Energy Consumption and Renewable Energy

Data Science Capstone Milestone 5

Angelene Arito
Xiaoxiao Deng
Steven Jordan
Huy Tran
Justin Winfield

Background

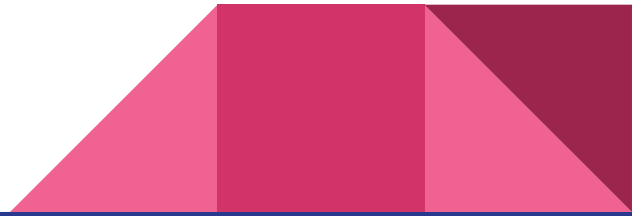
- Power City, USA is looking for ways to increase its reliability on sustainable energy from renewable sources.
- They need to understand their electricity production from photovoltaic solar and wind sources and their fluctuations
- Also need to understand electricity consumption across the different sectors as well as with respect to increase ownership of electric vehicles

Business Objectives

- Build predictive model(s) to calculate
  - Electrical energy production from wind
  - Electrical energy production from photovoltaic solar
  - Electrical energy consumption for Power City’s 8 sectors
    - Food Service, Grocery, Health Care, Education (K-12), Lodging, Office, Residential, and Stand-Alone Retail
- Provide recommendations to compensate for the consumption increase due to electric vehicles

Business Objectives

- Forecast the values for a set of diverse days in a scenario year:
  - March 15
  - June 26
  - July 3
  - October 13
  - November 19
  - December 25

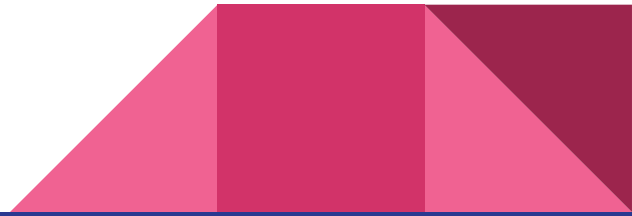


Original Data

Wind	Solar	Consumption	Scenario	Electrical Car
- Production (kW/hr) - Wind speed	- Production (kW/hr) - Solar angle - Weather conditions	- Consumption (kW/sq.ft.) - Weather conditions - Solar angle - Sector use (sq.ft./person), - Population (by age) - Calendar days	- Weather conditions - Calendar days	- Charging (kW/sq.ft.)

Methodology

- Pre-Processing
- Exploratory Analysis
- Model Building
- Results
- Conclusion
- Recommendation



Pre-Processing

Pre-processing

- 14 separate data files
- Converted “HolidayName” to Boolean (2 files)
- Converted Hour column from 1-24 scale to 0-23 (7 files)
- Added a date_time column by concatenating date and time columns (12 files)
- Wind_Speed was binned to 14 bins of 1.5 width (2 files)

Pre-processing | Missing Data

powercity_weather_scenario.csv:

- Cloud_Cover_Fraction: 2
- Dew_Point: 26
- Humidity_Fraction: 26
- Precipitation: 369
- Pressure: 625
- Temperature: 25
- Visibility: 2
- Wind_Speed: 4

solararray_weather.csv:

- Cloud_Cover_Fraction: 191
- Dew_Point: 270
- Humidity_Fraction: 270
- Precipitation: 12590
- Pressure: 15342
- Temperature: 150
- Visibility: 142
- Wind_Speed: 184

- For most missing values, we applied the median value of the attribute for the 2 weeks before and 2 weeks after the missing value

- For solararray_weather.csv, nearly all Precipitation and Pressure values were missing for 2013-2014. We used the mean of those values for their Corresponding time in the previous years.

Pre-processing | Merging Data

Sector_Use_Matrix and Powercity_Population do not have predictive data. These files were used to determine the size (in square feet) of each sector, which will be used in calculating the total consumption of the city.

Residential	84,832,407	sq ft
Office	9,999,731	sq ft
K-12	2,426,480	sq ft
Food Service	967,558	sq ft
Health Care	965,374	sq ft
Lodging	931,129	sq ft
Grocery	539,981	sq ft

Pre-processing | Merging Data

After individual file pre-processing, files were then merged into four files based on their category and relevant model:

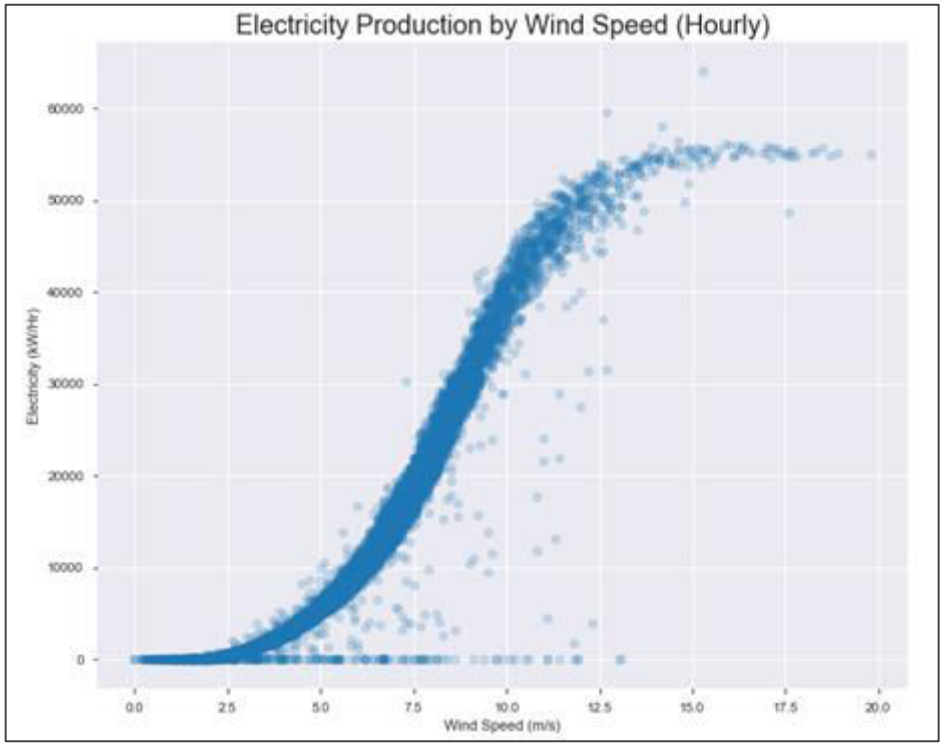
- **Power City Consumption:** 70,072 records
  - 19 features (6 Interval, 5 Nominal, 2 Boolean, 6 Ratio)
  - Used to build model to predict Power City Consumption
- **Solar Array Production:** 18,692 records
  - 17 features (9 Interval, 1 Nominal, 1 Ordinal, 6 Ratio)
  - Used to build model to predict Solar Energy Production
- **Wind Farm Production:** 15,385 records
  - 10 features (6 Interval, 1 Nominal, 1 Ordinal, 2 Ratio)
  - Used to build model to predict Wind Energy Production
- **Power City Scenario:** 8736 records
  - 21 features (7 Interval, 5 Nominal, 2 Boolean, 1 Ordinal, 6 Ratio)
  - Used to forecast the above models

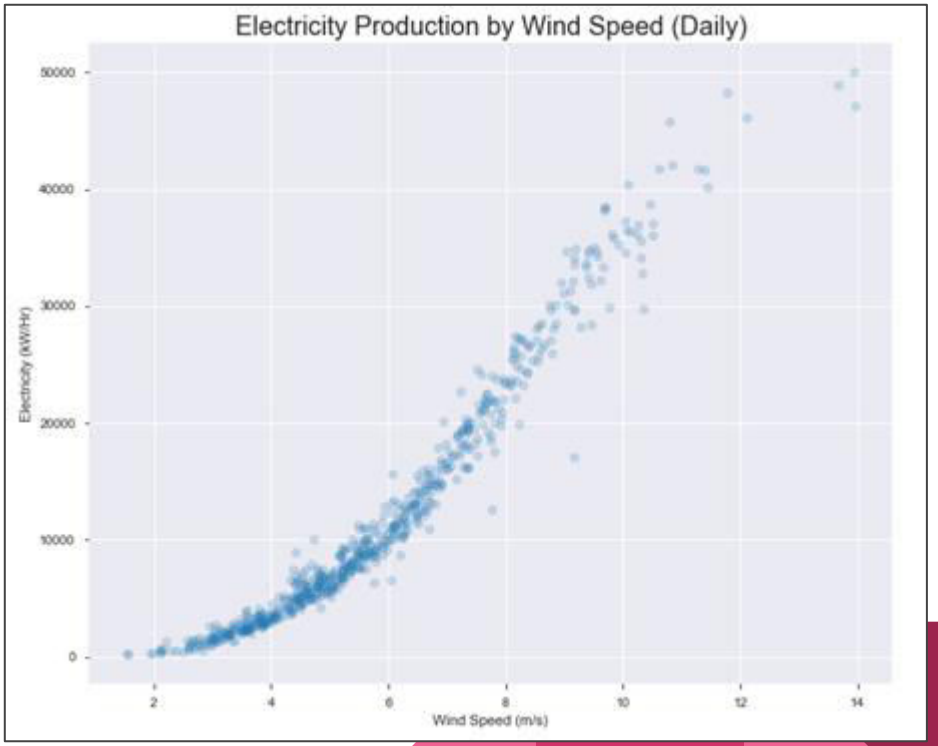
Preprocessing | Feature Handling & Train/Test

- Data was aggregated daily using mean values
- Dummy variables were generated based on certain calendar features
  - Month
  - Day
  - Day of Week
- Other calendar features were added
  - Weekend
  - Season
- Consumption file was broken into 8 separate files based on sector
- All 10 files were then split into 80/20 train/test
- Lasso feature selection was performed on all 10 training sets

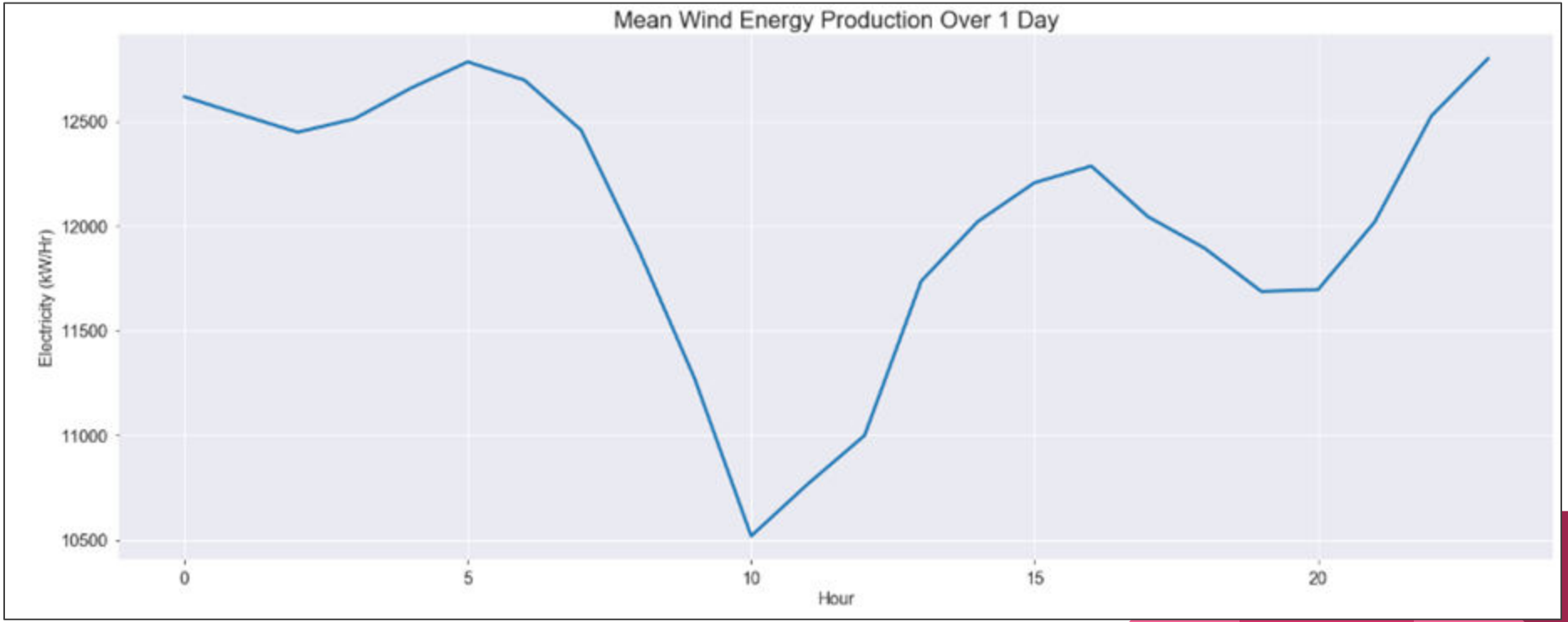
Exploratory Analysis

Exploratory Analysis | Wind Production

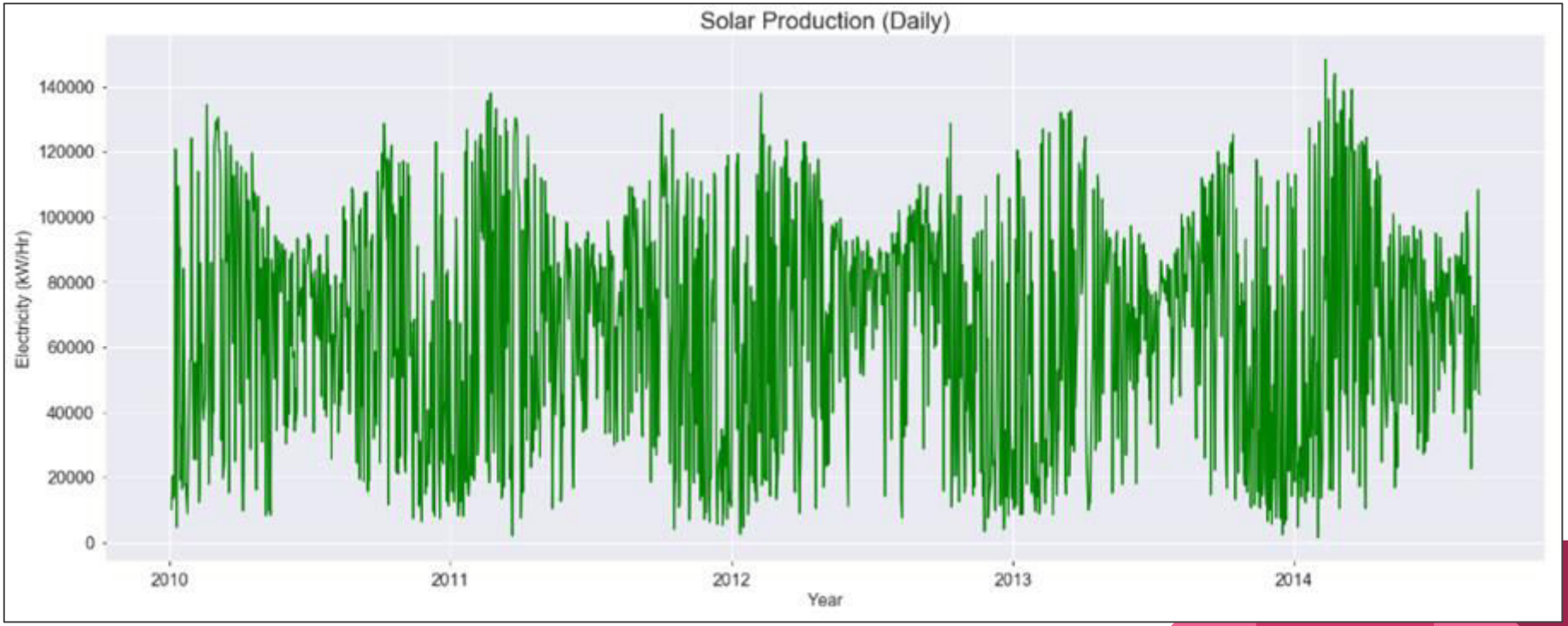




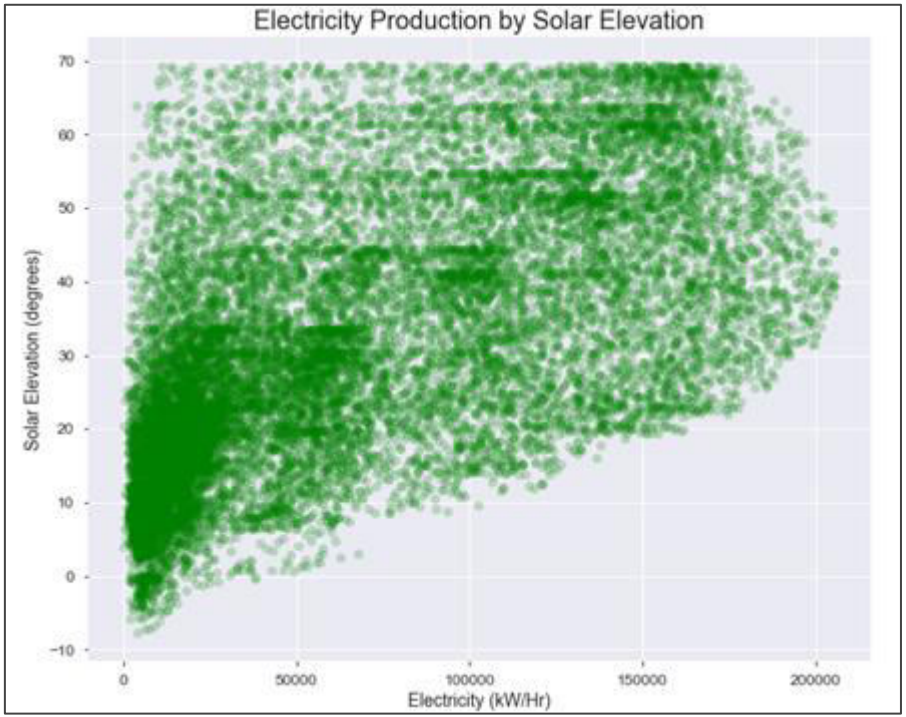
Exploratory Analysis | Wind Production

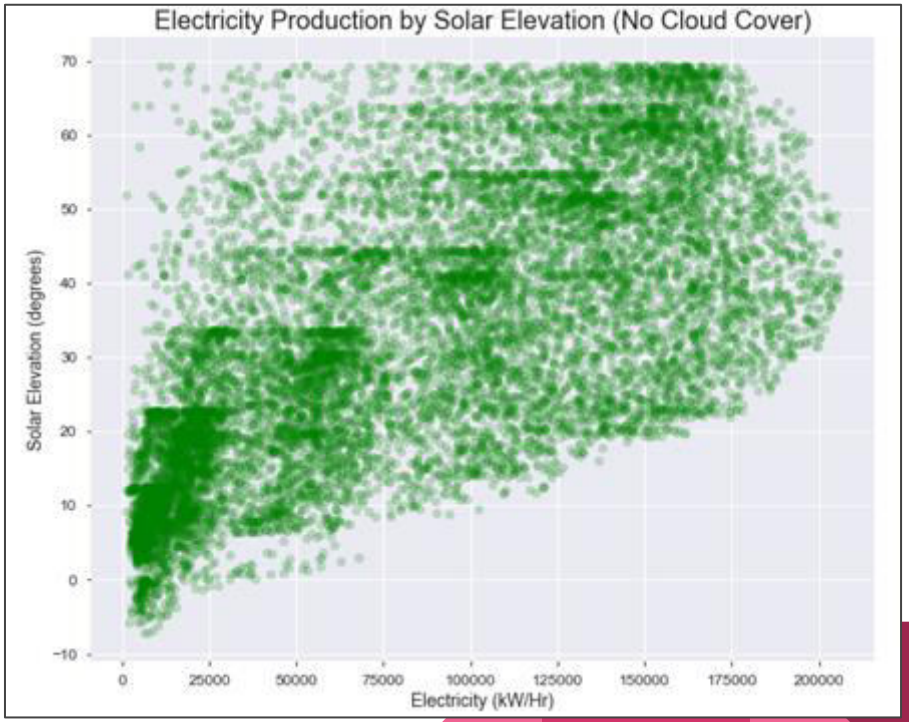


Exploratory Analysis | Solar Production

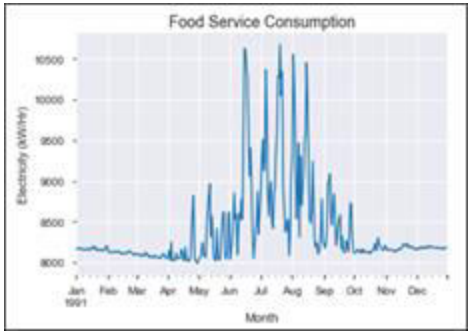


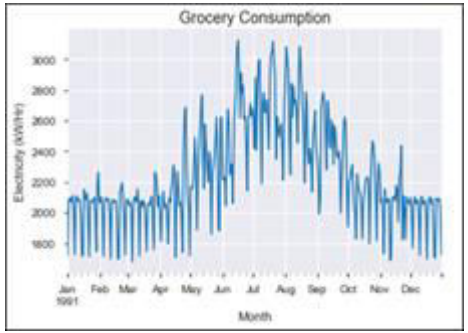
Exploratory Analysis | Solar Production

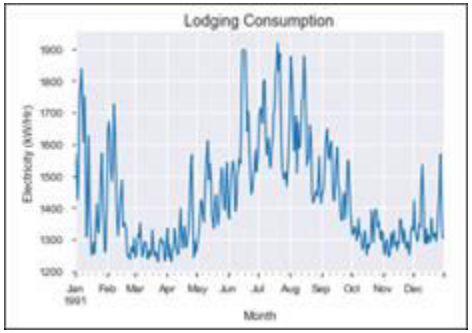


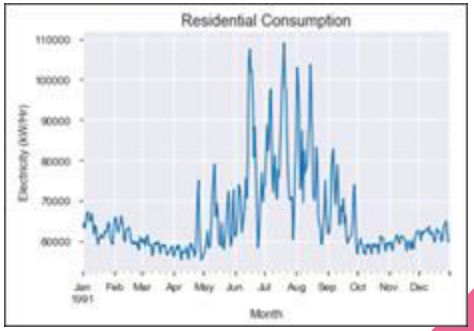


Exploratory Analysis | Consumption by Sector

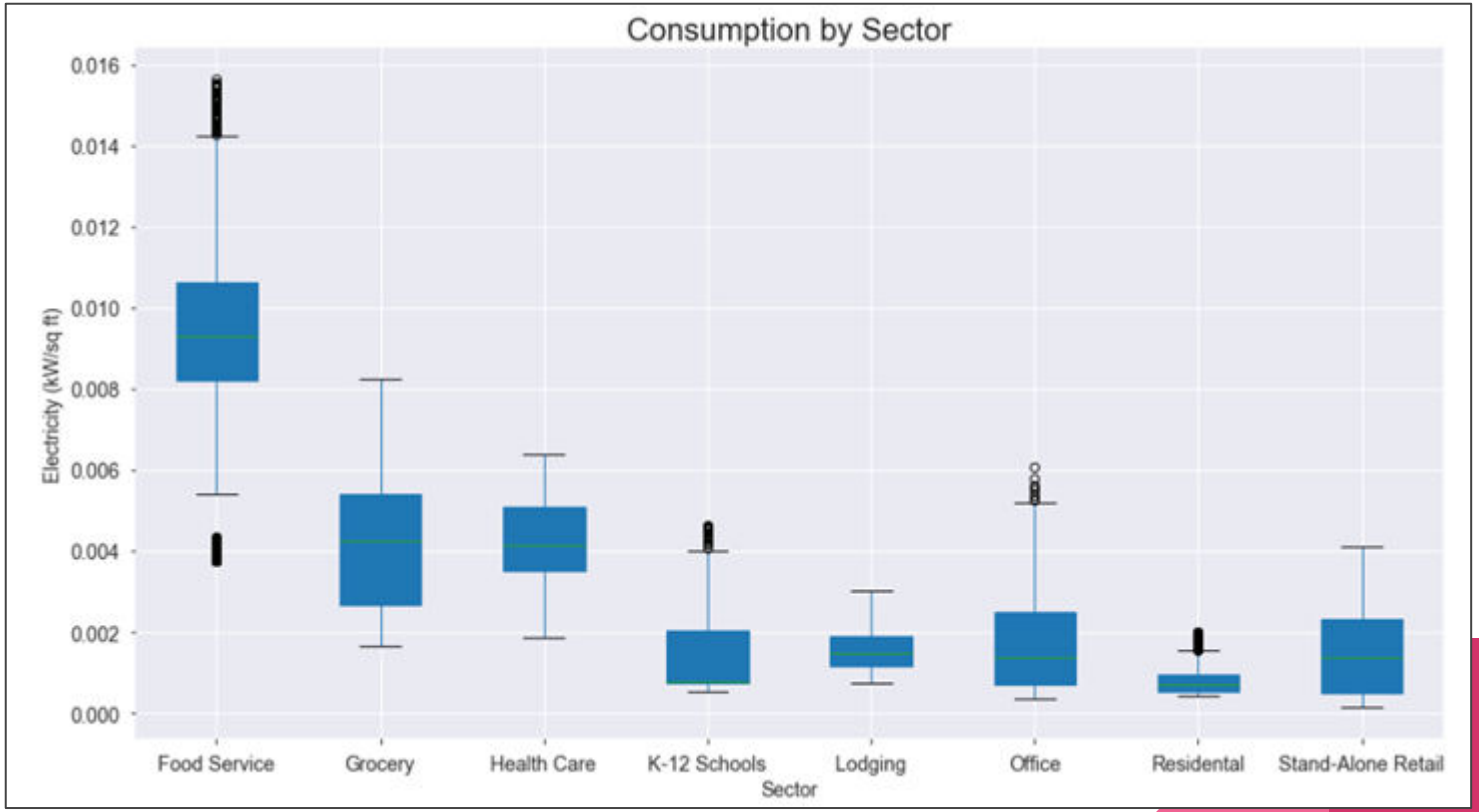








Exploratory Analysis | Consumption by Sector

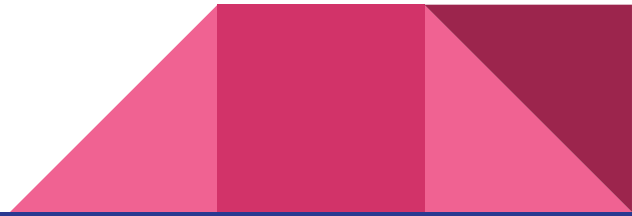


Model Building

Model Building | Overview

Various algorithms used to predict energy production and consumption

- AdaBoost
- Decision Trees
- Decision Trees + Bagging
- Gradient Boosting
- Multiple Linear Regression
- Neural Network (Keras)
- Neural Network (scikit-learn)
- Random Forest
- Support Vector Machine
- Time Series
- XGBoost



Model Building | Evaluation Chart

	Production				Consumption													
	Wind		Solar		Food Service		Grocery		Healthcare		K12		Lodging		Office		Residential		Retail	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Time Series-X	4607	0.85	19642	0.69	261	0.00	79	0.81	330	0.80	311	0.89	62	0.10	3423	0.77	7412	0.00	124	0.72
Neural Network	2064	0.96	18405	0.72	231	0.76	60	0.97	155	0.93	273	0.93	61	0.83	2257	0.87	2180	0.95	82	0.92
Multilinear Regression	2802	0.92	18033	0.63	244	0.70	76	0.95	158	0.92	307	0.92	66	0.80	2403	0.83	4620	0.77	94	0.87
Gradient Boosting	1701	0.97	17082	0.76	66	0.98	**42**	**0.99**	**99**	**0.97**	168	0.97	41	0.93	1868	0.91	1278	0.98	41	0.98
AdaBoosting	**1550**	**0.98**	18511	0.72	81	0.97	75	0.95	159	0.92	159	0.98	41	0.93	2169	0.88	1644	0.97	**36**	**0.98**
XGBoost	2047	0.96	17264	0.66	63	0.98	50	0.98	122	0.95	207	0.96	45	0.91	2087	0.88	1385	0.98	51	0.97
Decision Tree	2299	0.95	24703	0.50	85	0.97	86	0.94	168	0.92	128	0.99	43	0.92	1942	0.90	1631	0.97	92	0.90
DT + Bagging	1596	0.98	17499	0.75	73	0.98	71	0.96	173	0.91	285	0.93	43	0.92	1560	0.94	1705	0.97	48	0.97
Random Forest	1558	0.98	**16812**	**0.77**	**66**	**0.98**	59	0.97	160	0.92	**158**	**0.98**	**37**	**0.94**	**1426**	**0.95**	**1170**	**0.99**	44	0.98
SVM	9766	0.29	31136	0.20	273	0.67	158	0.81	343	0.69	1126	0.12	66	0.80	6519	0.08	7015	0.51	242	0.37

Wind Production

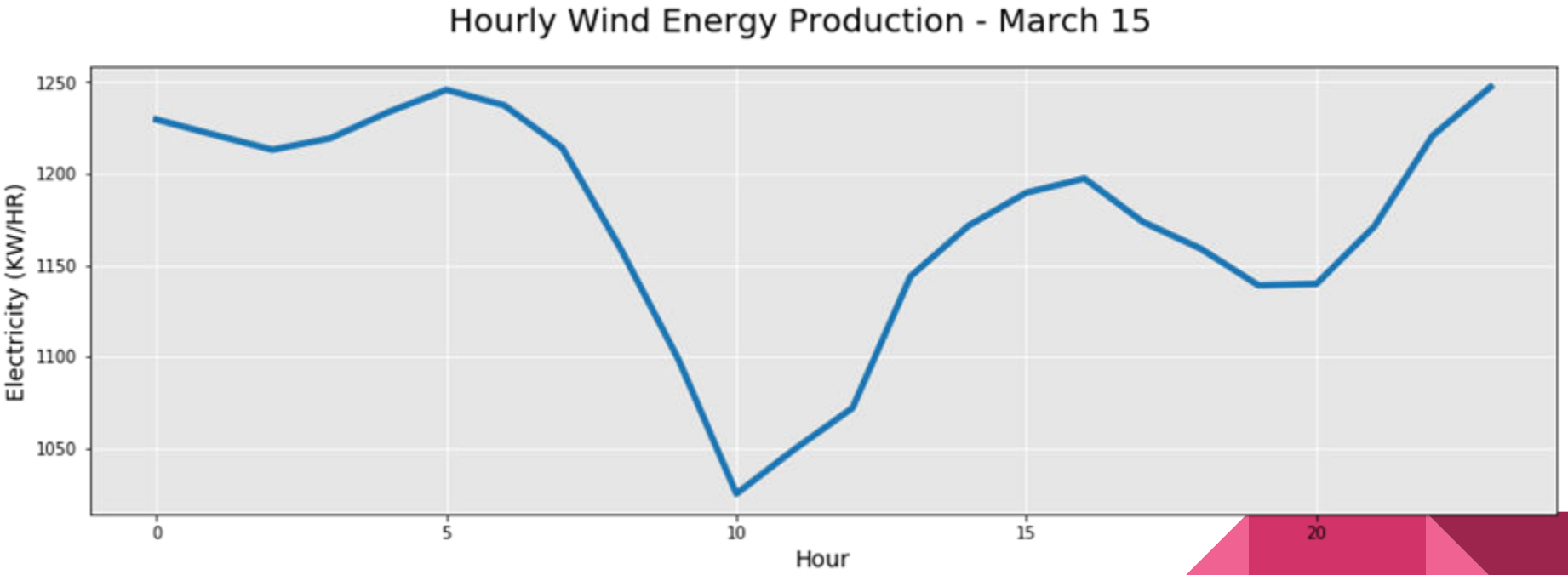
Wind Production | Best Model

- **Best Model**: Random Forest
  - min_samples_split = 7
  - n_estimators = 200
- **RMSE**: 1557.96
- **Top Important Feature (Daily Average)**:
  - Wind Speed 99%

Wind Production | Predictions

Date	Electricity Production (kW/hr)	Daily Electricity Production (kW)
March 15th	1,173	28,169
June 26th	1,176	28,224
July 3rd	469	11,261
October 13th	3,441	82,598
November 19th	8,777	210,666
December 25th	8,534	204,825

Wind Hourly Production March 15



Solar Production

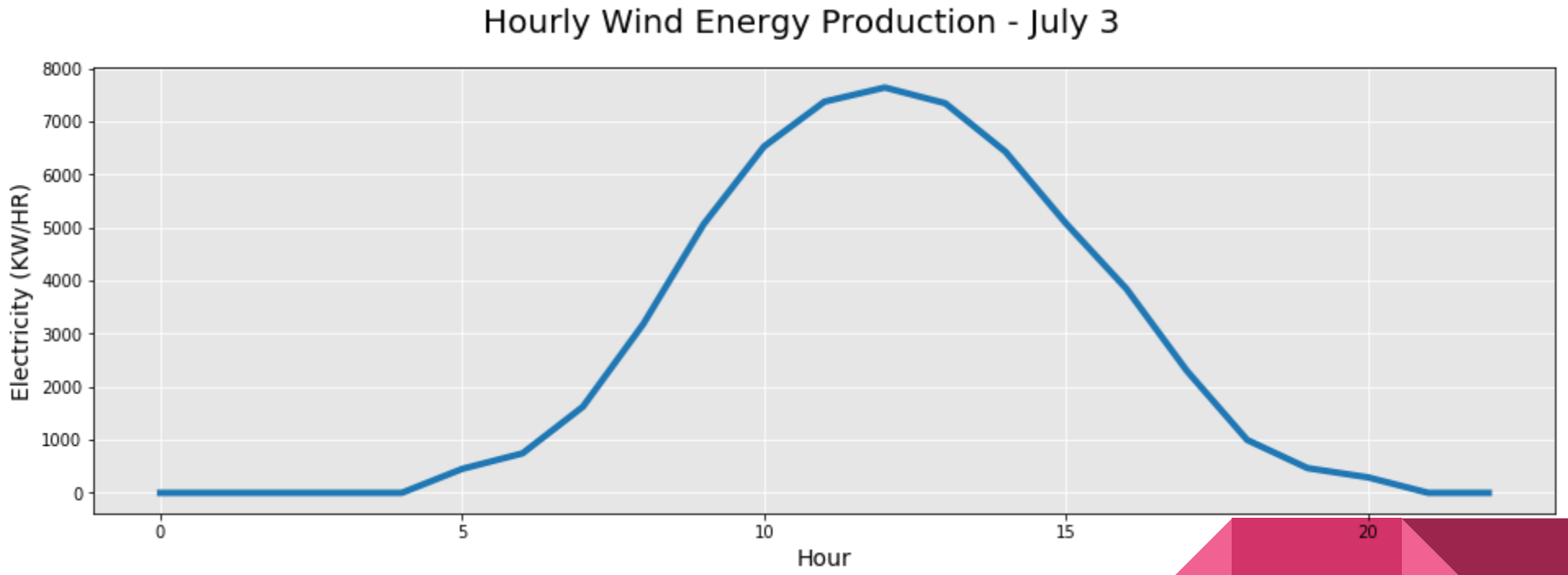
Solar Production | Best Model

- **Best Model**: Random Forest
  - min_samples_split = 6
  - n_estimators = 200
  - criterion split = Mean Absolute Error
- **RMSE**: 16811.71
- **Top 5 Important Features (Daily Average)**:
  - Cloud Coverage 55.2%
  - Humidity 9.4%
  - Solar Elevation 6.8%
  - Dew Point 4.7%
  - Temperature 4.4%

Solar Production | Predictions

Date	Electricity Production (kW/hr)	Daily Electricity Production (kW)
March 15th	59,364	1,424,748
June 26th	47,331	1,135,937
July 3rd	78,336	1,880,053
October 13th	41,291	990,976
November 19th	35,449	850,788
December 25th	15,487	371,694

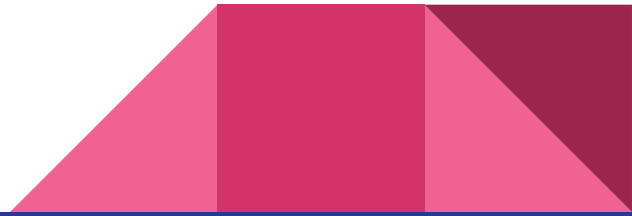
Solar Hourly Production July 3



Sector Consumption

Consumption | Challenge

- 8 different sectors consume electricity differently
- Work on individual sector’s energy consumption separately
- Find the best model for each sector
- Random forests is the winning algorithm across 8 sectors’ models
- Compare and fine-tune one Random forests model to fit all 8 sectors



Consumption | Evaluation Chart

	Consumption															
	Food Service		Grocery		Healthcare		K12		Lodging		Office		Residential		Retail	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Gradient Boosting	70.43	0.98	41.96	0.99	98.95	0.97	191.19	0.97	40.72	0.92	1868.20	0.91	1277.54	0.98	40.54	0.98
AdaBoosting	80.62	0.97	75.03	0.95	149.45	0.93	158.96	0.98	40.58	0.93	2468.26	0.82	1644.44	0.97	35.91	0.98
XGBoost	69.27	0.98	42.93	0.98	136.75	0.94	181.94	0.97	45.25	0.91	2180.78	0.87	1282.83	0.98	47.56	0.97
Random Forest	65.99	0.98	59.36	0.97	161.17	0.92	158.40	0.98	37.39	0.94	1425.54	0.95	1169.76	0.99	52.63	0.97

Consumption | Best Model

- Model Parameters: **RandomForestRegressor**(max_features=0.7, min_samples_leaf=1, min_samples_split=3, n_estimators=410, random_state=3)
- Average RMSE: 17,921.49

Consumption | Feature Importance

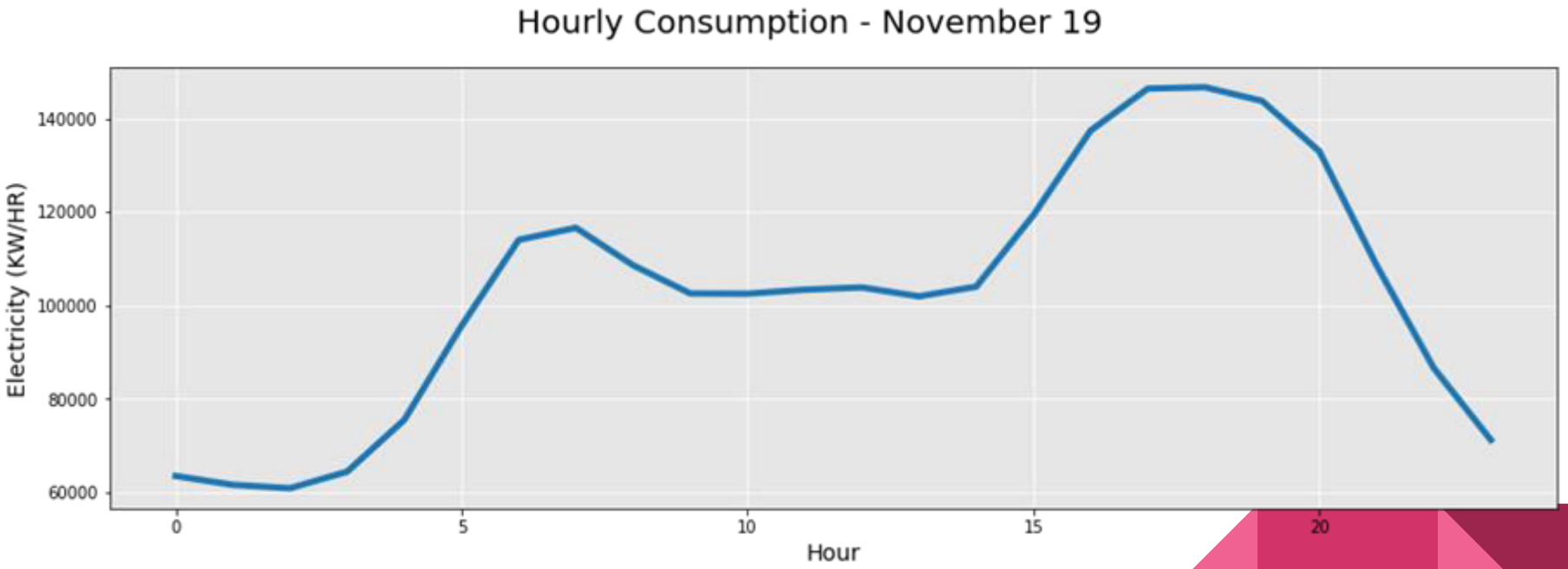
Food		Grocery		Healthcare		K-12	
Temperature	74.5%	Dew Point	51.7%	Temperature	51.6%	Weekend	61.8%
Dew Point	19.3%	Temperature	29.7%	Weekend	16.9%	Saturday	6.9%
Solar Elevation	2.6%	Sunday	8.5%	Dew Point	15.5%	Temperature	5.6%
Summer	0.9%	Weekend	3.4%	School Day	3.4%	Solar Elevation	5.3%
Humidity	0.7%	Solar Elevation	1.9%	Solar Elevation	3.3%	School Day	4.9%

Lodging		Office		Residential		Retail	
Temperature	69.1%	Weekend	45.6%	Temperature	71.9%	Sunday	48.5%
Dew Point	23.6%	Temperature	18.3%	Dew Point	20.9%	Temperature	26.1%
Solar Elevation	2.4%	Sunday	12.6%	Solar Elevation	2.4%	Dew Point	7.3%
Summer	2.0%	Dew Point	8.5%	Summer	1.6%	Weekend	7.2%
Cloud Cover	0.7%	Holiday	4.1%	Humidity	0.7%	Holiday	5.1%

Consumption | Predictions

Date	Electricity Consumption (kW/hr)	Electricity Consumption (kW)
March 15th	83,802	2,011,266
June 26th	129,677	3,112,260
July 3rd	97,558	2,341,405
October 13th	102,659	2,463,831
November 19th	103,032	2,472,786
December 25th	103,164	2,475,945

Hourly Consumption Nov 19



Electric Vehicles

Consumption w/ Electric Vehicles | Predictions

Date	Electricity Consumption (kW)	Consumption w/EVs (kW)
March 15th	2,011,266	2,169,449
June 26th	3,112,260	3,270,444
July 3rd	2,341,405	2,499,589
October 13th	2,463,831	2,622,014
November 19th	2,472,786	2,630,969
December 25th	2,475,945	2,634,129

Increase in 158,183 kW/daily per year

~7% consumption increase

Hourly Consumption w/ Electric Cars Nov 19

Das Bild kann zurzeit nicht angezeigt werden.

Conclusions & Recommendations

Predictions | Summary

Date	Wind Production (kW)	Solar Production (kW)	Consumption (kW)	Consumption w/EVs (kW)
March 15th	28,169	1,424,748	2,011,266	2,169,449
June 26th	28,224	1,135,937	3,112,260	3,270,444
July 3rd	11,261	1,880,053	2,341,405	2,499,589
October 13th	82,598	990,976	2,463,831	2,622,014
November 19th	210,666	850,788	2,472,786	2,630,969
December 25th	204,825	371,694	2,475,945	2,634,129

Concerns

- Without increased investment in renewable energy, the electric vehicles will increase your reliance on coal and non-renewables
- Growth in electric vehicles have a consistent increased demand on the electric grid
- Risk of *overproduction* of energy in the summer, wasting potential if Power City does not have adequate energy storage
- Renewable sources are not consistent supply to the electric grid
    - Subject to shocks with unexpected weather

Recommendations

- Three recommended investments:
    - Scaling up your wind farm annually by 30%
    - Scaling up your solar park annually by 20%
    - Increasing energy storage technology

- At *minimum*, increase wind and solar energy by 14% to compensate for EV demand alone (assuming normal weather and energy storage)

- Invest in more efficient energy technology (e.g. LEDs)

- Invest in reducing your citizens consumption

Questions?

Thank You!