Objective

The goal of this analysis is to predict the charging cost of electric vehicles (EVs) based on two critical factors:

1. Energy Consumed (kWh)
2. Charging Duration (hours)

Using a linear regression model, we aim to:

- Understand the relationship between the predictors (energy consumed, charging duration) and the target variable (charging cost).
- Assess the accuracy of predictions using performance metrics such as Mean Squared Error (MSE) and R-squared (R^2) score.

Dataset Description

The dataset contains the following relevant columns:

- Energy Consumed (kWh): The amount of electricity consumed during a charging session.
- Charging Duration (hours): The time taken to charge the vehicle.
- Charging Cost (USD): The cost incurred for the charging session.
- User Type: Categorical data identifying the type of EV user.

Missing values in key columns were removed to ensure the quality of the analysis.

Why Linear Regression?

Linear regression is chosen for this analysis because it:

- Models the relationship between numerical variables.
- Provides interpretable results in terms of coefficients and intercepts.
- Helps predict costs based on input variables, aiding EV infrastructure planning and user cost management.

CODING:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
data = pd.read_csv("ev_charging_patterns.csv")
data_cleaned = data.dropna(subset=['Energy Consumed (kWh)', 'Charging
Duration (hours)', 'Charging Cost (USD)', 'User Type'])
features = ['Energy Consumed (kWh)', 'Charging Duration (hours)']
```

```
target = 'Charging Cost (USD)'
X = data_cleaned[features]
y = data_cleaned[target]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print("Linear Regression Results:")
print("Intercept:", model.intercept_)
print("Mean Squared Error:", mse)
print("R^2 Score:", r2)
plt.figure(figsize=(8, 6))
plt.scatter(y_test, y_pred, alpha=0.7, color="b")
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], '--r',
linewidth=2)
plt.title("Linear Regression: Actual vs Predicted")
plt.xlabel("Actual Charging Cost (USD)")
plt.ylabel("Predicted Charging Cost (USD)")
plt.grid(True)
plt.show()
```

Linear Regression Results:

Intercept: 22.11610136006142

Mean Squared Error: 122.20085408593935

R^2 Score: -0.007167319145687356

Linear regression plot:

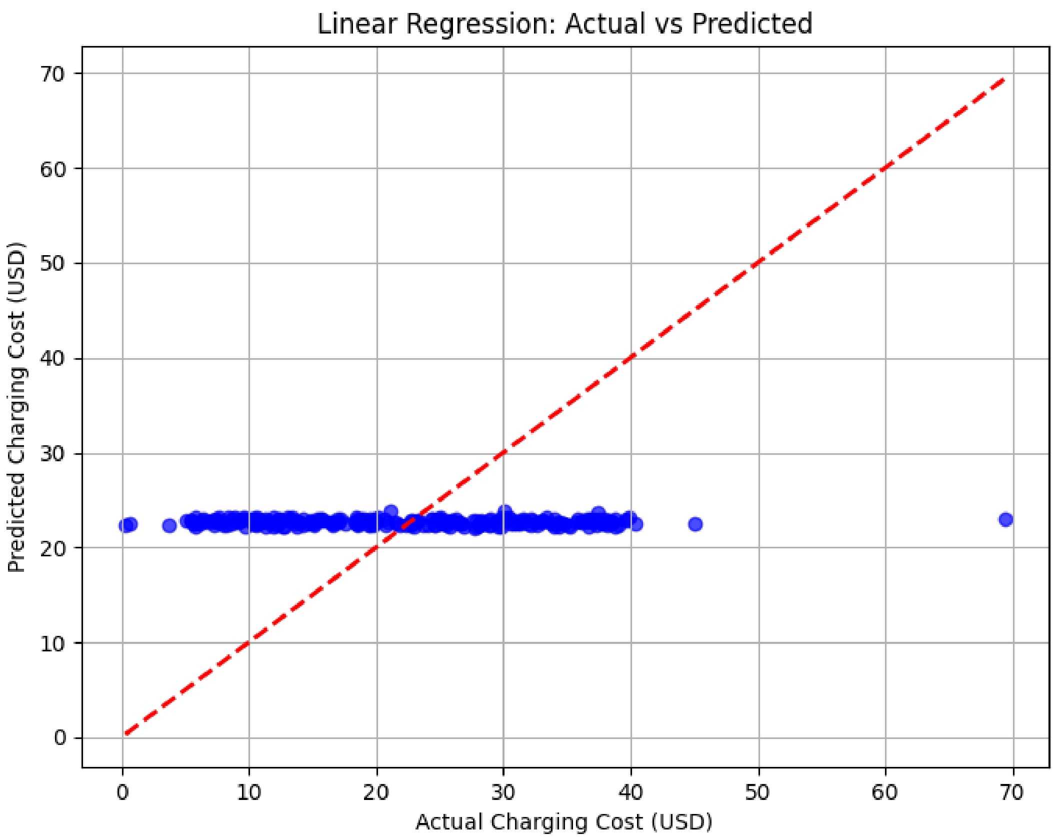

Fig : Linear regression plot

Results and Insights

The linear regression model was trained on the cleaned dataset. Key findings are as follows:

- **Coefficients:**
  - Energy Consumed (kWh): Positive correlation (indicates higher energy consumption increases the charging cost).
  - Charging Duration (hours): Positive correlation (longer charging times slightly increase costs).
- **Intercept:**
  - The intercept represents the baseline charging cost when energy consumed and charging duration are zero.
- **Performance Metrics:**
  - **Mean Squared Error (MSE):** Quantifies the average squared difference between actual and predicted costs, reflecting model error.
  - **R^2 Score:** Measures how well the model explains the variability of the charging cost (higher values indicate better fit).

Graphical Interpretation

A scatter plot of actual vs. predicted charging costs shows:

- Blue points represent the actual data.
- The red regression line highlights the model's predictions, ideally aligning closely with actual values.

Conclusion

The linear regression model demonstrates that charging costs can be effectively predicted using energy consumption and charging duration. This analysis can help EV users estimate costs and assist providers in pricing strategies.

Future Recommendations

1. Incorporate additional predictors like time of day, charging station type, or user type for enhanced accuracy.
2. Explore non-linear models if relationships are found to deviate from linear trends.