	 price_p1_var: price of price_p2_var: price of periodprice_p3_var: price of price_p1_fix: price of price_p2_fix: price of price_p3_fix: price of ml_case_training_da id: contact id activity_new: categor campaign_disc_elec: channel_sales: code of cons_12m: electricity 	f energy for the orice of energy power for the 2 power for the 3 power for the 3 ta.csv contacts of the compact code of the sales characteristics.	e 2nd for the 3rd per list period 2nd period 3rd period ins: any's activity. ectricity campa	ign the customer	· last subscribe	d to.			
	 cons_I2m: electricity cons_gas_12m: gas cons_last_month: ele date_activ: date of acceptance date_end: registered date_first_activ: date date_modif_prod: date date_renewal: date of forecast_base_bill_ele forecast_base_bill_ye forecast_bill_12m: for forecast_cons: forecast 	onsumption of ctricity consup- ctrivation of the date of the end of first contracte of last modifies the next contracte: forecasted ever: forecasted ecasted electrical consumptions of the contracted examples of the contracted examples of the contracted ecasted electrical consumptions of the consumption of the consum	the past 12 moderation of the lact of the contract of the client ication of the pract renewal electricity bill be lectricity bill baseling the process of the pract renewal electricity bill baseling the process of the pract renewal electricity bill baseling the process of the pract renewal electricity bill baseling the process of the practical p	onths st month ct product aseline for next m baseline for caler ne for 12 months					
	 forecast_cons_12m: forecast_cons_year: forecast_discount_en forecast_discount_en forecast_meter_rent_ forecast_price_energ forecast_price_energ forecast_price_pow_p has_gas: indicated if orecast_price_pow_en imp_cons: current pain margin_gross_pow_en margin_net_pow_ele: nb_prod_act: number 	forecasted electricity. Forecasted y_p1: forecasted y_p2: forecasted y_p2: forecasted policity also a quid consumption le: gross margin on	etricity consumed value of curved bill of meter ed energy price ed energy price power price for gas client on on power subscri	ption for next cal rent discount rental for the nex for 1st period e for 2nd period 1st period	endar year				
:	 net_margin: total net num_years_antig: ant origin_up: code of the pow_max: subscribed mport Librar import pandas as pd import numpy as np import seaborn as sn import matplotlib.py	margin iquity of the cli e electricity car l power ies	ient (in numbei	of years)	cribed to				
:	<pre>import missingno as from scipy.stats imp Load Data # list of dates dt_lst = ['date_acti pco_main = pd.read_c pco_hist = pd.read_c pco output = pd.read_c</pre>	msno ort zscore a v','date_end sv('ml_case_ sv('ml_case_	','date_firs training_dat training_his	a.csv', parse_ t_data.csv', p	dates=dt_lst)			
:	pd.set_option('displ Main Dataset pco_main.head() 0 48ada52261e7cf587152 1 24011ae4ebbe30351116	id 02705a0451c9	ns',None)	activity_new		NaN Imkebamo			а
	<pre>2 d29c2c54acc38ff3c061 3 764c75f661154dac3a6c 4 bba03439a292a1e166f8 pco_main.info() <class 'pandas.core.f<="" pre=""></class></pre>	4d0a653813dd 254cd082ea7d 80264c16191cb		Nan Nan Nan	1	NaN foosdfpt	fkusacimwkcsos fkusacimwkcsos caaclubfxadlmud	NaN sbicdxkicaua	N a
	RangeIndex: 16096 ent Data columns (total 3 # Column 0 id 1 activity_new 2 campaign_disc_el 3 channel_sales 4 cons_12m 5 cons_gas_12m 6 cons_last_month 7 date_activ 8 date_end 9 date_first_activ 10 date_modif_prod 11 date_renewal 12 forecast_base_bi 13 forecast_base_bi 14 forecast_bill_12 15 forecast_cons_12 17 forecast_cons_12 17 forecast_cons_12 17 forecast_cons_12 17 forecast_price_e 21 forecast_price_e 22 forecast_price_e 23 has_gas 24 imp_cons 25 margin_gross_pow 26 margin_net_pow_e 27 nb_prod_act 28 net_margin 29 num_years_antig 30 origin_up 31 pow_max dtypes: datetime64[ns memory usage: 3.9+ ME # Percentage of null # Percentage of null	ries, 0 to 1 22 columns): No 16 65 6 0 11 16 16 16 16 16 17 18 18 19 19 19 10 10 10 11 11 11 11	on-Null Count on-Null Count on-Null Count on-Null Count on-Null on-null	object object float64 object int64 int64 int64 datetime64[n datetime64[n datetime64[n datetime64[n datetime64[n float64 float65	ns] ns] ns]				
	missing_perc = pco_m print('Percentage of Percentage of Missing id activity_new campaign_disc_ele channel_sales cons_12m cons_gas_12m cons_last_month date_activ date_end date_first_activ date_modif_prod date_renewal forecast_base_bill_el forecast_base_bill_ye forecast_bill_12m forecast_cons forecast_cons_12m forecast_cons_12m forecast_cons_12m forecast_cons_12m forecast_cons_12m forecast_price_energy forecast_price_energy forecast_price_energy forecast_price_energy forecast_price_energy forecast_price_pow_pl has_gas imp_cons margin_gross_pow_ele margin_net_pow_ele nb_prod_act net_margin num_years_antig origin_up pow_max dtype: float64 # Descriptive statis pco_main.describe()	Missing Val (Values: (Va	.mean() * 10 ues:\n', mis 000000 800447 00000 805268 00000 00000 000000 12425						
	75% NaM max NaM Observations 14 columns have negative campaign_disc_el activity_new columns	1.609600e+0 1.948044e+0 1.948044e+0 1.948044e+0 1.252760e+0 1.533250e+0 1.533250e+0 1.609711e+0 1.609711e+0 1.609711e+0 1.609711e+0 1.609711e+0	1.609600e+ 1.775885e+ 1.775886e+	1.9461546 -05 8.2356766 -03 -9.1386006 -00 0.0000006 -00 9.0100006 -00 4.1270006 -06 4.5387206	e+04 e+04 e+04 e+04 e+00 e+02 e+03 e+06	3508.000000 335.843857 649.406000 -364.940000 0.000000 162.955000 396.185000 2566.080000	3508. 335 649. -364. 0. 162. 396 12566.	.000000 .843857 .406000 .940000 .000000 .185000 .080000	
:	• The date_first_accord forecast_cons constant forecast_constant forecast_const	lumns are each Oataset id	price_date pri				1_fix price_p2		
:	 038af19179925da21a25d 038af19179925da21a25d 038af19179925da21a25d 038af19179925da21a25d pco_hist.info() 	619c5a24b745 619c5a24b745 619c5a24b745	2015-02- 01 2015-03- 01 2015-04- 01 2015-05- 01	0.151367 0.151367 0.149626 0.149626	0.0 0.0 0.0 0.0	0.0 44.266 0.0 44.266 0.0 44.266	6931 6931	0.0 0.0 0.0	
1	0 id 19 1 price_date 19 2 price_p1_var 19 3 price_p2_var 19 4 price_p3_var 19 5 price_p1_fix 19 6 price_p2_fix 19 7 price_p3_fix 19 dtypes: datetime64[nsmemory usage: 11.8+ M	atries, 0 to columns): on-Null Count on-Null Count on-Null Count on-Null on-nul	Dtype Dtype Dobject datetime6 float64						
:	# Percentage of null missing_perc = pco_h print('Percentage of Percentage of Missing id	ity by columnist.isnull() Missing Val (Values: 00000 0000 1138 1138 1138 1138 1138	.mean() * 10						
	<pre># Descriptive statis pco_hist.describe() price_p1_var</pre>	price_p2_var	price_p3_var 191643.000000 0.030712 0.036335 0.000000 0.000000	price_p1_fix 191643.000000 43.325546 5.437952 -0.177779 40.728885 44.266930	price_p2_fix 191643.000000 10.698201 12.856046 -0.097752 0.0000000 0.0000000	price_p3_fix 191643.000000 6.455436 7.782279 -0.065172 0.000000 0.0000000			
(50% 0.146033 75% 0.151635 max 0.280700 Observations price_p1_var, privalues. price_p1_fix, privalues. 	0.101780 0.229788 rice_p2_var,	0.072558 0.114102 price_p3_v	44.444710 59.444710 ar, price_p1_	24.339581 36.490692 _fix , price_	16.226389 17.458221 p2_fix, price		_	
: [The Output D pco_output.head() 48ada52261e7cf587152 24011ae4ebbe30351116 2d29c2c54acc38ff3c0616 3764c75f661154dac3a6c bba03439a292a1e166f8	id 02705a0451c9 d65fa7c15bc57 4d0a653813dd 254cd082ea7d	churn 0 1 0 0 0 0						
:	pco_output.info() <class #="" 'pandas.core.f="" (total="" 0="" 1="" 16096="" 2="" 251.6+<="" churn="" column="" columns="" data="" dtypes:="" ent="" id="" int64(1),="" memory="" no="" non-null="" obj="" rangeindex:="" td="" usage:=""><td>Frame.DataFra Fries, 0 to 1 Columns): Count Dtyp on-null obje on-null int6 ect(1) KB</td><td>me'> .6095 Dect</td><td></td><td></td><td></td><td></td><td></td><td></td></class>	Frame.DataFra Fries, 0 to 1 Columns): Count Dtyp on-null obje on-null int6 ect(1) KB	me'> .6095 Dect						
	<pre># Percentage of null missing_perc = pco_o print('Percentage of Percentage of Missing id 0.0 churn 0.0 dtype: float64 # Descriptive statis pco_output.describe(</pre>	utput.isnull Missing Val Values:	().mean() *						
	count 16096.000000 mean 0.099093 std 0.298796 min 0.000000 25% 0.000000 50% 0.000000 75% 0.000000 max 1.000000								
] 	Observations Complete dataset. Data Cleaning Missing Data Types of missingne Missing Completely at R Missingness has no relation	ess andom (MCAR	- R)		ng				
N T N T	Missingness has no relation Missing at Random (MAR There is a systematic relation Missing Not at Random (There is a relationship bet The History Dat # Identify negative negative_cols = ['pr # Convert to positive]	columns columns cice_pl_fix', re the negati	en missingness ness and its val	s and other obserues, missing or not a simple of the contract	on-missing	ot the missing d	ata		
	# Convert to positive pco_hist[negative_co pco_hist.describe()] price_p1_var price_p1_var count 191643.000000 19 mean 0.140991 std 0.025117 min 0.000000 25% 0.125976	price_p2_var 0.054412 0.050033 0.000000 0.000000	price_p3_var 191643.000000 0.030712 0.036335 0.000000 0.0000000	price_p1_fix 191643.000000 43.325563 5.437816 0.000000 40.728885	price_p2_fix 191643.000000 10.698210 12.856039 0.000000 0.000000	price_p3_fix 191643.000000 6.455443 7.782273 0.000000 0.000000			
	50% 0.146033 75% 0.151635 max 0.280700 # Visualize the comp msno.bar(pco_hist) plt.show()	0.085483 0.101780 0.229788 Pleteness of	0.000000 0.072558 0.114102 the datafram	44.266930 44.444710 59.444710	0.000000 24.339581 36.490692	0.000000 16.226389 17.458221	2936 ⁴⁷³	19 ¹ 6 ⁴³	
(0.8 0.6 0.4								
= =	# Visualize the local sorted = pco_hist.somsno.matrix(sorted) plt.show()	tions of the	missing val	ues of the datice_date'])	aset			ee gant	
	# Visualize the corr msno.heatmap(pco_his plt.show() price_pl_var		reen the nume	ric variables	of the dataf.	rame		2	
	price_p2_var 1 price_p3_var 1 price_p1_fix 1	1		1					
	price_p2_fix 1 price_p3_fix 1	price pl. yar		1 1 yat price p	1 Art print	1 egzint	nice 93 ht		
	# Identify the index hist_NAN_index = pco # Obtain a dataframe pco_hist_missing = p # Glimpse at the NaN pco_hist_missing.hea 75 ef716222bbd97a8bd	hist[pco_hi with the mi co_hist.iloc cases of th d(10)	st.isnull(). ssing values [hist_NAN_in e pco_hist d id price_date 2015-04- 01	<pre>any(axis=1)].i dex,:]</pre>	ndex.values.		e_p1_fix price	e _p2_fix p NaN	pric
	 221 0f5231100b2febab8 377 2f93639de582fadfb 413 f83c1ab1ca1d1802b 461 3076c6d4a060e12a0 471 33bb3af90650ac2e9 472 33bb3af90650ac2e9 	e3e86ce1c8d8f3 b1df4d72820243 49d1700d9b09cf 9ecac6ff2c975a6	2015-06- 01 35 2015-06- 01 3c 2015-06- 01 5b 2015-04- 01	NaN NaN NaN NaN	NaN NaN NaN NaN	NaN NaN NaN NaN NaN	NaN NaN NaN NaN NaN	NaN NaN NaN NaN	
	<pre>475 33bb3af90650ac2e9 476 33bb3af90650ac2e9 874 0e90101b08183cc95 # extract the unique date_lst = pco_hist_ id_lst = pco_hist_mi # Create a time data</pre>	Decac6ff2c975a6 Decac6ff2c975a6 48e827e4b256f4 Adtes of mi missing['pri ssing['id']. frame with t	2015-08- 01 6b 2015-09- 01 47 2015-12- 01 ssing data ce_date'].un unique()	NaN NaN ique()	NaN NaN NaN	NaN NaN NaN	NaN NaN NaN	NaN NaN NaN	
	<pre># Glimpse the time d time_df.sort_values(price_date 9 2015-01-01 11 2015-02-01 8 2015-03-01 0 2015-04-01 2 2015-05-01</pre>	me(data=date	_lst, column						
	1 2015-06-01 10 2015-07-01 3 2015-08-01 4 2015-09-01 7 2015-10-01 6 2015-11-01 5 2015-12-01 Observations The columns containing p	rice	trop.	corro!	mies.	Jan	o Of *		
٧	The columns containing p We can use trend and cyc Filling Time series dat # Make a copy of pco pco_hist_ff = pco_hi # Print prior to imp print(pco_hist_ff.il # Fill NaNs using fo pco_hist_ff.fillna(m	licality when in Ca Chist datase st.copy(deep Outing missin oc[hist_NAN_ Orward fill ethod = 'ffi	t =True) g values index, 3:9].h	eries data. ead()) =True)	Jugness, si	oung a case	.arvAR.		
	75 0.000000 221 0.000000 377 0.087970 413 0.102239 461 0.000000	rice_p3_var NaN NaN NaN NaN NaN NaN O.000000 0.000000 0.000000 0.070381 0.000000	price_p1_fix NaN NaN NaN NaN Price_p1_fix 44.266931 44.266931 40.565969 44.266931	price_p2_fix NaN NaN NaN NaN NaN NaN One price_p2_fix O.000000 O.000000 O.000000 24.339581 O.000000	N N N N N N N N N N N N N N N N N N N	aN aN aN aN ix 00 00 00			
	# Merge output datas pco_hist_ff_merged = pco_hist_ff_merged.h 0 038af19179925da21a250 2 038af19179925da21a250 3 038af19179925da21a250	id 619c5a24b745 619c5a24b745	.merge(right		=['id'])	p3_var price_p1 0.0 44.266 0.0 44.266 0.0 44.266 0.0 44.266	6931 6931 6931	2_fix price 0.0 0.0 0.0 0.0	e_I
	4 038af19179925da21a250 The Main Datas # Visualize the comp msno.bar(pco_main) plt.show()	et pleteness of	01 2015-05- 01 the datafram	0.149626 e	0.0	0.0 44.266	6931	0.0	» .45
(0.6								
		tions of the	tole Wec	ues of the dat	aset		Pow Prod act naturn years	antio non por por	
	1 activity heavy transactions of the season	Les July Houth	end first active force?	by pase pill ele pill pear consecutive de la lectrone de la lectro	ns 2m year unteneral	c. J. Kr. B. J. B.	os pod ele parin de la parin d	as and like the trade	الساليان بأيايان أيايانا بالمالياتين أ
	msno.heatmap(pco_maiplt.show()	n)						19 •	11 11 11 11 1 1 1 1 1 1 1 1 1 1 1 1 1
	activity_new channel_sales date_end date_first_activ date_modif_prod date_renewal forecast_base_bill_ele forecast_base_bill_year forecast_bill_12m	0.2 -0.1 0.1 0.2 0.2 0.2 0.2 0.2	1 1 1 1 1	1 1					
	forecast_bill_12m forecast_cons forecast_discount_energy forecast_price_energy_p1 forecast_price_energy_p2 forecast_price_pow_p1	0.2	1		1 1 1 1 1		.5 0.5	,	
	margin_gross_pow_ele margin_net_pow_ele net_margin origin_up pow_max	date date hate	column cannirst_activ']			n dies bon es bon es		ð.	
	# Demonstrate why the activity = ['date_activity = pco_meters activity = pco_meters acti	tiv','date_f of interest ain[activity		elements.		'any',inplace=	False)		
	# Demonstrate why the activity = ['date_ac' # Filter the columns pco_activity_cc = pco_activity_cc = pco_activity_cc.date # Describe it pco_activity_cc.desc.	tiv','date_f of interest ain[activity mplete cases o_activity.d bjects conta activ.equal ribe(datetim date_activ 3508	ropna (subset in the same s (pco_activi e_is_numeric da 011-06-19 20:20	te_first_activ					
	margin_net_pow_ele net_margin origin_up pow_max # Demonstrate why the activity = ['date_ace # Filter the columns pco_activity = pco_m # Obtain only the co pco_activity_cc = pce # Test whether two of pco_activity_cc.date # Describe it pco_activity_cc.desce count mean 2011-09-03 07:45:0 min 2003-09- 25% 2010-10- 50% 2012-01- 75% 2012-08- max 2014-09 # Drop the column acc pco_main_drop = pco_ # Remove date_end dat brush = ['date_end',	tiv', 'date_f of interest lain[activity mplete cases o_activity.d bjects conta a_activ.equal ribe(datetim date_activ 3508 05.131128832 20 -23 00:00:00 -03 00:00:00 -03 00:00:00 -01 00:00:00 ctivity_new a main.drop(laid) tte_modif_pro 'date_modif_pro 'date_modif_pro 'date_modif_pro	ropna (subset in the same s (pco_activi) e_is_numeric da 011-06-19 20:20 2001-0 2011-1 2012-0 2014-0 nd campaign_bels= ['acti ddate_renew prod','date_ prod','date_	3508 :23.261117440 01-10 00:00:00 8-04 18:00:00 0-28 00:00:00 6-22 06:00:00 9-01 00:00:00 disc_elec vity_new','cam al origin_up p renewal','orig	oow_max margi in_up','pow_r	n_gross_pow_el max','margin_g	le margin_ne gross_pow_el	e',	e 1
	margin_net_pow_ele net_margin origin_up pow_max # Demonstrate why the activity = ['date_ac # Filter the columns pco_activity = pco_m # Obtain only the co pco_activity_cc = pc # Test whether two of pco_activity_cc.date # Describe it pco_activity_cc.desc count mean 2011-09-03 07:45:0 min 2003-09- 25% 2010-10- 50% 2012-01- 75% 2012-08- max 2014-09 # Drop the column ac pco_main_drop = pco_ # Remove date_end da brush = ['date_end',	tiv', 'date_f of interest lain[activity mplete cases o_activity.d bjects conta _activ.equal ribe(datetim date_activ 3508 05.131128832 20 -23 00:00:00 -06 00:00:00 -08 00:00:00 -01 00:00:00 ctivity_new a main.drop(laid te_modif_prod 'date_modifrow_ele', ' rice_energy_ (subset=brus _drop)	ropna (subset in the same s (pco_activi) e_is_numeric da 011-06-19 20:20 2010-0 2011-1 2012-0 2014-0 ind campaign_bels= ['acti ind date_renew prod','date_ net_margin', p2','forecas h, how='any'	3508 :23.261117440 01-10 00:00:00 8-04 18:00:00 0-28 00:00:00 6-22 06:00:00 9-01 00:00:00 disc_elec vity_new','cam al origin_up p renewal','orig 'forecast_disc t_price_pow_p1 ,inplace=True)	oow_max marginin_up','pow_nount_energy',	n_gross_pow_el max','margin_g ,'forecast_pri	e margin_ne gross_pow_el .ce_energy_p	e', 1',	e
	# Demonstrate why the activity = ['date_ace # Filter the columns pco_activity_cc = pco_activity_cc.date # Describe it pco_activity_cc.date # Describe it pco_activity_cc.desce	tiv', 'date_f of interest lain[activity mplete cases o_activity.d bjects conta _activ.equal ribe(datetim date_activ 3508 05.131128832 20 -23 00:00:00 -06 00:00:00 -08 00:00:00 -01 00:00:00 ctivity_new a main.drop(lai te_modif_pro 'date_modifrice_energy_ (subset=brus _drop)	ropna (subset in the same s (pco_activi) e_is_numeric da 011-06-19 20:20 2010-0 2011-1 2012-0 2014-0 ind campaign_bels= ['acti ind date_renew prod','date_ net_margin', p2','forecas h, how='any'	3508 :23.261117440 01-10 00:00:00 8-04 18:00:00 0-28 00:00:00 6-22 06:00:00 9-01 00:00:00 disc_elec vity_new','cam al origin_up p renewal','orig 'forecast_disc t_price_pow_p1 ,inplace=True)	oow_max marginin_up','pow_nount_energy',	n_gross_pow_el max','margin_g ,'forecast_pri	e margin_ne gross_pow_el .ce_energy_p	e', 1',	
	# Demonstrate why the activity = ['date_activity_cc = pco_meter of the pco_activity_cc = pco_meter of the pco_activity_cc describe it pco_activity_cc.describe it pco_acti	tiv', 'date_f of interest ain[activity mplete cases o_activity.d bjects conta activ.equal ribe(datetim date_activ 3508 05.131128832 2 -23 00:00:00 -03 00:00:00 -03 00:00:00 -03 00:00:00 -01 00:00:00 ctivity_new a main.drop(la te_modif_pro 'date_modif_ pow_ele', ' rice_energy_ (subset=brus drop) Cannot replace cannot	ropna (subset in the same s (pco_activi e_is_numeric da 011-06-19 20:20 2001-0 2010-0 2011-1 2012-0 2014-0 ad campaign bels= ['acti ad date_renew prod', 'date_ net_margin', p2', 'forecas h, how='any' correlation with can drop this correlation ce date_acti ce date_acti	ate_first_activ 3508 :23.261117440 21-10 00:00:00 8-04 18:00:00 0-28 00:00:00 6-22 06:00:00 9-01 00:00:00 disc_elec vity_new','cam al origin_up p renewal','orig 'forecast_disc t_price_pow_p1 , inplace=True) another variables. column. Suggests	Can drop this of that subscribe	column rs are not subsc	ribing through	e', 1', campaign campaign	

41: Se A valu Try us See th urning self count mean std	ttingWithCopyWarning: e is trying to be set on a ing .loc[row_indexer,col_i e caveats in the documenta -a-view-versus-a-copy [k1] = value[k2]	copy of a slice ndexer] = value i tion: https://par cons_last_month for 1.567400e+04 1.941588e+04 8.226881e+04	from a DataFrame instead ndas.pydata.org/p recast_cons_12m for 15674.000000 2359.676441 3979.605687	orecast_cons_year 15674.000000 1911.698354 5224.813531	de/user_guide/indexing.html# forecast_discount_energy forecast 15674.000000 0.976139 5.124103 0.000000
25% 50% 75% max # Con pco_m # Mer pco_m	0.000000e+00 0.000000e+00 5.893250e+03 0.000000e+00 1.522000e+04 0.000000e+00 4.953825e+04 0.000000e+00 1.609711e+07 4.154590e+06 vert the has_gas column to ain_cc['has_gas'] = pco_ma: ge the main dataset with thain_cc_merged = pco_main_cc vert the churn column to Characteristics	<pre>in_cc['has_gas']. he output dataset c.merge(right=pco</pre>			0.000000 0.000000 0.000000 50.000000
/tmp/i A valu Try us See th urning pco_	-a-view-versus-a-copy main_cc['has_gas'] = pco_m ain_cc_merged.head()	<pre>co_main_cc_merged y:2: SettingWithO copy of a slice ndexer] = value i tion: https://par ain_cc['has_gas']</pre>	CopyWarning: from a DataFrame instead ndas.pydata.org/p	eandas-docs/stab	<pre>0:'Stayed'}) ple/user_guide/indexing.html# date_end date_modif_prod date_</pre>
 d29 764 bba 	id da52261e7cf58715202705a0451c9 c2c54acc38ff3c0614d0a653813dd c75f661154dac3a6c254cd082ea7d 03439a292a1e166f80264c16191cb bb38a1afd7c0fc49c77b3789b59a3	309275 4660 544 1584	0 0 0	nonth date_activ 10025 2012-11- 07 0 2009-08- 21 0 2010-04- 16 0 2010-03- 30 12400 2010-04- 08	2016-11- 06 2012-11-07 20 2016-08- 30 2009-08-21 20 2016-04- 16 2010-04-16 20 2016-03- 30 2010-03-30 20
<pre># Obt categ # Obt dates # Obt</pre>	<pre>ain all the Date Variables = [column_name for column_ ain all the numeric column_ ic = [column_name for column_</pre>	lumn_name in pco_ riables except for column_name in variable s mn_name in variable n categorical id'	or id driables if pco_m es if pco_main_cc	olumns if colum ain_cc_merged[c	
The (and column_name != ': and column_name != ': and column_name not : a Visualization Output Dataset culate the zcores of tenure e_zcores = zscore(a=pco_ma: vert to absolute values	churn' in dates]	n_years_antig'])		
abs_t # Ext churn # Add churn # Rem churn # Vis vio = # Set vio.s	enure_zscores = np.abs(tenuract Columns of interest _tenure = pco_main_cc_merget _z-score column _tenure['z_score'] = list(atomic over outliers) ed_tenure_filtered = churnualize tenure by retained of sns.violinplot(y=churned) tings et(xlabel='Years', ylabel= et_title("Customer Attritice)	ed[['churn','num_ abs_tenure_zscore _tenure[churn_ten customer and chur _tenure_filtered[es) nure['z_score'] <		ered["num_years_antig"])
valury us ee th	pykernel_1269/726791132.py e is trying to be set on a ing .loc[row_indexer,col_i e caveats in the documenta -a-view-versus-a-copy n_tenure['z_score'] = list Customer Attrition	<pre>copy of a slice ndexer] = value i tion: https://par (abs_tenure_zscor</pre>	from a DataFrame instead ndas.pydata.org/p		ple/user_guide/indexing.html#
Churned		8 10			
• Cus • The	m.value_counts(subset=['or:	during the 4th year ners is 5 years ign [(pco_main_cc_mer		tayed') & (pco_	main_cc_merged['net_margin'];
xidpi xamkkx dkssx sapbe ewxeel dtype: # Hig print prigin xidpi dkssx	ddsbxsbosboudacockeimpuepw fxxuwbdslkwifmmcsiusiuosws wpmemidmecebumciepifcamkci pcfoloekilkwsdiboslwaxobdp celemmiwuafmddpobolfuxioce int64 hest netting electricity se (ele_nm.groupby('origin_up) _up ddsbxsbosboudacockeimpuepw wpmemidmecebumciepifcamkci	4188 3201 2 1 ubscription campa ')['net_margin']. 1541159.95 814230.02	nign agg('sum').sort_	values(ascendin	g=False))
aamkkx asapbe ewxeel Jame: acts • The cus	fxxuwbdslkwifmmcsiusiuosws pcfoloekilkwsdiboslwaxobdp celemmiwuafmddpobolfuxioce net_margin, dtype: float64 most popular electricity campa tomers. With a net margin of \$1,	717939.95 250.40 46.22 hign is lxidpiddsb. ,541,159.95 in 2015.	rgins		as brought 6,584 current 6 (pco_main_cc_merged['net_material)
top_c	10 customers by net marginustomers.sort_values(by=['1000000000000000000000000000000000000	id num_years_a 2b0dd 5f8861 69325 07b6d		d(10)	
10100 12028 6405 6850 13553 hese a		0fa49e :10ded /a267a d0312f	4 4148.99 3 4040.60 4 3744.72 4 3716.78 4 3407.65 ms of net margin. No	otet that most of tl	hem are within the likely tenure of