ECHO HAPPINESS SCORE

ARE WE MAKING OUR CUSTOMERS
HAPPY?
WILL THEY END UP RENEWING?

The Old

Can we use predictive analytics to focus the efforts of our retention team?

The proprosed features were:

Social Media
Activity
Practice Type
Patient
Interaction
Affluence

However...

Most of those features did not work

But...

PRACTICE MAKES PERFECT

The New

Happiness Score

Given:

Doctor Type
Price
Years
Share Usage
Device Usage
Previous Products

Decision Maker

We assume that people act similar to their peers

Can we predict if our customer will renew?

1

Data Cleaning

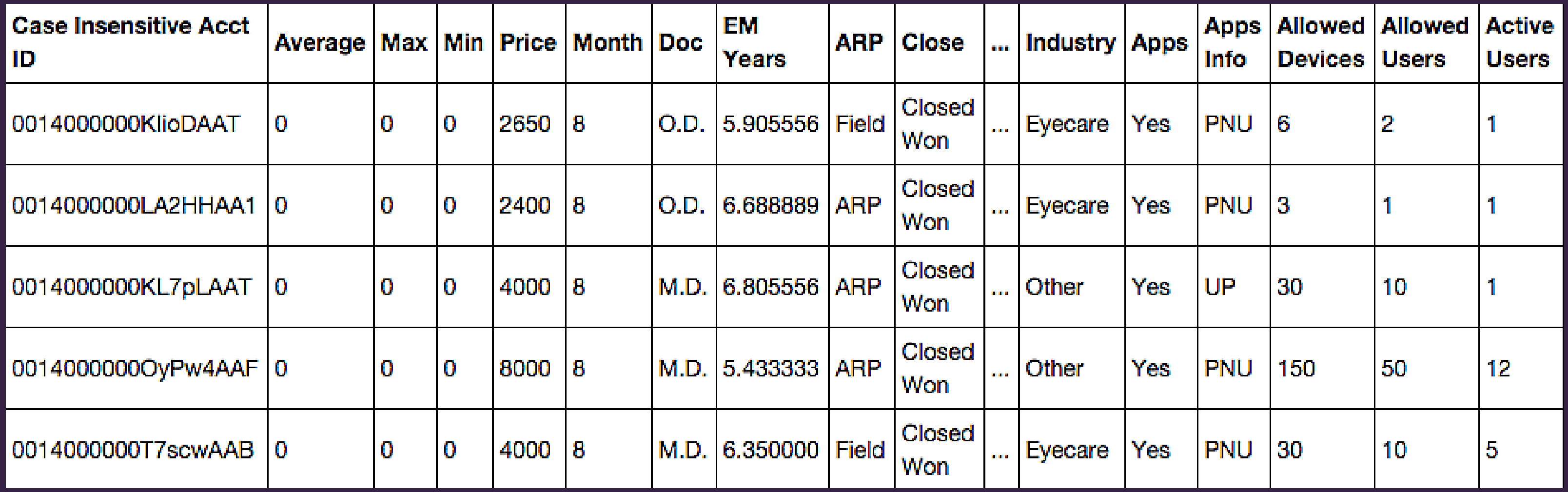
Used Salesforce Data that is populated from MySQL and DynamoDB

However....

Tons of data missing!

1

DATA CLEANING



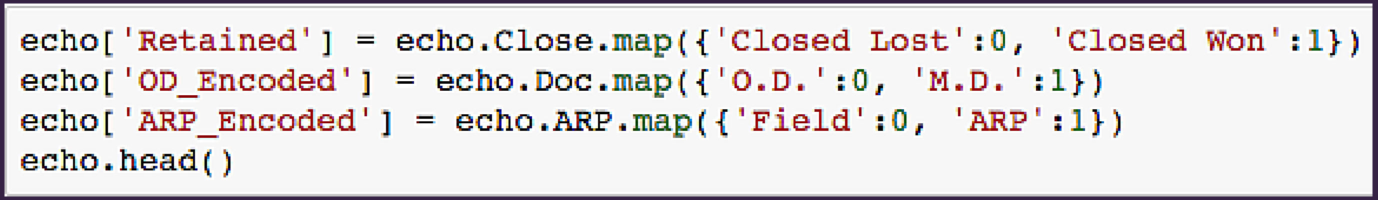
Case Insensitive Acct ID	Average	Max	Min	Price	Month	Doc	EM Years	ARP	Close	...	Industry	Apps	Apps Info	Allowed Devices	Allowed Users	Active Users
0014000000KlioDAAT	0	0	0	2650	8	O.D.	5.905556	Field	Closed Won	...	Eyecare	Yes	PNU	6	2	1
0014000000LA2HHAA1	0	0	0	2400	8	O.D.	6.688889	ARP	Closed Won	...	Eyecare	Yes	PNU	3	1	1
0014000000KL7pLAAT	0	0	0	4000	8	M.D.	6.805556	ARP	Closed Won	...	Other	Yes	UP	30	10	1
0014000000OyPw4AAF	0	0	0	8000	8	M.D.	5.433333	ARP	Closed Won	...	Other	Yes	PNU	150	50	12
0014000000T7scwAAB	0	0	0	4000	8	M.D.	6.350000	Field	Closed Won	...	Eyecare	Yes	PNU	30	10	5

Encoded Features:

Won or Lost Account

ARP or Not

Type of Doctor



```
echo['Retained'] = echo.Close.map({'Closed Lost':0, 'Closed Won':1})
echo['OD_Encoded'] = echo.Doc.map({'O.D.':0, 'M.D.':1})
echo['ARP_Encoded'] = echo.ARP.map({'Field':0, 'ARP':1})
echo.head()
```

2

Feature Engineering

EM_Years

How many years since the first purchase of our product

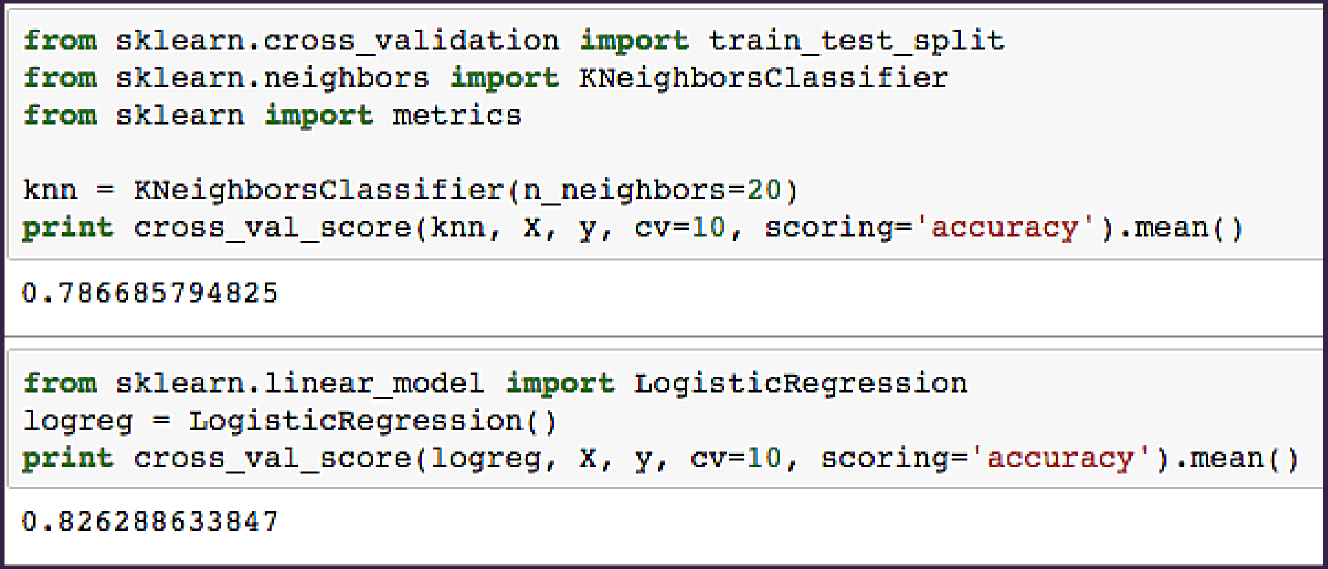
Closed

Whether or not an account was won or lost based on expiration and renewal dates

3

DATA MODELING

KNN



```
from sklearn.cross_validation import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn import metrics

knn = KNeighborsClassifier(n_neighbors=20)
print cross_val_score(knn, X, y, cv=10, scoring='accuracy').mean()
```

```
0.786685794825
```

```
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression()
print cross_val_score(logreg, X, y, cv=10, scoring='accuracy').mean()
```

```
0.826288633847
```

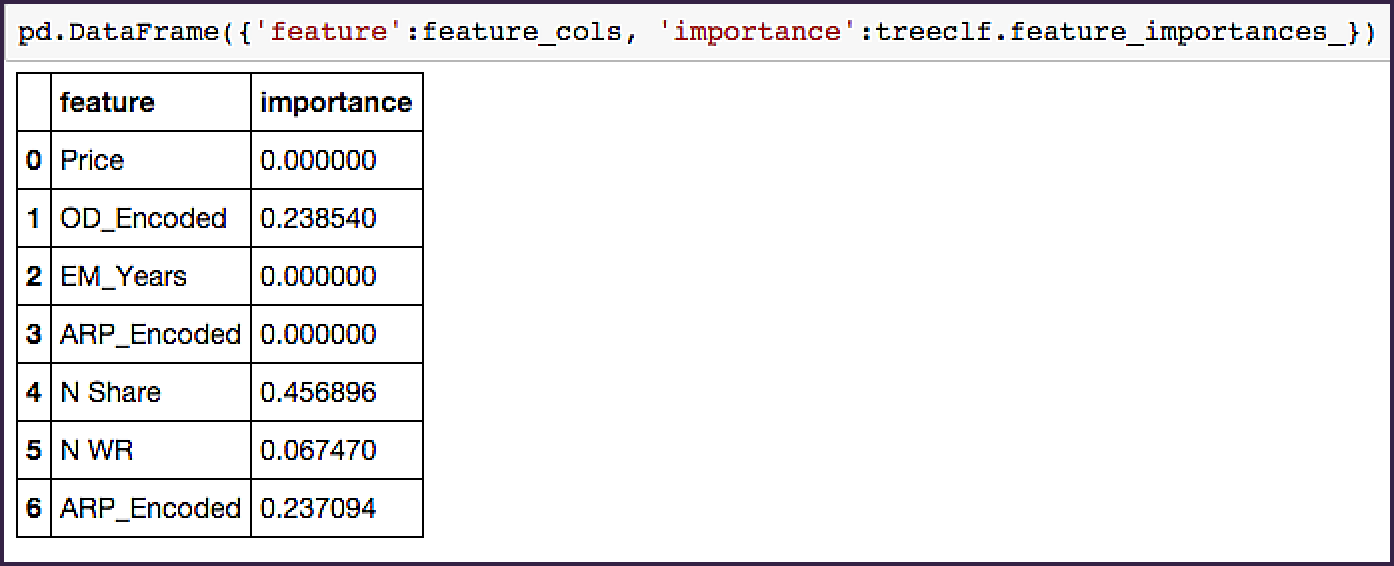
✓

Logistic Regression

3

DATA MODELING

Price paid for product, Alumni Status and whether they were a partner were not significant features



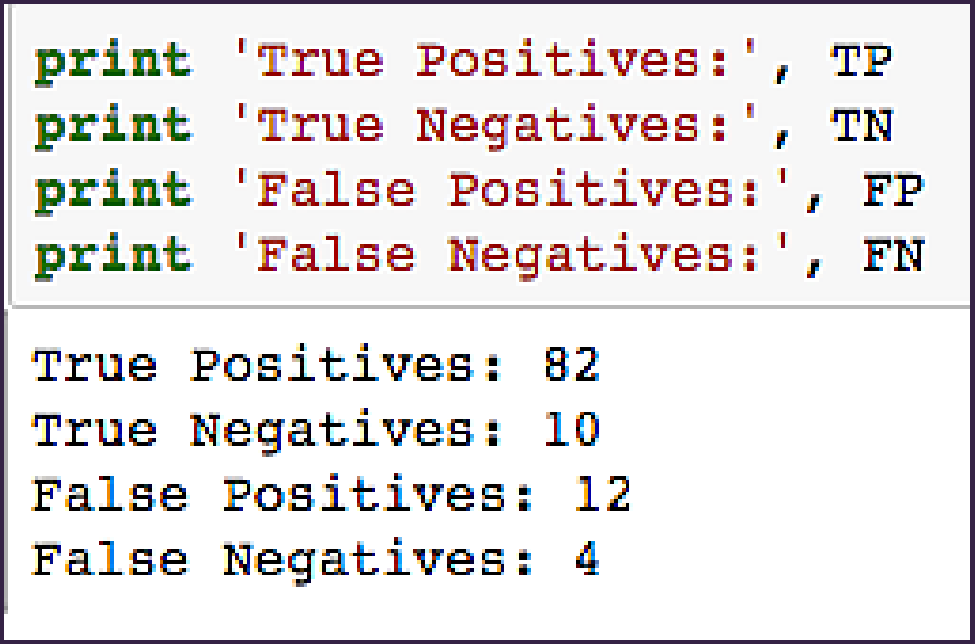
```
pd.DataFrame({'feature':feature_cols, 'importance':treeclf.feature_importances_})
```

	feature	importance
0	Price	0.000000
1	OD_Encoded	0.238540
2	EM_Years	0.000000
3	ARP_Encoded	0.000000
4	N Share	0.456896
5	N WR	0.067470
6	ARP_Encoded	0.237094

However, whether they were an MD or OD was important

3

DATA MODELING



```
print 'True Positives:', TP
print 'True Negatives:', TN
print 'False Positives:', FP
print 'False Negatives:', FN
```

```
True Positives: 82
True Negatives: 10
False Positives: 12
False Negatives: 4
```

What accounts would we miss?

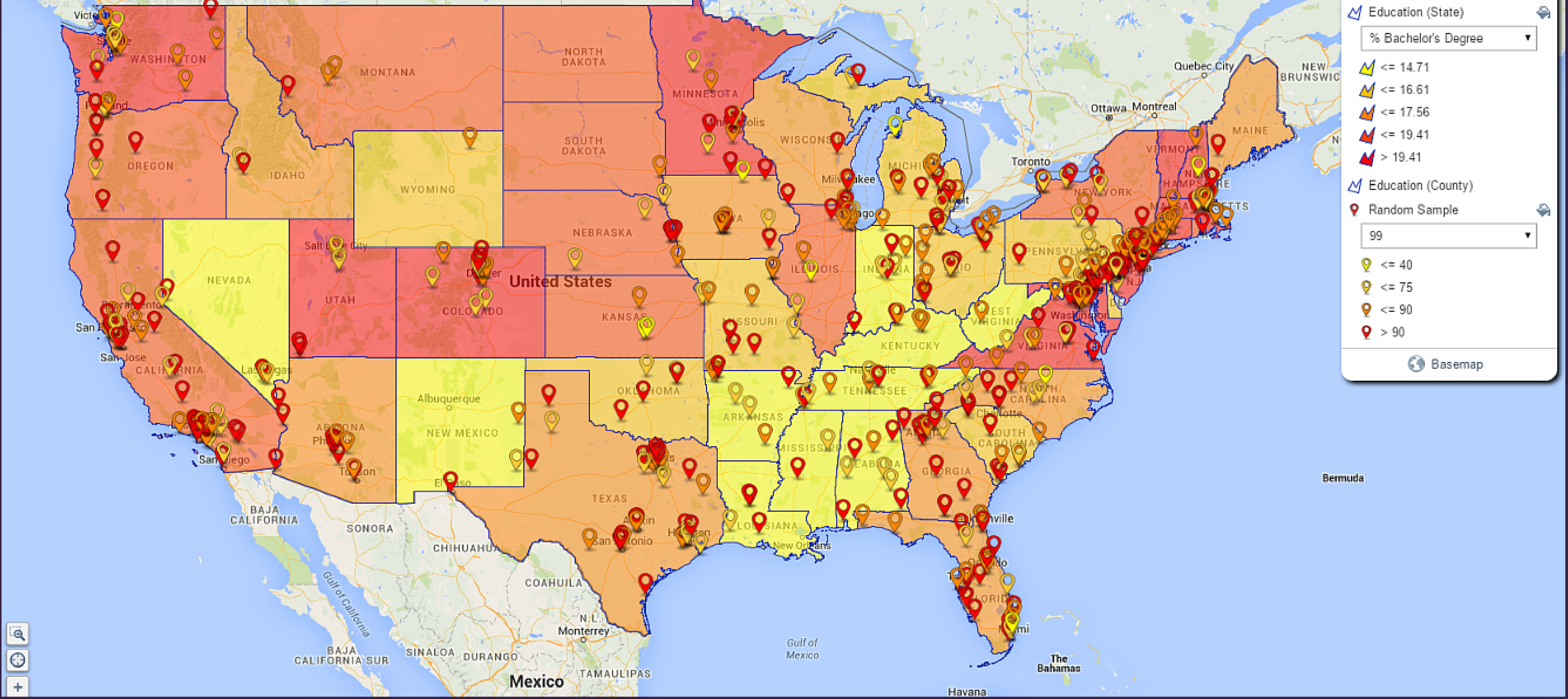
4

VISUALIZATIONS

Previously, I wanted to use affluence as a feature for predictive analysis

But...

Decided to use it for visualization solely instead



Education (State)

% Bachelor's Degree ▾

- <= 14.71
- <= 16.61
- <= 17.56
- <= 19.41
- > 19.41

Education (County)

Random Sample

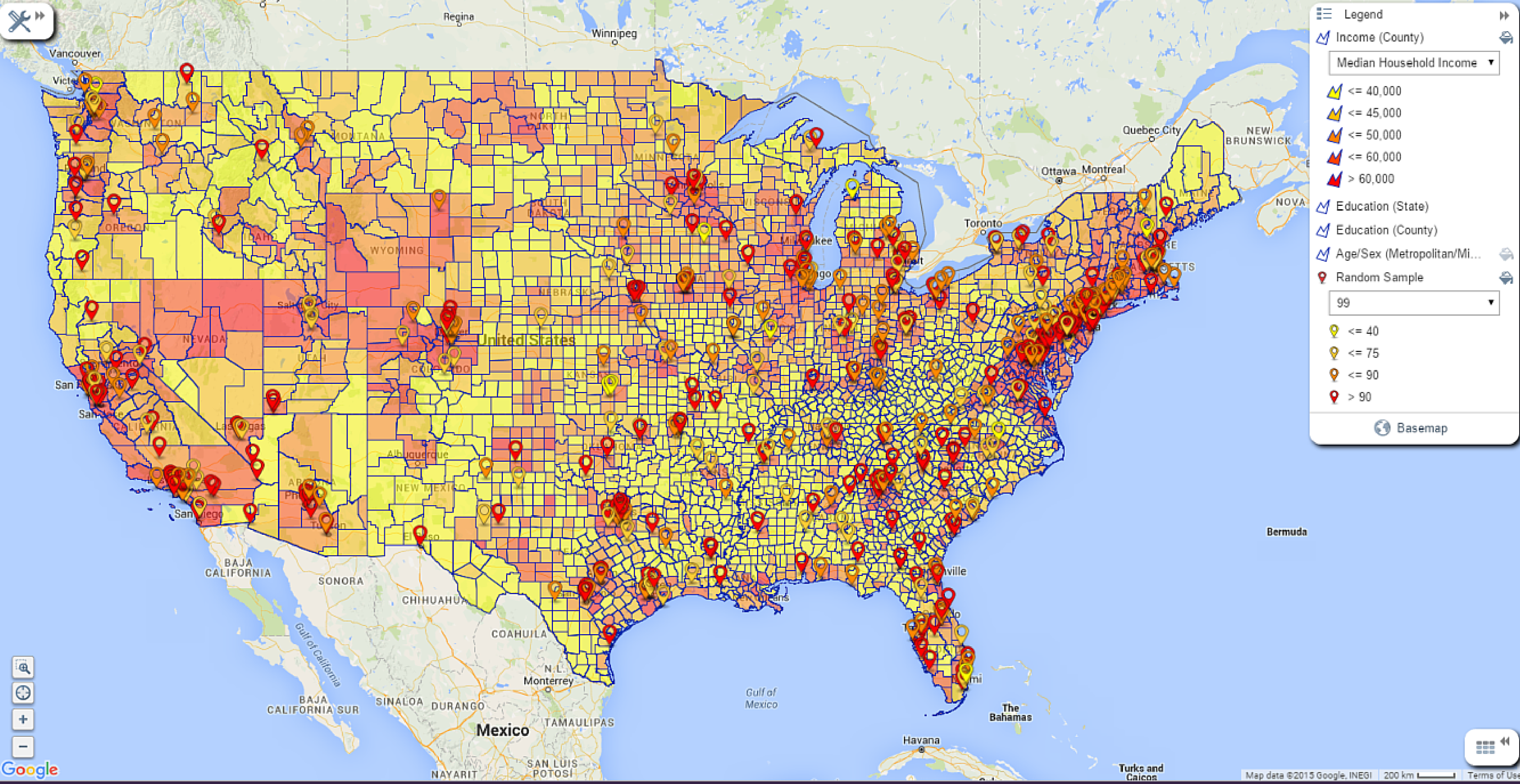
99 ▾

- <= 40
- <= 75
- <= 90
- > 90

Basemap

Account Holders in the US and Bachelor's Degrees

Suggests that our happier customers are in areas where residents are better educated.



Account Holders in the US and Median Household Income

4

Whats Next?

MORE
FEATURES!

Natural Language Processing

Webscraping

Share Tweets