

Explicación del set de datos

Deyban Pérez

January 14, 2016

Abstract

Este documento ofrece una descripción de la problemática y de las variables pertenecientes al set de datos. Realizar un análisis previo y tener documentadas las variables es fundamental para una mejor comprensión del ambiente por parte de la organización, el conjunto de datos fueron tomados dentro del ámbito universitario de la **Universidad Central de Venezuela** y corresponde a solicitud de renovación de becas estudiantiles por parte de los alumnos.

1 Introducción

A la hora de hacer minería, los pasos que llevan más tiempo y son más importantes son los de **recolección de la data y preprocesamiento de los datos**, dependiendo de la calidad de los aspectos señalados anteriormente, nuestro algoritmo podrá tener mejores resultados, ya que debemos tener una buena base sobre la que podamos empezar a realizar minería. En la problemática local referente a la renovación de becas estudiantiles ya se realizó el primer paso de **recolección de datos**, y lo que se busca es poder tener una información ordenada y limpia de los datos recolectados (**Vista Minable**) que permita a la organización una mejor comprensión del problema.

2 Explicativa del problema

La **Universidad Central de Venezuela** cuenta con un programa de becas para sus estudiantes que sirve como ayuda económica para aquellas personas que así lo necesiten y lo desee, la solicitudes de becas deben ser renovadas periódicamente con la finalidad de llevar un control sobre el rendimiento de los estudiantes que justifique la renovación de la beca o la destitución por parte de universidad para el estudiante, por lo anterior, se puede pensar en una forma de automatización del proceso previamente mencionado basándose en estadísticas extraídas durante un proceso de minería; la minería en este ambiente también podría servir para recolectar información útil sobre otros aspectos de la población de estudiantes que renueva las becas, tales como: **edad promedio, promedio de materias aprobadas, estrato social promedio**, etc. Y

que genere información confiable y viable en vistas de prestar un mejor servicio o hacer un estudio para mejorar el sistema de becas.

3 Explicación de los datos

1. **””**: No posee ningún nombre, por el contexto se cree que es un **identificador** que lleva el orden en el que se realizaron las solicitudes, este identificador en el sistema es incremental y único para cada usuario y es de tipo entero.
2. **Indique.el.Período.Académico.a.renovar**: Período académico en curso por el cuál se desea renovar la beca, este tipo de dato es **string**.
3. **Cédula.de.Identidad**: Documento de identidad ciudadano que identifica a la persona dentro de su país, esta variable es un **entero**.
4. **Fecha.de.Nacimiento..colocar.sólo.datos.numéricos**: Fecha de nacimiento de la persona, este tipo de dato es un **string** con formato de fecha **dd/mm/aaa**.
5. **Edad**: Edad de la persona, tipo de dato **entero**.
6. **Estado.Civil**: Puede adoptar los valores de **Soltero (a)**, **Casado (a)**, **Unido (a)**, **Viudo (a)**, es de tipo **categorico**.
7. **Sexo**: Género de la persona que adopta los valores de **Masculino o Femenino**, tipo **categorico**.
8. **Escuela**: Indica la escuela a la cuál pertenece la persona que desea renovar la beca y adopta los valores de **Bioanálisis** o **Enfermería**, es de tipo **categorico**.
9. **Año.de.Ingreso.a.la.UCV**: Tipo de dato **entero** que representa el año de ingreso de la persona a la universidad.
10. **Modalidad.de.Ingreso.a.la.UCV**: Indica la forma a la que se ingresó a la universidad, es un tipo de dato **categorico** que adopta los valores de **Asignado OPSU**, **Convenios Interinstitucionales (nacionales e internacionales)**, **Convenios Internos (Deportistas, artistas, hijos de empleados docente y obreros, Samuel Robinson)**.
11. **Semestre.que.cursa**: Indica el semestre en curso del estudiante, esta variable es de tipo **categorico** y adopta los valores de **1er sem**, **2do sem**, **3er sem**, **4to sem**, **5to sem**, **6to sem**, **7mo sem**, **8vo sem**, **9no sem**.
12. **Ha.cambiado.usted.de.dirección**: Representa el cambio de dirección con respecto a la renovación anterior, adopta los valores de **Si** o **No** y es un tipo de dato **categorico**.

13. **De.ser.afirmativo..indique.el.motivo:** Explicación del cambio de dirección en caso de que haya ocurrido, esta variable es de tipo **string**.
14. **Número.de.materias.inscritas.en.el.semestre.o.año.anterior:** Indica la cantidad de materias inscritas en el semestre actual, es de tipo **entero** y toma los valores en el intervalo **[0,7]**.
15. **Número.de.materias.aprobadas.en.el.semestre.o.año.anterior:** Representa el número de materias aprobadas en el último semestre o año, adopta valores **1, 1,2, 3, 4, 5, 6, 7, más de 10**, esta variable es de tipo **string**.
16. **Número.de.materias.retiradas.en.el.semestre.o.año.anterior:** Indica el número de materias retiradas en el semestre o año anterior, es de tipo **entero** y adopta valores en el rango de **[0,10]**.
17. **Número.de.materias.reprobadas.en.el.semestre.o.año.anterior:** Indica el número de materias reprobadas en el último semestre o año, es de tipo **entero** y adopta los valores en el rango **[0,5]**.
18. **Promedio.ponderado.aprobado:** Representa el promedio ponderado de materias aprobadas de los estudiantes, es de tipo **flotante**.
19. **Eficiencia:** Representa la eficiencia del estudiante, es de tipo **flotante**.
20. **Si.reprobó.una.o.más.materias.indique.el.motivo:** Argumento explicativo del motivo por el cual se reprobó alguna materia, es de tipo **string**.
21. **Número.de.materias.inscritas.en.el.semestre.en.curso:** Número de materias totales inscritas en el semestre en curso, adopta valores en el rango de **[0,7]** y es de tipo **entero**.
22. **X.Estás.realizando.tesis...trabajo.de.grado.o.pasantías.de.grado:** Indica si el estudiante está realizando tesis, trabajo especial de grado o pasantías de grado, adopta los valores **Si** o **No** y es de tipo **categorico**.
23. **Tesis...trabajo.de.grado.o.pasantías.de.grado:** Indica el número de veces que se ha inscrito la tesis, trabajo especial de grado o pasantía de grado, adopta los valores **Más de dos veces**, **Primera vez** o **Segunda vez** y es de tipo **categorico**.
24. **Procedencia:** Indica el lugar desde el cual proviene el estudiante, adopta los valores **Altos Mirandinos**, **Anzoategui**, **Apure**, **Aragua**, **Barinas**, **Barlovento**, **Bolívar**, **Delta Amacuro**, **Guarenas - Guatire**, **Guárico**, **Lara**, **Monagas**, **Municipio Baruta**, **Municipio Chacao**, **Municipio El Hatillo**, **Municipio Libertador Caracas**, **Municipio Sucre**, **Mérida**, **Nueva Esparta**, **Portuguesa**, **Sucre**, **Trujillo**, **Táchira**, **Valles del Tuy**, **Vargas**, **Yaracuy**, y es de tipo **categorico**.

25. **Lugar.donde.reside.mientras.estudia.en.el.Universidad:** Indica el lugar actual en el que reside el estudiante, adopta los valores de **Altos Mirandinos, Guarenas - Guatire, Municipio Baruta, Municipio Chacao, Municipio El Hatillo, Municipio Libertador Caracas, Municipio Sucre, Valles del Tuy.**
26. **Personas.con.las.cuales.usted.vive..mientras.estudia.en.la.universidad:** Señala las personas con la que vive el estudiante, adopta los valores **Dueños del apartamento donde alquilo habitación, Esposo (a) Hijos (as), Familiares maternos, Familiares paternos, Hermana, Madre, Madre y Hermanos, Madre y hermano, Madre, Hermana, Abuela, Madre, Hermano y Sobrina, Mamá y Abuela, Mi Mamá y mi hijo, OTROS INQUILINOS, Padre Padres, hermana y abuelos maternos, Residencia, Solo, abuela, ambos padres y dos hermanis, compañeros de habitación alquilada, dos hermanos, dueña del apartamento, hermana, hermanas, hermano, hermano, hermana y mi hijo, madre y hermana, madre y hermana, madre y su esposo,abuela,y mi esposo, madre y su esposo,abuela,y mi esposo, madrina, prima, residencia, residencia, residencia estudiantil, sola, y es de tipo categórico.**
27. **Tipo.de.vivienda.donde.reside.mientras.estudia.en.la.universidad:** Indica el tipo de vivienda en el que vive el estudiante, adopta los valores **Apartamento en edificio, Apartamento en quinta - casa quinta o casa, Casa de vecindad, Casa en barrio rural, Casa en barrio urbano, Habitación alquilada, Quinta o casa quinta, Residencia estudiantil, casa, conserjería, y es de tipo categórico.**
28. **En.caso.de.vivir.en.habitación.alquilada.o.residencia.estudiantil..indique.el.monto.mensual:** Representa el monto por costo del alquiler, su tipo es **string**.
29. **Dirección.donde.se.encuentra.ubicada.la.residencia.o.habitación.alquilada:** Dirección del lugar de residencia alquilada por el estudiante, es de tipo **string**.
30. **X.Contrajo.Matrimonio:** Indica si el estudiante ha contraído matrimonio, adopta los valores **Si** o **No**, y es de tipo **categórico**.
31. **X.Ha.solicitado.algún.otro.beneficio.a.la.Universidad.u.otra.Institución:** Indica si el estudiante ha solicitado algún beneficio a la universidad o a otra institución como forma de apoyo, adopta los valores **Si** o **No**, y es de tipo **categórico**.
32. **En.caso.afirmativo.señale..año.de.la.solicitud..institución.y.motivo:** Explicación del motivo por el cual se solicitó el beneficio anteriormente, colocando el año y la institución a la que se le solicitó dicho beneficio, tipo **string**.

33. **X.Se.encuentra.usted..realizando.alguna.actividad.que.le.genere.ingresos:**
Indica si realiza alguna actividad que provea dinero de ingreso para el estudiante, adopta los valores **Si** o **No**, y es de tipo **categorico**.
34. **En.caso.de.ser.afirmativo..indique.tipo.de.actividad.y.su.frecuencia:**
Indicaca el tipo de actividad que realiza el estudiante como forma de ingreso, y adicionalmente indica la frecuencia con la que la realiza, esta variable de tipo **string**.
35. **Monto.mensual.de.la.beca:** Monto actual de la beca asignada al estudiante, es de tipo **entero**.
36. **Aporte.mensual.que.le.brinda.su.responsable.económico:** Monto monetario que recibe el estudiante por parte de su responsable económico, es de tipo **entero**.
37. **Aporte.mensual.que.recibe.de.familiares.o.amigos:** Monto monetario que recibe el estudiante por parte de familiares o amigos, es de tipo **entero**;
38. **Ingreso.mensual.que.recibe.por.actividades.a.destajo.o.por.horas:**
Monto monetario que recibe el estudiante producto de actividades realizadas fuera de clases, tipo **entero**.
39. **Ingreso.mensual.total:** Cantidad total monetaria que un estudiante ingresa mensualmente, tipo de dato **entero**.
40. **Alimentación:** Cantidad de dinero utilizado con fines alimenticios por parte del estudiante, tipo de dato **entero**.
41. **Transporte.público:** Cantidad de dinero utilizado con fines de transporte por parte del estudiante, tipo de dato **entero**.
42. **Gastos.médicos:** Cantidad de dinero utilizado con fines médicos por parte del estudiante, tipo de dato **entero**.
43. **Gastos.odontológicos:** Cantidad de dinero utilizado con fines odontológicos, tipo de dato **entero**.
44. **Gastos.personales:** Cantidad de dinero empleado a otros gastos personales necesarios del estudiante, tipo de dato **entero**.
45. **Residencia.o.habitación.alquilada:** Cantidad de dinero empleado para pagar el alquiler de la residencia o habitación en la que se aloja el estudiante, en caso de vivir alquilado, tipo de dato **entero**.
46. **Materiales.de.estudio:** Cantidad de dinero empleado para el costeo de los materiales de estudio, tipo de dato **entero**.
47. **Recreación:** Cantidad de dinero empleado para el costeo de actividades recreativas por parte del estudiante, tipo de dato **entero**.

48. **Otros.gastos:** Cantidad de dinero empleado para cubrir otros gastos que no fueron contemplados en las opciones anteriores, tipo de dato **entero**.
49. **Total.egresos:** Cantidad total de dinero que gasta el estudiante mensualmente, tipo de dato **entero**.
50. **Indique.quién.es.su.responsable.económico:** Persona responsable de la principal fuente de ingreso del estudiante, adopta los valores **Am-
bos padres, Cónyugue, Familiares, Hermano, MI HERMANA,
Madre, Padre, Usted mismo, abuela, esposo, hermana, ninguno,
tia**, tipo de dato **string**.
51. **Carga.familiar:** Carga familiar del responsable económico, tipo de dato **entero**.
52. **Ingreso.mensual.de.su.responsable.económico:** Cantidad de dinero que ingresa el responsable mensualmente, tipo de dato **flotante**.
53. **Otros.ingresos:** Cantidad de dinero que ingresa el responsable económico fuera de su ingreso mensual por trabajo, tipo de dato **flotante**.
54. **Total.de.ingresos:** Cantidad total de dinero que ingresa el responsable económico mensualmente, tipo de dato **flotante**.
55. **Vivienda:** Cantidad de dinero que gasta el responsable económico en la vivienda, tipo de dato **string**.
56. **Alimentación.1:** Cantidad de dinero que gasta el responsable económico en comida, tipo de dato **string**.
57. **Transporte:** Cantidad de dinero que gasta el responsable económico en transporte, tipo de dato **string**.
58. **Gastos.médicos.1:** Cantidad de dinero que gasta el responsable económico referentes a los gastos médicos, tipo de dato **string**.
59. **Gastos.odontológicos:** Cantidad de dinero que gasta el responsable económico en gastos odontológicos, tipo de dato **string**.
60. **Gastos.educativos:** Cantidad de dinero que gasta el responsable económico en educación, tipo de dato **string**.
61. **Servicios.públicos.de.agua..luz..teléfono.y.gas:** Cantidad de dinero que gasta el responsable económico en pago de servicios públicos, tipo de dato **entero**.
62. **Condominio:** Cantidad de dinero que gasta el responsable económico en el pago del condominio, tipo de dato **string**.
63. **Otros.gastos.1:** Cantidad de dinero que gasta el responsable económico en otros aspectos que no se mencionaron anteriormente, tipo de dato **flotante**.

- 64. **Total.de.egresos:** Cantidad total de dinero que egresa mensualmente el responsable económico, tipo de dato **flotante**.
- 65. **Deseamos.conocer.la.opinión.de.nuestros.usuarios(...):** Valor que indica la calidad del servicio con un puntaje de valores en el rango **[1,5]**, donde **1** es el valor más bajo y **5** el más alto, tipo de dato **categorico**.
- 66. **Sugerencias.y.recomendaciones.para.mejorar.nuestra.atención:** Variable de tipo **string**, que recoje las sugerencias escritas de los usuarios.

4 Propuesta

Dada la explicación del conjunto en el punto anterior, podemos observar aspectos tales como: **datos desordenados**, **nombres de variables poco entendibles**, **exceso de variables** (producto de redundancia de información), **incompatibilidad en los tipos de datos**, etc. Dicho esto se propone realizar un estudio a las **66 columnas** de tal forma de poder **ordenar** mejor la información, **reducir** el número de variables, **unificar** los tipos de datos de cada variable y **auditar** los registros que presenten inconsistencias. Con esto se logrará un conjunto de datos mucho más entendible y útil para las labores de minería de datos.