1]:	Our client currently In order to determine	ratement: provides for Health ne if this can be do urrent policy holde	h Insurance, but they a one, the Insurance comp er would be interested i	oany has tasked our te	am with developing mo		
	<pre>import numpy a import pandas import matplot import seaborn import statsmo import imblear from mord impo</pre>	aline as np as pd clib.pylab as pl as sns odels.api as sm cn ort LogisticIT	lt import RandomOverSa	ampler			
	<pre>from sklearn.l from sklearn.t from sklearn.e from sklearn.m from sklearn.m from sklearn.m from sklearn.n from sklearn.l from sklearn.d from sklearn.d</pre>	meighbors import cree import Deci- ensemble import model_selection metrics import a preprocessing im- meural_network i discriminant_ana ensemble import ensemble import	port LogisticRegres t NearestNeighbors, isionTreeClassifier RandomForestClassi import train_test_ accuracy_score, roc mport MinMaxScaler import MLPClassifier alysis import Linea AdaBoostClassifier BaggingClassifier	KNeighborsClassi fier, GradientBoo split, cross_val_ c_auc_score, f1_sc er arDiscriminantAnal	fier stingClassifier score, GridSearchC ^v ore, auc	V	
2]:	<pre>from dmba impo from dmba impo from dmba impo from dmba impo import warning warnings.filte Import the cl testing sets. #import the tr train = pd.rea</pre>	ort plotDecision ort backward_eli ort adjusted_r2_ gs erwarnings('igno ient provideo raining set and	ionSummary, gainsCharree, exhaustive_simination, forward_score, AIC_score, ore') d datasets. They view the first few s/Leonard/Desktop/i	search _selection, stepwi BIC_score have already s v rows	plit the data int	o training and	
2]:	 0 1 Male 44 1 2 Male 70 2 3 Male 4 3 4 Male 2 4 5 Female 25 	4 1 6 1 7 1 1 9 1	Region_Code Previousl 28.0 3.0 28.0 11.0 41.0	0 > 2 Years 0 1-2 Year 0 > 2 Years 1 < 1 Year 1 < 1 Year	Yes No Yes No No	40454.0 33536.0 38294.0 28619.0 27496.0	1
	3	a. We will use the en y Data Ana is			validation will be used	3	e place
	Gender Age Driving_License Region_Code Previously_Insured Vehicle_Age			Unique code for the Vehicle Insural Age of the Vehicle	tomer loes not have DL, 1: Custom or the region of the custome lready has Vehicle Insurance nce	e, 0 : Customer doesn't have	
	Vehicle_Damage Annual_Premium Policy Sales Channel Vintage Response Let's check for miss that appropriately		data can impact analyse	get his/her vel The amount of Anonymized C Different Ager Number of Day 1: Customer is	ot his/her vehicle damaged in the past. Istomer needs to pay as presonde for the channel of outre ts, Over Mail, Over Phone, In the past of the channel of outre ts, Customer has been associated interested, 0: Customer is the present of the channel of outre ts, over Mail, Over Phone, In the past of the channel of the past of the	mium in the year aching to the customer ie. n Person, etc. ciated with the company not interested	
6]:	#check for mis train.isnull()	0 0 0 0 0 e 0 0 ured 0	he training set				
7]:	Vintage Response dtype: int64 Great! We have no #determine the dict=[] for i in train print(f'{i id has 381109 u Gender has 2 un Age has 66 unice	missing data! Now number of unique. has {train[i] unique values hique values que values	let us see what kind of que values for each	n column	ke on.		
4]:	Region_Code has Previously_Inst Vehicle_Age has Vehicle_Damage Annual_Premium Policy_Sales_Ch Vintage has 290 Response has 2 #determine the print("Gender print("Driving print("Previous	unique values can take on:\n" g_License can ta	ues ue values es alues ue values unique values	lue_counts(), "\n" Driving_License.va in.Previously_Insu	lue_counts(), "\n"; red.value_counts();		
	print ("Vehicle print ("Respons") Gender can take Male 2060 Female 17502 Name: Gender, continuing_License 1 380297 0 812 Name: Driving_I	e_Damage can take e can take on:\ e on: 089 20 dtype: int64	ke on:\n", train.Ve\n", train.Response	ehicle_Damage.valu	e_counts(), "\n")		
	1 174628 Name: Previousl Vehicle_Age car 1-2 Year 2 < 1 Year 16 > 2 Years 1 Name: Vehicle_A Vehicle_Damage Yes 192413 No 188696	200316 54786 16007 Age, dtype: inte can take on: Damage, dtype: i	64				
	also see that this is This is also very imb Previously_Insured in note that this is an mentioned here are	servations about the very imbalanced. Voalanced, but this is is also binary: 0 (no ordinal variable sind e continuous as the	e columns here. First, the Also see that Gende is something we would and 1 (yes). Vehicle_Alce the order of the valuey take on very many var	r is binary: either male expect since most insu age is categorized into ues matter. Finally, Veh	or female. Driving_Lice red individuals are licer three categories: < 1 ye	nse is either 0 (no) or 1 nsed drivers. ear, 1-2 years, > 2 year	1 (yes). rs. We
5]: 5]:	#frequency dis	ne frequency distribution gende	oution of each of our value	ariables			
	125000 - 100000 - 75000 - 50000 - 25000 -	Male Gen e seems balanced b	Female nder Detween male and fema	ale customers			
0]:	#frequency dis	stribution age		ille Customers			
	15000 - 10000 -			I II			
1]:	The majority of our #frequency dis sns.countplot(customers are between tribution drivi	x = 'Driving_Licens	Age ad 27 with a second pe		9 70 71 72 73 74 75 76 77 78 79 80	81 82 838
1]:	<pre></pre>	uriving_I	License', ylabel='d				
5]:	The vast majority of should have a licens #frequency dis plt.figure(fig	se. stribution regionsize=(20,6))	ve a driver's license. We	e would expect this as	customers who will be i	interested in car insura	ance
	100000 - 80000 - 40000 - 20000 -						
6]: 6]:	The majority of our #frequency dis sns.countplot(customers fall with stribution previ	nin region code = .028. iously_insured x = 'Previously_Insured', ylabe.	Region_Code	vo1.032.033.034.035.036.037.038.039.0	-v. u41.042.043.044.045.046.047.048.04	49.050.051
	125000 - 100000 - 75000 - 50000 - 25000 -	0 Previously ured flag also is rela	j y_Insured atively even with a sligh	t edge to not =	ly insured.		
7]: 7]:	#frequency dis	stribution vehic					
8]:	75000 - 50000 - 25000 - > 2 The car age of most #frequency dis	stribution vehic	e_Age are 2 years or newer. T		ring insurance as newei	r cars tend to be safer.	
8]:	sns.countplot((data = train, x					
8]:	The flag for whethe #frequency dis sns.distplot(t	stribution annua rain['Annual_Pr pins=int(180/5),	<pre>naged or not is roughly al_premium remium'], hist=True , color = 'darkblue color':'black'},</pre>	e, kde=True,			
8]:	<axessubplot:xl< td=""><td></td><td>remium', ylabel='De</td><td>ensity'></td><td></td><td></td><td></td></axessubplot:xl<>		remium', ylabel='De	ensity'>			
9]:	sns.distplot(t b h	Annual_Premi I_premium is rough stribution polic crain['Policy_Sa pins=int(180/5), nist_kws={'edgec	nly 30,564. cy_sales_channel ales_Channel'], his color = 'darkblue color':'black'},				
9]:	0.08 - 0.07 - 0.06 - 0.05 - 0.04 - 0.03 -	de_kws={'linewi		pel='Density'>			
8]:	0.02 0.01 0.00 The most frequent #frequency dis sns.distplot(t	Policy_Sales_C policy sales channe stribution vinta rain['Vintage']	el is 150 followed by 25 age , hist=True, kde=1	with 125 coming in 3r	d.		
8]:	b h k	pins=int(180/5), nist_kws={'edgeckde_kws={'linewi	<pre>, color = 'darkblue color':'black'},</pre>	e',			
	0.0030 -						
0]:	0.0030 - 0.0025 - 0.0025 - 0.0015 - 0.0015 - 0.0005 - 0.0005 - 0.0005 - 0.0005 - 0.0005 - 0.0005 - 0.0005 - 0.0000 - 0.0005 - 0.0000 - 0.0005 - 0.00000 - 0.000000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.000000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.000000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.00000 - 0.000000 - 0.000000 - 0.000000 - 0.00000000	_	ghout. This suggests that		any is pretty good with	getting new clients	
0]:	0.0030 - 0.0025 - 0.0020 - 0.0015 - 0.0010 - 0.0005 - 0.0000 - 0.0005 - 0.0000 - 0.0005 - 0.00000 - 0.0000 - 0.	Winta e is uniform throug ay. stribution responded	ghout. This suggests that		any is pretty good with	getting new clients	
	The vintage variable consistently each date of the vintage variable v	Response variable is need to use a technology. Wintage is uniform through ay. Response variable is need to use a technology.	induction in the suggests that the suggests the suggests that the suggests that the suggests that the suggests the suggests that the suggests the suggests that the suggests that the suggests that the suggests the suggests that t	here the majority of ouversampling.	or customers are not int	erested in car insurance	rs. For
	The vintage variable consistently each date of the vintage variable v	Response variable is need to use a technology and the light of the lig	induction of the state of the s	here the majority of ouversampling. ap Gender, Vehicle_Agge, < 1 year will be 0, 1	e and Vehicle_Damage -2 years will be 1 and >	to integer placeholder > 2 years will be 2. Fina	rs. For ally, es_Cha
0]:	The vintage variable consistently each date of the vintage variable va	Wintage is uniform throughay. Stribution responsed data = train, xelabel='Response' response variable is need to use a technology is need to use a technology is need to use a technology is need to and fend be 0 for no and 1 for a few or and 1 for a few or a few	inductory in the suggests that	there the majority of ouversampling. ap Gender, Vehicle_Agge, < 1 year will be 0, 1 Le': 0, 'Female': y_Insured Vehicle_Age 0 > 2 Years 0 1-2 Year 0 > 2 Years 1 < 1 Year 1 < 1 Year	r customers are not interest and Vehicle_Damage -2 years will be 1 and > 1}) Vehicle_Damage Annu Yes No Yes No No No 0, '1-2 Year': 1,	to integer placeholder 2 years will be 2. Fina 40454.0 33536.0 38294.0 28619.0 27496.0 '> 2 Years': 2})	rs. For ally, es_Char 1 1 1
<pre>0]: 4]: 4]:</pre>	The vintage variable consistently each disconsistently each disconsisten	Response variable is need to use a technology and the control of t	inductors and incomplete the suggests that the s	there the majority of ouversampling. ap Gender, Vehicle_Age ge, < 1 year will be 0, 1 Le': 0, 'Female': y_Insured Vehicle_Age 0 > 2 Years 0 1-2 Year 0 > 2 Years 1 < 1 Year 1 < 1 Year 1 < 1 Year 2	e and Vehicle_Damage -2 years will be 1 and > 1}) Vehicle_Damage Annu Yes No No No 0, '1-2 Year': 1, Vehicle_Damage Annu Yes No No No 0, '1-2 Year': 1, Vehicle_Damage Annu Yes No No No 0, '1-2 Year': 1,	to integer placeholder > 2 years will be 2. Final 40454.0 33536.0 38294.0 27496.0 33536.0 38294.0 28619.0 27496.0 33536.0 38294.0 28619.0 27496.0	es_Chai
<pre>3]: 4]:</pre>	0.0030	e is uniform throughay. Stribution responsed data = train, x data = train = t	phout. This suggests that phouse the series of the suggests that the suggests the suggests that the suggests the suggests that the suggest	there the majority of outversampling. ap Gender, Vehicle_Age ge, < 1 year will be 0, 1 1 year wil	r customers are not interest and Vehicle_Damage -2 years will be 1 and > 1}) Vehicle_Damage Annu Yes No No No 0, '1-2 Year': 1, Vehicle_Damage Annu Yes No No O, '2-2 Year': 1, Vehicle_Damage Annu Yes No No O, '1-2 Annu O O O O O O O O O O O O O O O O O O	to integer placeholder > 2 years will be 2. Final Aud Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Policy_Sale 40454.0 28619.0 27496.0	rs. For ally, es_Char 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
<pre>0]: 3]: 4]: 5]:</pre>	### O 1	e is uniform throughay. Stribution response (data = train, x label='Response') response variable is need to use a technology is perfect to use a technology is perfect to an and 1 is label to 0 and fend to be 0 for no and 1 is label to 0 i	age shout. This suggests that sonse k = 'Response') ', ylabel='count'> sheavily imbalanced with inique called random of the values. Here, we will make to 1. For vehicle age for yes. der'].replace({'Mail Region_Code Previousless.0 28.0 11.0 41.0 ('Vehicle_Age'].region_Code Previousless.0 28.0 11.0 41.0 ain['Vehicle_Damage Region_Code Previousless.0 28.0 11.0 41.0 41.0	there the majority of outversampling. ap Gender, Vehicle_Age ge, < 1 year will be 0, 1 1 year wil	r customers are not interest and Vehicle_Damage -2 years will be 1 and > 1}) Vehicle_Damage Annu Yes No No No 0, '1-2 Year': 1, Vehicle_Damage Annu Yes No No O, '2-2 Year': 1, Vehicle_Damage Annu Yes No No O, '1-2 Annu O O O O O O O O O O O O O O O O O O	to integer placeholder > 2 years will be 2. Final Aud Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Policy_Sale 40454.0 28619.0 27496.0	rs. For ally, es_Chai 1 1 1 es_Chai
<pre>0]: 3]: 4]: 5]:</pre>	### vehicle attrain['Vehicle train.head() ###################################	Response variable is need to use a technology and the control of t	age phout. This suggests that phouse	there the majority of outversampling. ap Gender, Vehicle_Age ge, < 1 year will be 0, 1 year on 1-2 year 0	r customers are not interest and Vehicle_Damage	to integer placeholder > 2 years will be 2. Final Aud Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Policy_Sale 40454.0 28619.0 27496.0	rs. For ally, es_Chai 1 1 1 es_Chai
0]: 3]: 4]: 5]:	0.0030 - 0.0025 - 0.0025 - 0.0025 - 0.0020 - 0.0015 - 0.0015 - 0.0015 - 0.0000 - 0.0	Vintage is uniform throughay. Is tribution responsed to the secondary of	age shout. This suggests that some Response	the insurance companies the insurance companies the majority of outersampling. ap Gender, Vehicle_Age	r customers are not interest and Vehicle_Damage	to integer placeholder > 2 years will be 2. Final Moderate Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 33536.0 33536.0 33536.0 33536.0 3294.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 40454.	rs. For ally, es_Chai 1 1 1 es_Chai
0]: 3]: 4]: 5]:	### Vehicle attrain['Vehicle train.head() id Gender Age 1 2 0 76 2 3 0 4 3 4 0 2 ##################################	Vintage is uniform throughay. Is tribution responsed to the secondary of	age whout. This suggests that the property of the property o	the insurance companies the insurance companies the majority of outersampling. ap Gender, Vehicle_Age	r customers are not interest and Vehicle_Damage	to integer placeholder > 2 years will be 2. Final Moderate Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 33536.0 33536.0 33536.0 33536.0 3294.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 28619.0 27496.0 Moderate Policy_Sale 40454.0 40454.	rs. For ally, es_Chai 1 1 1 es_Chai
0]: 3]: 4]: 5]:	#map vehicle atrain.head() id Gender Agronal Gender, we will may vehicle damage will #map vehicle atrain.head() id Gender Agronal Gender A	winta e is uniform throug ay. stribution respondence (data = train, x Label='Response' label-'Response' label='Response' label-'Response' label-'Respo	thout. This suggests that the suggests the suggests that the suggests the suggests that the suggests t	there the majority of ouversampling. ap Gender, Vehicle_Age	re and Vehicle_Damage -2 years will be 1 and 3 1}) Vehicle_Damage Annu Yes No No No 0, '1-2 Year': 1, Vehicle_Damage Annu Yes No No 1 0 0 1 0 "Gender") . add_legend rthermore, we see the seemone.	to integer placeholder > 2 years will be 2. Final Mal_Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Mal_Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Sweet spot for custome	es_Chai
0]: 3]: 4]: 4]:	### vehicle a train.head() id Gender Aground	winta e is uniform throug ay. stribution respondence (data = train, x Label='Response' label-'Response' label='Response' label-'Response' label-'Respo	shout. This suggests that shout. This suggests that shout. This suggests that shout. This suggests that shout is suggested as the suggest of	there the majority of ouversampling. ap Gender, Vehicle_Age	re and Vehicle_Damage -2 years will be 1 and 3 1}) Vehicle_Damage Annu Yes No No No 0, '1-2 Year': 1, Vehicle_Damage Annu Yes No No 1 0 0 1 0 "Gender") . add_legend rthermore, we see the seemone.	to integer placeholder > 2 years will be 2. Final Mal_Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Mal_Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0 Sweet spot for custome	es_Chai
0]: 3]: 4]: 4]:	### vehicle at train.head() id Gender Aground Agrange ve	winta e is uniform throug ay. stribution responsed data = train, x tabel='Responsed data = train, y tabel	age shout. This suggests that shout.	the insurance companies the insurance companies the majority of output of the control of the con	rcustomers are not interest and Vehicle_Damage	to integer placeholder > 2 years will be 2. Final 40454.0	es_Chai
0]: 3]: 4]: 4]:	#map gender train['Gender' train.head() #map gender train['Gender' train.head() #map vehicle atrain['Vehicle train.head() #map vehicle atrain['Vehicle train.head() #map vehicle atrain['Vehicle train.head() #map vehicle atrain['Nehicle	winta e is uniform throug ay. stribution responsed data = train, x tabel='Responsed data = train, y tabel	age shout. This suggests that shout. This suggests that shout. This suggests that the suggests the suggests that the suggests the suggests the the suggests the suggests the suggests the the suggests the suggests the the suggests the suggests the suggests the the suggests the the suggests the suggests the the suggests the suggests the the	the insurance companies the insurance companies the majority of output of the control of the con	rcustomers are not interest and Vehicle_Damage	to integer placeholder > 2 years will be 2. Final 40454.0	es_Chai
0]: 3]: 4]: 4]:	### Age of the consistent of t	winta e is uniform throug ay. et uniform at uniform a	age shout. This suggests that the suggests the the suggests that the suggests the the suggests that the suggests the suggests the the suggests the the suggests the	there the majority of outersampling. ap Gender, Vehicle_Age	read Vehicle_Damage	to integer placeholder > 2 years will be 2. Fina tal_Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0	es_Chai
0]: 3]: 4]: 4]:	### Public and Standy S	winta e is uniform throug ay. tribution response stribution response data = train, x label='Response' label='Response' proving_License pr	age shout. This suggests that some e	the insurance companies there the majority of outersampling. ap Gender, Vehicle_Age	read Vehicle_Damage	to integer placeholder > 2 years will be 2. Fina tal_Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0	es_Chai
4]: 3]: 5]: 5]: 7]: 7]:	### See that No see that No seaborn. axisgrates for day of seaborn. axisgrates for seaborn. axisgrates	wintale is uniform through ay. stribution responded at a train, with the control of the control	age shout. This suggests that some e	there the majority of outersampling. ap Gender, Vehicle Age	"Vehicle_Damage Annu Yes No No No No No No No Yes No	to integer placeholder > 2 years will be 2. Fina tal_Premium Policy_Sale 40454.0 33536.0 38294.0 28619.0 27496.0	es_Chai
4]: 4]: 5]: 5]: 7]: 5]:	### See that Consistently each of the consistent each of the consistently each of the consistent ea	wintale is uniform through ay. stribution responsed at a train, with a train, with a train, with a train and a tr	age shout. This suggests the shout. This suggests the should suppose the should be sho	the insurance company application of the insurance company application of the insurance company the insurance company application of	r customers are not interested and Vehicle_Damage	erested in car insurance to integer placeholder > 2 years will be 2. Final August	es_Chai
4]: 3]: 4]: 5]: 7]:	### Company of the consistent	o Response variable is need to use a technology and a second a se	age shout. This suggests the shout. This suggests the contact e "Response") "Yelse and	the insurance company application of the insurance company application of the insurance company the insurance company application of	r customers are not interested and Vehicle_Damage	erested in car insurance to integer placeholder > 2 years will be 2. Final August	es_Chai
4]: 3]: 4]: 5]: 7]:	### Profice of Sanalysis for S	o Response variable is need to use a technolysis response variable is need to	age shout. This suggests the property of the p	there the majority of outersampling. ap Gender, Vehicle_Age_ge, < 1 year will be 0, 2 Years	r customers are not interest and vehicle. Damage Annu Yes No	erested in car insurance to integer placeholder > 2 years will be 2. Final August	es_Chai
4]: 3]: 5]: 5]: 8]: 6]:	### See that Interest of the Sanalysis for S	o Response variable is need to use a technolysis response variable is need to	spended the transpose of transpose of the transpose of the transpose of transpose of the transpose of transpose	there the majority of outersampling. ap Gender, Vehicle_Age_ge, < 1 year will be 0, 2 Years	r customers are not interest and vehicle. Damage Annu Yes No	erested in car insurance to integer placeholder > 2 years will be 2. Final August	es_Chai
4]: 3]: 5]: 5]: 8]: 6]:	### See that int ### Se	o Response variable is uniform throug ay. Intribution respondata = train, x Intribut	spender state of the state of t	the insurance companies of the insurance companies of the insurance companies of the insurance companies of the insurance consurance in the insurance consumption of the insurance consumption	r customers are not interest and vehicle. Damage Annu Yes No	erested in car insurance to integer placeholder > 2 years will be 2. Final August	es_Chai
4] 3] 4] 5] 6] 7] 8] 8]	### See that in ### See that i	e is uniform through a provided the service of the	special decision of the control of t	there the majority of outersampling. ap Gender, Vehicle_Age	r customers are not interest and Vehicle_Damage	erested in car insurance to integer placeholder 2 years will be 2. Final 40454.0 33524.0 28619.0 27496.0 *** 2 Years*: 2)) *** 2 Years*: 2) *** 33524.0 33524.0 28619.0 27496.0 **** 40454.0 33524.0	es_Chai
4]: 3]: 5]: 5]: 8]: 6]:	### ### ### ### ### ### ### ### ### ##	e is uniform through a provided the service of the	spender this suggests the chourt his suggests the chourt his suggests the consequence of	there the majority of outersampling. ap Gender, Vehicle_Age	r customers are not interest and Vehicle_Damage	erested in car insurance to integer placeholder 2 years will be 2. Final 40454.0 33524.0 28619.0 27496.0 *** 2 Years*: 2)) *** 2 Years*: 2) *** 33524.0 33524.0 28619.0 27496.0 **** 40454.0 33524.0	es_Chai
4]: 3]: 5]: 5]: 8]: 6]:	### See that In Paramy price for any stage f	de is uniform through a version de la componse variable is need to usta el componse de la composización de la composización de la componse variable is need to	spender this suggests the chourt his suggests the chourt his suggests the consequence of	there the majority of outersampling. ap Gender, Vehicle_Age	r customers are not interest and Vehicle_Damage	erested in car insurance to integer placeholder 2 years will be 2. Final 40454.0 33524.0 28619.0 27496.0 *** 2 Years*: 2)) *** 2 Years*: 2) *** 33524.0 33524.0 28619.0 27496.0 **** 40454.0 33524.0	es_Chai
4]: 3]: 5]: 5]: 8]: 6]:	## See that for an analysis for an alysis for an analysis for	de is uniform through a version de la componse variable is need to usta el componse de la composización de la composización de la componse variable is need to	spender this suggests the chourt his suggests the chourt his suggests the consequence of	anthe insurance compose there the majority of outersampling. ap Gender, Vehicle_Age	r customers are not interest and Vehicle_Damage	erested in car insurance to integer placeholder 2 years will be 2. Final 40454.0 33536.0 28619.0 27496.0 al Premium Policy Sale 40454.0 33536.0 38294.0 28619.0 27496.0 end () sweet spot for custome signeration () sale gend () signeration () si	es_Char
4]: 3]: 5]: 5]: 8]: 6]:	### See that No	what is a series of the series	age thout This suggests the thirty age of the control of the con	and the insurance companies the insurance companies and th	residence are not interested and vehicle Damage Annu Yes No	erested in Car insurance to integer placeholder to 2 years will be 2. Final all Premium Policy Sale 4.0540 38580 387840 287960 27960 387840 287960 27960 388940 287960 27960 388940 287960 388940 287960 388940 3889	es_Char
4]: 3]: 5]: 5]: 8]: 6]:	## Control of the state of the	Response variable side state of the state of	and the suggest she will be a suggest that the suggest she will be a suggest she will be	are the insurance compared to the comp	restormers are not interested. Vehicle Damage Annu Yes No	erested in car insurance to integer placeholder to 2 years will be 2. Fina al Premium Policy Sale 404540 335360 32240 28490 274960	ers not
4]: 5]: 6]: 7]: 8]:	The vintage variable of the vi	with the state of	and the suggest she will be a suggest that the suggest she will be a suggest she will be	and the insurance company application of the insurance company the insurance company application of the insurance company the insurance company application of the insurance compa	restormers are not interested. Vehicle Damage Annu Yes No	and premium Policy, Sall Anish P	ers not

<pre>#apply the X, y = tra print(X.sh.</pre>	oversample inOver.fit_ ape, y.shap	rSampler(sa e resample(tr e)	ampling_stra	ntegy='minor in_y)	enced. If we lea erefore, we wan				_
#display to y.value_con 1 195292 0 195292 Name: Respon Ve now have a controlling a) (390584,) he count of unts() nse, dtype: a balanced res and model valid	int64 sponse variable dation. The find	le. We are now nal model will	ready to beging the used to pre	dict the values	of the test se	et provided	by the client.	
rest baseline power will consider the well consider the consider the consider the consider the consideration that the consideration the consideration that the consideration the consideration that the consid	egression ogistic RegogisticRegre (X, y) ressionCV(c	Regression, Notes and the state of the state	ned for optima Neural Network del enalty='12', iter=500, n_ n model	vill run a baselinal perfomance of the control of t	and that mode minant Analysi fgs', cv=10,	l will be subm	nitted as th prests and	e final model. Decision Trees	The mo
Confusion M Pred Actual 0 7782 1 50 This model yiel Neural Ne #Run the N nn = MLPC1 nn.fit(X,	datrix (Acculication 0 1 3 52519 8 17466 dds terrible reservork	sults. 64% is not model dden_layer_	oot what we are	e shooting for.	Logistic regres	solver='l			=1)
#Analyze to classification Modern Pred Actual 0 1303 1 179 Neural network nsurance com	solver=' he Neural N tionSummary Matrix (Accu iction 0 1 42 0 74 0	lbfgs') Network mode (valid_y, r aracy 0.8788) better accura	el nn.predict (v 8) acy than Logist	_layer_sizes			t predict aı	ny positive val	ues. For
#Run the L lda = Line lda.fit(X, LinearDiscr #Analyze t classifica Confusion M Pred Actual 0 7752	DA model arDiscrimin y) iminantAnal he LDA mode	antAnalysis Lysis() el (valid_y, l	s() lda.predict	(valid_X))					
Decision T #Run the C classTree : classTree. DecisionTre #Analyze to classificate	Regression, LI Tree lassificati DecisionTric(X, y) eClassifier he Classifi	on Tree modereeClassification_state cation Tree (valid_y, o	del ier(random_s ate=1) e model classTree.pr	uracy and ther state=1)		e used.			
Our decision tr Random F #Run the R rf = Randor rf.fit(X,	andom Fores mForestClas. y)	t model sifier(n_es	stimators=50	lded an accuration of the second of the seco	tate=1, n_jc	bbs=-1)	ooth positiv	ve and negativ	e value:
Confusion M Pred Actual 0 1211 1 133 Random Forest	Matrix (Accu iction 0 1 47 9195 15 4659 t turned out to	o be our best	rf.predict(v	odel at 85%. Ti	nerefore, this w	vill be our cho	osen model	I to fine tune a	and try t
<pre>param_grid 'max_d 'min_s 'min_in } gridSearch gridSearch print('Ini print('Ini</pre>	<pre>initial gu = { epth': [None amples_splin mpurity_dec = GridSeare .fit(X, y) tial score: tial parame</pre>	e, 10, 15, t': [10, 20 crease': [0, cchCV(rf, pa ', gridSea ters: ', gr	0, 30, 40, 0, 0.0005, 0. aram_grid, carch.best_scridSearch.be	0], .001, 0.005, ev=5, n_jobs					
#adapt ini param_grid 'max_d 'min_in } gridSearch gridSearch print('Imp	<pre>tial grid s = { epth': list amples_splimpurity_dec = GridSear .fit(X, y) roved score roved parameter: 0.893</pre>	<pre>c'max_depth rearch (range(23, t': list(range): [0, rease': [0, rehCV(rf, parents): ', gridSenters: ', gri</pre>	27)), ange(8, 12)), 0.001, 0.0 aram_grid, c earch.best_s gridSearch.k	ov=5, n_jobs score_) pest_params_	=-1))				
<pre>#use the n rf = Randon rf.fit(X, RandomFores #Analyze t. classifica Confusion M</pre>	ew paramete mForestClas y) tClassifier he Random F	sifier(n_es n_jo r(max_depth= min_sample random_sta	n the random stimators=50 obs=-1) =26, min_imp es_split=8, ate=1) I rf.predict(v	n_impurity_d n_forest mod 00, random_s purity_decre n_estimator valid_X))	el tate=1, max_ ase=0,	depth=26,			se=0, 1
Results a We ran a variet Random Forest able to obtain negatives (pred the advertisem advertisements	and Fina ty of models in ts performed to a model that we dicted negative tent for car insector.	ncluding Logis the best with will be useful. re but actually surance but w	stic Regression a baseline acc While the acc positive). This ould actually b	n, Neural Netwo uracy of 85%. A uracy decrease s means our mo be interested. In	After tuning the d to 79%, we wo del will minim	e model and to control	finding the Irastically re er of custo	best paramet educe the nun omers who do	ers, we not recent
#import th test = pd. test.head(id Ge 0 381110 1 381111	e test set read_csv('C	ers who we and view the company of t	will be intended to the first few conard/Deskton Region_Code 11.0 28.0	erested in v rows pp/insurance Previously_Insu	_test.csv')	ge Vehicle_D ear	amage An No Yes Yes	35786.0 33762.0 40050.0	Policy_
#map gende test['Gende test.head(id Ge 0 381110 1 381111	er'] = test) ender Age December 0 25 0 40	riving_License	28.0 .replace({'N Region_Code 11.0 28.0	Male': 0, 'F Previously_Insu	1 < 1 Ye 0 1-2 Ye	ge Vehicle_D ear	No Yes	35786.0 33762.0	Policy_
test.head(cle_Age'] =)		27.0 28.0 icle_Age'].r Region_Code	replace({'<		ear ear '1-2 Year'			
id Ge 0 381110	cle_Damage') ender Age Dr 0 25		27.0 28.0 Vehicle_Dama Region_Code	age'].replac	ured Vehicle_A			35786.0	Policy_
<pre>#display t</pre>	onse'] = rf	.predict(te	27.0 28.0 hich custome est[predicto	ers will be ors]) pe intereste		1 1 0 0 in car insu	1 1 1 0	33762.0 40050.0 37356.0 59097.0	
After applying hose should b nterested. We	e interested. V	the unseen da We also expec ximize the nui	ct that only a s	ied 31,523 cust mall subset of mers who will v	the 95,514 cust	tomers not se	ent mailers	would have b	een