1. fecha\_dato

days = 365\*(y-2015)+31\*(m)+d

range:59~548

days = days/548

2. ncodpers

skip

3. ind\_empleado

onehot

{'N': 0, 'A': 1, 'B': 2, 'F': 3, 'S': 4}

4. pais\_residencia

onehot

{'ES': 0, 'CA': 1, 'CH': 2, 'CL': 3, 'IE': 4, 'AT': 5,

'NL': 6, 'FR': 7, 'GB': 8, 'DE': 9, 'DO': 10, 'BE': 11,

'AR': 12, 'VE': 13, 'US': 14, 'MX': 15, 'BR': 16, 'IT': 17,

'EC': 18, 'PE': 19, 'CO': 20, 'HN': 21, 'FI': 22, 'SE': 23,

'AL': 24, 'PT': 25, 'MZ': 26, 'CN': 27, 'TW': 28, 'PL': 29,

'IN': 30, 'CR': 31, 'NI': 32, 'HK': 33, 'AD': 34, 'CZ': 35,

'AE': 36, 'MA': 37, 'GR': 38, 'PR': 39, 'RO': 40, 'IL': 41,

'RU': 42, 'GT': 43, 'GA': 44, 'NO': 45, 'SN': 46, 'MR': 47,

'UA': 48, 'BG': 49, 'PY': 50, 'EE': 51, 'SV': 52, 'ET': 53,

'CM': 54, 'SA': 55, 'CI': 56, 'QA': 57, 'LU': 58, 'PA': 59,

'BA': 60, 'BO': 61, 'AU': 62, 'BY': 63, 'KE': 64, 'SG': 65,

'HR': 66, 'MD': 67, 'SK': 68, 'TR': 69, 'AO': 70, 'CU': 71,

'GQ': 72, 'EG': 73, 'ZA': 74, 'DK': 75, 'UY': 76, 'GE': 77,

'TH': 78, 'DZ': 79, 'LB': 80, 'JP': 81, 'NG': 82, 'PK': 83,

'TN': 84, 'TG': 85, 'KR': 86, 'GH': 87, 'RS': 88, 'VN': 89,

'PH': 90, 'KW': 91, 'NZ': 92, 'MM': 93, 'KH': 94, 'GI': 95,

'SL': 96, 'GN': 97, 'GW': 98, 'OM': 99, 'CG': 100, 'LV': 101,

'LT': 102, 'ML': 103, 'MK': 104, 'HU': 105, 'IS': 106, 'LY': 107,

'CF': 108, 'GM': 109, 'KZ': 110, 'CD': 111, 'BZ': 112, 'ZW': 113,

'DJ': 114, 'JM': 115, 'BM': 116, 'MT': 117}

5. sexo

H ->0; V->1

6. age

NA -> 23

100 up -> 100

/= 100

7. fecha\_alta

days = 365\*(y-1995)+31\*(m)+d

8. ind\_nuevo

unchange

9. antiguedad

NA -> 12

/=256

10. indrel

/= 100

11. ult\_fec\_cli\_1t

drop(too much empty)

12. indrel\_1mes

onehot

{1.0, 1, P -> 0; 2.0, 2 ->1; 3.0, 3 -> 2; 4.0, 4 ->3}

13. tiprel\_1mes

onehot

{'A': 0, 'I': 1, 'P': 2, 'R': 3, 'N': 0}

14. indresi

{'S': 0, 'N': 1}

15. indext

{'S': 0, 'N': 1}

16. conyuemp

drop(too much empty)

17. canal\_entrada

onehot

{'KHL': 0, 'KHE': 1, 'KHD': 2, 'KFA': 3, 'KFC': 4, 'KAT': 5, 'KAZ': 6, 'RED': 7,

'KHC': 8, 'KHK': 9, 'KGN': 10, 'KHM': 11, 'KHO': 12, 'KDH': 13, 'KEH': 14, 'KAD': 15,

'KBG': 16, 'KGC': 17, 'KHF': 18, 'KFK': 19, 'KHN': 20, 'KHA': 21, 'KAF': 22, 'KGX': 23,

'KFD': 24, 'KAG': 25, 'KFG': 26, 'KAB': 27, 'KCC': 28, 'KAE': 29, 'KAH': 30, 'KAR': 31,

'KFJ': 32, 'KFL': 33, 'KAI': 34, 'KFU': 35, 'KAQ': 36, 'KFS': 37, 'KAA': 38, 'KFP': 39,

'KAJ': 40, 'KFN': 41, 'KGV': 42, 'KGY': 43, 'KFF': 44, 'KAP': 45, 'KDE': 46, 'KFV': 47,

'013': 48, 'K00': 49, 'KAK': 50, 'KCK': 51, 'KCL': 52, 'KAY': 53, 'KBU': 54, 'KDR': 55,

'KAC': 56, 'KDT': 57, 'KCG': 58, 'KDO': 59, 'KDY': 60, 'KBQ': 61, 'KDA': 62, 'KBO': 63,

'KCI': 64, 'KEC': 65, 'KBZ': 66, 'KES': 67, 'KDX': 68, 'KAS': 69, '007': 70, 'KEU': 71,

'KCA': 72, 'KAL': 73, 'KDC': 74, 'KAW': 75, 'KCS': 76, 'KCB': 77, 'KDU': 78, 'KDQ': 79,

'KCN': 80, 'KCM': 81, '004': 82, 'KCH': 83, 'KCD': 84, 'KCE': 85, 'KEV': 86, 'KBL': 87,

'KEA': 88, 'KBH': 89, 'KDV': 90, 'KFT': 91, 'KEY': 92, 'KAO': 93, 'KEJ': 94, 'KEO': 95,

'KEI': 96, 'KEW': 97, 'KDZ': 98, 'KBV': 99, 'KBR': 100, 'KBF': 101, 'KDP': 102, 'KCO': 103,

'KCF': 104, 'KCV': 105, 'KAM': 106, 'KEZ': 107, 'KBD': 108, 'KAN': 109, 'KBY': 110, 'KCT': 111,

'KDD': 112, 'KBW': 113, 'KCU': 114, 'KBX': 115, 'KDB': 116, 'KBS': 117, 'KBE': 118, 'KCX': 119,

'KBP': 120, 'KBN': 121, 'KEB': 122, 'KDS': 123, 'KEL': 124, 'KDG': 125, 'KDF': 126, 'KEF': 127,

'KCP': 128, 'KDM': 129, 'KBB': 130, 'KDW': 131, 'KBJ': 132, 'KFI': 133, 'KBM': 134, 'KEG': 135,

'KEN': 136, 'KEQ': 137, 'KAV': 138, 'KFH': 139, 'KFM': 140, 'KAU': 141, 'KED': 142, 'KFR': 143,

'KEK': 144, 'KFB': 145, 'KGW': 146, 'KFE': 147, 'KGU': 148, 'KDI': 149, 'KDN': 150, 'KEE': 151,

'KCR': 152, 'KCQ': 153, 'KEM': 154, 'KCJ': 155, 'KHQ': 156, 'KDL': 157, '025': 158, 'KHP': 159,

'KHR': 160, 'KHS': 161}

18. indfall

{'S': 0, 'N': 1}

19. tipodom

drop (all = 1.0, no use)

20. cod\_prov

onehot

{'29.0': 0, '13.0': 1, '50.0': 2, '45.0': 3, '24.0': 4, '20.0': 5, '10.0': 6,

'17.0': 7, '49.0': 8, '8.0': 9, '37.0': 10, '9.0': 11, '22.0': 12, '31.0': 13,

'5.0': 14, '40.0': 15, '27.0': 16, '25.0': 17, '28.0': 18, '3.0': 19, '42.0': 20,

'41.0': 21, '39.0': 22, '7.0': 23, '47.0': 24, '36.0': 25, '46.0': 26, '44.0': 27,

'15.0': 28, '32.0': 29, '23.0': 30, '16.0': 31, '48.0': 32, '12.0': 33, '26.0': 34,

'2.0': 35, '6.0': 36, '30.0': 37, '11.0': 38, '4.0': 39, '19.0': 40, '34.0': 41,

'35.0': 42, '14.0': 43, '21.0': 44, '18.0': 45, '33.0': 46, '38.0': 47, '52.0': 48,

'43.0': 49, '1.0': 50, '51.0': 51}

21. nomprov

drop(exactly the same with cod\_prov, redundent)

22. ind\_actividad\_cliente

skip

(already 0.0 and 1.0)

23. renta

28894395:max rebta

if renta >= 28894395/100:

renta = 28894395/100

/= (28894395/100)

24. segmento

onehot

{'02 - PARTICULARES': 0, '03 - UNIVERSITARIO': 1, '01 - TOP': 2}