

```
EV project - Data Cleaning & Pre-processing

In [12]: import pandas as pd

In [13]: df = pd.read_csv(r'C:\Users\Owli\OneDrive\Desktop\Electric_Vehicle_Population_Data.csv')

In [14]: print(df)

0      VIN (1-10)      County      City      State      Postal Code      \
1  1C4JXZP66P      Kitsap      Poulsbo      WA      98370.0      ...
2  1G1FXB8X8K      Snohomish      Lake Stevens      WA      98258.0
3  WBV1Z2C58F      King      Seattle      WA      98116.0
4  5YJ3E1B8XK      King      Seattle      WA      98178.0
... ..
233990 78AYGDEE4H      Pierce      Puyallup      WA      98374.0
233991 WBV9P2000W      Snohomish      Lake Stevens      WA      98258.0
233992 UN1AD0CF3B      Pierce      University Place      WA      98466.0
233993 5YJ3E1B2A8      Pierce      Puyallup      WA      98374.0
233994 WBV9P2C57E      King      Woodinville      WA      98072.0

... ..
0      Model Year      Make      Model      \
1      2023      JEEP      WRANGLER
2      2019      CHEVROLET      BOLT EV
3      2015      BMW      I3
4      2019      TESLA      MODEL 3
... ..
233990 2024      TESLA      MODEL Y
233991 2021      BMW      I3
233992 2011      HISSAN      LEAF
233993 2024      TESLA      MODEL 3
233994 2019      BMW      I3

... ..
0      Electric Vehicle Type      \
1      Plug-in Hybrid Electric Vehicle (PHEV)
2      Battery Electric Vehicle (BEV)
3      Battery Electric Vehicle (BEV)
4      Battery Electric Vehicle (BEV)
... ..
233990 Battery Electric Vehicle (BEV)
233991 Battery Electric Vehicle (BEV)
233992 Battery Electric Vehicle (BEV)
233993 Battery Electric Vehicle (BEV)
233994 Plug-in Hybrid Electric Vehicle (PHEV)

... ..
0      Clean Alternative Fuel Vehicle (CAFEV) Eligibility      Electric Range      \
1      Not eligible due to low battery range      21.0
2      Clean Alternative Fuel Vehicle Eligible      238.0
3      Clean Alternative Fuel Vehicle Eligible      81.0
4      Clean Alternative Fuel Vehicle Eligible      220.0
... ..
233990 Eligibility unknown as battery range has not b...      ...
233991 Eligibility unknown as battery range has not b...      0.0
233992 Clean Alternative Fuel Vehicle Eligible      73.0
233993 Eligibility unknown as battery range has not b...      0.0
233994 Clean Alternative Fuel Vehicle Eligible      126.0

... ..
0      Base MSRP      Legislative District      DOL Vehicle ID      \
1      0.0      23.0      258127145
2      0.0      44.0      4735426
3      0.0      37.0      477309682
4      0.0      15.0      258112970
... ..
233990 0.0      2.0      264662359
233991 0.0      44.0      157728188
233992 0.0      28.0      261730363
233993 0.0      25.0      215283487
233994 0.0      45.0      267288801

... ..
0      Vehicle Location      \
1      POINT (-122.64683 47.73689)
2      POINT (-122.06402 48.01497)
3      POINT (-122.23825 47.49461)
4      POINT (-120.53145 46.65405)
... ..
233990 POINT (-122.27575 47.13959)
233991 POINT (-122.06402 48.01497)
233992 POINT (-121.57756 47.23165)
233993 POINT (-122.27575 47.13959)
233994 POINT (-122.15545 47.75448)

... ..
0      Electric Utility      2020 Census Tract
1      PUGET SOUND ENERGY INC      5.303509e+10
2      PUGET SOUND ENERGY INC      5.304105e+10
3      CITY OF SEATTLE - (WA)CITY OF TACOMA - (WA)      5.303301e+10
4      CITY OF SEATTLE - (WA)CITY OF TACOMA - (WA)      5.303301e+10
... ..
233990 PUGET SOUND ENERGY INCICITY OF TACOMA - (WA)      5.305107e+10
233991 PUGET SOUND ENERGY INC      5.304105e+10
233992 BORNEVILLE POWER ADMINISTRATIONICITY OF TACOMA -...      5.305107e+10
233993 PUGET SOUND ENERGY INCICITY OF TACOMA - (WA)      5.305107e+10
233994 PUGET SOUND ENERGY INCICITY OF TACOMA - (WA)      5.303301e+10

(233995 rows x 17 columns)

In [15]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 233995 entries, 0 to 233994
Data columns (total 17 columns):
 #   Column      Non-Null Count  Dtype
---  ---
0   VIN (1-10)      233995 non-null    object
1   County      233992 non-null    object
2   City      233992 non-null    object
3   State      233992 non-null    object
4   Postal Code      233992 non-null    float64
5   Model Year      233995 non-null    int64
6   Make      233995 non-null    object
7   Model      233995 non-null    object
8   Electric Vehicle Type      233995 non-null    object
9   Clean Alternative Fuel Vehicle (CAFEV) Eligibility      233995 non-null    object
10  Electric Range      233977 non-null    float64
11  Base MSRP      233977 non-null    float64
12  Legislative District      233992 non-null    float64
13  DOL Vehicle ID      233992 non-null    float64
14  Vehicle Location      233995 non-null    object
15  Electric Utility      233995 non-null    object
16  2020 Census Tract      233992 non-null    float64
dtypes: float64(5), int64(2), object(10)
memory usage: 29.1+ MB

Out[16]:

VIN (1-10)      County      City      State      Postal Code      Model Year      Make      Model      Electric Vehicle Type      Clean Alternative Fuel Vehicle (CAFEV) Eligibility      Electric Range      Base MSRP      Legislative District      DOL Vehicle ID      Vehicle Location      Electric Utility      2020 Census Tract

0      1C4JXZP66P      Kitsap      Poulsbo      WA      98370.0      2023      JEEP      WRANGLER      Plug-in Hybrid Electric Vehicle (PHEV)      Not eligible due to low battery range      21.0      0.0      23.0      258127145      POINT (-122.64681 47.73689)      PUGET SOUND ENERGY INC      5.303509e+10

1      1G1FXB8X8K      Snohomish      Lake Stevens      WA      98258.0      2019      CHEVROLET      BOLT EV      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      238.0      0.0      44.0      4735426      POINT (-122.06402 48.01497)      PUGET SOUND ENERGY INC      5.306105e+10

2      WBV1Z2C58F      King      Seattle      WA      98116.0      2015      BMW      I3      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      81.0      0.0      34.0      272697666      POINT (-122.01407 47.57894)      CITY OF SEATTLE - (WA)CITY OF TACOMA - (WA)      5.303301e+10

3      5YJ3E1B8XK      King      Seattle      WA      98178.0      2019      TESLA      MODEL 3      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      220.0      0.0      37.0      477309682      POINT (-122.23825 47.49461)      CITY OF SEATTLE - (WA)CITY OF TACOMA - (WA)      5.303301e+10

4      5YJSA1V24F      Yakima      Selah      WA      98942.0      2015      TESLA      MODEL S      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      208.0      0.0      15.0      258112970      POINT (-120.53145 46.65405)      PACIFICCORP      5.307700e+10

In [17]: df.describe()

Out[17]:

Postal Code      Model Year      Electric Range      Base MSRP      Legislative District      DOL Vehicle ID      2020 Census Tract
count      233992.000000      233995.000000      233977.000000      233977.000000      233521.000000      2.239950e+05      2.239920e+05
mean      98176.491165      2021.264408      47.736187      829.894386      28.87361      2.329328e+08      5.287997e+10
std      2544.240509      2.889676      84.967140      7372.508049      14.911023      6.884329e+07      1.531491e+09
min      1731.000000      1999.000000      0.000000      0.000000      1.000000      4.385000e+03      1.001020e+09
25%      98062.000000      2020.000000      0.000000      0.000000      17.000000      2.008002e+08      5.303301e+10
50%      98126.000000      2022.000000      0.000000      0.000000      32.000000      2.482992e+08      5.303303e+10
75%      98374.000000      2023.000000      39.000000      0.000000      42.000000      2.673973e+08      5.305307e+10
max      99577.000000      2025.000000      337.000000      845000.000000      49.000000      4.792549e+08      5.602100e+10

In [18]: df.shape

Out[18]: (233995, 17)

In [19]: print(df.isnull().sum())

VIN (1-10)      0
County      3
City      3
State      0
Postal Code      3
Model Year      0
Make      0
Model      0
Electric Vehicle Type      0
Clean Alternative Fuel Vehicle (CAFEV) Eligibility      0
Electric Range      18
Base MSRP      8
Legislative District      474
DOL Vehicle ID      0
Vehicle Location      10
Electric Utility      3
2020 Census Tract      3
dtype: int64

In [100]: missing_values = df.isnull().sum()
total_values = len(df)
missing_percentage = (missing_values / total_values) * 100
print(missing_percentage)

VIN (1-10)      0.000000
County      0.001339
City      0.001339
State      0.000000
Postal Code      0.001339
Model Year      0.000000
Make      0.000000
Model      0.000000
Electric Vehicle Type      0.000000
Clean Alternative Fuel Vehicle (CAFEV) Eligibility      0.000000
Electric Range      0.008036
Base MSRP      0.008036
Legislative District      0.211612
DOL Vehicle ID      0.000000
Vehicle Location      0.004669
Electric Utility      0.001339
2020 Census Tract      0.001339
dtype: float64

In [112]: # remove duplicates
print(df.duplicated().sum())
df.drop_duplicates(inplace=True)
Duplicate rows: 0

In [122]: print(df.dtypes)

VIN (1-10)      object
County      object
City      object
State      object
Postal Code      float64
Model Year      int64
Make      object
Model      object
Electric Vehicle Type      object
Clean Alternative Fuel Vehicle (CAFEV) Eligibility      object
Electric Range      float64
Base MSRP      float64
Legislative District      float64
DOL Vehicle ID      int64
Vehicle Location      object
Electric Utility      object
2020 Census Tract      float64
dtype: object

In [133]: df['Model Year'] = df['Model Year'].astype(int)
df['Postal Code'] = df['Postal Code'].astype(str) #conversion of data types

In [144]: print(df.dtypes) #verification

VIN (1-10)      object
County      object
City      object
State      object
Postal Code      object
Model Year      int32
Make      object
Model      object
Electric Vehicle Type      object
Clean Alternative Fuel Vehicle (CAFEV) Eligibility      object
Electric Range      float64
Base MSRP      float64
Legislative District      float64
DOL Vehicle ID      int64
Vehicle Location      object
Electric Utility      object
2020 Census Tract      float64
dtype: object

In [153]: df.head()

Out[153]:

VIN (1-10)      County      City      State      Postal Code      Model Year      Make      Model      Electric Vehicle Type      Clean Alternative Fuel Vehicle (CAFEV) Eligibility      Electric Range      Base MSRP      Legislative District      DOL Vehicle ID      Vehicle Location      Electric Utility      2020 Census Tract

0      1C4JXZP66P      Kitsap      Poulsbo      WA      98370.0      2023      JEEP      WRANGLER      Plug-in Hybrid Electric Vehicle (PHEV)      Not eligible due to low battery range      21.0      0.0      23.0      258127145      POINT (-122.64681 47.73689)      PUGET SOUND ENERGY INC      5.303509e+10

1      1G1FXB8X8K      Snohomish      Lake Stevens      WA      98258.0      2019      CHEVROLET      BOLT EV      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      238.0      0.0      44.0      4735426      POINT (-122.06402 48.01497)      PUGET SOUND ENERGY INC      5.306105e+10

2      WBV1Z2C58F      King      Seattle      WA      98116.0      2015      BMW      I3      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      81.0      0.0      34.0      272697666      POINT (-122.01407 47.57894)      CITY OF SEATTLE - (WA)CITY OF TACOMA - (WA)      5.303301e+10

3      5YJ3E1B8XK      King      Seattle      WA      98178.0      2019      TESLA      MODEL 3      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      220.0      0.0      37.0      477309682      POINT (-122.23825 47.49461)      CITY OF SEATTLE - (WA)CITY OF TACOMA - (WA)      5.303301e+10

4      5YJSA1V24F      Yakima      Selah      WA      98942.0      2015      TESLA      MODEL S      Battery Electric Vehicle (BEV)      Clean Alternative Fuel Vehicle Eligible      208.0      0.0      15.0      258112970      POINT (-120.53145 46.65405)      PACIFICCORP      5.307700e+10

In [161]: #standardizing categorical data
df['Make'] = df['Make'].str.strip().str.title()
df['State'] = df['State'].str.upper()

In [171]: df.to_csv('cleaned_ev_data.csv', index=False)

EV Project Step 2 - Exploratory Data Analysis

EV count by state

In [23]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
top_states = df['State'].value_counts().head(10)
top_states_df = top_states.reset_index()
top_states_df.columns = ['State', 'EV Count']

plt.figure(figsize=(12, 6))
sns.barplot(x=top_states, y='EV Count', hue=top_states, data=top_states_df, palette='viridis', legend=False)
plt.title('Top 10 States with Highest EV Adoption', fontsize=14)
plt.xlabel('State', fontsize=12)
plt.ylabel('Number of EVs', fontsize=10)
plt.xticks(rotation=45, fontsize=10)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()

Top 10 States with Highest EV Adoption

Number of EVs

200000
150000
100000
50000
0

WA      CT      VA      MD      TX      NC      CO      OR      IL      IN

Insight | Washington (WA) dominates with 223,521 EVs, while other states have significantly lower numbers (e.g., CA: 114, VA: 59). This large difference causes WA to overshadow the rest in the bar chart, making them barely visible.

In [30]: # Correlation heatmap for numerical features
plt.figure(figsize=(8, 5))
sns.heatmap(df[['Base MSRP', 'Electric Range']].corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Between EV Price and Range')
plt.show()

Correlation Between EV Price and Range

Base MSRP      Electric Range
Base MSRP      1      0.12
Electric Range      0.12      1

In [34]: import seaborn as sns
import matplotlib.pyplot as plt
sns.scatterplot(x=df['Base MSRP'], y=df['Electric Range'])
plt.title('Scatter Plot of Base MSRP vs Electric Range')
plt.show()

Scatter Plot of Base MSRP vs Electric Range

Electric Range

350
300
250
200
150
100
50
0

0      200000      400000      600000      800000

Base MSRP

In [42]: top_brands = df['Make'].value_counts().head(10)
print(top_brands)
top_brands_df = top_brands.reset_index()
top_brands_df.columns = ['Brand', 'EV Count']

plt.figure(figsize=(10, 5))
sns.barplot(x=top_brands, y='EV Count', data=top_brands_df, hue='Brand', palette='viridis', legend=False)
plt.title('Top 10 Most Popular EV Brands', fontsize=14)
plt.xlabel('Brand', fontsize=12)
plt.ylabel('Number of EVs', fontsize=10)
plt.xticks(rotation=45, fontsize=10)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()

Make      Tesla      96180
Chevrolet      16405
Nissan      15329
Ford      11930
Audi      10760
BMW      9171
Toyota      8955
Hyundai      6878
Rivian      6236
Jeep      5773
Name: count, dtype: int64

Top 10 Most Popular EV Brands

Number of EVs

100000
80000
60000
40000
20000
0

Tesla      Chevrolet      Nissan      Ford      Audi      BMW      Toyota      Hyundai      Rivian      Jeep

Brand

In [44]: # Group data by model year and find average electric range
avg_range_yearly = df.groupby('Model Year')['Electric Range'].mean()

# Plot
plt.figure(figsize=(10, 5))
avg_range_yearly.plot(marker='o', linestyle='--', color='b')
plt.title('EV Electric Range Evolution Over Time')
plt.xlabel('Model Year')
plt.ylabel('Average Electric Range (miles)')
plt.grid(True)
plt.show()

EV Electric Range Evolution Over Time

Average Electric Range (miles)

200
150
100
50
0

2000      2005      2010      2015      2020      2025

Model Year
```