Analyzing EV Charging Infrastructure: A Data-Driven Approach

Report File

Group 2

Rohan Verma

Priyansh Nileshbhai Vagadiya

617 459 9214

857 339 8070

verma.rohan@northeastern.edu

vagadiya.p@northeastern.edu

Percentage of Effort Contributed by Student1 : 50%

Percentage of Effort Contributed by Student2 : 50%

Signature of Student 1 : Rohan Verma

Signature of Student 2 : Priyansh Nileshbhai Vagadiya

Executive Summary

The primary objective of this study was to design and implement a comprehensive relational database system tailored to the management and optimization of electric vehicle (EV) charging stations. This system addresses critical challenges, including station compatibility with various EV models, compliance with regulations, accessibility, operational efficiency, and environmental impact. By centralizing and structuring diverse data points, the database enhances decision-making and supports the expansion of sustainable EV infrastructure.

The dataset used for this project is sourced from the Alternative Fueling Station Locations, provided by the U.S. federal government. This dataset focuses on EV charging stations across the United States, offering detailed information about their locations, capabilities, and availability. The data was cleaned and prepared using Python in the Jupyter Notebook. Data preprocessing included handling missing values, standardizing geographical and station-specific details, and filtering out incomplete records. These steps ensure data quality and reliability, enabling meaningful analyses such as pricing effects on user behavior, energy consumption, CO2 reduction from EV usage, and optimal station locations for infrastructure development.

The database design involved the creation of Enhanced Entity-Relationship (EER) and Unified Modeling Language (UML) diagrams to conceptualize the system, which was subsequently implemented in MySQL. Selected relationships were also prototyped in MongoDB to demonstrate adaptability in a NoSQL environment.

Future improvements include implementing robust data governance measures and expanding the database to integrate nationwide EV infrastructure. By leveraging the capabilities of this relational database, the project supports the development of a cleaner, more efficient EV charging network, promoting sustainable energy usage and reducing carbon emissions across the United States.

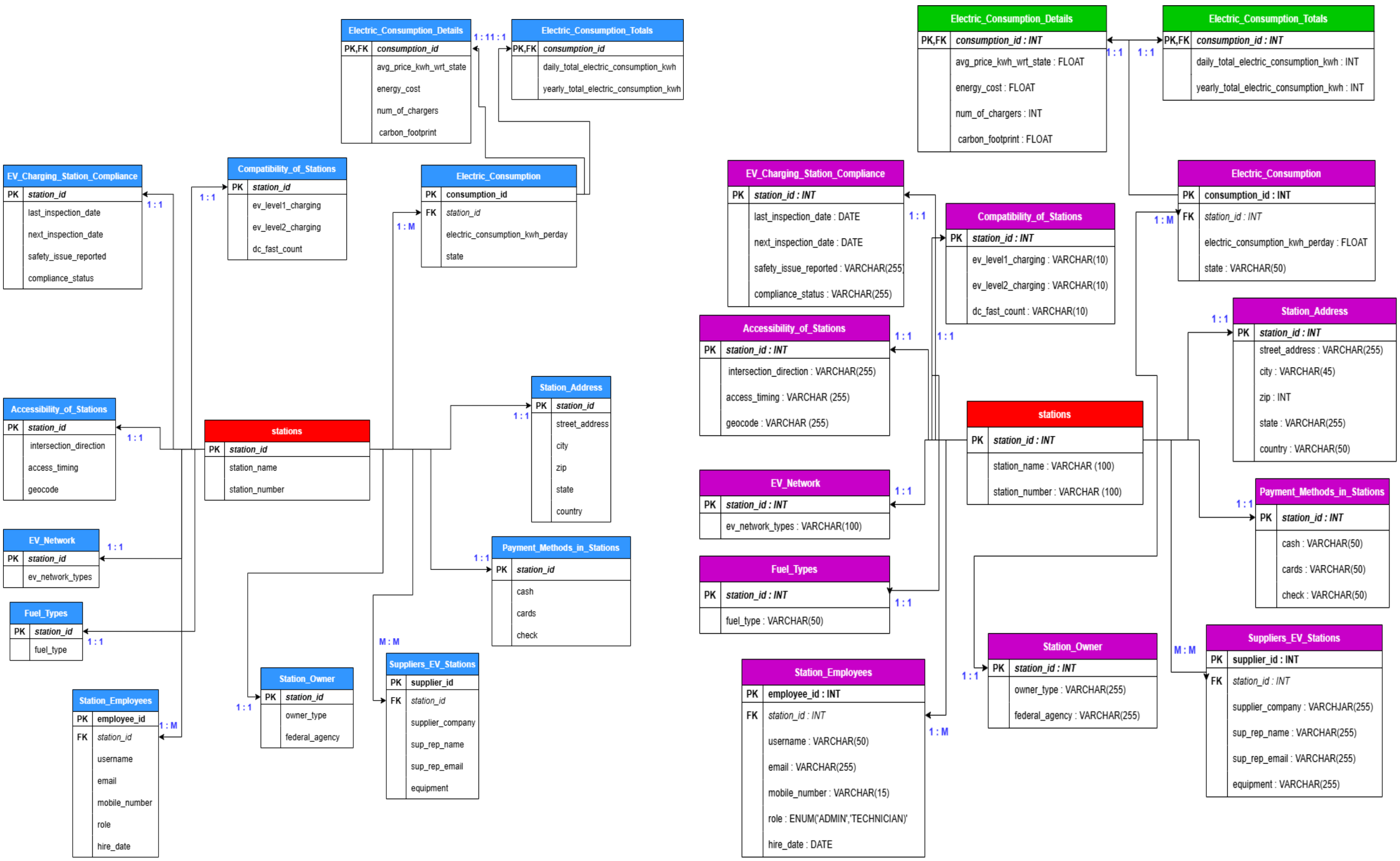
Introduction

The rise in electric vehicle (EV) adoption worldwide has significantly increased the demand for EV charging infrastructure. In the United States, particularly in states like Washington, where environmental sustainability is a priority, the establishment of EV charging stations is a critical step toward achieving clean energy goals. However, managing and optimizing this infrastructure requires a comprehensive database system to address key challenges such as accessibility, compatibility, compliance, and operational efficiency.

This project aims to develop a relational database system to address these challenges by consolidating all critical aspects of EV charging station management. The database will include schemas for stations, station owners, station employees, station addresses, suppliers of EV stations, accessibility of stations, compatibility of stations, electric consumption, EV charging station compliance, EV networks, fuel types, and payment methods at stations. Each schema is designed to eliminate data redundancy, streamline information management, and facilitate data-driven decision-making.

With this centralized database, station operators can efficiently manage their facilities, monitor energy consumption, and ensure compatibility with evolving EV standards. Suppliers can track compliance and support station operations, while users benefit from better accessibility and a seamless charging experience. By integrating real-time data from multiple entities, the database will also enable predictive analytics, such as forecasting energy demand and optimizing station deployment based on usage patterns.

Conceptual Data Modeling : Primary Key - Bold , Foreign Key - Italicized
EER and UML Diagram



Mapping Conceptual Model to Relational Model
Tables and Attributes

1. Stations (***station_id***, station_name, station_number)
2. Accessibility_of_Stations (***station_id***, intersection_direction, access_timing, geocode)
3. Compatibility_of_Stations (***station_id***, ev_level1_charging, ev_level2_charging, dc_fast_count)
4. Electric_Consumption (***id***, state, electric_consumption_kwh_perday, daily_total_electric_consumption_kwh, yearly_total_electric_consumption_kwh, avg_price_kwh_wrt_state, energy_cost, carbon_footprint, num_of_chargers)
5. EV_Charging_Station_Compliance (***station_id***, last_inspection_date, next_inspection_date, safety_issue_reported, compliance_status)
6. EV_Network (***station_id***, ev_network_types)

7. Fuel_Types (***station_id***, fuel_type)
8. Payment_Methods_in_Stations (***station_id***, cash, cards, checks)
9. Station_Address (***station_id***, street_address, city, zip, state, country)
10. Station_Employees (**employee_id**, *station_id*, user_name, email, mobile_number, role, hire_date)
11. Station_Owner (**station_id**, owner_type, federal_agency)
12. Suppliers_EV_Stations (**supplier_id**, *station_id*, supplier_company, sup_rep_name, sup_rep_email, equipment)

We normalized the **Electric_Consumption** entity into three related tables to eliminate redundancy and ensure each table focuses on a single theme: **Electric_Consumption** (daily consumption details), **Electric_Consumption_Totals** (aggregated totals), and **Electric_Consumption_Details** (pricing and environmental factors). All other tables in the schema are already in **4NF**, and this completes the normalization of the entire scheme.

	TABLE NAME	PRIMARY KEY	ATTRIBUTES
1.	Stations	station_id	station_name, station_number
2.	Accessibility_of_Stations	station_id	intersection_direction, access_timing, geocode
3.	Compatibility_of_Stations	station_id	ev_level1_charging, ev_level2_charging, dc_fast_count
4.	Electric_Consumption	consumption_id	station_id (FK), state, electric_consumption_kwh_perday
5.	Electric_Consumption_Totals	consumption_id (FK)	daily_total_electric_consumption_kwh, yearly_total_electric_consumption_kwh
6.	Electric_Consumption_Details	consumption_id (FK)	avg_price_kwh_wrt_state, energy_cost, carbon_footprint, num_of_chargers
7.	EV_Charging_Station_Compliance	station_id	last_inspection_date, next_inspection_date, safety_issue_reported, compliance_status
8.	EV_Network	station_id	ev_network_types
9.	Fuel_Types	station_id	fuel_type
10	Payment_Methods_in_Stations	station_id	cash, cards, checks
11	Station_Address	station_id	street_address, city, zip, state, country
12	Station_Employees	employee_id	station_id (FK), user_name, email, mobile_number, role, hire_date
13	Station_Owner	station_id	owner_type, federal_agency
14	Suppliers_EV_Stations	station_id	station_id (FK), supplier_company, sup_rep_name, sup_rep_email, equipment

Implementation of Relation Model via MySQL

- Q1 : query retrieves name of the station and there corresponding id from the Stations table :

SELECT station_name, station_id FROM Stations ;

```
select station_name, station_id
from stations
```

Result Grid | Filter Rows: | Edit:

station_name	station_id
Con Edison - E 16th Street Service Center	125
PG&E Santa Cruz Service Center	798
PG&E Richmond Service Center	826
Holiday Stationstore #73	16241
U-Haul	19360
Schindler Gas	20182
U-Haul	20551
AmeriGas	21945

- Q2 : This aggregate query calculates the total electric consumption per state from the Electric_Consumption table.

```
SELECT state, SUM(electric_consumption_kwh_perday) AS total_consumption
FROM Electric_Consumption
GROUP BY state;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

state	total_consumption
NY	316.8
CA	996
MN	158.4
MD	0
TX	0
MA	0
WI	158.4
NE	0
VA	0
PA	0
MI	158.4
OK	0
WA	520.8

SELECT state,
SUM(electric_consumption_kwh_perday)
AS total_consumption
FROM electric_consumption GROUP BY state

- Q3 : query retrieves a list of stations along with their corresponding address details by performing an inner join between the Stations and Station_Address tables.

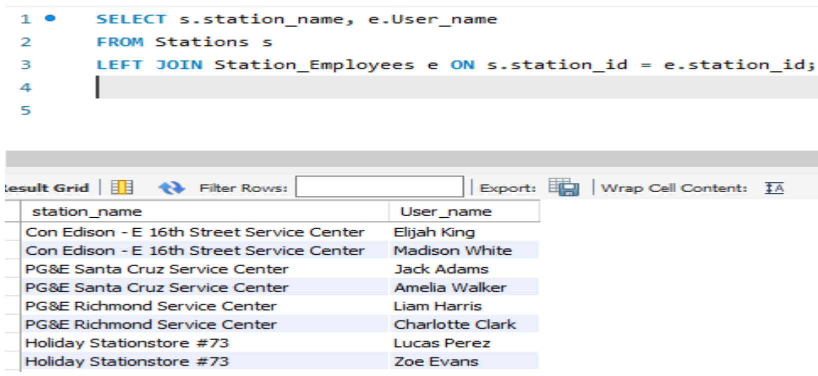
```
SELECT s.station_name, a.street_address, a.city
FROM Stations s
INNER JOIN Station_Address a ON s.station_id = a.station_id;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

station_name	street_address	city
Con Edison - E 16th Street Service Center	700 E 16th St	New York
PG&E Santa Cruz Service Center	615 7th Ave	Santa Cruz
PG&E Richmond Service Center	1100 S 27th St	Richmond
Holiday Stationstore #73	8401 Lyndale Ave S	Bloomington
U-Haul	4111 W Northern Pkwy	Baltimore
Schindler Gas	6072 Highway 90A	Eagle Lake
U-Haul	626 S Anaheim Blvd	Anaheim
AmeriGas	1162 Park St	Palmer
Southern California Edison - Valencia Service Ce...	25625 W Rye Canyon Rd	Valencia
Holiday Stationstore #306	4167 Ball Rd NE	Blaine

SELECT s.station_name, a.street_address, a.city
FROM Stations s INNER JOIN
Station_Address a ON s.station_id = a.station_id;

- Q4 : This query retrieves a list of all stations along with their employee details, including stations that do not have any employees (LEFT JOIN).



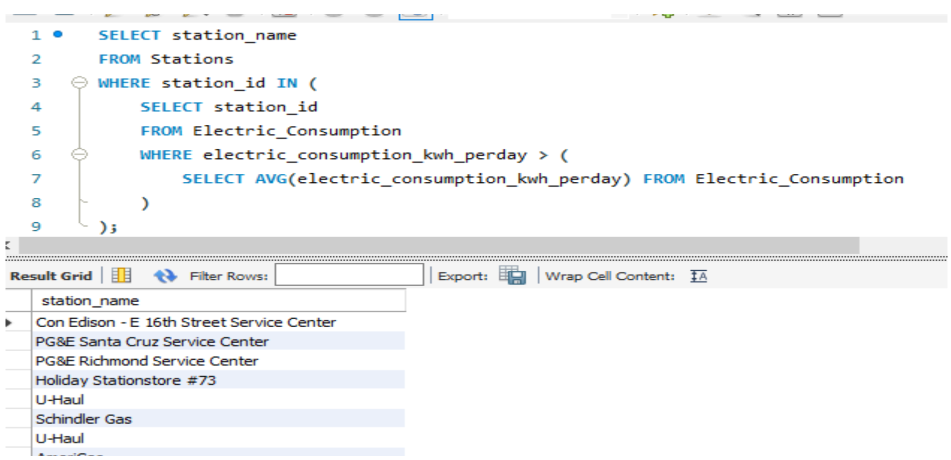
```
SELECT s.station_name, e.User_name
FROM Stations s
LEFT JOIN Station_Employees e ON s.station_id = e.station_id;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

station_name	User_name
Con Edison - E 16th Street Service Center	Elijah King
Con Edison - E 16th Street Service Center	Madison White
PG&E Santa Cruz Service Center	Jack Adams
PG&E Santa Cruz Service Center	Amelia Walker
PG&E Richmond Service Center	Liam Harris
PG&E Richmond Service Center	Charlotte Clark
Holiday Stationstore #73	Lucas Perez
Holiday Stationstore #73	Zoe Evans

SELECT s.station_name, e.User_name
FROM Stations s
LEFT JOIN Station_Employees e
ON s.station_id = e.station_id;

- Q5 : This nested query retrieves stations that have a higher electric consumption than the average consumption for all stations.



```
SELECT station_name
FROM Stations
WHERE station_id IN (
    SELECT station_id
    FROM Electric_Consumption
    WHERE electric_consumption_kwh_perday > (
        SELECT AVG(electric_consumption_kwh_perday) FROM Electric_Consumption
    )
);
```

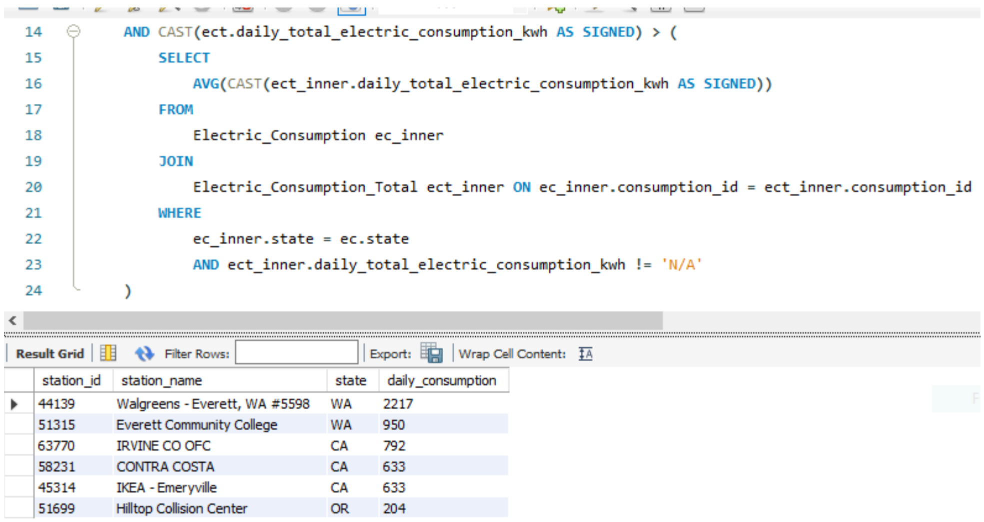
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

station_name
Con Edison - E 16th Street Service Center
PG&E Santa Cruz Service Center
PG&E Richmond Service Center
Holiday Stationstore #73
U-Haul
Schindler Gas
U-Haul

SELECT station_name FROM Stations
WHERE station_id IN
(SELECT station_id
FROM Electric_Consumption
WHERE electric_consumption_kwh_perday >
(SELECT AVG(electric_consumption_kwh_perday)
FROM Electric_Consumption)

);

- Q6 : This correlated subquery identifies stations where the daily electric consumption is higher than the average daily consumption of all stations in the same state.



```
AND CAST(ect.daily_total_electric_consumption_kwh AS SIGNED) > (
    SELECT
        AVG(CAST(ect_inner.daily_total_electric_consumption_kwh AS SIGNED))
    FROM
        Electric_Consumption ec_inner
    JOIN
        Electric_Consumption_Total ect_inner ON ec_inner.consumption_id = ect_inner.consumption_id
    WHERE
        ec_inner.state = ec.state
        AND ect_inner.daily_total_electric_consumption_kwh != 'N/A'
)
```

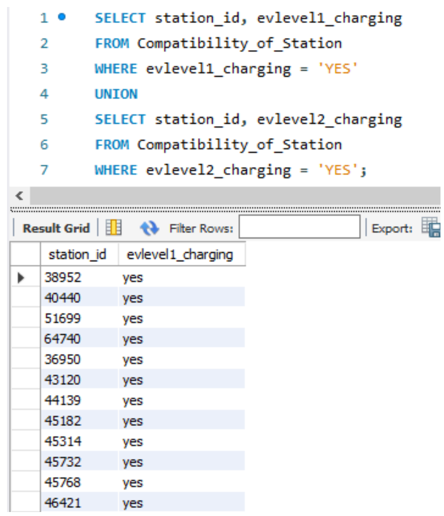
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

station_id	station_name	state	daily_consumption
44139	Walgreens - Everett, WA #5598	WA	2217
51315	Everett Community College	WA	950
63770	IRVINE CO OFC	CA	792
58231	CONTRA COSTA	CA	633
45314	IKEA - Emeryville	CA	633
51699	Hilltop Collision Center	OR	204

```
SELECT
    ec.station_id,
    s.station_name,
    ec.state,
    ect.daily_total_electric_consumption_kwh
FROM
    Electric_Consumption ec
JOIN
    Electric_Consumption_Total ect ON ec.consumption_id = ect.consumption_id
JOIN
    Stations s ON ec.station_id = s.station_id
WHERE
    ect.daily_total_electric_consumption_kwh > (
        SELECT
            AVG(ect_inner.daily_total_electric_consumption_kwh)
        FROM
            Electric_Consumption ec_inner
        JOIN
            Electric_Consumption_Total ect_inner ON ec_inner.consumption_id = ect_inner.consumption_id
        WHERE
            ec_inner.state = ec.state
    )
ORDER BY
    ect.daily_total_electric_consumption_kwh DESC
```

- Q7 : This query retrieves a list of stations with either EV Level 1 or EV Level 2 charging compatibility (using UNION).

```
SELECT station_id, evlevel1_charging
FROM Compatibility_of_Station
WHERE evlevel1_charging = 'YES'
 UNION
SELECT station_id, evlevel2_charging
FROM Compatibility_of_Station
WHERE evlevel2_charging = 'YES';
```



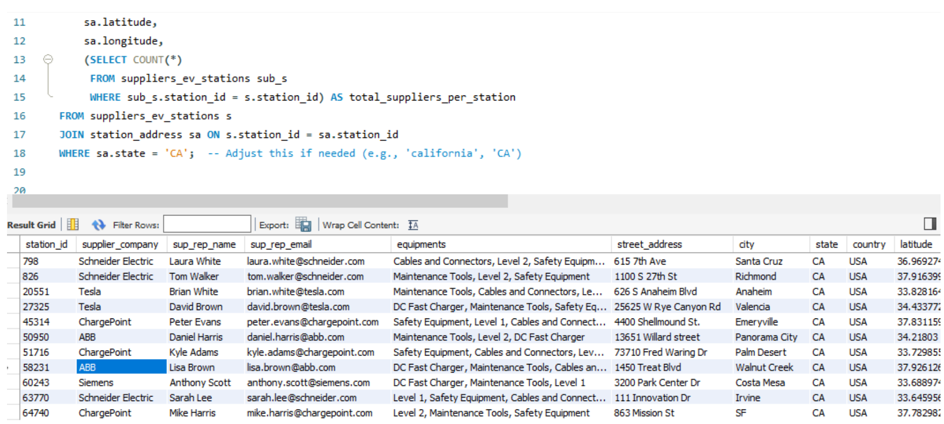
```
SELECT station_id, evlevel1_charging
FROM Compatibility_of_Station
WHERE evlevel1_charging = 'YES'
UNION
SELECT station_id, evlevel2_charging
FROM Compatibility_of_Station
WHERE evlevel2_charging = 'YES';
```

Result Grid | Filter Rows: | Export:

station_id	evlevel1_charging
38952	yes
40440	yes
51699	yes
64740	yes
36950	yes
43120	yes
44139	yes
45182	yes
45314	yes
45732	yes
45768	yes
46421	yes

- Q8 : In this query, we retrieve detailed information about electric vehicle (EV) charging stations in California, including the supplier information and station location using Subqueries in Select and From.

```
SELECT S.station_id, s.supplier_company,
s.sup_rep_name, s.sup_rep_email, S.equipments,
sa.street_address, sa.city,
sa.state, sa.country,
sa.latitude, sa.longitude,
(SELECT COUNT(*)
FROM suppliers_ev_stations
sub_s
WHERE sub_s.station_id =
s.station_id)
AS total_suppliers_per_station
FROM suppliers_ev_stations s
JOIN station_address sa
ON s.station_id = sa.station_id
WHERE sa.state = 'CA';
```



```
        sa.latitude,
        sa.longitude,
        (SELECT COUNT(*)
         FROM suppliers_ev_stations sub_s
         WHERE sub_s.station_id = s.station_id) AS total_suppliers_per_station
    FROM suppliers_ev_stations s
    JOIN station_address sa ON s.station_id = sa.station_id
    WHERE sa.state = 'CA';  -- Adjust this if needed (e.g., 'california', 'CA')
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

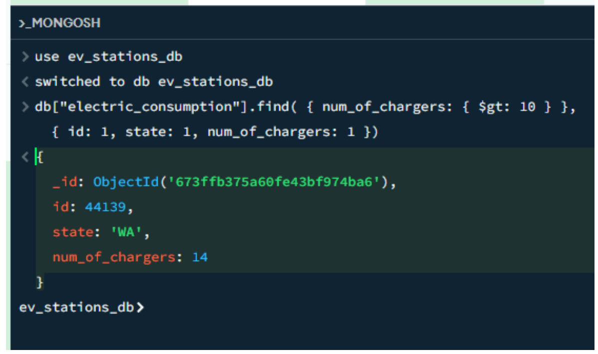
station_id	supplier_company	sup_rep_name	sup_rep_email	equipments	street_address	city	state	country	latitude
798	Schneider Electric	Laura White	laura.white@schneider.com	Cables and Connectors, Level 2, Safety Equipm...	615 7th Ave	Santa Cruz	CA	USA	36.96927
826	Schneider Electric	Tom Walker	tom.walker@schneider.com	Maintenance Tools, Level 2, Safety Equipment	1100 S 27th St	Richmond	CA	USA	37.91639
20551	Tesla	Brian White	brian.white@tesla.com	Maintenance Tools, Cables and Connectors, Le...	626 S Anaheim Blvd	Anaheim	CA	USA	33.82816
27325	Tesla	David Brown	david.brown@tesla.com	DC Fast Charger, Maintenance Tools, Safety Eq...	25625 W Rye Canyon Rd	Valencia	CA	USA	34.43377
45314	ChargePoint	Peter Evans	peter.evans@chargepoint.com	Safety Equipment, Level 1, Cables and Connect...	4400 Shellmound St.	Emeryville	CA	USA	37.83115
50950	ABB	Daniel Harris	daniel.harris@abb.com	Maintenance Tools, Level 2, DC Fast Charger	13651 Willard street	Panorama City	CA	USA	34.21803
51716	ChargePoint	Kyle Adams	kyle.adams@chargepoint.com	Safety Equipment, Cables and Connectors, Lev...	73710 Fred Waring Dr	Palm Desert	CA	USA	33.72985
58231	ABB	Lisa Brown	lisa.brown@abb.com	DC Fast Charger, Maintenance Tools, Cables an...	1450 Treat Blvd	Walnut Creek	CA	USA	37.92612
60243	Siemens	Anthony Scott	anthony.scott@siemens.com	DC Fast Charger, Maintenance Tools, Level 1	3200 Park Center Dr	Costa Mesa	CA	USA	33.68897
63770	Schneider Electric	Sarah Lee	sarah.lee@schneider.com	Level 1, Safety Equipment, Cables and Connect...	111 Innovation Dr	Irvine	CA	USA	33.64595
64740	ChargePoint	Mike Harris	mike.harris@chargepoint.com	Level 2, Maintenance Tools, Safety Equipment	863 Mission St	SF	CA	USA	37.78298

Implementation of Relation Model via NoSQL

Twelve collections from the **evstation schema** were migrated to MongoDB for efficient EV station data analysis.

- Q1 : db.electric_consumption.find(
  { num_of_chargers: { $gt: 10 } },
  { id: 1, state: 1, num_of_chargers: 1 });

  The query focuses on retrieving the charging stations that have more than 10 chargers, the query returns key information such as the station ID, state, and the number of chargers, helping to focus on stations with a significant capacity.



```
>_MONGOSH
> use ev_stations_db
< switched to db ev_stations_db
> db["electric_consumption"].find( { num_of_chargers: { $gt: 10 } },
    { id: 1, state: 1, num_of_chargers: 1 })
< {
    _id: ObjectId('673ffb375a60fe43bf974ba6'),
    id: 44139,
    state: 'WA',
    num_of_chargers: 14
  }
ev_stations_db>
```

- Q2 : db["ev_chargingstation_compliance"].find(
  { compliance_status: "Pending Investigation",
  safety_issue_reported: "Fire Hazard" },
  { station_id: 1, compliance_status: 1,
  last_inspection_date: 1, next_inspection_date: 1,
  safety_issue_reported: 1, _id: 0 })

  This query retrieves details of all electric charging stations that have the compliance status "Pending Investigation" and have reported the safety issue "Fire Hazard".

  This information is essential for identifying stations that require immediate attention due to fire hazards while ensuring regulatory compliance and safety standards are maintained.

```
> db["ev_chargingstation_compliance"].find({
      compliance_status: "Pending Investigation",
      safety_issue_reported: "Fire Hazard"
    },
    {
      station_id: 1,
      compliance_status: 1,
      last_inspection_date: 1,
      next_inspection_date: 1,
      safety_issue_reported: 1,
      _id: 0
    })
< {
    station_id: 53316,
    compliance_status: 'Pending Investigation',
    last_inspection_date: '1/8/2015',
    next_inspection_date: '1/8/2016',
    safety_issue_reported: 'Fire Hazard'
  }
ev_stations_db>
```

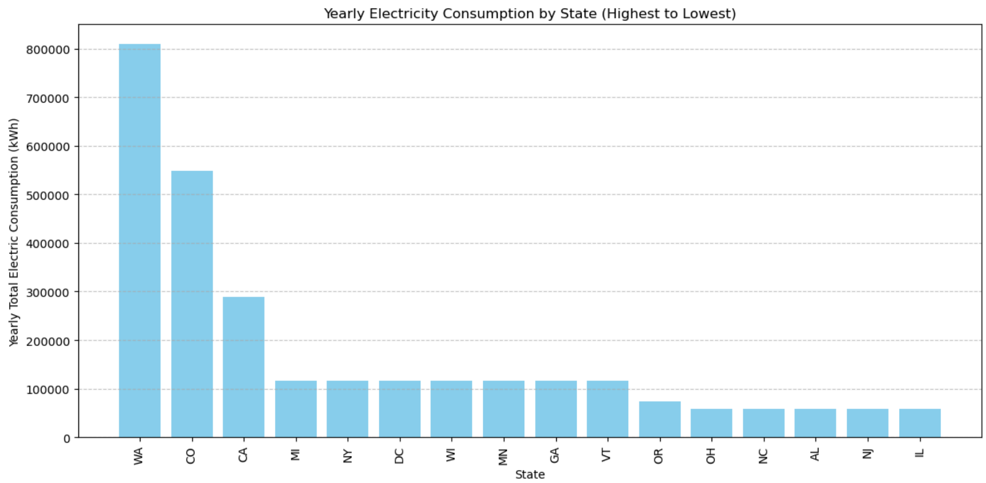
- Q3 : db["station_employees"].aggregate([
  { $group: {_id: "$station_id", // Group by station_id
  employee_count: { $sum: 1 } // Count the number of employees per station} },{
  $sort: { employee_count: -1 } // Optional: Sort by employee count in descending order
  }])

  This query is to aggregate the data, to count the number of employees per station or group by the role of employees.

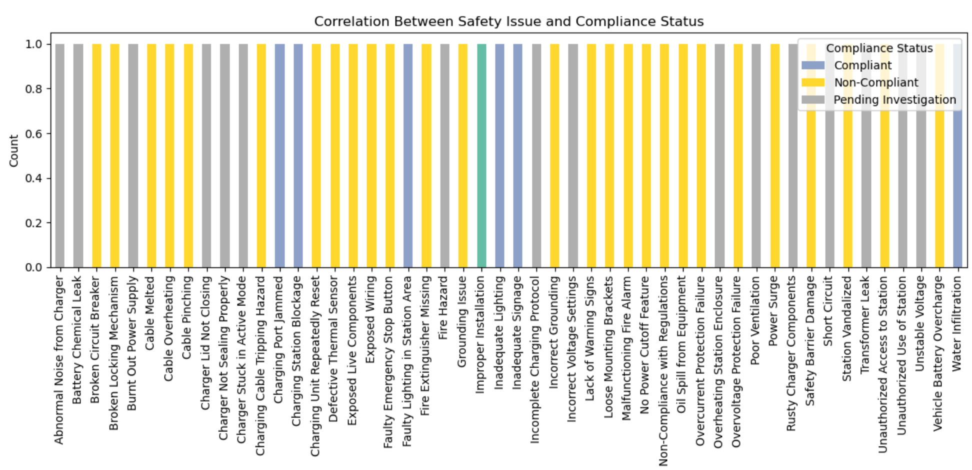
```
> db["station_employees"].aggregate([
  {
    $group: {
      _id: "$station_id",  // Group by station_id
      employee_count: { $sum: 1 }  // Count the number of employees per station
    }
  },
  {
    $sort: { employee_count: -1 }  // Optional: Sort by employee count in descending order
  }
])
< {
  _id: 51699,
  employee_count: 2
}
{
  _id: 63524,
  employee_count: 2
}
{
  _id: 53316,
  employee_count: 2
}
```

Database Access via Python

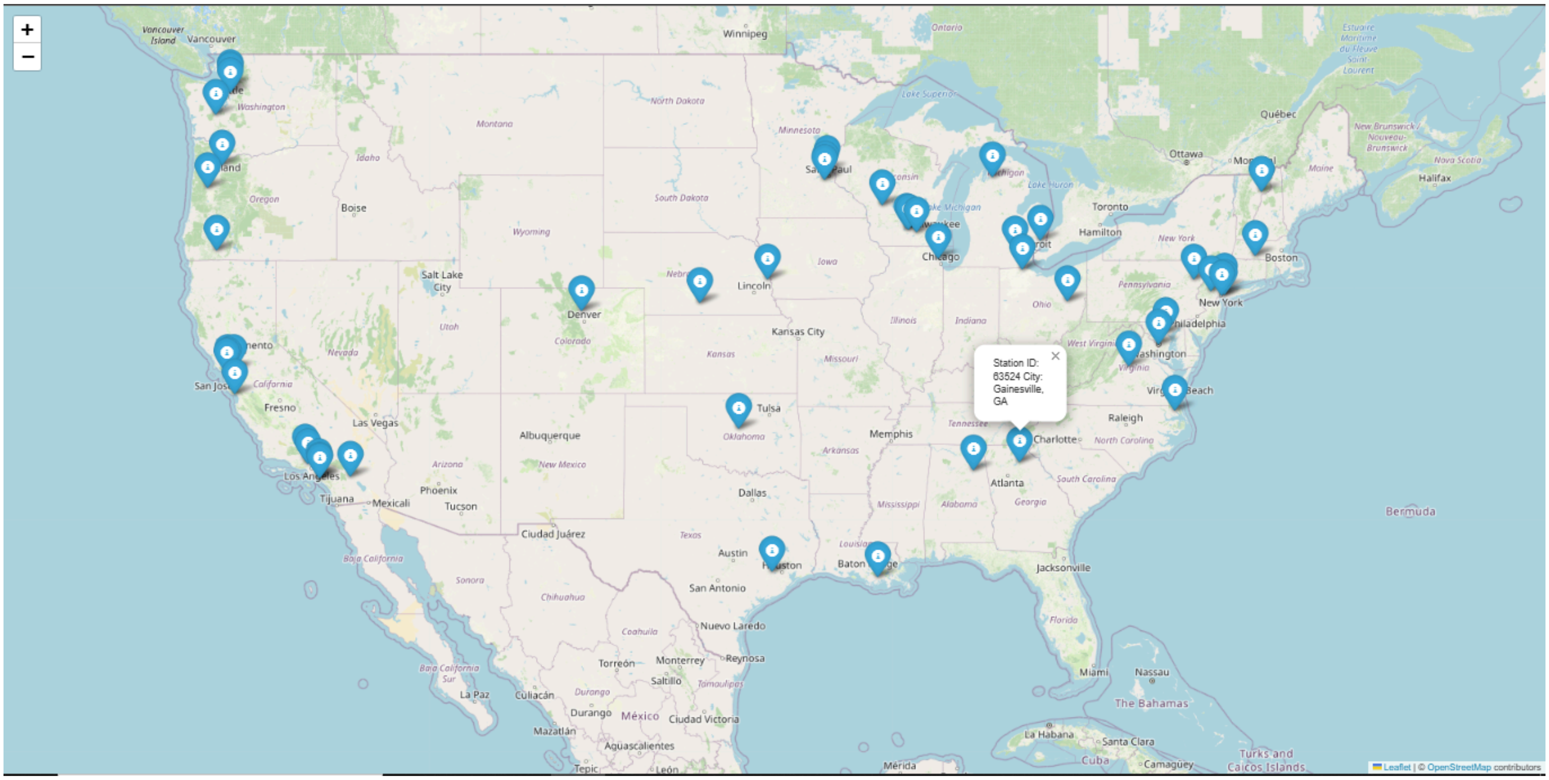
The database is accessed using Python and visualization of analyzed data is shown below. The connection of MySQL to Python is done using mysql.connector, followed by cursor.excecute to run and fetchall from query, followed by converting the list into a dataframe using pandas library and using matplotlib to plot the graphs for the analytics.



Graph 1: This graph displays states ranked from highest to lowest based on their yearly electricity consumption, with the top 5 states being Washington, California, Colorado, New York, Oregon, and Wisconsin. This graph aids in **targeting expansion**, **providing market insights**, **strategic planning**, and **identifying potential government support and incentives** for EV station businesses in high electricity consumption states.



Graph 2: The chart highlights a significant problem with compliance for reported safety issues. Most of the issues are either non-compliant or pending investigation, **signaling that improvements are needed in safety standards or enforcement.**



Graph 3: Using Python libraries like Folium and GeoPandas, along with the MySQL database schema for EV stations, we have created a map displaying **station_id, city, and state across the USA**, covering a total of 50 EV stations. If trained properly, this model can help in optimizing the placement of future EV stations, improving accessibility, and identifying underserved regions for better expansion planning.

Summary and Recommendation

In conclusion, this project successfully designed and implemented a relational database system tailored to the management and optimization of EV charging stations. By leveraging data from the U.S. government's Alternative Fueling Station Locations dataset, the database integrates essential information on station locations, energy consumption, station compatibility, and more. Through careful data cleaning and preprocessing, we ensured that the system could provide meaningful insights into EV infrastructure and support decision-making for future expansion.

As a recommendation, future developments should focus on enhancing the system with additional features such as real-time data integration and predictive analytics to better forecast energy consumption and demand. Implementing stronger data governance practices will also be crucial to maintaining data quality and security as the system scales. Additionally, expanding the database to include more comprehensive nationwide EV infrastructure data will support broader sustainability goals and help in optimizing the deployment of future charging stations across the country. By refining these elements, the database can play a pivotal role in advancing a more efficient and sustainable EV charging network.