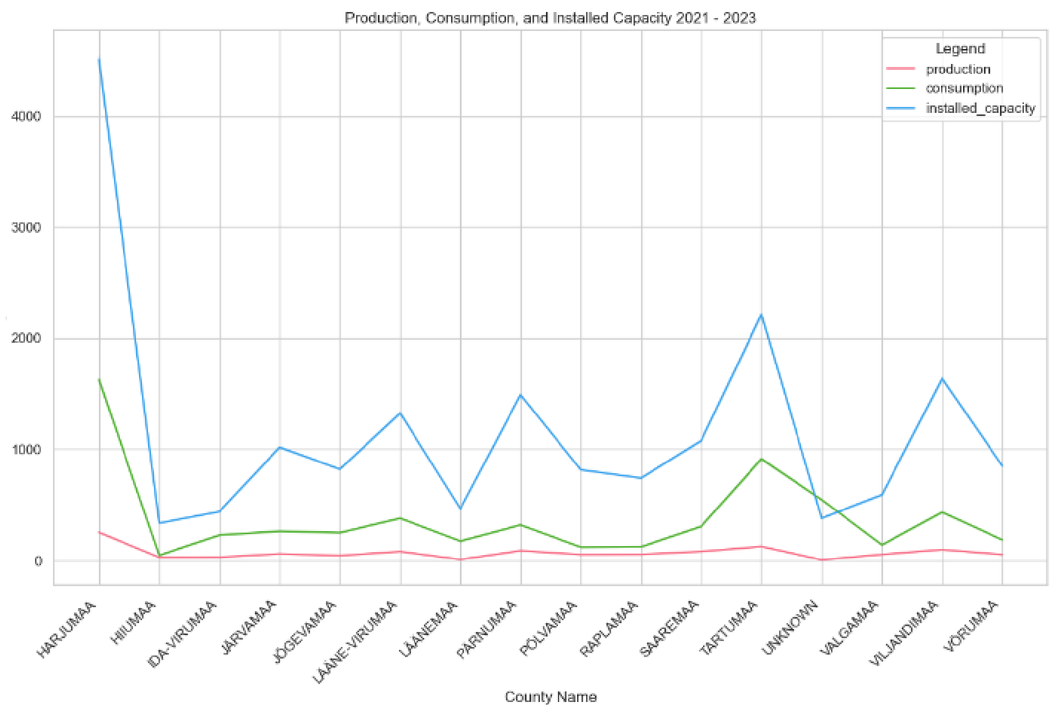
Result and Conclusion

A. Result

The primary objective is to differentiate between the consumption and production patterns of consumers, which serves as the focal point for the model training. (The data saved under csv file named merged_production_consumption_installed_capacity_datetime _solar_radiation).

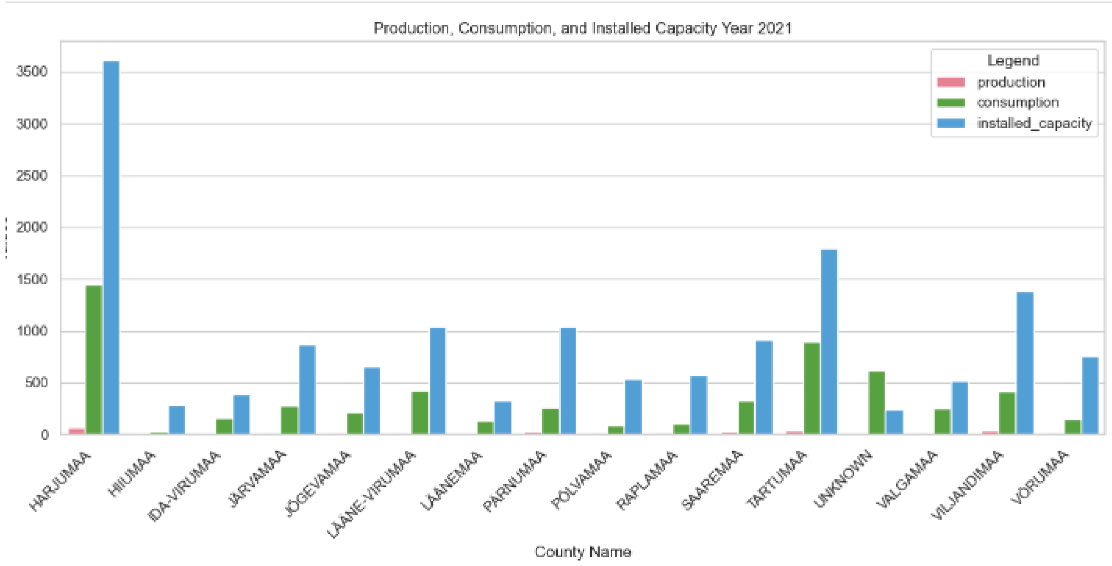
For a comprehensive understanding of the data, visualizations were created as follows:

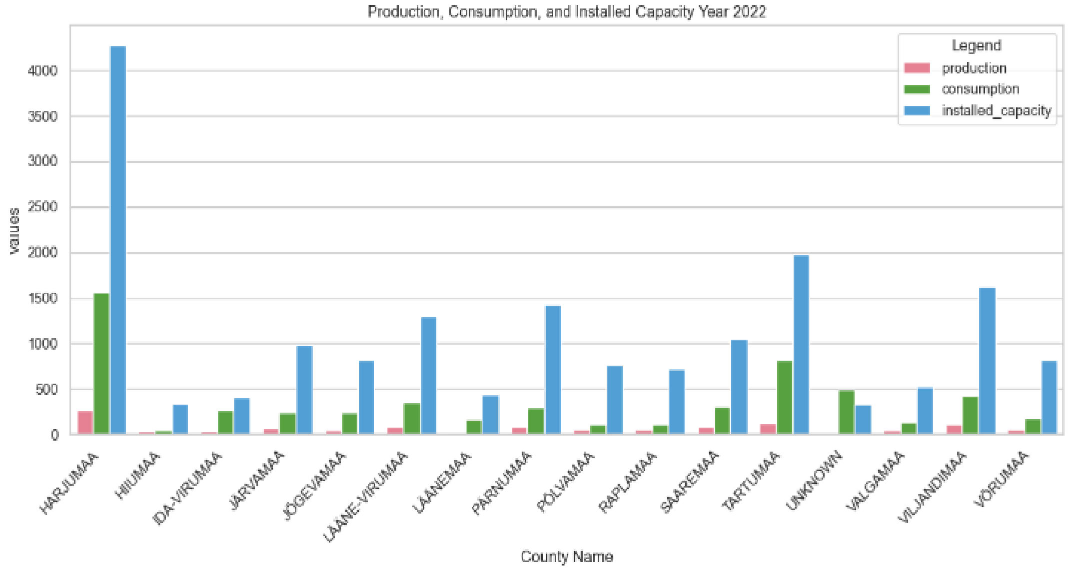
1. Consumption, Production and Installed Capacity All Year for each County

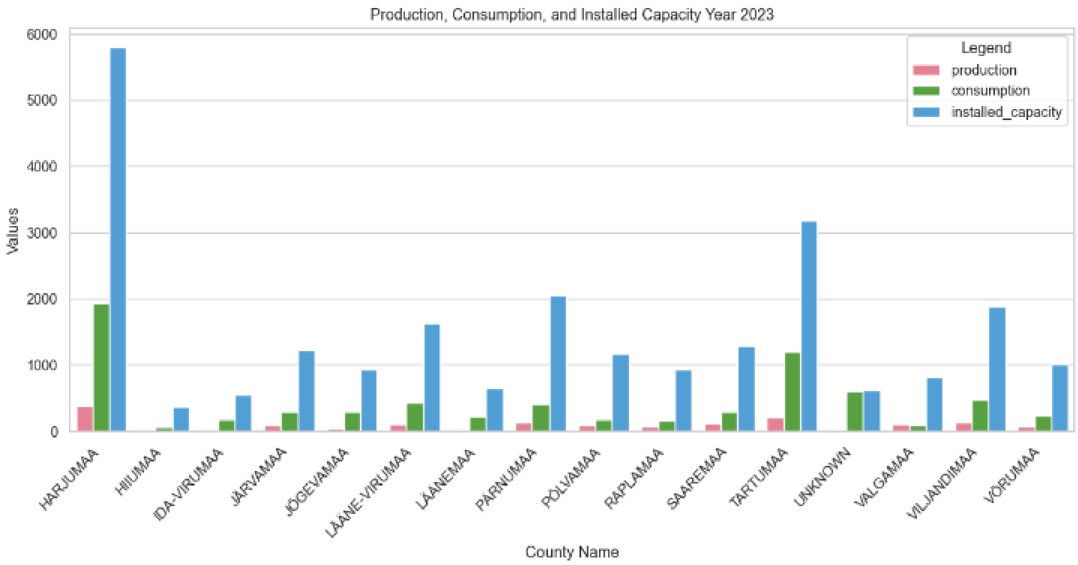


- Reveals the highest and lowest characteristics of the data, providing valuable insights into production and consumption patterns.
- Displays installed capacity, offering insights for assessing the feasibility of adding solar panels.

2. Consumption, Production, and Installed Capacity per Year for each County

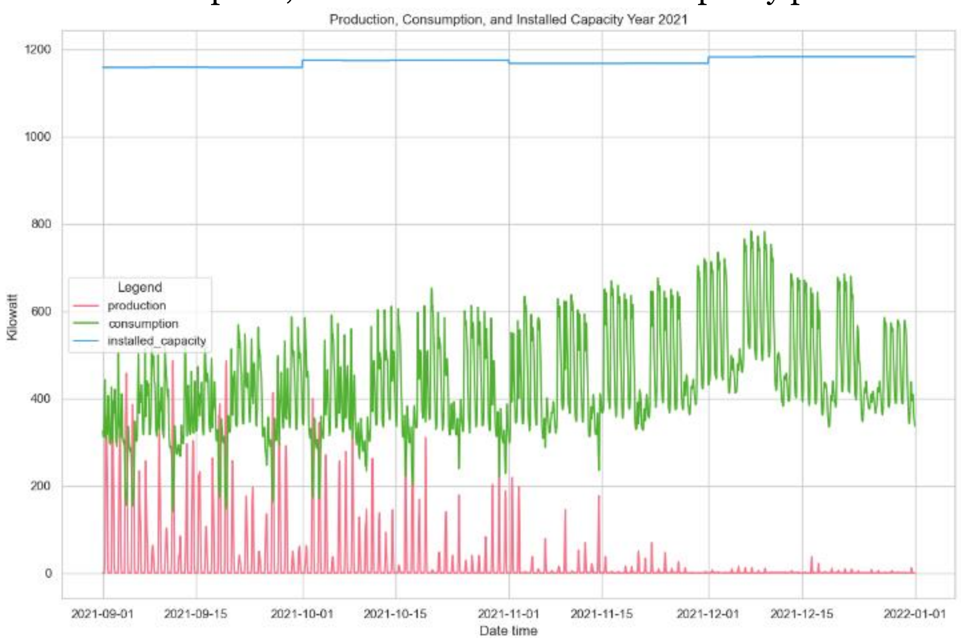


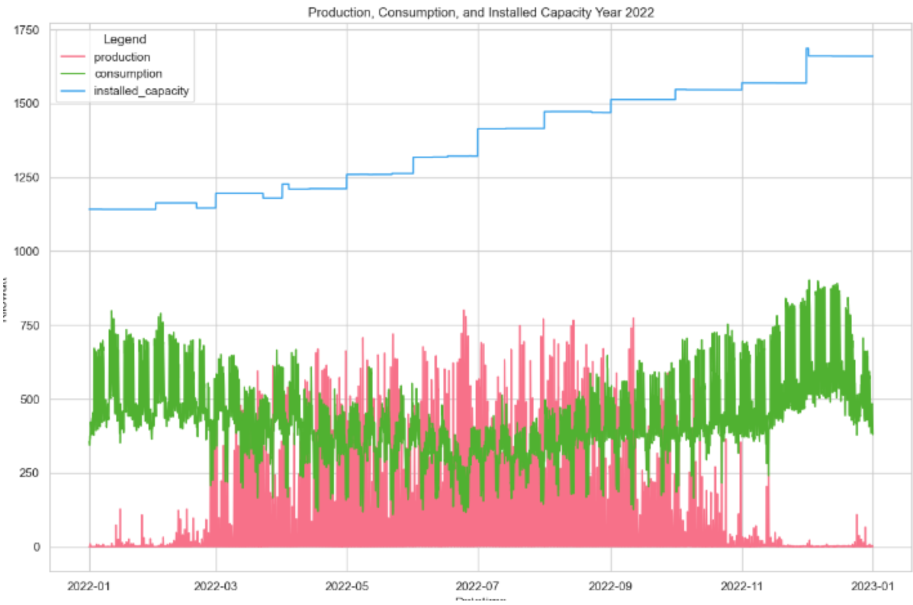


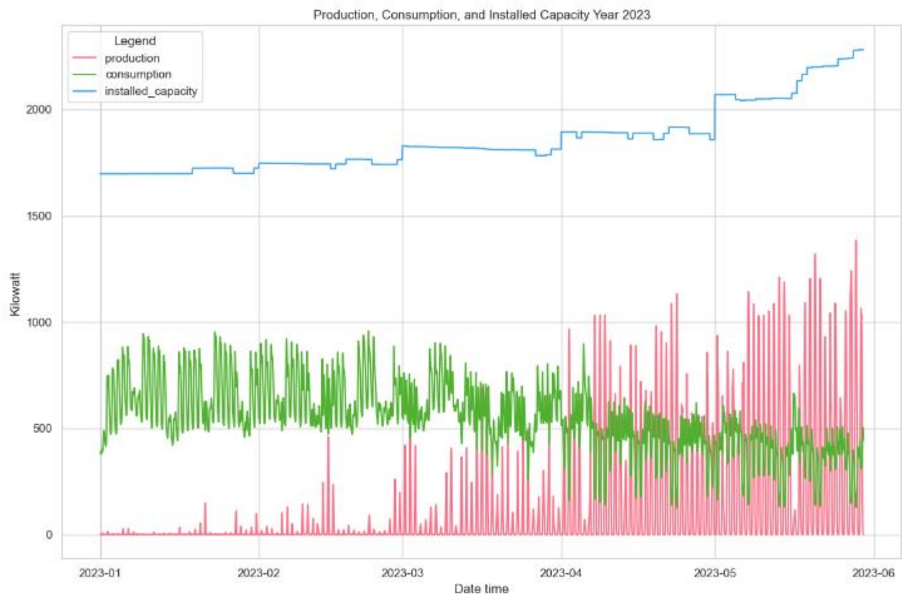


- Offers a detailed view of patterns, facilitating the observation of changes over each year.

3. Mean Consumption, Production and Installed Capacity per Year







- Aids in subsequent stages by visualizing patterns of production and consumption per day in a specific year, facilitating a comprehensive understanding of trends over time.

The consumption and production patterns were initially separated through data wrangling. This data served as the foundational material for entering the feature engineering phase. In the current context, the available data is employed to forecast patterns for the next 24 hours. The obtained results are summarized as follows:

```
df_train = df_train.dropna()
df_train
```

	county	is_business	datetime	is_consumption	target	product_type	target_lag24h	hour	day	weekday	week	month	quarter
2928	0	0	2021-09-02 00:00:00	0	1.687	1	0.713	0	2	3	35	9	3
2929	0	0	2021-09-02 00:00:00	1	109.366	1	96.590	0	2	3	35	9	3
2930	0	0	2021-09-02 00:00:00	0	0.000	2	0.000	0	2	3	35	9	3
2931	0	0	2021-09-02 00:00:00	1	21.008	2	17.314	0	2	3	35	9	3
2932	0	0	2021-09-02 00:00:00	0	1.003	3	2.904	0	2	3	35	9	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2011579	15	1	2023-05-29 23:00:00	1	188.167	0	173.048	23	29	0	22	5	2
2011580	15	1	2023-05-29 23:00:00	0	0.000	1	0.000	23	29	0	22	5	2
2011581	15	1	2023-05-29 23:00:00	1	31.484	1	35.217	23	29	0	22	5	2
2011582	15	1	2023-05-29 23:00:00	0	0.000	3	0.000	23	29	0	22	5	2
2011583	15	1	2023-05-29 23:00:00	1	177.056	3	161.650	23	29	0	22	5	2

2006688 rows × 13 columns

Both the consumption and production datasets have been effectively modeled using LightGBM, a choice driven by its efficiency, speed, and capability to handle larger datasets. LightGBM coupled with GPU acceleration, stems from its commendable efficacy in capturing intricate relationships within the data, offering both accuracy and efficiency in predicting energy patterns. The models are assessed through the Mean Absolute Error (MAE). Here are the results for both MAE :

A. LGBM (First Code)

For Consumption, MAE : 58.01

```
from sklearn.metrics import mean_absolute_error
y_predictions_C = model.predict(X_consumption_test)
MAE_C = mean_absolute_error(y_consumption_test, y_predictions_C)
print(f'MAE_C: {MAE_C}')
```

```
MAE_C: 58.012500816391224
```

For Production, MAE : 39.72

```
from sklearn.metrics import mean_absolute_error
y_predictions = model.predict(X_production_test)
MAE_P = mean_absolute_error(y_production_test, y_predictions)
print(f'MAE_P: {MAE_P}')
```

```
MAE_P: 39.72140607492467
```

B. LGBM (Second Code)

For Consumption, MAE : 56.58

```
from sklearn.metrics import mean_absolute_error
y_predictions = model_C.predict(X_consumption_test)
MAE = mean_absolute_error(y_consumption_test, y_predictions)
print(f'MAE: {MAE}')
```

```
MAE: 56.58965984131783
```

For Production, MAE : 36.54

```
from sklearn.metrics import mean_absolute_error
y_predictions = model_P.predict(X_production_test)
MAE = mean_absolute_error(y_production_test, y_predictions)
print(f'MAE: {MAE}')
```

```
MAE: 36.54153896185464
```

B. Conclusion

In conclusion, the utilization of the LightGBM Regressor for predicting energy consumption and production patterns has yielded valuable insights. Meticulous data wrangling and feature engineering were undertaken, successfully separating, and enriching the datasets, thereby laying the foundation for meaningful modeling.

The resulting visualizations have played a crucial role in highlighting nuanced patterns and trends, contributing to a comprehensive understanding of energy dynamics. These findings not only address the immediate goal of predicting 24-hour patterns but also establish a groundwork for informed decision-making in the energy management of Estonia.

The predictive models, employed on both consumption and production datasets, demonstrated commendable efficacy in capturing intricate relationships within the data. The assessment, performed through the Mean Absolute Error (MAE), served to quantify the accuracy of the models. The achieved MAE result for production, specifically 36.54 using the second code, attests to the reliability of the models in predicting energy patterns.