## data cleaning project

## December 17, 2023

```
[1]: import pandas as pd
     import numpy as np
[2]: df=pd.read_csv('EV_Registration_Dataset.csv')
     df.head
[2]: <bound method NDFrame.head of
                                           Identifier
                                                                 City Postal Code
     Model Year
                       Make
                               Tacoma
     0
           KL8CL6S05E
                                              98407
                                                           2014
                                                                  CHEVROLET
                       Port Townsend
     1
           JTDKARFPOH
                                              98368
                                                           2017
                                                                     TOYOTA
     2
           JHMZC5F34J
                               Vashon
                                              98070
                                                           2018
                                                                      HONDA
     3
           KMHC65LD0K
                           Ridgefield
                                                           2019
                                                                    HYUNDAI
                                              98642
     4
                            Bremerton
           1N4AZ1CP6J
                                              98337
                                                           2018
                                                                     NISSAN
                                                             •••
     4222 5YJSA1E22L
                              Belfair
                                              98528
                                                           2020
                                                                      TESLA
     4223
           7SAYGDEEXP
                             Kirkland
                                              98033
                                                           2023
                                                                      TESLA
     4224 WPOAA2Y10N
                              Redmond
                                                           2022
                                                                    PORSCHE
                                              98053
     4225
           5YJXCBE27J
                              Seattle
                                                                      TESLA
                                              98102
                                                           2018
     4226
           JHMZC5F32J
                              Poulsbo
                                              98370
                                                           2018
                                                                      HONDA
                 Model
                                           Electric Vehicle Type
     0
                 SPARK
                                 Battery Electric Vehicle (BEV)
                        Plug-in Hybrid Electric Vehicle (PHEV)
     1
           PRIUS PRIME
     2
               CLARITY
                         Plug-in Hybrid Electric Vehicle (PHEV)
     3
                         Plug-in Hybrid Electric Vehicle (PHEV)
                 IONIQ
                                 Battery Electric Vehicle (BEV)
     4
                  LEAF
     4222
               MODEL S
                                 Battery Electric Vehicle (BEV)
     4223
               MODEL Y
                                 Battery Electric Vehicle (BEV)
     4224
                                 Battery Electric Vehicle (BEV)
                TAYCAN
     4225
               MODEL X
                                 Battery Electric Vehicle (BEV)
     4226
               CLARITY
                        Plug-in Hybrid Electric Vehicle (PHEV)
           Clean Alternative Fuel Vehicle (CAFV) Eligibility Electric Range
     0
                     Clean Alternative Fuel Vehicle Eligible
                                                                           82.0
     1
                       Not eligible due to low battery range
                                                                           25.0
     2
                     Clean Alternative Fuel Vehicle Eligible
                                                                           47.0
```

3	Not eligible due to low battery range	29.0				
4	Clean Alternative Fuel Vehicle Eligible	151.0				
4222 4223 4224 4225 4226	Clean Alternative Fuel Vehicle Eligible Eligibility unknown as battery range has not b Eligibility unknown as battery range has not b Clean Alternative Fuel Vehicle Eligible NaN	330.0 0.0 0.0 238.0 NaN				
	Base MSRP Legislative District Vehicle ID \					
0	0.0 27.0 259013639.0					
1	0.0 24.0 220589904.0					
2	0.0 34.0 109177015.0					
3	0.0 18.0 196216282.0					
4	0.0 26.0 187956821.0					
4222	0.0 35.0 6123659.0					
4223	0.0 48.0 225767643.0					
4224	0.0 45.0 219090882.0					
4225	0.0 43.0 116650978.0 NaN NaN NaN					
4226	NaN NaN NaN					
0 1 2 3	Vehicle Location \ POINT (-122.5113356 47.29238280000004)  POINT (-122.76441969999996 48.11958740000006)  POINT (-122.458675199999996 47.448468700000035)  POINT (-122.74348499999996 45.81886000000003)					
4	POINT (-122.63590499999998 47.570970000000045)					
4222 4223 4224 4225 4226	POINT (-122.85516469999999 47.44957850000003) POINT (-122.20263999999999 47.6785000000004) POINT (-122.02219999999994 47.69557000000003) POINT (-122.32234499999998 47.639830000000075) NaN					
0 1 2 3 4	Electric Utility BONNEVILLE POWER ADMINISTRATION   CITY OF TACOM BONNEVILLE POWER ADMINISTRATION   PUGET SOUND E PUGET SOUND ENERGY INC   CITY OF TACOMA - (WA) BONNEVILLE POWER ADMINISTRATION   PUD NO 1 OF C PUGET SOUND ENERGY INC	2020 Census Tract 5.305306e+10 5.303195e+10 5.303303e+10 5.301104e+10 5.303508e+10				
4222 4223 4224 4225 4226	BONNEVILLE POWER ADMINISTRATION   CITY OF TACOM  PUGET SOUND ENERGY INC   CITY OF TACOMA - (WA)  PUGET SOUND ENERGY INC   CITY OF TACOMA - (WA)  CITY OF SEATTLE - (WA)   CITY OF TACOMA - (WA)  Nan	5.304596e+10 5.303302e+10 5.303303e+10 5.303301e+10 NaN				

[5]: df.shape

[5]: (4227, 15)

[6]: df.describe

[0].	ur . uo	501150					
[6]:	<box< th=""><th>d method NDF</th><th>rame.describe of</th><th>Identifie</th><th>r</th><th>City Posta</th><th>ıl Code</th></box<>	d method NDF	rame.describe of	Identifie	r	City Posta	ıl Code
	Model	Year	Make \			·	
	0	KL8CL6S05E	Tacoma	98407	2014	CHEVROLET	
	1	JTDKARFPOH	Port Townsend	98368	2017	TOYOTA	
	2	JHMZC5F34J	Vashon	98070	2018	HONDA	
	3	KMHC65LD0K	Ridgefield	98642	2019	HYUNDAI	
	4	1N4AZ1CP6J	Bremerton	98337	2018	NISSAN	
		<b></b>					
	4222		Belfair	98528	2020	TESLA	
	4223	7SAYGDEEXP	Kirkland	98033	2023	TESLA	
	4224	WPOAA2Y10N	Redmond	98053	2022	PORSCHE	
	4225		Seattle	98102	2018	TESLA	
	4226	JHMZC5F32J	Poulsbo	98370	2018	HONDA	
		Model		Electric Vehi	cle Type	\	
	0	SPARK		Electric Vehic			
	1	PRIUS PRIME	v				
	2	CLARITY	•				
	3	IONIQ	· ·				
	4	LEAF	•	Electric Vehic			
	•••	•••			•••		
	4222	MODEL S	Battery	Electric Vehic	le (BEV)		
	4223	MODEL Y	Battery	Electric Vehic	le (BEV)		
	4224	TAYCAN	Battery	Electric Vehic	le (BEV)		
	4225	MODEL X	Battery	Electric Vehic	le (BEV)		
	4226	CLARITY	Plug-in Hybrid E	lectric Vehicl	e (PHEV)		
		Clean Alter	native Fuel Vehicl	e (CAFV) Eligi	bilitv :	Electric Range	\
	0		lean Alternative F	_	•	82.0	•
	1		Not eligible due		•	25.0	
	2	C	lean Alternative F	•	•	47.0	
	3		Not eligible due		•	29.0	
	4	C	lean Alternative F			151.0	
						•••	
	4222	C	lean Alternative F	uel Vehicle El	igible	330.0	
	4223		unknown as batter		_	0.0	
	4224	-	unknown as batter	-		0.0	
	4225	-	lean Alternative F	-		238.0	

4226 NaNNaNBase MSRP Legislative District Vehicle ID 0 0.0 27.0 259013639.0 0.0 24.0 220589904.0 1 0.0 2 34.0 109177015.0 0.0 18.0 196216282.0 3 0.0 4 26.0 187956821.0 4222 0.0 6123659.0 35.0 4223 0.0 48.0 225767643.0 4224 0.0 45.0 219090882.0 4225 0.0 43.0 116650978.0 4226 NaN NaN NaN Vehicle Location POINT (-122.5113356 47.29238280000004) 0 1 POINT (-122.76441969999996 48.11958740000006) POINT (-122.45867519999996 47.448468700000035) 3 POINT (-122.74348499999996 45.81886000000003) POINT (-122.63590499999998 47.570970000000045) 4 4222 POINT (-122.85516469999999 47.44957850000003) 4223 POINT (-122.20263999999997 47.67850000000004) 4224 POINT (-122.02219999999994 47.69557000000003) 4225 POINT (-122.32234499999998 47.639830000000075) 4226 NaN Electric Utility 2020 Census Tract 0 BONNEVILLE POWER ADMINISTRATION | CITY OF TACOM ... 5.305306e+10 1 BONNEVILLE POWER ADMINISTRATION | PUGET SOUND E... 5.303195e+10 2 PUGET SOUND ENERGY INC | CITY OF TACOMA - (WA) 5.303303e+10 3 BONNEVILLE POWER ADMINISTRATION | PUD NO 1 OF C... 5.301104e+10 4 PUGET SOUND ENERGY INC 5.303508e+10 BONNEVILLE POWER ADMINISTRATION | CITY OF TACOM... 5.304596e+10 PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA) 4223 5.303302e+10 4224 PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA) 5.303303e+10 4225 CITY OF SEATTLE - (WA) | CITY OF TACOMA - (WA) 5.303301e+10 4226 NaN NaN

[4227 rows x 15 columns]>

[10]: df.isnull

[10]: <bound method DataFrame.isnull of Identifier City Postal Code Model Year Make \

```
0
      KL8CL6S05E
                          Tacoma
                                         98407
                                                       2014
                                                             CHEVROLET
1
      JTDKARFPOH
                   Port Townsend
                                         98368
                                                       2017
                                                                 TOYOTA
2
      JHMZC5F34J
                          Vashon
                                         98070
                                                       2018
                                                                  HONDA
3
      KMHC65LD0K
                      Ridgefield
                                         98642
                                                       2019
                                                                HYUNDAI
4
      1N4AZ1CP6J
                       Bremerton
                                                       2018
                                                                NISSAN
                                         98337
4222
      5YJSA1E22L
                                                       2020
                                                                  TESLA
                         Belfair
                                         98528
4223
      7SAYGDEEXP
                        Kirkland
                                         98033
                                                       2023
                                                                  TESLA
4224
                         Redmond
                                                                PORSCHE
      WPOAA2Y10N
                                         98053
                                                       2022
4225
      5YJXCBE27J
                         Seattle
                                                                  TESLA
                                         98102
                                                       2018
4226
                         Poulsbo
                                                                  HONDA
      JHMZC5F32J
                                         98370
                                                       2018
            Model
                                      Electric Vehicle Type
0
            SPARK
                            Battery Electric Vehicle (BEV)
                    Plug-in Hybrid Electric Vehicle (PHEV)
1
      PRIUS PRIME
2
          CLARITY
                    Plug-in Hybrid Electric Vehicle (PHEV)
3
                    Plug-in Hybrid Electric Vehicle (PHEV)
            IONIQ
4
             LEAF
                            Battery Electric Vehicle (BEV)
4222
          MODEL S
                            Battery Electric Vehicle (BEV)
4223
          MODEL Y
                            Battery Electric Vehicle (BEV)
4224
                            Battery Electric Vehicle (BEV)
           TAYCAN
4225
                            Battery Electric Vehicle (BEV)
          MODEL X
4226
                    Plug-in Hybrid Electric Vehicle (PHEV)
          CLARITY
      Clean Alternative Fuel Vehicle (CAFV) Eligibility
                                                            Electric Range
                 Clean Alternative Fuel Vehicle Eligible
                                                                       82.0
0
1
                   Not eligible due to low battery range
                                                                       25.0
2
                 Clean Alternative Fuel Vehicle Eligible
                                                                       47.0
3
                   Not eligible due to low battery range
                                                                       29.0
4
                 Clean Alternative Fuel Vehicle Eligible
                                                                      151.0
4222
                 Clean Alternative Fuel Vehicle Eligible
                                                                      330.0
4223
      Eligibility unknown as battery range has not b...
                                                                      0.0
4224
      Eligibility unknown as battery range has not b...
                                                                      0.0
4225
                 Clean Alternative Fuel Vehicle Eligible
                                                                      238.0
4226
                                                       NaN
                                                                        NaN
      Base MSRP
                  Legislative District
                                          Vehicle ID
0
            0.0
                                   27.0
                                         259013639.0
1
            0.0
                                   24.0
                                         220589904.0
2
            0.0
                                   34.0
                                         109177015.0
3
            0.0
                                   18.0
                                         196216282.0
4
            0.0
                                   26.0
                                         187956821.0
4222
            0.0
                                   35.0
                                           6123659.0
4223
            0.0
                                   48.0
                                         225767643.0
```

```
0.0
      4224
                                       45.0 219090882.0
      4225
                  0.0
                                       43.0
                                             116650978.0
      4226
                  NaN
                                        NaN
                                                      NaN
                                          Vehicle Location
                    POINT (-122.5113356 47.29238280000004)
      0
             POINT (-122.76441969999996 48.11958740000006)
      1
      2
            POINT (-122.45867519999996 47.448468700000035)
             POINT (-122.74348499999996 45.81886000000003)
      3
            POINT (-122.63590499999998 47.570970000000045)
      4
      4222
             POINT (-122.85516469999999 47.44957850000003)
      4223
             POINT (-122.20263999999997 47.67850000000004)
      4224
             POINT (-122.02219999999994 47.69557000000003)
      4225
           POINT (-122.32234499999998 47.639830000000075)
      4226
                                                        NaN
                                              Electric Utility 2020 Census Tract
      0
            BONNEVILLE POWER ADMINISTRATION | CITY OF TACOM ...
                                                                   5.305306e+10
            BONNEVILLE POWER ADMINISTRATION | PUGET SOUND E...
                                                                   5.303195e+10
      1
      2
                PUGET SOUND ENERGY INC | CITY OF TACOMA - (WA)
                                                                     5.303303e+10
      3
            BONNEVILLE POWER ADMINISTRATION | PUD NO 1 OF C...
                                                                   5.301104e+10
      4
                                       PUGET SOUND ENERGY INC
                                                                     5.303508e+10
      4222
            BONNEVILLE POWER ADMINISTRATION | CITY OF TACOM ...
                                                                   5.304596e+10
      4223
                PUGET SOUND ENERGY INCICITY OF TACOMA - (WA)
                                                                     5.303302e+10
                PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)
                                                                     5.303303e+10
      4224
      4225
                 CITY OF SEATTLE - (WA) | CITY OF TACOMA - (WA)
                                                                     5.303301e+10
      4226
                                                           NaN
                                                                              NaN
      [4227 rows x 15 columns]>
[11]: df.isna()
[11]:
            Identifier
                        City Postal Code Model Year
                                                        Make Model
      0
                 False False
                                     False
                                                  False False False
                 False False
                                     False
                                                  False False False
      1
                 False False
      2
                                     False
                                                  False
                                                        False False
      3
                 False False
                                     False
                                                  False False False
      4
                 False False
                                     False
                                                  False
                                                        False False
      4222
                 False False
                                     False
                                                  False
                                                        False False
      4223
                 False False
                                     False
                                                  False
                                                        False False
      4224
                 False False
                                     False
                                                  False False False
      4225
                 False False
                                     False
                                                  False False False
      4226
                 False False
                                     False
                                                  False False False
```

```
Electric Vehicle Type \
0
                       False
1
                       False
2
                       False
3
                       False
4
                       False
4222
                       False
4223
                       False
4224
                       False
4225
                       False
4226
                       False
      Clean Alternative Fuel Vehicle (CAFV) Eligibility Electric Range \
0
                                                     False
                                                                      False
1
                                                     False
                                                                      False
2
                                                     False
                                                                      False
3
                                                     False
                                                                      False
4
                                                     False
                                                                      False
4222
                                                     False
                                                                      False
4223
                                                     False
                                                                      False
4224
                                                     False
                                                                      False
4225
                                                     False
                                                                      False
4226
                                                      True
                                                                       True
      Base MSRP Legislative District Vehicle ID Vehicle Location \
0
          False
                                  False
                                               False
                                                                  False
1
          False
                                  False
                                               False
                                                                  False
2
          False
                                  False
                                                                  False
                                               False
3
          False
                                  False
                                                                  False
                                               False
4
          False
                                  False
                                               False
                                                                  False
4222
          False
                                  False
                                               False
                                                                  False
4223
          False
                                  False
                                               False
                                                                  False
4224
          False
                                  False
                                               False
                                                                  False
4225
          False
                                  False
                                               False
                                                                  False
4226
           True
                                   True
                                                True
                                                                   True
      Electric Utility
                         2020 Census Tract
0
                  False
                                      False
                  False
                                      False
1
2
                  False
                                      False
3
                  False
                                      False
4
                                      False
                  False
4222
                  False
                                      False
```

	4220	raise	
	4226 True	True	
	[4227 rows x 15 columns]		
[12]:	df.isna().any()		
[12]:	Identifier		False
	City		False
	Postal Code		False
	Model Year		False
	Make		False
	Model		True
	Electric Vehicle Type		False
	Clean Alternative Fuel Ve	hicle (CAFV) Eligibility	True
	Electric Range		True
	Base MSRP		True
	Legislative District		True
	Vehicle ID		True
	Vehicle Location		True
	Electric Utility		True
	2020 Census Tract		True
	dtype: bool		
[13]:	df.isna().sum()		
[13]:	Identifier		0
	City		0
	Postal Code		0
	Model Year		0
	Make		0
	Model		7
	Electric Vehicle Type		0
	Clean Alternative Fuel Ve	chicle (CAFV) Eligibility	1
	Electric Range		1
	Base MSRP		1
	Legislative District		5
	Vehicle ID		1
	Vehicle Location		1
	Electric Utility		1
	2020 Census Tract		1
	dtype: int64		
[15]:	df.isna().any().sum()		

False

False

False

4223

4224

4225

False

False

False

## [15]: 9

```
[17]: df.duplicated()
[17]: 0
              False
      1
              False
      2
              False
      3
              False
              False
      4
      4222
              False
      4223
              False
      4224
              False
      4225
              False
      4226
              False
     Length: 4227, dtype: bool
[]:
```