Impact of Gas Prices on Electric Vehicle Adoption: A Data Pipeline Approach

1. Question

This project explores the relationship between gas prices and electric vehicle (EV) adoption in the United States. The hypothesis is that rising gas prices encourage EV adoption due to cost-saving motivations. A data pipeline was developed to preprocess, transform, and merge gas price and EV sales datasets for comprehensive analysis.

2. Data Sources

2.1 Gas Price Data

Source: U.S. Energy Information Administration (EIA). **Format:** Weekly time-series data (XLS). **Description:** Weekly U.S. gas prices (1997–present).
   Challenges:

- Inconsistent weekly intervals.

- Multiple entries for some weeks.

2.2 EV Registration Data

Source: Atlas EV Hub. **Format:** Event-based data (CSV). **Description:** EV registration records (2012–2021 , vary with states).
   Challenges:

- Aggregation to monthly intervals.

- Missing data for specific states or months.

3. Data Pipeline

The pipeline dynamically handles inconsistent data and logs issues for future reference.

3.1 Gas Price Data Transformation

1. **Sorting and Consistency Check:** Data is sorted chronologically. Deviations from 7-day intervals are logged.

2. **Normalizing Weekly Data:** Multiple entries for a week are averaged. **Example:** Week 1 prices $[2.50, 2.55] \rightarrow 2.525$.

3. **Monthly Aggregation:** Weekly prices grouped to compute monthly averages. **Example:** January 2021 weekly prices $[2.50, 2.60, 2.70, 2.80] \rightarrow$ Monthly Average: 2.65 USD.

3.2 EV Registration Data Transformation

1. **Column Filtering:** Retain only `registration_date` and `vehicle_name`.

2. **Monthly Aggregation:** Count total registrations per month. **Example:** January 2021: Tesla (500), Ford (300), Porsche (200) $\rightarrow$ 1000 total registrations.

3. **Timeframe Adjustment:** Align with gas data, e.g., 2012–2021.

3.3 Merging Datasets

1. **Outer Join on Timestamps:** Gas price and EV data merged by month. Missing values are logged and filled with `NIL`.

2. **Cleaning:** Rows where both `price` and `volume` are `NIL` are removed.

3. **Final Dataset Example:**

Month	Gas Price (USD)	EV Volume
31.01.2021	2.65	1000
28.02.2021	2.56	1200
31.03.2021	2.80	3000

Table 1: Merged Dataset Example

3.4 Pipeline

The pipeline integrates gas price and EV data through various stages, ensuring data consistency and alignment. The overall workflow is depicted below:
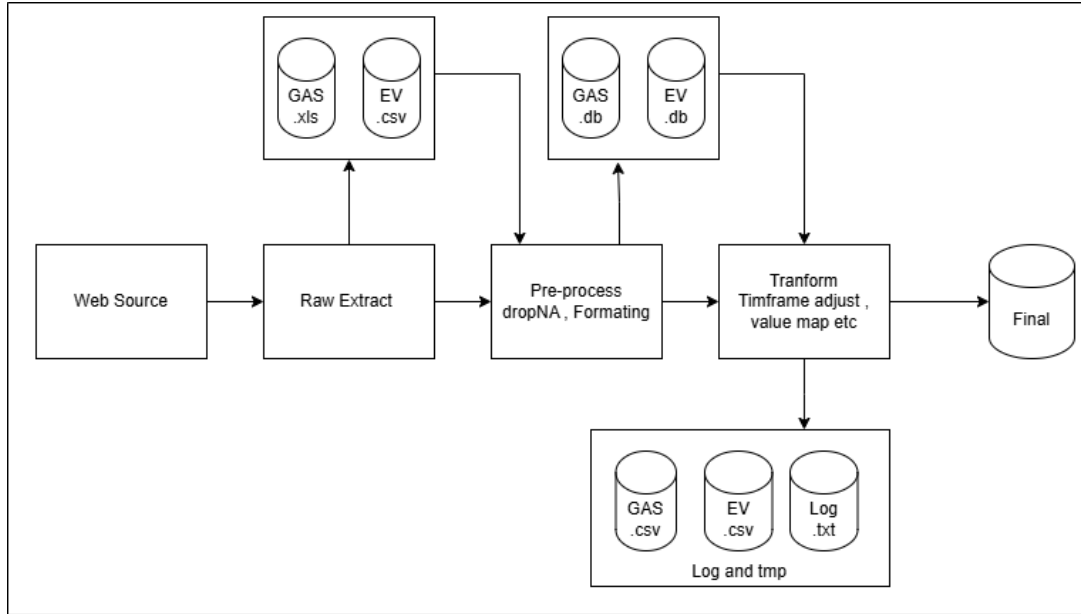
Figure 1: Overview of the Data Pipeline Workflow

4. Results and Limitations

4.1 Output Data

Format: Unified CSV and SQLite database. **Columns:** `timestamp`, `price`, `volume`.

4.2 Quality Improvements

- Addressed weekly inconsistencies in gas data.

- Aggregated EV registrations into uniform monthly intervals.

- Logged all inconsistencies for future reference.

4.3 Limitations

- Missing months are flagged, not interpolated.

- External factors (e.g., subsidies) are not considered.

5. Conclusion and Future Work

5.1 Achievements

- Built a reproducible pipeline to preprocess and merge datasets .

- Pipeline is kept as dynamic to handles any dataset (diff states dataset)

- Provided a transparent logging mechanism for inconsistencies.

5.2 Basic Analysis

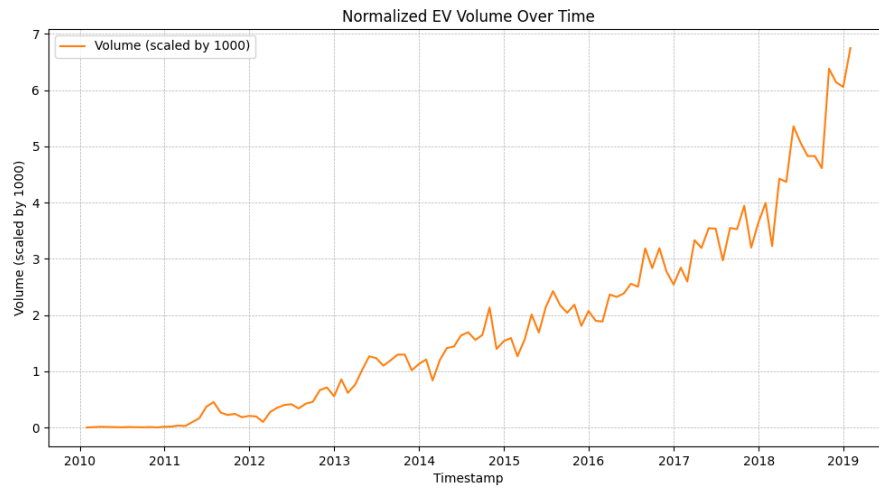- Have made a Basic analysis with some basic plotting to check data distribution

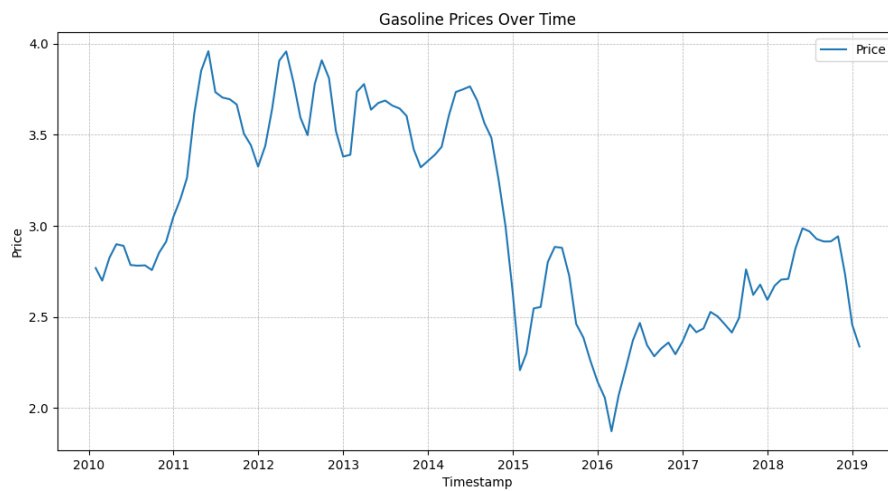Figure 2: EV data - sales in volume vs Time

Figure 3: EV data - sales in volume vs Time

5.3 References

- EV Dataset - atlasevhub.com

- Gas Dataset - eia.gov