a single plot of land for the IPP.

Data:

The World Bank Group has worked extensively on Pakistan's electricity transmission grid data. A GEOJSON file listing all nodes in the national grid is available at their website https://energydata.info/dataset/pakistan-electricity-transmission-network-2017 .I have used that data in this study.

I used Foursuare API to get venues nearby the grid stations. The purpose was to explore how many venues are there within a range of 15 kms of the Grid Station and how what was the mean distance of those venues from the Grid locations. Greater the number of venues near a grid and smaller mean distance of those venues from the Grid Location would mean that the Grid station is in an urban area and land development costs would be much higher.

Finally, I used the Global Solar Atlas website https://globalsolaratlas.info/map?c=30.637912,68.994141,5&r=PAK to get the Specific photovoltaic power output (PVOUT) values of the grid locations. I had to build the dataset of PVOUT values for each grid location by entering the coordinates one by one and getting the PVOUT values.

The resulting data frame has coordinates of all grid locations, their nearby venues count, mean distances from the venues and the Specific photovoltaic power output (PVOUT) values for each location.

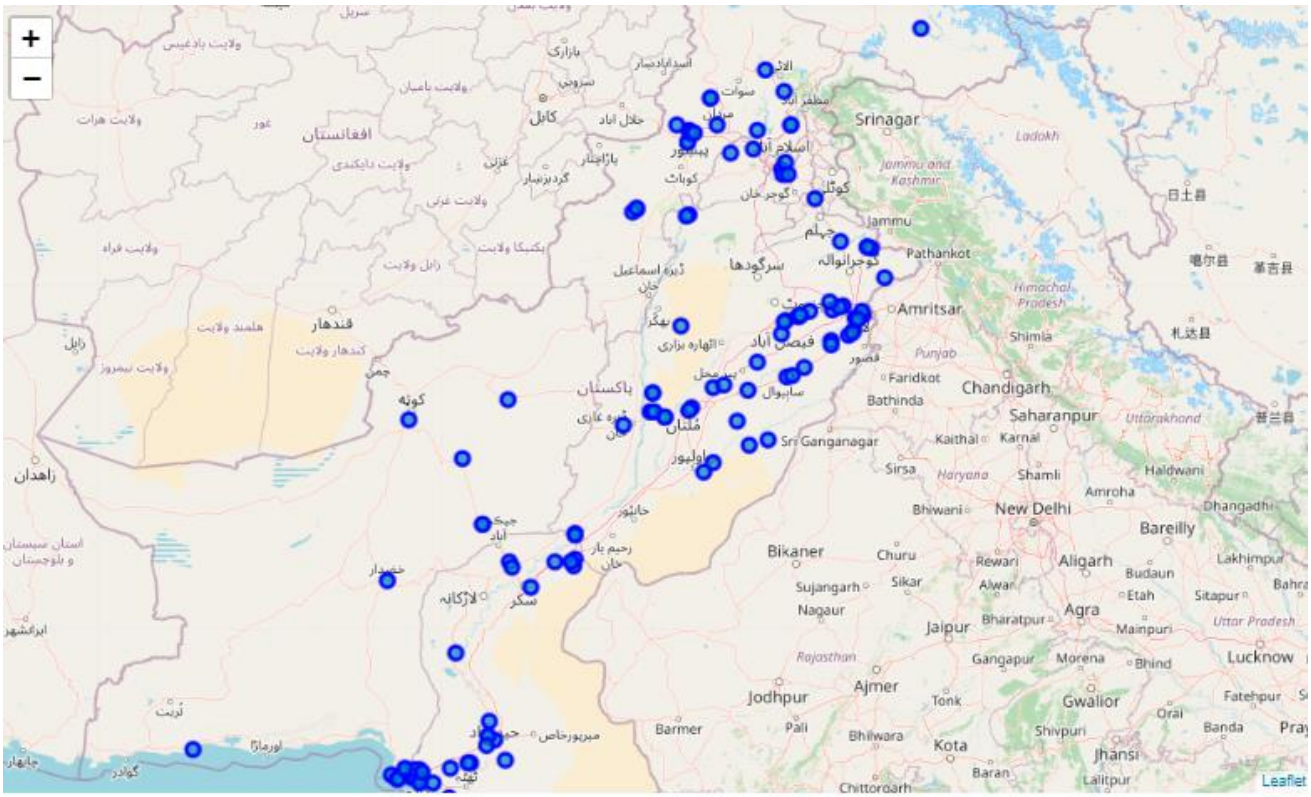
Methodology:

The main dataframe that I got from the Pakistan Transmission Network GEOJSON file contained the geo coordinates, description, Names and descriptions of all the nodes on Pakistan’s Electricity Transmission Network. Out of this dataset I only require the data of major grid stations so I dropped all un-named nodes from my master dataset.

The resulting dataset contained only the required data for 120 grid stations in Pakistan’s transmission network.

	Node name	Description	Latitude	Longitude
0	11KV	"power"=>"line"	33.014250	70.699914
1	500 KV Transmission line	"power"=>"line","cables"=>"3","voltage"=>"5000...	28.430650	69.692011
2	AES Lalpir 362 MW	"power"=>"sub_station","voltage"=>"220000","su...	30.098233	71.162417
3	AES Pak Gen Oil Power Plant	"power"=>"plant","generator:method"=>"thermal"...	30.175593	70.925320
4	Anoud Power Generation	"power"=>"generator","generator:method"=>"ther...	24.846694	67.131269
5	Atlas Power Plant	"power"=>"generator","generator:method"=>"ther...	31.677432	74.081654
6	Attock Generation Oil Power Plant	"power"=>"generator","generator:method"=>"ther...	33.556579	73.072533
7	BROTHA	"power"=>"sub_station"	33.777866	72.251038
8	Bahria Town	"power"=>"substation","voltage"=>"220000","ope...	33.485745	73.097298
9	Balloki Power Plant	"power"=>"plant","generator:method"=>"thermal"...	31.190073	73.887848
10	Bannu	"power"=>"substation","voltage"=>"220000","ope...	32.970531	70.632295

Then I used folium library to visualize the locations of the grid stations by creating a map of Pakistan and superimposing the Grid Station coordinates on the map of Pakistan.

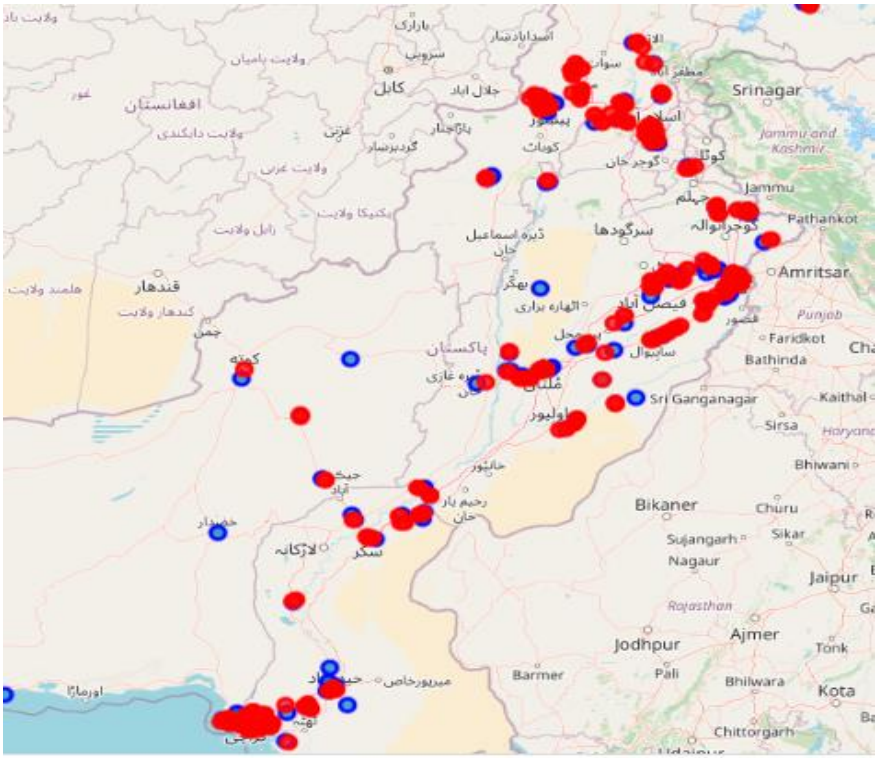


Next I used the Foursquare API to get the number of venues around the Grid stations. The purpose of the exercise was to get the number of venues around the grid stations. This helps us

identify which grid stations are located in or near populated urban centers. The resulting data set of all venues had 2191 rows i.e. the 120 Grid stations had 2192 venues around them.

	Node name	Node Latitude	Node Longitude	Venue	Venue Latitude	Venue Longitude	Venue distance	Venue Category
0	11KV	33.014250	70.699914	karam in hotel Bannu	32.995400	70.621770	7590	Motel
1	500 KV Transmission line	28.430650	69.692011	Daewoo Express Bus Station	28.444069	69.605316	8616	Bus Station
2	500 KV Transmission line	28.430650	69.692011	Guddu toll plaza	28.331517	69.788758	14545	Toll Booth
3	AES Lalpir 362 MW	30.098233	71.162417	Nazar Balauch Doodh Point	30.073277	71.185741	3572	Juice Bar
4	AES Lalpir 362 MW	30.098233	71.162417	shafiq tiles	30.070488	71.136145	3992	Arts & Crafts Store
5	AES Lalpir 362 MW	30.098233	71.162417	Iqra Shopping Centee	30.072492	71.193095	4116	Convenience Store
6	AES Lalpir 362 MW	30.098233	71.162417	Chenab River	30.074881	71.290266	12585	River
7	AES Pak Gen Oil Power Plant	30.175593	70.925320	Coffee Shop MCR	30.170641	70.947067	2164	Coffee Shop
8	Anoud Power Generation	24.846694	67.131269	Pakistan Air Force (PAF) Museum	24.869072	67.094364	4483	History Museum
9	Anoud Power Generation	24.846694	67.131269	Xander's	24.866432	67.077803	5830	Café
10	Anoud Power Generation	24.846694	67.131269	Tutti Frutti	24.874902	67.088258	5360	Frozen Yogurt Shop
11	Anoud Power Generation	24.846694	67.131269	14th Street Pizza	24.910596	67.096607	7928	Pizza Place

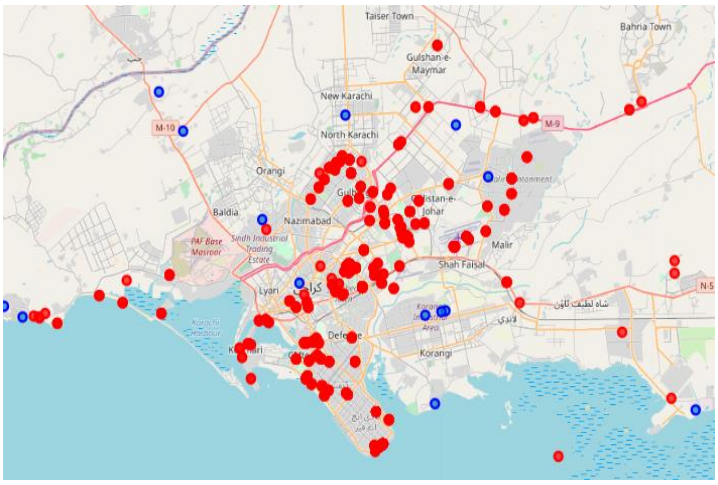
The resulting venues data set also had the important feature of venue distance which shows how far or near the grid is to the surrounding venue. I would later use this attribute to calculate the mean distance of each grid station from its surrounding venue cluster. But for now let's plot the venues on the aforementioned map to visualize the venue data and see which grid stations have higher venue density and which have lower number of clusters surrounding them.



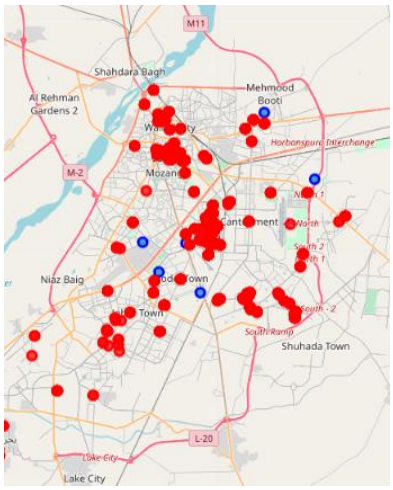
We can see from the map that the Grid Stations located in the densely populated cities of Pakistan have the greates venue clusters around them. Look at the dense red clusters around Karachi, Lahore and Islamabad!

Let's have a look at Karachi, the most populated city of Pakistan. Or Lahore which is the 2nd largest city by population.

Karachi Map



Lahore Map



Compared to cities like Quetta or Peshawar these

cities are very densely populated.

Next I use the group_by clause to get the total number of venues around each grid station and add it to the master data frame.

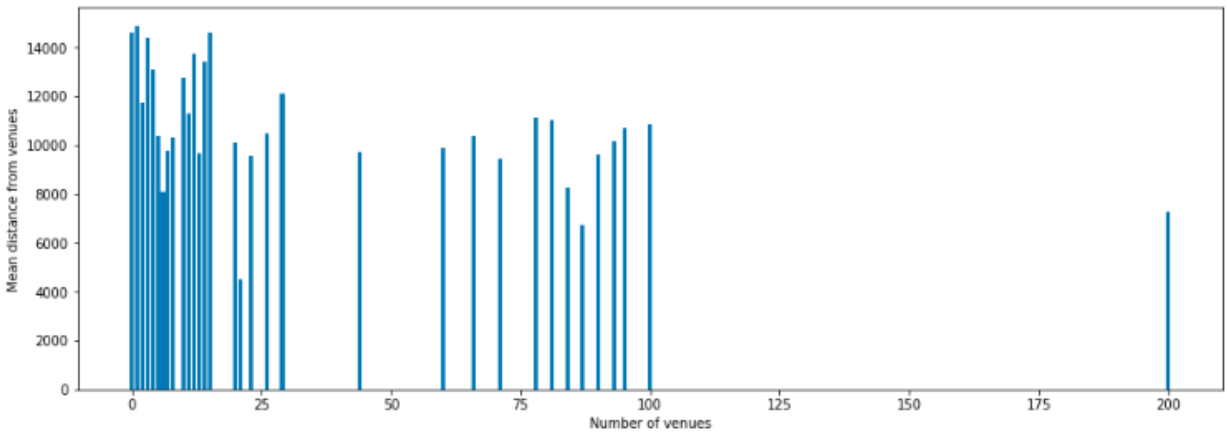
Then I calculated the mean distance for all grid stations having venues around them and added it to the main dataset. It was observed that df_g which shows number of venues per grid station had only 107 rows. Which means that13 Grid station do not have any venues in 15 km radius. I assigned 0 to the remaining grid stations venues. Also no venues mean no mean distance from venues. For the sake of analysis, I assigned those nodes the max mean distance.

The resulting dataset looked something like this

level_0	Node name	Description	Latitude	Longitude	No of Venues	Mean distance
0	11KV	"power"=>"line"	33.014250	70.699914	1	7590.000000
1	500 KV Transmission line	"power"=>"line","cables"=>"3","voltage"=>"5000...	28.430650	69.692011	2	11580.500000
2	AES Lalpir 362 MW	"power"=>"sub_station","voltage"=>"220000","su...	30.098233	71.162417	4	6066.250000
3	AES Pak Gen Oil Power Plant	"power"=>"plant","generator:method"=>"thermal"...	30.175593	70.925320	1	2164.000000
4	Anoud Power Generation	"power"=>"generator","generator:method"=>"ther...	24.846694	67.131269	81	10183.716049
5	Atlas Power Plant	"power"=>"generator","generator:method"=>"ther...	31.677432	74.081654	4	11408.500000
6	Attock Generation Oil Power Plant	"power"=>"generator","generator:method"=>"ther...	33.556579	73.072533	87	6708.804598
7	BROTHA	"power"=>"sub_station"	33.777866	72.251038	4	11629.500000
8	Bahria Town	"power"=>"substation","voltage"=>"220000","ope...	33.485745	73.097298	71	9445.830986
9	Balloki Power Plant	"power"=>"plant","generator:method"=>"thermal"...	31.190073	73.887848	4	4165.750000
10	Bannu	"power"=>"substation","voltage"=>"220000","ope...	32.970531	70.632295	2	1642.000000
11	Bannu 132 KV	"power"=>"substation","voltage"=>"132000"	33.019101	70.714575	1	9056.000000

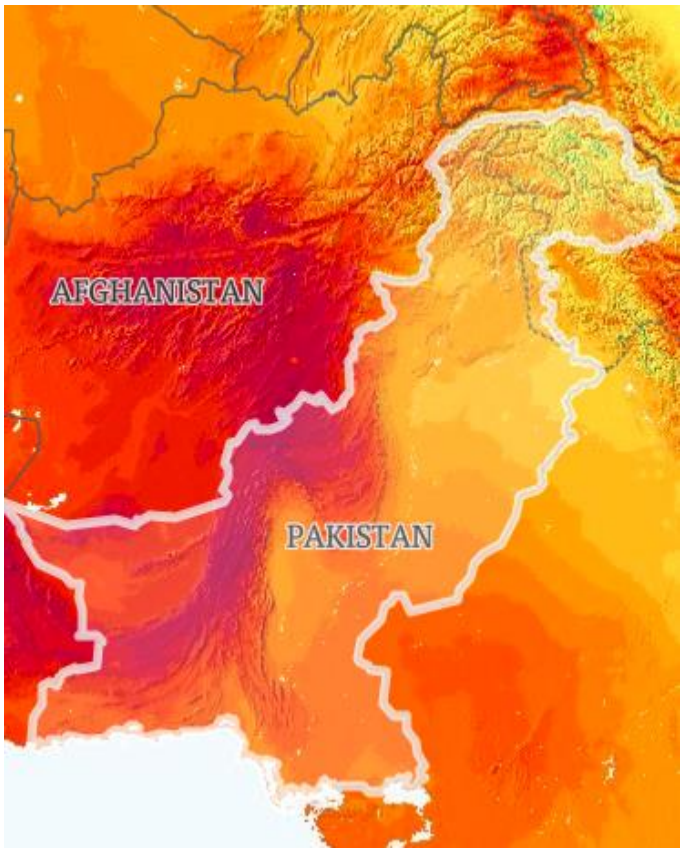
The dataset has Node names, description, Latitude, Longitude, No of Venues(nearby) and Mean distance(from venues).

Let’s plot the mean distance Vs the number of nodes and observe their relationship



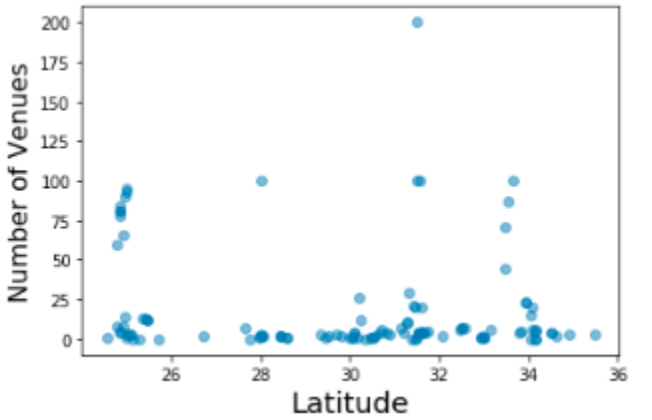
We can see here that the nodes with lower number of venues surrounding them have higher mean distances from the venues

Now in the next step we would add the Specific Photovoltaic Power Output for these grid stations. Unfortunately there is no dataset available for these values so I had no choice but to make one by entering the coordinates of the grid stations one by one into Global Solar Atlas and getting their PVOUT values. You can also check out data related to solar radiation such as PVOUT, DNI and GHI values for your hometown by searching it. Following is the link https://globalsolaratlas.info/map?r=PAK&c=30.626947,68.99585,5. I add all the PVOUT values to the final dataframe.



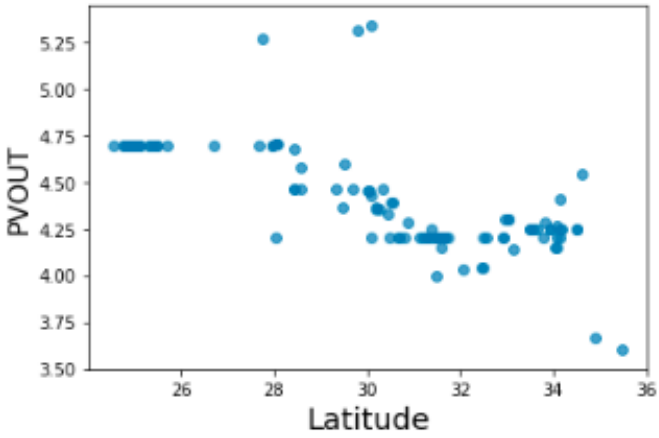
Let's also look at the solar irradiation map of Pakistan. What we can observe from the two maps of Pakistan is that the regions with the highest Direct Normal Irradiance (DNI) have very few grid stations.

Also we get very interesting observations about the way Pakistan's population density vary if we plot the Number Of Venues against Latitude



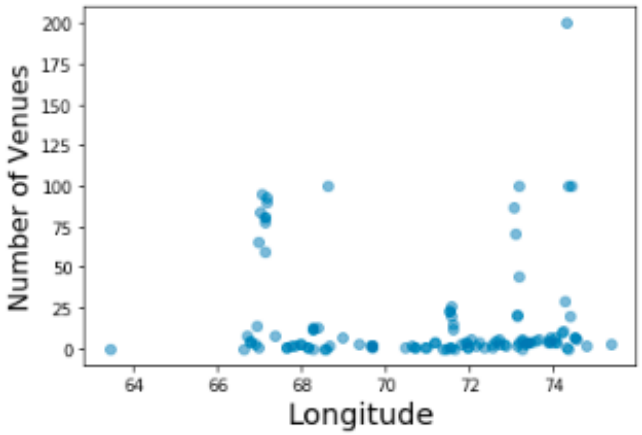
Observe that there is a greater concentration of nodes and their venues in the norther half of the country.
This is because the Northern half of Pakistan is more populated than its Southern half

But the PVOUT values show an opposite trend. We can observe that the PVOUT values decrease with increasing latitude. Hence location in the south of Pakistan receive more sunlight than locations in the north.

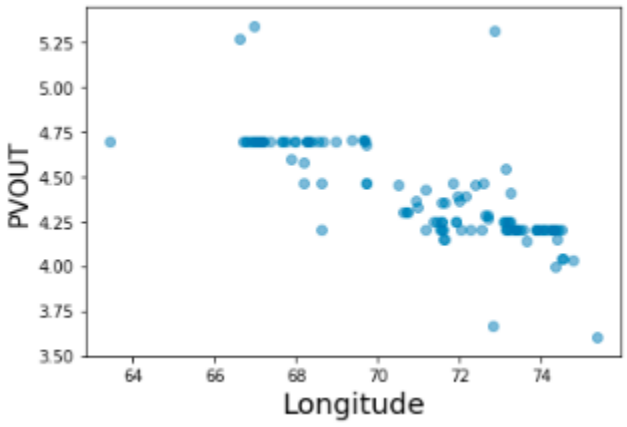


Let's also observe these two attributes in comparison with longitude.

We can see that the number of locations and venues are more concentrated along the eastern boundary of Pakistan

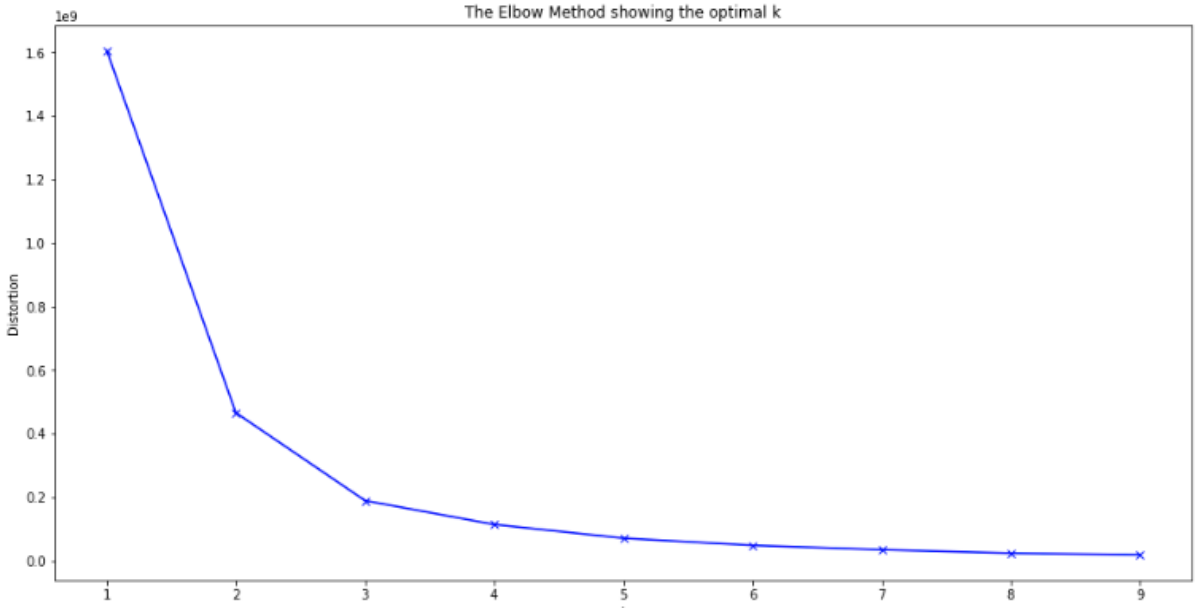


Once again the trend shows that the PVOUTs for locations in the western half of Pakistan are higher than for the locations in the eastern half



Next I apply the K-Means clustering algorithm to cluster the dataset. For that I first decide on how many clusters should be used to get optimum results by using the elbow method which shows distortion against the number of clusters for a dataset.

The plot shows that beyond a K value of 3 there is no significant improvement in the distortion value. So I am going to go with 3 clsuters for our dataset.



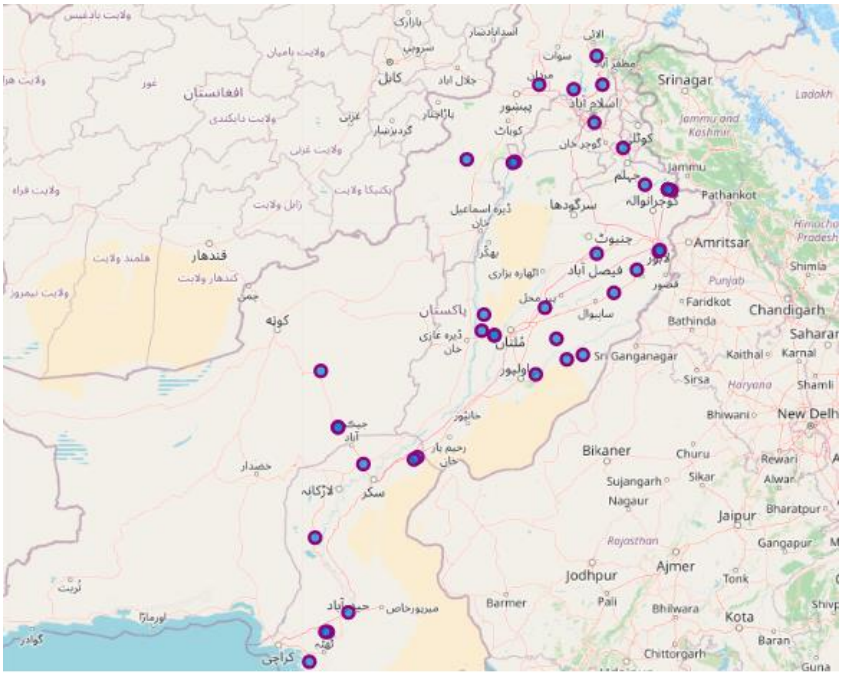
Results

After generating Labels, I add them to the main dataset and it looks like this

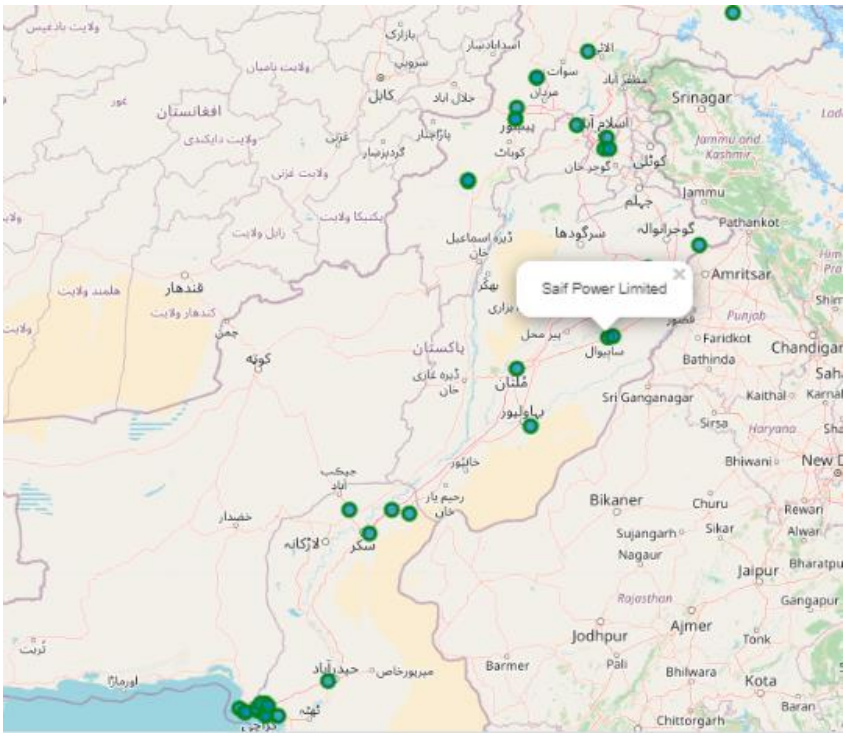
	Node name	Description	Latitude	Longitude	No of Venues	Mean distance	PVOUT	Clus_km
0	11KV	"power"=>"line"	33.014250	70.699914	1	7590.000000	4.300	1
1	500 KV Transmission line	"power"=>"line","cables"=>"3","voltage"=>"5000...	28.430650	69.692011	2	11580.500000	4.460	2
2	AES Lalpir 362 MW	"power"=>"sub_station","voltage"=>"220000","su...	30.098233	71.162417	4	6066.250000	4.425	0
3	AES Pak Gen Oil Power Plant	"power"=>"plant","generator:method"=>"thermal"...	30.175593	70.925320	1	2164.000000	4.367	0
4	Anoud Power Generation	"power"=>"generator","generator:method"=>"ther...	24.846694	67.131269	81	10183.716049	4.690	1
5	Atlas Power Plant	"power"=>"generator","generator:method"=>"ther...	31.677432	74.081654	4	11408.500000	4.200	1
6	Attock Generation Oil Power Plant	"power"=>"generator","generator:method"=>"ther...	33.556579	73.072533	87	6708.804598	4.250	0
7	BROTHA	"power"=>"sub_station"	33.777866	72.251038	4	11629.500000	4.200	2
8	Bahria Town	"power"=>"substation","voltage"=>"220000","ope...	33.485745	73.097298	71	9445.830986	4.250	1
9	Balloki Power Plant	"power"=>"plant","generator:method"=>"thermal"...	31.190073	73.887848	4	4165.750000	4.200	0
10	Bannu	"power"=>"substation","voltage"=>"220000","ope...	32.970531	70.632295	2	1642.000000	4.300	0

Let’s plot the 3 clusters on a map.

CLUSTER 0:



CLUSTER 1:



CLUSTER 2:

Discussion

Let's observe the mean values of our generated clusters to get a sense of what these clusters essentially represent.

CLUSTER 0:

No of Venues	8.526316
Mean distance	4492.825195
PVOUT	4.406816
Clus_km	0.000000

CLUSTER 1:

No of Venues	34.042553
Mean distance	9495.860348
PVOUT	4.369787
Clus_km	1.000000

CLUSTER 2:

No of Venues	4.375000
Mean distance	13449.408687
PVOUT	4.448531
Clus_km	2.000000

This means that Cluster 2 grid stations have the best PVOUT while having lowest number of Venues and greatest distances from those venues.

Since this study covers a vast geographic area (Pakistan!), the model can be further improved by using data such as district population, land prices, terrain, ambient and meteorological data etc. because those features also effect the costs, generation, feasibility and ROIs of Large scale power plants.

Conclusions

Pakistan's solar market needs more investment and investors. As an energy professional I have observed a lot of activity in solar power projects at enterprise level (Captive level) but we need more interest in Solar IPPs if we have to realize the goals set out by the new Renewable Energy Development Policy. I hope that this study becomes a conversation starter about a data driven approach to Pakistan's energy issues. This study can also be replicated for Wind power plants which are a relatively new phenomenon in Pakistan's energy landscape.