

Name: Joshua Chang

## Assignment 4: Feature Selection

1. Executed Notebook (attached)
2. The list of final variables along with their univariate KS's:

wrapper order		variable	filter score
0	1	max_count_by_address_30	0.359215
1	2	max_count_by_ssn_dob_7	0.228401
2	3	max_count_by_homephone_3	0.224757
3	4	max_count_by_fulladdress_30	0.359914
4	5	zip5_count_3	0.224706
5	6	max_count_by_ssn_dob_30	0.240836
6	7	max_count_by_homephone_7	0.232235
7	8	fulladdress_count_0_by_30	0.290722
8	9	max_count_by_fulladdress_homephone_30	0.249724
9	10	ssn_dob_day_since	0.228626
10	11	max_count_by_address_7	0.343335
11	12	address_day_since	0.334140
12	13	fulladdress_day_since	0.333269
13	14	max_count_by_fulladdress_3	0.329538
14	15	max_count_by_address_3	0.329445
15	16	address_count_14	0.322436
16	17	fulladdress_count_14	0.321953
17	18	max_count_by_address_1	0.315332
18	19	max_count_by_fulladdress_1	0.315253
19	20	address_count_7	0.301735

3. The plot showing model performance vs # variables: where num\_filter = 240 and num\_wrapper = 20

