

Informe

Deyban Pérez

January 14, 2016

Abstract

A la hora de hacer **minería de datos** es primordial tener los datos bien organizados y estructurados, de tal manera que el proceso a la hora de seleccionar el algoritmo a utilizar sea más natural y fácil, adicionalmente, un buen conjunto de datos, será la base de buenos resultados extraídos del set de datos.

En el siguiente documento se explican las actividades realizadas sobre el conjunto de datos **data.csv** para generar una vista minable mucho más amigable para la organización de **Becas Crema**.

1 Eliminación de variables redundantes

Este punto corresponde al primer paso a realizar, que consiste en eliminar aquellas variables que poseen información redundante, que puede ser calculada mediante otras columnas, este proceso reduce la cantidad de variables, dejando sólo aquellas que poseen mayor información, cabe destacar, que este proceso se realizó con la menor cantidad de pérdida de información posible.

1.1 Actividades realizadas

1. Eliminar la columna **””**: Esta columna correspondía a un identificador del sistema que no estaba ordenado y que era redundante, ya que cada estudiante tiene un identificador unívoco que es su **Cédula de indentidad**, y este campo lo hace unívoco dentro del set de datos.
2. Eliminar la columna **Edad**: Esta columna no aporta información relevante debido a que existe una variable llamada **Fecha.de.Nacimiento..colocar.solo.datos.numéricos** con la cual se puede obtener la misma información.
3. Eliminar la columna **De.ser.afirmativo..indique.el.motivo**: Debido a que existe otra columna llamada **Ha.cambiado.usted.de.dirección** que adopta valores de **Si** o **No** y que en caso de ser **Si** la respuesta, el estudiante debe dar una explicación; vemos como ambas columnas tienen una relación y este caso la estrategia a adoptar fue la siguiente: Unificar el

valor de las columnas **Ha.cambiado.usted.de.dirección** con la columna **De.ser.afirmativo..indique.el.motivo**, de tal forma que en la columna **Ha.cambiado.usted.de.dirección** se mantengan los valores de **No** y los de **Si** sean reemplazados por la explicación, de esta forma se maximiza el contenido de la información de una sola columna y podemos eliminar la restante.

4. Eliminar la columna **Número.de.materias.inscritas.en.el.semestre.o.ano.anterior**: Esta variable proporciona información redundante debido a que se puede calcular con la sumatoria de las variables **Número.de.materias.aprobadas.en.el.semestre.o.ano.anterior**, **Número.de.materias.retiradas.en.el.semestre.o.ano.anterior** y **Número.de.materias.reprobadas.en.el.semestre.o.ano.anterior**.
5. Eliminar la columna **Tesis...trabajo.de.grado.o.pasantías.de.grado**: Esta columna tiene otra relacionada cuyo nombre es **X.Estas.realizado.tesis...trabajo.de.grado.o.pasantías.de.grado** cuyos valores adoptados son **Si** o **No**, la estrategia que se utilizó fue la de unificar las columnas de tal manera que toda la información quedara en la columna **X.Estas.realizado.tesis...trabajo.de.grado.o.pasantías.de.grado**, manteniendo los valores **No** y reemplazando los **Si** por el valor de la otra columna. Hay que mencionar que en esta columna hubo registros que tenían valor **No** y el número de veces que habían inscrito, en esos casos se dejó el valor **No**. Hubo otros casos en los que El valor **Si** no poseía el número de inscripciones, en ese caso se modificó dicho valor por la moda de los valores que tenían el valor **Si**.
6. Eliminar la columna **En.caso.de.vivir.en.habitacion(..)mensual**: Existe otra columna llamada **Residencia.o.habitacion.alquilada** que posee la misma información y más, por lo tanto se hizo una intersección entre las dos tablas, colocando como valor final el mayor entre los dos valores en la columna **Residencia(...)**.
7. Eliminar la columna **X.Contrajo.Matrimonio**: Esta columna presenta redundancia debido a que existe una columna con el nombre **Estado.Civil** de la cual se puede extraer la información.
8. Eliminar la columna **En.caso.afirmativo.señale..año.de.la.solicitud.institución.y.motivo**: Existe una columna relacionada llamada **X.Ha.solicitado.algún.otro.beneficio.a.la.Universidad.u.otra.Institución** que adopta los valores de **Si** o **No**, y al igual que el punto anterior, la estrategia utilizada fue unificar ambas columnas en una sola, que en este caso será la columna **X.Ha.solicitado.algún.otro.beneficio.a.la.Universidad.u.otra.Institucion**, dejando en ella los valores con el valor **No** y reemplazando los valores **Si** por la descripción de la otra columna. En este caso no fue necesario imputar, debido a que no habían datos anómalos.

9. Eliminar la columna **En.caso.de.ser.afirmativo..indique.tipo.de.actividad.y.su.frecuencia**: Existe una columna relacionada llamada **X.Se .encuentra.usted..realizando.alguna.actividad.que.le.genere.ingresos** que adopta valores de **Si** o **No**, la estrategia utilizada fue la de unificar ambas columnas en una sola, que en este caso será **X.Se .encuentra.usted..realizando.alguna.actividad.que.le.genere.ingresos**, manteniendo los valores **No** y reemplazando los valores **Si** por la información de la otra columna. En este caso no fue necesario imputar los datos.
10. Eliminar la columna **Ingreso.mensual.total**: Esta columna posee información redundante ya que se puede calcular de las otras columnas que indican el valor de los ingresos separados del estudiante.
11. Eliminar la columna **Total.egresos**: Esta columna es redundante, ya que la información se puede calcular con la sumatoria de las otras columnas que tienen los gastos de forma separada del estudiante.
12. Eliminar la columna **Total.de.ingresos** Esta columna es redundante, ya que la información se puede calcular con la sumatoria de las otras columnas que tienen los ingresos por separado del responsable económico.
13. Eliminar la columna **Total.de.egresos**: Esta columna es redundante, ya que la información se puede calcular con la sumatoria de las otras columnas que tienen los egresos por separado del responsable económico.

2 Renombramiento de las columnas

Se cambió el nombre de las columnas restantes, debido a que los nombres que presentaban las variables eran poco intuitivos y con un formato que no era el adecuado, los cambios fueron los siguientes:

1. Indique.el.Perodo.Academico.a.renovar ->**pRenovar**
2. Cedula.de.Identidad ->**cIdentidad**
3. Fecha.de.Nacimiento..colocar.solo.datos.numéricos ->**fNacimiento**
4. Estado.Civil ->**eCivil**
5. Sexo ->**sexo**
6. Escuela ->**escuela**
7. Año.de.Ingreso.a.la.UCV ->**aIngreso**
8. Modalidad.de.Ingreso.a.la.UCV ->**mIngreso**
9. Semestre.que.cursa ->**sCurso**
10. Ha.cambiado.usted.de.direccion ->**cDireccion**

11. Número(...)aprobadas(...)semestre(...)anterior ->**mAprobadas**
12. Número(...)retiradas(...)semestre(...)anterior ->**mRetiradas**
13. Número(...)reprobadas(...)semestre(...)anterior ->**mReprobadas**
14. Promedio.ponderado.aprobado ->**pAprobado**
15. Eficiencia ->**eficiencia**
16. Si.reprobo.una.o.mas.materias.indique.el.motivo ->**jReprobadas**
17. Número.de.materias.inscritas.en.el.semestre.en.curso ->**mInscritas**
18. X.Estás.realizado.tesis...trabajo.de.grado.o.pasantías.de.grado ->**tGrado**
19. Procedencia ->**lProcedencia**
20. Lugar.donde.reside.mientras.estudia.en.el.Universidad ->**lResidencia**
21. Personas.con.las.cuales.usted.vive..mientras.estudia(...) ->**pReside**
22. Tipo.de.vivienda.donde.reside.mientras.estudia(...) ->**tVivienda**
23. En.caso.de.vivir.en.habitación.alquilada(...)monto.mensual ->**gAlquiler**
24. Direccion(...)habitacion.alquilada->**dHAbitacion**
25. X.Ha.solicitado.algún.otro.beneficio(...) ->**oSolicitudes**
26. X.Se.encuentra.usted..realizando.alguna.actividad(...)ingresos ->**aEconomica**
27. Monto.mensual.de.la.beca ->**beca**
28. Aporte.mensual.que.le.brinda.su.responsable.económico ->**aResponsable**
29. Aporte.mensual.que.recibe.de.familiares.o.amigos ->**aOtros**
30. Ingreso.mensual.que.recibe.por.actividades.a.destajo.o.por.horas ->**iActividades**
31. Alimentación->**gAlimentacion**
32. Transporte.público->**gTransporte**
33. Gastos.médicos->**gMedicos**
34. Gastos.odontologicos ->**gOdontologicos**
35. Gastos.personales ->**gPersonales**
36. Materiales.de.estudio ->**gEstudios**
37. Recreacion ->**gRecreacion**
38. Otros.gastos ->**gOtros**

39. Indique.quien.es.su.responsable.economico ->**rEconomico**
40. Carga.familiar ->**crFamiliar**
41. Ingreso.mensual.de.su.responsable.economico->**irMensual**
42. Otros.ingresos ->**irOtros**
43. Vivienda ->**grVivienda**
44. Alimentacion.1->**grAlimentacion**
45. Transporte ->**grTransporte**
46. Gastos.medicos.1 ->**grMedicos**
47. Gastos.odontologicos ->**grOdontologicos**
48. Gastos.educativos ->**grEducativos**
49. Servicios.publicos.de.agua..luz..telefono.y.gas ->**grServicios**
50. Condominio ->**grCondominio**
51. Otros.gastos.1 ->**grOtros**
52. Deseamos.conocer.la.opinion.de.nuestros.usuarios(...) ->**rating**
53. Sugerencias.y.recomendaciones.para.mejorar.nuestra.atencion ->**sugerencias**

3 Reordenamiento de las columnas

De acuerdo al paper de **Tidy Data** que se mencionó en el enunciado de la tarea, las columnas fueron reordenadas para que se pudieran agrupar bajo un criterio que haga que su entendimiento sea más directo. La información se organizó en **siete** grupos, que son los siguientes:

1. **Datos Personales:** cIdentidad, fNacimiento, sexo, eCivil, escuela, aIngreso, mIngreso, sCurso, tGrado, mInscritas, pAprobado, eficiencia, mAprobadas, mRetiradas, mReprobadas, jReprobadas.
2. **Datos Renovación:** pRenovar, beca, lProcedencia, lResidencia, tVivienda, pReside, dHabitacion, cDireccion, oSolicitudes, aEconómica, rEconomico, crFamiliar.
3. **Ingresos Estudiante:** aResponsable, iActividades, aOtros.
4. **Gastos Estudiante:** gAlimentacion, gTransporte, gMedicos, gOdontologicos, gAlquiler, gEstudios gPersonales, gRecreacion, gOtros.
5. **Ingresos Responsable:** irMensual, irOtros.
6. **Gastos Responsable:** grAlimentacion, grTransporte, grMedicos, grOdontologicos, grEducativos, grVivienda, grServicios, grCondominio, grOtros.

4 Criterios de Estandarización e Imputación de los Datos

Los valores de las columnas tenían algunas inconsistencia con respecto a sus valores, los criterios adoptados para la estandarización e imputación de los datos fue la siguiente:

- **Numerización de los elementos categóricos:** Todas las variables con tipo de dato categórico fueron enumeradas.
- **Reemplazamiento de valores faltantes:** Si la variable es de tipo **categórico**, entonces el valor se reemplaza por la moda, si la variable es de tipo **entero** o **real**, el valor se reemplaza por el elemento neutro, que en este caso es el cero.

5 Explicación de los datos

1. **cIdentidad:** Cédula de Identidad del estudiante, tipo **entero**.
2. **fNacimiento:** Fecha de Nacimiento del estudiante, tipo **date** (dd/mm/aaaa)
3. **sexo:** Sexo del estudiante, adopta los valores **0** para femenino y **1** para masculino, tipo **binario**.
4. **eCivil:** Estado Civil del estudiante, adopta los valores **0** (Soltero), **1** (Casado), **2** (Viudo), **3** (Unido), tipo **categórico**.
5. **escuela:** Escuela a la que pertenece el estudiante, adopta los valores **0** (Enfermería) y **1** (Bionálisis), tipo **textbfbinario**.
6. **aIngreso:** Año de Ingreso a universidad por parte del estudiante, tipo **entero**.
7. **mIngreso:** Modalidad de Ingreso a la universidad por parte del estudiante, adopta los valores **0** (Asignado OPSU), **1** (Convenios Interinstitucionales (nacionales e internacionales)), **2** (Convenios Internos (Deportistas, artistas, hijos empleados docente y obreros, Samuel Robinson)), **3** (Prueba Interna y/o propedeútico), tipo **categórico**.
8. **sCurso:** Semestre en Curso del estudiante, adopta valores en intervalo **[1,10]**, tipo **entero**.
9. **tGrado:** Tesis de Grado, indica el número de veces que ha sido inscrita por el estudiante, adopta los valores **0** (Ninguna), **1** (Primera vez), **2** (Segunda Vez), **3** (Más de dos), tipo **categórico**.
10. **mInscritas:** Número de Materias Inscritas por el estudiante en el semestre actual, adopta valores en el rango **[1,7]**, tipo **entero**.

11. **pAprobado:** Promedio Ponderado Aprobado del estudiante, tipo **real**.
12. **eficiencia:** Eficiencia del estudiante, tipo **real**.
13. **mAprobadas:** Número de Materias Aprobadas por el estudiante en el último semestre o año, adopta valores en el rango **[0,11]**, donde entre **[0,10]** tiene un valor literal del número y el valor **11** expresa: **Más de 10 materias aprobadas**, tipo **entero**.
14. **mRetiradas:** Número de Materias Retiradas por el estudiante en el semestre o año anterior, tipo **entero**.
15. **mReprobadas:** Número de Materias Reprobadas por el estudiante en el semestre o año anterior, tipo **entero**.
16. **jReprobadas:** Justificación de las Materias Reprobadas, adopta los valores **No** en caso no poseer justificación y cualquier cadena de caracteres en caso contrario, tipo **string**.
17. **pRenovar:** Período a Renovar del estudiante, adopta los valores **0** (I-2014), **1** (II-2014), **2** (I-2015), **3** (II-2015), tipo **categorico**.
18. **beca:** Monto de la Beca del estudiante, tipo **entero**.
19. **lProcedencia:** Lugar de Procedencia del estudiante, adopta los valores **0** (Altos Mirandinos), **1** (Anzoategui), **2** (Apure), **3** (Aragua), **4** (Barinas), **5** (Barlovento), **6** (Bolívar), **7** (Delta Amacuro), **8** (Guarenas-Guatire), **9** (Guárico), **10** (Lara), **11** (Monagas), **12** (Municipio Baruta), **13** (Municipio Chacao), **14** (Municipio El Hatillo), **15** (Municipio Libertador), **16** (Municipio Sucre), **17** (Mérida), **18** (Nueva Esparta), **19** (Portuguesa), **20** (Sucre), **21** (Trujillo), **22** (Táchira), **23** (Valles del Tuy), **24** (Vargas), **25** (Yaracuy), tipo **Categorico**.
20. **lResidencia:** Lugar de Residencia del estudiante, adopta los valores **0** (Altos Mirandinos), **1** (Guarenas-Guatire), **2** (Municipio Baruta), **3** (Municipio Chacao), **4** (Municipio El Hatillo), **5** (Municipio Libertador), **6** (Municipio Sucre), **7** (Valles del Tuy).
21. **tVivienda:** Tipo de Vivienda donde reside el estudiante, adopta los valores **0** (Casa), **1** (Apartamento), **2** (Habitación), **3** (Residencia Estudiantil), **4** (Conserjería).
22. **pReside:** Personas con las que Reside el estudiante, adopta los valores **0** (Solo), **1** (Ambos Padres), **2** (Mamá), **3** (Padre), **4** (Hermano (a)), **5** (Esposo (a)), **6** (Familiares), **7** (Otros), tipo **categorico**.
23. **dHabitacion:** Dirección de Habitación Alquilada, adopta los valores **No** en caso de no vivir alquilado y cualquier otra cadena de caracteres en caso contrario, tipo **string**.

- 24. **cDireccion:** Cambio de Dirección, adopta los valores **No** en caso de no vivir alquilado y cualquier otra cadena de caracteres en caso contrario, tipo **string**.
- 25. **oSolicitudes:** Otras Solicitudes, adopta los valores **No** en caso de no vivir alquilado y cualquier otra cadena de caracteres en caso contrario, tipo **string**.
- 26. **aEconomica:** Actividad Económica que ejerce el estudiante, adopta los valores **No** en caso de no vivir alquilado y cualquier otra cadena de caracteres en caso contrario, tipo **string**.
- 27. **rEconomico:** Responsable Económico, adopta los valores **0** (Nadie), **1** (Ambos Padres), **2** (Mamá), **3** (Papá), **4** (Hermano (a)), **5** (Cónyugue), **6** (Familiar), tipo **categorico**.
- 28. **crFamiliar:** Carga Familiar del Responsable Económico, tipo **entero**.
- 29. **aResponsable:** Aporte del Responsable Económico al estudiante, tipo **entero**.
- 30. **iActividades:** Ingreso de Actividades realizadas por el estudiante, tipo **entero**.
- 31. **aOtros:** Otros Aportes, tipo **entero**.
- 32. **gAlimentacion:** Gastos de Alimentación del estudiante, tipo **entero**.
- 33. **gTransporte:** Gastos de Transporte del estudiante, tipo **entero**.
- 34. **gMedicos:** Gastos Médicos del estudiante, tipo **entero**.
- 35. **gOdontologicos:** Gastos Odontológicos del estudiante, tipo **entero**.
- 36. **gAlquiler:** Gastos de Alquiler del estudiante, tipo **entero**.
- 37. **gEstudios:** Gastos de Estudios del estudiante, tipo **entero**.
- 38. **gPersonales:** Gastos Personales del estudiantre, tipo **entero**.
- 39. **gRecreacion:** Gastos de Recreación del estudiante, tipo .
- 40. **gOtros:** Otros Gastos del estudiante, tipo **entero**.
- 41. **irMensual:** Ingreso Mensual del responsable económico, tipo **real**.
- 42. **irOtros:** Otros Ingresos del Responsable Económico, tipo **real**.
- 43. **grAlimentacion:** Gastos de Alimentación del Responsable Económico, tipo **real**.
- 44. **grTransporte:** Gatos de Transporte del Responsable Económico, tipo **real**.

45. **grMedicos**: Gastos Meédicos del Responsable Económico, tipo **real**
46. **grOdontologicos**: Gastos Odontológicos del Responsable Económico, tipo **entero**.
47. **grEducativos**: Gastos Educativos del Responsable Económico, tipo **entero**.
48. **grVivienda**: Gastos de Vivienda del Responsable Económico, tipo **real**.
49. **grServicios**: Gastos de Servicios del Responsable Económico, tipo **real**.
50. **grCondominio**: Gastos del Condominio por parte del Responsable Económico, tipo **real**.
51. **grOtros**: Otros Gastos por parte del Responsable Económico, tipo **real**.
52. **rating**: Rating del servicio, adopta valores en el intervalo $[1,5]$, donde 1 es lo **peor** y 5 es lo **mejor**, tipo **categorico**.
53. **sugerencias**: Cadena de caracteres que brinda una sugerencia al servicio, tipo **string**.

6 Análisis Exploratorio de los datos

| | cIdentidad | sexo | eCivil | escuela | aIngreso |
|-------|------------|------------|------------|------------|-------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 95.500000 | 0.142105 | 0.036842 | 0.678947 | 2010.294737 |
| std | 54.992424 | 0.350081 | 0.279286 | 0.468114 | 2.007255 |
| min | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 2005.000000 |
| 25% | 48.250000 | 0.000000 | 0.000000 | 0.000000 | 2009.000000 |
| 50% | 95.500000 | 0.000000 | 0.000000 | 1.000000 | 2010.000000 |
| 75% | 142.750000 | 0.000000 | 0.000000 | 1.000000 | 2012.000000 |
| max | 190.000000 | 1.000000 | 3.000000 | 1.000000 | 2014.000000 |

Figure 1: Tabla 1.

| | mIngreso | sCurso | tGrado | mInscritas | pAprobado |
|-------|------------|------------|------------|------------|------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 1.826316 | 5.336842 | 0.410526 | 3.510526 | 13.580033 |
| std | 1.435138 | 2.337233 | 0.809922 | 1.612581 | 1.759438 |
| min | 0.000000 | 1.000000 | 0.000000 | 1.000000 | 0.600000 |
| 25% | 0.000000 | 4.000000 | 0.000000 | 2.000000 | 12.669250 |
| 50% | 3.000000 | 6.000000 | 0.000000 | 4.000000 | 13.582500 |
| 75% | 3.000000 | 7.000000 | 1.000000 | 5.000000 | 14.587750 |
| max | 3.000000 | 9.000000 | 3.000000 | 7.000000 | 17.220000 |

Figure 2: Tabla 2.

| | eficiencia | mAprobadas | mRetiradas | mReprobadas | pRenovar |
|-------|------------|------------|------------|-------------|------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 0.899466 | 3.315789 | 0.215789 | 0.363158 | 1.384211 |
| std | 0.148417 | 1.561868 | 0.903118 | 0.809801 | 0.577133 |
| min | 0.357100 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.834425 | 2.000000 | 0.000000 | 0.000000 | 1.000000 |
| 50% | 0.972200 | 4.000000 | 0.000000 | 0.000000 | 1.000000 |
| 75% | 1.000000 | 4.000000 | 0.000000 | 0.000000 | 2.000000 |
| max | 1.000000 | 11.000000 | 10.000000 | 5.000000 | 3.000000 |

Figure 3: Tabla 3.

| | beca | lProcedencia | lResidencia | tVivienda | rEconomico |
|-------|--------------|--------------|-------------|------------|------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 1566.315789 | 13.463158 | 4.268421 | 0.657895 | 2.521053 |
| std | 987.861302 | 6.380046 | 2.014905 | 0.737374 | 1.157940 |
| min | 200.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 1500.000000 | 8.000000 | 3.250000 | 0.000000 | 2.000000 |
| 50% | 1500.000000 | 15.000000 | 5.000000 | 1.000000 | 2.000000 |
| 75% | 1500.000000 | 16.000000 | 5.000000 | 1.000000 | 3.000000 |
| max | 15000.000000 | 25.000000 | 7.000000 | 4.000000 | 6.000000 |

Figure 4: Tabla 4.

| | crFamiliar | aResponsable | iActividades | aOtros | gAlimentacion |
|-------|------------|--------------|--------------|-------------|---------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 3.121053 | 1618.947368 | 107.894737 | 331.763158 | 1232.368421 |
| std | 1.264526 | 1414.871485 | 419.360548 | 718.473302 | 2319.777441 |
| min | 1.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 2.000000 | 600.000000 | 0.000000 | 0.000000 | 412.500000 |
| 50% | 3.000000 | 1000.000000 | 0.000000 | 0.000000 | 1000.000000 |
| 75% | 4.000000 | 2000.000000 | 0.000000 | 500.000000 | 1500.000000 |
| max | 8.000000 | 9000.000000 | 3000.000000 | 5000.000000 | 30000.000000 |

Figure 5: Tabla 5.

| | gTransporte | gMedicos | gOdontologicos | gAlquiler | gEstudios |
|-------|-------------|--------------|----------------|-------------|-------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 608.084211 | 345.000000 | 194.736842 | 392.578947 | 1040.894737 |
| std | 588.287366 | 869.69176 | 443.529237 | 1164.303358 | 702.272004 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 100.000000 |
| 25% | 250.000000 | 0.000000 | 0.000000 | 0.000000 | 500.000000 |
| 50% | 500.000000 | 0.000000 | 0.000000 | 0.000000 | 1000.000000 |
| 75% | 750.000000 | 500.000000 | 0.000000 | 0.000000 | 1500.000000 |
| max | 3500.000000 | 10000.000000 | 2500.000000 | 6500.000000 | 4000.000000 |

Figure 6: Tabla 6.

| | gPersonales | gRecreacion | gOtros | irMensual | irOtros |
|-------|-------------|-------------|--------------|----------------|--------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 569.000000 | 274.315789 | 326.500000 | 76005.882158 | 2250.997263 |
| std | 557.436071 | 446.952414 | 861.794601 | 651399.661392 | 3273.834547 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 200.000000 | 0.000000 | 0.000000 | 6507.250000 | 0.000000 |
| 50% | 500.000000 | 0.000000 | 200.000000 | 8500.000000 | 0.000000 |
| 75% | 800.000000 | 500.000000 | 300.000000 | 12347.735000 | 4000.000000 |
| max | 3000.000000 | 2500.000000 | 10900.000000 | 8773759.000000 | 17879.000000 |

Figure 7: Tabla 7.

| | grAlimentacion | grTransporte | grMedicos | grEducativos | grVivienda |
|-------|----------------|--------------|--------------|--------------|-------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 6113.068421 | 1326.778947 | 1364.497474 | 954.600000 | 387.357895 |
| std | 4402.714218 | 1526.241052 | 1787.681958 | 1640.818394 | 1051.649900 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 3500.000000 | 500.000000 | 0.000000 | 0.000000 | 0.000000 |
| 50% | 5000.000000 | 1000.000000 | 1000.000000 | 0.000000 | 0.000000 |
| 75% | 8000.000000 | 2000.000000 | 2000.000000 | 1500.000000 | 0.000000 |
| max | 36000.000000 | 12960.000000 | 12000.000000 | 10000.000000 | 6000.000000 |

Figure 8: Tabla 8.

| | grServicios | grCondominio | grOtros | rating |
|-------|-------------|--------------|---------------|------------|
| count | 190.000000 | 190.000000 | 190.000000 | 190.000000 |
| mean | 770.700000 | 420.686579 | 3432.277684 | 4.547368 |
| std | 768.695686 | 711.307975 | 32199.561638 | 0.701750 |
| min | 0.000000 | 0.000000 | 0.000000 | 1.000000 |
| 25% | 300.000000 | 0.000000 | 157.500000 | 4.000000 |
| 50% | 600.000000 | 0.000000 | 700.000000 | 5.000000 |
| 75% | 1000.000000 | 500.000000 | 1500.000000 | 5.000000 |
| max | 5650.000000 | 3500.000000 | 444482.000000 | 5.000000 |

Figure 9: Tabla 9.

7 Gráficos

Se graficaron las variables numéricas.

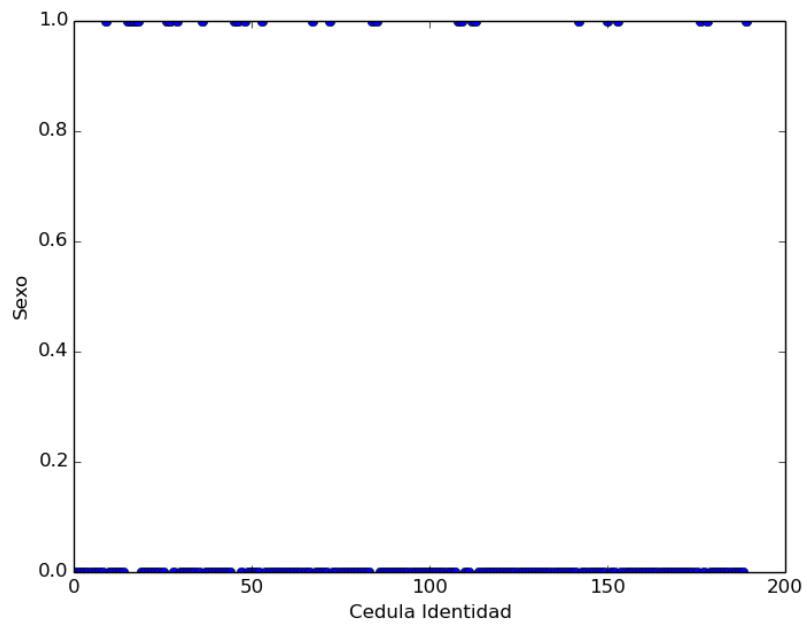


Figure 10: sexo.

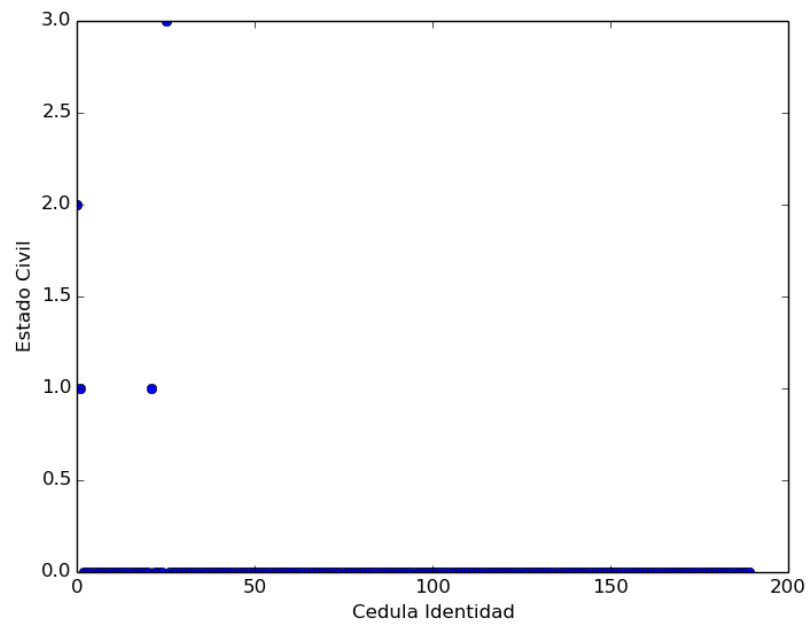


Figure 11: eCivil.

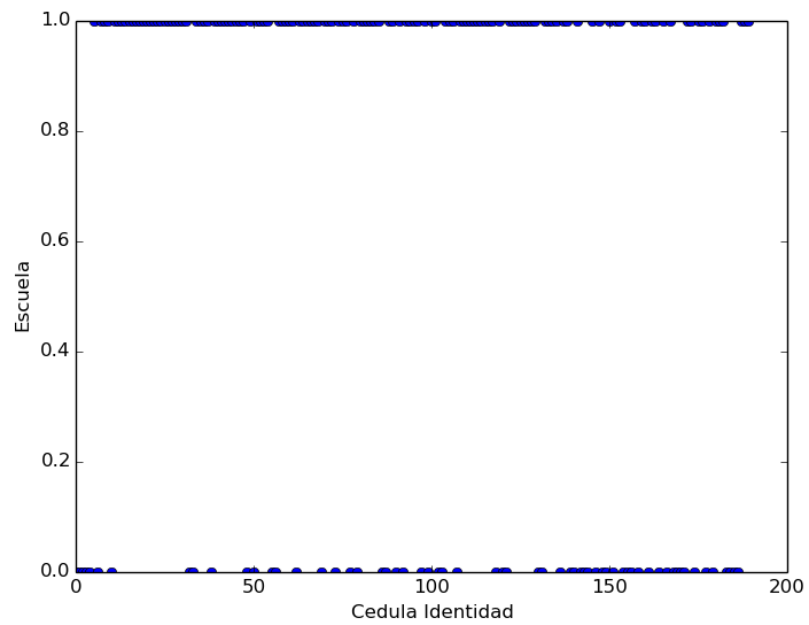


Figure 12: escuela.

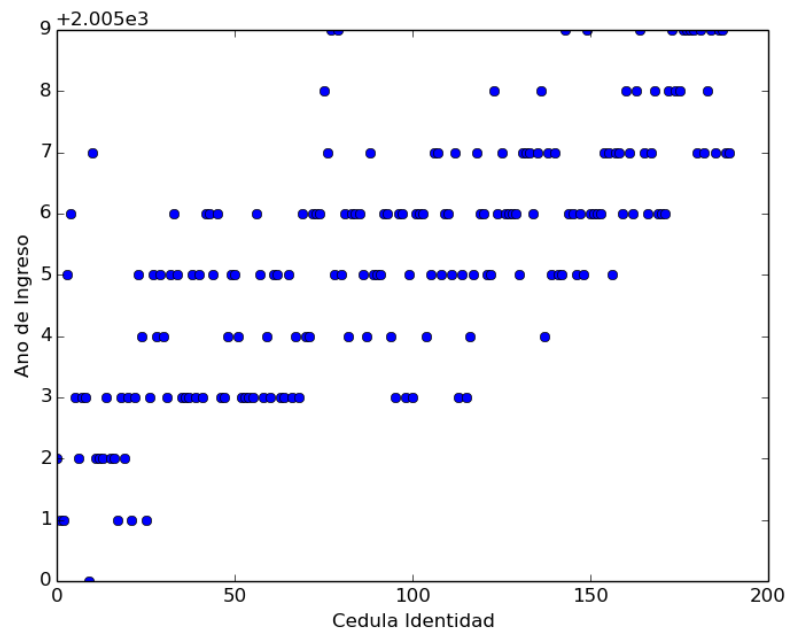


Figure 13: aIngreso.

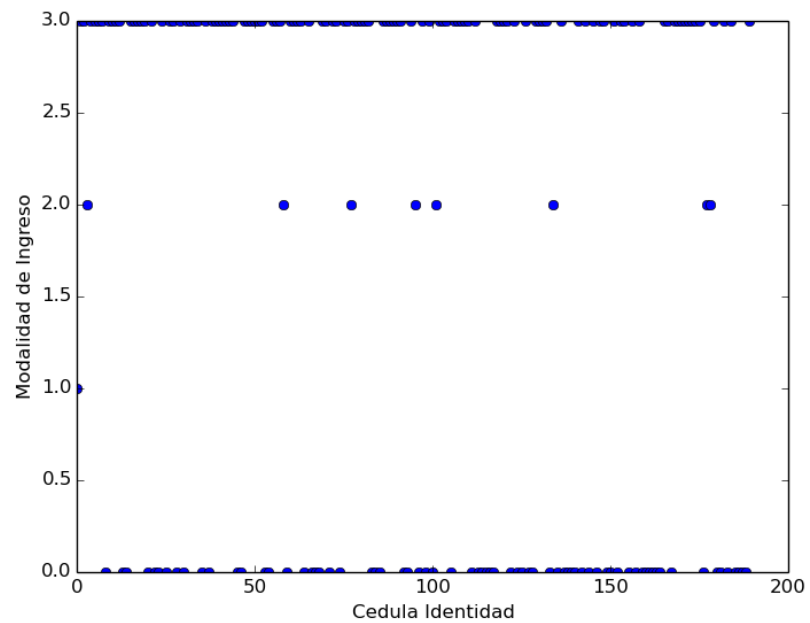


Figure 14: mIngreso.

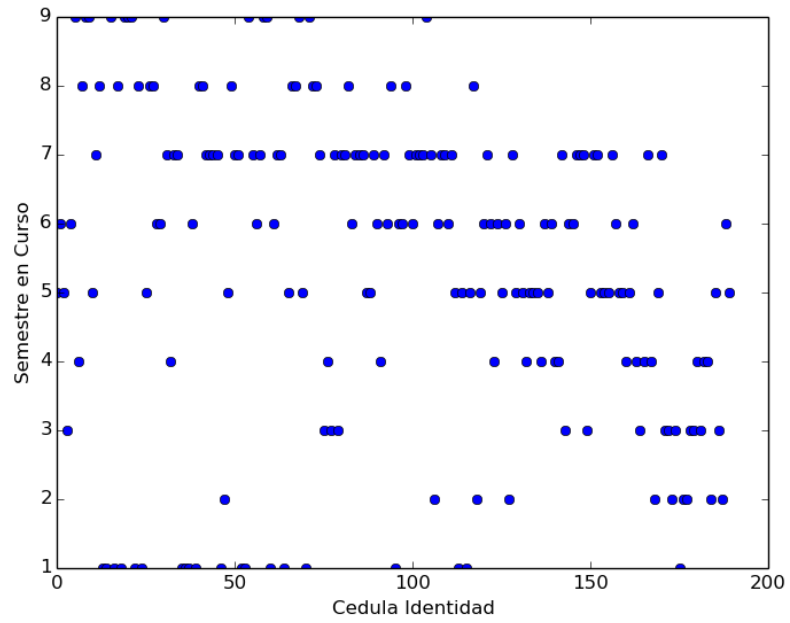


Figure 15: sCurso.

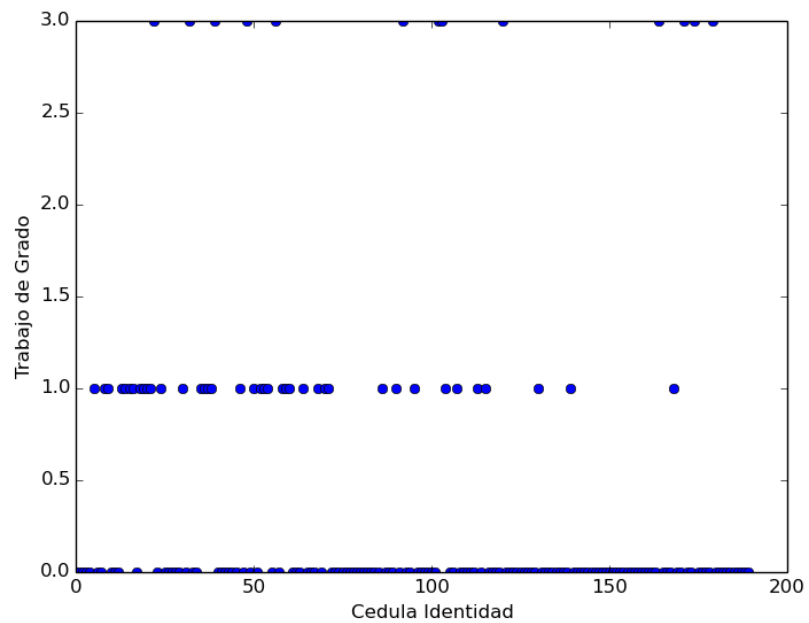


Figure 16: tGrado.

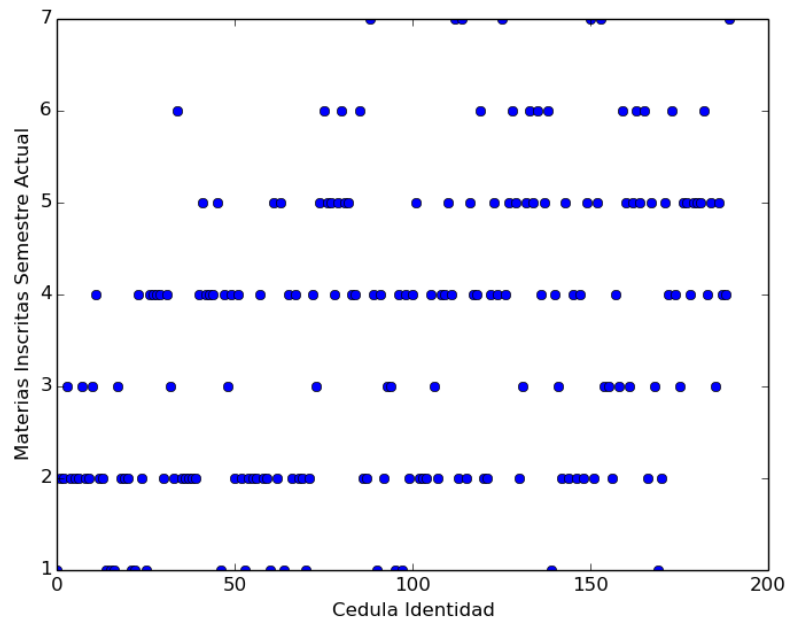


Figure 17: mInscritas.

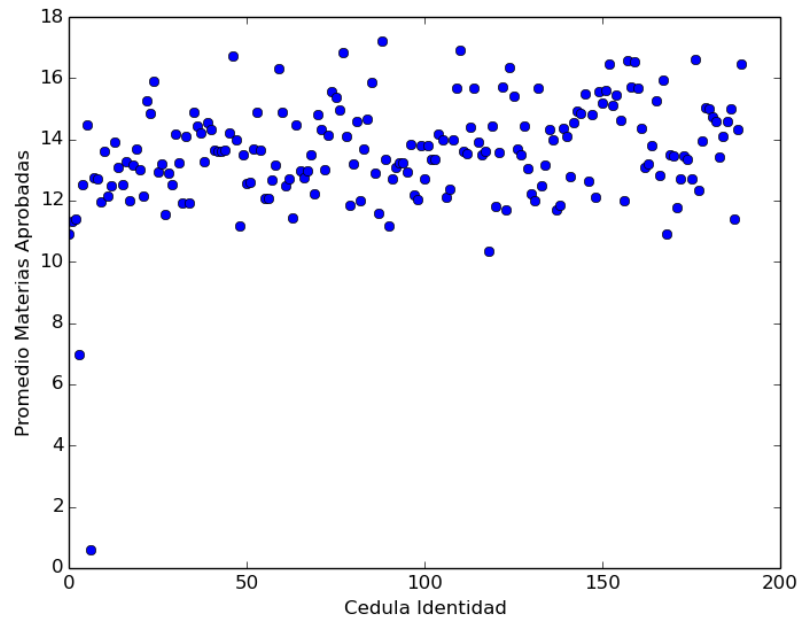


Figure 18: pAprobado.

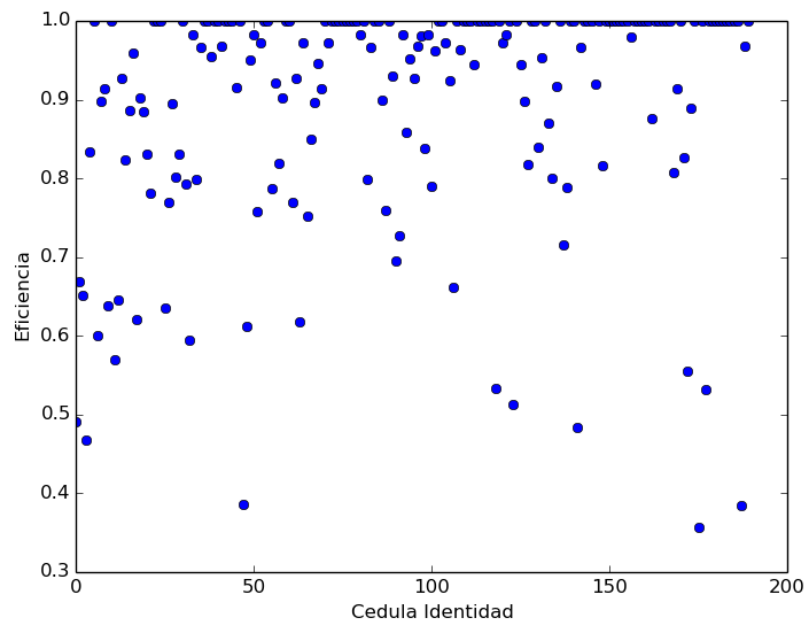


Figure 19: eficiencia.

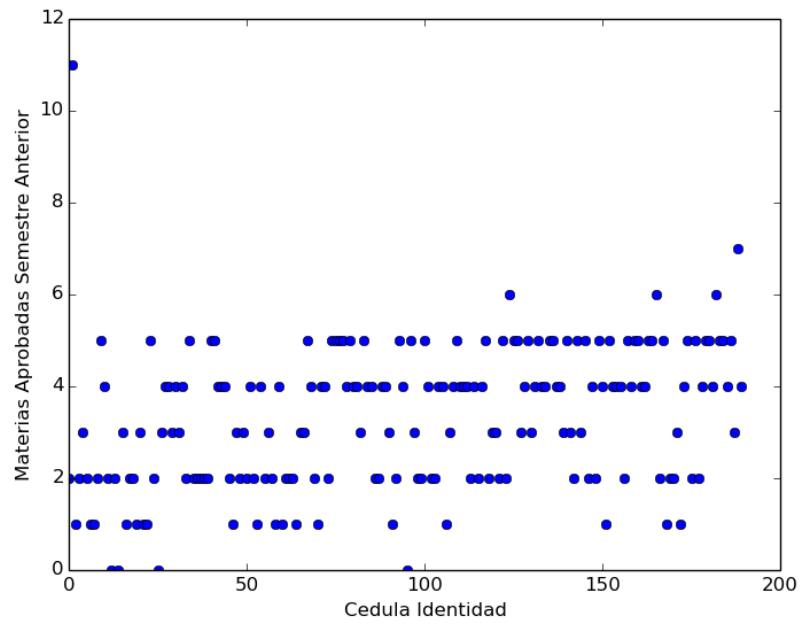


Figure 20: mAprobadas.

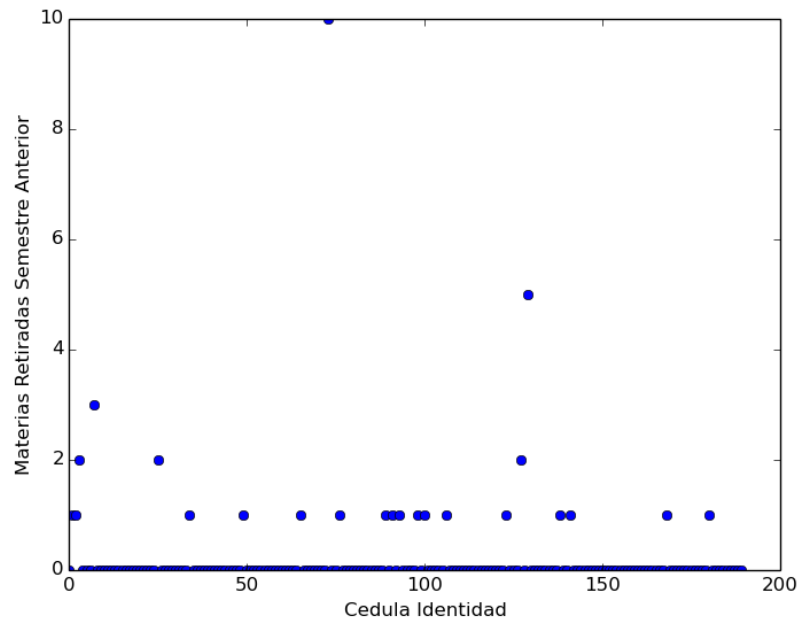


Figure 21: mRetiradas.

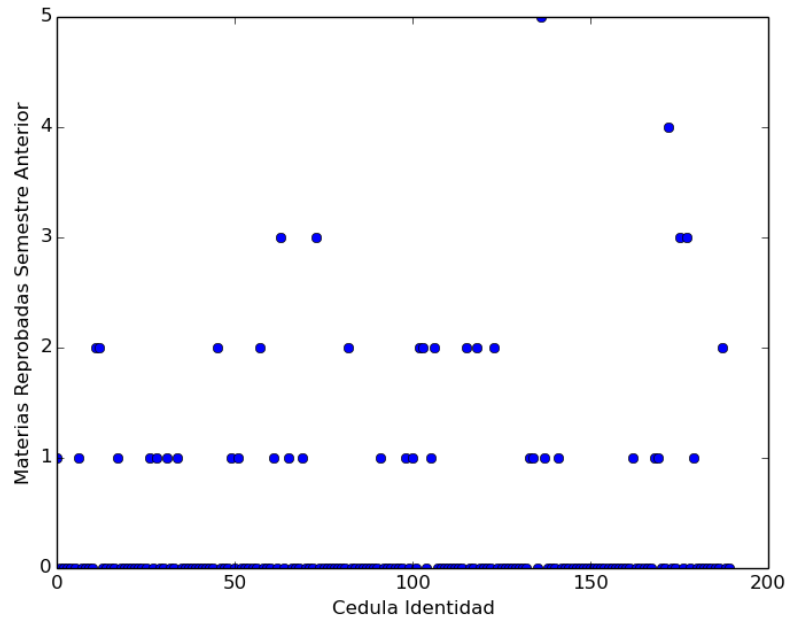


Figure 22: mReprobadas.

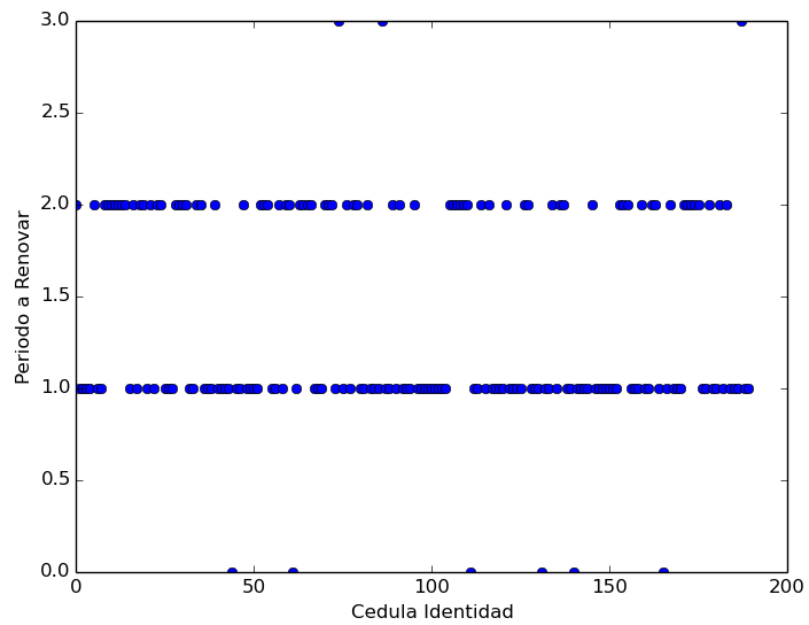


Figure 23: pRenovar.

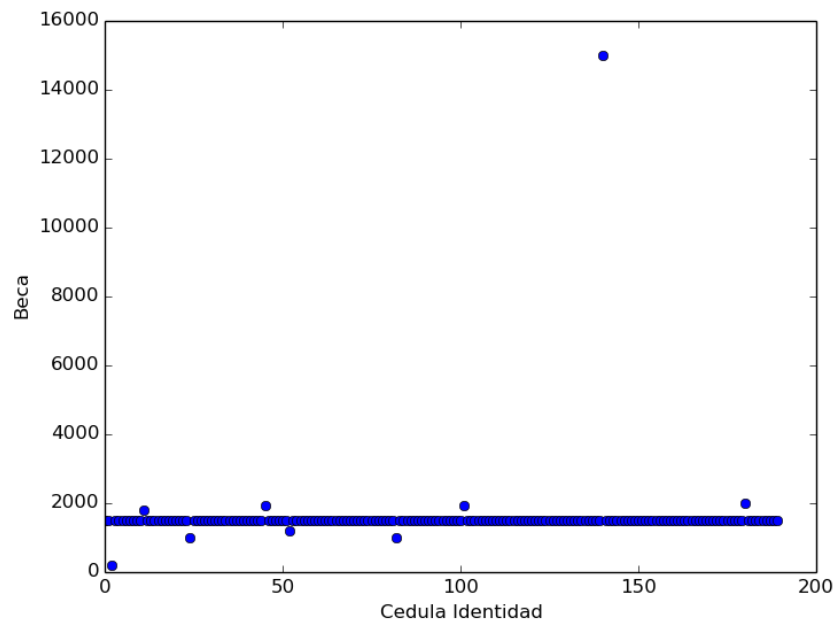


Figure 24: beca.

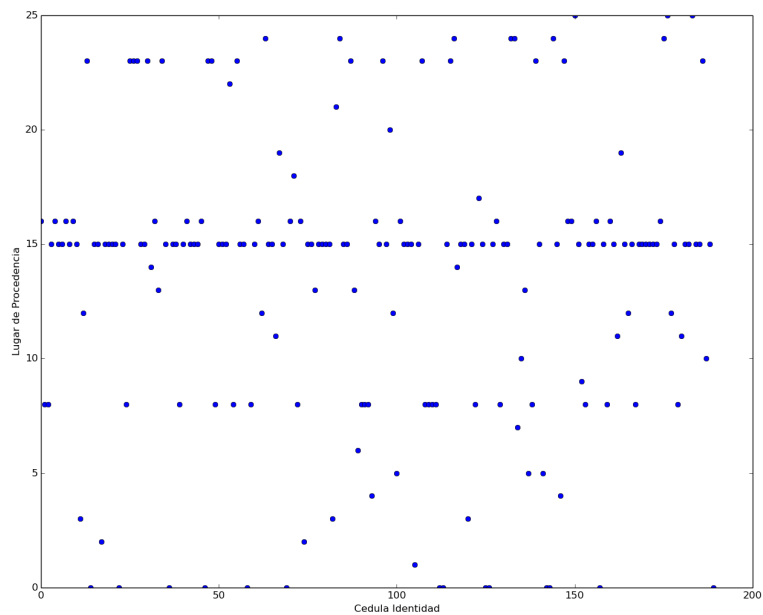


Figure 25: lProcedencia.

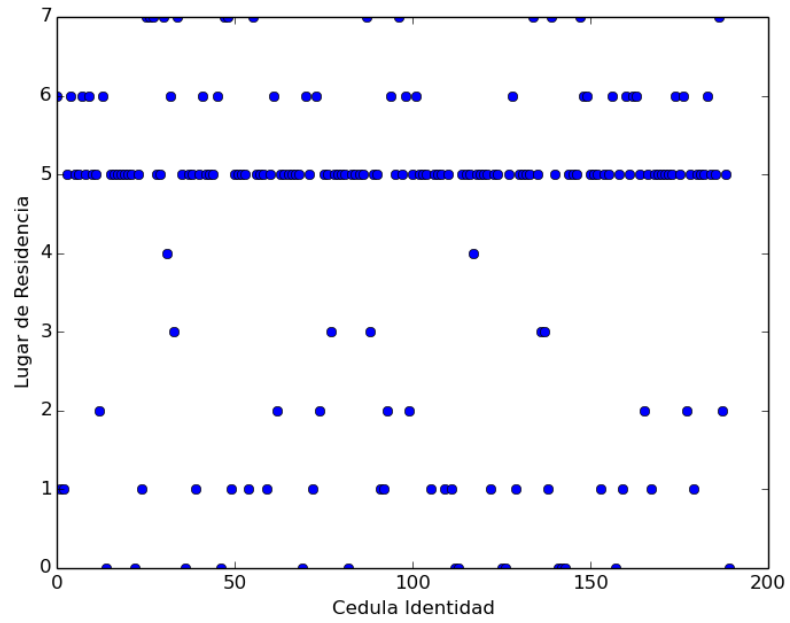


Figure 26: lResidencia.

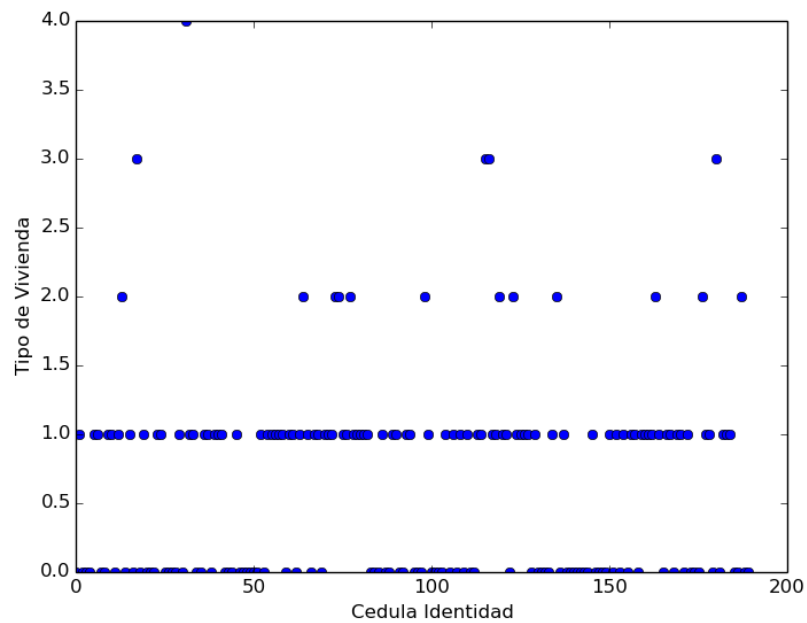


Figure 27: tVivienda.

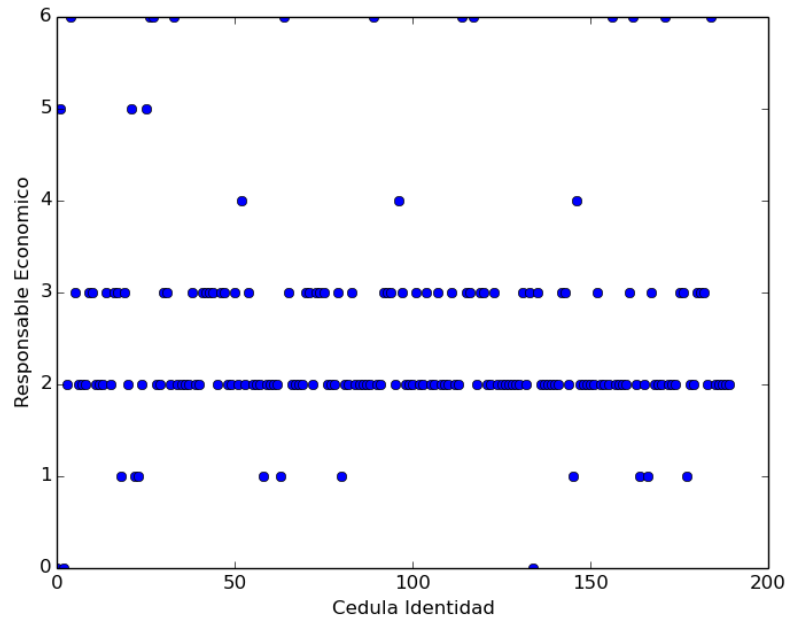


Figure 28: rEconomico.

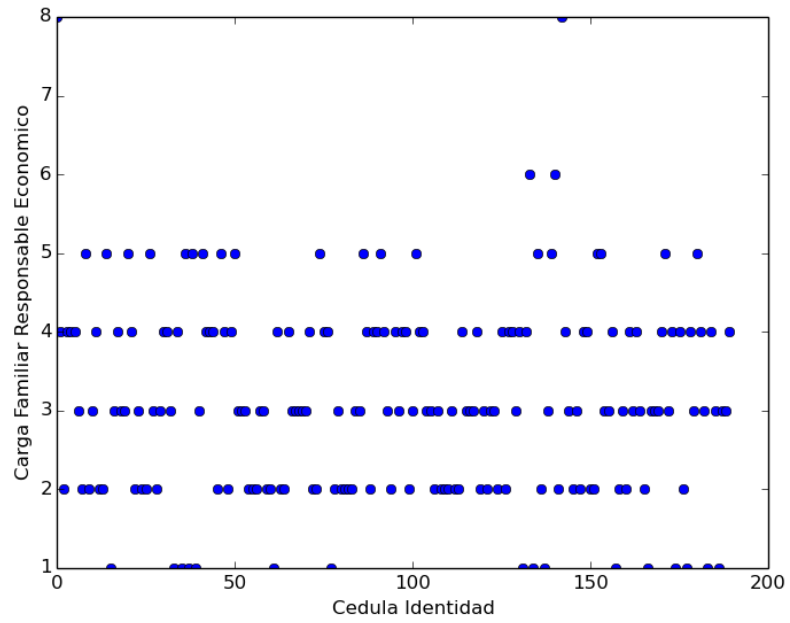


Figure 29: crFamiliar.

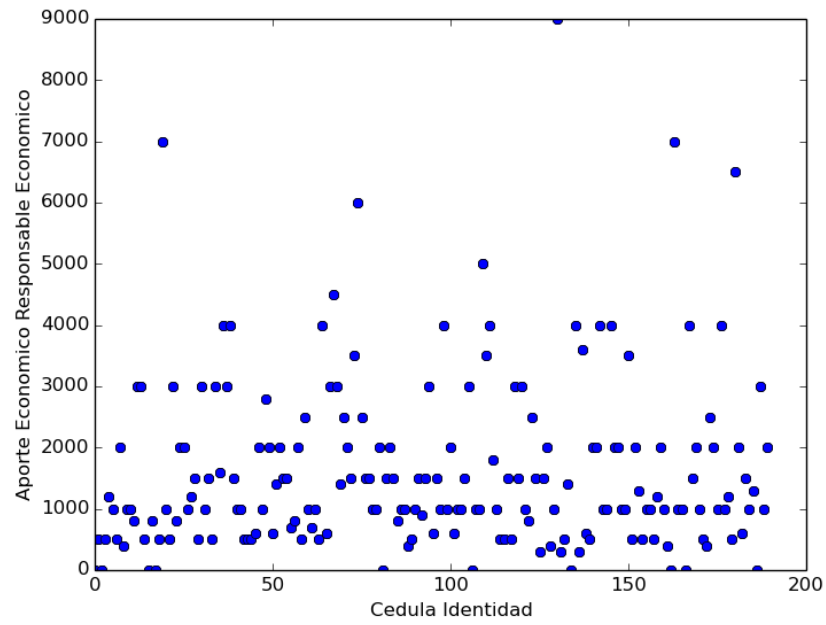


Figure 30: aResponsable.

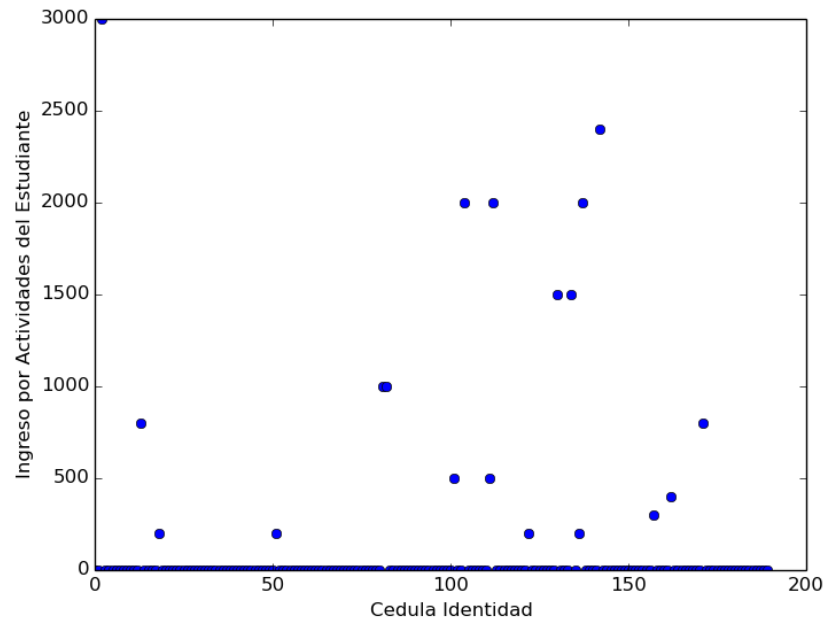


Figure 31: iActividades.

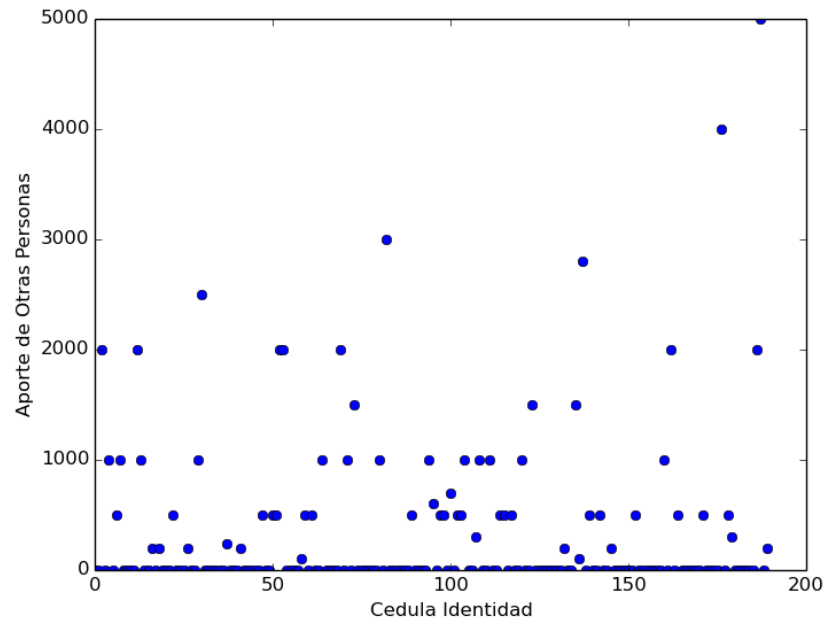


Figure 32: aOtros.

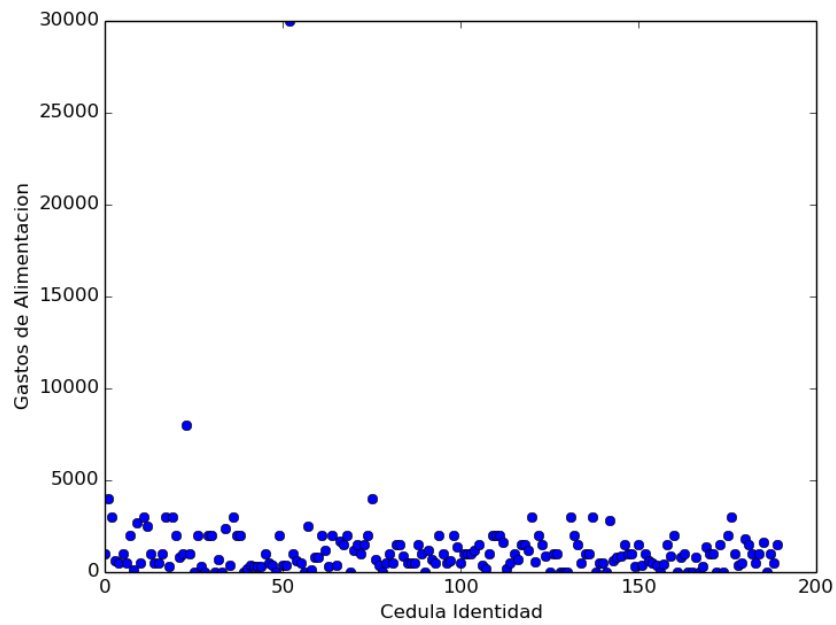


Figure 33: gAlimentacion.

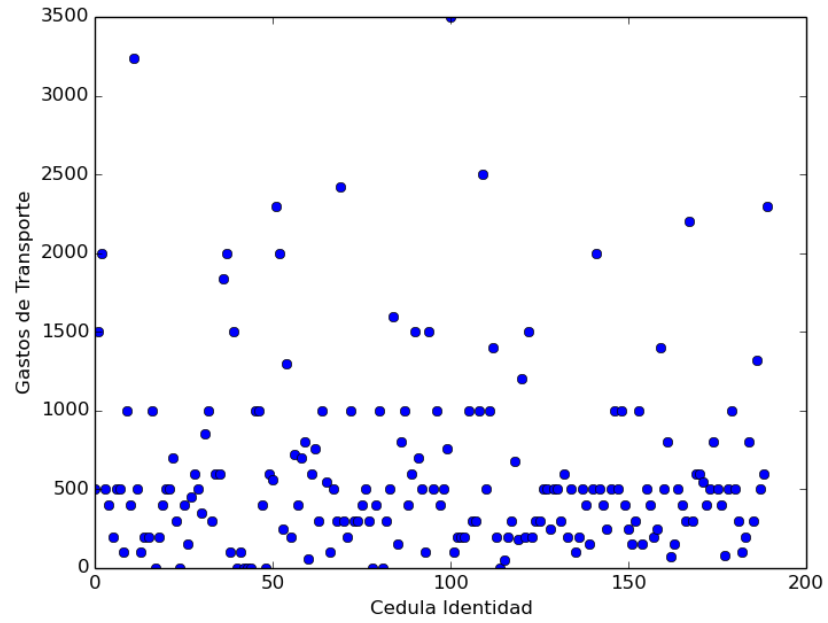


Figure 34: gTransporte.

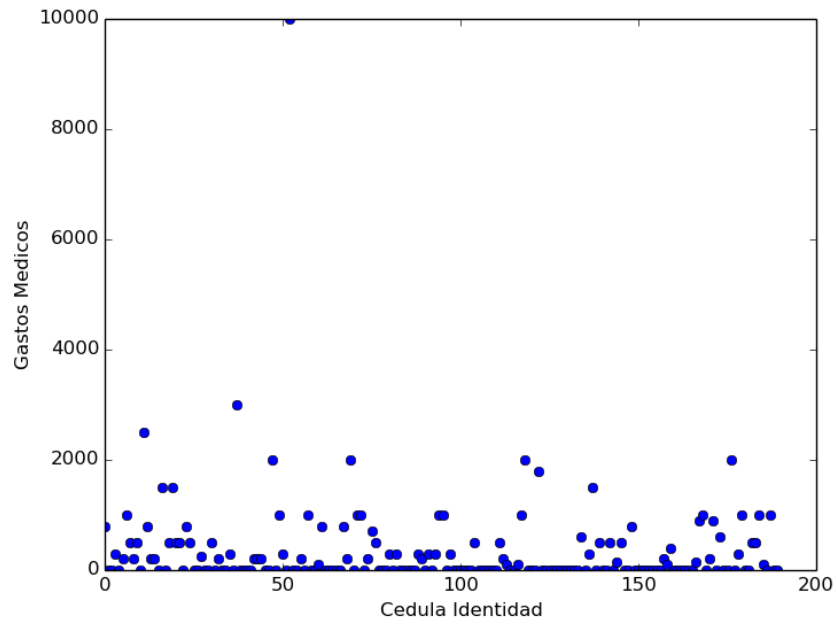


Figure 35: gMedicos.

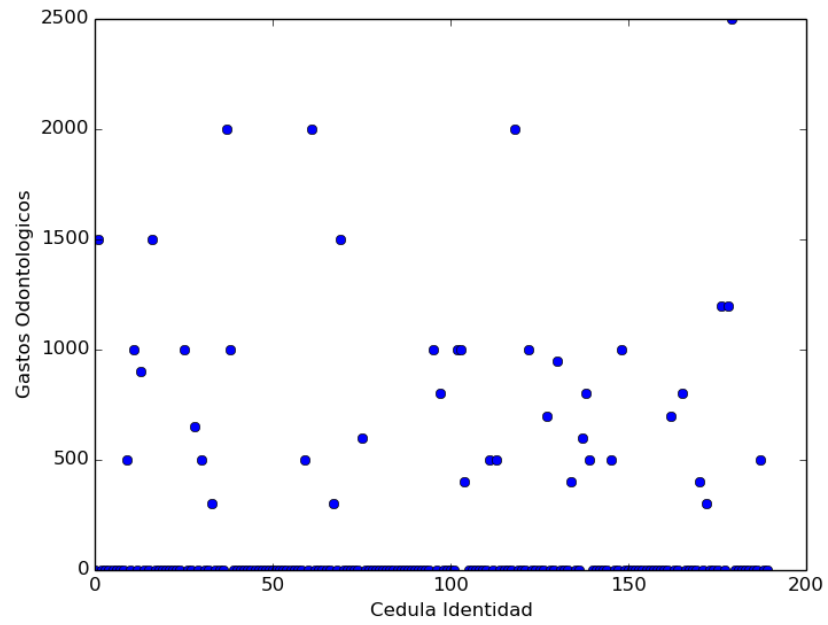


Figure 36: gOdontologicos.

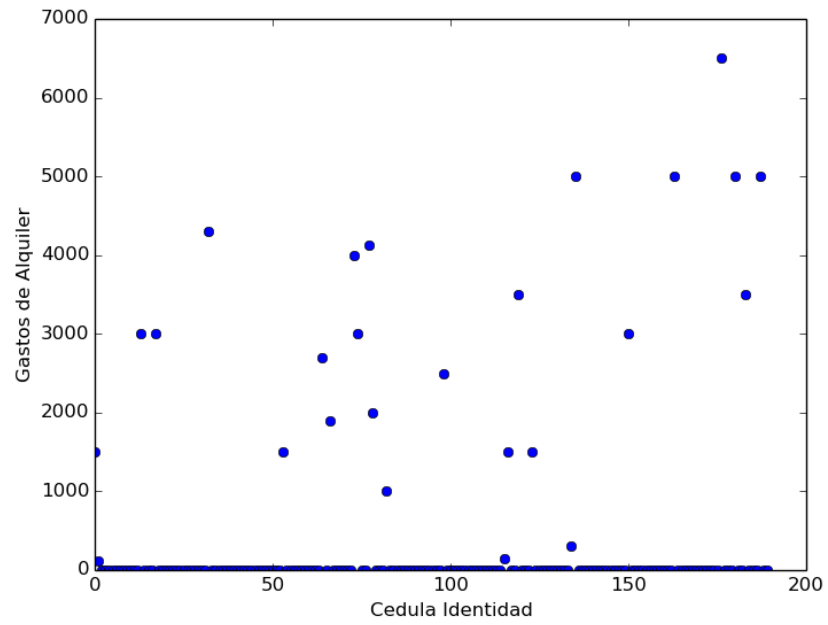


Figure 37: gAlquiler.

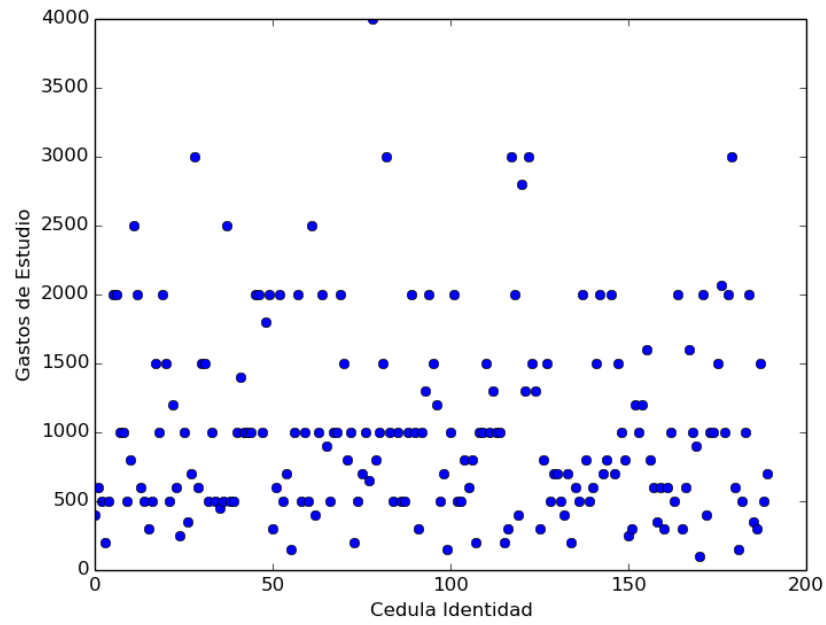


Figure 38: gEstudios.

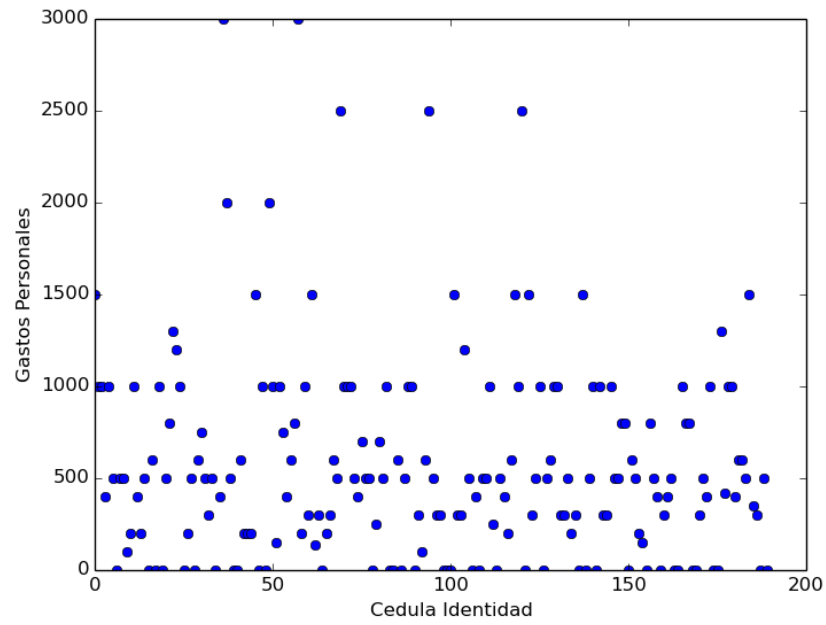


Figure 39: gPersonales.

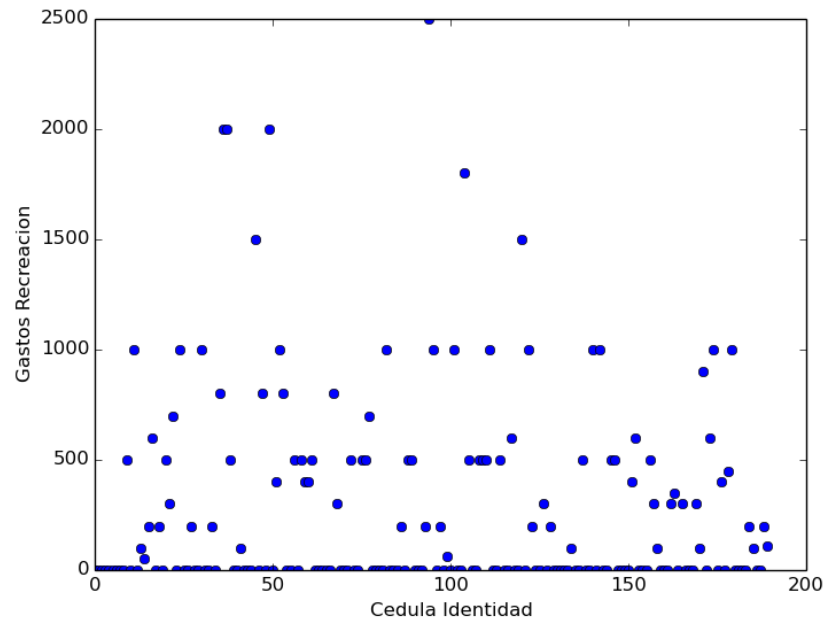


Figure 40: gRecreacion.

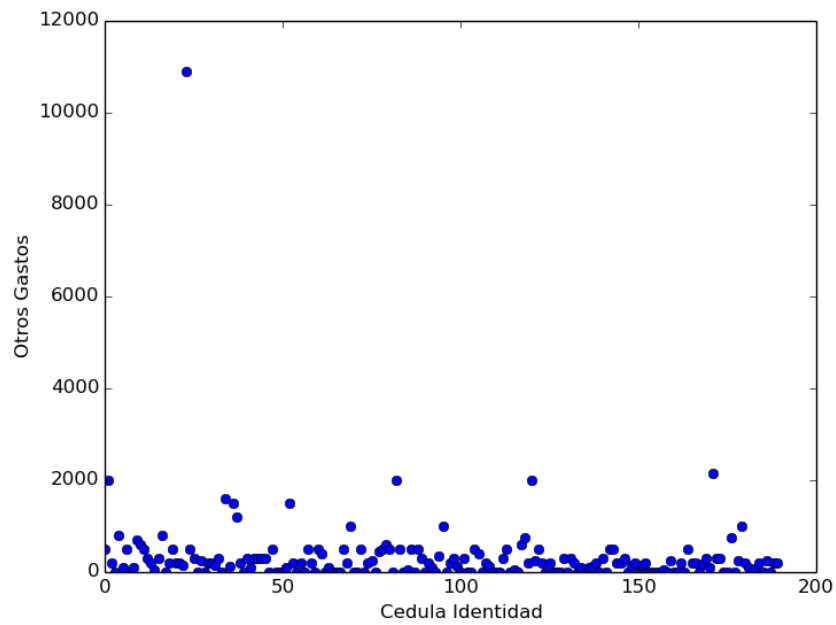


Figure 41: gOtros.

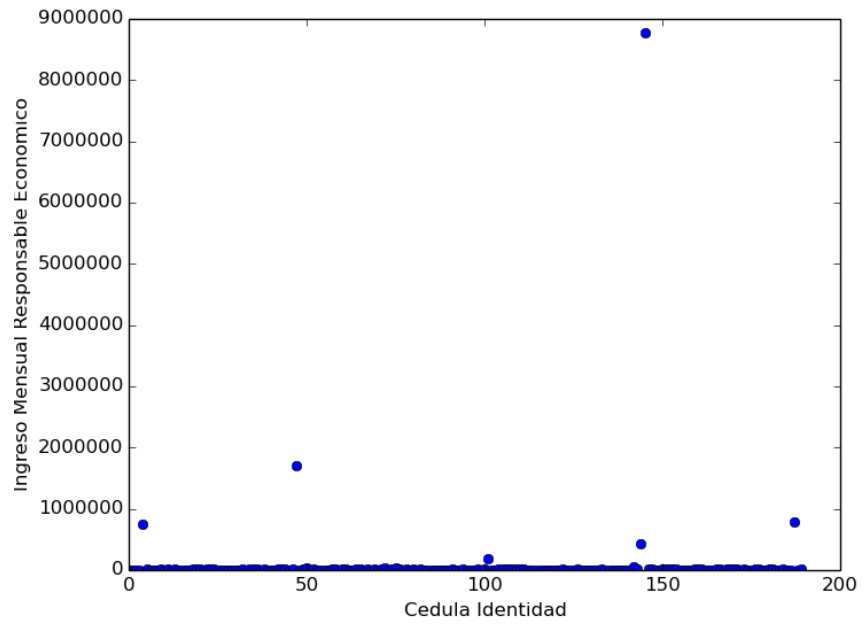


Figure 42: irMensual.

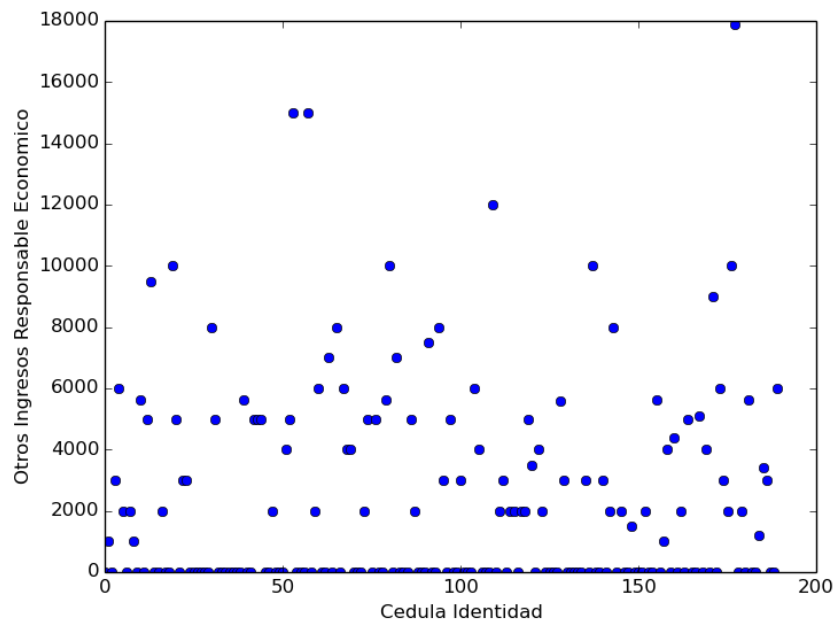


Figure 43: irOtros.

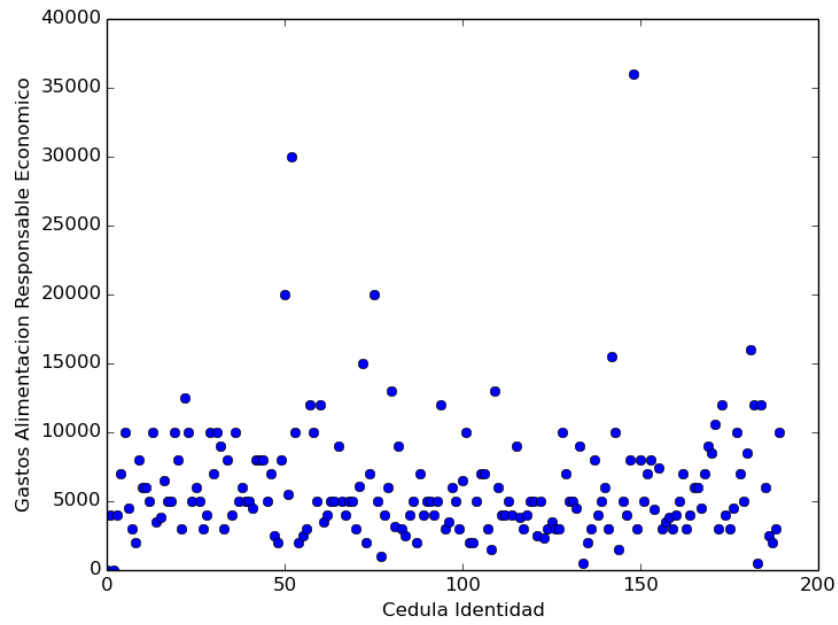


Figure 44: grAlimentacion.

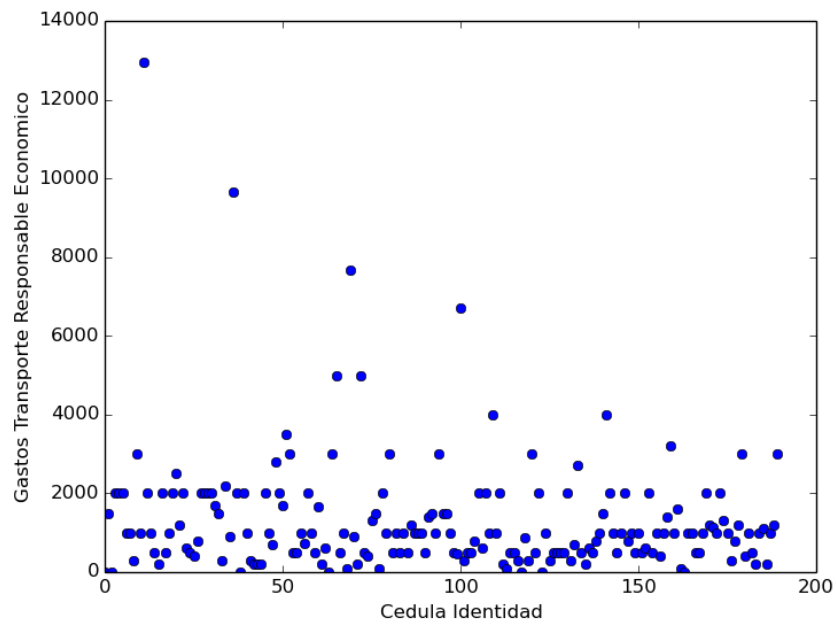


Figure 45: grTransporte.

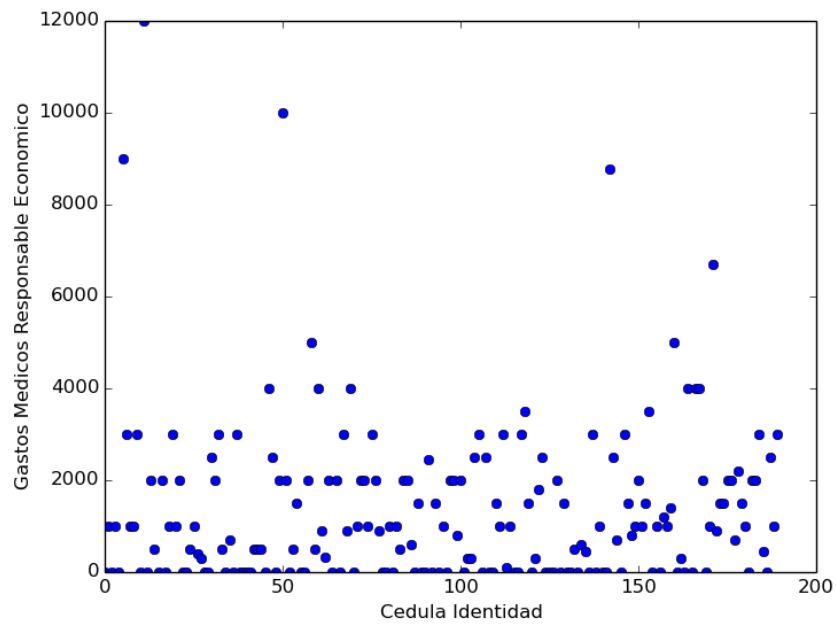


Figure 46: grMedicos.

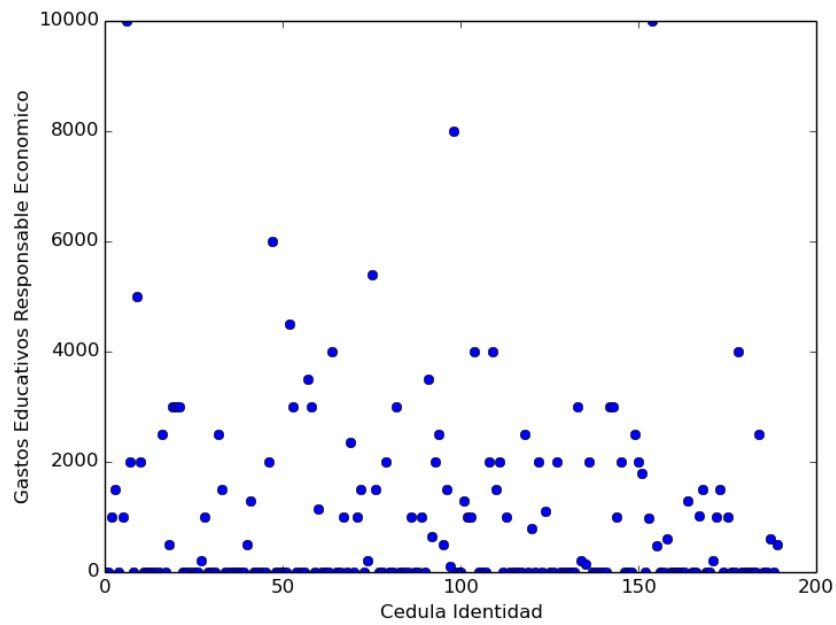


Figure 47: grEducativos.

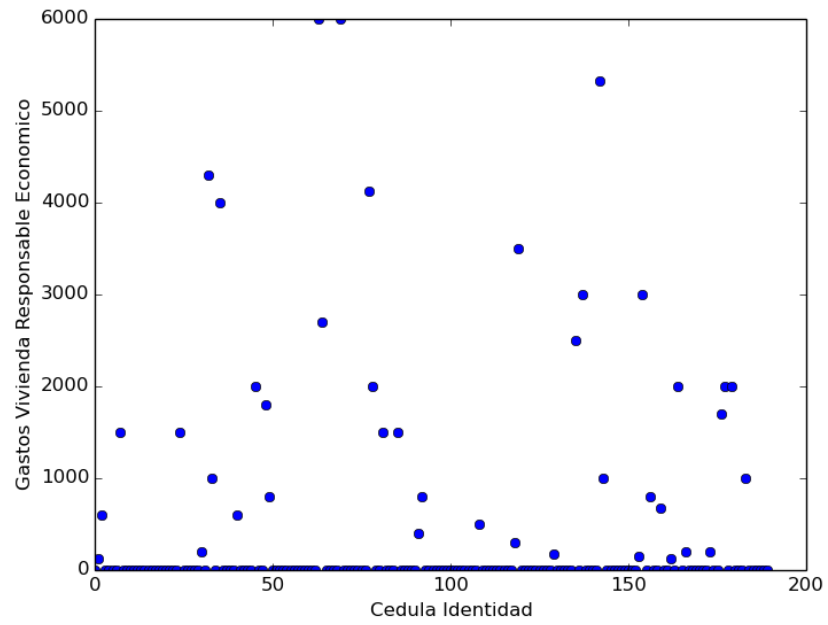


Figure 48: grVivienda.

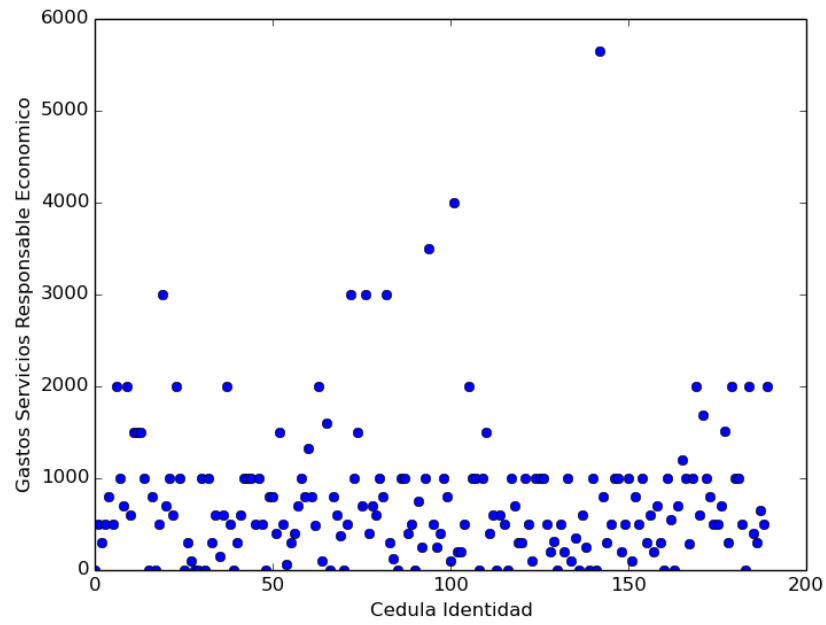


Figure 49: grServicios.

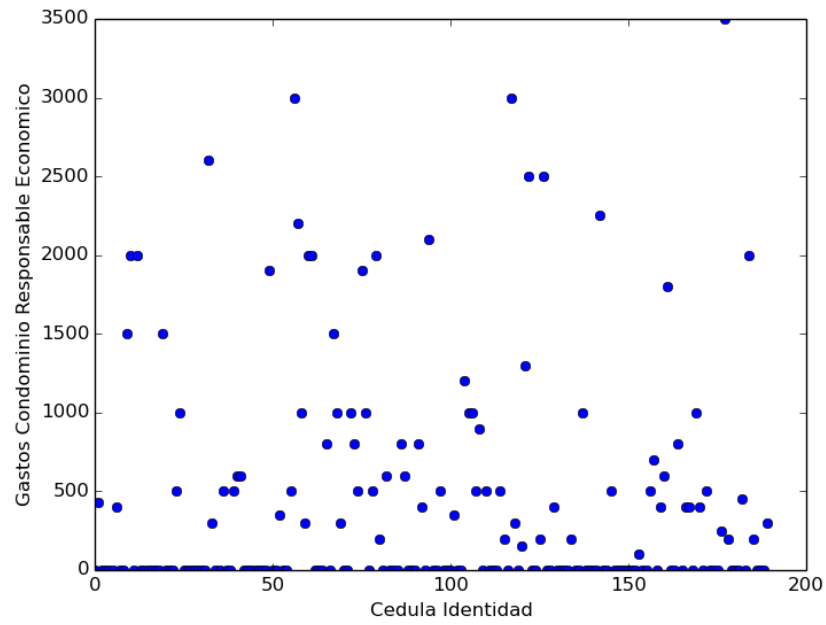


Figure 50: grCondominio.

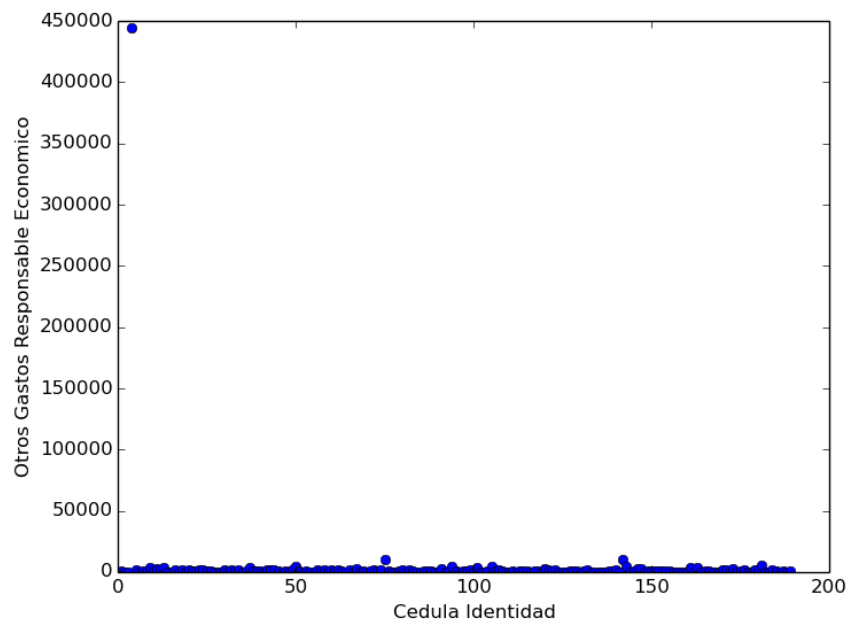


Figure 51: grOtros.

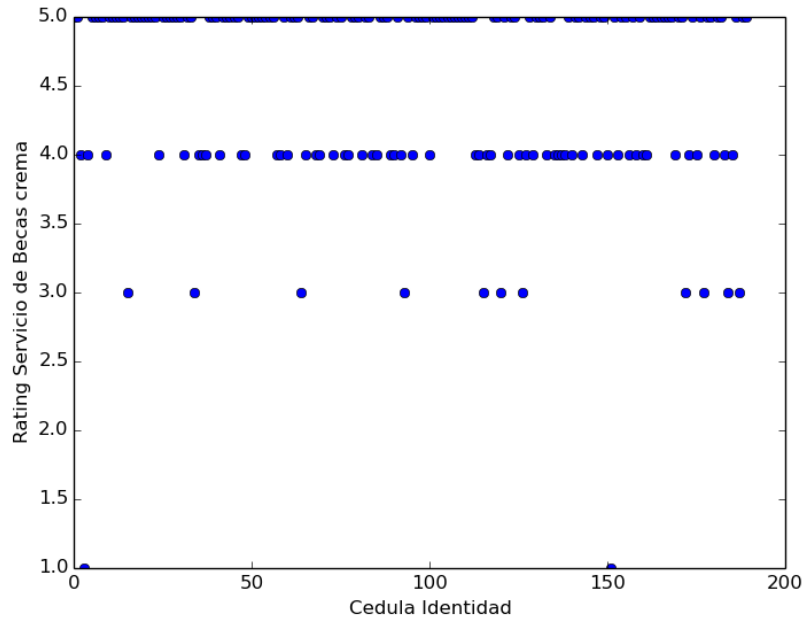


Figure 52: rating.

8 Valores Anómalos

Dada la sección anterior, se pueden estudiar las diferentes gráficas de las variables junto con el significado de la misma para determinar los valores anómalos.

9 Conclusión

Se logró el objetivo de crear una vista minable que permita que los datos estén más organizados, simples y unificados, de tal manera que el set de datos pueda servir para aplicar un algoritmo de minería de datos dependiendo del problema que se quiera atacar, **minable.csv** es el archivo resultante luego de realizar la limpieza de los datos.