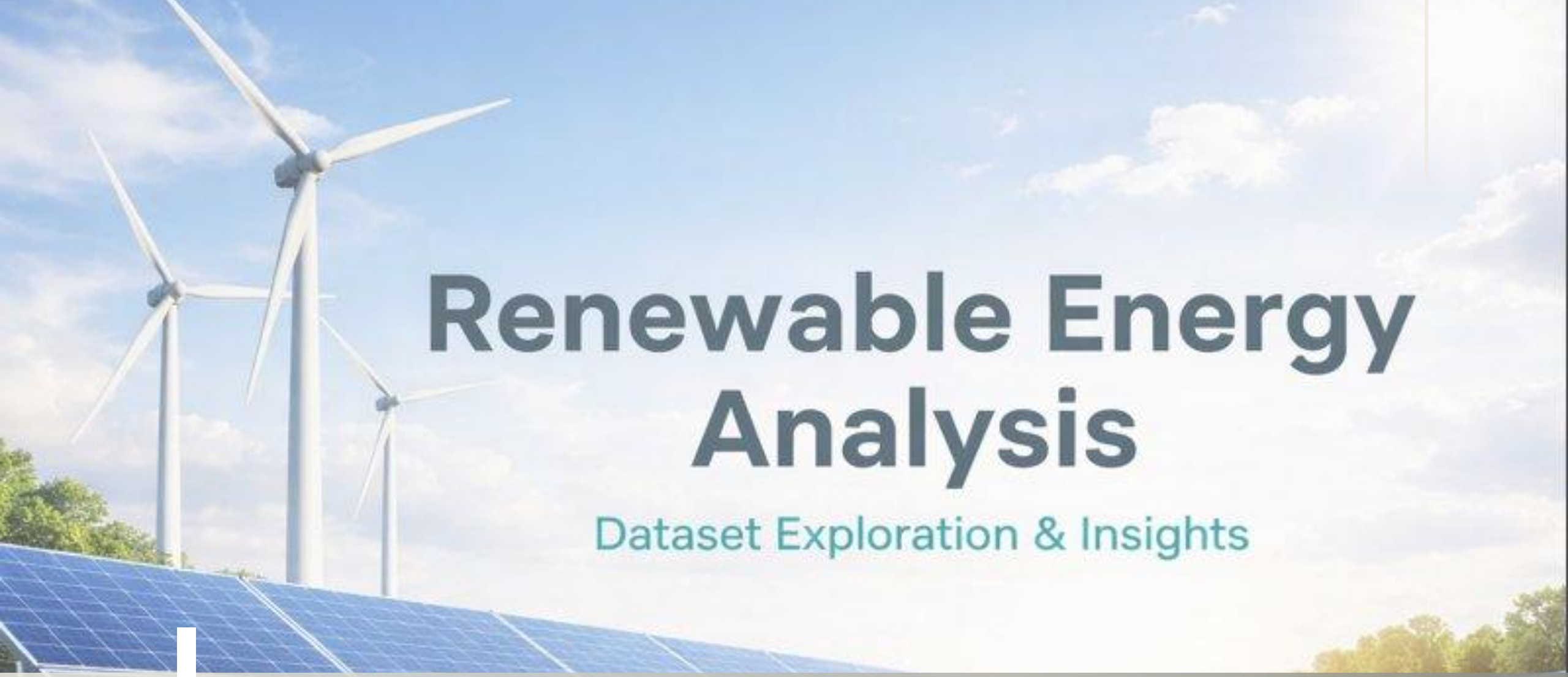
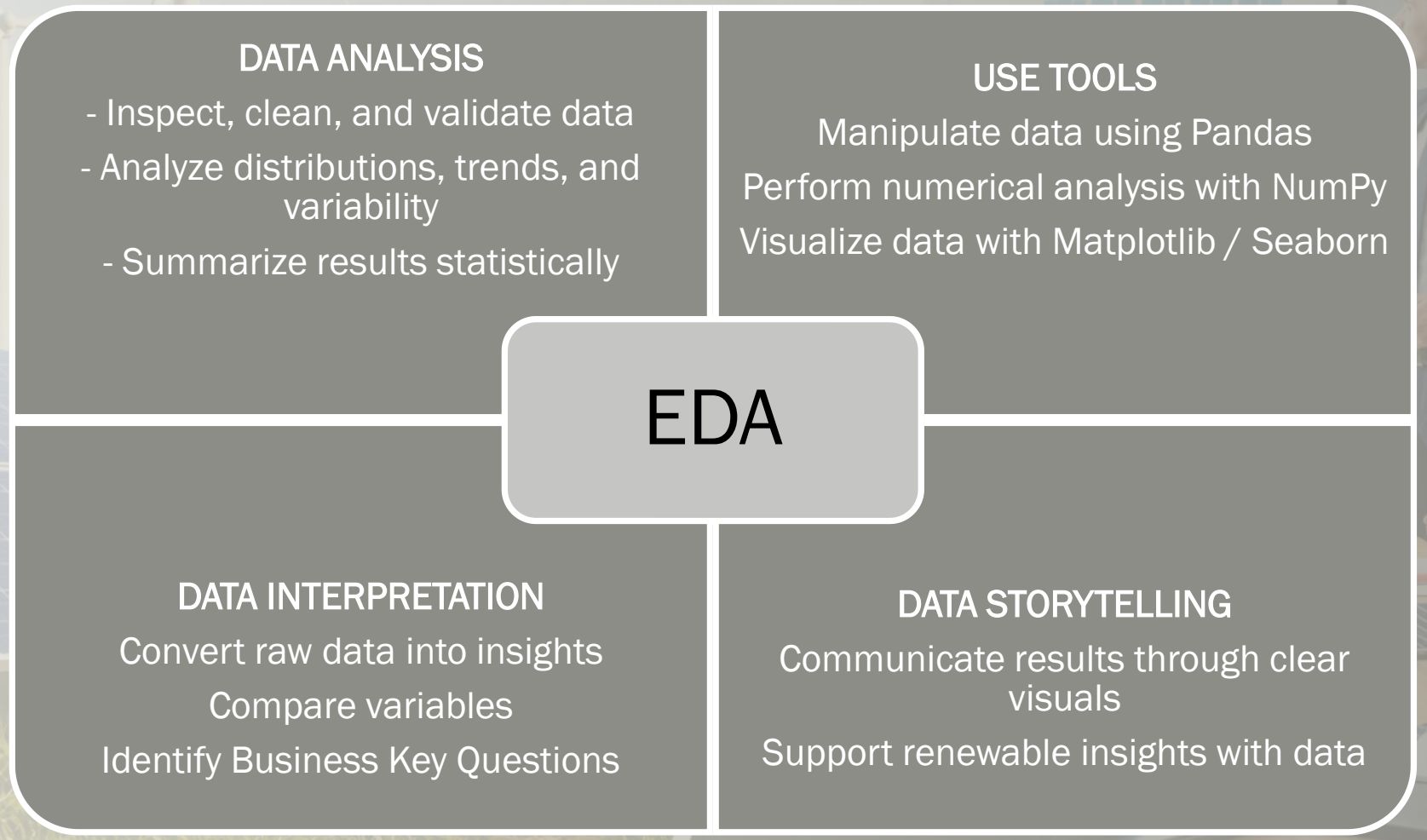



# Data Analysis & Visualization: Exploratory Data Analysis

Anne Valvezan and Suzana Souza

# Learning Objectives





# WIND AND SOLAR PRODUCTION

Dataset contains:

- Hourly wind and solar generation
- Data from France January 2020 to November 2025
- 51,864 complete records with 9 key columns.


**Production**
- Season
- Source
- Trend Source

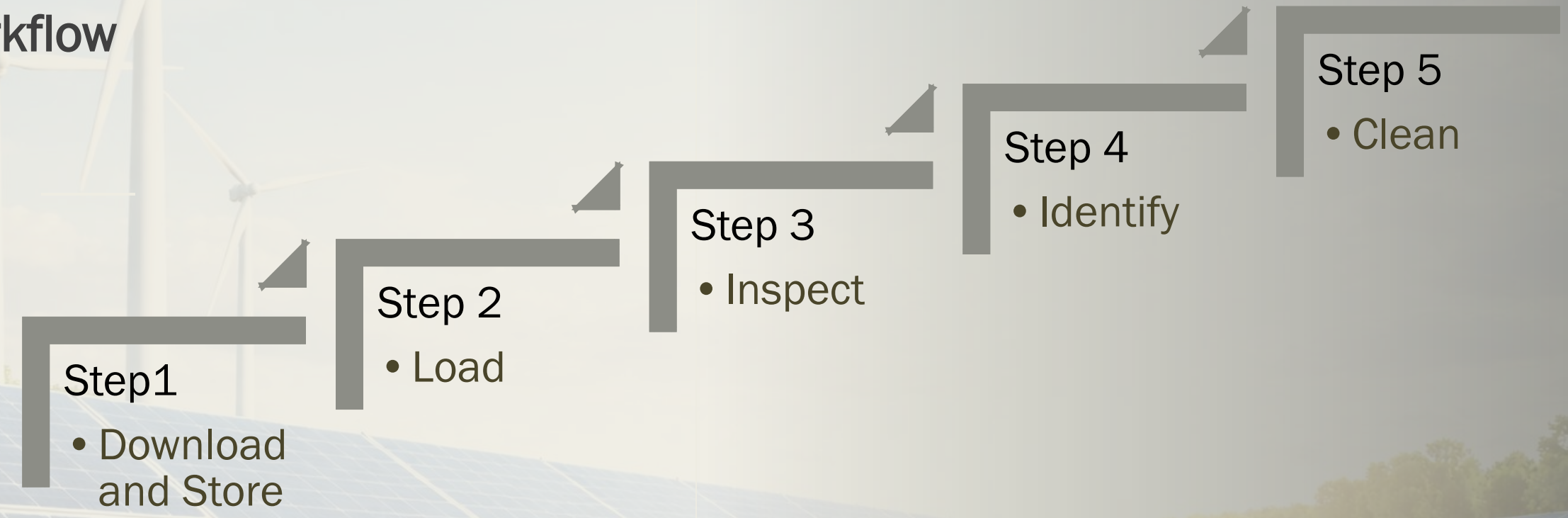
**Year Trends**
- Increasing
- Decreasing

**Business Insights**
- Recommendations
- Strategies

# WIND AND SOLAR PRODUCTION



## Workflow



# DATA EXPLORATION, ORGANIZATION AND CLEANING



```
df = data.copy()
df.columns
df.head()
df.describe()
df.info()
df.Date
```

## Verifying hours of the day

No missing values were found considering the days.

```
#Verifying if all days have 24 hours of measurements
measurements= df.groupby("Date")["Start_Hour"].unique()
```

```
df.duplicated().any()
```

```
np.False_
```

```
# Hours without 24 hours of measurements
measurements = measurements.reset_index()
non_standard_measurements = measurements[measurements["S
non_standard_measurements
```

```
if (df.duplicated()).any():
    print("Found duplicated values!")
else:
    print("Did not find any duplicated values!")
```

```
Did not find any duplicated values!
```

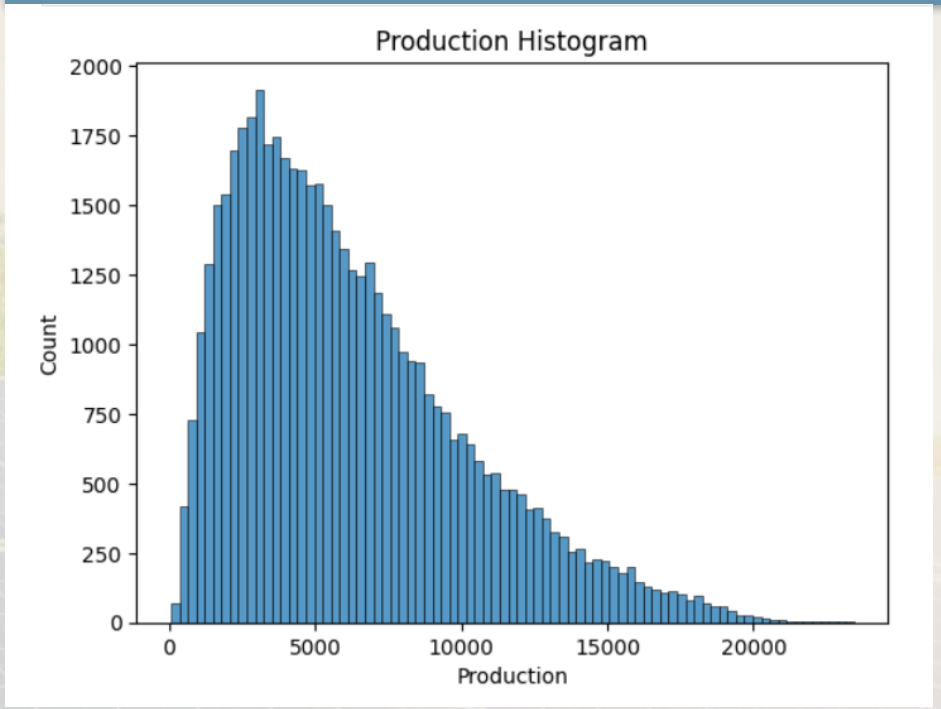
```
df.info()
```

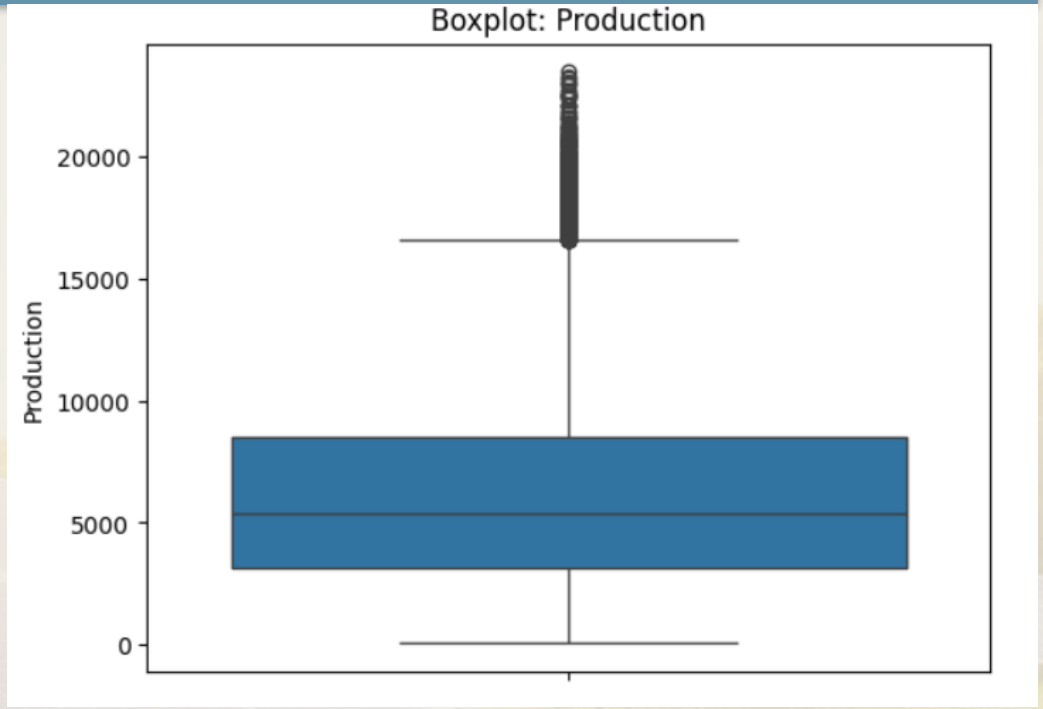
## CONCLUSIONS BEFORE UNIVARIATE AND BIVARIATE ANALYSIS:

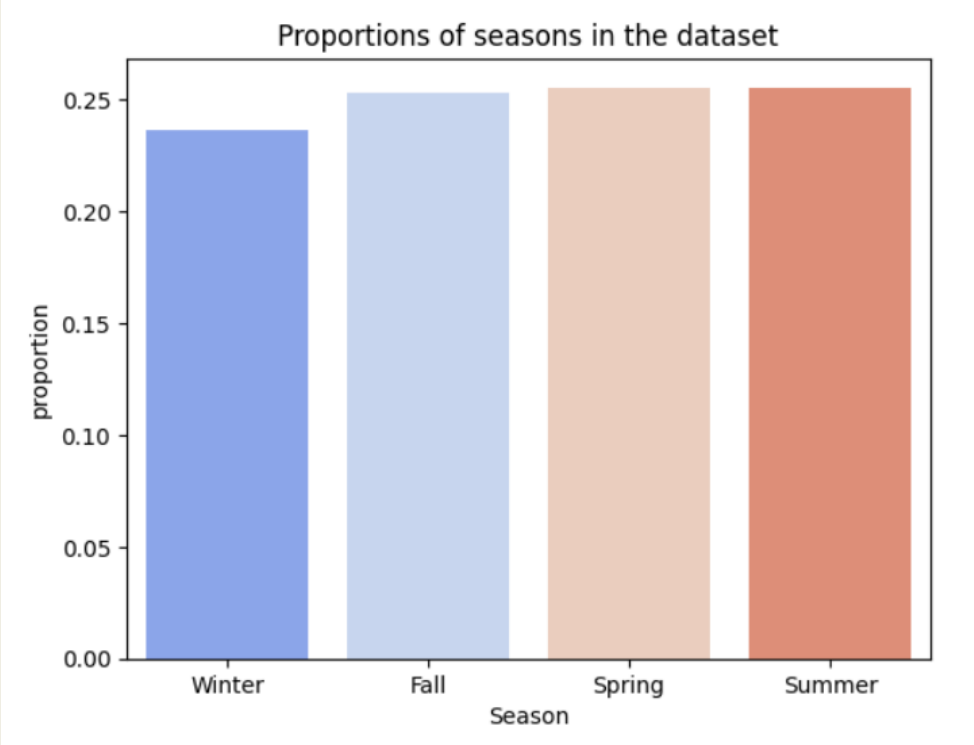
- No duplicated data
- No Null Values
- Truncated values
- Hours Anomalies
- Impact on Analyses

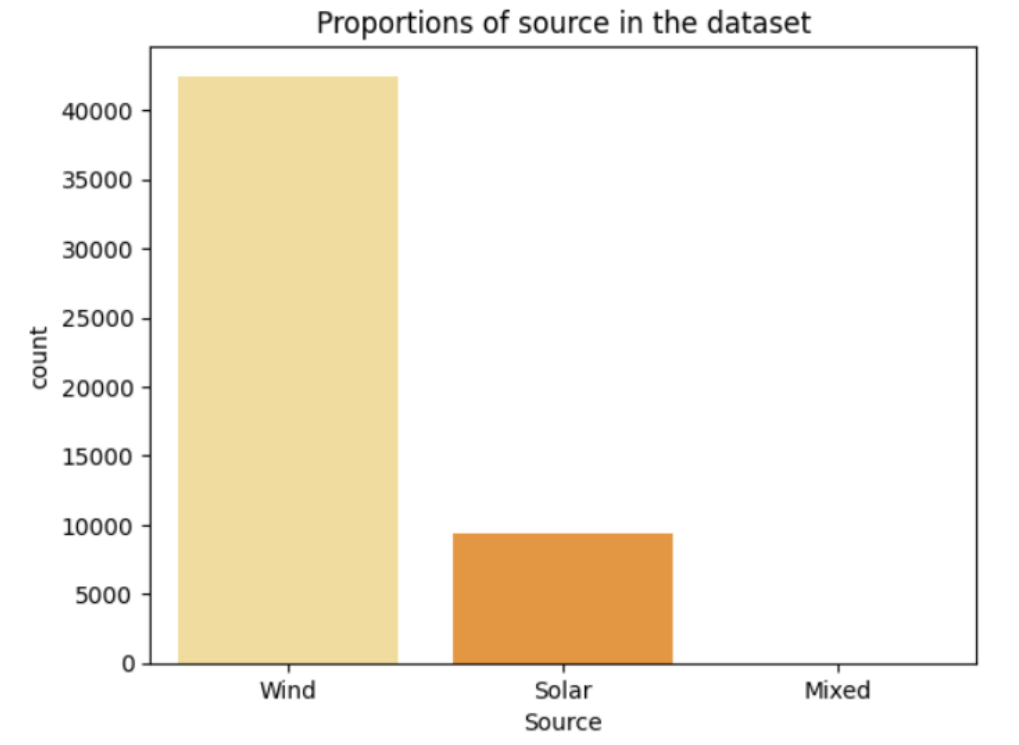
# EDA - UNIVARIATED VARIABLES

## Production





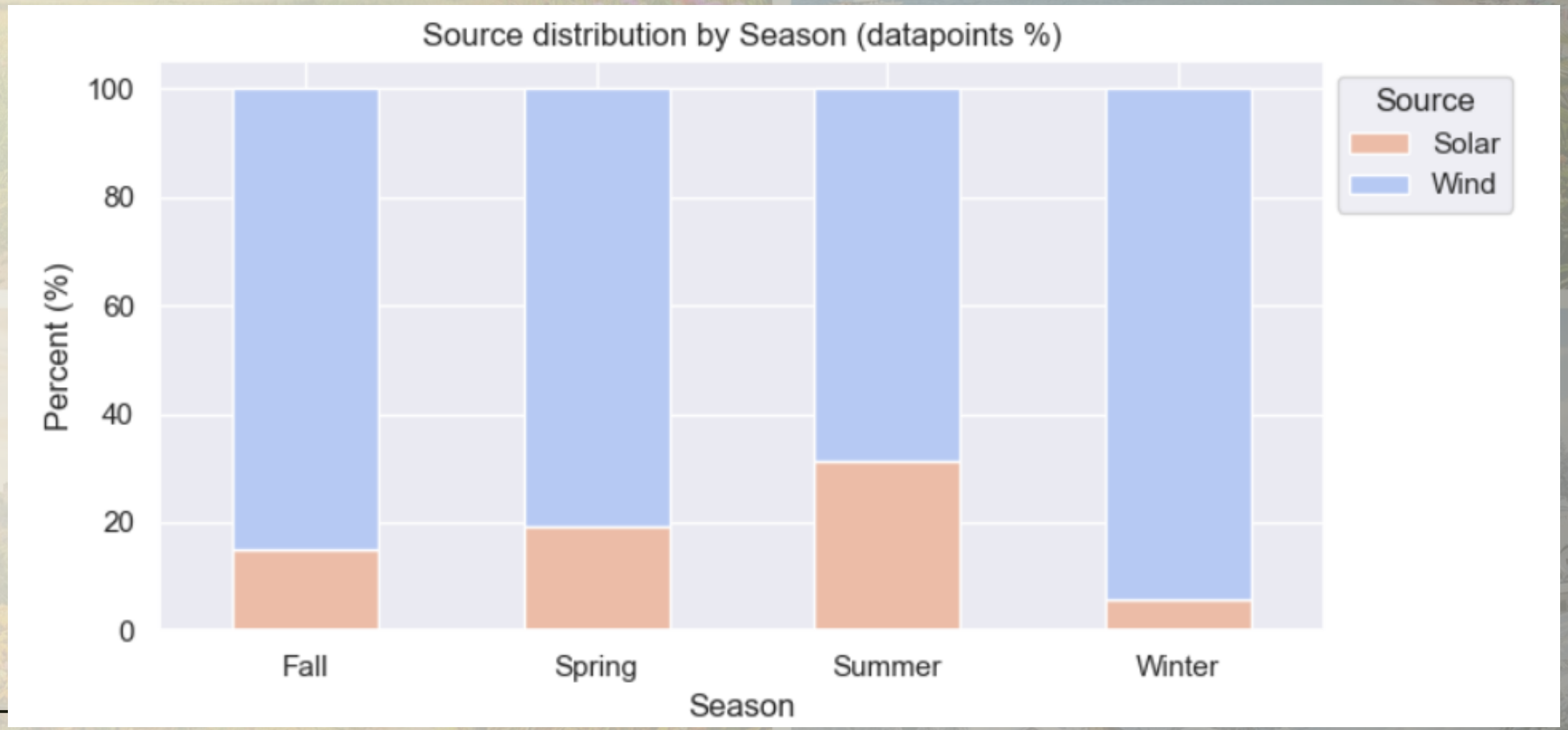



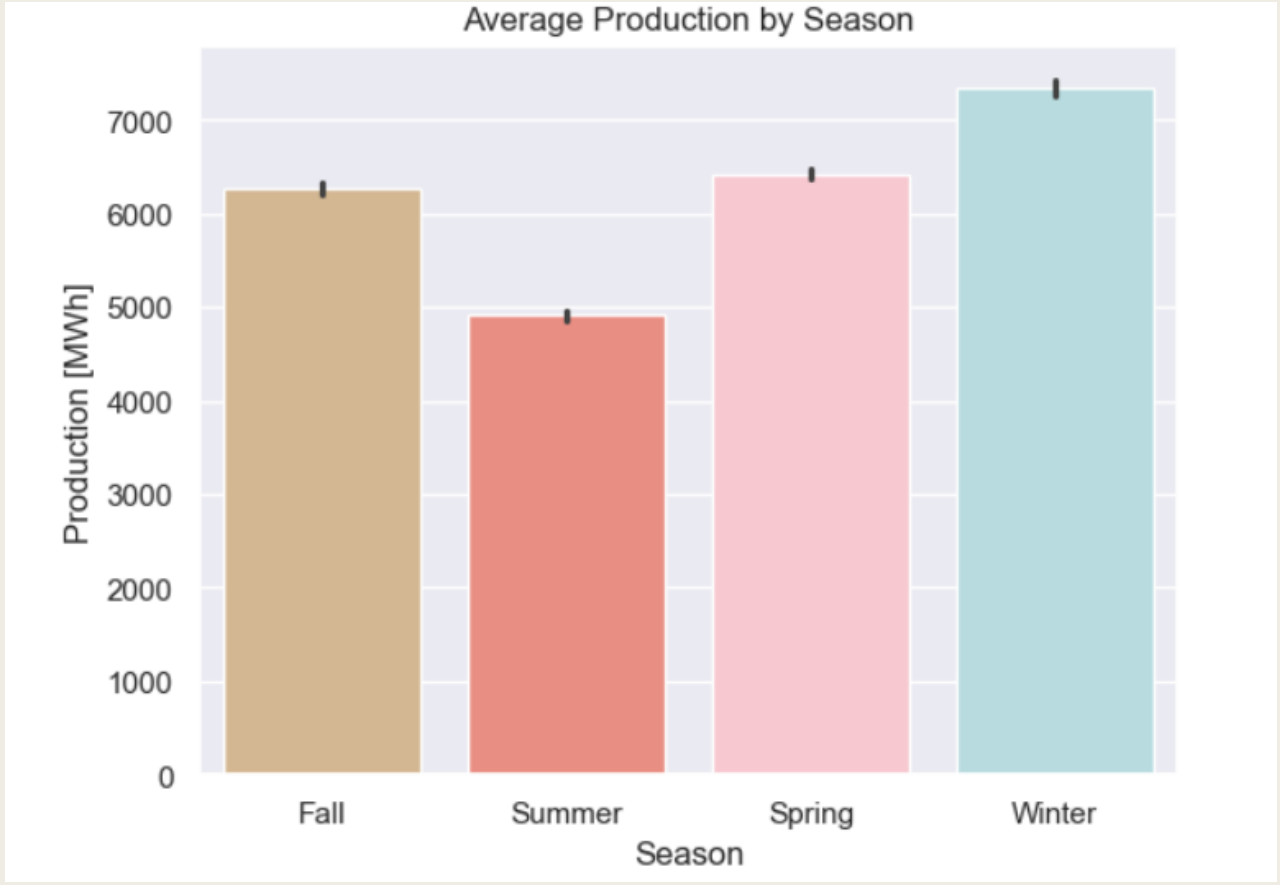

# SEASON AND SOURCE

*2026*

# EDA – BIVARIATED VARIABLES







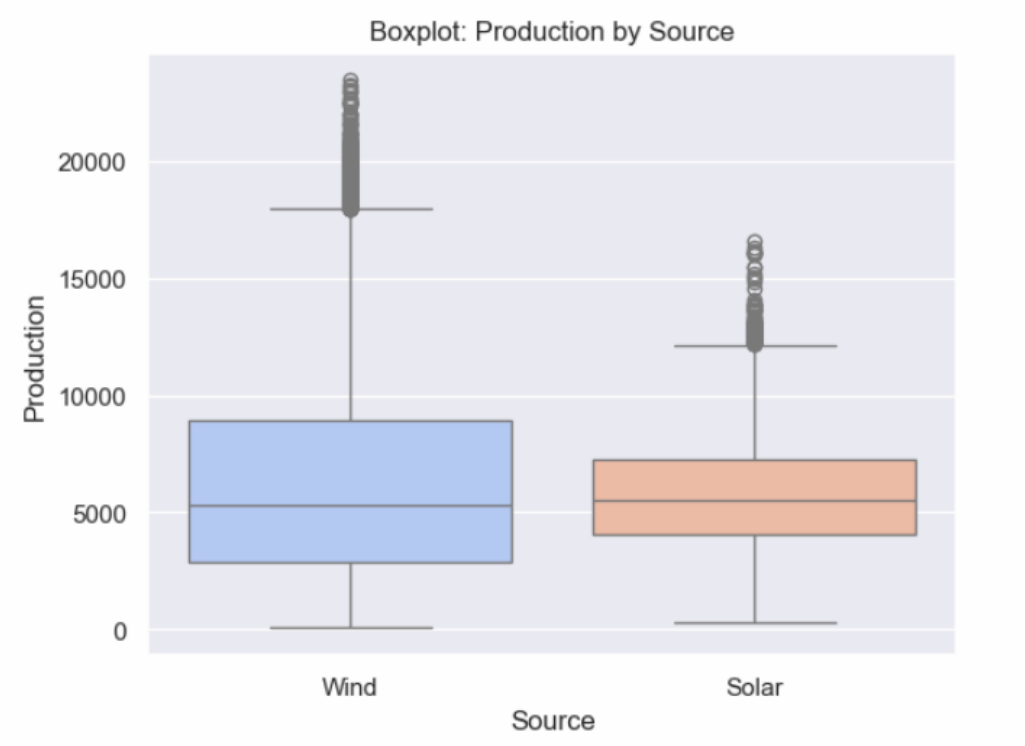

# EDA – BIVARIATED VARIABLES

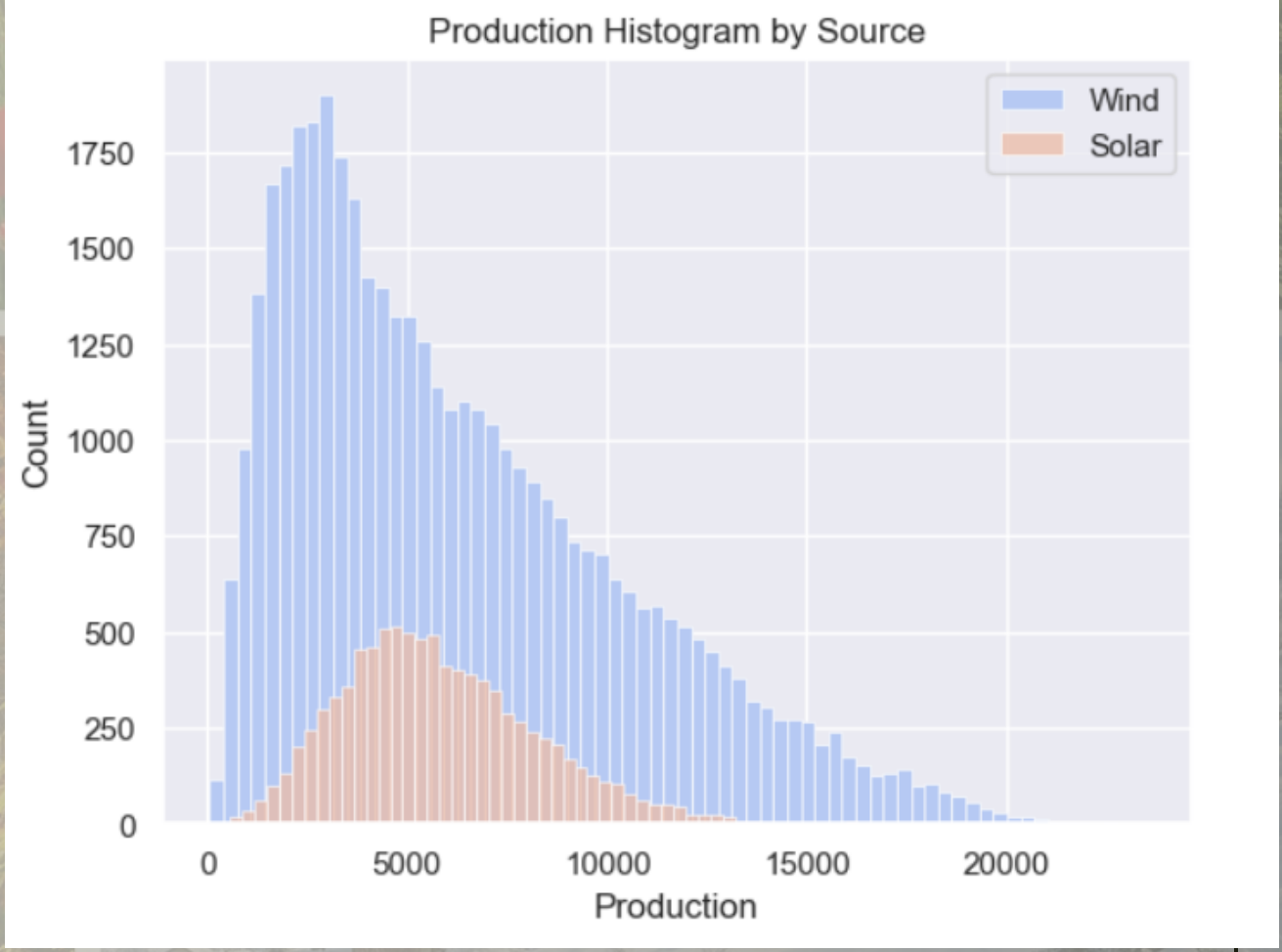
*2026*

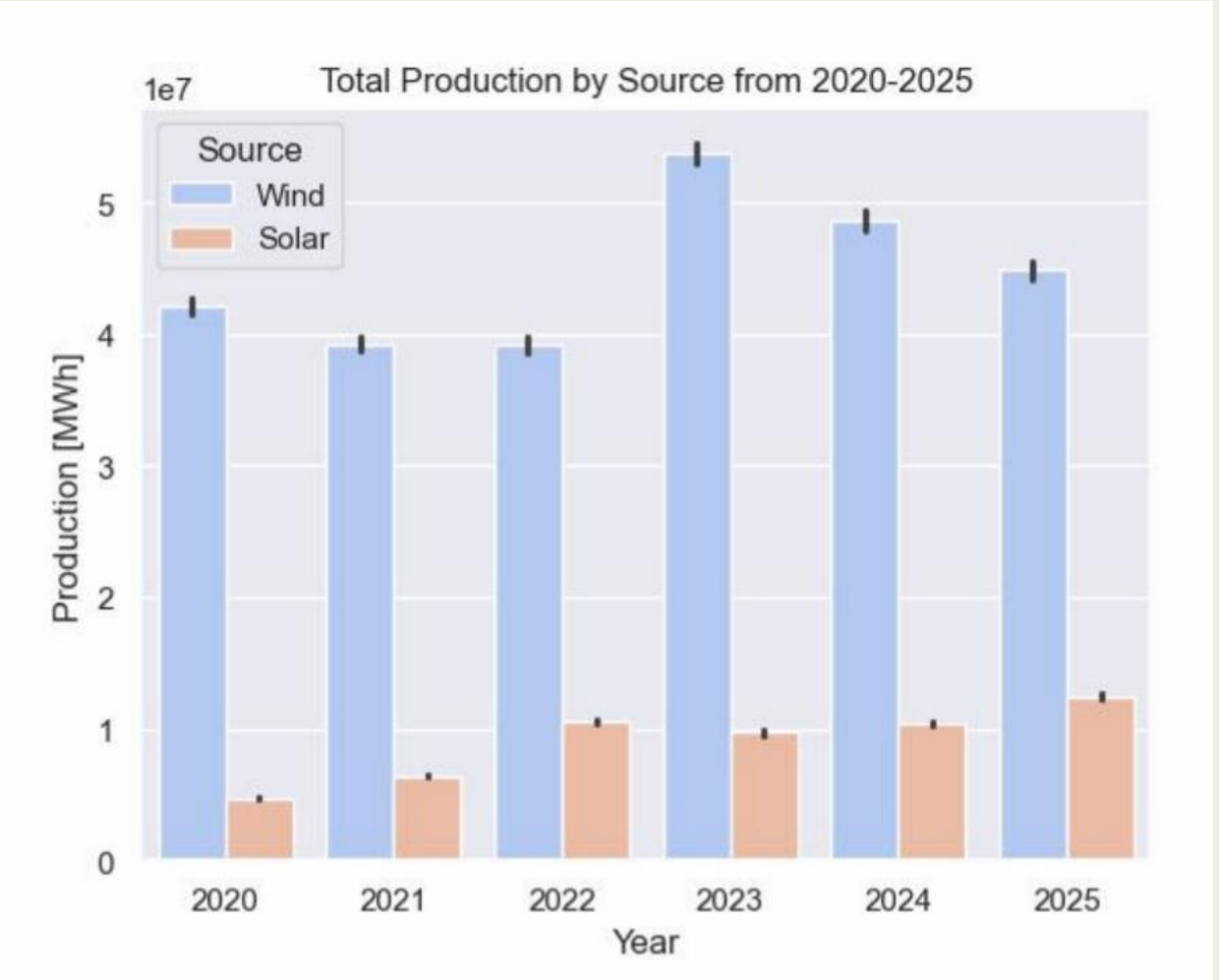
# EDA – BIVARIATED VARIABLES



- Distribution and Skewness
- Amplitude



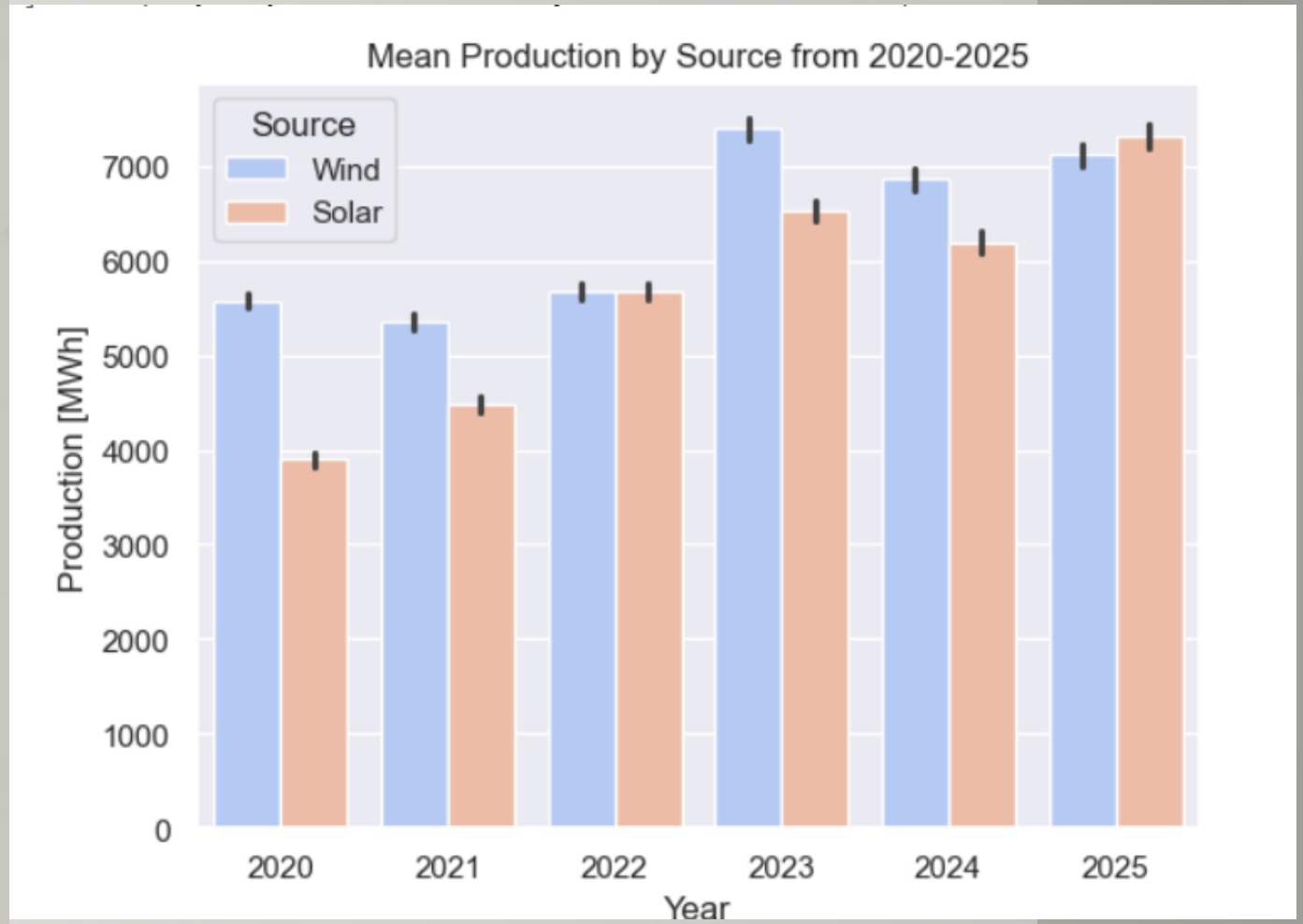






# EDA – BIVARIATED VARIABLES

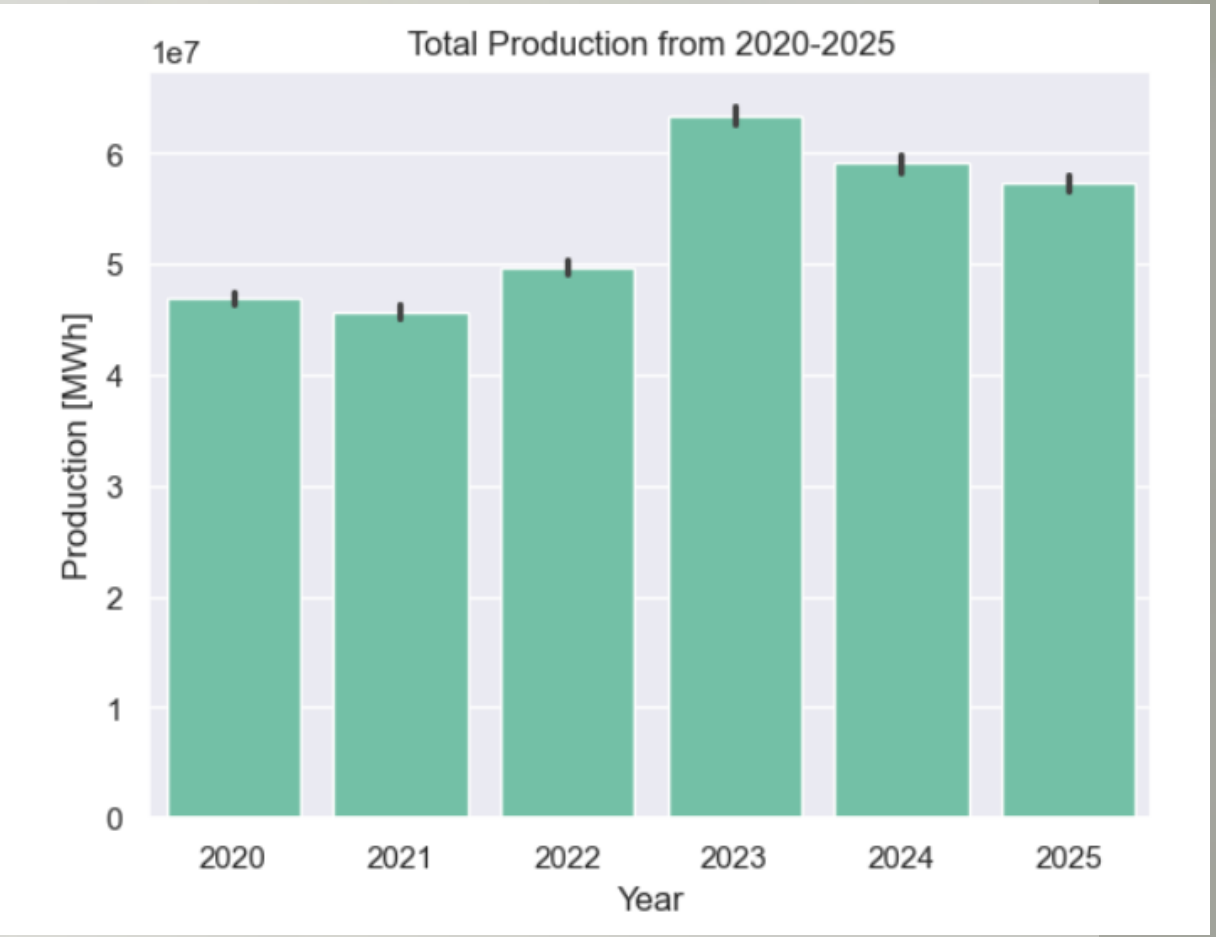
*2026*

- Solar trend increase
- Wind volatility oscillation
- Production comparable in magnitude



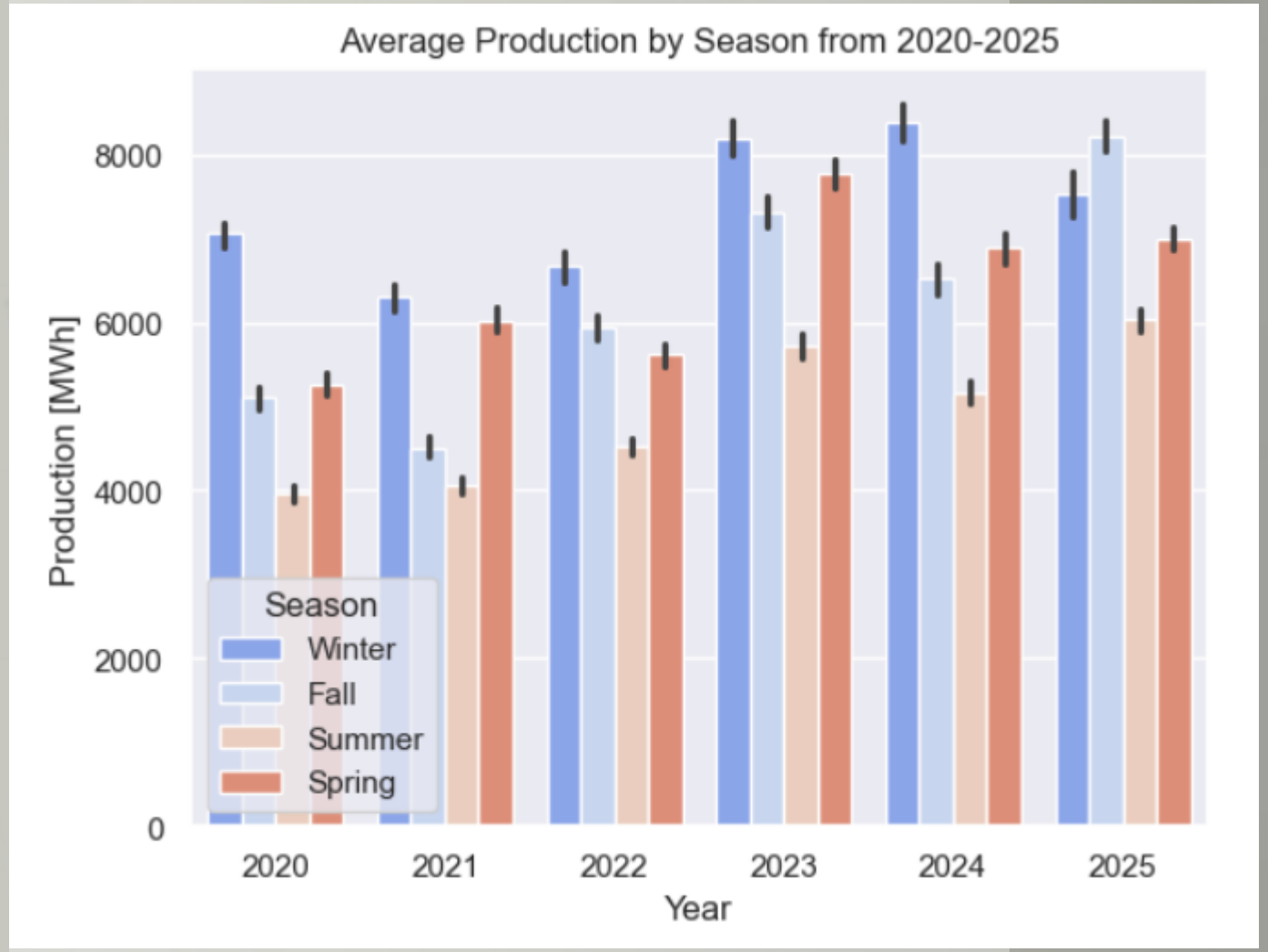
The production is consistent with wind distribution

Peak production



# Energy Average Production

- Winter fall spring oscillation
- Summer trend up
- Stability Wind vs Solar

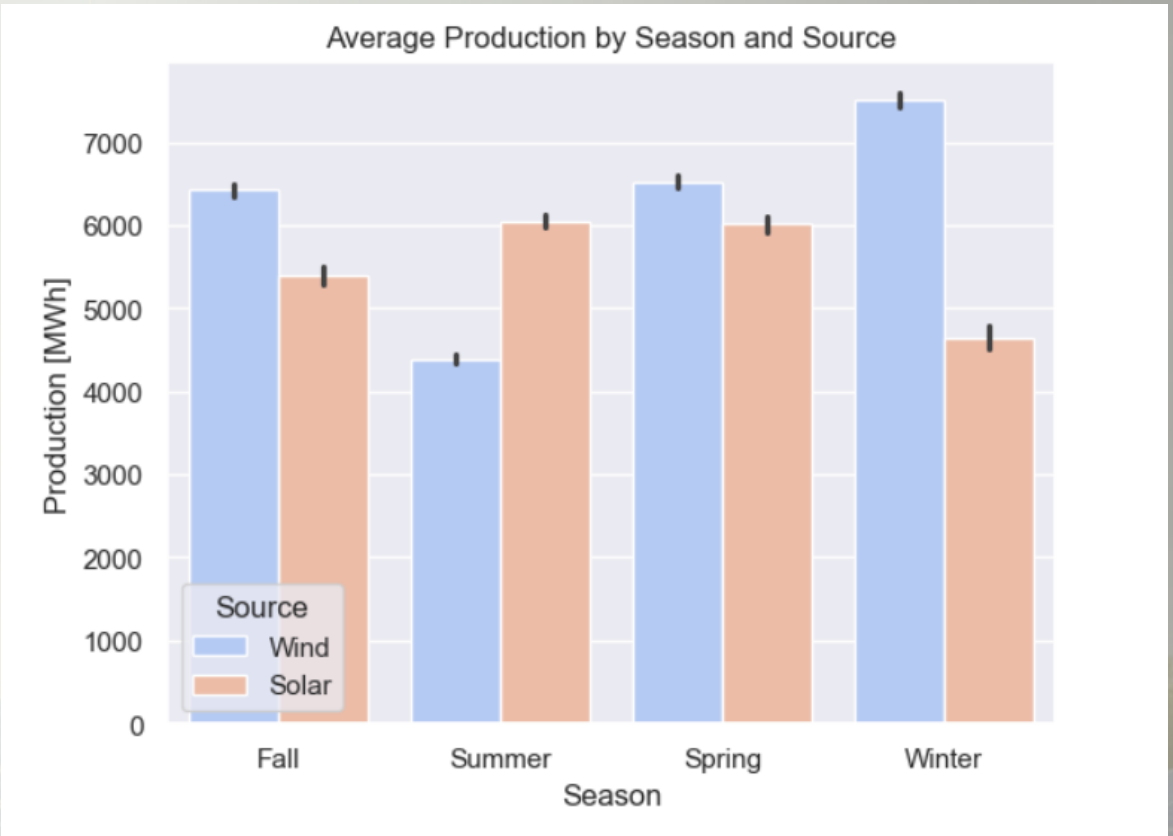


# MULTIVARIATED VARIABLES



## EXPLANATION

CONFIRMS THAT SOLAR PRODUCTION DOMINATE IN SUMMER AND WIND PRODUCTION DOMINATE IN WINTER



# GENERAL CONCLUSION

## Wind Energy:

Exhibits strong seasonal variation with peak output during winter months and significant decline in summer

High volatility characterized by frequent outliers and right-skewed distribution

Inherent unpredictability poses challenges for load forecasting and dispatch planning

## Solar Energy:

Demonstrates near-Gaussian distribution with more predictable behavior

Winter production decreases are moderate compared to wind's seasonal swings

Lower outlier frequency suggests higher reliability for baseline capacity planning

# RECOMENDATIONS



- **Short-term: Integrate dispatchable sources (battery storage, gas/coal sources) to address intermittency gaps and meet demand surges**

- **Medium-term: Expand solar capacity to improve generation predictability and reduce portfolio variance or explore alternative dispatchable long-term sources, e.g. nuclear or hydroelectric.**

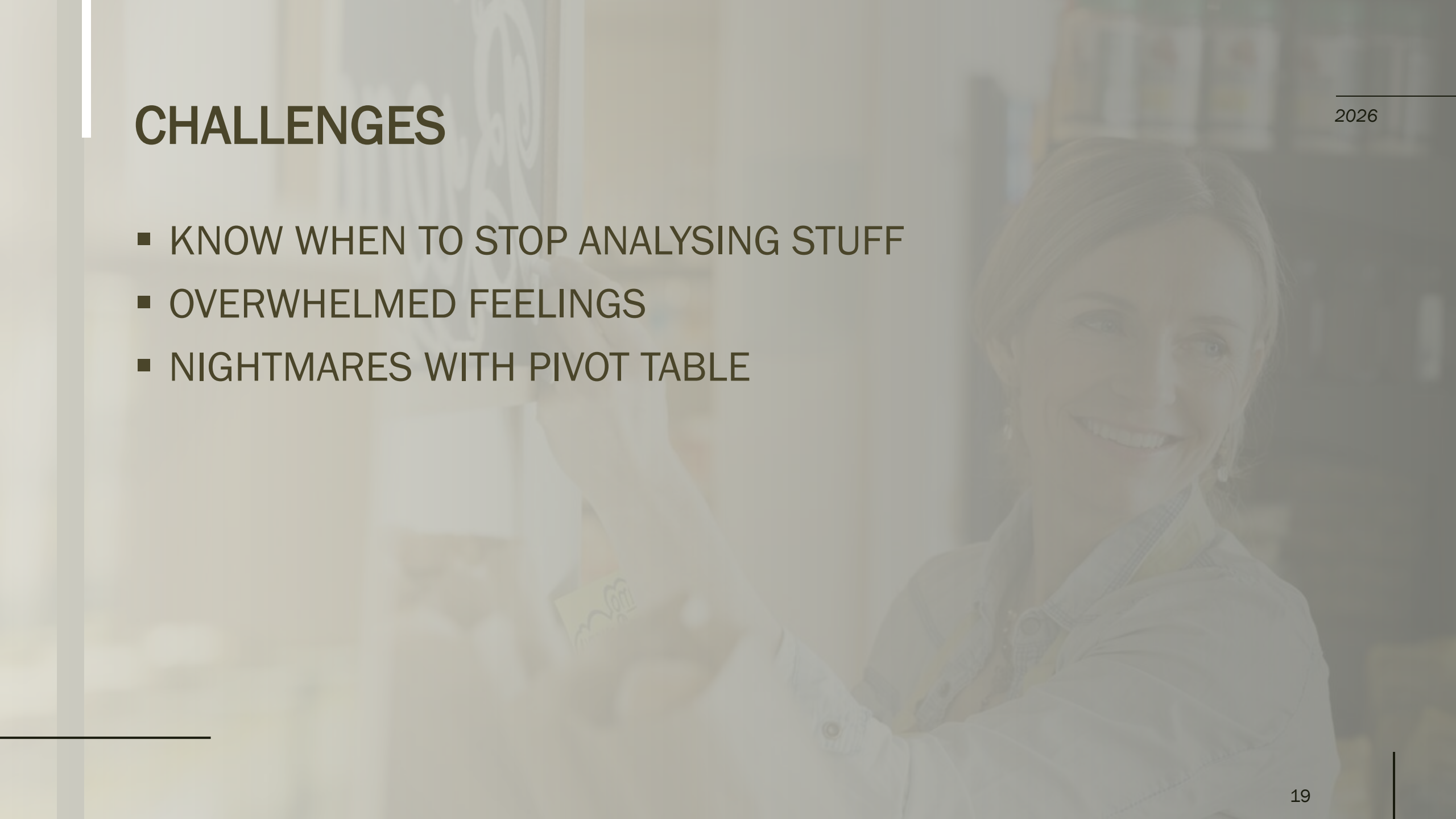
# Business Value

These findings provide a foundation for:

- capacity planning
- Grid modernization
- Energy production strategies

# CHALLENGES

- KNOW WHEN TO STOP ANALYSING STUFF
- OVERWHELMED FEELINGS
- NIGHTMARES WITH PIVOT TABLE



# Thanks

Anne Valvezan and Suzana Souza

Jan-2026