id	age	sex	region	income	married	children	car	save_act	current_act	mortgage	pep
ID12101	48	FEMALE	INNER_CITY	17546.00	NO	1	NO	NO	NO	NO	YES
ID12102	40	MALE	TOWN	30085.10	YES	3	YES	NO	YES	YES	NO
ID12103	51	FEMALE	INNER_CITY	16575.40	YES	0	YES	YES	YES	NO	NO
ID12104	23	FEMALE	TOWN	20375.40	YES	3	NO	NO	YES	NO	NO
ID12105	57	FEMALE	RURAL	50576.30	YES	0	NO	YES	NO	NO	NO
ID12106	57	FEMALE	TOWN	37869.60	YES	2	NO	YES	YES	NO	YES
ID12107	22	MALE	RURAL	8877.07	NO	0	NO	NO	YES	NO	YES
ID12108	58	MALE	TOWN	24946.60	YES	0	YES	YES	YES	NO	NO
ID12109	37	FEMALE	SUBURBAN	25304.30	YES	2	YES	NO	NO	NO	NO
ID12110	54	MALE	TOWN	24212.10	YES	2	YES	YES	YES	NO	NO

Table 1: Un-transformed data snapshot

age	sex	region	income	married	children	car	save_act	current_act	mortgage	pep
forties	FEMALE	INNER_CITY	tier 1	single	1 kid	no car	no savings	not current	no mortgage	pep
thirties	MALE	TOWN	tier 2	married	3 kids	car	no savings	current	mortgage	no pep
fifties	FEMALE	INNER_CITY	tier 1	married	no kids	car	savings	current	no mortgage	no pep
twenties	FEMALE	TOWN	tier 1	married	3 kids	no car	no savings	current	no mortgage	no pep
fifties	FEMALE	RURAL	tier 3	married	no kids	no car	savings	not current	no mortgage	no pep
fifties	FEMALE	TOWN	tier 2	married	2 kids	no car	savings	current	no mortgage	pep
twenties	MALE	RURAL	tier 1	single	no kids	no car	no savings	current	no mortgage	pep
fifties	MALE	TOWN	tier 2	married	no kids	car	savings	current	no mortgage	no pep
thirties	FEMALE	SUBURBAN	tier 2	married	2 kids	car	no savings	not current	no mortgage	no pep
fifties	MALE	TOWN	tier 1	married	2 kids	car	savings	current	no mortgage	no pep

Table 2: Transformed data snapshot

$$m = \min(\text{income}) = \$5,014.21$$

$$M = \max(\text{income}) = \$63,130.10$$

$$\text{group size} = \frac{M - m}{3} =$$

$$\frac{\$63,130.10 - \$5,014.21}{3} = \$19,371.96$$

$$\text{Supp}(X, Y) = P(X, Y) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$

$$\text{Conf}(X, Y) = P(Y|X) = \frac{\text{Supp}(X, Y)}{\text{Supp}(X)}$$

$$\text{Lift}(X, Y) = \frac{\text{Supp}(X, Y)}{\text{Supp}(X) \cdot \text{Supp}(Y)}$$

$$0 \to \text{Zero}$$

$$1 \to \text{One}$$

$$\vdots$$

$$9 \to \text{Nine}$$

Let A and B be events. Then

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Confidence \ Support	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
0.9	0	0	0	0	0	0	0	945	22258
0.8	0	0	0	0	0	0	0	945	23169
0.7	0	0	0	0	0	0	0	945	25589
0.6	0	0	0	0	0	0	0	945	25617
0.5	0	0	0	0	0	0	0	945	25618
0.4	0	0	0	0	0	0	0	945	25618
0.3	0	0	0	0	0	0	0	945	25618
0.2	0	0	0	0	0	0	0	945	25618
0.1	0	0	0	0	0	0	0	945	25618

Table 3: Number of rules with varying support and confidence

Confidence \ Support	0.5	0.4	0.3	0.2	0.1
0.9	0	0	0	0	0
0.8	0	0	0	0	6
0.7	0	0	0	0	9
0.6	0	0	0	0	9
0.5	0	0	0	1	21
0.4	0	0	2	12	76
0.3	0	0	2	13	93
0.2	0	0	2	13	93
0.1	0	0	2	13	93

Table 4: Number of rules implying PEP purchase

Linear SVN prediction \ Actual values	0	1	2	3	4	5	6	7	8	9
0	**19**	0	0	0	0	0	0	0	0	0
1	0	**31**	1	2	0	1	1	0	2	0
2	0	0	**24**	0	0	0	0	0	1	0
3	0	0	0	**29**	0	0	0	0	0	0
4	0	0	0	0	**22**	0	0	0	0	1
5	1	0	0	1	0	**27**	2	0	3	0
6	0	0	1	1	0	0	**33**	0	1	0
7	0	0	0	0	0	0	0	**22**	0	1
8	0	0	0	2	0	0	0	0	**19**	0
9	0	0	1	0	0	0	0	1	1	**29**

Matching predicted values

	0	1	2	3	4	5	6	7	8	9
Count	75	92	10	18	16	0	45	15	27	55

If $\mathbf{X} = \{X_1, X_2, \ldots, X_d\}$, then

$$P(\mathbf{X}|Y = y) = \prod_{i=1}^{d} P(X_i|Y = y)$$

$$P(Y|\mathbf{X}) = \frac{P(Y)\prod_{i=1}^{d} P(X_i|Y)}{P(\mathbf{X})}$$

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

For $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$,

$$x_i \rightarrow \frac{x_i}{\|\mathbf{x}\|}$$

€10
Accuracy: 79.1%

UW \ Cluster	1	2	3	4
No	144	14	0	0
Yes	15	28	10	4

	min	max	range	median	mean	var	std.dev
TotalPop	2208.00	45986.00	43778.00	17931.00	19925.07	110016354.41	10488.87
Men	1100.00	22691.00	21591.00	8801.00	9881.68	26155736.77	5114.27
Women	1108.00	23295.00	22187.00	9025.00	10043.39	29073515.88	5391.99
Hispanic	0.00	5316.00	5316.00	322.00	499.74	428665.84	654.73
White	2188.00	41437.00	39249.00	16542.00	18066.39	87180578.76	9337.05
Black	0.00	10986.00	10986.00	391.00	929.83	2861883.17	1691.71
Native	0.00	364.00	364.00	24.00	38.55	2205.81	46.97
Asian	0.00	843.00	843.00	34.00	67.71	9426.39	97.09
Pacific	0.00	166.00	166.00	0.00	5.39	319.66	17.88
Citizen	1791.00	34520.00	32729.00	13745.00	15240.73	62401475.08	7899.46
Income	20306.09	63618.55	43312.45	36810.82	37265.64	56620845.69	7524.68
IncomePerCap	10879.72	30490.00	19610.28	19285.43	19694.04	10191312.35	3192.38
Poverty	486.00	11615.00	11129.00	3995.00	4308.85	5395673.10	2322.86
ChildPoverty	905.00	16537.00	15632.00	5624.00	5861.36	9937319.37	3152.35
Professional	693.00	15346.00	14653.00	4661.00	5351.50	9385950.79	3063.65
Service	431.00	8486.00	8055.00	3020.00	3428.01	3392649.51	1841.91
Office	250.00	10644.00	10394.00	3958.00	4395.09	6337444.95	2517.43
Construction	342.00	6223.00	5881.00	2399.00	2511.06	1568738.27	1252.49
Production	367.00	12373.00	12006.00	3670.00	4239.73	5815386.57	2411.51
Drive	1718.00	39163.00	37445.00	14669.00	16609.08	78997823.76	8888.07
Carpool	333.00	6301.00	5968.00	1878.00	2094.52	1429394.84	1195.57
Transit	0.00	434.00	434.00	32.00	54.92	4583.96	67.70
Walk	6.00	2515.00	2509.00	217.00	328.33	121127.95	348.03
OtherTransp	0.00	1072.00	1072.00	164.00	239.71	37956.55	194.82
WorkAtHome	15.00	2162.00	2147.00	484.00	599.17	169245.93	411.40
MeanCommute	17.18	37.68	20.50	26.37	26.56	19.15	4.38
Employed	806.00	21103.00	20297.00	6792.00	7662.43	20130162.68	4486.66
PrivateWork	1731.00	37088.00	35357.00	13502.00	15050.57	67956720.96	8243.59
PublicWork	406.00	8740.00	8334.00	2978.00	3343.20	2999311.09	1731.85
SelfEmployed	71.00	4807.00	4736.00	1321.00	1487.05	715025.04	845.59
FamilyWork	0.00	314.00	314.00	26.00	44.65	2957.10	54.38
Unemployment	148.00	5709.00	5561.00	1754.00	1967.86	1296715.31	1138.73

consider	dining	dishes	else	good	high	know	least
0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042
0	0	0.03	0	0	0	0	0
0	0	0.1	0	0	0	0	0
0	0	0	0	0.1	0	0	0.05
0	0	0	0	0	0	0	0
0	0	0	0	0	0.021	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

	No	Yes
No	**27**	2
Yes	6	**8**

	No	Yes
No	**21**	2
Yes	8	**12**

	No	Yes
No	**35**	4
Yes	1	**3**

	N	P
N	**11**	2
P	2	**3**

	N	P
N	**7**	2
P	1	**8**

	N	P
N	**10**	0
P	0	**8**

Model	Method	Accuracy	Label	Recall	Precision	$F1$
Model 1	Naive Bayes	0.86	No	0.86	0.97	0.91
			Yes	0.83	0.50	0.63
	Decision Tree	0.81	No	0.93	0.82	0.87
			Yes	0.57	0.80	0.67
	SVM	0.88	No	0.87	1.00	0.93
			Yes	1.00	0.50	0.67
	$k$nn	0.88	No	0.91	0.94	0.93
			Yes	0.78	0.70	0.74
	Random Froest	0.79	No	0.90	0.82	0.86
			Yes	0.54	0.70	0.61
Model 2	Naive Bayes	0.81	No	0.80	0.97	0.88
			Yes	0.88	0.50	0.64
	Decision Tree	0.77	No	0.91	0.72	0.81
			Yes	0.60	0.86	0.71
	SVM	0.86	No	0.90	0.90	0.90
			Yes	0.79	0.79	0.79
	$k$nn	0.88	No	0.88	0.97	0.92
			Yes	0.91	0.71	0.80
	Random Froest	0.86	No	0.90	0.90	0.90
			Yes	0.79	0.79	0.79
Model 3	Naive Bayes	0.88	No	0.90	0.97	0.93
			Yes	0.43	0.75	0.55
	Decision Tree	0.88	No	0.90	0.97	0.93
			Yes	0.43	0.75	0.55
	SVM	0.86	No	0.86	1.00	0.92
			Yes	1.00	0.14	0.25
	$k$nn	0.91	No	0.92	0.97	0.95
			Yes	0.80	0.57	0.66
	Random Froest	0.88	No	0.94	0.92	0.93
			Yes	0.63	0.71	0.67

Method	Model Used	Accuracy	Label	Recall	Precision	$F1$
Naive Bayes	Model 3	0.54	No	0.77	0.40	0.53
			Yes	0.42	0.79	0.55
Decision Tree	Model 3	0.57	No	0.42	0.74	0.53
			Yes	0.79	0.48	0.60
SVM	Model 1	0.51	No	0.39	0.84	0.54
			Yes	0.81	0.35	0.49
$k$nn	Model 3	0.44	No	0.36	0.84	0.50
			Yes	0.75	0.24	0.37
Random Froest	Model 3	0.59	No	0.44	0.77	0.56
			Yes	0.82	0.50	0.62

CensusTract	State	County	TotalPop	Men	Women
Hispanic	White	Black	Native	Asian	Pacific
Citizen	Income	IncomeErr	IncomePerCap	IncomePerCapErr	Poverty
ChildPoverty	Professional	Service	Office	Construction	Production
Drive	Carpool	Transit	Walk	OtherTransp	WorkAtHome
MeanCommute	Employed	PrivateWork	PublicWork	SelfEmployed	FamilyWork
Unemployment					

	min	max	range	median	mean	std.dev
TotalPop	5097	933267	928170	31454	69164	128384
Men	2505	443176	440671	15448	33675	61475
Women	2592	490091	487499	16188	35489	66919
Hispanic	4	65127	65123	768	3353	9142
White	4959	373598	368639	26624	52029	71405
Black	28	488999	488971	1022	11258	53375
Native	0	1830	1830	76	152	256
Asian	0	23264	23264	120	1060	3481
Pacific	0	769	769	0	31	99
Citizen	4138	658353	654215	24156	51423	91952
Income	26924	100922	73998	38844	41079	9920
IncomePerCap	12810	44671	31861	20721	21485	4233
Poverty	679	200484	199805	6048	12319	24804
ChildPoverty	1249	292883	291634	8354	17228	35615
Professional	1530	307794	306264	8016	22509	46391
Service	792	172226	171434	5455	12057	22889
Office	982	236838	235856	6868	17087	33167
Construction	506	64346	63840	3520	6314	9016
Production	800	152137	151337	6840	11199	18051
Drive	4378	764379	760001	26906	57847	105048
Carpool	338	98465	98127	2852	6515	12856
Transit	0	17894	17894	62	603	2573
Walk	21	13300	13279	350	920	2089
OtherTransp	0	14798	14798	372	859	1833
WorkAtHome	115	30213	30098	924	2421	4806
MeanCommute	17	37	20	26	26	4
Employed	1840	425283	423443	12534	30752	61756
PrivateWork	3277	754970	751693	24254	54501	104155
PublicWork	911	129780	128869	5042	9796	17090
SelfEmployed	445	47782	47337	2610	4755	7402
FamilyWork	0	771	771	42	114	156
Unemployment	250	101708	101458	3198	6039	11961

	min	max	range	median	mean	std.dev
TotalPop	156752	933267	776515	294000	393438	269458
Men	75894	443176	367282	144768	190273	128292
Women	80858	490091	409233	149231	203166	141190
Hispanic	2590	65127	62537	17149	22089	19258
White	121655	525414	403759	215494	264132	140217
Black	2977	488999	486022	37973	88250	130505
Native	236	1830	1594	521	691	452
Asian	908	23264	22356	7742	9464	7791
Pacific	25	769	744	125	181	189
Citizen	120604	658353	537749	211166	287378	193210
Income	42512	100922	58410	54236	56826	13756
IncomePerCap	22914	44671	21758	28256	28513	5119
Poverty	10150	200484	190334	46043	66581	56258
ChildPoverty	13000	292883	279883	56142	91898	80991
Professional	49755	307794	258039	118199	143229	93812
Service	21747	172226	150479	51005	67726	48861
Office	42359	236838	194479	73426	99883	67154
Construction	7476	64346	56870	21806	28104	17990
Production	11222	152137	140915	35534	54501	44823
Drive	135560	764379	628819	239843	322906	216997
Carpool	13584	98465	84881	28122	35879	26866
Transit	319	24410	24091	4114	7714	9085
Walk	951	17288	16337	4801	7237	6218
OtherTransp	1347	14798	13451	3834	5590	4687
WorkAtHome	3778	30213	26435	13094	14105	8921
MeanCommute	20	28	8	22	23	3
Employed	66313	425283	358970	149090	187852	129631
PrivateWork	127449	754970	627521	233856	320084	224721
PublicWork	17215	129780	112565	47398	51163	33559
SelfEmployed	5209	47782	42573	17068	21684	13064
FamilyWork	155	1001	846	439	514	272
Unemployment	8383	101708	93325	20130	32837	27963