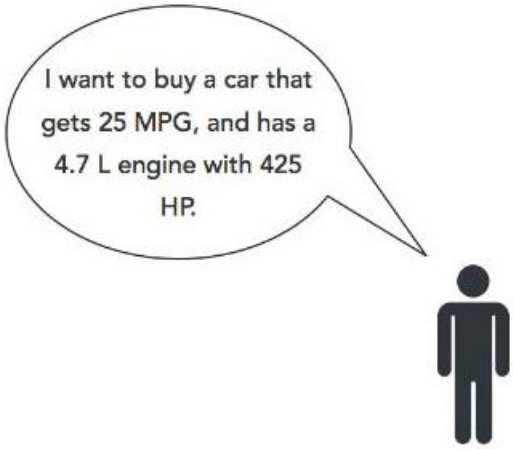
Content Based Recommender Systems

Content Based Recommenders recommend an item based on its features and how similar those are to features of other items in a dataset.

Here we use Nearest neighbor algorithm to build a content-based recommender. Imagine you are a car dealer. you get a customer that comes in and tells you that he wants a car that gets 25 miles per gallon and has a 4.7 liter,425 horsepower engine. you have a dataset that describes all of these spaces on the cars in your inventory.

car id no.	mpg	hp	engine size (L)
1	39	184	2.4
2	20	449	4.7
3	31	185	2.4
4	16	454	4.7



so you could use the nearest neighbor algorithm to identify the single best matching car from among all cars in your favorite inventory database.

To do this you'd create a single test point that represents your customer's desired specifications, when you run the model it will memorize all of the data points in your inventory. It will then calculate the single data point that is quantitatively most similar to your test point. The car that this data point represents would be the one you'd recommend to your customer.

Let me show in the coding demostration:

Step1: import libraries

```
import numpy as np
import pandas as pd
import sklearn
from sklearn.neighbors import NearestNeighbors
```

we use the **mtcars dataset**, source of dataset is: *Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.*

Step2:load dataset:

```
cars=pd.read_csv('mtcars.csv')
cars.columns=['car_names','mpg','cy1','disp','hp','drat','wt','qsec','vs','am','gear','carb']
print(cars.head())
```

	car_names	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2

So imagine that a customer walks in and tells you that he's looking for a car that weights 3.2 tons, gets at least 15 miles per gallon, has an engine with a displacement size of 300 cubic inches, and a power of 160 horsepower. Let's take a test point to represents the shopper's specifications.

Step3:set test data point

```
t=[15,300,160,3.2]
X=cars.ix[:,(1,3,4,6)].values
print(X[0:5])
```

```
array([[  21.   ,  160.   ,  110.   ,    2.62 ],
       [  21.   ,  160.   ,  110.   ,    2.875],
       [  22.8  ,  108.   ,   93.   ,    2.32 ],
       [  21.4  ,  258.   ,  110.   ,    3.215],
       [  18.7  ,  360.   ,  175.   ,    3.44 ]])
```

see how the nearest neighbor algorithm can be used to recommend a car for this shopper based on his requirements.

Step4: Use NearestNeighbors

```
NN=NearestNeighbors(n_neighbors=1)
nbrs=NN.fit(X)
print(nbrs.kneighbors([t]))
print(cars)
```

```
(array([[ 10.77474942]]), array([[22]], dtype=int64))
```

And let's look at cars dataset to see 22 rows of data:

```
print(cars)
```

10	Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
11	Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
12	Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
13	Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
14	Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
15	Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
16	Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
17	Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
18	Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
19	Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
20	Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
21	Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
22	AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2

So according to our nearest neighbor model,you should recommend athe shopper to take a closer look at this AMC Javelin car, because it's the most similar car to the shopper's specifications of all cars that the car dealer has on his lot.

CODE

```
 import numpy as np
import pandas as pd
import sklearn
from sklearn.neighbors import NearestNeighbors

cars=pd.read_csv('mtcars.csv')
cars.columns=['car_names','mpg','cy1','disp','hp','drat','wt','qsec','vs','am','gear','carb']

t=[15,300,160,3.2]
X=cars.ix[:,(1,3,4,6)].values


NN=NearestNeighbors(n_neighbors=1)
nbrs=NN.fit(X)
print(nbrs.kneighbors([t]))
print(cars)
```