EV Charger Analysis Summary Report

1.Data Collection

Being explore the latest related data on the internet. However due to the information limitation provided by the Australian government. In this case, I would like to explore existing database.
https://github.com/Chameleon-company/EVCFLO/blob/main/notebooks/T1_2023/EV%20density%20clustering%20model%20T1%202023/Cleaned_Australian_EV_Charging_Stations.csv

In this database only cover Victoria's EV Charger station. (391 rows x 24 columns).

2. Data Pre-processing

- Handling missing value (prediction for the missing value)

As I can see there's quite a few missing values in the database.

```
# Check for missing values
print(df.isnull().sum())
```

```
Unnamed: 0                     0
Location Name                  0
Latitude                       0
Longitude                      0
Town                           0
Postal Code                    0
City                           0
Address                        0
Plugs_Type2                   30
Plugs_Three_Phase             73
Plugs_CHAdeMO                 18
Plugs_CCS/SAE                 17
Plugs_Tesla                   40
Plugs_J-1772                  73
Plugs_Caravan_Mains_Socket    74
Plugs_wall_AU/NZ              73
Power 1                      264
charging_stations             74
Nearby EVStations              0
Hospitals                      0
Parks                          0
Restaurants                    0
Malls                          0
Supermarkets                   0
dtype: int64
```

Therefore, i would like to use KNN prediction model to handle the missing value in this case. The KNN algorithm can make high accurate predictions.

```
import pandas as pd
from sklearn.impute import KNNImputer

# Read the CSV file
df = pd.read_csv('/Users/jenniferyau/Documents/Deakin unit/2023 T2/SIT764/Database/Cleaned_/

#
numeric_columns = ['Plugs_Type2', 'Plugs_Three_Phase', 'Plugs_CHAdeMO', 'Plugs_CCS/SAE',
                   'Plugs_Tesla', 'Plugs_J-1772', 'Power 1', 'charging_stations']
df[numeric_columns] = df[numeric_columns].apply(pd.to_numeric, errors='coerce')

# Impute missing values using KNN
knn_imputer = KNNImputer(n_neighbors=5)
df[numeric_columns] = knn_imputer.fit_transform(df[numeric_columns])


categorical_columns = ['Plugs_Caravan_Mains_Socket', 'Plugs_wall_AU/NZ']
categorical_imputer = SimpleImputer(strategy='most_frequent')
df[categorical_columns] = categorical_imputer.fit_transform(df[categorical_columns])

# Check for missing values after imputation
print(df.isnull().sum())
```

```
Unnamed: 0                    0
Location Name                 0
Latitude                      0
Longitude                     0
Town                          0
Postal Code                   0
City                          0
Address                       0
Plugs_Type2                   0
Plugs_Three_Phase             0
Plugs_CHAdeMO                 0
Plugs_CCS/SAE                 0
Plugs_Tesla                   0
Plugs_J-1772                  0
Plugs_Caravan_Mains_Socket    0
Plugs_wall_AU/NZ              0
Power 1                       0
charging_stations             0
Nearby EVStations             0
Hospitals                     0
Parks                         0
Restaurants                   0
Malls                         0
Supermarkets                  0
dtype: int64
```

2. Correlation Analysis

From the overview, the "Charging station" and "Nearby EVStations" have wide distribution. "plugs_Type2", "Plugs_Three_Phase", "Plugs-CHAdeMO" are having low values, and some outliners.
- "Hospitals", "Parks", "Restaurant", "Malls", "Supermarkets" are having high means, shows that many locations have these amenities.

2.1 Pearson's correlation matrix

		Power 1	Charging stations	Nearby EV stations	Hospitals	Parks	Restaurants	Malls	Supermarkets
1	Plugs_Type2	-0.175649	0.469643	0.121290	0.148496	0.12254	0.095230	0.142760	0.116831
2	Plugs_Three_Phase	-0.007964	0.004764	-0.016156	-0.026227	-0.022585	0.035344	0.006154	0.061438
3	Plugs_CHAdeMO	0.576256	0.359830	-0.203750	-0.113157	-0.179521	-0.059274	-0.209457	-0.168542
4	Plugs_CCS/SAE	0.550339	0.362185	-0.200636	-0.122305	-0.188770	-0.031902	-0.208932	-0.167341
5	Plugs_Tesla	0.272111	0.149152	0.071600	0.025303	0.021628	0.062412	0.037826	0.057626
6	Plugs_J-1772	-0.035759	0.005798	0.072338	0.098549	0.110906	0.087603	0.141537	0.090624
7	Plugs_Caravan_Main_Socket	-0.007720	-0.018334	-0.039188	-0.029830	-0.013574	-0.010198	-0.020333	-0.037411
8	Plugs_wall_AU/NZ	0.006642	0.494344	0.176230	0.001634	0.169834	0.065161	0.065089	0.178193

Summary for different types of EV plugs and various facilities

Power 1 shows a varied correlation ranging from strong positive with Plugs CHAdeMO to slight negative Plugs_Type2.
Other facilities like hospitals, parks, restaurants, malls, and supermarkets show moderate to slight correlation depending on the plug types.

2.2 Spearman's correlation matrix

		Power 1	Charging stations	Nearby EV stations	Hospitals	Parks	Restaurants	Malls	Supermarkets
1	Plugs_Type2	-0.380785	0.348695	0.121575	0.144171	0.074491	0.044035	0.035945	0.062519
2	Plugs_Three_Phase	0.059889	0.033626	0.035768	-0.008910	-0.003173	0.052484	0.033199	0.034789
3	Plugs_CHAdeMO	0.661819	0.429215	-0.155510	-0.136457	-0.226141	-0.157935	-0.265785	-0.215199
4	Plugs_CCS/SAE	0.669237	0.412855	-0.153567	-0.150209	-0.226696	-0.143265	-0.269395	-0.225689
5	Plugs_Tesla	-0.071358	0.058701	0.086490	0.001987	0.007982	0.051756	0.086379	0.045466
6	Plugs_J-1772	-0.120482	-0.017820	0.076562	0.063015	0.091717	0.069889	0.106375	0.099363
7	Plugs_Caravan_Main_Socket	0.032388	-0.016715	-0.075617	-0.044835	-0.012131	-0.041858	0.005033	-0.032058
8	Plugs_wall_AU/NZ	0.175451	-0.136253	0.090730	0.100569	0.176877	0.053684	0.127801	0.129121

Summary for different types of EV plugs and various facilities

Charging stations usually have moderate to strong positive correlations with most plug types.
Power 1 has varied relationships with different plug types, ranging from strong positive to moderate negative correlations.
Other facilities like hospitals, parks, restaurants, malls, and supermarkets generally show low to moderate correlations depending on the plug type.

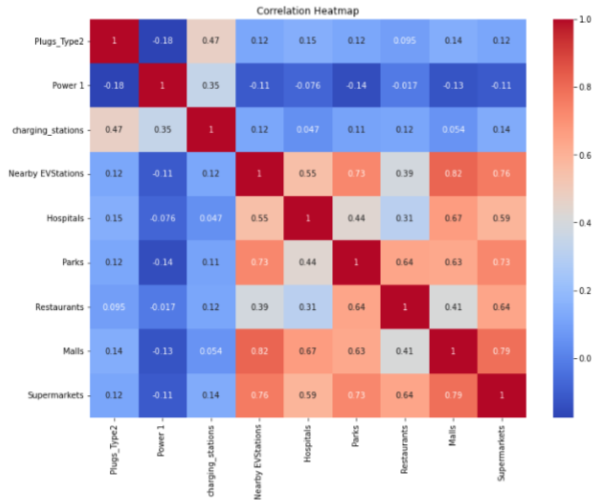
2.3 Heatmap

1:Plus_Type2

In [33]:
```
numeric_columns = [
    'Plugs_Type2','Power 1', 'charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

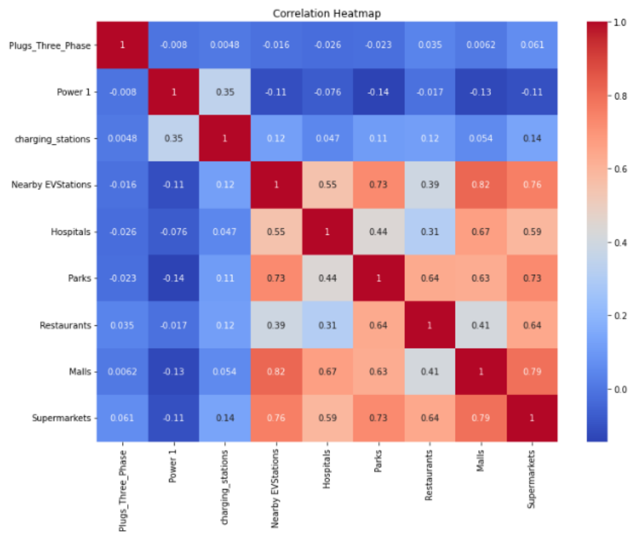


2:Plugs_Three_Phase

In [21]:
```
numeric_columns = [
    'Plugs_Three_Phase','Power 1', 'charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

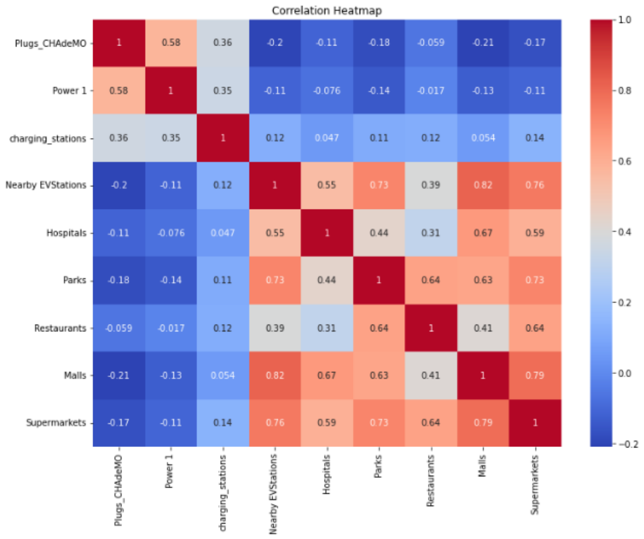


3:Plugs_CHAdeMO

In [22]:
```python
numeric_columns = [
    'Plugs_CHAdeMO', 'Power 1','charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

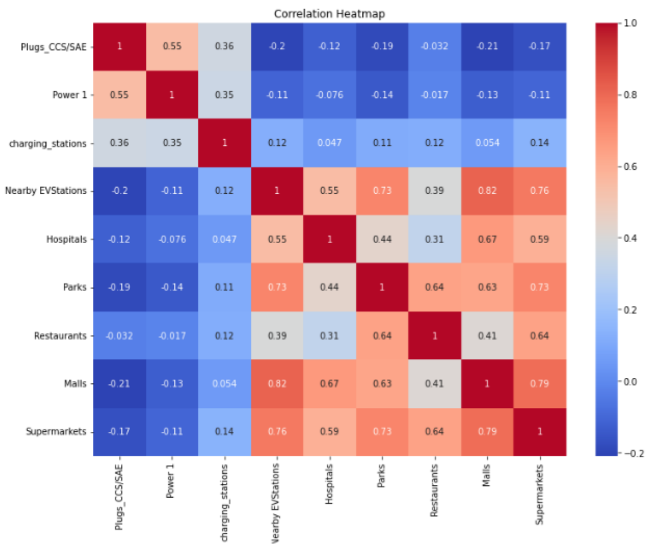


4:Plugs_CCS/SAE

In [23]:
```python
numeric_columns = [
    'Plugs_CCS/SAE','Power 1', 'charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



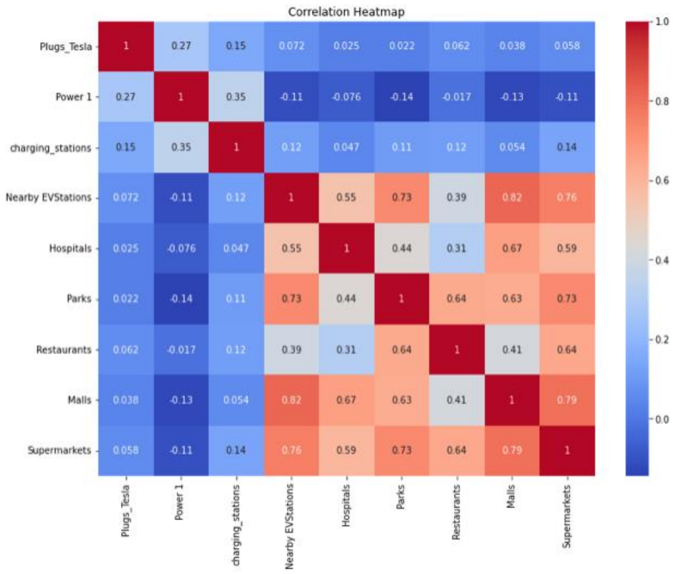
5:Plugs_Tesla

In [24]:

```python
numeric_columns = [
    'Plugs_Tesla', 'Power 1','charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



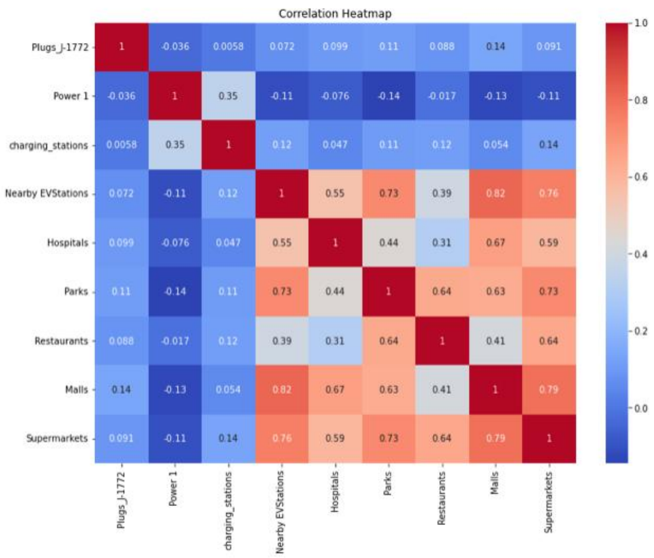
6:Plugs_J-1772

In [25]:

```python
numeric_columns = [
    'Plugs_J-1772', 'Power 1','charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



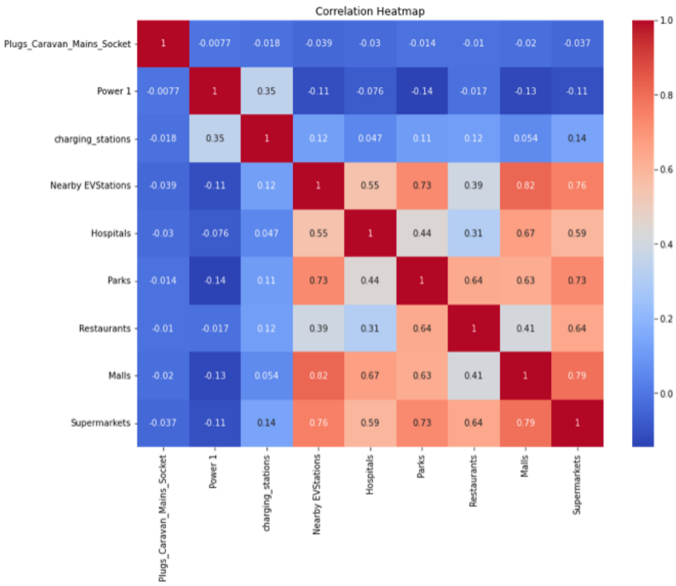
7:Plugs_Caravan_Mains_Socket

In [27]:

```python
numeric_columns = [
    'Plugs_Caravan_Mains_Socket', 'Power 1','charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



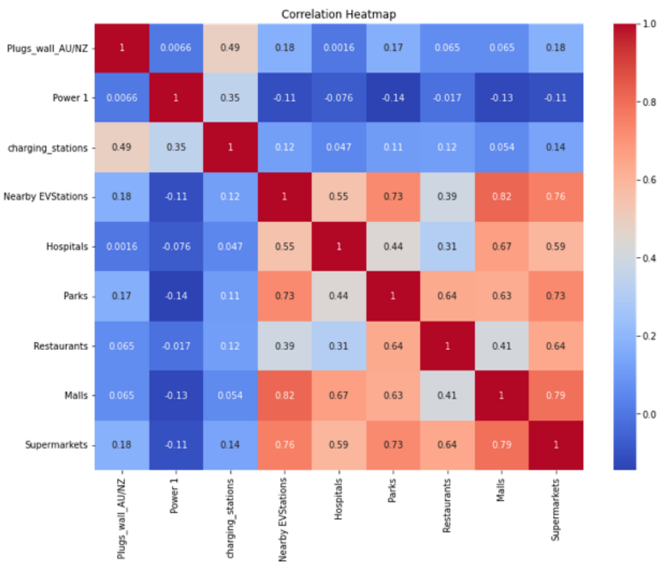
8:Plugs_wall_AU/NZ

In [28]:

```python
numeric_columns = [
    'Plugs_wall_AU/NZ', 'Power 1','charging_stations',
    'Nearby EVStations', 'Hospitals', 'Parks', 'Restaurants', 'Malls', 'Supermarkets'
]


# Calculate the correlation matrix
correlation_matrix = df[numeric_columns].corr()

# Plot the heatmap
plt.figure(figsize=(12, 9))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



3. Regression analysis

In my analysis I would like to try different type of models (Linear Regression, Polynomial regression and Random forest) to compare the result.

3.1 Linear Regression

		R-squared	Mean Squared Error
1	Plugs_Type2	0.23172478143364983	3.4074144630450434
2	Plugs_Three_Phase	-4.014938265980382	0.0025070673593749052
3	Plugs_CHAdeMO	0.07925519868617648	0.45557762321055706
4	Plugs_CCS/SAE	0.18704101403464768	0.7060146937882084
5	Plugs_Tesla	0.09023160014988751	1.5787200401180443
6	Plugs_J-1772	-0.06555887217729217	0.09033945048012315
7	Plugs_Caravan_Main_Socket	0.0	2.661426631853063e-05
8	Plugs_wall_AU/NZ	0.3777445637420255	4.17203276314352

3.2 Linear Regression with MinMaxScaler

		R-squared	Intercept
1	Plugs_Type2	0.23172478143364938	0.6926061544530675
2	Plugs_Three_Phase	-4.01493826598038	0.02623513334681194
3	Plugs_CHAdeMO	0.07925519868617592	0.0845265340787062
4	Plugs_CCS/SAE	0.18704101403464757	0.07206487757079799
5	Plugs_Tesla	0.09023160014988763	0.1416160231212688
6	Plugs_J-1772	-0.06555887217729195	0.0406339305303009
7	Plugs_Caravan_Main_Socket	0.0	0.004514637420014581
8	Plugs_wall_AU/NZ	0.37774456374202536	-0.027022127158421116

R-squared values suggest that linear models may not be the best fit for predicting these types of electric vehicle plugs based on the available independent variables.

3.3 Polynomial regression

		R-squared	Mean Squared Error
1	Plugs_Type2	0.6014536389272575	1.7676121812797398
2	Plugs_Three_Phase	-54.535461532587135	0.027763281522459826
3	Plugs_CHAdeMO	-0.2625699433447506	0.6247101402096763
4	Plugs_CCS/SAE	0.07992574568464694	0.7990390095159771
5	Plugs_Tesla	-0.5990679876896285	2.7748608086330186
6	Plugs_J-1772	-0.1950110372261562	0.10131457138553174
7	Plugs_Caravan_Main_Socket	0.0	0.000162090101993577
8	Plugs_wall_AU/NZ	0.8578378303578298	0.9531539539347821

Polynomial regression improved the model for some plug types, like Plugs_Type2 and Plugs_wall_AU/NZ. Plugs_Type2 and Plugs_wall_AU/NZ show a decent R-squared value, suggesting the model fits well for these types.
Models for Plugs_Three_Phase, Plugs_CHAdeMO, Plugs_Tesla, and Plugs_J-1772 have negative R-squared values, suggesting poor fits and possibly, incorrect model specifications.

3.4 Random Forest

		R-squared	Mean Squared Error
1	Plugs_Type2	0.6469500030347084	1.5658290632911394
2	Plugs_Three_Phase	-5.112371794871798	0.0030556962025316456
3	Plugs_CHAdeMO	0.7736998950777202	0.11197159493670886
4	Plugs_CCS/SAE	0.8734544767527676	0.10989853164556962
5	Plugs_Tesla	0.31901600258541074	1.181710734177215
6	Plugs_J-1772	-0.4525411324463262	0.12314830379746834
7	Plugs_Caravan_Main_Socket	0.0	1.7721518987341772e-05
8	Plugs_wall_AU/NZ	0.7022940015294906	1.9960278481012663

	Feature importance	Power 1	Charging stations	Nearby EV stations	Hospitals	Parks	Restaurants	Malls	Supermarkets
1	Plugs_Type2	0.28214319	0.55460491	0.02371025	0.04749269	0.0365990 1	0.01652751	0.0228173 1	0.01610513
2	Plugs_Three_Phase	0.37289231	0.11773888	0.07595182	0.03609721	0.1473879	0.05235517	0.0729625 3	0.12461418
3	Plugs_CHAdeMO	0.55003946	0.3684972	0.02345293	0.01205618	0.0242688 9	0.00312927	0.0123369 7	.00621909
4	Plugs_CCS/SAE	0.50479696	0.4039602	0.01558058	0.01681263	0.0411752 1	0.00367935	0.0070532 4	0.00694183
5	Plugs_Tesla	0.65598947	0.14387207	0.04207457	0.02445433	0.0520006	0.01703172	0.0300359 5	0.0345413
6	Plugs_J-1772	0.48565624	0.08195872	0.04435134	0.05677675	0.0819385 1	0.01647013	0.1430271 7	0.08982114
7	Plugs_Caravan_Main_Socket	0.31410027	0.08174816	0.03090947	0.	0.1516223 2	0.14756957	0.1059222 4	0.16812797
8	Plugs_wall_AU/NZ	0.18460403	0.53149666	0.10925812	0.02770226	0.0648262 6	0.00895484	0.0447229 8	0.02843486

Overall, the Random Forest model seems to offer the most consistently good fits across multiple types of plugs. Most plug types have positive R-squared values, with Plugs_CCS/SAE showing the best fit. Feature importance varies by plug type, but 'Charging Stations' and 'Power 1' are often the most significant features.

4.1 Cross-validation

	Feature importance	Power 1	Charging stations	Nearby EV stations	Hospitals	Parks	Restaurants	Malls	Supermarkets
1	Plugs_Type2	26.67%	52.50%	2.65%	5.68%	7.13%	0.99%	2.20%	2.18%
2	Plugs_Three_Phase	43.06%	12.18%	5.57%	3.52%	12.45%	4.98%	7.28%	10.97%
3	Plugs_CHAdeMO	57.38%	31.00%	4.29%	0.60%	3.88%	0.61%	1.02%	1.22%
4	Plugs_CCS/SAE	53.96%	36.80%	3.54%	0.74%	3.03%	0.3%	0.87%	0.71%
5	Plugs_Tesla	54.95%	15.45%	4.05%	3.22%	7.74%	2.65%	4.75%	7.19%
6	Plugs_J-1772	55.07%	6.80%	3.74%	4.73%	5.64%	1.50%	7.23%	15.28%
7	Plugs_Caravan_Main_Socket	41.87%	4.65%	2.42%	0	9.92%	15.32%	9.02%	16.79%
8	Plugs_wall_AU/NZ	9.67%	75.29%	4.02%	2.05%	4.07%	0.61%	2.46%	1.82%

Based on the result, would recommend integrating more Plugs_Three_Phase and Plugs_Caravan_Main_Socket near supermarkets and other high-importance features to meet user demands for these plug types. Businesses could focus on placing high-speed chargers where users of Plugs like Type2, CHAdeMO, and Tesla are likely to need them. Also understand that the feature importance can help in resource allocation. Therefore, would recommend not to invest heavily in placing Plugs_Type2 near hospitals if that feature has low importance in the model.

4.2 Coefficient Analysis

		Power 1	Charging stations	Nearby EV stations	Hospitals	Parks	Restaurants	Malls	Supermarkets
1	Plugs_Type2	-0.018474	0.533544	-0.032778	0.071359	0.003829	0.012373	0.035459	-0.032669
2	Plugs_Three_Phase	-0.000085	0.001018	-0.002813	-0.005770	-0.004526	0.000119	-0.000363	0.008296
3	Plugs_CHAdeMO	0.009156	0.086694	-0.013569	0.005846	0.000318	-0.003654	-0.011375	0.002494
4	Plugs_CCS/SAE	0.009072	0.098171	-0.009402	0.003547	-0.007543	0.004411	-0.010690	-0.003069
5	Plugs_Tesla	0.005023	0.032624	0.020303	0.009796	-0.007979	0.007092	-0.003426	-0.006271
6	Plugs_J-1772	-0.000343	0.008645	-0.013437	-0.001244	0.007973	0.002126	0.022306	-0.007793
7	Plugs_Caravan_Main_Socket	-0.000007	-0.000365	-0.000698	-0.000402	0.000412	0.000082	0.000682	-0.000559
8	Plugs_wall_AU/NZ	-0.003724	0.241092	0.042729	-0.059108	0.006724	-0.021505	-0.034750	0.042084

Plugs_Type2: "Charging Stations" has the highest positive coefficient (0.533544), implying a strong positive relationship. However, "Power 1" and "Nearby EV Stations" have small negative coefficients.

Plugs_Three_Phase: All coefficients are near zero, indicating that the features have minimal linear impact on this plug type.

Plugs_CHAdeMO: A positive coefficient for "Charging Stations" and "Power 1" suggests that these factors encourage the use of CHAdeMO plugs. Negative coefficients for "Nearby EV Stations" and "Malls" suggest otherwise.

Plugs_CCS/SAE: Similar to CHAdeMO, but "Power 1" and "Charging Stations" have stronger positive coefficients.

Plugs_Tesla: "Charging Stations" and "Nearby EV Stations" both have positive coefficients, implying a positive relationship. Negative coefficients for features like "Parks" suggest these may not be as important.

Plugs_J-1772: "Malls" have the highest positive coefficient (0.022306), followed by "Parks" and "Charging Stations." Negative coefficients for "Nearby EV Stations" and "Supermarkets" suggest they might reduce the usage of J-1772 plugs.

Plugs_Caravan_Main_Socket: All coefficients are nearly zero, indicating negligible linear relationship with the dependent variable.

Plugs_wall_AU/NZ: "Charging Stations" have an extremely strong positive impact (0.241092). Negative coefficients for "Hospitals" and "Restaurants" suggest that these factors might not encourage the usage of these plug types.

5. Conclusion

Based on the result, would like to recommend the specific plug types near amenities like malls or parks, based on the feature importance and coefficient values. Also, would recommend to focus on increasing the number of charging stations, especially for Plugs_Type2 and Plugs_CCS/SAE.
This analysis provides valuable insights into the relationships between various features and the usage of different EV plug types. A deeper understanding of user behaviour can inform more effective strategies for increasing the adoption and utilization of EV charging facilities.