Electric Vehicles:
Impact on our Environment & Health

Team Tesla
2.4.20

Team Tesla Project

Purpose : Determine if the increased investment towards Electric Vehicles in the State of California has had a positive effect on the environment and the populations overall respiratory health.

Hypothesis: With YoY sales generally increasing we expect to see some improvement in California's air quality.

EV Sales & Consumers

Ethnicity County Population Data Much Too Granular

With the aid of data key as provided by US Census worked to identify ethnicity group, age groups and year.

Duplicate data columns were dropped and those remaining were summed together and converted into 4 general ethnic groups: Black, Asian, White, Other.

Also population data was reformatted from being strings to integers using "astype"

In [11]:

```python
# Sum White ("WA_MALE" + "WA-FEMALE"), BLACK ("BA_MALE" + "BA_FEMALE"), etc,
#ethnic population numbers and create new ethnic pop columns ('white_tot', black_tot, asian_tot, etc.)

ca_2018_new['white_tot'] = ca_2018_new.loc[:, 'WA_MALE':'WA_FEMALE'].sum(1)
ca_2018_new['black_tot'] = ca_2018_new.loc[:, 'BA_MALE':'BA_FEMALE'].sum(1)
ca_2018_new['asian_tot'] = ca_2018_new.loc[:, 'AA_MALE':'AA_FEMALE'].sum(1)
ca_2018_new['wht/blk/asn_tot'] = ca_2018_new.loc[:, 'white_tot':'asian_tot'].sum(1)
ca_2018_new['nonhisp_tot'] = ca_2018_new.loc[:, 'NH_MALE':'NH_FEMALE'].sum(1)
ca_2018_new['hisp_tot'] = ca_2018_new.loc[:, 'H_MALE':'H_FEMALE'].sum(1)
ca_2018_new['other_tot'] = ca_2018_new['TOT_POP'] - ca_2018_new['wht/blk/asn_tot']
ca_2018_new
```

Out[11]:

	SUMLEV	STATE	COUNTY	STNAME	CTYNAME	YEAR	AGEGRP	TOT_POP	TOT_MALE	TOT_FEMALE	WA_MALE	WA_FEMALE	BA_MALE	BA_FEM
190	50	6	1	California	Alameda County	2018	All Ages	1666753	820045	846708	417023	411943	88566	9
191	50	6	1	California	Alameda County	2018	Age <20 yrs	96716	49506	47210	23976	22753	5016	
192	50	6	1	California	Alameda County	2018	Age <20 yrs	95425	48711	46714	22773	21852	5073	
193	50	6	1	California	Alameda County	2018	Age <20 yrs	95370	48947	46423	23178	21725	5197	
194	50	6	1	California	Alameda County	2018	Age <20 yrs	93570	47278	46292	23006	22384	5475	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
12117	50	6	115	California	Yuba County	2018	Age 60-69 yrs	3580	1766	1814	1484	1526	68	
12118	50	6	115	California	Yuba County	2018	Age 70+ yrs	2628	1221	1407	1047	1201	41	
12119	50	6	115	California	Yuba County	2018	Age 70+ yrs	1706	846	860	744	725	22	
12120	50	6	115	California	Yuba County	2018	Age 70+ yrs	1086	470	616	408	540	10	
12121	50	6	115	California	Yuba County	2018	Age 70+ yrs	955	404	551	361	479	8	

1102 rows × 33 columns

In [14]:

```python
ctMax=grpD.loc[(grpD.CTYNAME=='San Francisco County')\
          |(grpD.CTYNAME=='Marin County')\
          |(grpD.CTYNAME=='Santa Clara County')\
          |(grpD.CTYNAME=='San Mateo County')\
          |(grpD.CTYNAME=='Alameda County')]

ctMax=ctMax.loc[ctMax.AGEGRP=="All Ages"]
ctMax
```

Out[14]:

	CTYNAME	AGEGRP	TOT_POP	white_tot	black_tot	asian_tot	other_tot
7	Alameda County	All Ages	1666753	828966	185930	529210	122647
167	Marin County	All Ages	259666	221914	7321	16847	13584
303	San Francisco County	All Ages	883305	467444	49184	317204	49473
327	San Mateo County	All Ages	769545	462008	21123	231957	54457
343	Santa Clara County	All Ages	1937570	1028134	54961	742295	112180

Data fields (in order of appearance):

VARIABLE	DESCRIPTION
SUMLEV	Geographic Summary Level
STATE	State FIPS code
COUNTY	County FIPS code
STNAME	State Name
CTYNAME	County Name
YEAR	Year
AGEGRP	Age group
TOT_POP	Total population
TOT_MALE	Total male population
TOT_FEMALE	Total female population
WA_MALE	White alone male population
WA_FEMALE	White alone female population
BA_MALE	Black or African American alone male population
BA_FEMALE	Black or African American alone female population
IA_MALE	American Indian and Alaska Native alone male population
IA_FEMALE	American Indian and Alaska Native alone female population
AA_MALE	Asian alone male population
AA_FEMALE	Asian alone female population
NA_MALE	Native Hawaiian and Other Pacific Islander alone male population
NA_FEMALE	Native Hawaiian and Other Pacific Islander alone female population
TOM_MALE	Two or More Races male population
TOM_FEMALE	Two or More Races female population
WAC_MALE	White alone or in combination male population
WAC_FEMALE	White alone or in combination female population
BAC_MALE	Black or African American alone or in combination male population
BAC_FEMALE	Black or African American alone or in combination female population
IAC_MALE	American Indian and Alaska Native alone or in combination male population
IAC_FEMALE	American Indian and Alaska Native alone or in combination female population

EV Sales Data by Model and Needed to be Aggregated:

Sales data was by model

First executed a groupby to aggregate totals by EV Vehicle Type.

Next summed two rows to get to an overall sales total for each year.

In [32]:
```
len(ev_sales_pd.index)
```
Out[32]: 57

In [33]:
```
ev_sales_pd2 = ev_sales_pd.drop(columns=['Total', 'Unnamed: 12', 'Unnamed: 13', 'Unnamed: 14', 'Unnamed: 15', 'Unnamed: 16',
                                         'Unnamed: 17'])
ev_sales_pd2.head()
```
Out[33]:

	Vehicle	Type	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	Chevy Volt	PHEV	767+C2:C411	23461.0	23094.0	18805.0	15393.0	24739.0	20349.0	18306.0	4915.0
1	Nissan Leaf	EV	9674	9819.0	22610.0	30200.0	17269.0	14006.0	11230.0	14715.0	12365.0
2	Smart ED	EV	342	139.0	923.0	2594.0	1387.0	657.0	544.0	1219.0	680.0
3	Mitsubishi I EV	EV	76	588.0	1029.0	196.0	115.0	94.0	6.0	0.0	0.0
4	BMW Active E	EV	0	673.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

In [34]:
```
ev_grpby = ev_sales_pd2.groupby('Type')
ev_grpby2 = ev_grpby.sum()
ev_grpby2
```
Out[34]:

Type	2012	2013	2014	2015	2016	2017	2018	2019
EV	14587.0	48094.0	63525.0	71064.0	86731.0	104492.0	238823.0	241912.0
PHEV	38584.0	49008.0	55357.0	42959.0	72885.0	91089.0	122492.0	84732.0

In [57]:
```
ev_grpby2.sum(axis=1)
inc_sum = ca_inc_rev['Income'].sum()

```

In [58]:
```
ev_grpby2.loc['Total']=ev_grpby2.sum()
ev_grpby2=ev_grpby2.reset_index()

```

In [59]:
```
ev_grpby3 = ev_grpby2.loc[ev_grpby2.Type=='Total']
ev_grpby3
```
Out[59]:

	Index	Type	2012	2013	2014	2015	2016	2017	2018	2019
2	2	Total	53171.0	97102.0	118882.0	114023.0	159616.0	195581.0	361315.0	326644.0

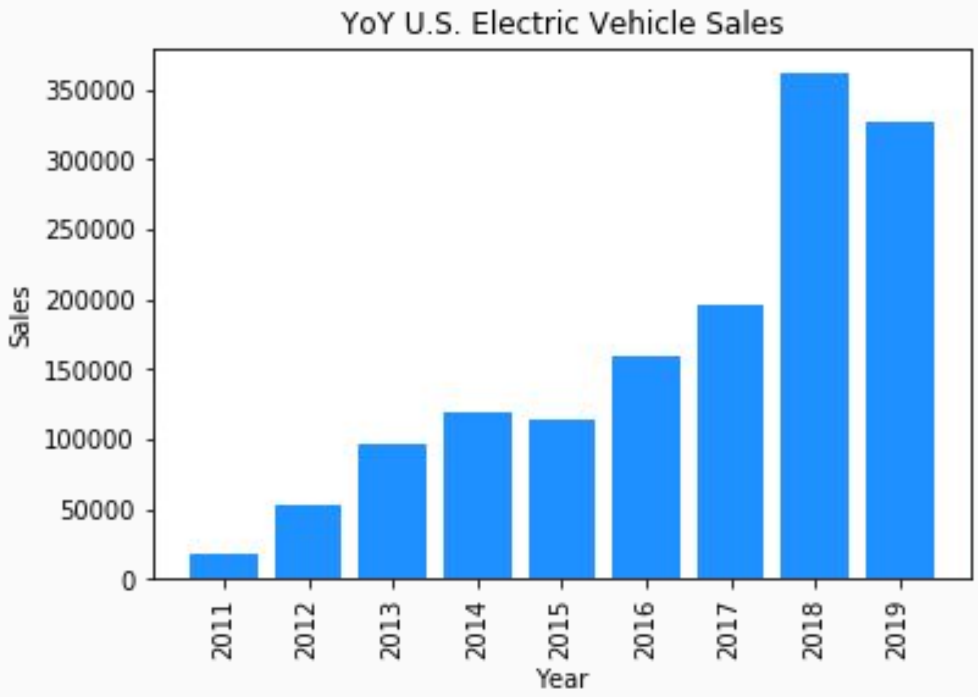
U.S. EV Sales

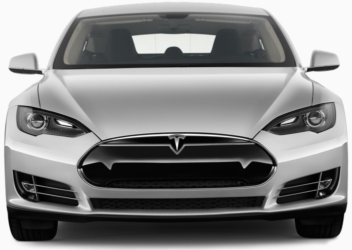
2011-2018: YoY EV sales in the US increased an average of 65%

2017 to 2018: EV sales spiked by 85% primarily due to Tesla's intro of its Model 3 (retail price @ $38,990)

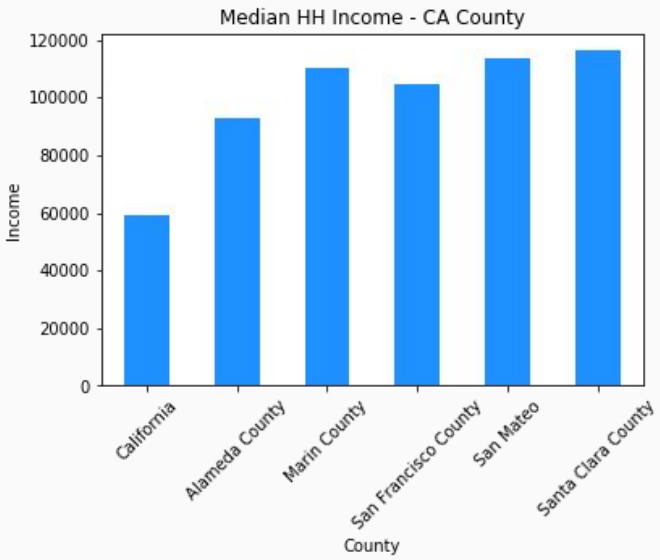
Last year almost 63% of prospective U.S. car buyers considered an EV

Cheaper batteries, more models and manufacturers ramping up assembly lines continue to drive EV prices down & spur demand.

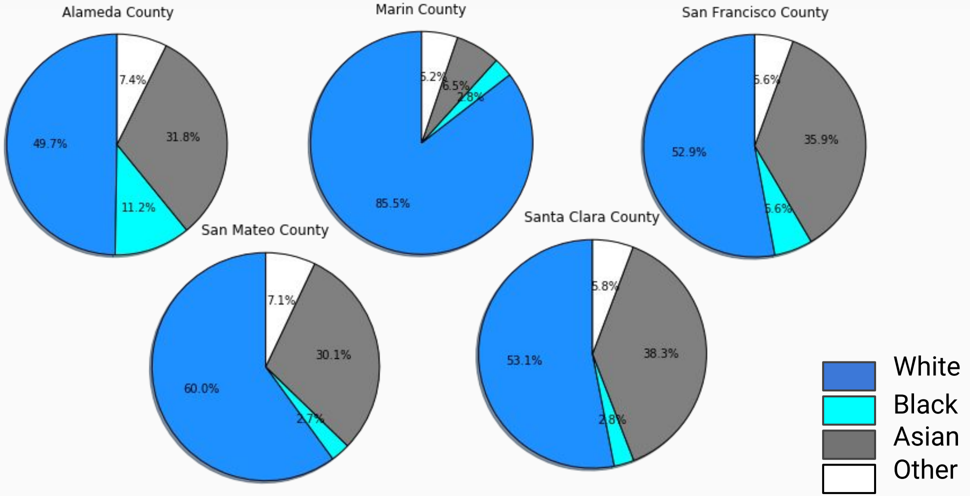




Who's Purchasing EV's In CA?



Affluent: Media HH Income 44-80% higher than CA average ($92.6K - $116.2k)



Based on County Demographics, EV owners most likely to be White or Asian

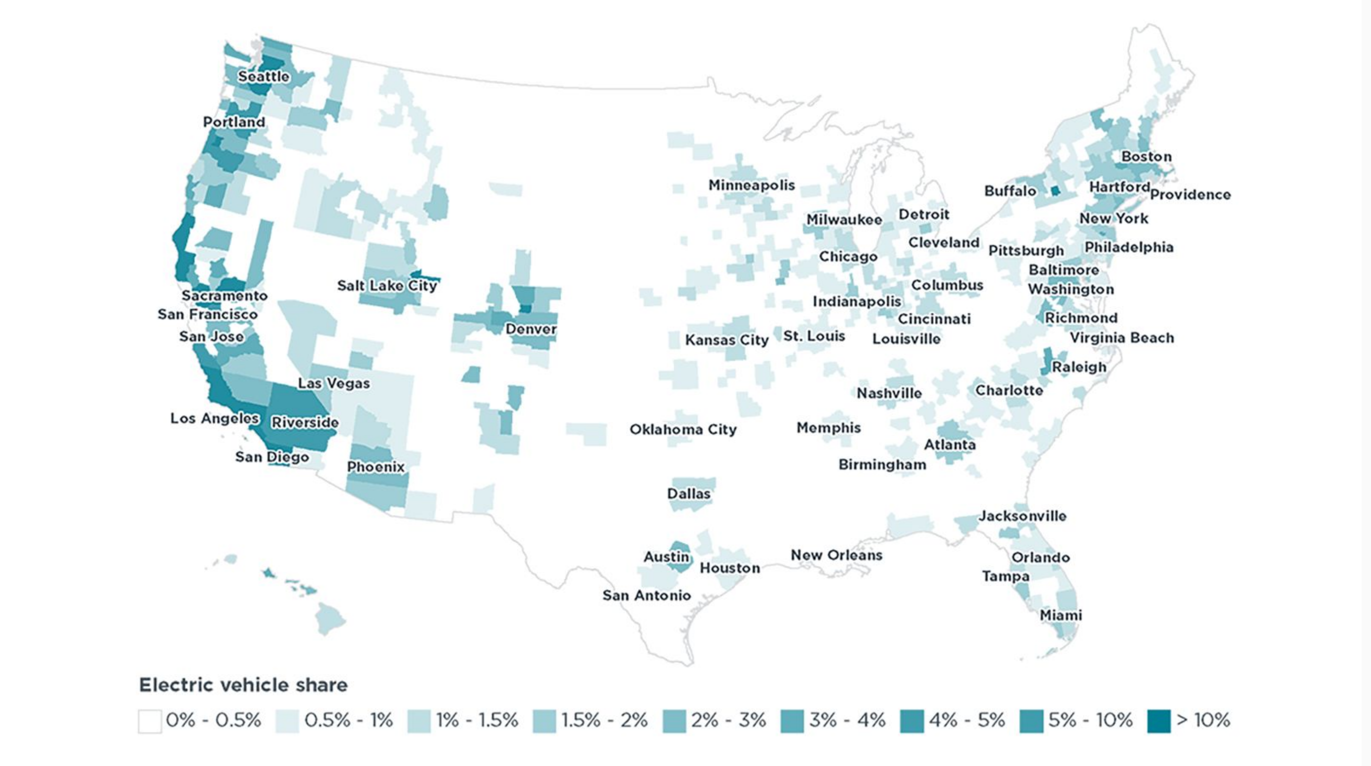
EV’s Impact on Air Quality

Problem: Limited sources of data for EVs

Limited years

Limited states

Limited adoption of EVs by geographical region



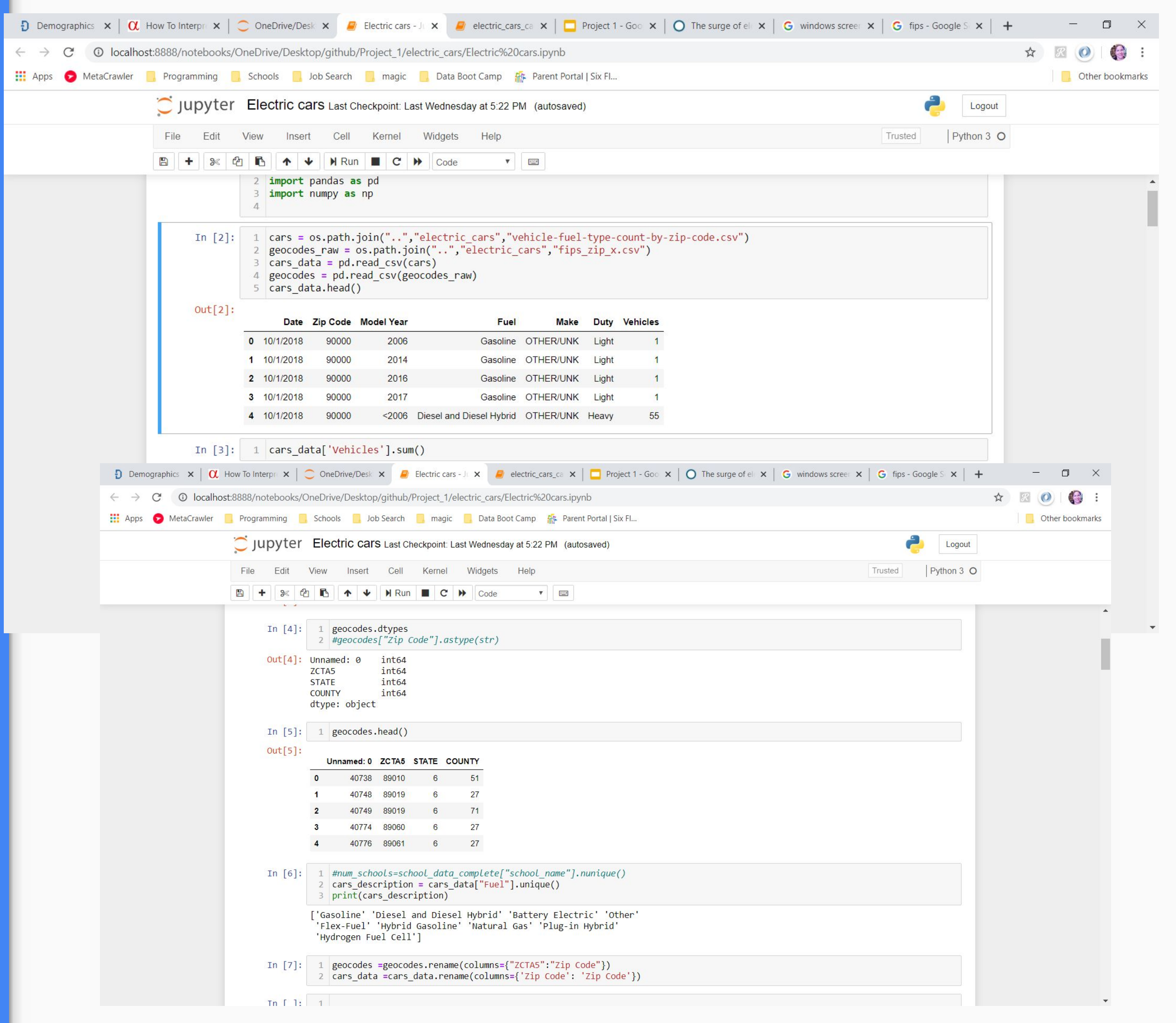
Problem: EV data is organized by zip code

All other data to do our comparisons were by county.

We used a county number csv to merge to the EV data set and ended up with cars being counted several times because of counties overlapping with zip codes.

Number of cars before merge: 30,411,713

Number of cars after merge: 33,859,585



Introducing a weighted mapping system

Found how many counties a single zip mapped into using a groupby(zip code) function

```
#getting the number of zips per county, by grouping by zip and then counting number of counties
geocodes = geocodes.rename(columns = {'ZCTA5': 'Zip Code'})
n_zip_county = geocodes.groupby('Zip Code').count().sort_values(['COUNTY'],ascending = False).reset_index()

#creating weights by doing 1/n
n_zip_county['Weights'] = 1/n_zip_county['COUNTY']
n_zip_county.head()
```

	Zip Code	Unnamed: 0	STATE	COUNTY	Weights
0	93252	4	4	4	0.250000
1	95329	3	3	3	0.333333
2	95230	3	3	3	0.333333
3	93461	3	3	3	0.333333
4	95960	3	3	3	0.333333

Weight = 1/COUNTY

Mapped zip and weights back into the zip -> FIPS code to get the county code

```
#mapping n_zips back into geocodes
geocodes_weights = pd.merge(geocodes,n_zip_county,
                            on = 'Zip Code',how = 'inner')
geocodes_weights
```

	Unnamed: 0	Zip Code	STATE	COUNTY	Weights
0	40738	89010	6	51	1.0
1	40748	89019	6	27	0.5
2	40749	89019	6	71	0.5

Weighted System continued

Merged zip/fips with weights to vehicle data

```
#merging weights/number of car regristrations by zip code into fips code
cars_merge_county = pd.merge(cars_data, geocodes_weights, how="inner", on=["Zip Code"])
cars_merge_county.sort_values(['Weights']).head()
```

	Date	Zip Code	Model Year	Fuel	Make	Duty	Vehicles	Unnamed: 0	STATE	COUNTY	Weights
298101	10/1/2018	93252	2007	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	4	41696	6	83	0.25
298290	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41697	6	111	0.25
298289	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41696	6	83	0.25
298288	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41695	6	79	0.25
298287	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41694	6	29	0.25

Then multiplied the weight by the number of vehicles, to get adjusted vehicles

```
#multiplying the weights by number of vehicles
cars_merge_county['Adj Veh'] = cars_merge_county['Vehicles']*cars_merge_county['Weights']

cars_merge_county.sort_values(['Weights'], ascending = True).head(5)
```

	Date	Zip Code	Model Year	Fuel	Make	Duty	Vehicles	Unnamed: 0	STATE	COUNTY	Weights	Adj Veh
298101	10/1/2018	93252	2007	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	4	41696	6	83	0.25	1.00
298290	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41697	6	111	0.25	0.25
298289	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41696	6	83	0.25	0.25
298288	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41695	6	79	0.25	0.25
298287	10/1/2018	93252	2017	Diesel and Diesel Hybrid	OTHER/UNK	Heavy	1	41694	6	29	0.25	0.25

Validation Process

Unique Zip codes in Cars data: 2913

Unique Zip Codes in crosswalk file: 2448

The difference in these zips led to a an insignificant amount in the data < 1%

```python
sum_cars = cars_merge_county[['Vehicles','Adj Veh','Fuel','COUNTY']].groupby(['Fuel','COUNTY']).


difference = round(cars_data['Vehicles'].sum() - sum_cars['N Adj Veh'].sum())
percentage = round(difference/cars_data['Vehicles'].sum(),4)

print('the original number of cars is ' + str(cars_data['Vehicles'].sum()))
print('The number of unadjusted vehicles are ' + str(round(sum_cars['Veh'].sum())))
print('The number of adjusted vehicles are ' + str(round(sum_cars['N Adj Veh'].sum())))
print('The difference between the original cars and the adjusted cars is ' + str(difference))
print('The percentage of this change is '+ str(percentage) + ' of the data')

sum_cars.head()
```

```
the original number of cars is 30411372
The number of unadjusted vehicles are 33859585
The number of adjusted vehicles are 30386851.0
The difference between the original cars and the adjusted cars is 24521.0
The percentage of this change is 0.0008 of the data
```

Choosing vehicle categories to use

In vehicle registration data there were 9 vehicle types to choose from.

Originally we chose 2 categories:

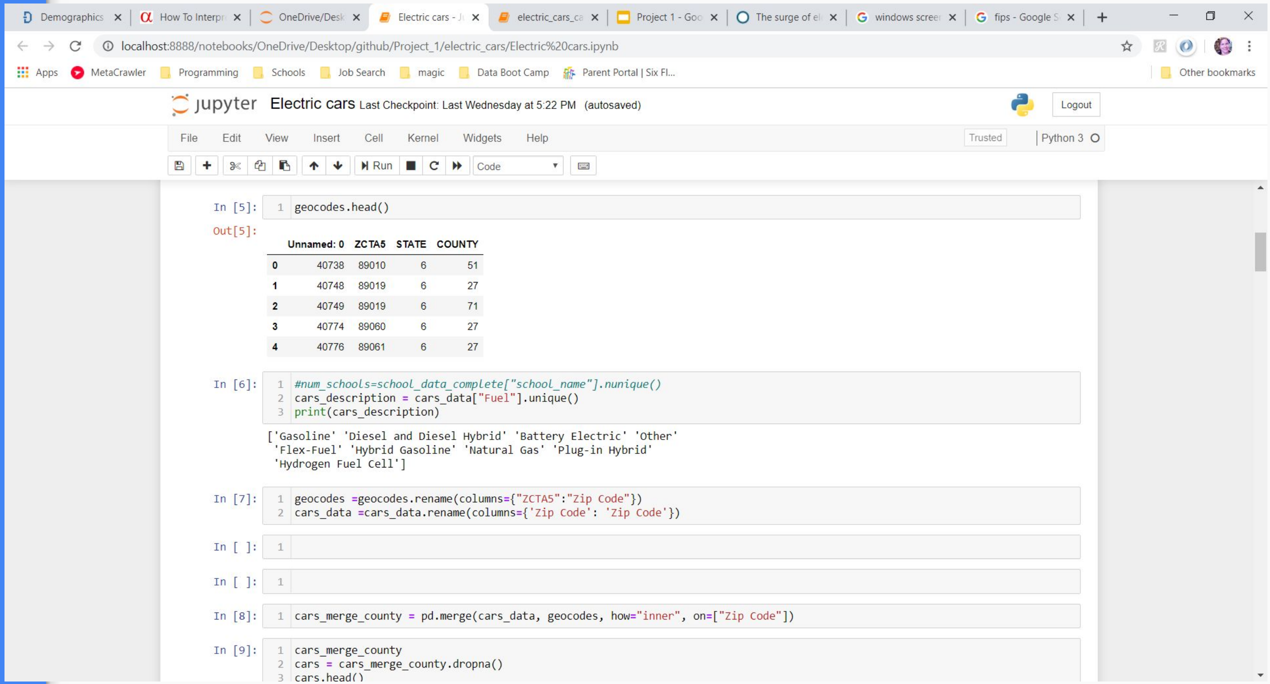
Battery Electric

Plug-In Hybrid

After counting the number of these cars it was apparent that it wouldn't be statistically significant.

>.01% of cars in CA

Added Hybrid gasoline cars resulting in some counties containing ~10%.



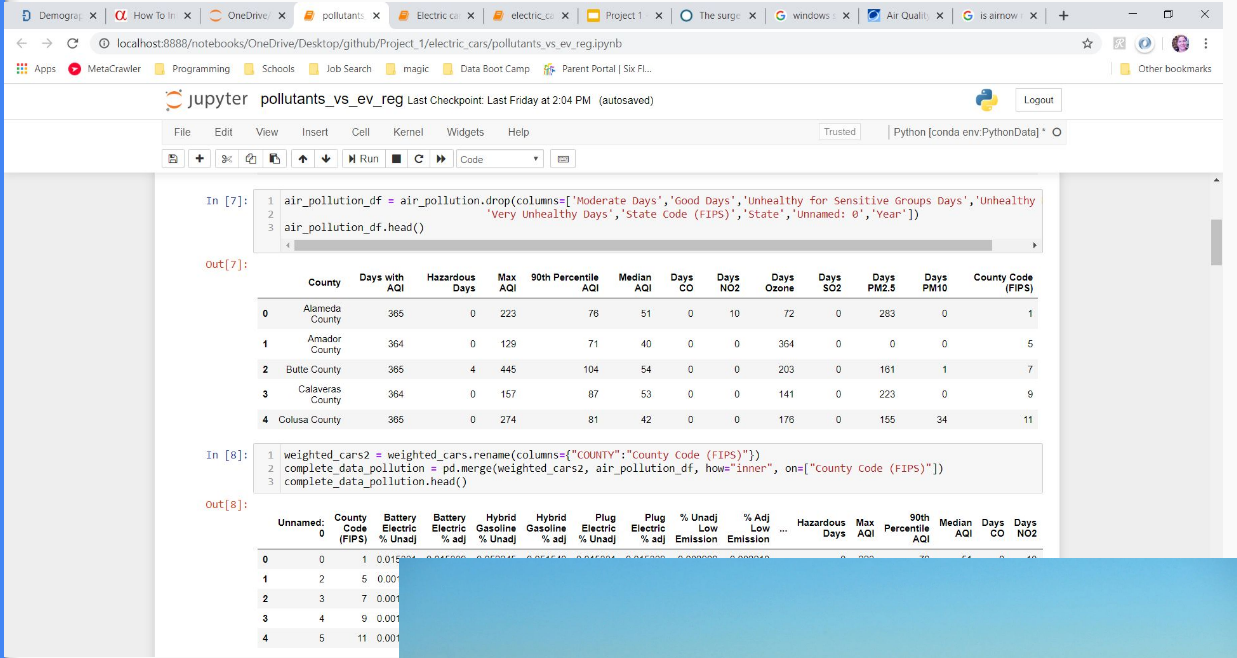
Choosing pollutants as a measure

EPA's data contains several parameters.

Air pollutants vary in potency, and the function used to convert from air pollutant concentration to AQI varies by pollutant.

Median AQI was used.

Since different countries use different models to compute median AQI it was decided to consider PM10 and ozone as measures of gasoline/diesel car emissions.



Linear regression for each pollutant

Linear regression models showed a decrease in median AQI, ozone and PM10.

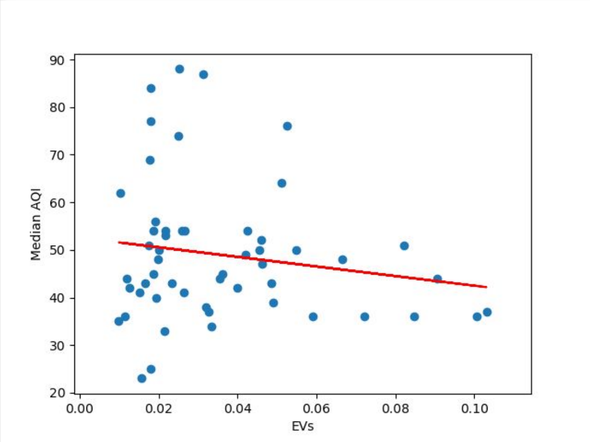
R^2 values were:

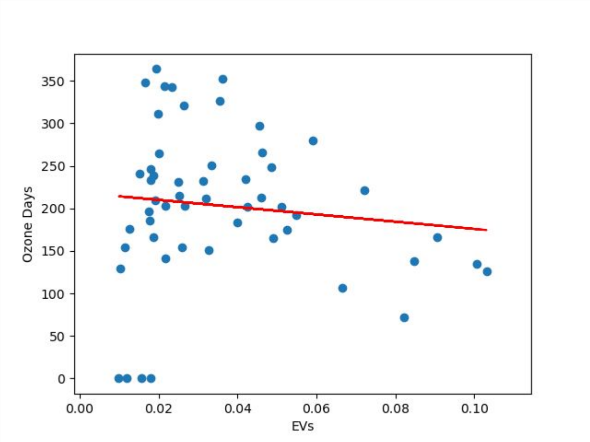
-.3116 for PM10

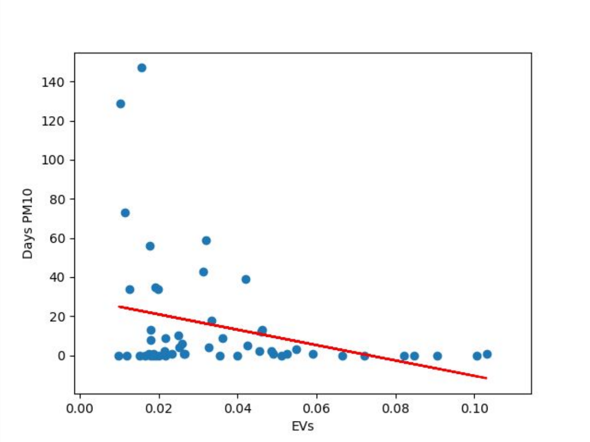
-.114 for ozone

-.1642 for median AQI

Based on this given data and models used there is no correlation between percentage of EVs in each county of CA and air quality factors. No linear relationship.







Linear regression for median AQI

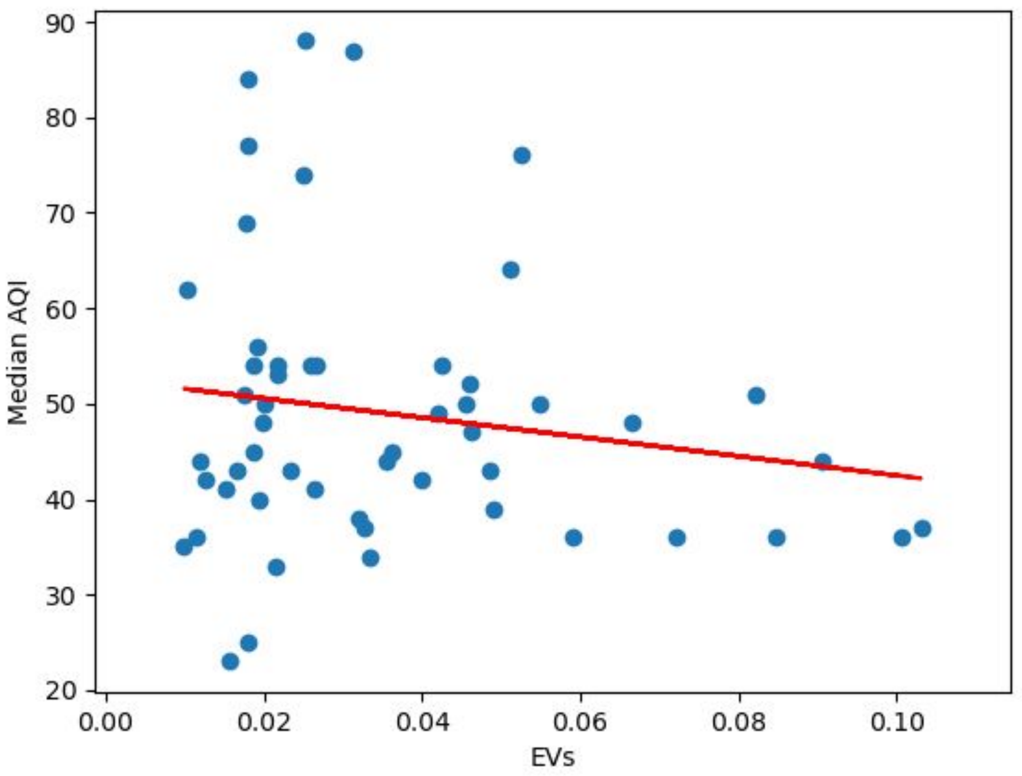
Graph shows a decrease in median AQI with an increase in EVs per county

Again a terrible R^2 value

-.1642

Model does not fit

Based on this given data and models used there is no correlation between percentage of EVs in each county of CA and air quality factors. No linear relationship.



EV’s Impacts on Health

Question: impact EV's have on health outcomes

Motivation: EV's provide lower emissions than other vehicle fuel types, therefore air quality should be better and respiratory rates should decrease

Data source: California Health and Human Services Open Data

Theory

EV ↑ Air Quality ↑ Asthma Instances ↓

Outcome Variable

Outcome Variable: Emergency Department (ED) visits as a percentage of total county population

Can be repeat people, can be people outside of the county

On per 10,000 people basis, percentage of 10% is actually .0001%

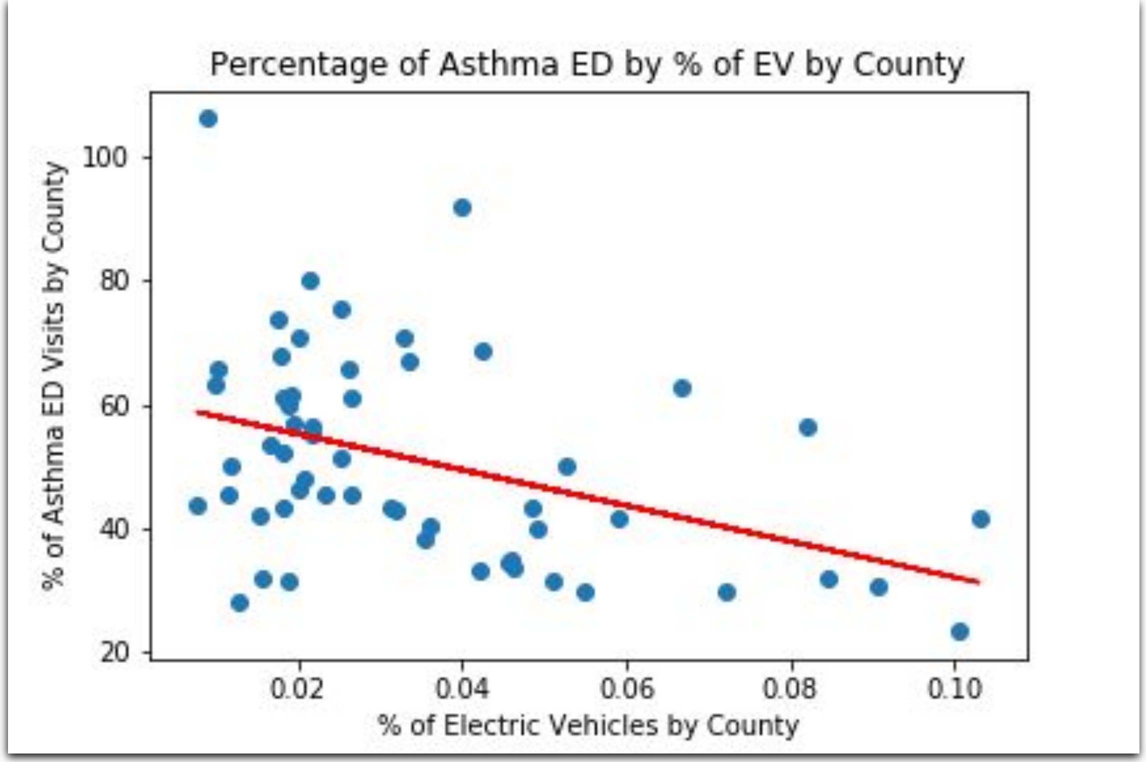
$$\%ED\ Visits_{asthma} = \frac{Number\ of\ ED\ Visits\ for\ County}{Total\ Population\ by\ County} \times 10{,}000$$

Linear Regression Analysis

R Square Value: -.4

P-Value: .002

Conclusion: visual inspection shows a negative relationship exists, but negative r-square indicates no linear relationship



Sources of Power in CA

Background

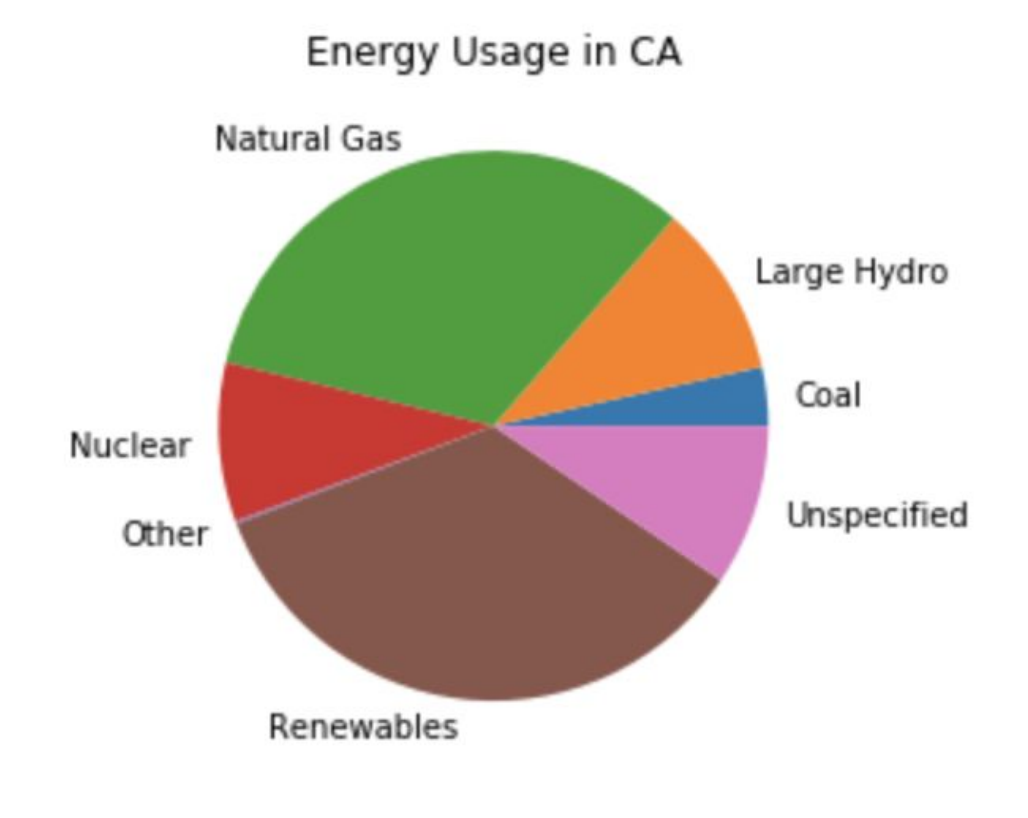
Trends seemed relatively weak

Correlated in the right direction but not strong

Why would we see this?

While EVs can help the environment
-> Energy needed to power them are still mostly non-renewable

While on the rise
-> Renewables are not the minority



Code

```python
Coal = 0
Large_Hydro = 0
Natural_Gas = 0
Nuclear = 0
Oil = 0
Other = 0
Renewables = 0
Unspecified = 0


for i in range(len(df_energy)):
    Coal += df.iloc[i,1]
    Large_Hydro += df.iloc[i,2]
    Natural_Gas += df.iloc[i,3]
    Nuclear += df.iloc[i,4]
    Oil += df.iloc[i,5]
    Other += df.iloc[i,6]
    Renewables += df.iloc[i,7]
    Unspecified += df.iloc[i,8]
```

	County	Coal	Large Hydro	Natural Gas	Nuclear	Oil	Other	Renewables	Unspecified
0	Los Angeles County	1877.0	6999.0	19578.0	7599.0	6.00	102.73	26205.0	6119.5
1	San Diego County	590.0	2370.0	6870.0	1670.0	2.99	26.00	6299.0	1972.0
2	Orange County	843.0	1956.0	5793.0	1886.0	3.00	25.00	7878.0	1813.0
3	Riverside County	490.0	1784.0	5907.0	1399.0	2.40	20.50	4818.0	1836.0
4	San Bernardino County	678.0	1814.0	5107.0	1546.0	3.90	19.50	4668.0	1798.0

```python
pie_df = pd.DataFrame({"Usage": [Coal,
                                  Large_Hydro,
                                  Natural_Gas,
                                  Nuclear,
                                  Other,
                                  Renewables,
                                  Unspecified]},
                       index = ["Coal",
                                "Large Hydro",
                                "Natural Gas",
                                "Nuclear",
                                "Other",
                                "Renewables",
                                "Unspecified"])


pie_chart = pie_df.plot.pie(y="Usage", legend = None)

pie_chart.set_ylabel("")
pie_chart.set_title("Energy Usage in CA")
```

Hypothesis

While CA as a whole uses about Renewable Energy around 32%

-> It's possible that in areas where there is a greater level of electric vehicle adoption, we also see greater usage of Renewable Energies

Used:
Energy Usage by Type and County
EV adoption by county

Sorted by adoption rate and created separate groups
-> Top 10 Counties
-> Rest of CA

```
topEVaverage = topTen_EVCounty['Renewables %'].mean()
topEVaverage
```

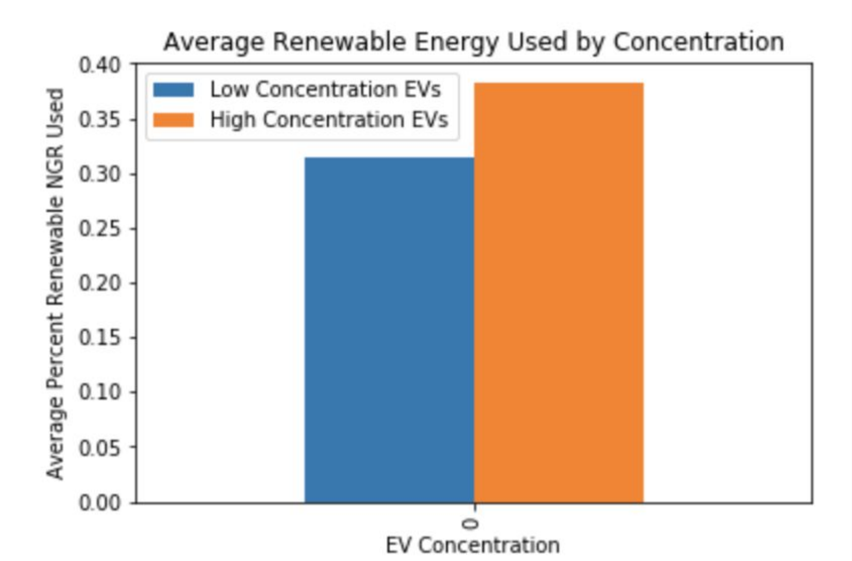
```
0.38163391895439486
```

```
bottomEVaverage = Bottom_EVCounty['Renewables %'].mean()
bottomEVaverage
```

```
0.3135976197317883
```

```
renew_pect = pd.DataFrame({"Low Concentration EVs": [bottomEVaverage],
                           "High Concentration EVs": [topEVaverage]})

renew_pect
chart = renew_pect.plot(kind='bar', title = 'Average Renewable Energy Used by Concentration')
chart.set_xlabel("EV Concentration")
chart.set_ylabel("Average Percent Renewable NGR Used")
```



Independent T-Test

The Bar Chart seemed promising

Seems to be 2 Distinct Groups that use Renewables at different rates

Test Statistic: 9.22

PValue: 3.3453e-06

Pvalue is below confidence Level

Reject Null Hypothesis

Conclusion:
The 10 Counties that use EVs at the highest rate also use Renewable Energy at a statistically different rate then the rest of CA

```
energyAndCar = pd.merge(energy_countyCodes,car_percentage, on = 'COUNTY',how = 'inner')
energyAndCar.head()
```

```
sortedDown_NRGandCAR = energyAndCar.sort_values(["% Adj Low Emission"],
                                    ascending = False).reset_index()
```

```
topTen_EVCounty = sortedDown_NRGandCAR.head(10)
```

```
sortedUp_NRGandCAR = energyAndCar.sort_values(["% Adj Low Emission"],
                                    ascending = True).reset_index()
```

```
Bottom_EVCounty = sortedUp_NRGandCAR.head(40)
```

```
cleanTop = topTen_EVCounty[["COUNTY", "County","Renewables %","% Adj Low Emission"]]
cleanTop.head()
```

	COUNTY	County	Renewables %	% Adj Low Emission
0	75	San Francisco County	0.371807	0.103046
1	41	Marin County	0.402522	0.100525
2	85	Santa Clara County	0.379077	0.090614
3	81	San Mateo County	0.370169	0.084710
4	1	Alameda County	0.380242	0.082218

```
stats.ttest_ind(cleanTop['% Adj Low Emission'],
                cleanBottom['% Adj Low Emission'],
                equal_var=False)
```

Conclusion

It is great to see EVs on the rise and their benefits to society

But we are still behind

In order to power these cars, we still need Gas and Coal

We are too reliant on non-renewables

In order to see true change, we need to increase the number of Renewable energy power plants

DATA SOURCE	DATA
Transportation Research Center at Argonne Laboratory	YoY US EV Sales Data
Alliance of Auto Manufacturers	2018 EV Sales By State Data
2018 U.S. Census Data	CA County Population Data By Ethnic Group Data
Robert Wood Johnson Foundation/ UW Population Health Institute	CA County Median HH Income Data
EPA	Air Quality Index Report
Zips to FIPS crosswalk	HUD
California Open Data	EV registration
California Health and Human Services Open Data	Asthma emergency department rates
CA.gov - Total System Electric Generation	2018 - Energy usage by type
CA.go - Electricity consumption by County	2018 - Energy Usage by County