

report

June 28, 2023

1 Analyzing Electric Vehicle Charging Patterns and Infrastructure in Germany

This data science project explored the availability and usage patterns of electric vehicle (EV) charging facilities in Germany. The project aims to obtain essential insights into EV charging behaviour, determine usage patterns, and examine the existing charging infrastructure. The analysis seeks to provide significant findings and recommendations for improving the EV charging ecosystem by leveraging statistical or machine-learning methodologies.

The project collects data from two primary sources: the Electric Charging Station dataset and the E-Ladesäulenregister dataset. These datasets contain information about charging station operators, types of charging facilities, power capacity, location details, and operational dates.

The question that interests us is: - What are the usage patterns of electric vehicle charging stations in Germany? - How does the number of charging points and power capacity vary across different types of charging stations? - Which areas in Germany have the highest concentration of charging stations, and which areas have limited coverage? - Are there specific types of charging stations that are more prevalent in certain regions or urban areas?

1.1 Install dependencies

Initially, install all required dependencies.

```
[ ]: %pip install pandas
      %pip install plotly
      %pip install 'SQLAlchemy==1.4.46'
      %pip install nbformat
      %pip install pysqlite
      %pip install seaborn
```

1.2 Data Sources

Datasource 1: Electric Charging Station

- Metadata URL: [Link to Metadata](#)
- Data URL: [Link to Data](#)
- Data Type: CSV

This data source provides information on electric charging stations in Germany. The metadata URL contains additional details about the dataset, including its source and description. The data

can be accessed using the provided data URL, which leads to a CSV file. The dataset contains relevant information about the charging stations, such as the operator, type of charging facility, number of charging points, power capacity, plug types, location details, and operational dates.

Datasource 2: E-Ladesäulenregister

- Metadata URL: [Link to Metadata](#)
- Data URL: [Link to Data](#)
- Data Type: xlsx

This data source refers to the E-Ladesäulenregister, which provides information on electric charging stations in Germany. The metadata URL contains additional details about the dataset. The data can be accessed using the provided data URL, which leads to an Excel file (XLSX) containing the relevant information. The dataset includes details about the charging station operator, address, postal code, city, federal state, latitude, longitude, power capacity, type of charging facility, number of charging points, and installation dates.

1.3 Load data

Create a pandas dataframe using the local sqlite file.

```
[ ]: import pandas as pd
import sqlite3 as db

import matplotlib.pyplot as plt
import seaborn as sns

conn = db.connect('../data/database/data.db')

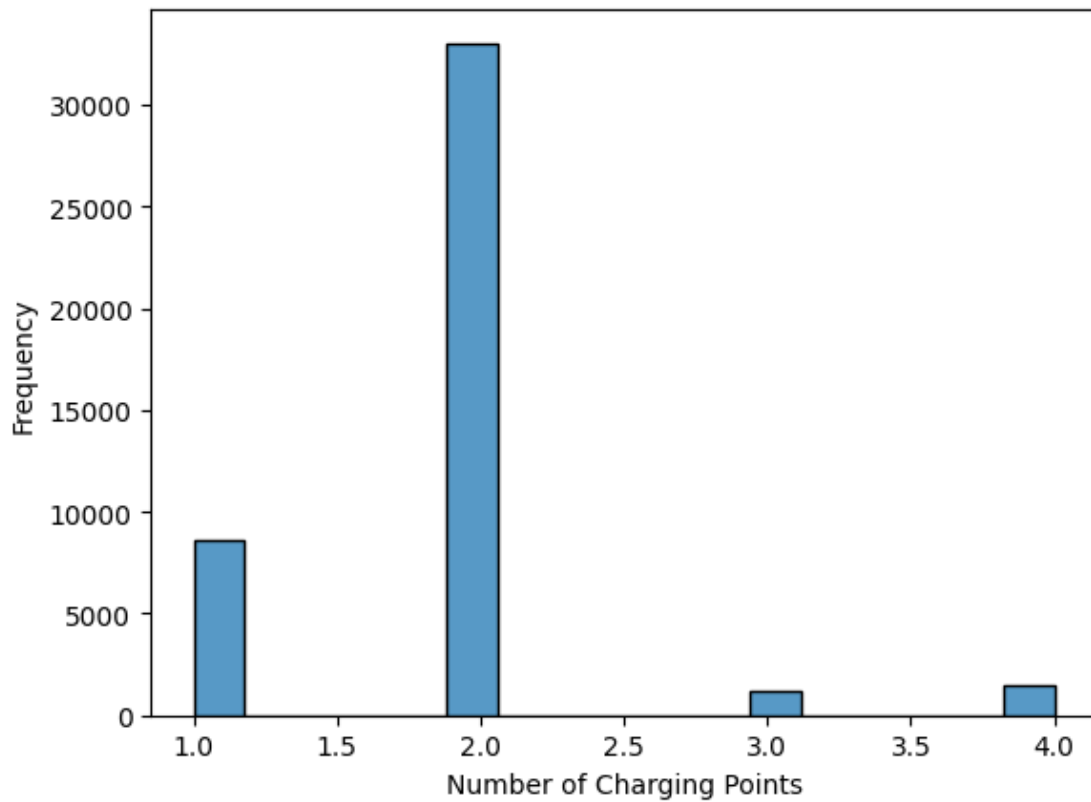
# Setting queries for loading data
query_1 = f"SELECT * FROM table1"
query_2 = f"SELECT * FROM table2"

# load data from the database into pandas dataframe
df_1 = pd.read_sql_query(query_1, conn)
df_2 = pd.read_sql_query(query_2, conn)

[ ]: # Merging the both data
merged_df = pd.concat([df_1, df_2], ignore_index=True)
# Step 2: Data cleaning and preprocessing
merged_df = merged_df.dropna(subset=['Postleitzahl'])
# Convert to datetime
merged_df['Inbetriebnahmedatum'] = pd.
    .to_datetime(merged_df['Inbetriebnahmedatum'])

[ ]: # Step 3: Exploratory Data Analysis
# Anzahl Ladepunkte: Number of charging points
sns.histplot(merged_df['Anzahl Ladepunkte'])
plt.xlabel('Number of Charging Points')
```

```
plt.ylabel('Frequency')
plt.show()
```

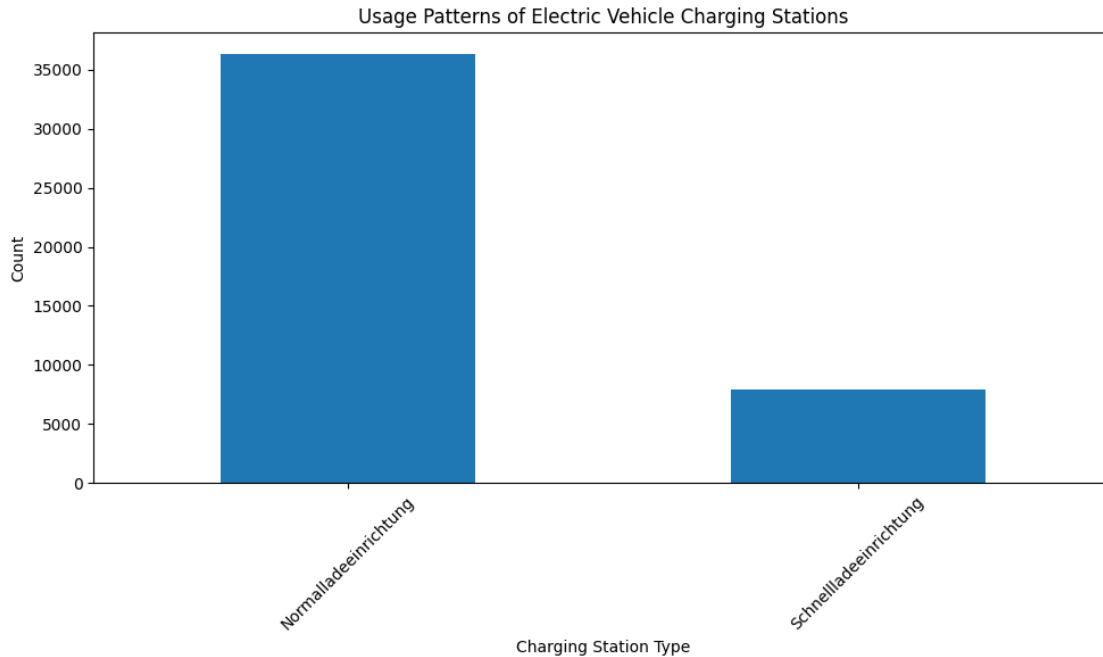


1.3.1 What are the usage patterns of electric vehicle charging stations in Germany?

```
[ ]: # Art der Ladeeinrichtung: Type of charging device
usage_patterns = merged_df['Art der Ladeeinrichtung'].value_counts()

# Normalladeeinrichtung: Normal charging device
# Schnellladeeinrichtung: Fast charger

plt.figure(figsize=(10, 6))
usage_patterns.plot(kind='bar')
plt.xlabel('Charging Station Type')
plt.ylabel('Count')
plt.title('Usage Patterns of Electric Vehicle Charging Stations')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

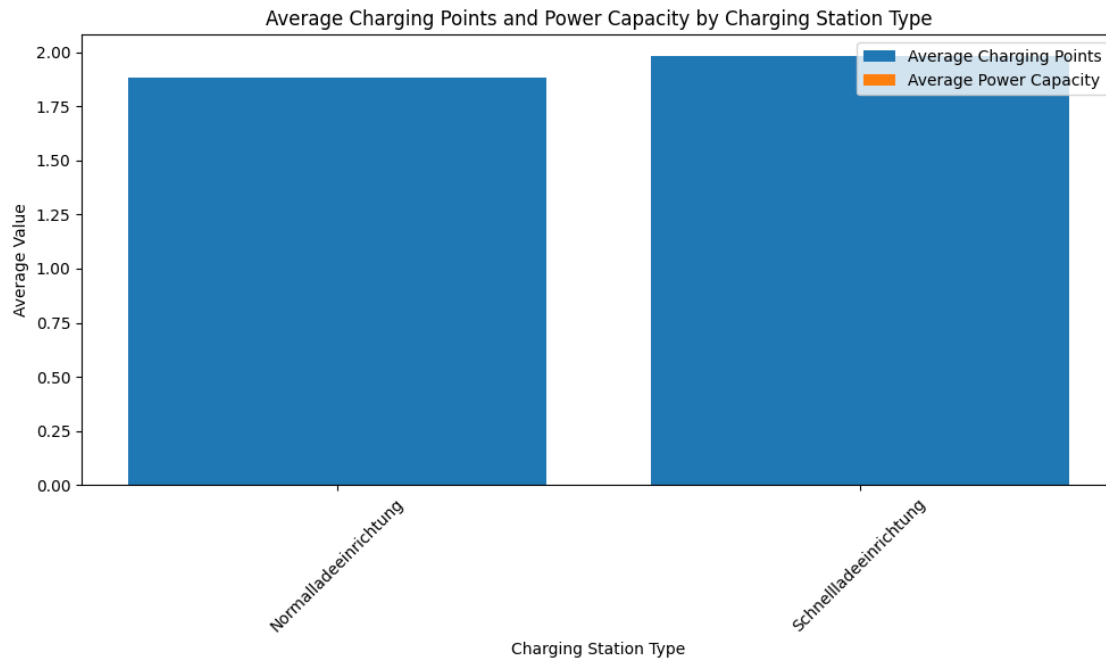


1.3.2 How does the number of charging points and power capacity vary across different types of charging stations?

```
[ ]: # anschlussleistung: connected load
charging_points_by_type = merged_df.groupby('Art der Ladeeinrichtung')['Anzahl_Ladepunkte'].mean()
power_capacity_by_type = merged_df.groupby('Art der Ladeeinrichtung')['anschlussleistung'].mean()

# Plotting
plt.figure(figsize=(10, 6))
plt.bar(charging_points_by_type.index, charging_points_by_type.values, label='Average Charging Points')
plt.bar(power_capacity_by_type.index, power_capacity_by_type.values, label='Average Power Capacity')

plt.xlabel('Charging Station Type')
plt.ylabel('Average Value')
plt.title('Average Charging Points and Power Capacity by Charging Station Type')
plt.xticks(rotation=45)
plt.legend()
plt.tight_layout()
plt.show()
```

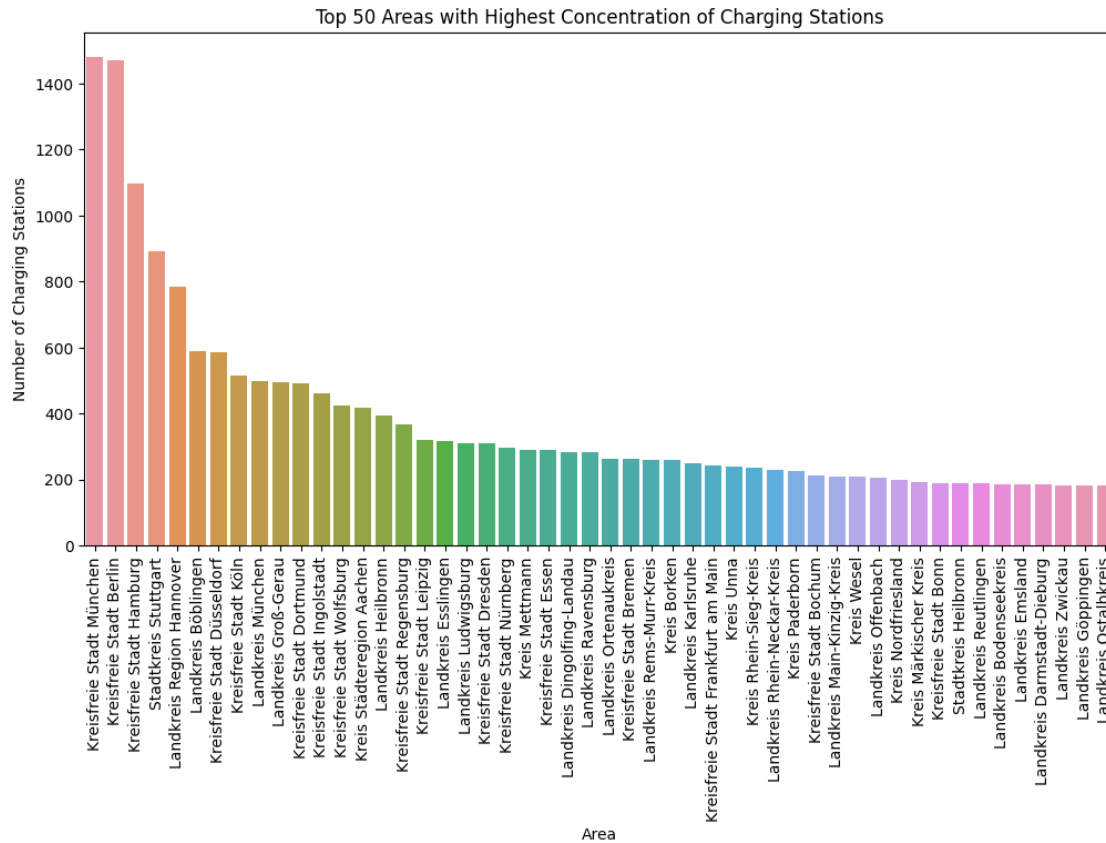


1.3.3 Which areas in Germany have the highest concentration of charging stations, and which areas have limited coverage?

```
[ ]: import seaborn as sns
import matplotlib.pyplot as plt

# finding data
charging_station_counts = merged_df['Kreis/kreisfreie Stadt'].value_counts()
# specify the number of top areas
top_n = 50
top_n_areas = charging_station_counts.head(top_n)

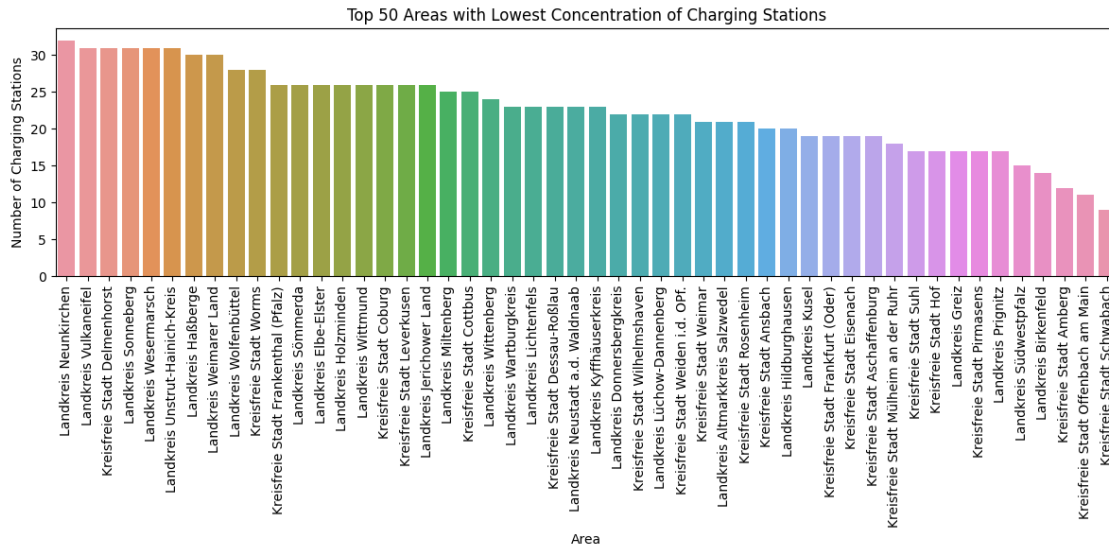
# create bar plot
plt.figure(figsize=(12, 6))
sns.barplot(x=top_n_areas.index, y=top_n_areas.values)
plt.title(f"Top {top_n} Areas with Highest Concentration of Charging Stations")
plt.xlabel("Area")
plt.ylabel("Number of Charging Stations")
plt.xticks(rotation=90)
plt.show()
```



```
[ ]: import seaborn as sns
import matplotlib.pyplot as plt

# Finding data
charging_station_counts = merged_df['Kreis/kreisfreie Stadt'].value_counts()
# Specify the number of bottom areas
bottom_n = 50
bottom_n_areas = charging_station_counts.tail(bottom_n)

# Create bar plot using seaborn to plot
plt.figure(figsize=(12, 6))
sns.barplot(x=bottom_n_areas.index, y=bottom_n_areas.values)
plt.title(f"Top {bottom_n} Areas with Lowest Concentration of Charging Stations")
plt.xlabel('Area')
plt.ylabel('Number of Charging Stations')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```

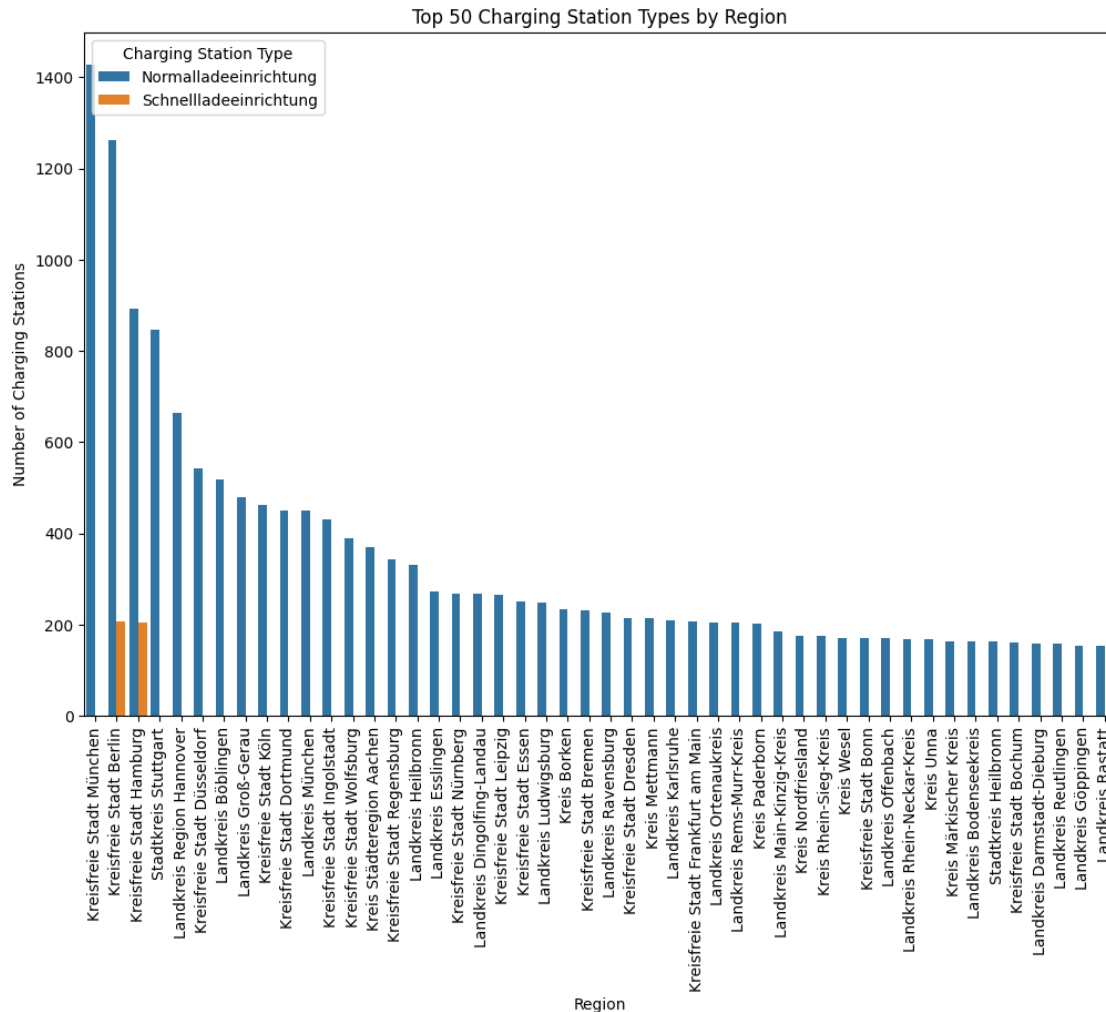


1.3.4 Are there specific types of charging stations that are more prevalent in certain regions or urban areas?

```
[ ]: # Charging station types by region
charging_station_types_by_region = merged_df.groupby('Kreis/kreisfreie_
↳Stadt')['Art der Ladeeinrichtung'].value_counts().nlargest(50)

# Convert the data to a DataFrame
data = charging_station_types_by_region.reset_index(name='Count')

# Plotting the data using seaborn
plt.figure(figsize=(12, 8))
sns.barplot(x='Kreis/kreisfreie Stadt', y='Count', hue='Art der_
↳Ladeeinrichtung', data=data)
plt.title('Top 50 Charging Station Types by Region')
plt.xlabel('Region')
plt.ylabel('Number of Charging Stations')
plt.xticks(rotation=90)
plt.legend(title='Charging Station Type')
plt.show()
```



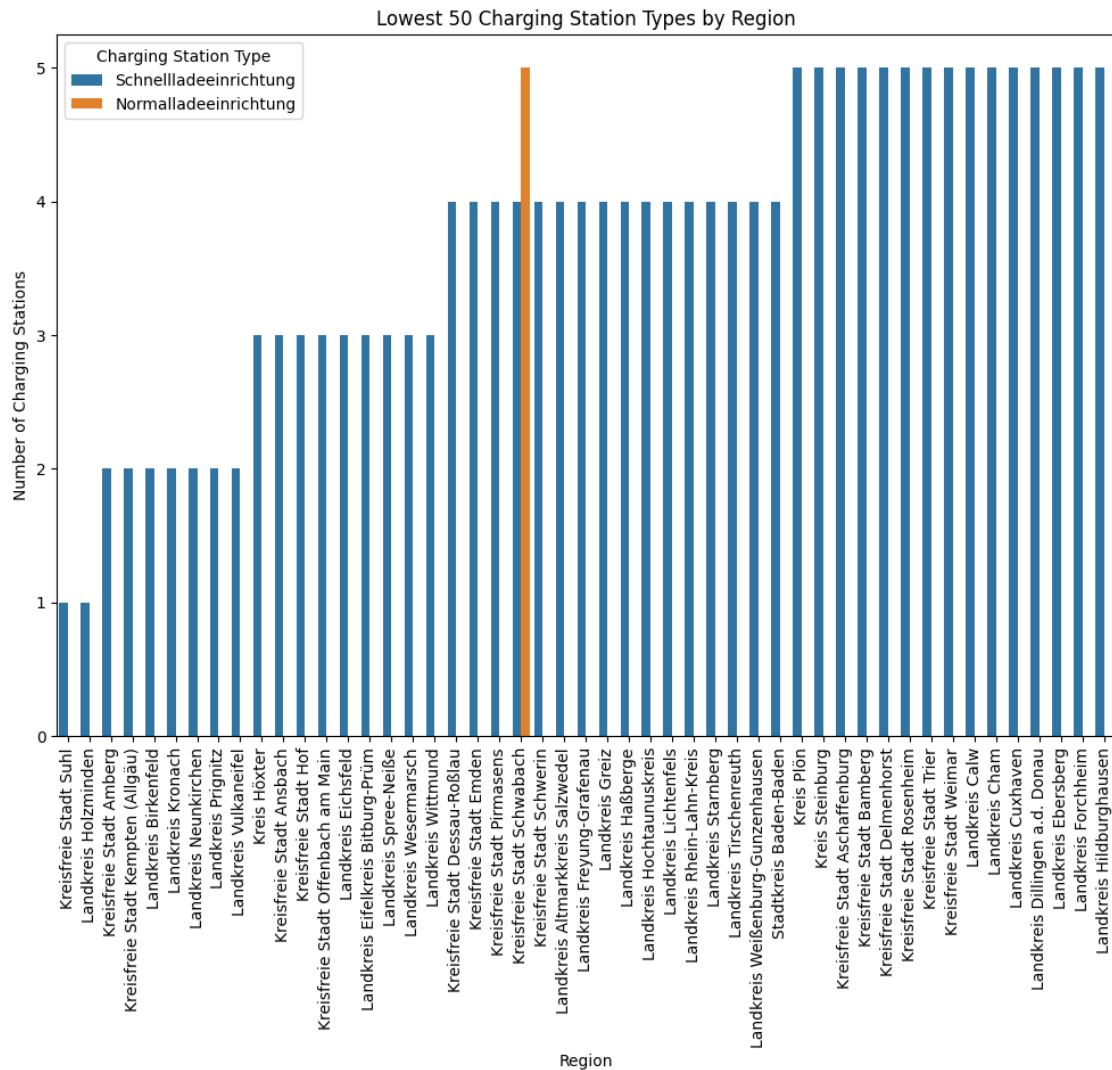
```
[ ]: import seaborn as sns
import matplotlib.pyplot as plt

charging_station_types_by_region = merged_df.groupby('Kreis/kreisfreie_
↳Stadt')['Art der Ladeeinrichtung'].value_counts().nsmallest(50)
data = charging_station_types_by_region.reset_index(name='Count')

# Plotting the data using seaborn
plt.figure(figsize=(12, 8))
sns.barplot(x='Kreis/kreisfreie Stadt', y='Count', hue='Art der_
↳Ladeeinrichtung', data=data)
plt.title('Lowest 50 Charging Station Types by Region')
plt.xlabel('Region')
plt.ylabel('Number of Charging Stations')
plt.xticks(rotation=90)
plt.legend(title='Charging Station Type')
```



```
plt.show()
```



1.4 Limitations and Conclusion

1.4.1 Limitations

- The analysis is based on the available mentioned data sources, which may have limitations such as incomplete or outdated information. When interpreting the results, it is essential to consider the data quality and potential biases.
- It is worth noting that the dataset I used represents registered charging stations, and there may be unregistered or privately owned charging points that need to incorporate into the analysis. This could result in a potential underestimation of the overall charging infrastructure.

- The absence of population density data in the provided dataset limits exploring the relationship between charging station distribution and population density. This aspect could provide valuable insights into higher-demand areas and help prioritize charging infrastructure expansion.
- The analysis focuses on the current state of charging stations and usage patterns and does not provide future projections or predictions. Considering the dynamic nature of the electric vehicle market and evolving infrastructure needs is essential when planning for future developments.

1.4.2 Conclusion

In conclusion, the analysis enlightens the usage patterns and distribution of electric vehicle charging stations in Germany. The findings are useful for policymakers, urban planners, and other stakeholders interested in promoting sustainable transportation. Furthermore, the investigation emphasizes the need for further expansion of the charging network, particularly in areas with limited coverage. Addressing this gap will be crucial to support the widespread adoption of electric vehicles and ensuring easy access to charging infrastructure. However, based on the available data and analysis, it is necessary to recognize the project's findings and recommendations.. To keep up with the evolving electric vehicle landscape and identify future opportunities for expansion and improvement, charging station data will need to be monitored and analyzed on an ongoing basis.