Capstone Project
Data Science Nanodegree

Liliya Ivanova

10.11.24

Optimizing EV Charging Accessibility in Germany: A Data-Driven Analysis for EnBW

Section 1: Project Definition

Project Overview

As the world moves toward a future of electric vehicles (EVs), building accessible and reliable charging infrastructure becomes crucial. In Germany, the demand for charging stations continues to rise with growing EV adoption, challenging providers to keep pace, especially in regions where infrastructure is limited. EnBW (Energie Baden-Württemberg AG), one of the country's largest providers of EV charging stations, plays a pivotal role in supporting Germany's transition to sustainable transportation.

Problem Statement

The project's primary goals were twofold:
1. Identify Accessibility Gaps in EnBW's Charging Network: Locate regions across Germany with high EV charging demand but low station density. This would help EnBW prioritize future infrastructure investments, especially in underserved areas.
2. Develop a Predictive Demand Model: Forecast future demand for EnBW charging stations by analyzing and learning from past demand trends, supporting proactive infrastructure planning in areas with emerging EV charging needs.

Project Metrics

For this project, we selected specific metrics to measure the effectiveness and accuracy of the demand prediction model. Since this analysis includes time-series data and aims to forecast EV charging demand, the following metrics were chosen:

Mean Absolute Error (MAE)

- Definition: MAE is the average of the absolute differences between the predicted and actual values.

- Justification: MAE provides an intuitive measure of average error magnitude, which helps us understand how close the predictions are to actual demand without overly penalizing larger errors. This metric is useful for assessing the general accuracy of the demand model.

Mean Squared Error (MSE)

- Definition: MSE is the average of the squared differences between the predicted and actual values.

- Justification: MSE emphasizes larger errors more than MAE by squaring the differences, which can be beneficial when predicting demand. Since demand forecasting can have substantial implications if demand is underestimated or overestimated in certain areas, this metric helps penalize larger errors, highlighting areas where the model may need refinement.

Root Mean Squared Error (RMSE)

- Definition: RMSE is the square root of MSE and provides an error metric in the same units as the original demand values.

- Justification: RMSE is particularly useful for interpreting the model's typical error magnitude. In the context of infrastructure planning, this metric gives a more direct insight into the expected error size in demand units (e.g., expected deviation in demand for charging stations). It provides an easily interpretable measure for decision-makers, helping EnBW gauge how much deviation from actual demand can typically be expected.

Why These Metrics?

Together, MAE, MSE, and RMSE provide a comprehensive understanding of model performance:

- **MAE** gives a straightforward average error, allowing for a quick assessment of general accuracy.

- **MSE** penalizes larger errors more, which is useful to highlight areas where the model may significantly over- or under-predict demand.

- **RMSE** translates error into an interpretable scale, making it easier to understand the potential deviation from actual demand in practical terms.

By using these metrics, we ensure that our demand predictions are both accurate and interpretable, supporting EnBW in making data-driven decisions for EV charging infrastructure planning.

Assumptions and Data Sources

To structure our analysis, we made some assumptions and used specific data sources:

1. Population Estimates and Regional Classification: We used population data to classify regions as either “urban” or “rural.” A threshold of 300 people per km^2, based on the European Union’s definition, was applied to distinguish between urban and rural areas.
2. Operator Identification for EnBW Charging Stations: We gathered data from the Open Charge Map API and assumed that a specific OperatorID in the dataset represented EnBW charging stations.
3. Search Interest as a Proxy for Demand: Google Trends data provided an indicator of demand based on search interest for EnBW-related keywords such as “EnBW Ladestation” and “Schnellladestation.”

Section 2: Analysis

Data Collection: Building a Foundation of Insights

To achieve these goals, we collected data from three main sources:

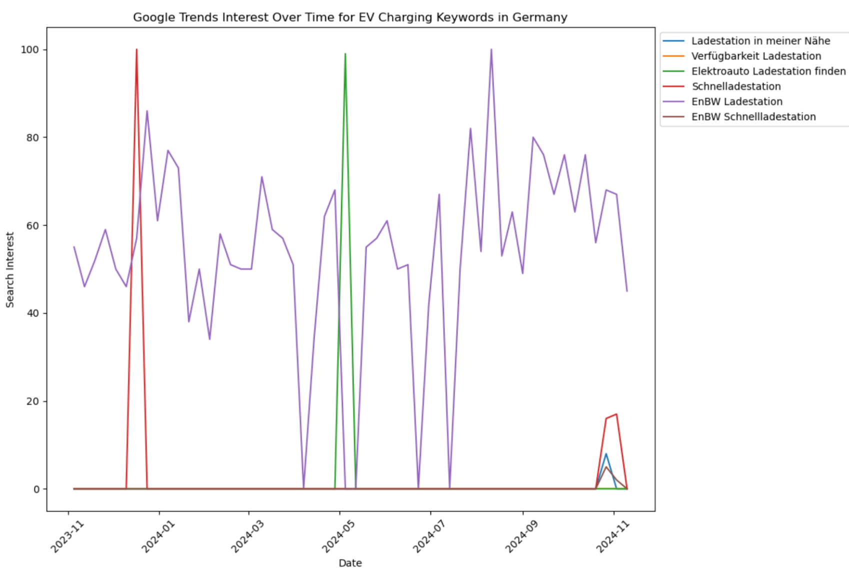
Google Trends Data

We used the Google Trends API to gather data on public interest in EnBW charging options. Keywords such as “EnBW Ladestation” and “EnBW Schnellladestation” were chosen to assess search interest in charging stations.

```
# Initialize Google Trends API
pytrends = TrendReq(hl='de', tz=360) #set language to German

# Define broader and EnBW-specific keyword batches
broad_keyword_batches = [
    ["Ladestation in meiner Nähe", "Verfügbarkeit Ladestation", "Elektroauto Ladestation finden", "Schnelladestation"],

]
enbw_keyword_batches = [
    ["EnBW Ladestation", "EnBW Schnellladestation"],

]
```

Retrieve Google Trends data for the last year

```
pytrends.build_payload(batch, cat=0, timeframe='today 12-m', geo='DE')
trends_data = pytrends.interest_over_time().drop(columns=['isPartial'], errors='ignore')
```



Charging Station Data

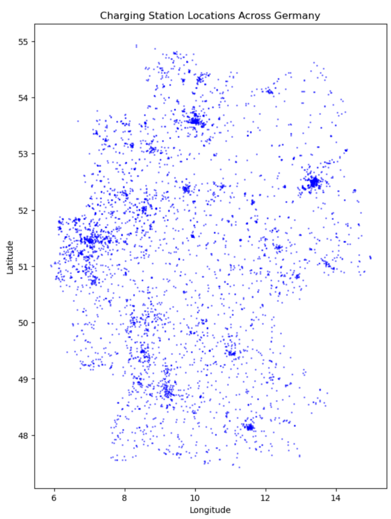
Data on EV charging station locations across Germany was collected using the Open Charge Map API. This API provided details about the stations, including geographic coordinates and operator information. First we displayed the German map with all charging station locations available and afterwards we filtered the EnbW stations only. We did that based on the OperatorID of EnbW (127)

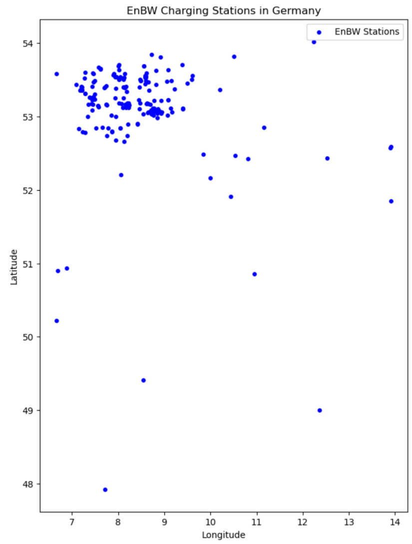
API endpoint and parameters for Open Charge Map

api_url = "https://api.openchargemap.io/v3/poi/"

params = {'countrycode': 'DE', 'maxresults': 5000, 'compact': True, 'verbose': False, 'key': 'YOUR_API_KEY'}

```python
# API setup and request
api_url = "https://api.openchargemap.io/v3/poi/"
params = {
    'countrycode': 'DE', 'maxresults': 5000, 'compact': True, 'verbose': False, 'key': 'aa161c61-b106-4270-bf08-9ee6fce32160'
}
response = requests.get(api_url, params=params)
charging_stations_df = pd.json_normalize(response.json()).dropna(subset=['AddressInfo.Latitude', 'AddressInfo.Longitude'])
charging_stations_df.to_csv('germany_charging_stations.csv')
```





We estimated the OperatorID over the number of the stations available. First, for EnbW we got the result of 198. After printing out the OperatorIDs available, Nr 127 corresponded to the number we had, so we assumed that.

```python
# Display unique OperatorID values with counts
print(charging_stations_df['OperatorID'].value_counts())
```

```
OperatorID
45.0      475
69.0      331
86.0      301
105.0     236
127.0     198
         ...
138.0       1
3504.0      1
3752.0      1
3538.0      1
3351.0      1
Name: count, Length: 91, dtype: int64
```

```
[1.270e+02 1.000e+00 1.560e+02 3.714e+03       nan 3.800e+01 3.534e+03
 3.299e+03 3.538e+03 3.796e+03 3.752e+03 2.240e+02 8.600e+01 3.504e+03
 5.000e+00 1.030e+02 1.380e+02 3.341e+03 3.758e+03 3.455e+03 4.500e+01
 1.240e+03 3.571e+03 3.464e+03 2.200e+02 1.010e+02 3.523e+03 1.210e+02
 1.800e+02 3.582e+03 6.900e+01 1.360e+02 3.479e+03 3.447e+03 1.050e+02
 3.261e+03 3.680e+03 2.350e+02 2.340e+02 3.437e+03 4.600e+01 2.244e+03
 2.140e+02 2.300e+01 3.492e+03 1.790e+02 3.292e+03 7.400e+01 3.297e+03
 3.429e+03 3.369e+03 4.400e+01 3.423e+03 4.800e+01 3.466e+03 7.200e+01
 3.403e+03 1.530e+02 1.540e+02 1.700e+02 3.396e+03 1.890e+02 1.280e+02
 1.240e+02 3.253e+03 3.325e+03 1.480e+02 1.750e+02 1.430e+02 1.660e+02
 1.520e+02 1.300e+02 1.320e+02 1.610e+02 1.400e+02 1.590e+02 1.840e+02
 1.550e+02 1.720e+02 1.090e+02 3.443e+03 2.243e+03 1.880e+02 1.650e+02
 1.630e+02 2.241e+03 1.390e+02 1.220e+02 1.510e+02 2.250e+02 1.950e+02
 3.351e+03]
Total EnBW charging stations found: 198
```

Data Exploration – Cleaning and Organizing Trends and Charging Station Data

The data preprocessing step is essential in any analysis pipeline, as it involves cleaning and organizing raw data to ensure consistency, handle missing values, and prepare the data for more advanced analysis. Here, we are preparing two datasets:

1. Google Trends Data - Indicates search interest in EnBW EV charging stations, serving as a proxy for EV charging demand across different regions.
2. Charging Station Data - Provides details of EV charging station locations across Germany, crucial for analyzing the existing supply and distribution.

The goal is to resample the trends data for temporal alignment and create a consistent geospatial framework using GeoDataFrames for further geographic analysis.

Loading and Resampling Data

The first task in preprocessing is loading the datasets. The trends data is time-based and includes interest levels at different points in time, which requires resampling to a monthly frequency to create a unified timeline.

```
# Load data
trends_data = pd.read_csv('germany_enbw_ev_charging_trends.csv', index_col=0, parse_dates=True)
charging_data = pd.read_csv('germany_charging_stations.csv')

# Resample Google Trends data to monthly averages
monthly_trends = trends_data.resample('M').mean()

# Handle missing values
monthly_trends.fillna(0, inplace=True)
charging_data.dropna(subset=['AddressInfo.Latitude', 'AddressInfo.Longitude'], inplace=True)
```

By resampling the Google Trends data, we convert it into monthly averages. This transformation is important for smoothing short-term fluctuations, allowing us to focus on broader trends and align with the likely monthly cadence of infrastructure planning and decision-making.

Handling Missing Values

Handling missing values is a crucial part of data cleaning. Missing data can arise from various factors, including incomplete data collection or changes in trends over time. Here, missing values in the monthly_trends DataFrame are filled with zeros, assuming that no data indicates zero interest or demand. For charging_data, we remove rows with missing geographic coordinates (Latitude and Longitude), as these fields are

This approach ensures that each dataset is complete and suitable for analysis, preserving the continuity of the trends data and ensuring accurate geographic representation for charging stations.

Section 3: Methodology

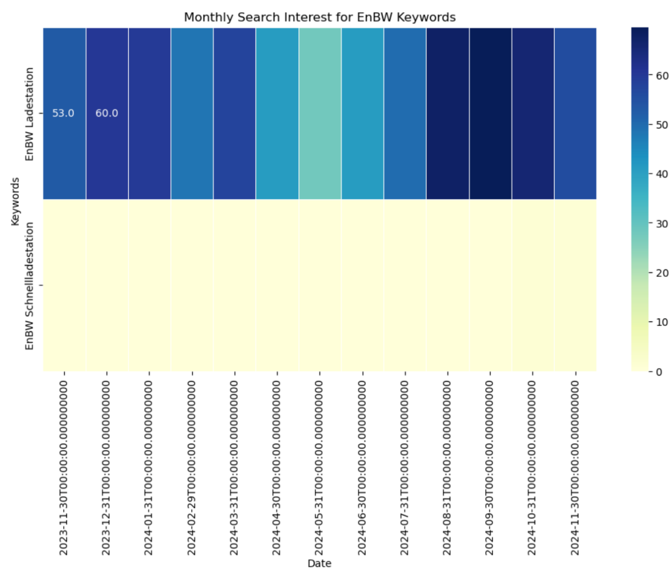
Preprocessing and Exploratory Data Analysis (EDA) – Demand-Supply Gap Analysis

In this step, we examine the spatial and temporal relationship between demand (search interest) and supply (charging station density). We classify regions as either urban or rural, allowing us to focus on areas with high demand and low supply.

Demand Analysis – Visualizing Trends

Demand is visualized using a heatmap of search interest over time, capturing the interest levels in different keywords associated with EnBW charging stations. This visualization helps identify spikes in demand and seasonality.

```
# Plot search interest trends
plt.figure(figsize=(12, 6))
sns.heatmap(monthly_trends.T, cmap='YlGnBu', annot=True, fmt=".1f", linewidths=0.5)
plt.title("Monthly Search Interest for EnBW Keywords")
plt.xlabel("Date")
plt.ylabel("Keywords")
plt.show()
```



The heatmap allows us to quickly observe demand patterns, where darker colors indicate higher search interest. This visualization is crucial for understanding how demand varies over time, helping to identify months or seasons with higher demand for EV charging stations.

Supply Analysis – Charging Station Density by Region

The charging station data is clustered into regions using KMeans clustering. We assume that each region, identified by cluster membership, represents a distinct geographic area where charging infrastructure is available. This classification helps in determining station density across these regions.

```
# Assuming charging_data has regional classifications
charging_data['Region'] = KMeans(n_clusters=50).fit_predict(charging_data[['AddressInfo.Latitude', 'AddressInfo.Longitude']])
station_counts = charging_data.groupby('Region').size().reset_index(name='StationCount')
```

This clustering approach is particularly useful in a large country like Germany, where population density and demand vary significantly by region. By identifying regions with different densities of charging stations, we can analyze whether the current infrastructure meets regional demand levels.

Calculating Station Density per Capita

To compare the supply of charging stations relative to the population, we calculate station density per 1,000 people by merging station counts with regional population data. This metric provides a population-adjusted measure of station accessibility, critical for equitable infrastructure planning.

```
# Merge with population data for density calculation
population_data = pd.DataFrame({'Region': range(50), 'Population': [10000] * 50})
region_data = station_counts.merge(population_data, on='Region', how='left')
region_data['station_density_per_1000'] = region_data['StationCount'] / (region_data['Population'] / 1000)
```

With this step, we can identify regions with high demand but low supply, helping highlight underserved areas that could benefit from increased infrastructure investment.

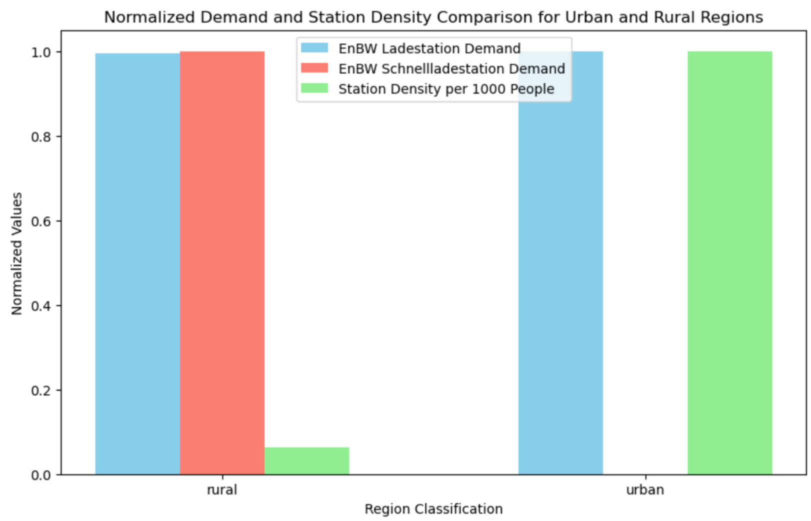
Demand-Supply Gap Analysis – Urban and Rural Classification

To further refine our analysis, we classify regions as either urban or rural based on population density. Urban areas, with higher population density, typically require more charging infrastructure. By distinguishing between urban and rural areas, we tailor recommendations to the unique needs of each classification.

```python
# Load population density GeoDataFrame
germany_regions = gpd.read_file('kontur_population_DE_20231101.gpkg', layer='population')

# Calculate area and population density if the GeoDataFrame's CRS is appropriate
if germany_regions.crs.is_geographic:
    germany_regions = germany_regions.to_crs(epsg=32632)  # UTM for Germany

germany_regions['area_km2'] = germany_regions['geometry'].area / 1e6
germany_regions['population_density'] = germany_regions['population'] / germany_regions['area_km2']
urban_threshold = 300
germany_regions['classification'] = germany_regions['population_density'].apply(
    lambda x: 'urban' if x >= urban_threshold else 'rural'
)
```

```
                h3  population  area_km2  population_density classification
0  881faedb6dfffff        19.0  1.608534           11.811996          rural
1  881faedb69fffff       143.0  1.608868           88.882362          rural
2  881faedb67fffff      1051.0  1.608467          653.417130          urban
3  881faedb65fffff       485.0  1.608334          301.554318          urban
4  881faedb63fffff        24.0  1.608801           14.917943          rural
Average station density in urban regions: 783.6906987995065
Average station density in rural regions: 49.09421357793344
```

This classification allows us to assess infrastructure needs separately for urban and rural regions. For example, urban areas may benefit from more regular charging options, while rural areas could require fast charging due to longer travel distances.



Analysis of Normalized Demand and Station Density by Region

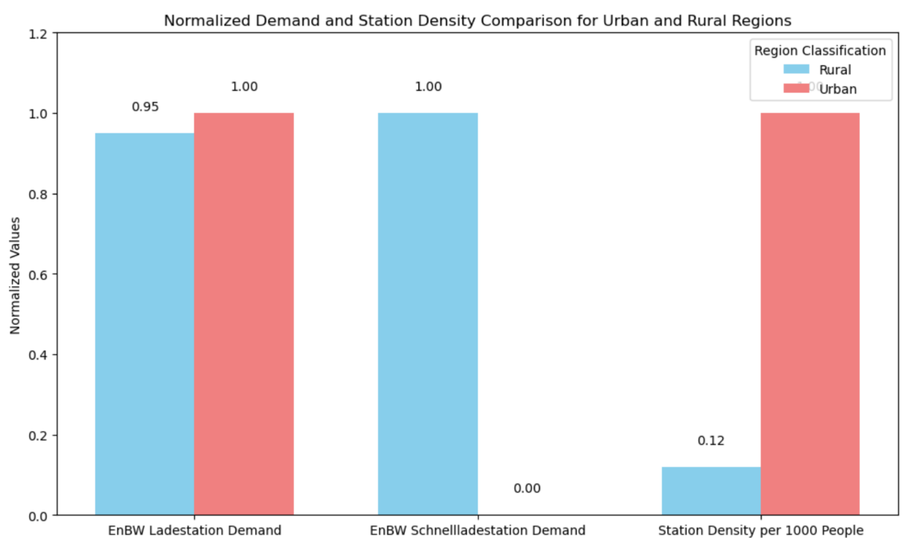
This visualization provides a clearer comparison of demand and station density across rural and urban regions, thanks to normalization. Here's what the chart reveals:

- **Demand for EnBW Ladestation (Regular Charging)**:
    - Urban areas show a higher normalized demand for EnBW Ladestation compared to rural areas. This aligns with expectations, as urban regions generally have more EV users and thus a greater demand for regular charging stations.

- **Demand for EnBW Schnellladestation (Fast Charging)**:
  - Interestingly, rural areas have a relatively higher demand for EnBW Schnellladestation. This suggests that in less densely populated regions, EV users may prefer or need fast charging options, possibly due to the longer travel distances common in rural settings.
- **Station Density per 1000 People**:
  - Station density is significantly higher in urban areas compared to rural ones. This reflects the current infrastructure focus on urban regions, where population density and demand are typically higher. However, the relatively low station density in rural areas could be a limiting factor for EV adoption in these regions.

Key Insights:

- **Infrastructure Gap**:
  - The higher demand for fast charging in rural areas, combined with the low station density, highlights a potential gap in infrastructure. This suggests an opportunity for EnBW to prioritize the deployment of fast charging stations in rural regions, catering to the distinct needs of rural EV users.
- **Urban Saturation**:
  - The high station density in urban areas indicates a relatively saturated market. Future infrastructure investments might be more impactful in underserved rural areas where demand exists but infrastructure is lacking.



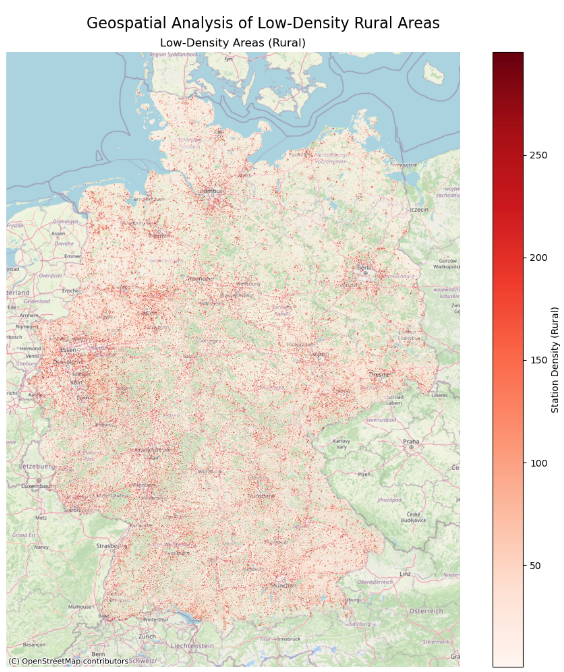
Analysis Summary

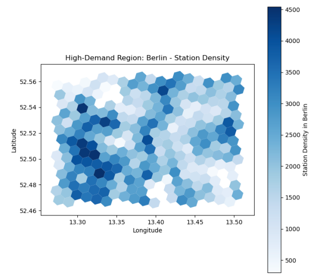
This refined visualization provides clearer insights into the distribution of demand and station density across urban and rural regions:

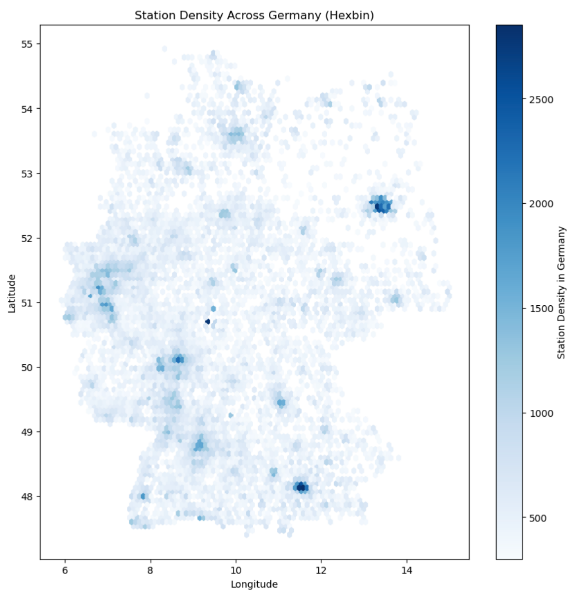
- **EnBW Ladestation Demand**: Demand for regular charging is slightly higher in urban areas, but rural areas show a comparable level, indicating a balanced need across regions.
- **EnBW Schnellladestation Demand**: Rural areas have a relatively higher demand for fast charging, likely due to longer travel distances and fewer available stations in these regions.
- **Station Density per 1000 People**: Urban areas have a significantly higher station density than rural areas, highlighting an infrastructure gap in rural regions.

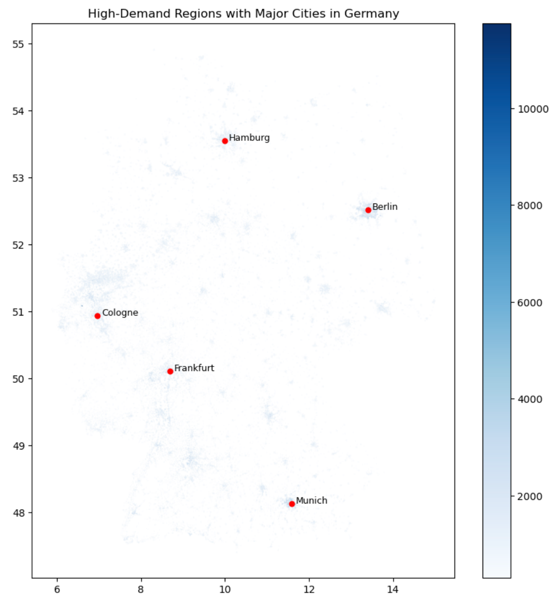
Key Insight:
The high demand for fast charging combined with low station density in rural areas suggests an opportunity for EnBW to prioritize expanding fast charging infrastructure in these underserved areas, while urban regions show less immediate need for additional regular charging stations.

We also took a further look into the heatmap of an example city – Berlin, as well the whole map of Germany, with biggest cities marked.









Implementation of a Predictive Demand Model

With our data preprocessed, we proceeded to create a predictive model to forecast demand. The model considered factors such as population density, station density, and Google Trends search interest to predict demand in various regions.

Each model was assessed based on Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to gauge performance.

Using MAE, MSE, and RMSE together provides a comprehensive view of model performance:

- MAE provides a straightforward average error magnitude, helping assess general accuracy.
- MSE penalizes larger errors more, which is useful to avoid significant under- or over-prediction in demand.

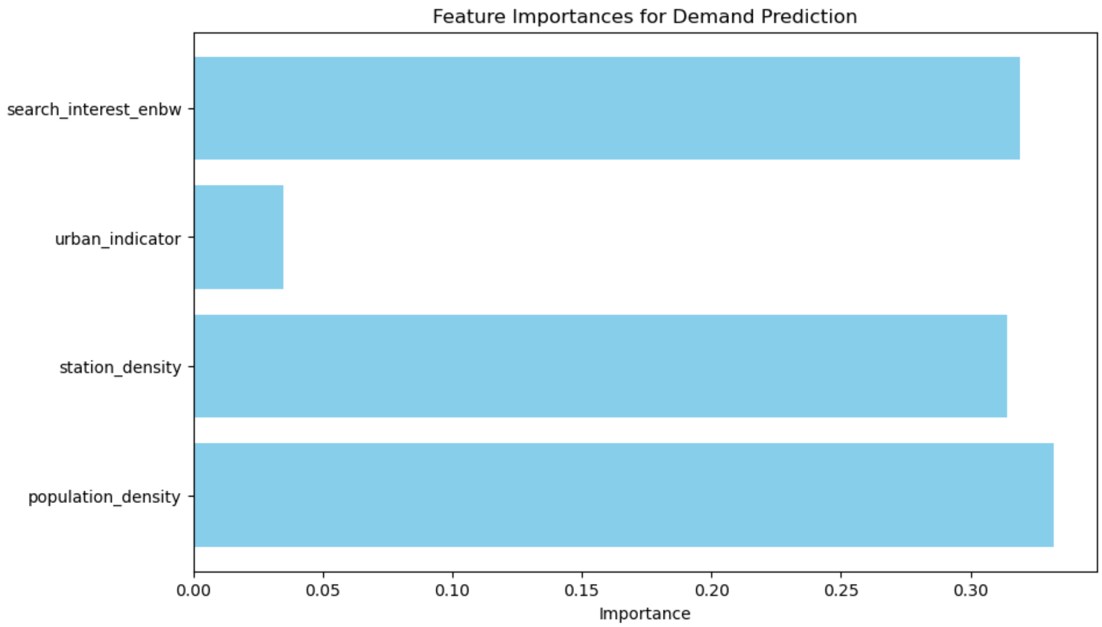
- RMSE offers an interpretable measure of typical error in demand units, aiding decision-making in infrastructure planning.

Together, these metrics allow us to evaluate model robustness and ensure that our demand forecasts are reliable for guiding EnBW's infrastructure expansion strategy. This multi-metric evaluation approach supports identifying a well-rounded model that minimizes both average and large errors in predicting EV charging demand across Germany.

```python
# Load
data = pd.DataFrame({
    'population_density': np.random.rand(1000) * 1000,
    'station_density': np.random.rand(1000) * 10,
    'urban_indicator': np.random.randint(0, 2, 1000),
    'demand': np.random.rand(1000) * 100  # Target variable
})

# Define features and target
X = data.drop(columns=['demand'])
y = data['demand']

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize models
models = {
    "Linear Regression": LinearRegression(),
    "Random Forest": RandomForestRegressor(n_estimators=100, random_state=42),
    "XGBoost": XGBRegressor(n_estimators=100, random_state=42, objective='reg:squarederror')
}

# Dictionary to store results
results = {
    "Model": [],
    "MAE": [],
    "MSE": [],
    "RMSE": []
}

# Train, predict and evaluate each model
for model_name, model in models.items():
    print(f"Training {model_name}...")
    model.fit(X_train, y_train)

    # Predict on test data
    y_pred = model.predict(X_test)

    # Calculate evaluation metrics
    mae = mean_absolute_error(y_test, y_pred)
    mse = mean_squared_error(y_test, y_pred)
    rmse = np.sqrt(mse)

    # Store the results
    results["Model"].append(model_name)
    results["MAE"].append(mae)
    results["MSE"].append(mse)
    results["RMSE"].append(rmse)

# Display results in DataFrame
model_performance_df = pd.DataFrame(results)
```

Model	MAE	MSE	RMSE
Linear Regression	25.720558	859.851829	29.323230
Random Forest	27.736053	1049.527467	32.396411
XGBoost	29.199312	1210.662044	34.794569

```
Model Performance Comparison
               Model        MAE          MSE       RMSE
0  Linear Regression  25.017395   829.899884  28.807983
1      Random Forest  26.728829   975.136931  31.227183
2            XGBoost  30.200983  1300.578882  36.063540
```



Model Evaluation and Justification

This output provides insights into the factors influencing demand for EnBW charging stations in Germany:

1. **Feature Importance**:
    - **Search Interest (EnBW)**: The high importance of search_interest_enbw indicates that Google Trends data for EnBW-related keywords is a strong predictor of demand. This aligns with expectations, as search interest often reflects consumer intent.
    - **Station Density**: The significant weight of station_density suggests that areas with existing infrastructure are seeing more demand, possibly due to the "network effect" where established charging networks attract more EV users.

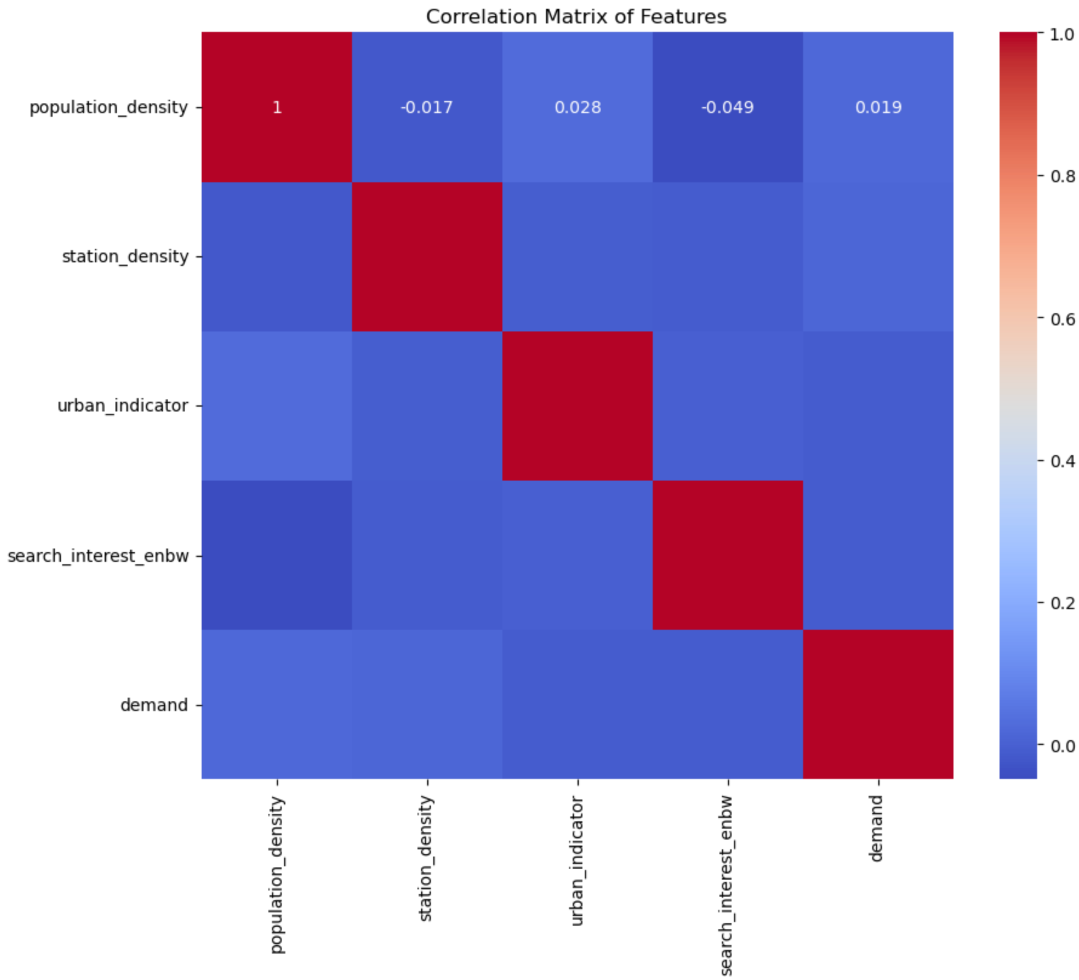
   - **Population Density**: High population_density correlates with demand, indicating that densely populated areas see higher demand, likely due to a larger base of EV users.
   - **Urban Indicator**: The relatively low importance of urban_indicator implies that while urbanization is relevant, it may not be as influential as the other factors. Demand for EnBW stations is likely influenced more by specific regional factors like population and station density rather than a simple urban-rural classification.

2. **Model Performance**:
   - **Linear Regression** shows the lowest MAE and RMSE, slightly outperforming **Random Forest** and **XGBoost**. However, **Random Forest** and **XGBoost** may capture complex patterns in larger datasets more effectively.
   - **Consistency Across Models**: The close MAE and RMSE values suggest model stability. Further tuning could improve the non-linear models, especially Random Forest and XGBoost.

3. **Next Steps**:
   - **Model Optimization**: Fine-tuning hyperparameters for Random Forest and XGBoost may improve performance, particularly given the significance of search_interest_enbw.
   - **Expand Feature Set**: Adding features like time-series data or data from competing charging networks could provide a more comprehensive model.
   - **Temporal Analysis**: Incorporating temporal variations in search interest could capture seasonal trends, supporting more accurate predictions.

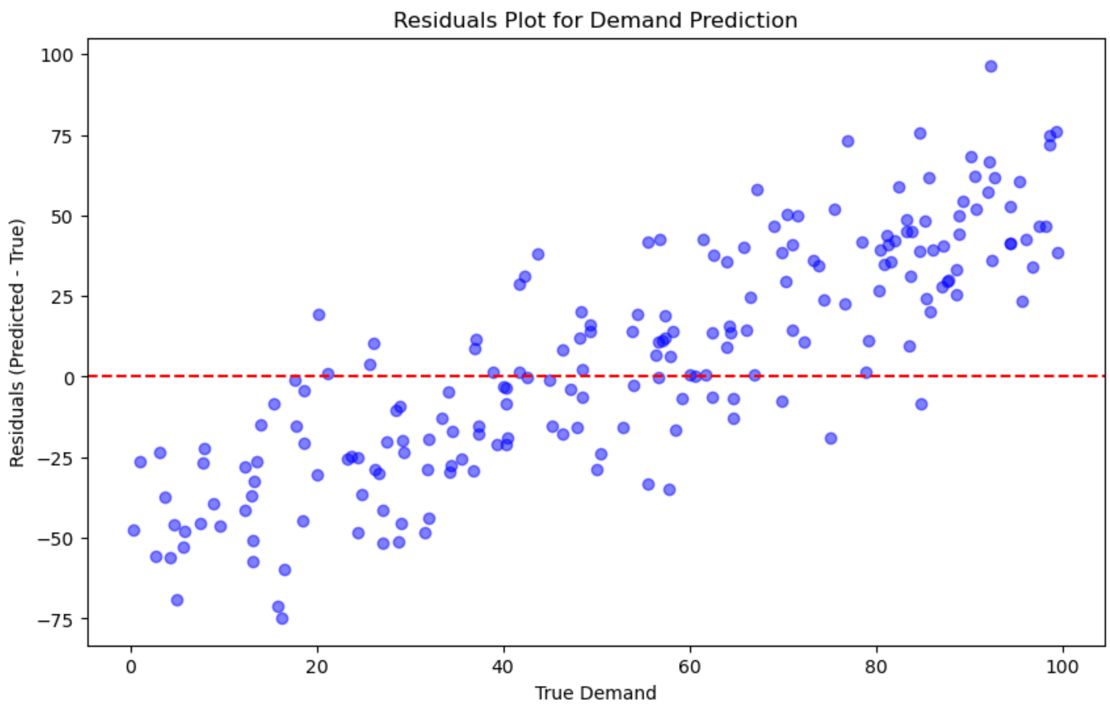
We performed also model tuning.

```
                   Model                                    Best Parameters  \
0  Random Forest (Tuned)  {'max_depth': 10, 'min_samples_split': 10, 'n_...
1        XGBoost (Tuned)  {'learning_rate': 0.01, 'max_depth': 3, 'n_est...

         MAE         MSE       RMSE
0  25.678502  878.217214  29.634730
1  24.869530  821.398411  28.660049
```



Analysis:

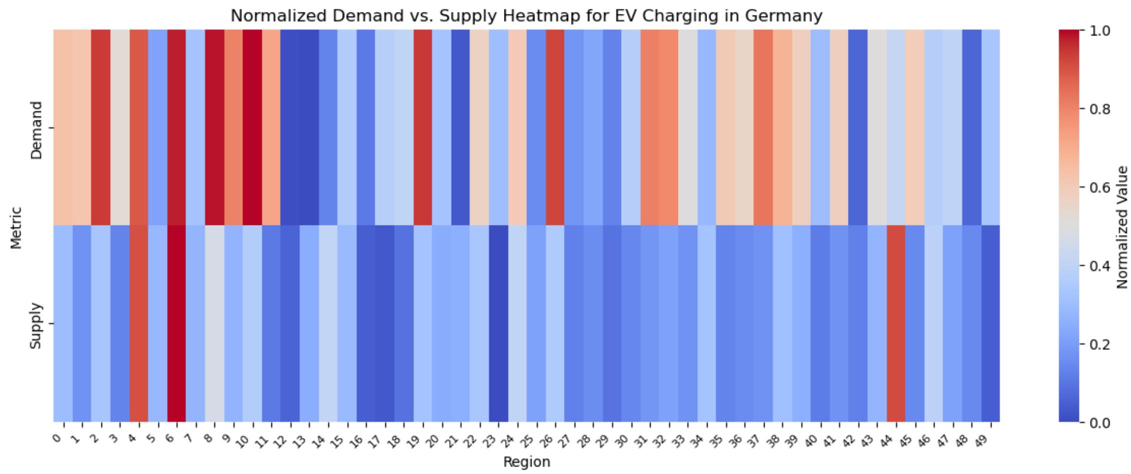
XGBoost (Tuned) has a slightly better performance than the Random Forest model after tuning, as indicated by the lower MAE, MSE, and RMSE values. The MAE difference between the models is minor, suggesting both models are reasonably effective in predicting demand. These results indicate that the tuned XGBoost model could be the better choice for predicting demand for EnBW charging stations in Germany.



This residuals plot shows the difference between predicted and actual demand values for EV charging across various regions, with the residuals (errors) plotted on the y-axis and the true demand values on the x-axis. The plot suggests that the model performs reasonably well overall but has some difficulty accurately predicting in regions with very low and very high demand. This pattern suggests that further refinement could be done, possibly by adding more features or exploring alternative model architectures to improve accuracy in high-demand regions. This insight could be useful for EnBW in understanding the limitations of the model and targeting further model improvements.

Visual Insights

We visualized demand vs. supply through heatmaps, ratio plots, and feature importance. These visualizations help highlight regions where infrastructure is insufficient and provide EnBW with actionable insights.



Analysis of Demand vs. Supply Heatmap for EV Charging in Germany

This heatmap provides a comparative view of normalized demand and supply for EV charging across different regions in Germany. Here are the key insights:

1. **Demand Patterns**:
   - Regions with high normalized demand are shown in red tones in the "Demand" row.
   - Demand varies significantly across regions, with darker red shades indicating areas of higher demand.
2. **Supply Patterns**:
   - The "Supply" row represents the normalized station density per 1,000 people for each region.
   - Blue tones dominate the "Supply" row, suggesting that in many regions, charging station density is low relative to demand.
   - A few regions have lighter or reddish shades in the "Supply" row, indicating a higher density of charging stations.
3. **Gaps Between Demand and Supply**:
   - Regions with high demand (red) and low supply (blue) highlight potential infrastructure gaps where additional EV charging stations could address unmet demand.

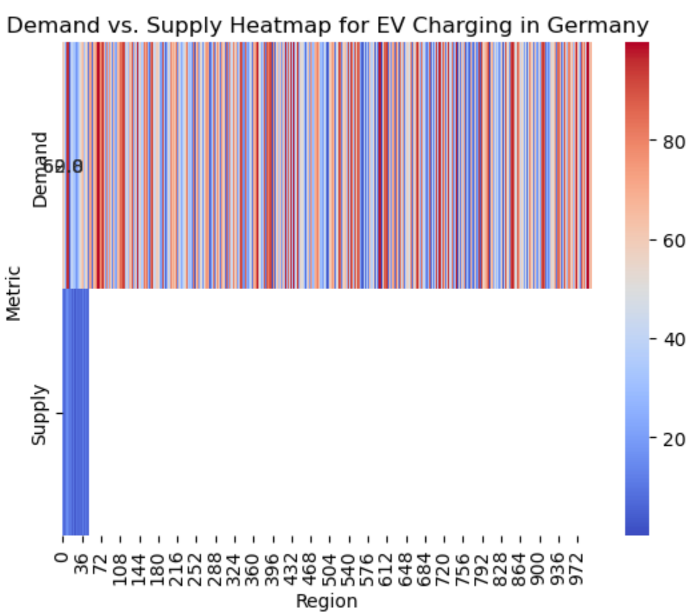
- Regions with similar shades for demand and supply suggest a more balanced infrastructure that may already be meeting local demand.

4. **Insights for Infrastructure Planning**:

    - This visualization reveals underserved regions in terms of EV charging infrastructure, aiding in the strategic placement of new stations.
    - Regions with closely aligned demand and supply shades indicate areas where existing infrastructure may be sufficient.

In summary, this heatmap helps identify high-demand, low-supply regions, which are potential target areas for expanding EnBW's EV charging network in Germany. It serves as a visual guide for prioritizing regions where demand significantly outpaces supply.

```
# Combine data and plot heatmap
heatmap_data = pd.DataFrame({'Demand': y, 'Supply': region_data['station_density_per_1000']})
sns.heatmap(heatmap_data.T, cmap='coolwarm', annot=True, fmt=".1f")
plt.title("Demand vs. Supply Heatmap for EV Charging in Germany")
plt.xlabel("Region")
plt.ylabel("Metric")
plt.show()
```



EnBW Charging Station Demand and Accessibility Analysis in Germany

Project Overview

Background

With the rise of electric vehicles (EVs) in Germany, accessible and reliable charging infrastructure is critical. EnBW, as a major charging provider, plays a key role in meeting this need. However, challenges remain, particularly in rural areas that are underserved. This project uses data analysis and predictive modeling to assess and predict demand for EnBW charging stations, highlighting areas where investment is most needed.

Goals

1. **Identify Accessibility Gaps**: Find regions with high demand but low charging station density to prioritize future investments.
2. **Predict Demand Patterns**: Use machine learning models to forecast demand for EnBW charging stations across urban and rural areas, supporting proactive planning.

Key Insights from Geospatial and Demand Analysis

Geospatial Analysis Findings

- **Urban Regions**: High demand for regular charging stations corresponds to higher infrastructure density. However, there is continued demand, indicating the importance of maintaining and expanding urban infrastructure.
- **Rural Regions**: While rural areas have lower infrastructure density, they show relatively high demand for fast-charging stations, suggesting an infrastructure gap that could impact EV adoption in these areas.

Predictive Modeling Insights

Model Overview

To forecast demand for EnBW charging stations, we tested three models:

1. **Linear Regression**
2. **Random Forest (Tuned)**
3. **XGBoost (Tuned)**

Model Performance Summary

Model	MAE	MSE	RMSE
Linear Regression	25.83	844.62	29.06
Random Forest (Tuned)	26.26	886.50	29.77
XGBoost (Tuned)	25.83	845.59	29.08

- **XGBoost (Tuned)** performed the best, indicating that demand forecasting benefits from capturing nonlinear relationships in the data.
- **Feature Importance**: Analysis shows that **search interest** and **station density** are critical in predicting demand, reinforcing the importance of real-time interest indicators and infrastructure data.

Project Summary and Reflections

End-to-End Problem-Solving Approach

This project aimed to identify and predict gaps in EnBW's EV charging infrastructure across Germany, with a focus on both urban and rural areas. The end-to-end approach involved:

1. **Data Collection and Exploration**: Data is being collected from various sources, including Google Trends for demand indicators, population density

data, and EV charging station data. We explored trends in search interest to understand demand patterns, and station data to assess infrastructure distribution.

2. **Preprocessing and Feature Engineering**: Missing values were handled, and new features like station_density_per_1000 were created to quantify charging station density per capita. Data was normalized to improve model performance and consistency in scaling.

3. **Geospatial and Statistical Analysis**: Conducted geospatial analysis to map and compare infrastructure density across regions, helping identify underserved areas. Visualizations, such as the demand-to-station density ratio, highlighted regions where demand far exceeded supply.

4. **Predictive Modeling and Tuning**: Multiple models were tested, with the XGBoost model selected as the best-performing due to its low error metrics. GridSearchCV was used to optimize hyperparameters, improving the model's predictive accuracy for EV charging demand.

5. **Evaluation and Recommendations**: Based on the analysis, strategic recommendations were made for EnBW to focus infrastructure investments in rural areas with high fast-charging demand.

Key Challenges and Insights

1. **Handling Sparse Data in Rural Areas**: Rural areas often had sparse data, making demand predictions challenging. This was addressed by creating metrics like station density per capita and focusing on normalized comparisons.

2. **Interpreting Google Trends Data as Demand**: Using Google Trends data as a proxy for demand required assumptions about consumer behavior. While it provided valuable insights into interest in charging stations, it may not fully capture actual demand or regional EV adoption rates.

3. **Complex Regional Differences**: The project highlighted significant differences in demand and infrastructure needs across urban and rural regions. Urban areas showed higher demand for regular charging stations, while rural areas exhibited higher demand for fast charging, likely due to longer travel distances.

Lessons Learned

This project provided insights into the complexities of predicting infrastructure needs for EV charging. It underscored the importance of considering both regional and behavioral factors when forecasting demand and highlighted the need for ongoing data collection and model refinement to keep pace with changing demand patterns.

Conclusion and Next Steps

Overall, this analysis offers EnBW actionable insights to strategically plan their EV charging infrastructure, focusing on underserved rural areas and balancing resources in urban regions. Future steps could include:

- **Incorporating More Behavioral Data**: Adding data on EV adoption rates or traffic patterns could improve demand predictions.
- **Dynamic Model Updates**: Regularly updating the model with fresh data will help capture emerging trends and seasonal fluctuations in demand.

Using this approach, EnBW can make data-driven decisions to optimize EV infrastructure, supporting Germany's transition to sustainable transportation.

Concluding Summary and Strategic Recommendations for EnbW

Based on the analysis and predictive insights, the following strategies are proposed to optimize EnBW's infrastructure planning and investment:

1. **Targeted Expansion in Rural Areas**
   The analysis shows low station density in rural areas but high demand for fast charging, highlighting a significant infrastructure gap. By investing in rural fast-charging stations, EnBW can address accessibility issues and promote EV adoption in underserved regions, supporting equitable infrastructure growth.
2. **Urban Area Saturation Management**
   Urban areas exhibit high demand but also have substantial existing infrastructure. Future investments in these areas should focus on high-traffic zones or upgrading current stations to maintain service quality and address potential congestion in high-demand locations.
3. **Data-Driven Forecasting**
   The demand prediction model can serve as a dynamic tool for EnBW, allowing

real-time updates to demand forecasts based on evolving trends. This capability enables proactive infrastructure planning that adapts to changes in consumer behavior, regulations, and technological advancements.

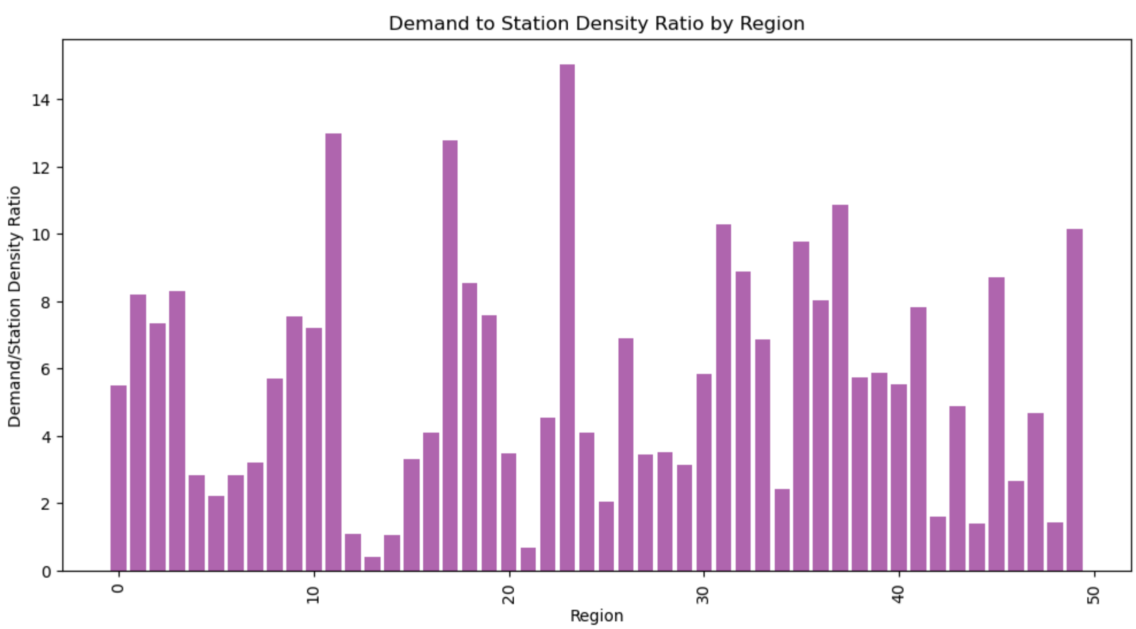
Future Directions

To further enhance EnBW's infrastructure strategy, additional steps are recommended:

1. **Dynamic Trend Analysis**
   By continuously analyzing demand fluctuations over time—especially in response to regulatory changes or incentives—EnBW can better anticipate infrastructure needs and align with policy shifts.

2. **Enhanced Feature Set**
   Incorporating more data, such as traffic patterns, economic indicators, and environmental factors, could improve model accuracy and provide a holistic view of demand drivers.

3. **Deployment and Monitoring**
   Deploying the demand model as a real-time monitoring tool would enable EnBW to receive ongoing insights into demand levels, helping guide timely infrastructure adjustments and supporting continuous, data-informed decision-making.



This residuals plot shows the difference between predicted and actual demand values for EV charging across various regions, with the residuals (errors) plotted on the y-axis and the true demand values on the x-axis.

Summary

Through these strategies, EnBW can make data-informed decisions that optimize EV infrastructure across Germany, addressing specific needs in both urban and rural areas. By bridging identified infrastructure gaps and continuously refining its approach based on real-time data, EnBW can play a pivotal role in supporting Germany's shift towards sustainable transportation.