<https://chatgpt.com/c/67cd4650-d618-8003-8761-82e184e623f8>

Key Factors Influencing Has\_EV

1️⃣ Charging Infrastructure & Vehicles

✅ Charging\_Type

* If "Level 1", "Level 2", or "DC Fast" → 100% chance of EV
* If "None" → Reduced probability

✅ Num\_Vehicles

* More vehicles → Higher probability of an EV

✅ Charging\_Cycles\_Per\_Week

* More charging cycles → Higher probability of an EV

✅ Nearby\_Charging\_Station\_Distance\_km

* Farther charging stations → Lower probability of EV

2️⃣ Household Characteristics

✅ Total\_Electricity\_Usage\_kWh

* Higher electricity usage → Higher probability of an EV

✅ Nighttime\_Usage\_kWh

* Higher nighttime consumption → Higher probability of an EV (Since EVs often charge at night)

✅ Peak\_Load\_Night\_kW

* Higher peak nighttime load → More likely to have an EV

✅ Num\_Residents

* More residents → Slightly higher probability of an EV (as larger families might have multiple cars)

✅ Household\_Income\_kUSD

* Higher income → Higher probability of an EV
* Low income (<50k) → Lower probability

✅ Home\_Type

* Detached Houses → Higher probability of EV ownership (Easier to install home chargers)
* Apartments → Lower probability

✅ Has\_Solar\_Panels

* If solar panels are present → Higher probability of EV ownership

✅ Geographic\_Location\_Score

* Higher score → More likely to own an EV (Better infrastructure & adoption rates)

3️⃣ Weather & Environmental Factors

✅ Avg\_Temperature\_C

* Extremely cold (< -5°C) or very hot (> 35°C) → Lower probability of EV ownership (Battery performance issues)
* Moderate temperatures (5°C - 30°C) → Higher probability

✅ Sunlight\_Hours\_Per\_Day

* More sunlight → Higher probability (correlated with solar panel ownership & EV adoption)

✅ Rainfall\_mm & Snowfall\_cm

* Heavy rain (>30mm) or snowfall (>20cm) → Slightly lower probability (EVs have lower adoption in extreme weather regions)

🔹 Summary

The strongest predictors of Has\_EV are:  
🚀 Charging\_Type, Total\_Electricity\_Usage\_kWh, Household\_Income\_kUSD, Home\_Type, and Has\_Solar\_Panels

Other secondary influencers:  
📊 Num\_Vehicles, Geographic\_Location\_Score, Nighttime\_Usage\_kWh, Avg\_Temperature\_C, Nearby Charging Station Distance, and Sunlight Hours

====================================================================

Negative influencers

====================================================================

Yes! In real life, several additional factors could negatively impact the likelihood of owning an EV. Here are some realistic features that, if added to the dataset, could make the Has\_EV prediction more accurate by introducing negative influences:

🔹 Additional Realistic Negative Factors for Has\_EV

1️⃣ Financial & Economic Barriers

✅ High Cost of Living (Cost\_of\_Living\_Index)

* If a household is in an area with a high cost of living, they might prioritize other expenses over an EV.
* Example: Cities like New York or San Francisco have high expenses, so middle-income households may avoid EVs.

✅ Existing Vehicle Debt (Auto\_Loan\_Balance\_kUSD)

* If a household already has a car loan, they may be less likely to switch to an EV.
* Example: Someone with a $30,000 car loan on a gas vehicle is less likely to buy an EV soon.

✅ Limited Government Incentives (EV\_Subsidy\_Available)

* If a region does not offer subsidies or tax breaks for EVs, ownership rates tend to be lower.
* Example: Some U.S. states (e.g., Wyoming, North Dakota) have fewer incentives for EV buyers.

✅ High Electricity Costs (Electricity\_Cost\_Per\_kWh)

* If electricity is expensive, people may avoid EVs due to higher charging costs.
* Example: Some countries or states (e.g., Germany, California) have high electricity prices that discourage EV ownership.

2️⃣ Infrastructure & Location Issues

✅ Lack of Home Parking (Has\_Private\_Parking)

* People living in apartments without private parking find it harder to charge an EV.
* Example: A New York City resident who relies on street parking is unlikely to own an EV.

✅ Unreliable Power Grid (Grid\_Reliability\_Score)

* If a location has frequent power outages, EV adoption may be low due to charging concerns.
* Example: Some rural or developing areas have unstable electricity supply, discouraging EVs.

✅ Rural vs. Urban (Population\_Density)

* Very rural areas (low population density) might have fewer charging stations, discouraging EVs.
* Example: Farmers or ranchers in Montana may prefer gas vehicles for long distances.

✅ Cold Climate Impact on Battery (Winter\_Temperature\_C)

* EV range drops significantly in extremely cold regions (batteries perform poorly in subzero temps).
* Example: Nordic countries invest heavily in charging infrastructure to compensate.

3️⃣ Driving & Lifestyle Factors

✅ High Daily Commute Distance (Daily\_Commute\_Distance\_km)

* If a person drives long distances daily (e.g., 200+ km), they may hesitate to own an EV due to range limitations.
* Example: A salesperson driving long distances across a rural state may prefer gasoline cars.

✅ Frequent Towing or Heavy Load Use (Towing\_Requirement)

* EVs are not ideal for towing large trailers or heavy loads.
* Example: Pickup truck users who need to tow boats or trailers may avoid EVs.

✅ Preference for Gasoline Cars (Prefers\_ICE\_Vehicles)

* Some people simply prefer gasoline/diesel vehicles due to tradition or driving habits.
* Example: Many car enthusiasts, truck owners, or sports car drivers avoid EVs.

✅ Lack of EV Awareness (EV\_Knowledge\_Score)

* Some people are not well-informed about EV benefits and hesitate to switch.
* Example: Older generations or people in remote areas may not trust EV technology.

🔹 Summary: Key New Negative Predictors

🚫 Financial Barriers: High cost of living, existing car loan, no EV subsidies, high electricity costs  
🚫 Infrastructure Issues: No home parking, unreliable power grid, no charging stations, extreme cold weather  
🚫 Lifestyle Barriers: Long commute distance, heavy towing needs, gasoline preference, lack of EV awareness

Would you like me to modify the dataset to include some of these negative factors? 🚗⚡

====================================================================

**Key Updates in this Version:**

1️⃣ **New Columns Added**:

* Daily\_Commute\_Distance\_km (Longer commute → Less likely to own an EV)
* Has\_Private\_Parking (No private parking → Less likely to own an EV)
* Cost\_of\_Living\_Index (Higher cost → Less likely to own an EV)
* Electricity\_Cost\_Per\_kWh (Higher electricity cost → Less likely to own an EV)

======================================================================

Sure! Below is the **complete list of features** used in the dataset so far:

**🔹 Household-Level Features (Static)**

1. **Household\_ID** → Unique identifier for each household
2. **Total\_Electricity\_Usage\_kWh** → Total daily electricity consumption
3. **Nighttime\_Usage\_kWh** → Electricity usage during night hours
4. **Peak\_Load\_Night\_kW** → Peak power demand during night hours
5. **Charging\_Cycles\_Per\_Week** → How many times an EV is charged per week
6. **Power\_Factor\_Variation** → Variability in power factor (indicates efficiency of electricity usage)
7. **Num\_Vehicles** → Number of vehicles owned by the household
8. **Household\_Income\_kUSD** → Household annual income (in thousand USD)
9. **Has\_Solar\_Panels** → Whether the household has solar panels (0 = No, 1 = Yes)
10. **Geographic\_Location\_Score** → A score (0 to 1) indicating how EV-friendly the location is

**🔹 Weather & Environmental Factors**

1. **Avg\_Temperature\_C** → Average daily temperature in °C
2. **Sunlight\_Hours\_Per\_Day** → Average number of sunlight hours per day
3. **Rainfall\_mm** → Daily rainfall in mm
4. **Snowfall\_cm** → Daily snowfall in cm

**🔹 EV Infrastructure & Charging Access**

1. **Nearby\_Charging\_Station\_Distance\_km** → Distance to the nearest public EV charging station (km)

**🔹 Household Characteristics**

1. **Num\_Residents** → Number of residents in the household
2. **Daily\_Commute\_Distance\_km** → Average daily commute distance for the household (km)
3. **Has\_Private\_Parking** → Whether the household has private parking (0 = No, 1 = Yes)
4. **Cost\_of\_Living\_Index** → Cost of living index (higher values mean more expensive areas)
5. **Electricity\_Cost\_Per\_kWh** → Cost of electricity per kWh (USD)

**🔹 Target Variable**

1. **Has\_EV** → Whether the household owns an EV (0 = No, 1 = Yes)

**🔹 Time-Series Electricity Data (Dynamic)**

1. **Timestamp** → 15-minute interval timestamps (spanning 1 month)
2. **Electricity\_Consumption\_kWh** → Electricity consumption for the household at that specific timestamp

**💡 Summary**

* ✅ **21 static household-level features**
* ✅ **1 dynamic feature (electricity consumption every 15 min)**
* ✅ **1 target variable (Has\_EV)**
* ✅ **Total: 23 features**

======================================================================

Good question! Let's analyze whether any feature in the dataset might **not be significantly impacting Has\_EV**.

**🔍 Features with Strong Impact on Has\_EV**

These features have a **clear impact** (either positive or negative) on EV ownership:  
✅ **Strong Positive Impact:**

* Charging\_Cycles\_Per\_Week → Directly related to EV usage
* Num\_Vehicles → More vehicles, higher chance of having an EV
* Household\_Income\_kUSD → Higher income, more likely to afford an EV
* Has\_Solar\_Panels → Solar owners might prefer EVs for lower-cost charging
* Nearby\_Charging\_Station\_Distance\_km → If far away, less likely to own an EV

✅ **Strong Negative Impact:**

* Daily\_Commute\_Distance\_km → Longer commutes make EV ownership less likely
* Has\_Private\_Parking → No parking = No easy EV charging
* Cost\_of\_Living\_Index → High cost of living reduces disposable income for EVs
* Electricity\_Cost\_Per\_kWh → Expensive electricity discourages EV adoption

**🔍 Features That Might Have Little or No Impact on Has\_EV**

Some features might not directly influence EV ownership or have **a weak correlation**:

❓ **Possible Weak Impact:**  
🔹 **Power\_Factor\_Variation** → Measures electricity efficiency but has no direct link to EV ownership.  
🔹 **Peak\_Load\_Night\_kW** → May indicate high energy use but doesn’t necessarily mean an EV is present.  
🔹 **Rainfall\_mm & Snowfall\_cm** → EV adoption is **slightly** impacted by weather, but the effect is small.  
🔹 **Sunlight\_Hours\_Per\_Day** → Affects solar panels but **doesn’t directly impact EV ownership** unless combined with Has\_Solar\_Panels.  
🔹 **Geographic\_Location\_Score** → Could be refined further. If it's too generic, it may not strongly impact Has\_EV.  
🔹 **Num\_Residents** → May not have a strong effect unless larger families prefer multiple gas-powered vehicles.

**🔍 Features That Might Need Refinement**

Some features **could** impact Has\_EV, but they **might need adjustments** to be more relevant:  
🔸 **Avg\_Temperature\_C** → Cold weather **can reduce battery performance**, but alone, it may not strongly determine EV ownership.  
🔸 **Total\_Electricity\_Usage\_kWh** → A household could use a lot of electricity for reasons other than EV charging (e.g., heating, large appliances).

======================================================================

Feature Engineering Steps:

======================================================================

**1. Handling Categorical Features**

**A. Encoding Categorical Variables**

Some features in your dataset are categorical (Has\_Private\_Parking, House\_Type, Electricity\_Tariff\_Type). These need to be converted into numerical format:  
✅ **One-Hot Encoding** (For non-ordinal categories like House\_Type)  
✅ **Label Encoding** (For ordinal categories like Electricity\_Tariff\_Type)

🔹 **Example (One-Hot Encoding for House\_Type)**

This will convert House\_Type into separate columns like:

* House\_Type\_Townhouse
* House\_Type\_SingleFamilyHome

🔹 **Example (Label Encoding for Electricity\_Tariff\_Type)**

If Electricity\_Tariff\_Type has values like **Fixed, Time-of-Use, Variable**, they will be encoded as **0, 1, 2**.

**🔹 2. Creating New Features**

**A. Energy Consumption Patterns**

* Nighttime\_Energy\_Ratio = Nighttime\_Usage\_kWh / Total\_Electricity\_Usage\_kWh
  + **Why?** If the ratio is **high**, it's a sign of **EV charging at night**.
* Peak\_Load\_Difference = Peak\_Load\_Night\_kW - (Total\_Electricity\_Usage\_kWh / 24)
  + **Why?** A large peak at night might indicate an **EV charging session**.

**B. Transportation Behavior**

* Cost\_Per\_Commute = (Daily\_Commute\_Distance\_km / Fuel\_Cost\_Per\_Liter\_USD) \* 10
  + **Why?** Higher commuting costs may encourage EV adoption.
* Public\_Transport\_Access = Public\_Transport\_Score \* Nearby\_Charging\_Station\_Distance\_km
  + **Why?** If **public transport is good & charging stations are far**, EV adoption is less likely.

---------------------------

**3. Feature Scaling & Transformation**

Some features have very different ranges (Household\_Income\_kUSD vs. Nighttime\_Usage\_kWh).  
✅ Use **MinMax Scaling** or **Standard Scaling** for better ML performance.

🔹 **Example (Using Standard Scaling)**

**4. Handling Missing Values**

✅ **Numerical Features** → Fill with **mean/median**  
✅ **Categorical Features** → Fill with **mode (most frequent value)**

**5. Feature Selection (Remove Unimportant Features)**

If some features **don’t impact EV ownership**, they can be removed using techniques like:  
✅ **Correlation Analysis** → Remove highly correlated features  
✅ **Feature Importance (Random Forest / XGBoost)** → Drop least important features

🔹 **Example (Using Random Forest Feature Importance)**

**🚀 Final Thoughts**

✅ **Apply One-Hot/Label Encoding for categorical data**  
✅ **Create new features from energy usage & commute behavior**  
✅ **Normalize numerical features to improve model accuracy**  
✅ **Remove unimportant features using feature importance analysis**