

INCREASE ONLINE EMR ACCESS



PRATHAMESH BAPAT

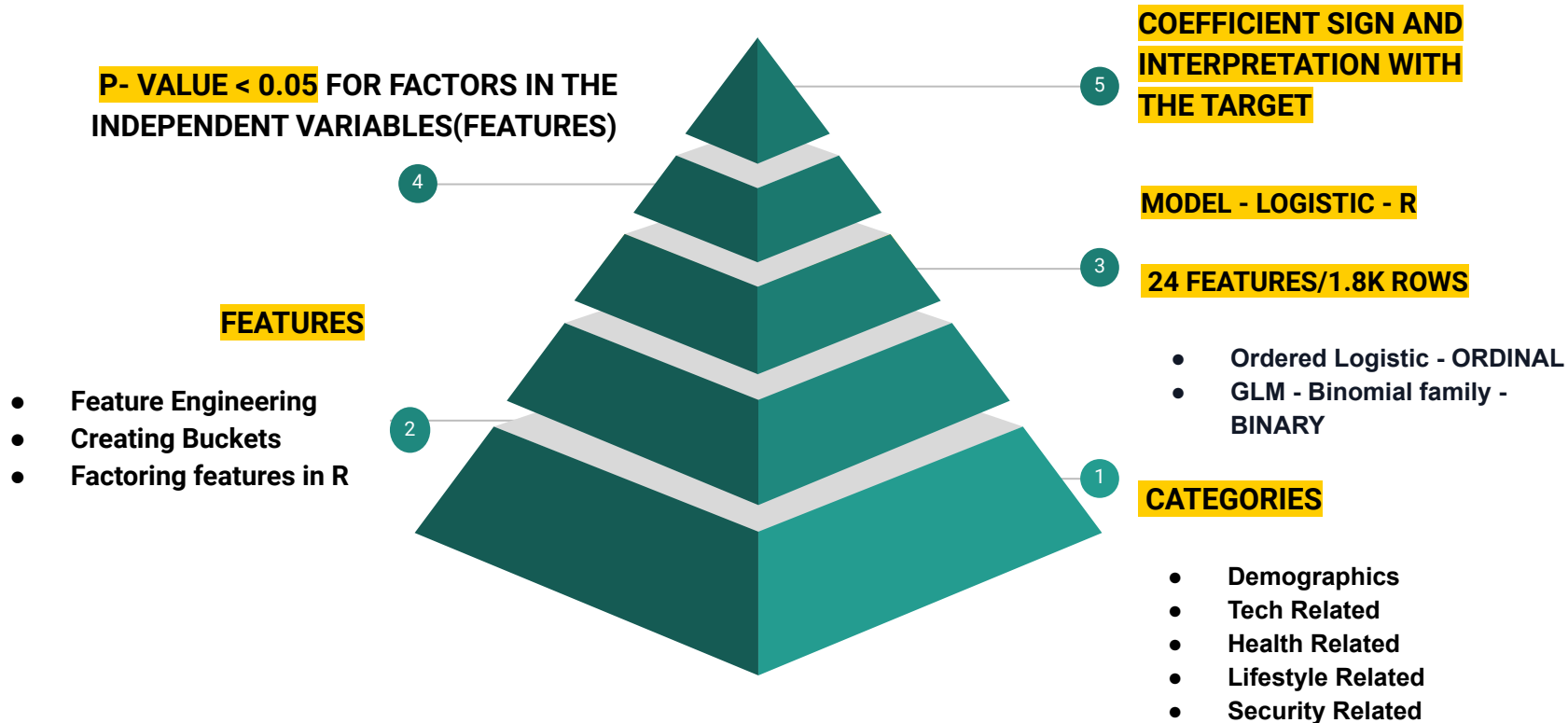
TARGET FEATURE D6

- Original => 0, 1 - (1 to 2 times), 2 - (3 to 5 times), 3 - (6-9 times), 4 - (10+ times)

LET'S CONVERT THEM TO TWO POSSIBLE WAYS

- BINARY => 0 - NO ACCESS, 1 - ACCESS
- ORDINAL => 0 - NO ACCESS → 1 - (1-5 times) → 2 - 6+ times

LET'S LOOK AT THE PROCESS



FEATURE ENGINEERING

GROUP	FEATURES	QUESTION	Feature Engineering
Demographics	age	O1	Random Integer Values - Divide into buckets <30, 30-45, 45-60 and 60>, Then factor them.
	occupationstatus	O2	-9,-5, 1-6,91 -- Factor them then convert the negatives to 91 and keep everything as it is.
	yActiveDutyArmedForces	O3	-9,1,2,3,4,5 -- Factor them then convert (1,2,3) to 1 and (4,5) to 0.
	MaritalStatus	O5	-9,-5,1-6 -- Factor them then convert (-9,-5) to 0.
	education	O6	-9,1-7 -- Factor them then convert -9 to 0, (1,2) - 1 and keep rest as it is.
	BornInUSA	O7	-9,1,2 -- Factor them then convert -9 to 1.
	SpeakEnglish	O9	-9,-5,1-4 -- Factor them then convert (-9,-5,3,4) to 0 and (1,2) to 1.
	NotHispanic, Mexican, PuertoRican, Cuban, OtherLatino	O10	Just use hisp_cat and convert hispanic/non hispanic.
	White, Black, AmerInd, AsInd, Chinese, Filipino	O11	Just use race_cap2 and convert to White/Black/Others.
	IncomeRanges	O17	-9,(1-9) -- Convert (1-4) to low, (5-6) to medium and (7-9) to High.
Tech Related	UseInternet	B1	-9,1,2 -- Convert -9 to 1.
	Electronic_SelfHealthInfo, Electronic_HealthInfo	B5	make new column if all answers to electronic are no, get 0 score in new column else 1
	OtherDevTrackHealth	B9	DROP THIS SINCE ONLY ONE VALUE AFTER DROPPING ROWS.
	IntRsn_VisitedSocNet, IntRsn_SharedSocNet,	B11	make new column if all answers to health conditions are no, get 0 score in new column
Health Related	EverOfferedAccessRec	D4	DROP THIS SINCE ONLY ONE VALUE AFTER DROPPING ROWS.
	mostroutinecheckup2	C2	-9,-5,(1-6) -- Convert 1 -recent, 2-3 - sometime and rest to longtime.
	FreqGoProvider	C3	-9,-5,0,1,2,3,4,5,6 -- Convert to buckets -9,0,1 --> "Not often", 2,3,4 --> "Somewhat often", 5,6 --> "Often"
	GeneralHealth	G1	-9,-5,1,2,3,4,5 -- Convert to buckets -9,-5,5 --> "poor", 4,3 --> Okay, 1,2 --> "great"
	MedConditions_Diabetes, MedConditions_HighBloodPressure	G3	Make new column if all answers to health conditions are no, get 0 score in new column
Lifestyle Related	TimesModerateExercise	I1	-9,0,1,2,3,4,5,6,7 -- -9,0,1 --> "rare", 2,3,4, --> "medium", 5,6,7 --> "Often"
	TimesStrengthTraining	I3	-9,-5,0,1,2,3,4,5,6,7 -- -9,-5,0,1 --> "rare", 2,3,4, --> "medium", 5,6,7 --> "Often"
	AverageTimeSitting	I4	-9,-5, 0-->20 -- -9,-5, 0-3 --> 4-7 --> 8-10 --> 10+ -->
Security Related	ConfidentInfoSafe	D2	-9,-5,1,2,3, -9,-5,2 -->2
	WithheldInfoPrivacy	D3	-9,1,2 -9,2 -->2

LOGISTIC - RESULTS - SUMMARY

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.73787	1.17061	-1.485	0.13765
occupationstatus1	0.82884	0.43885	1.889	0.05893 .
occupationstatus2	1.23098	0.57106	2.156	0.03111 *
occupationstatus3	0.68818	0.52382	1.314	0.18892
occupationstatus4	1.32643	0.71815	1.847	0.06475 .
occupationstatus5	0.85330	0.44599	1.913	0.05572 .
occupationstatus6	0.68388	0.50493	1.354	0.17561
activedutyarmedforces0	0.51414	0.47027	1.093	0.27427
activedutyarmedforces1	0.68356	0.50012	1.367	0.17169
maritalstatus1	-0.68091	0.82060	-0.830	0.40667
maritalstatus2	-0.60530	0.90806	-0.667	0.50504
maritalstatus3	-0.76353	0.82666	-0.924	0.35568
maritalstatus4	-0.37844	0.83937	-0.451	0.65209
maritalstatus5	-1.85435	0.94572	-1.961	0.04990 *
maritalstatus6	-0.54286	0.83234	-0.652	0.51426
education1	-1.17104	1.00442	-1.166	0.24366
education2	-0.49073	0.96196	-0.510	0.60996
education3	-0.72936	0.98011	-0.744	0.45678
education4	-0.60028	0.95659	-0.628	0.53032
education5	-0.39343	0.95586	-0.412	0.68064
education6	-0.63955	0.96037	-0.666	0.50545
borninusa2	0.18708	0.20248	0.924	0.35551
speakenglish1	0.02353	0.33924	0.069	0.94470
hisp_catNot Hispanic	-0.15561	0.27210	-0.572	0.56739
hisp_catHispanic	-0.10914	0.32766	-0.333	0.73908
race_cat2White	-0.11334	0.19103	-0.593	0.55298
race_cat2Black	0.20161	0.24131	0.835	0.40344
incomerangesLow	-0.56470	0.24822	-2.275	0.02291 *
incomerangesMedium	-0.28250	0.22771	-1.241	0.21475
incomerangesHigh	0.02568	0.22622	0.114	0.90963
useinternet2	-1.21018	0.23850	-5.074	0.000000389 ***
mostrecentcheckup2Sometime	-0.37560	0.16171	-2.323	0.02019 *

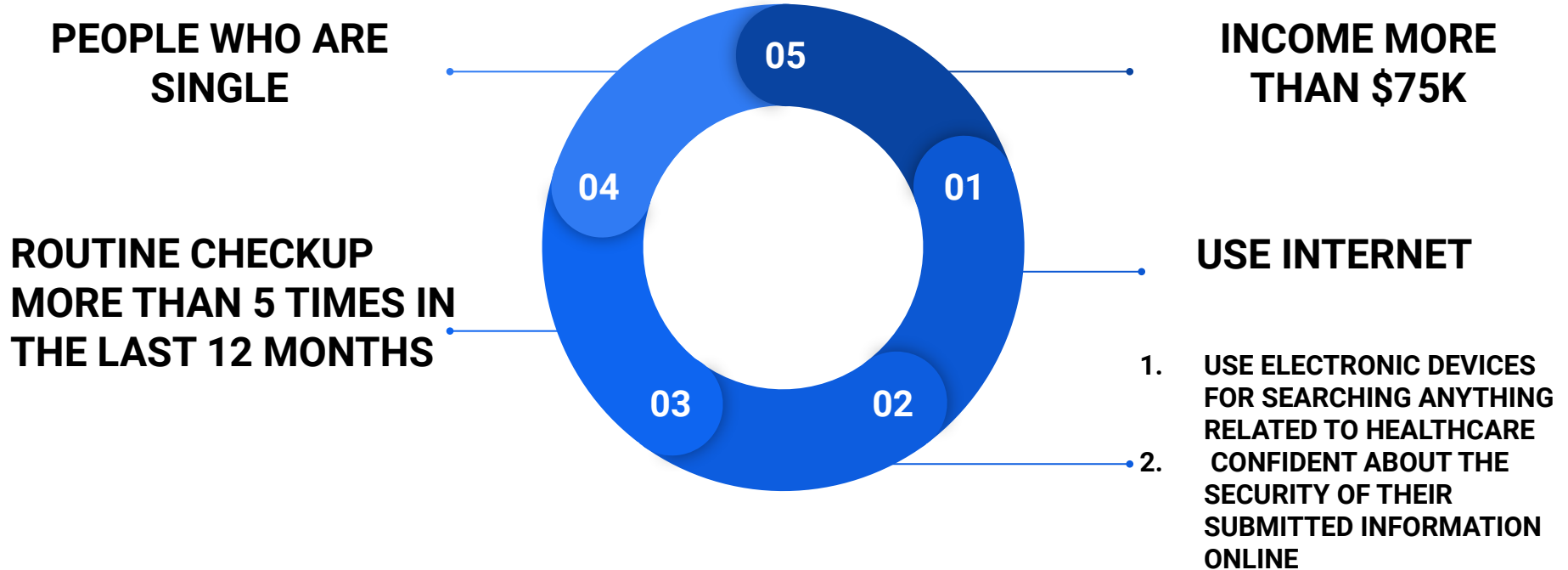
- Get the coefficients for both Binary and Ordinal.
- Take out the common Feature levels and then Interpret



FEATURE LEVEL COEFFICIENTs – INTERPRETATION

STATUS - SEPARATED	LESS LIKELY TO ACCESS	STATUS SINGLE/RELATIONSHIP
LOW INCOME	LESS LIKELY TO ACCESS	HIGH INCOME
DON'T USE INTERNET	LEAST LIKELY TO ACCESS	USE INTERNET
SOMETIME ROUTINE CHECKUP	LESS LIKELY TO ACCESS	OFTEN CHECKUP
OFTEN PROVIDER CONTACT	MOST LIKELY TO ACCESS	RARE PROVIDER CONTACT
VERY CONFIDENT ONLINE INFO	MOST LIKELY TO ACCESS	NOT CONFIDENT
ELECTRONIC USE FOR MEDICAL SEARCHES	MOST LIKELY	NOT USING ANY ELECTRONICS FOR MEDICAL SEARCHES

TARGET AUDIENCE





Thanks!

Any questions ?