

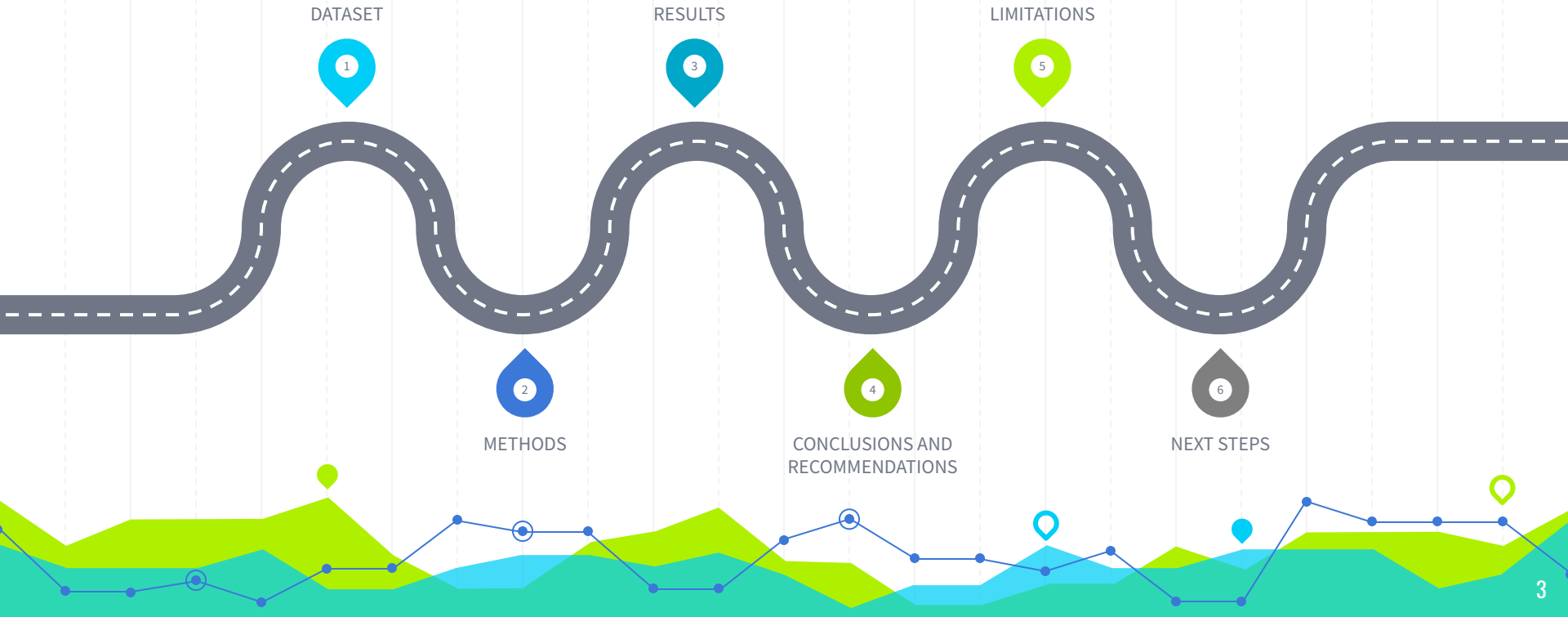
# TIME SERIES - ENERGY PRODUCTION IN THE US

# Business Problem

- Electricity - more in demand for tech sector growth
- Data Centers, Hospitals, Distribution Centers
- Need for available and reliable sources of power
- Investing in electricity - identify growing sources



# ROADMAP





# Dataset:

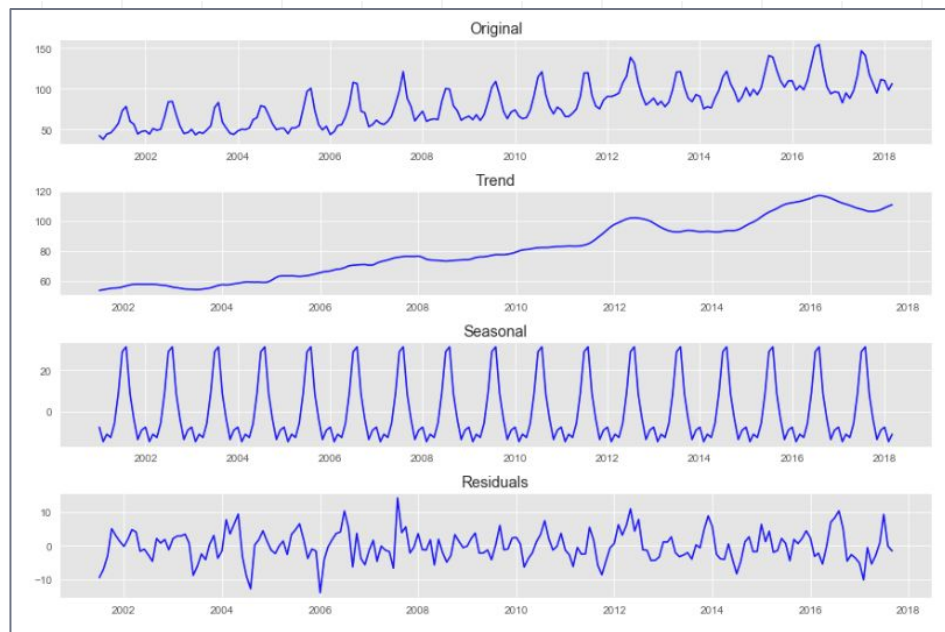
- Dataset from Kaggle, adapted from the US EIA website
- Production of electricity via different sources in all 50 states
- Data spans from January 2001 to May 2022, one record per month

# Methods

- Data Preparation
- Naive Baseline Model
  - ARMA models
  - SARIMAX models
- Predicting on the Test Set
- Forecasting next 3 years
  - Study of Seasonality
- Study of production by State

# Models

## Series Decomposition:



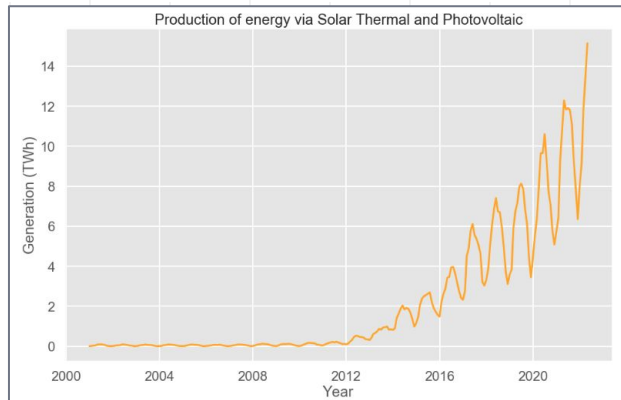
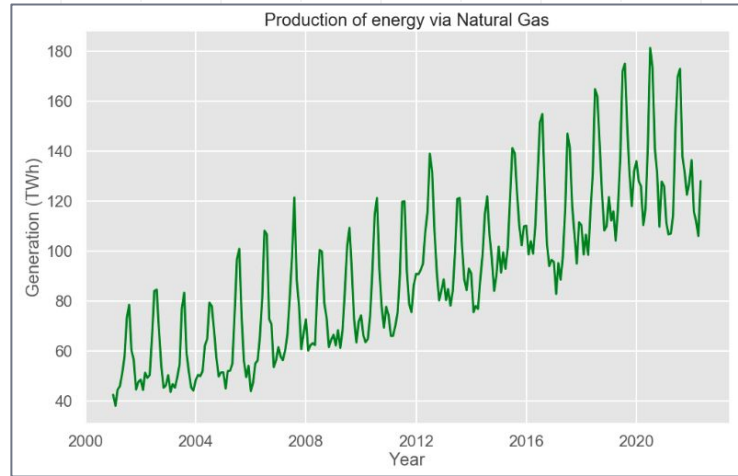
## ARMA and SARIMA models:

$$y_t = c + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$



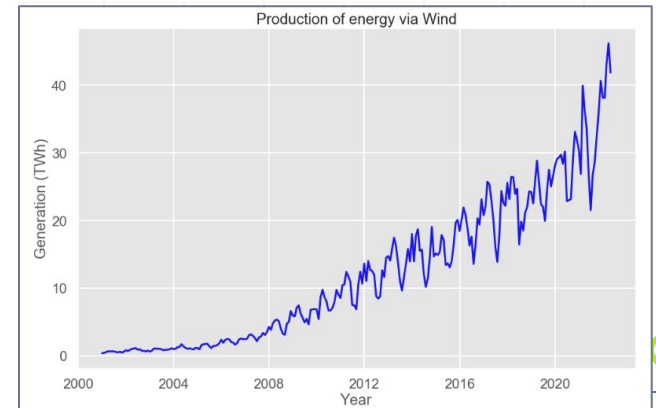
# Results

# TOP 3 SOURCES



Natural gas, solar and wind all showed upward trends

We decided to focus on natural gas for its greater reliability





# PREDICTIONS ON TEST

Best model:

SARIMAX model parameters

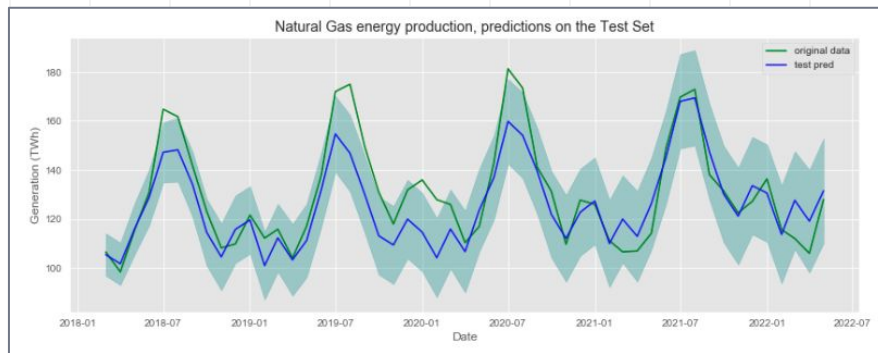
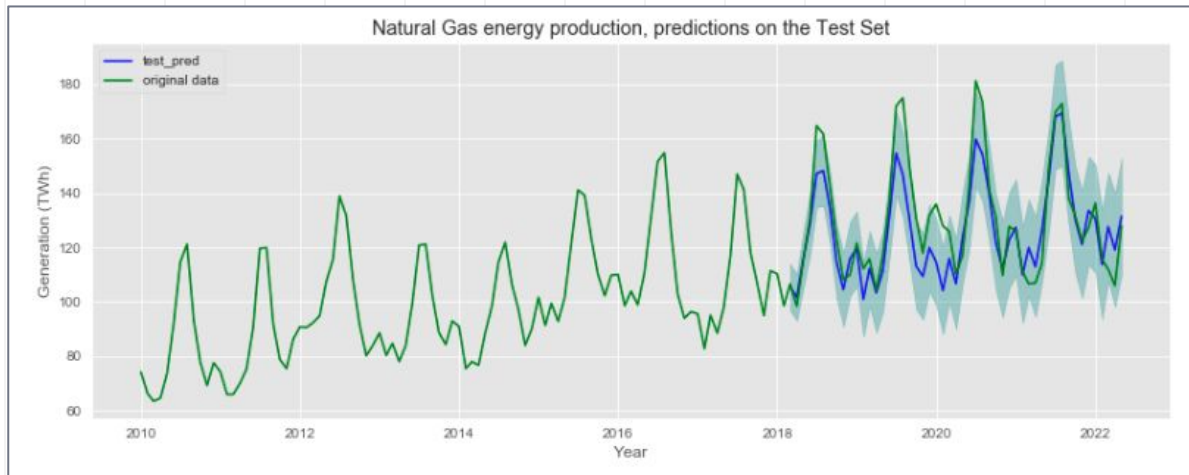
(8, 1, 2)(12, 2, 1, 12)

AIC of 173.5

With this model:

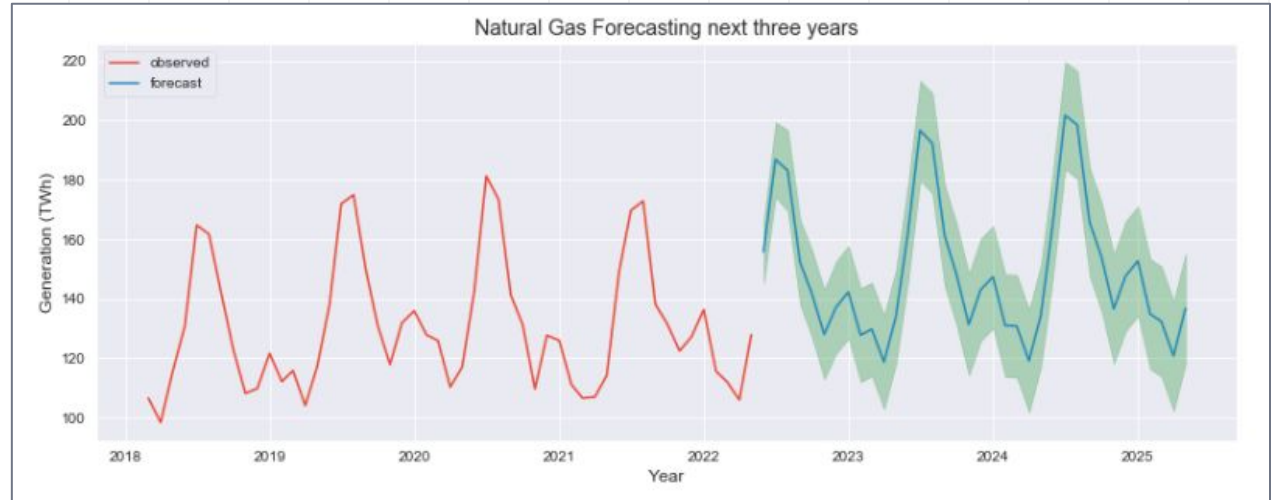
predictions on the test set  
for 51 months with an error

RMSE = 10.9 TWh

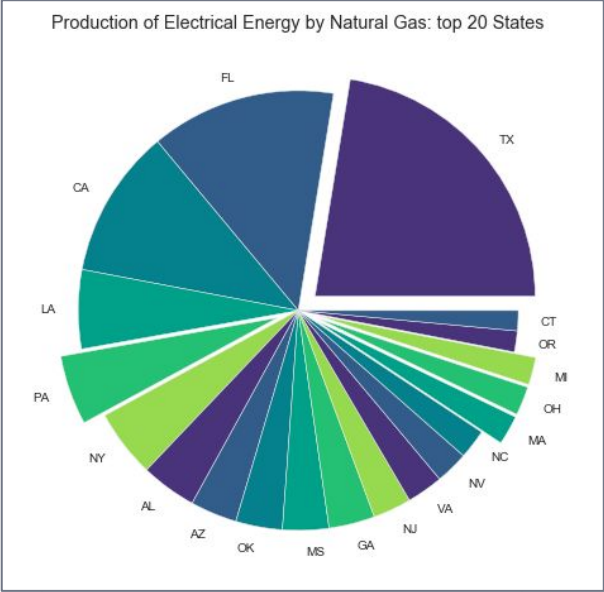


# FORECASTING

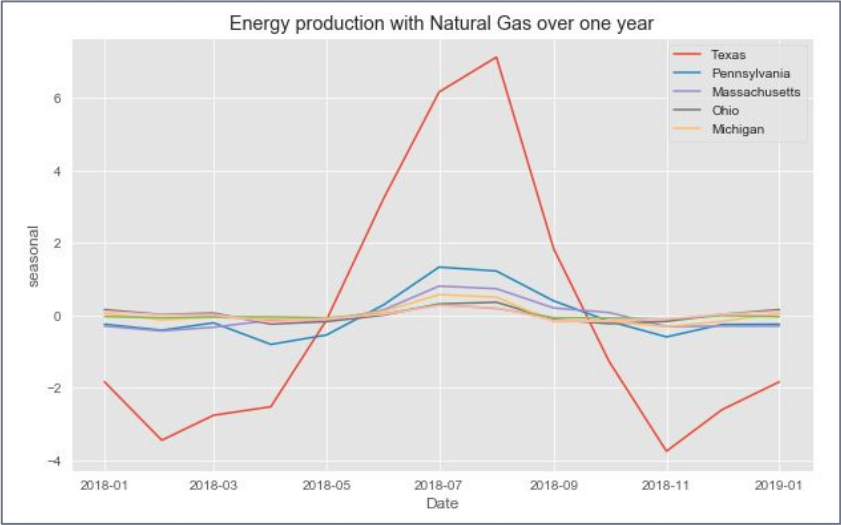
We forecasted our data  
for 3 years finding a growth  
of up to 16.7%  
between 2021 and 2024  
with an error of 4%



# STUDY BY STATE AND SEASON



Study of pattern over a year for Texas, Pennsylvania, Massachusetts, Ohio and Michigan



## CONCLUSIONS and RECOMMENDATIONS:

1

**Nat Gas: growing and reliable**

17% growth between 2021 and 2024

2

**12 Months Seasonality**

Repeating patterns probably due to weather fluctuations

3

**Best states: Texas and PA**

TX: state with highest production  
PA: lower spike over summer season

# LIMITATIONS

## Data Preparation:

Different dataset could require formatting and filling missing data

## Parameters and Models:

Different parameters and models might not perform as well

## Running Time:

Searches for best parameters take a long time to run



## Next Steps and Future Improvements:

- Studying solar and wind generated energy trends
- More data: farther in the past or records by day
- Deeper study trend and seasonality and influential factors
- More powerful tools: Prophet, AWS SageMaker's DEEPAR

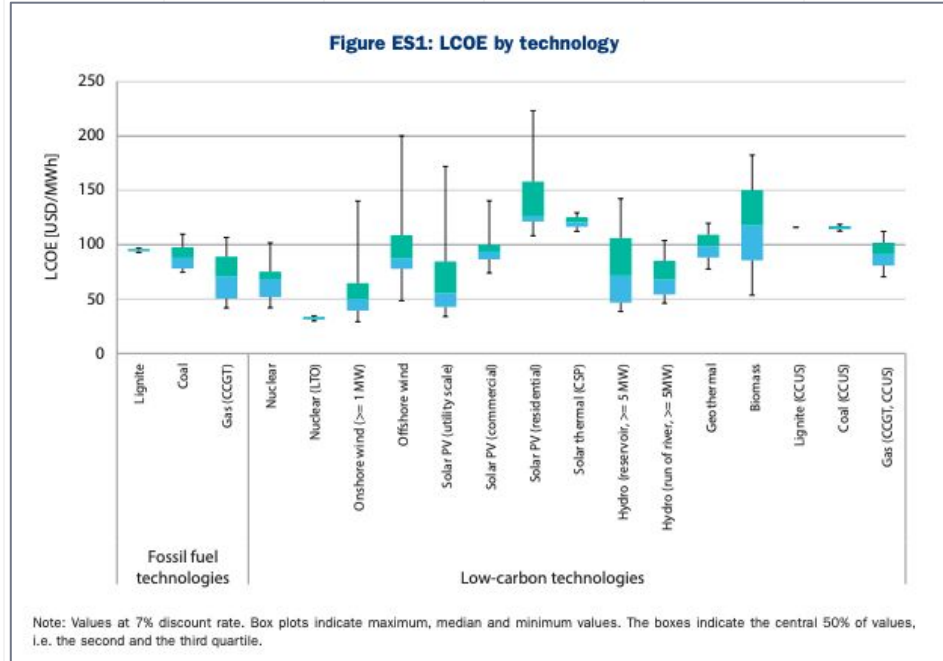


# THANKS!

**Any questions?**

You can find me at  
[marianlkuzmin@gmail.com](mailto:marianlkuzmin@gmail.com)

# Production Cost by source:



Source: from the full report found on <https://www.iea.org/reports/projected-costs-of-generating-electricity-2020>