FINDING THE CORRECT AUDIENCE ONLINE

General Assembly Data Science Course Final Project by Yuani Chen . 5 August 2017

Every Consumer is Unique

Audience of ONE

Programmatic marketing leverages rich, detailed insights allowing you to know your distinctive consumer. This knowledge informs strategy, implementation, optimization and a smarter way to spend on Ads.

Why does it matter?

Not Knowing Your Audience is a Big Problem

Fragmented view of data

Targeting the **WRONG** person
at the **WRONG** time

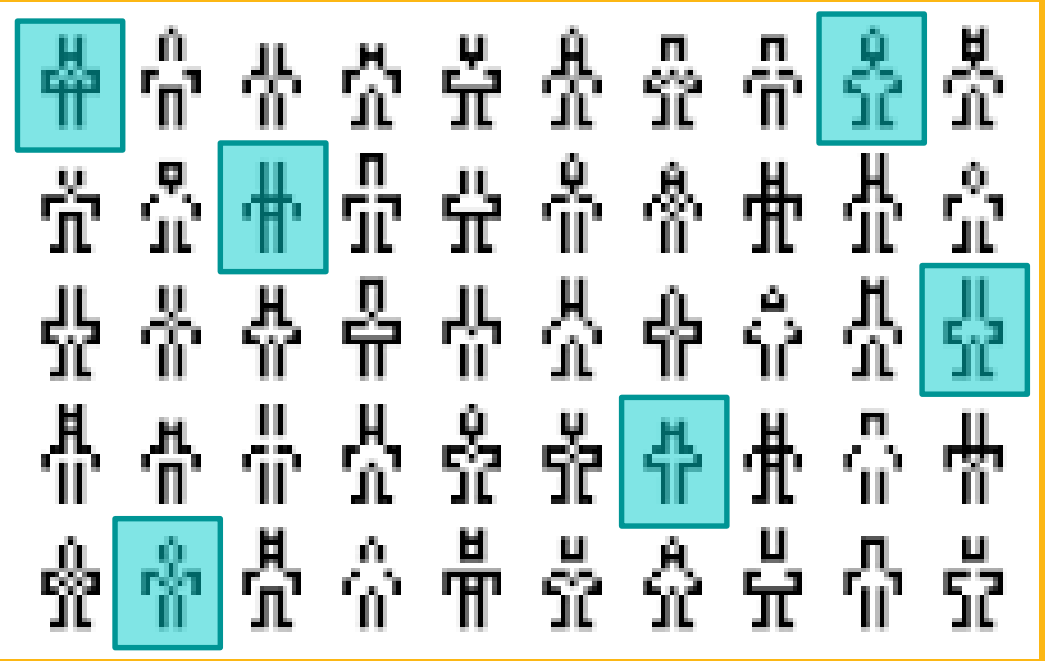
Wasted Ad Spend

Lower Returns on Investment (ROI)

3rd Party Data Providers profile users online providing Advertisers with richer Data

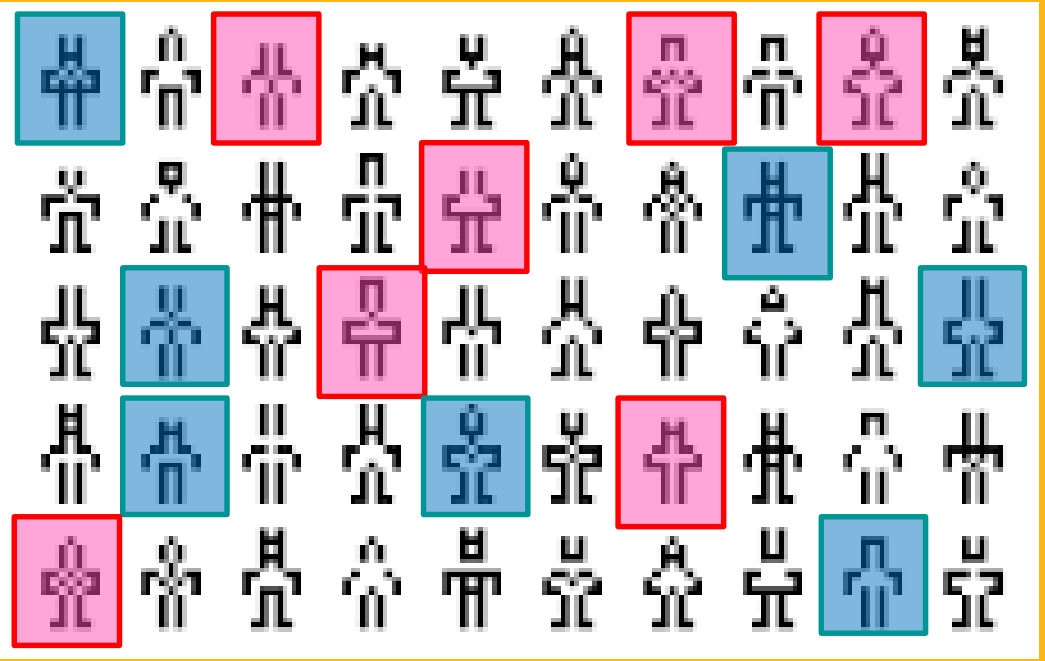
CAN WE DIFFERENTIATE:

HIGH POTENTIAL CUSTOMERS VS LOW POTENTIAL CUSTOMERS?



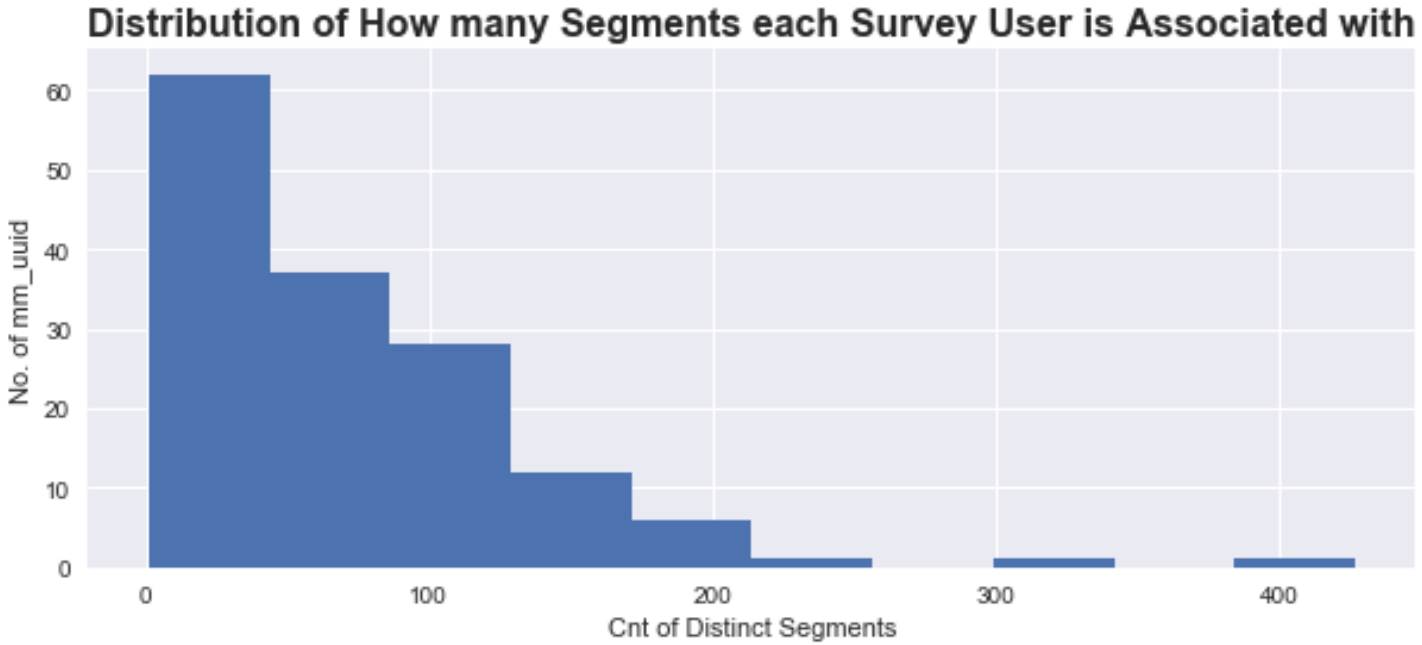
CAN WE DIFFERENTIATE:

MALE VS FEMALE USERS

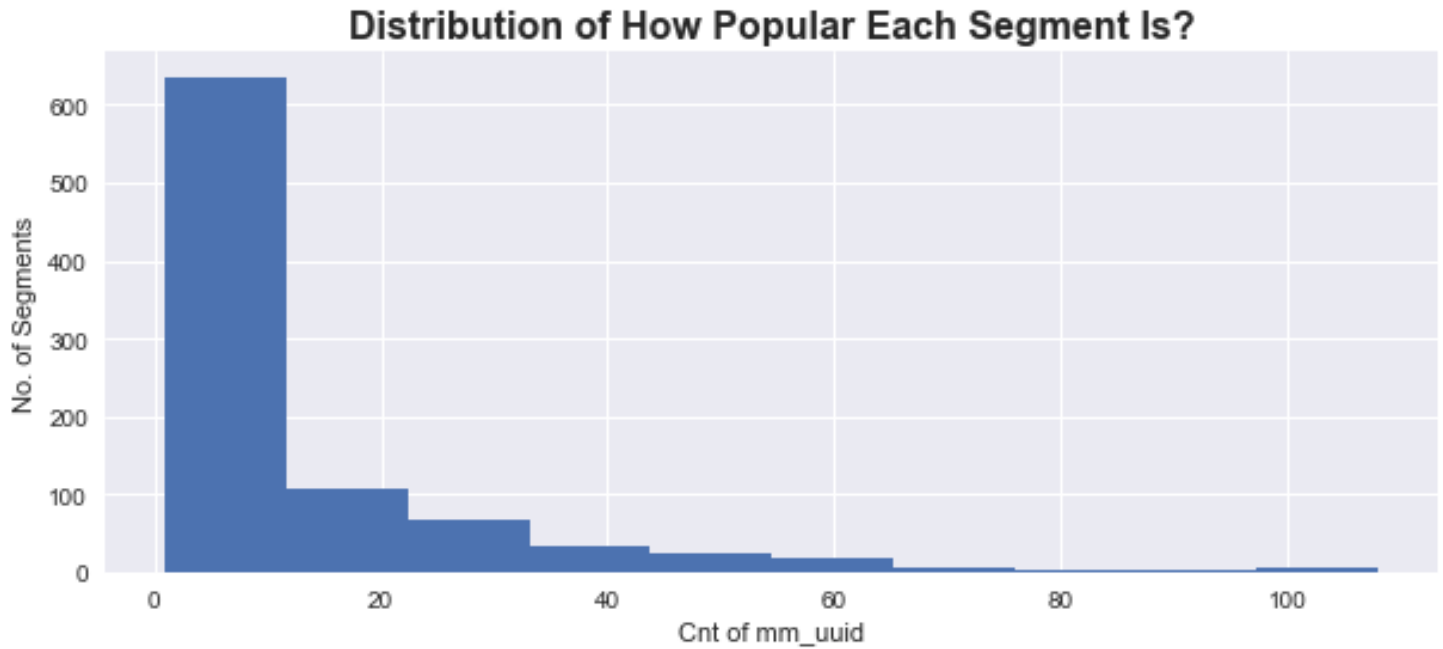


HOW WELL CAN WE SEGMENT USERS USING ONLINE ATTRITBUTES?

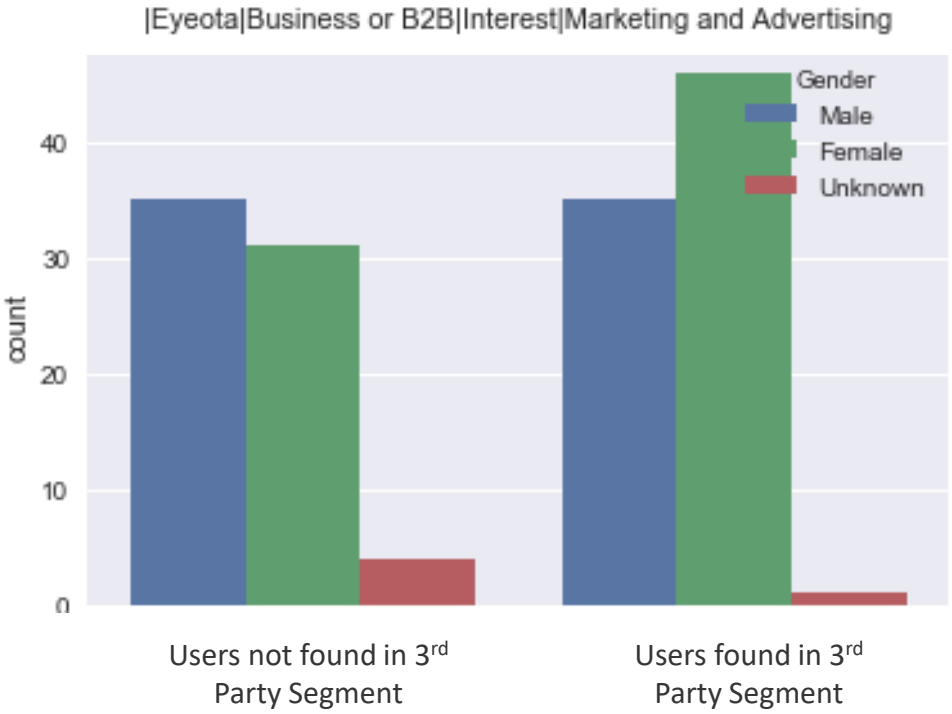
ANALYZED 148 SURVEY USERS (F: 77 ; M: 70 ; ?: 1)



899 SEGMENTS PROVIDED BY 3rd PARTY DATA COMPANIES



MODEL FEATURES



Example of a Model Feature: Users profiled in Marketing & Advertising by 3rd Party Audiences are more likely to be Female.

CORRELATION TO GENDER

	Segment	Corr	Corr_Abs
899	Gender_1_0	1.000000	1.000000
777	|Eyeota|Interest|Food Enthusiasts|Cocktails and Drinks	0.286143	0.286143
892	|Eyeota|Seasonal|Sport Events|Football Matches|FIFA World Cup 2018	-0.262948	0.262948
805	|Eyeota|Interest|Sports|Football	-0.262948	0.262948
531	|Eyeota|Branded Data|Global excluding US and EMEA ShareThis|Food and Drink|Beverages|Coffee and Tea	0.259608	0.259608
781	|Eyeota|Interest|Food Enthusiasts|Vegetarians and Vegans	0.249004	0.249004
711	|Eyeota|Intent|Shopping|CPG / FMCG|Grocery	0.247273	0.247273
857	|Eyeota|Regional Taxonomies|Southeast Asia (SEA)|Interest|Sports|Football	-0.246512	0.246512
778	|Eyeota|Interest|Food Enthusiasts|Cooking	0.240786	0.240786
710	|Eyeota|Intent|Shopping|CPG / FMCG	0.223172	0.223172

BUILDING A MODEL

ANALYZED 148 SURVEY USERS

Survey Data

	aac_uuid	ccs_uid	Gender
0	e09d5775-1aa2-4f00-af1b-3b0383b6a6e7	SGP16200003077	Female
1	bfa45769-3bcd-4000-96c8-e32efea73c5c	SGP16200003092	Male
2	ef5b573b-e9b6-4400-a9ab-ec12a0da6d25	SGP16200003071	Male
3	e15c55b4-98f3-4600-866d-496c962744bb	SGP16200003046	Female
4	7a175753-c3b9-4d00-a599-059eb41a860f	SGP16200003065	Female

3rd Party Data

	mm_uuid	provider	segment id	full path	date
17228	ff7756d5-73d8-4e00-a738-ccdfbed7e3be	ey	391	|Eyeota|Interest|Entertainment|Games and Toys	20/5/2017
17229	ff7756d5-73d8-4e00-a738-ccdfbed7e3be	ey	21949	|Eyeota|Interest|Entertainment|Music	20/5/2017
17230	ff7756d5-73d8-4e00-a738-ccdfbed7e3be	ey	459	|Eyeota|Interest|Sports|Football	20/5/2017
17231	ff7756d5-73d8-4e00-a738-ccdfbed7e3be	ey	21738	|Eyeota|Regional Taxonomies|Southeast Asia (SE...	20/5/2017
17232	ff7756d5-73d8-4e00-a738-ccdfbed7e3be	ey	21946	|Eyeota|Interest|Culture and Arts	20/5/2017

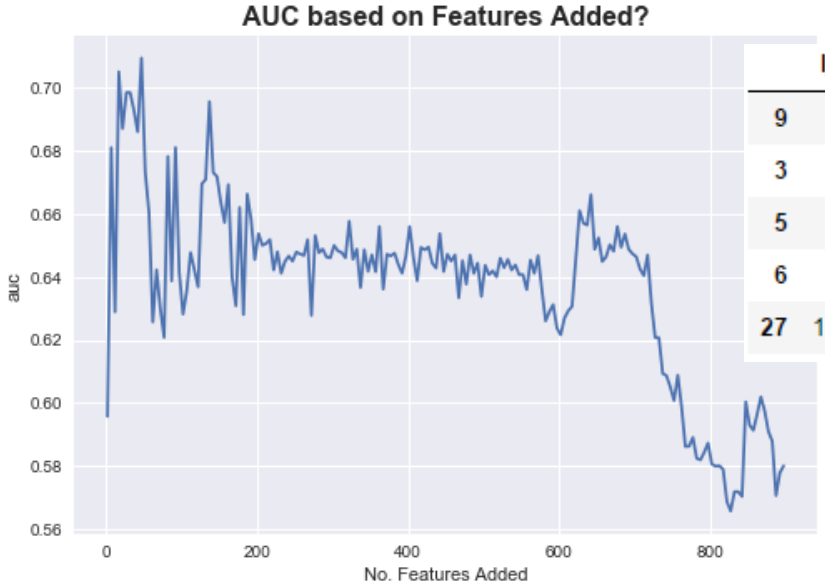
MODEL FEATURES:

- 899 Audience Segments
- Most popular Segments:

|Eyeota|Business or B2B	108
|Eyeota|Branded Data|Global Bombora|B2B|Professional Groups|Business Professional	102
|Eyeota|Interest|Tech Enthusiasts	101
|Eyeota|Business or B2B|Professions|Business Professionals	100
|Eyeota|Regional Taxonomies|Southeast Asia (SEA)|Interest|Tech Enthusiasts	97
|Eyeota|Business or B2B|Interest|Marketing and Advertising	81
|Eyeota|Interest|Entertainment	79
|Eyeota|Regional Taxonomies|Southeast Asia (SEA)|Interest|Arts and Entertainment	77
|Eyeota|Business or B2B|Interest|IT and Technology	74
|Eyeota|Branded Data|Global Bombora|B2B|Install Data|Web-Oriented Architecture|Web and Portal Technology	73
|Eyeota|Demographic|Employment Industry|HR and Recruitment	72
|Eyeota|Business or B2B|Interest|HR and Recruitment	69
|Eyeota|Branded Data|Global Bombora|B2B|Professional Groups|IT Professional	67
|Eyeota|Branded Data|Global Bombora|B2B|B2B Predictive Signals|Technology	66
|Eyeota|Demographic|Employment Industry|IT	65

MODEL 1: LOGISTIC REGRESSION

- Using a ranked list of features from the correlation between Gender & each segment, iterate through logistic model.



	l_n	l_auc
9	47	0.709437
3	17	0.705065
5	27	0.698467
6	32	0.698467
27	137	0.695588

Segment	Corr	Corr_Abs
Gender_1_0	1.000000	1.000000
|Eyeota|Interest|Food Enthusiasts|Cocktails an...	0.286143	0.286143
|Eyeota|Seasonal|Sport Events|Football Matches...	-0.262948	0.262948
|Eyeota|Interest|Sports|Football	-0.262948	0.262948
|Eyeota|Branded Data|Global excluding US and E...	0.259608	0.259608
|Eyeota|Interest|Food Enthusiasts|Vegetarians ...	0.249004	0.249004
|Eyeota|Intent|Shopping|CPG / FMCG|Grocery	0.247273	0.247273
|Eyeota|Regional Taxonomies|Southeast Asia (SE...	-0.246512	0.246512
|Eyeota|Interest|Food Enthusiasts|Cooking	0.240786	0.240786
|Eyeota|Intent|Shopping|CPG / FMCG	0.223172	0.223172

MODEL 2: RANDOM FOREST

- Ranking of 899 features
- Top Features:

	ntrees	depth	auc
56	121	3	0.639556
55	121	2	0.638543
37	81	2	0.635630
64	141	2	0.634293
65	141	3	0.631404

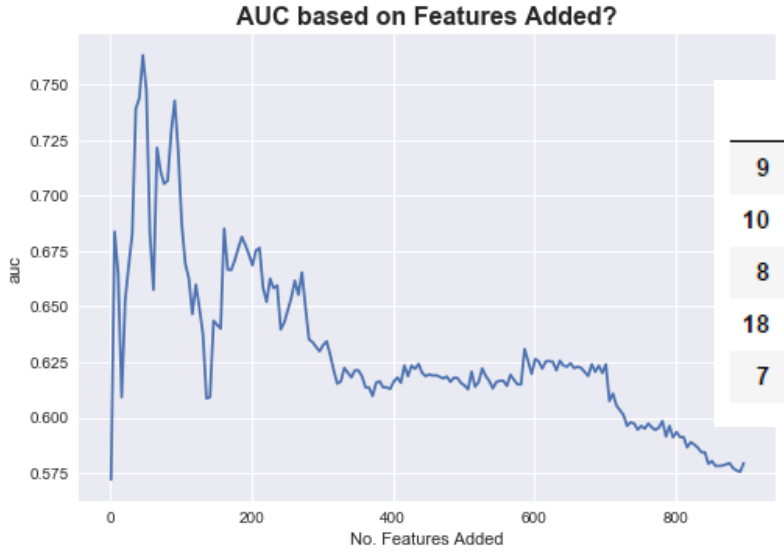
```
model = RandomForestClassifier(n_estimators = 121, max_depth=3, random_state=50)
X = data[data.columns[3:902]]
y = data['Gender_1_0']
model.fit(X, y)
scores = cross_val_score(model, X, y, scoring='roc_auc')
print('CV AUC {}, Average AUC {}'.format(scores, scores.mean()))
```

```
CV AUC [ 0.53525641  0.67558528  0.70782609], Average AUC 0.6395559271646228
```

	Features	Importance Score
805	|Eyeota|Interest|Sports|Football	0.029144
898	segment_cnt	0.020877
777	|Eyeota|Interest|Food Enthusiasts|Cocktails an...	0.019256
781	|Eyeota|Interest|Food Enthusiasts|Vegetarians ...	0.015991
857	|Eyeota|Regional Taxonomies|Southeast Asia (SE...	0.015597

MODEL 3: LOGISTIC REGRESSION

- Using a ranked list of features from the decision tree feature importance derived in Model 2, iterate through logistic model.



	l_n	l_auc
9	46	0.763189
10	51	0.747230
8	41	0.743830
18	91	0.742703
7	36	0.738993