

Explicación del set de datos

Leonardo Santella

January 2016

Introducción

Para el preprocesamiento de los datos proporcionados por la Organización de Becas Crema, se ha solicitado un documento en el cual se infiera el contexto de las características que se encuentran en el set de datos (data.csv) debido a que fue extraviada la explicación de este. En este documento se explicarán cada una de las características en el caso que exista el contexto suficiente, en caso contrario simplemente se mencionará la ausencia de contexto en dicho conjunto de datos (grupo de filas o columnas).

Explicación

Filas

Al observar el set de datos es apreciable que el número de filas es igual a 190. En principio, podríamos inferir que cada fila es una persona que esta solicitando una beca a la Organización de Becas Crema.

Columnas

Del mismo modo, al observar las comlumnas del set de datos podemos inferir que cada una de las columnas es una posible característica de un individuo a tomar en cuenta para una futura vista minable.

En este set de datos hay 66 columnas, de cada columna podemos inferir algunas cosas con tan solo ver los valores. El primer valor de cada columna proporciona cierto contexto para inferir un posible nombre para la columna (o característica) además de un significado semántico. Existen columnas que proporcionan un contexto inconsistente con los valores representados en estas, en estos caso haremos una observación al respecto. además de lo anterior, es posible contabilizar los elementos no-nulos de cada columna y en principio, tambien es posible inferir un tipo de dato para cada columna (es posible que el tipo de dato cambie a la hora de generar la vista minable a través del script requerido). Los tipos de datos que contienen letras, simbolos y/o números serán considerados como tipo String (para explicar el set de datos únicamente).

En el set de datos se pueden observar columnas que contienen valores que son NA (Not Available), serán tomados en cuenta como valores no-nulos en un principio, para diferenciarlos de la ausencia de valores en las columnas.

Existen además columnas con un 1er valor idéntico, sin embargo, su contenido es diferente, en esos casos se infiere su significado través de la información proporcionada por las columnas adyacentes, debido a que es observable algún tipo de relación 'ingenua' entre ellas.

A cada columna se le agregara un nombre relacionado a su significado semántico, para acceder de una manera más cómoda a los valores en la estructura de datos (pandas DataFrame).

A continuación, la explicación de cada una de las columnas (inferencia a través del contexto):

Table 1: Columnas 1-14

# Columna	1er Valor de la columna	# Valores no-nulos	Tipo	Nombre
1	""	190	Numérico	No-Name
2	Indique.el.Periodo. a.renovar	190	String	Periodo
3	Cédula.de.Identidad	190	Numérico	Cedula
4	Fecha.de.Nacimiento.. colocar.solo.datos. Numéricos	190	Fecha	F_Nacimiento
5	Edad	190	String	Edad
6	Estado.Civil	190	String	E_Civil
7	Sexo	190	String	Sexo
8	Escuela	190	String	Escuela
9	Año.de.Ingreso.a.la.UCV	190	Numérico	A_Ingreso
10	Modalidad.de. Ingreso.a.la.UCV	190	String	Mod_Ingreso
11	Semestre.que.cursa	190	String	S_Cursa
12	Ha.cambiado.usted. de.dirección	190	String	C_Direccion
13	De.ser.afirmativo. indique.el. motivo	9	String	Motivo_C
14	Número.de.materias. inscritas.en.el.semestre .o.año.anterior	190	Numérico	M_Inscritas

Table 2: Columnas 15-28

# Columna	1er Valor	# Valores no-nulos	Tipo	Nombre
15	Número.de.materias .aprobadas.en.el.semestre.o.año. anterior	190	String	M_Aprobadas
16	Número.de.materias. retiradas.en.el.semestre.o.año. anterior	190	Numérico	M_Retiradas
17	Número.de.materias. reprobadas.en.el.semestre.o.año. anterior	190	Numérico	M_Reprobadas
18	Promedio.ponderado .aprobado	190	Numérico	Promedio_Pond
19	Eficiencia	190	Numérico	Eficiencia
20	Si.reprobo.una.materia .o.más.indique.el.motivo	42	String	Reprobacion_Motiv
21	Número.de.materias. inscritas.en.el.semestre.en.curso	190	Numérico	Materias_Actuales
22	X.Estás.realizando. tesis...trabajo.de.gradp.o.pasantias.de .grado	190	String	Tesis
23	Tesis.trabajo.de.grado .o.pasantias.de.grado	49	String	Veces_Tesis
24	Procedencia	190	String	Procedencia
25	Lugar.donde.reside. mientras.estudia.en.el.Universidad	172	String	Residencia
26	Personas.con.las. cuales.usted.vive.. mientras.estudia.en.la.universidad	190	String	Acompanantes
27	Tipo.de.vivienda.don de.reside.mientras. estudia.en.la.universi dad	190	String	Tipo_Vivienda
28	En.caso.de.vivir.en. habitación.alquilada. o.residencia.estudian til.indique.el.monto. mensual	23	String	Monto_Alquiler_Res

Table 3: Columnas 29-42

# Columna	1er Valor	# Valores no-nulos	Tipo	Nombre
29	Dirección.donde.se.en cuenta.ubicada.la.residencia.o.habitación.alquilada	45	String	Direccion_Res_Alq
30	X.Contrajo.Matrimonio	190	String	Matrimonio
31	X.Ha.solicitado.algún.otro.beneficio.a.la.Universidad.u.otra.Institucion	190	String	Otro_Beneficio
32	En.caso.afirmativo.senale..año.de.la.solicitud..institucion.y.motivo	13	String	F_Motiv_Sol_Benef
33	X.Se.encuentra.usted..realizando.alguna.actividad.que.le.genere.ingresos	190	String	Actividad_Ingresos
34	En.caso.afirmativo..indique.tipo.de.actividad.y.su.frecuencia	4	String	Tipo_Frec_Actividad
35	Monto.mensual.de.la.beca	190	Numérico	Monto_Mensual_Beca
36	Aporte.mensual.que.le.brinda.su.responsable.económico	186	Numérico	Mensual_Aporte_Resp
37	Aporte.mensual.que.recibe.de.familiares.o.amigos	113	Numérico	Mensual_Aporte_Amigos
38	Ingreso.mensual.que.recibe.por.actividades.a.destajo.o.por.horas	88	Numérico	Aporte_Destajo
39	Ingreso.mensual.total	190	Numérico	Mensual_Ingreso_Total
40	Alimentación	172	Numérico	Alimentacion
41	Transporte.público	180	Numérico	T_Publico
42	Gastos.médicos	112	Numérico	Gastos_Medicos

Table 4: Columnas 43-56

# Columna	1er Valor	# Valores no-nulos	Tipo	Nombre
43	Gastos.odonto lógicos	89	Numérico	Gastos_Odonto
44	Gastos.personales	153	Numérico	Gastos_Personales
45	Residencia.o.habita ción.alquilada	69	Numérico	Gastos_Alquiler
46	Materiales.de.estudio	190	Numérico	Gastos_Materiales
47	Recreación	106	Numérico	Gastos_Recreacion
48	Otros.Gastos	140	Numérico	Gastos_Otros
49	Total.Egresos	190	Numérico	Total_Egresos
50	Indique.quien.es.su. responsable.econó mico	190	String	Responsable_Economico
51	Carga.familiar	190	Numérico	Carga_Familiar
52	Ingreso.mensual.de su.responsable.eco nómico	190	String	Ingreso_Responsable
53	Otros.ingresos	112	String	Ingreso_Resp_Otros
54	Total.ingresos	190	String	Total_Ingresos_Resp
55	Vivienda	86	String	Vivienda
56	Alimentación.1	190	String	Gastos_Alimentacion_Resp

Table 5: Columnas 57-66

# Columna	1er Valor	# Valores no-nulos	Tipo	Nombre
57	Transporte	190	String	Gastos_Transporte_Resp
58	Gastos.médicos.1	144	String	Gastos_Medicos_Resp
59	Gastos.odontológicos	89	String	Gastos_Odonto_Resp
60	Gastos.Educativos	111	Numérico	Gastos_Educativos_Resp
61	Servicios.públicos.de agua..luz..teléfono.y. gas	170	String	Gastos_SP_Gas
62	Condominio	107	String	Condominio_Resp
63	Otros.Gastos.1	153	Numérico	Otros_Gastos_Resp
64	Total.de.egresos	190	String	Total_Egresos_Resp
65	Deseamos.conocer.la .opinión.de.nuestros. usuarios.para.me jorar.la.caliad.de.los. servicios.ofrecidos. por.el.Dpto..de.Traba jo.Social.OBE	190	Numérico	Opinion_Usuarios
66	Sugerencias.y.reco mendaciones.para. mejorar.nuestra.aten ción	190	String	Sugerencia_Recomendaciones

Observaciones

1. La primera columna carece de un primer valor, y por los valores adyacentes, a priori no se puede inferir un significado