In [63]:

```python
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
plt.style.use('fivethirtyeight')
import folium
import plotly.express as px
```

Exploratory Data Analysis

In [121…

```python
# Exporting data into the Notebook.

ev = pd.read_csv('Electric_Vehicle_Population_Data.csv')
```

In [122…

```python
ev
```

Out[122]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	Vehic
0	5YJXCAE26J	Yakima	Yakima	WA	98908.0	2018	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	238	0	14.0	14115
1	JHMZC5F37M	Kitsap	Poulsbo	WA	98370.0	2021	HONDA	CLARITY	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	47	0	23.0	17156
2	5YJ3E1EB0K	King	Seattle	WA	98199.0	2019	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	220	0	36.0	942
3	1N4AZ0CP5D	King	Seattle	WA	98119.0	2013	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	75	0	36.0	21180
4	5YJSA1E21H	Thurston	Lacey	WA	98516.0	2017	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	210	0	22.0	18581
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
130438	7SAYGDEE6P	Pierce	Gig Harbor	WA	98335.0	2023	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	26.0	23113
130439	1N4BZ1CV7N	Pierce	Tacoma	WA	98408.0	2022	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	29.0	18581
130440	5YJYGDEE8M	King	Seattle	WA	98109.0	2021	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	36.0	17654
130441	5YJXCBE22L	Island	Camano Island	WA	98282.0	2020	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	293	0	10.0	10283
130442	5YJ3E1EA5M	Pierce	Puyallup	WA	98375.0	2021	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	2.0	18047

130443 rows × 17 columns

In [4]:

```python
for i in ev.columns:
```

```
    print(i)
```

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

In [5]:
```
ev.shape
```

Out[5]:
```
(130443, 17)
```

In [6]:
```
# Total Information about the Dataset

ev_info = pd.DataFrame(columns = ['Name of col.', 'no. of Null Val.', 'DataType', 'Unique Value'])

for i in range(0, len(ev.columns)):
    ev_info.loc[i] = [ev.columns[i],
                      ev[ev.columns[i]].isnull().sum(),
                      ev[ev.columns[i]].dtype,
                      ev[ev.columns[i]].nunique()]
ev_info
```

Out[6]:

	Name of col.	no. of Null Val.	DataType	Unique Value
0	VIN (1-10)	0	object	8827
1	County	3	object	166
2	City	3	object	656
3	State	0	object	46
4	Postal Code	3	float64	787
5	Model Year	0	int64	21
6	Make	0	object	35
7	Model	222	object	121
8	Electric Vehicle Type	0	object	2
9	Clean Alternative Fuel Vehicle (CAFV) Eligibility	0	object	3
10	Electric Range	0	int64	102
11	Base MSRP	0	int64	31
12	Legislative District	305	float64	49
13	DOL Vehicle ID	0	int64	130443
14	Vehicle Location	33	object	773
15	Electric Utility	3	object	75
16	2020 Census Tract	3	float64	2042

In [7]:
```
ev.drop(columns = ['2020 Census Tract', 'DOL Vehicle ID'], inplace= True)
```

Analysis: 1.The Dataset contains 130443 rows and 17 columns 2.The columns contains geographical as well as various model details about the Vehicles registered under Washingtion DOL Department 3.DOL Vehicle ID is the unique identifier for our data 4.VIN (1-10) column is not unique which might seem counterintutive but in reality it is truncated and the first 9 digits explain world manufacturer info and vehicle descriptor section and the 10th digit encodes the model year. 5.We also notice few missing values. The model and legislative distict columns in particular have quite a few of them and we will look to impute it or drop later. 6.We also drop DOL Vehicle ID and 2020 Census Tract column as it is not relevant to our anaylsis.

Managing Missing Data

In [8]:
```
ev[ev['City'].isna()]
```

Out[8]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	Vehicle Location
102	1N4AZ0CP1D	NaN	NaN	AP	NaN	2013	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	75	0	NaN	NaN
103	5YJ3E1EA5K	NaN	NaN	BC	NaN	2019	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	220	0	NaN	NaN
127670	5YJRE11B48	NaN	NaN	BC	NaN	2008	TESLA	ROADSTER	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	220	98950	NaN	NaN

In [9]:
```python
ev.dropna(subset = ['City'], inplace=True)
```

In [10]:
```python
ev[ev['Model'].isna()]['VIN (1-10)'].apply(lambda x: x[2 : 6]).unique()
# applies a lambda function to each value in the selected column. The lambda function takes a single argument,
#   x, and returns the substring of x starting at index 2 and ending at index 6.
```

Out[10]:
```
array(['4ED3'], dtype=object)
```

In [11]:
```python
# Analyzing the EV with same model code

subset_4ED3 = ev[ev['VIN (1-10)'].str.contains('4ED3')]
```

In [12]:
```python
subset_4ED3[['Model Year', 'Make', 'Model', 'Electric Vehicle Type']]
```

Out[12]:

	Model Year	Make	Model	Electric Vehicle Type
533	2023	VOLVO	NaN	Battery Electric Vehicle (BEV)
578	2022	VOLVO	XC40	Battery Electric Vehicle (BEV)
1023	2022	VOLVO	XC40	Battery Electric Vehicle (BEV)
1070	2021	VOLVO	XC40	Battery Electric Vehicle (BEV)
1167	2022	VOLVO	XC40	Battery Electric Vehicle (BEV)
...	...	...	...	...
129509	2021	VOLVO	XC40	Battery Electric Vehicle (BEV)
129597	2023	VOLVO	C40	Battery Electric Vehicle (BEV)
129736	2022	VOLVO	XC40	Battery Electric Vehicle (BEV)
130238	2021	VOLVO	XC40	Battery Electric Vehicle (BEV)
130379	2023	VOLVO	C40	Battery Electric Vehicle (BEV)

963 rows × 4 columns

In [13]:
```python
#The missing values are of Volvo which have launches two models XC40 and C40

print(subset_4ED3['Make'].unique())
print(subset_4ED3['Model'].unique())
```

```
['VOLVO']
[nan 'XC40' 'C40']
```

In [14]:
```python
#All the missing values belong to the year 2023

subset_4ED3[subset_4ED3['Model'].isna()]['Model Year'].value_counts()
```

Out[14]:
```
2023    222
Name: Model Year, dtype: int64
```

In [15]:
```python
#Analyse the year associated with the model variant and impute accordingly

subset_4ED3.groupby(['Model'])['Model Year'].value_counts()
```

Out[15]:
```
Model  Model Year
C40    2022          128
       2023          118
XC40   2021          247
       2022          242
       2023            6
Name: Model Year, dtype: int64
```

In [16]:
```
ev.fillna({'Model':'C40'},inplace = True)
```

In [17]:
```
ev.head(3)
```

Out[17]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	Vehicle Location	
0	5YJXCAE26J	Yakima	Yakima	WA	98908.0	2018	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	238	0	14.0	POINT (-120.56916 46.58514)	P.
1	JHMZC5F37M	Kitsap	Poulsbo	WA	98370.0	2021	HONDA	CLARITY	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	47	0	23.0	POINT (-122.64681 47.73689)	
2	5YJ3E1EB0K	King	Seattle	WA	98199.0	2019	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	220	0	36.0	POINT (-122.40092 47.65908)	C

In [18]:
```
ev['Legislative District'].mean()
```

Out[18]:
```
29.577025926324364
```

In [19]:
```
ev['Legislative District'] = ev['Legislative District'].fillna(29.0)
```

In [20]:
```
ev['Vehicle Location'].mode()
```

Out[20]:
```
0    POINT (-122.13158 47.67858)
Name: Vehicle Location, dtype: object
```

In [21]:
```
ev['Vehicle Location'] = ev['Vehicle Location'].fillna('POINT (-122.13158 47.67858)')
```

In [22]:
```
ev.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 130440 entries, 0 to 130442
Data columns (total 15 columns):
 #   Column                                             Non-Null Count   Dtype  
---  ------                                             --------------   -----  
 0   VIN (1-10)                                         130440 non-null  object 
 1   County                                             130440 non-null  object 
 2   City                                               130440 non-null  object 
 3   State                                              130440 non-null  object 
 4   Postal Code                                        130440 non-null  float64
 5   Model Year                                         130440 non-null  int64  
 6   Make                                               130440 non-null  object 
 7   Model                                              130440 non-null  object 
 8   Electric Vehicle Type                              130440 non-null  object 
 9   Clean Alternative Fuel Vehicle (CAFV) Eligibility  130440 non-null  object 
 10  Electric Range                                     130440 non-null  int64  
 11  Base MSRP                                          130440 non-null  int64  
 12  Legislative District                               130440 non-null  float64
 13  Vehicle Location                                   130440 non-null  object 
 14  Electric Utility                                   130440 non-null  object 
dtypes: float64(2), int64(3), object(10)
memory usage: 15.9+ MB
```

Result: 1.All Missing Data has been treated. 2.The Dataset also has hidden missing values disguised as 0. 3.The missing values in Model column can be imputed using the VIN column which encodes the model information. 4.A few columns have only 3 missing values and can be directly dropped.

Understanding Growth of an EV

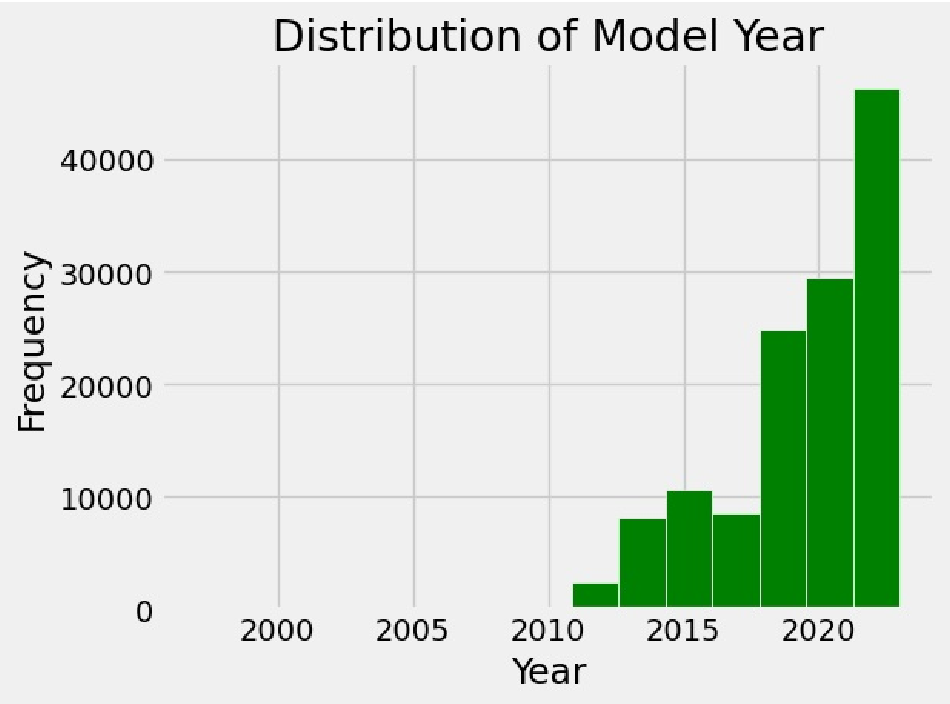
In [105…
```
ev.head()
```

Out[105]:

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID
0	5YJXCAE26J	Yakima	Yakima	WA	98908.0	2018	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	238	0	14.0	141151601
1	JHMZC5F37M	Kitsap	Poulsbo	WA	98370.0	2021	HONDA	CLARITY	Plug-in Hybrid Electric Vehicle (PHEV)	Clean Alternative Fuel Vehicle Eligible	47	0	23.0	171566447
2	5YJ3E1EB0K	King	Seattle	WA	98199.0	2019	TESLA	MODEL 3	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	220	0	36.0	9426525
3	1N4AZ0CP5D	King	Seattle	WA	98119.0	2013	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	75	0	36.0	211807760
4	5YJSA1E21H	Thurston	Lacey	WA	98516.0	2017	TESLA	MODEL S	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	210	0	22.0	185810306

In [24]:

```
# Graphical representation of "Model Year" using Histogram.

plt.hist(x= ev['Model Year'], edgecolor='white', bins= 15, color='g')
plt.title('Distribution of Model Year')
plt.xlabel('Year')
plt.ylabel('Frequency')
plt.tight_layout()
```



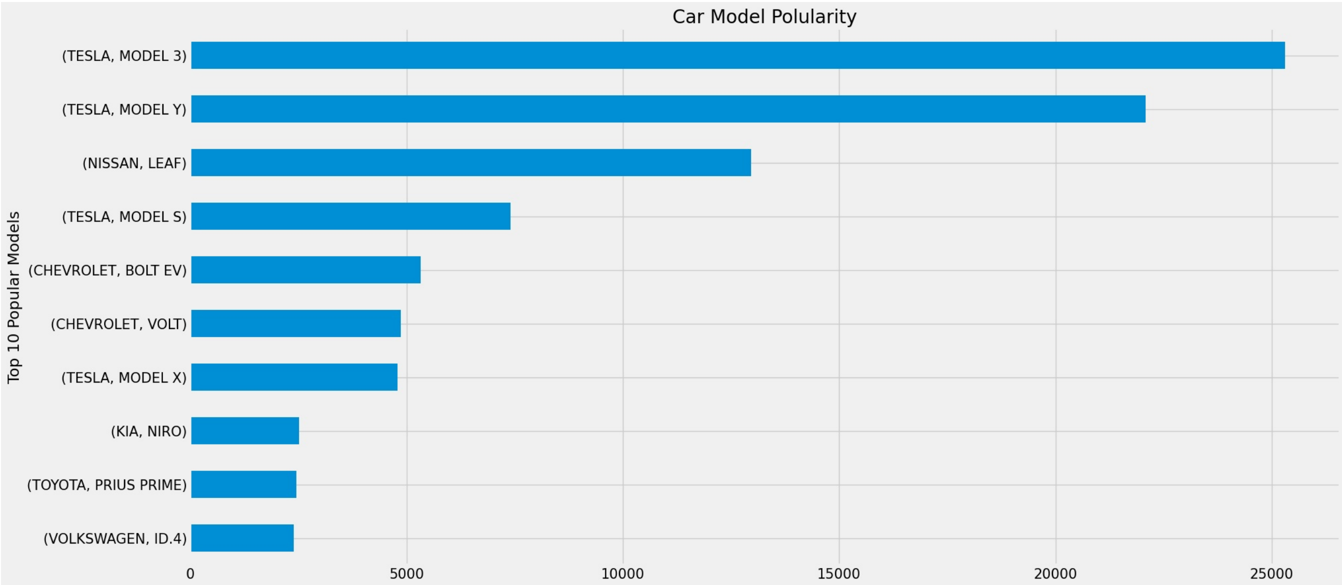
In [30]:

```
# The below code will group the count of the intersection Models and Make together.

car_popularity = ev.value_counts(['Make', 'Model'],ascending = True)

car_popularity
```

Out[30]:
```
Make        Model
CHEVROLET   S-10 PICKUP        1
PORSCHE     918                1
BENTLEY     FLYING SPUR        1
BMW         745LE              2
BENTLEY     BENTAYGA           2
                             ...
CHEVROLET   BOLT EV         5335
TESLA       MODEL S         7399
NISSAN      LEAF           12960
TESLA       MODEL Y        22078
            MODEL 3        25310
Length: 121, dtype: int64
```

In [39]:
```
car_popularity.tail(10).plot(kind='barh', title='Car Model Polularity', figsize = (20,10),
ylabel ='Total Number of Cars',
xlabel ='Top 10 Popular Models', fontsize = 15)
```

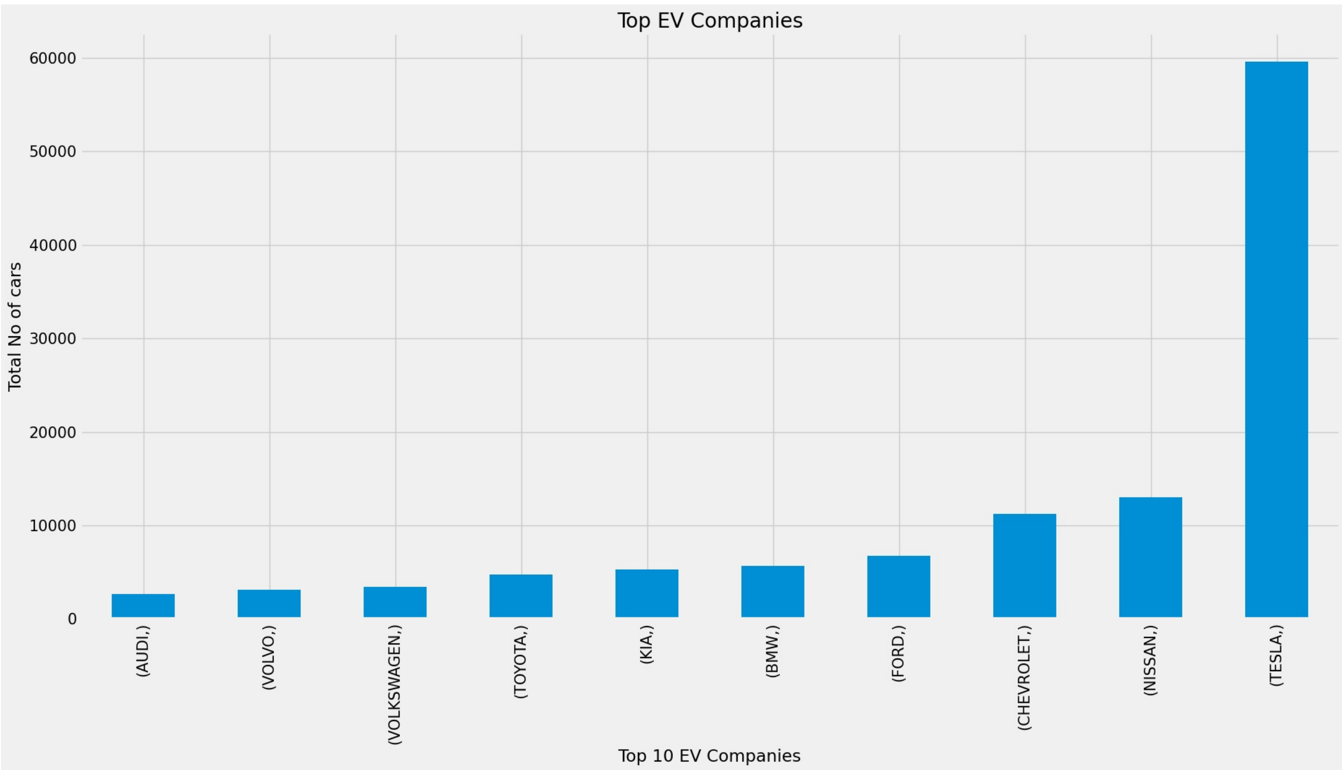
Out[39]:
```
<AxesSubplot:title={'center':'Car Model Polularity'}, ylabel='Top 10 Popular Models'>
```



In [59]:
```
top_companies = ev.value_counts(['Make'] ,ascending = True)
```

In [60]:
```
top_companies.tail(10).plot(kind= 'bar', title = 'Top EV Companies', figsize = (20, 10), fontsize = 15,
xlabel = 'Top 10 EV Companies',
ylabel = 'Total No of cars',)
```

Out[60]:
```
<AxesSubplot:title={'center':'Top EV Companies'}, xlabel='Top 10 EV Companies', ylabel='Total No of cars'>
```



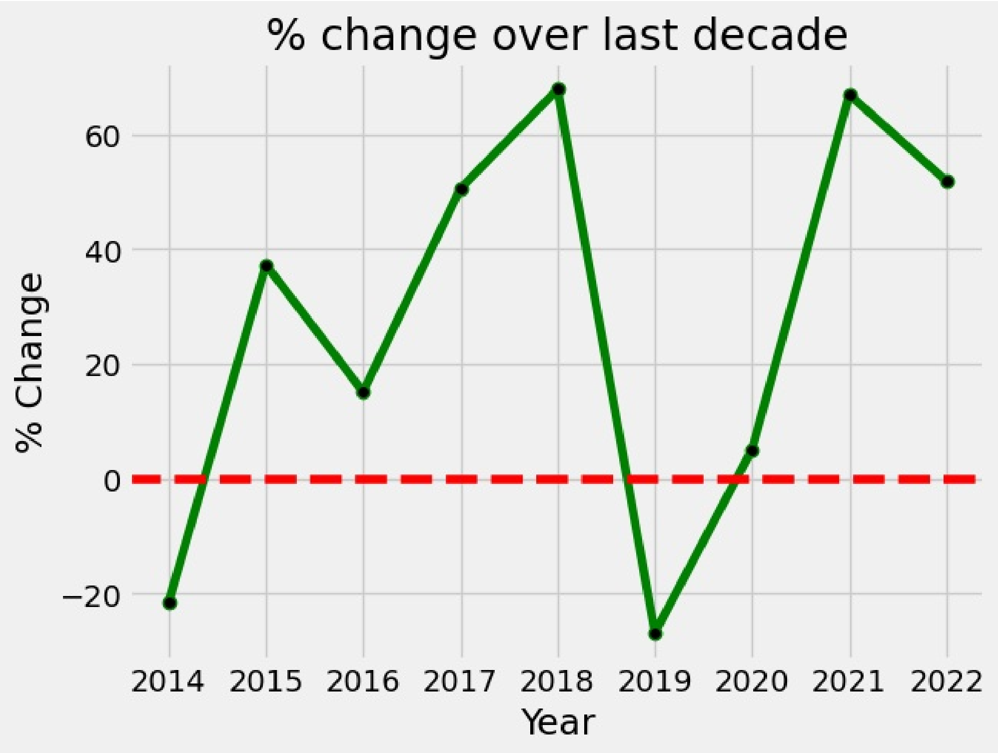
Conclusion : By concluding both graphs we can assume that the 'Model 3' of 'Tesla car comapany' has sold most of the cars in electric vehicle market followed by Nissan, Chevrolet, Ford and BMW.

In [45]:
```
model_year_data = ev['Model Year'].value_counts().sort_index(ascending = False).head(11)
```

In [45]:
```
model_year_data = ev['Model Year'].value_counts().sort_index(ascending = False).head(11)
decade_change = model_year_data.pct_change(periods=-1) * 100
```

In [47]:
```
# Plotting percentage change. (Except 2023 as the year is ongoing)

plt.plot(decade_change[1:], color= 'g', marker = 'o', markerfacecolor = 'k')
plt.axhline(y=0, color='red', ls= '--')
plt.title('% change over last decade')
plt.xlabel('Year')
plt.ylabel('% Change')
```

Out[47]:
```
Text(0, 0.5, '% Change')
```



Insights: 1.The EV Industry has been drastically rising over the last decade. 2.The percent change over the decade shows a consistent rise, except the year 2019 which saw a sharp decline but a quick recovery afterwards. 3.Note that the % change graph does not show data for the year 2023 as the year is still going on but can expect the trend to going on upwards.

Analyzing Features of EV Models.

In [48]:
```
# Calculating missing (0) values % for 'Electric Range'.

(ev['Electric Range'] == 0).sum()/len(ev)*100
```

Out[48]:
```
40.973627721557804
```

In [109…
```
ev.rename(columns={'Base MSRP': 'Base_MSRP'}, inplace=True)
```

In [111…
```
# Calculating missing (0) values % for 'Base MSRP'.

(ev['Base_MSRP'] == 0).sum()/len(ev)*100
```

Out[111]:
```
97.37969841233335
```

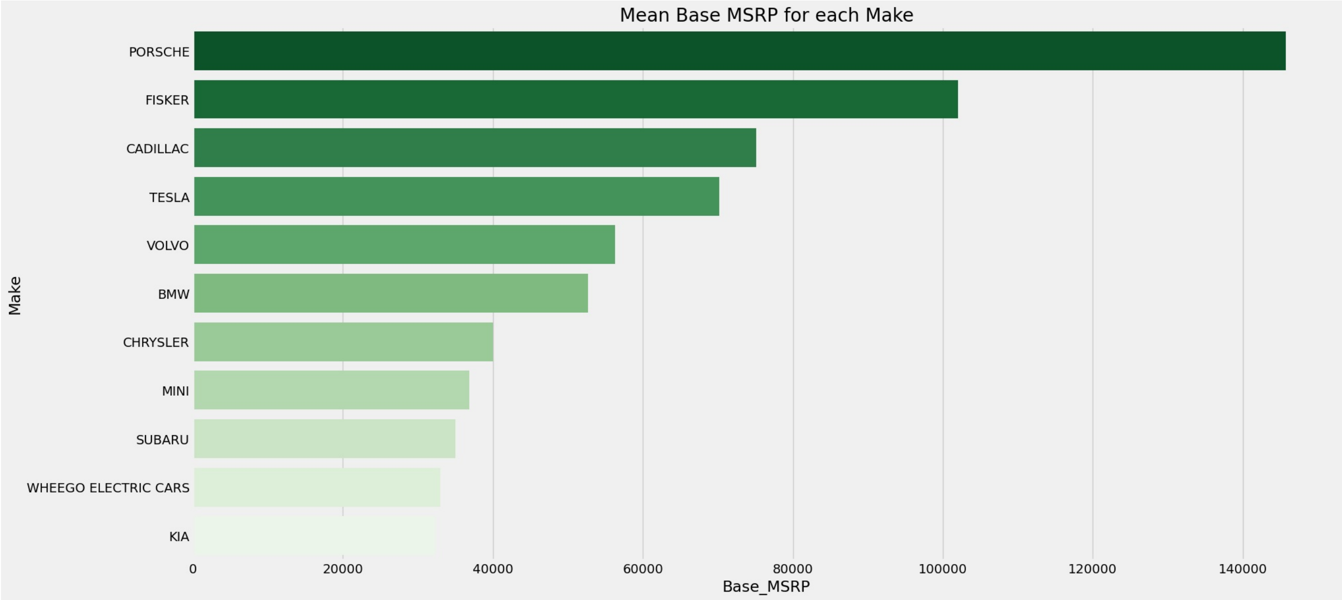
Caution : 1.The 'Electric range' and 'Base MSRP' has missing values disguised as 0. 2.The Electric range has around 41% and Base MSRP has around 97% missing values. 3.We must ignore this values or else this will disrupt our analysis.

In [112…
```
# Filtering out the missing data.

new_MSRP = ev[ev['Base_MSRP']>0]
```

In [114…
```
# Calculating the mean price of each Make.

ev2 = pd.DataFrame(new_MSRP.groupby(['Make'])['Base_MSRP'].mean())
ev2.sort_values(by= 'Base_MSRP', ascending= False, inplace= True)
```

In [129…
```
plt.figure(figsize=(20,10))
sns.barplot( y=ev2.index, x=ev2.Base_MSRP,palette='Greens_r')
plt.title("Mean Base MSRP for each Make")
```

Out[129]: Text(0.5, 1.0, 'Mean Base MSRP for each Make')

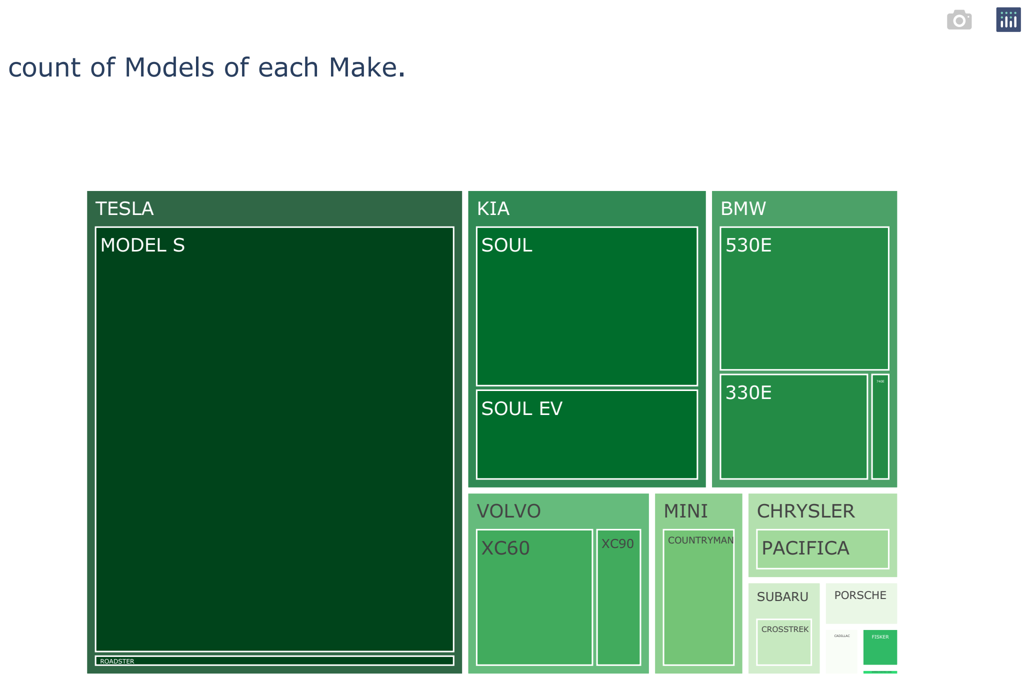


In [132…

```
# Summary for counting values of each 'model' in "Treemap".

summary = pd.DataFrame(new_MSRP.groupby(['Make'])['Model'].value_counts())
summary.columns = ['count']
summary = summary.reset_index()
```

In [137…

```
px.treemap(data_frame = summary, path=['Make' , 'Model'], values= 'count',
           color_discrete_sequence = px.colors.sequential.Greens_r,
          title = 'count of Models of each Make.')
```

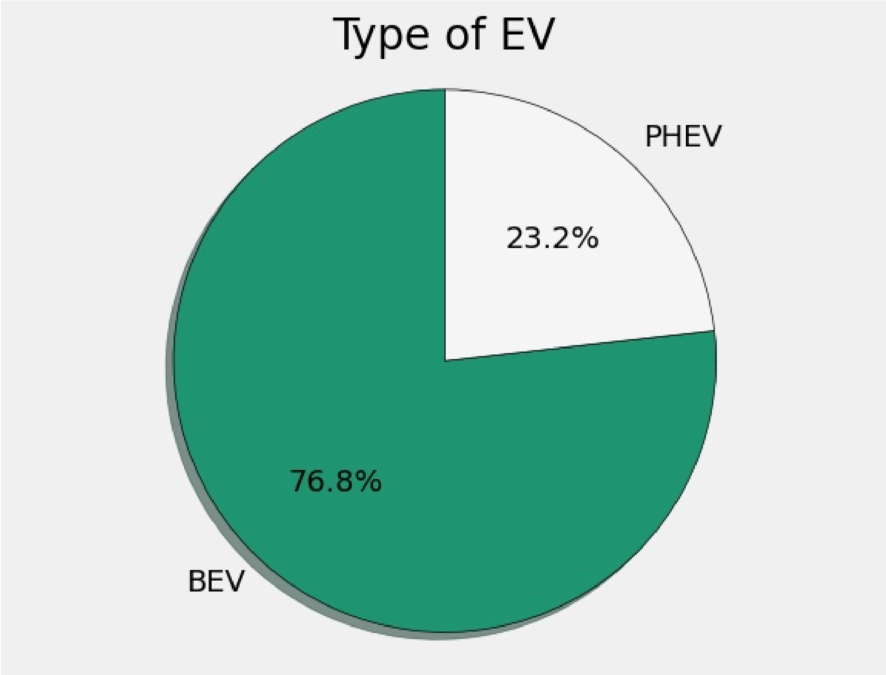


Insights: 1.The data has only around information for retail price for 11 models. 2.Porche is most expensive of all brands. 3.We may use the treemap below to understand the dispersion of different model counts because the average MSRP greatly depends on the number of instances and the number of models each Make has. 4.Since this analysis is based on the 3% of actual data, it is impractical to set the actuall price. (The analysis is to show the way to visualize this similar kind of problems.)

In [149…

```
# Pie-chart for type of EV.

fig, ax = plt.subplots()
sizes = ev['Electric Vehicle Type'].value_counts()
labels = 'BEV', 'PHEV'
ax.pie(sizes,labels=labels,autopct='%1.1f%%',shadow=True, startangle=90,wedgeprops={'edgecolor':'black'},
        colors=['#1e9570','#F5F5F5'])
ax.axis('equal')
ax.set_title('Type of EV')
```

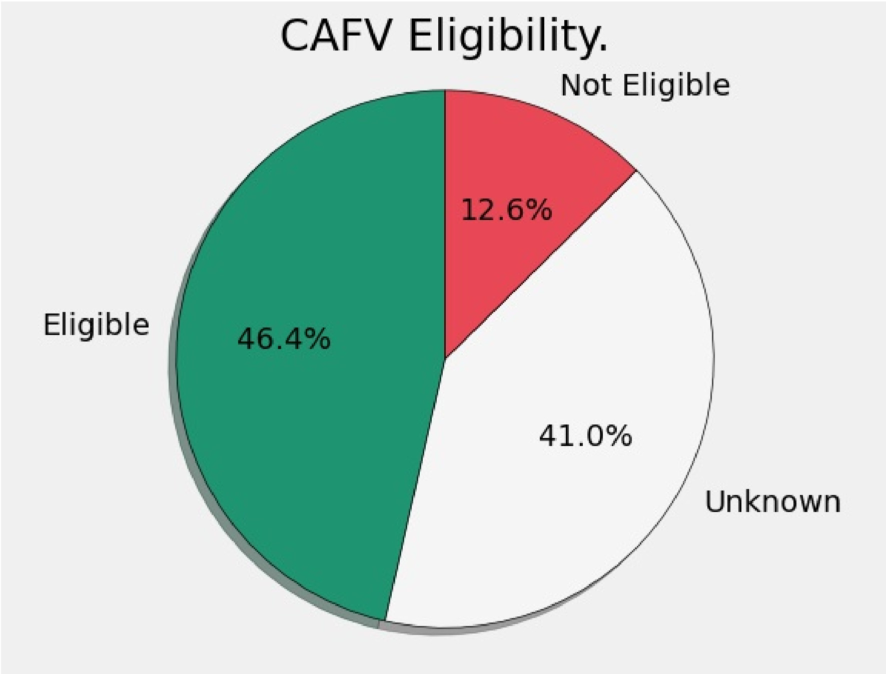
Out[149]: Text(0.5, 1.0, 'Type of EV')



In [148…

```
# Pie-chart for CAFV Eligibility.

fig, ax = plt.subplots()
sizes = ev['Clean Alternative Fuel Vehicle (CAFV) Eligibility'].value_counts()
labels = 'Eligible', 'Unknown', 'Not Eligible'
ax.pie(sizes,labels=labels,autopct='%1.1f%%',shadow=True, startangle=90,wedgeprops={'edgecolor':'black'},
      colors=['#1e9570','#F5F5F5','#E84855'])
ax.axis('equal')
ax.set_title('CAFV Eligibility.')
```

Out[148]: Text(0.5, 1.0, 'CAFV Eligibility.')



Insights:

1. There are two types of EVs with Battery Electric Vehicle being the dominant one.

2. BEV have also far superior range averaging around 200 miles while PHEV have range of only 20-50 miles.

3. We here do not have enough data to prove, but according to present data, PHEV are much cheaper than BEV.

4. Barely 12% of the vehicles do not have clean alternative feul eligibility which indicates a greener initiative.

Story :

```
The Dataset contains 130443 rows and 17 columns. The columns contains geographical as well as
various model
details about the Vehicles registered under Washingtion DOL DepartmentDOL Vehicle ID is the
unique identifier for our data
VIN (1-10) column is not unique which might seem counterintutive but in reality it is
truncated and the first 9 digits
explain world manufacturer info and vehicle descriptor section and the 10th digit encodes the
model year.
We also notice few missing values. The model and legislative distict columns in particular
have quite a few of them
and we will look to impute it or drop later. We also drop DOL Vehicle ID and 2020 Census Tract
column as it is not relevant
to our anaylsis.

         'Model 3' of 'Tesla car comapany' has sold most of the cars in electric vehicle
market followed by Nissan,
Chevrolet, Ford and BMW.
         The EV Industry has been drastically rising over the last decade.The percent change
over the decade shows
a consistent rise, except the year 2019 which saw a sharp decline but a quick recovery
afterwards.
Note that the % change graph does not show data for the year 2023 as the year is still going
on but can
expect the trend to going on upwards.
         Porche is most expensive of all brands. We may use the treemap below to understand
the dispersion of different
model counts because the average MSRP greatly depends on the number of instances and the
number of models each Make has.
Since this analysis is based on the 3% of actual data, it is impractical to set the actuall
price. (The analysis is to
show the way to visualize this similar kind of problems.) There are two types of EVs with
Battery Electric Vehicle being
the dominant one. BEV have also far superior range averaging around 200 miles while PHEV have
range of only 20-50 miles.
We here do not have enough data to prove, but according to present data, PHEV are much cheaper
than BEV. Barely 12% of
the vehicles do not have clean alternative feul eligibility which indicates a greener
initiative.
```