

# Explicación

---

Alejandro Acosta

15/01/2015

## INTRODUCCIÓN

En este documento se analizarán las características del dataset de salida creado por el script adjunto con este documento. Por cada característica se dirá que decisión se tomó a la hora de crear la vista minable.

## DECISIONES TOMADAS DE LAS CARACTERÍSTICAS

1. Índice: Este índice se encontraba no solo desordenado sino además incompleto y dado que no proporcionaba ninguna información útil se decidió en no utilizarla en la vista minable.
2. Cédula de identidad: Se decidió cambiar la posición de esta característica por el índice ya que la cédula es única por individuo.
3. Período académico a renovar: Se decidió, dado que las escuelas son Enfermería y Bioanálisis, ambas pertenecientes a la Facultad de Medicina y que esta lleva sus períodos académicos por año, que se dejara sólo el año en vez de el semestre. Aquellas filas que decían en que semestre estaba el estudiante mas no el año en el que se encontraba estudiando se eliminaron.
4. Fecha de nacimiento: Se decidió unificar el formato de la fecha a dd/mm/aaaa. Para lograr esto se decidió verificar cada fecha y si ésta tenía caracteres como '/', '-' o ' ' entonces se hacían las verificaciones pertinentes. en caso de que los numeros no estuviesen separados se incluía esta fila a las filas por eliminar.

5. Edad: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato
6. Estado civil: Se decidió transformar los posibles valores de esta característica de la siguiente manera:
  - Casado tomará el valor 1.
  - Soltero tomará el valor 2.
  - Viudo tomará el valor 3.
  - El caso aislado Unido se tomó como si fuese Casado, es decir toma el valor 1.
7. Sexo: Se decidió transformar los posibles valores de esta característica de la siguiente manera:
  - Femenino tomará el valor 0.
  - Masculino tomará el valor 1.
8. Escuela: Se decidió transformar los posibles valores de esta característica de la siguiente manera:
  - Enfermería tomará el valor 0.
  - Bioanálisis tomará el valor 1.
9. Año de ingreso a la UCV: Se decidió no cambiar esta característica pues es sólo un dato numérico simple.
10. Modalidad de ingreso a la UCV: Se decidió transformar los posibles valores de esta característica de la siguiente manera:
  - Prueba Interna y/o propedéutico tomará el valor 1.
  - Convenios Interinstitucionales (nacionales e internacionales) tomará el valor 2.
  - Convenios Internos (Deportistas, artistas, hijos empleados docente y obreros, Samuel Robinson) tomará el valor 3.
  - Asignado OPSU tomará el valor 4.
11. Semestre que cursa: Se decidió que para esta columna sólo se tomara el valor numérico que indica el semestre.
12. Ha cambiado usted de dirección: Se decidió transformar los posibles valores de esta característica de la siguiente manera:
  - No, tomará el valor 0.
  - Si, tomará el valor 1.
13. De ser afirmativo indique el motivo: Se decidió no utilizar esta característica debido a que las razones podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.

14. Número de materias inscritas en el semestre o año anterior: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato. Este total es verificado y si la suma de las siguientes 3 columnas es mayor que esta la fila será eliminada.
15. Número de materias aprobadas en el semestre o año anterior: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato.
16. Número de materias retiradas en el semestre o año anterior: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato.
17. Número de materias reprobadas en el semestre o año anterior: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato.
18. Promedio ponderado aprobado: Se decidió tomar el valor del dato sin necesidad de agregarle o quitarle nada. Se asume que los datos mayores a 20 tienen los dos primeros números enteros y el resto son decimales y el punto fue omitido, por lo tanto estos números corregidos.
19. Eficiencia: Se decidió tomar el valor del dato sin necesidad de agregarle o quitarle nada. Si el valor es mayor a 1 se asume que el valor es el lado decimal y se le agrega un 0 a la izquierda y el punto decimal, exceptuando los casos en que el primer número es 1 y el resto 0.
20. Si reprobó una o más materias indique el motivo: Se decidió no utilizar esta característica debido a que las razones podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.
21. Número de materias inscritas en el semestre en curso: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato.
22. Si está realizando tesis, trabajo de grado, pasantías de grado: Se decidió transformar los posibles valores de esta característica de la siguiente manera:  
No, tomará el valor 0.  
Si, tomará el valor 1.
23. Tesis, trabajo de grado o pasantías de grado: Se decidió transformar los posibles valores de esta característica de la siguiente manera:  
Primera vez, tomará el valor 1.  
Segunda vez, tomará el valor 2.  
más de dos, tomará el valor 3.  
El valor nulo tomará el valor nulo o vacío.

24. Procedencia: Se decidió transformar los posibles valores de esta característica de la siguiente manera:  
Si es de algún municipio, tomará el valor 0.  
En caso de que viva en otro estado, tomará el valor 1
25. Lugar donde reside mientras estudia en la universidad: Se decidió transformar los posibles valores de esta característica de la siguiente manera:  
Si es de algún municipio, tomará el valor 0.  
En caso de que viva en otro estado, tomará el valor 1
26. Personas con las que vive mientras estudia en la universidad: Se decidió no utilizar esta característica debido a que el número de respuestas podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.
27. Tipo de vivienda donde reside mientras estudia en la universidad: Se decidió no utilizar esta característica debido a que el número de respuestas podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.
28. En caso de vivir en habitación alquilada o residencia estudiantil indique el monto mensual: Se decidió no utilizar esta característica debido a que este monto se ve en los egresos del estudiante y se le considera redundante.
29. Dirección donde se encuentra ubicada la residencia o habitación alquilada: Se decidió no utilizar esta característica debido a que el número de respuestas podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.
30. Contrajo matrimonio: Se decidió transformar los posibles valores de esta característica de la siguiente manera:  
No, tomará el valor 0.  
Si, tomará el valor 1.
31. Ha solicitado algún otro beneficio a la universidad u otra institución: Se decidió transformar los posibles valores de esta característica de la siguiente manera:  
No, tomará el valor 0.  
Si, tomará el valor 1.
32. En caso afirmativo señale año de la solicitud, institución y motivo: Se decidió no utilizar esta característica debido a que el número de respuestas podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.
33. Se encuentra usted realizando alguna actividad que le genere ingresos: Se decidió transformar los posibles valores de esta característica de la siguiente manera:  
No, tomará el valor 0.  
Si, tomará el valor 1.

34. En caso afirmativo indique tipo de actividad y su frecuencia: Se decidió no utilizar esta característica debido a que el número de respuestas podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.
35. Monto mensual de la beca: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato. Si el valor supera los 2000 este se tomará como erróneo y se eliminará la fila. De igual manera si es menor a 1000.
36. Aporte mensual que le brinda su responsable económico: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
37. Aporte mensual que recibe de familiares o amigos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
38. Ingreso mensual que recibe por actividades a destajo o por horas: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
39. Ingreso mensual total: Para evitar errores y eliminar filas cuyo total sea distinto de la suma de las 4 características anteriores se decidió calcular dicho total.
40. Alimentación: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
41. Transporte público: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
42. Gastos médicos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
43. Gastos odontológicos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
44. Gastos personales: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
45. Residencia o habitación alquilada: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.

46. Materiales de estudio: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
47. Recreación: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
48. Otros gastos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA que se transforma a 0.
49. Total egresos: Para evitar errores y eliminar filas cuyo total sea distinto de la suma de las 9 características anteriores se decidió calcular dicho total.
50. Indique quién es su responsable económico: Se decidió no utilizar esta característica debido a que el número de respuestas podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.
51. Carga familiar: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato.
52. Ingreso mensual de su responsable económico: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
53. Otros ingresos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
54. Total de ingresos: Para evitar errores y eliminar filas cuyo total sea distinto de la suma de las 9 características anteriores se decidió calcular dicho total.
55. Vivienda: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
56. Alimentación: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
57. Transporte: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
58. Gastos médicos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.

59. Gastos odontológicos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
60. Gastos educativos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
61. Servicios públicos de agua, luz, teléfono y gas: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
62. Condominio: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
63. Otros gastos: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato incluyendo valor NA o el valor nulo " que se transforma a 0.
64. Total de egresos: Para evitar errores y eliminar filas cuyo total sea distinto de la suma de las 9 características anteriores se decidió calcular dicho total.
65. Deseamos conocer la opinión de nuestros usuarios para mejorar la calidad de los servicios ofrecidos por el Dpto. de Trabajo Social OBE: Se decidió tomar el dato justo como se encuentra. Aunque cabe decir que se verificó tomar sólo el valor numérico del dato.
66. Sugerencias y recomendaciones para mejorar nuestra atención: Se decidió no utilizar esta característica debido a que las razones podrían ser infinitas y sería muy difícil agruparlas o hacer búsquedas por éstas.