Target Solar Panel Market in Michigan

Team Members:
Eleazar Luis Escalante
Safari NDE
Kimberly Wagner-Dabbour
Kevin Paree

MSU Data Analytics Final Project

Solar Panels - Next Big Boom?

Overview:

Our team is interested in discovering our target market for residential solar panels in specific regional Michigan cities. With solar panels being the cheapest form of clean and renewable energy, we are looking to focus on areas with ideal solar radiation and a minimum base income to ensure that solar panels are cost effective for the customer.

Questions:

- How does education level affect solar panel ownership?
- Does home ownership affect the level of solar panel ownership?
- What is the average number of Solar Panel Systems per region?
- Do highly populated regions have more Solar Systems?

Problem Statement

Renewable energy accounted for about 11% of Michigan's total in-state electricity net generation in 2021. 60% of Michigan's renewable electricity comes from wind. Only .64% of Michigan’s electricity comes from Solar.

+

With the increasing interest and incentives from the federal government to increase the use and generation of renewable energy, there is a need to develop a model that will help predict the residential solar system count in Michigan.

→

Using data analytics tools, and a machine learning model to build a prediction model of residential solar panel systems count in Michigan based on demographic trends.

Source Data Description

Data:

- Average Household Income MI - Census Bureau
- Homeownership rate in MI - Census Bureau
- Renewable Energy Sector in MI - Census Bureau
- NREL National Solar Radiation Database 2015 - 2020
- DeepSolar by Stanford - 2018

Examination of the Datasets

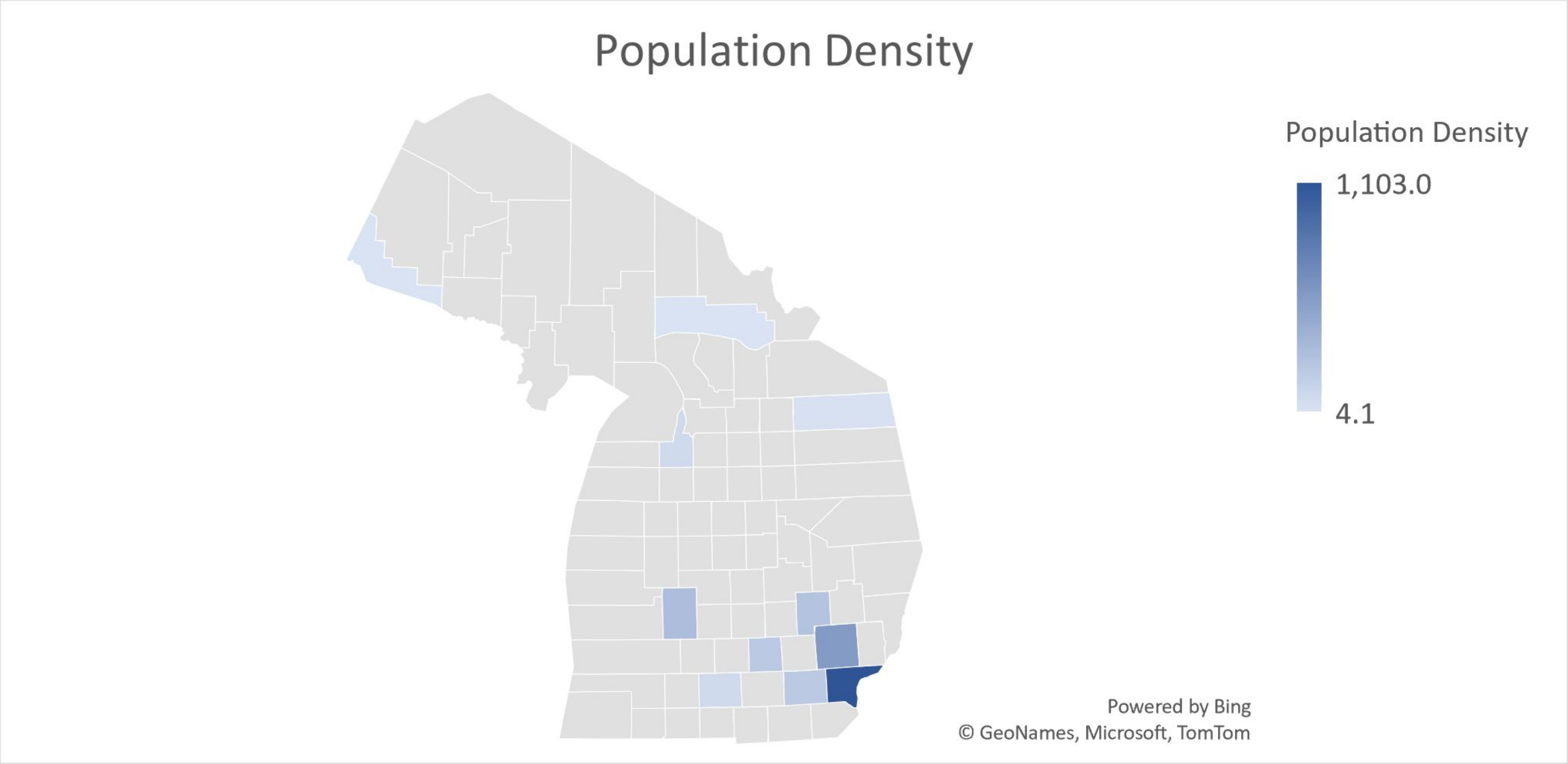
- We would test numerous different models with different variables to detect the p-values of each independent variable and use the r-squared value to find the effectiveness of the model.
- The counties included in the research are:
  - Alpena County
  - Calhoun County
  - Genesee County
  - Gogebic County
  - Grand Traverse County
  - Ingham County
  - Mackinac County
  - Oakland County
  - Washtenaw County
  - Wayne County
- We hypothesized that either Oakland County, due to its higher income per household, or Washtenaw County, due to higher education levels, would be the most pivotal factor in solar adaptation.

Variables used in OLS Regression

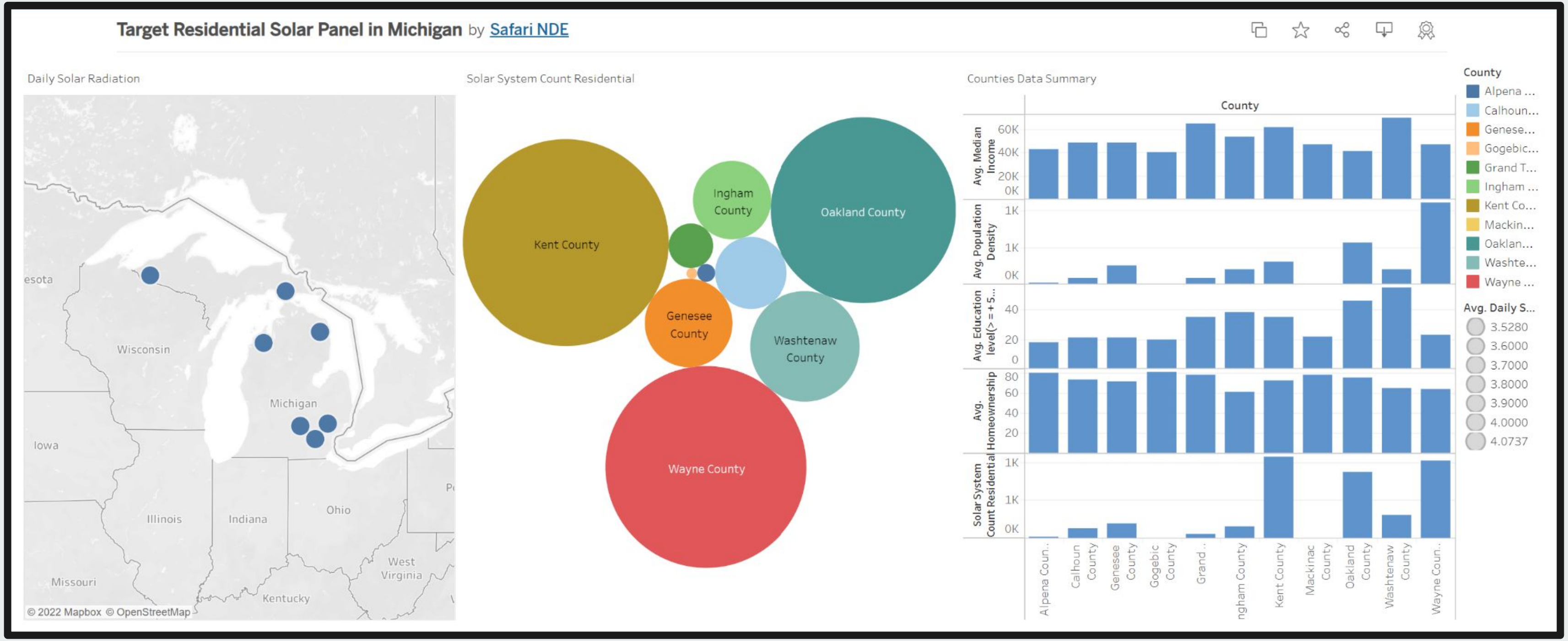
Tile Count	LAT
Solar Panel System Count	LON
Total Panel Area	Daily Solar Radiation
FIPS	Number of Solar Panel Systems per Household
Solar Panel Area Divided by Area	Median Income
Solar Panel Area Per capita	Education Level
Tile Count Residential	Home Ownership
Total Panel Area Residential	Population Density

Exploratory Data Analysis

What is the Population Density ?



Interactive Dashboard



https://public.tableau.com/app/profile/safari.nde/viz/TargetResidentialSolarPanelinMichigan/TargetSolarPanelMarketinMI?publish=yes

Correlation Analysis

Based on the correlation level, the Solar System Count Residential is positively correlated with Median income, educational level, and homeownership. This means that if one of these variables decreases the solar system count residential will decrease. The population density is negatively correlated with the Residential Solar System count this indicates that if the population density increase the Residential Solar system count will decrease.

	solar_system_count_resi	panel_area_reside	lat	lon	ily_solar_radiatic	solar_system_per	Median Income	tion level(>=+5)	Homeownership	opulation Densit	tile_count	olar_system_count	total_panel_area	nel_area_divided	panel_area_per	e_count_residential
solar_system_count_residential	1															
total_panel_area_residential	0.96	1														
lat	-0.03	-0.07	1													
lon	0.26	0.26	0.06	1												
daily_solar_radiation	0.04	0.08	-0.93	-0.16	1											
number_of_solar_system_per_househc	0.11	0.10	-0.06	-0.01	0.05	1										
Median Income	0.22	0.22	-0.01	0.08	0.09	0.06	1									
Education level(>=+5 years)	0.20	0.17	0.19	0.02	-0.17	0.04	0.25	1								
Homeownership	0.15	0.12	0.07	-0.05	0.05	-0.01	-0.15	0.38	1							
Population Density	-0.17	-0.15	0.00	0.13	-0.17	-0.06	-0.46	-0.44	-0.60	1						
tile_count	0.39	0.43	-0.10	0.17	0.11	0.04	0.10	0.04	0.03	-0.04	1					
solar_system_count	0.80	0.82	-0.07	0.23	0.08	0.08	0.18	0.14	0.11	-0.13	0.81	1				
total_panel_area	0.20	0.25	-0.10	0.13	0.11	0.02	0.06	0.01	0.01	-0.01	0.96	0.66	1			
solar_panel_area_divided_by_area	0.45	0.45	0.04	0.10	-0.06	0.05	0.15	0.17	-0.01	-0.05	0.11	0.31	0.03	1		
solar_panel_area_per_capita	0.06	0.06	-0.09	-0.01	0.08	0.85	0.03	0.00	0.02	-0.05	0.02	0.04	0.01	0.02	1	
tile_count_residential	0.99	0.97	-0.04	0.29	0.05	0.10	0.22	0.19	0.14	-0.16	0.41	0.81	0.23	0.44	0.05	1

Machine Learning Model - Ordinary Least Squares

- Preliminary investigation sought three distinct datasets -
  - household data (income, education, work experience, etc),
  - solar irradiance data for Michigan
  - dataset of household solar panel installations.

- First two were found without difficulty (ACS and NREL), household solar panel installations data was found from a deep learning dataset created from Stanford using satellite imagery.

- Machine learning model would be a ordinary least squares linear regression.

In [609]:

```
vif=pd.DataFrame()
vif['Features']=X.columns
vif['VIF']=[variance_inflation_factor((X.values).astype(float), i).astype(float) for i in range(X.shape[1])]
vif['VIF']=round(vif['VIF'],2)
vif=vif.sort_values(by="VIF", ascending=False)
vif
```

Out[609]:

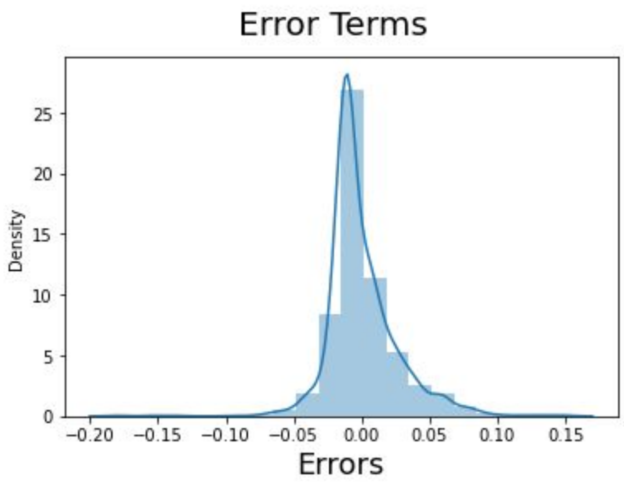
	Features	VIF
3	education_level	2.84
4	homeownership	2.72
2	total_panel_area_residential	1.77
1	solar_panel_area_divided_by_area	1.40
0	total_panel_area	1.19

In [610]:

```python
import seaborn as sns
y_train_solar_system_count_residential=lr_12.predict(X_train_lm)
fig=plt.figure()
sns.distplot((y_train-y_train_solar_system_count_residential),bins=20)
fig.suptitle('Error Terms', fontsize=20)
plt.xlabel('Errors', fontsize=18)
```

```
/Users/eleazarescalante/opt/anaconda3/lib/python3.9/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[610]: Text(0.5, 0, 'Errors')



```
In [612]:
y_test = df_test.pop('solar_system_count_residential')
X_test = df_test
X_test_m4=sm.add_constant(X_test)
X_test_m4 = X_test_m4.drop(['lon','lat','tile_count','tile_count_residential','fips','solar_system_count','daily
y_pred_m4=lr_12.predict(X_test_m4)
```

```
/Users/eleazarescalante/opt/anaconda3/lib/python3.9/site-packages/statsmodels/tsa/tsatools.py:142: FutureWarning: I
n a future version of pandas all arguments of concat except for the argument 'objs' will be keyword-only
  x = pd.concat(x[::order], 1)
```

```
In [621]:
from sklearn.metrics import r2_score
r2_score=r2_score(y_true=y_test, y_pred = y_pred_m4)
print("R-Squared Value: " + str(r2_score))
```

```
R-Squared Value: 0.929661012296137
```

```
In [617]:
from sklearn.metrics import mean_absolute_error as mae
error=mae(y_true=y_test,y_pred=y_pred_m4)
print("Mean Absolute Error : " + str(error))
```

```
Mean Absolute Error : 0.019702950098455466
```

```
In [618]:
from sklearn.metrics import mean_squared_error
error2=mean_squared_error(y_true=y_test, y_pred=y_pred_m4)
print("Mean Squared Error: " + str(error2))
```

```
Mean Squared Error: 0.0008856228995905903
```

```
In [619]:
from sklearn.metrics import mean_squared_error
root_mean_squared=mean_squared_error(y_true=y_test, y_pred=y_pred_m4, squared=False)
print("Root Mean Squared: " + str(root_mean_squared))
```

```
Root Mean Squared: 0.029759416990098955
```

Results

- 16 variables were included in our analysis to help provide insight into which factors affected the likelihood of someone installing solar panels living in Michigan.
- The variables with the highest correlation to determining our target markets were:

	Features	VIF
3	education_level	2.84
4	homeownership	2.72
2	total_panel_area_residential	1.77
1	solar_panel_area_divided_by_area	1.40
0	total_panel_area	1.19

Results Cont.

```
                                 OLS Regression Results
==============================================================================
Dep. Variable:     solar_system_count_residential   R-squared:                       0.919
Model:                                        OLS   Adj. R-squared:                  0.918
Method:                             Least Squares   F-statistic:                     2800.
Date:                            Tue, 23 Aug 2022   Prob (F-statistic):               0.00
Time:                                    00:50:31   Log-Likelihood:                 2773.9
No. Observations:                            1246   AIC:                            -5536.
Df Residuals:                                1240   BIC:                            -5505.
Df Model:                                       5
Covariance Type:                        nonrobust
==============================================================================================
                                    coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                             0.0077      0.001      5.295      0.000       0.005       0.011
total_panel_area              -1.812e-06   3.02e-07     -6.000      0.000    -2.4e-06   -1.22e-06
solar_panel_area_divided_by_area 2.549e-05 1.06e-05      2.409      0.016    4.73e-06    4.62e-05
total_panel_area_residential      0.0009   9.43e-06     99.435      0.000       0.001       0.001
education_level                   0.0092      0.003      3.266      0.001       0.004       0.015
homeownership                     0.0066      0.003      2.079      0.038       0.000       0.013
==============================================================================
Omnibus:                      255.350   Durbin-Watson:                   1.930
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2331.295
Skew:                           0.673   Prob(JB):                         0.00
Kurtosis:                       9.565   Cond. No.                     1.30e+04
==============================================================================
```

Findings

$$Y = 0.0077 - 1.812e\text{-}06\ X1 + 2.549e\text{-}05\ X2 + 0.009X3 + 0.0092\ X4 + 0.0066\ X5$$

- X1= Total Panel Area
- X2 = Solar Panel area divided by area
- X3 = Residential total panel area
- X4 = Education level (=>Bachelor Degree)
- X5 = Homeownership

- **Education and Home Ownership** were clearly the largest factors in targeting regions for solar panel installation.

Recommendations for Future Analysis

- Collect More Detailed Data:
  - Credit Data
  - Acceptance Data
- Look at external market factors
  - What areas are already highly targeted?
  - What options are there for purchasing the panels?
- More Incentives by the state and federal government
- A database for the number of solar panel installation over the year