



ugr

Universidad
de Granada

TRABAJO FIN DE GRADO
INGENIERÍA EN ...

Análisis de procesos

Subtitulo del Proyecto

Autor

María Isabel Ruiz Martínez (alumno)

Directores

Luis Castillo Vidal (tutor1)

Nombre Apellido1 Apellido2 (tutor2)



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, mes de 201



Análisis de procesos

Subtítulo del proyecto.

Autor

María Isabel Ruiz Martínez (alumno)

Directores

Luis Castillo Vidal (tutor1)

Nombre Apellido1 Apellido2 (tutor2)

Análisis de procesos: Subtítulo del proyecto

Nombre Apellido1 Apellido2 (alumno)

Palabras clave: palabra_clave1, palabra_clave2, palabra_clave3,

Resumen

Poner aquí el resumen.

Project Title: Project Subtitle

First name, Family name (student)

Keywords: Keyword1, Keyword2, Keyword3,

Abstract

Write here the abstract in English.

Yo, **María Isabel Ruiz Martínez**, alumno de la titulación **TITULACIÓN de la Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 75576979Z, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: María Isabel Ruiz Martínez

Granada a X de mes de 201 .

D. **Luis Castillo Vidal (tutor1)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

D. **Nombre Apellido1 Apellido2 (tutor2)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado ***Análisis de procesos, Subtítulo del proyecto***, ha sido realizado bajo su supervisión por **María Isabel Ruiz Martínez (alumno)**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a X de mes de 201 .

Los directores:

Luis Castillo Vidal (tutor1) **Nombre Apellido1 Apellido2 (tutor2)**

Agradecimientos

Poner aquí agradecimientos...

Índice general

1. Introducción, motivaciones, objetivos y estructura	1
1.1. Introducción	1
1.2. Motivaciones	2
1.3. Objetivos	2
2. Planificación	3
2.1. Herramientas	3
3. Análisis	5
3.1. Planteamiento del problema	5
4. Implementación	7
4.1. Análisis ANOVA y primeras conclusiones	7
4.1.1. Descripción del dataset	7
4.1.2. Introducción	10
4.1.3. Two way ANOVA	11
4.1.4. Segmentación de los datos	13
4.2. Extracción de procesos ocultos en los datasets	15
4.3. Análisis de los procesos por distintas categorías	15
4.4. Análisis de los motivos en los cambios de comportamiento	15
4.5. Análisis de si hay cambios inducidos por los éxitos de los demás grupos	15

Índice de figuras

4.1. Diagramas de caja y bigotes de las variables <i>map</i> y <i>action</i> . .	8
4.2. Distribuciones para diferentes niveles de los factores	8
4.3. Distribuciones de las variables <i>map</i> y <i>action</i> dependiendo del valor de <i>year</i>	9
4.4. Interacción entre las variables <i>map</i> y <i>action</i>	12
4.5. Gráficas diagnósticas del modelo ANOVA	12

Índice de cuadros

4.1. Muestra del primer dataset	9
4.2. Resumen del primer dataset que tenemos	9
4.3. Muestra del segundo dataset	10
4.4. Resumen del segundo dataset que tenemos	10
4.5. Resultados del test two way ANOVA	11
4.6. Coeficientes del modelo	11
4.7. Dataset 2 tras ser preparado para hacer join	13
4.8. Dataset 2 tras agrupar por grupo y año	13
4.9. Unión de ambos datasets	14
4.10. Muestra del dataset de los grupos suspensos . Consta de 2397 registros con 7 campos cada uno.	14
4.11. Muestra del dataset de los grupos aprobados . Consta de 7053 registros con 7 campos cada uno.	14
4.12. Muestra del dataset de los grupos con notable . Consta de 18840 registros con 7 campos cada uno.	15
4.13. Muestra del dataset de los grupos con sobresaliente . Consta de 17891 registros con 7 campos cada uno.	15
4.14. Muestra del dataset de los grupos con matrícula de honor . Consta de 1646 registros con 7 campos cada uno.	16

Capítulo 1

Introducción, motivaciones, objetivos y estructura

1.1. Introducción

La necesidad de comprender el proceso de aprendizaje y de personalizar la enseñanza para realizar una mejor adaptación a las necesidades del individuo ha motivado la *Analítica de Aprendizaje* o *Learning Analytics*, disciplina que consiste en la recogida de datos de un entorno de aprendizaje y el análisis de los mismos cuyo objetivo es asistir en el proceso de aprendizaje del alumnado.

Además, el uso de laboratorios virtuales y remotos en la enseñanza está en auge. Entre muchas de sus ventajas tenemos una mayor privacidad para el alumnado, accesos planificados a los mismos o soporte para reportar la actividad de los alumnos y la calificación de los mismos.

En este trabajo fin de grado se usarán datos de cinco cursos académicos obtenidos en el laboratorio virtual para sistemas multiagente de la asignatura del cuarto curso académico Desarrollo Basado en Agentes del grado de Ingeniería Informática de la Universidad de Granada (España).

El laboratorio virtual diseñado para la asignatura recoge el trabajo diario de los alumnos almacenando las interacción entre los diferentes agentes y obteniendo así un extenso dataset que nos proporciona una base sólida para el uso de diversas analíticas de aprendizaje.

Así pues, se empleará un enfoque “*data-driven*” o *impulsado por datos*, tomando decisiones estratégicas basándose en el análisis de los datos y en la interpretación de los mismos.

1.2. Motivaciones

El buen maestro hace que el mal estudiante se convierta en bueno y el buen estudiante en superior.

Maruja Torres

Como decía Maruja Torres, un buen profesor puede influenciar positivamente en el desarrollo académico de sus alumnos. Así pues, uno de las principales motivaciones de este trabajo es, precisamente, ayudar a identificar buenas y malas prácticas de los alumnos precozmente para que el profesorado pueda asistirles mejor durante su proceso de aprendizaje y mejorar su rendimiento académico.

Así pues, la motivación última de los estudios expuesto en este documento es ayudar a aprender al alumnado.

Asimismo, creo que es de gran interés práctico conocer cuáles son las mejores estrategias a la hora de programar un sistema multiagente que desconoce el entorno que le rodea y que intenta conseguir un objetivo preestablecido.

1.3. Objetivos

Tras un primer análisis de los datos, se pretende tratar de relevar posibles estrategias escondidas en los mismos utilizando técnicas de minería de procesos, donde se considerará que una estrategia es el proceso seguido por los alumnos hasta que llegan a su objetivo.

Es decir, se pretende identificar patrones de comportamiento de los drones en los diferentes mundos virtuales con la finalidad de ayudar al profesorado de la asignatura a guiar al alumnado.

Capítulo 2

Planificación

2.1. Herramientas

En la realización de un proyecto investigación, la elección de las herramientas para su desarrollo es clave. A continuación se expone una lista con las que se han utilizado este trabajo fin de grado:

- En el desarrollo del software se han empleado los lenguajes de programación `R` y `C++`.
- Se ha empleado la herramienta de minería de procesos `DISCO`.
- Se ha utilizado `GitHub` para alojar todo el contenido del proyecto y gestionar las distintas versiones del mismo.
- Para la redacción de la memoria se ha empleado el editor gratuito `Texmaker` utilizando el sistema de composición de textos `LATEX`.
- Para el seguimiento del proyecto se han utilizado las `Hojas de cálculo de Google`.

Capítulo 3

Análisis

3.1. Planteamiento del problema

El servidor contiene varios mundos virtuales. Cada mundo virtual es una matriz cuadrada que representa espacios abiertos (en color blanco), obstáculos (en negro) y objetivos (en rojo) tal y como se muestra en la Figura ???. Los agentes de los alumnos deben entrar en uno de esos mundos virtuales, percibir su vecindario, navegar a través de los espacios abiertos (empleando alguna clase de heurística exploratoria), evitar obstáculos y tratar de llegar al objetivo.

La percepción del agente de su entorno es crítica para resolver estos mundos. En este laboratorio virtual los alumnos pueden configurar cuál de los siguientes sensores estarán enchufados en sus agentes (cualquier combinación de ellos):

- Un **GPS** que indica al agente sus coordenadas (x, y) en el mundo virtual.
- Un **sensor de batería**. Cada agente está alimentado con una batería cuya capacidad es limitada y cuya carga decrece conforme el agente realiza algún movimiento. La batería nunca debe ser vaciada por completo.
- Un **sensor radar** que informa al agente acerca de los tipos de celdas que lo rodean con una percepción local de 5x5 (observar Figura ??).
- Un **sensor escáner** que actúa como *detector del objetivo* e indica al agente la distancia al objetivo medida desde cada una de las celdas de su entorno 5x5 (observar Figura ??).

Basados en su percepción del mundo virtual, cada agente decidirá ejecutar alguna de las siguientes acciones en su entorno implementando cualquier heurística o proceso de búsqueda.

- LOGIN. Entrar en cualquiera de los mundos virtuales.
- MOVE. Mover al agente a una de las 8 celdas adyacentes y gastar una cierta cantidad de batería. Si la celda destino es un obstáculo o el agente se queda sin batería, el agente se rompe y se sale del mundo virtual.
- REFUEL. El agente recarga completamente su batería. A los agentes se les permite recargar su batería tantas veces como deseen.

Capítulo 4

Implementación

4.1. Análisis ANOVA y primeras conclusiones

4.1.1. Descripción del dataset

Los datasets que se van a usar han sido recopilados por Luis Castillo Vidal y corresponden a la actividad de sus alumnos en la asignatura [Desarrollo Basado en Agentes](#).

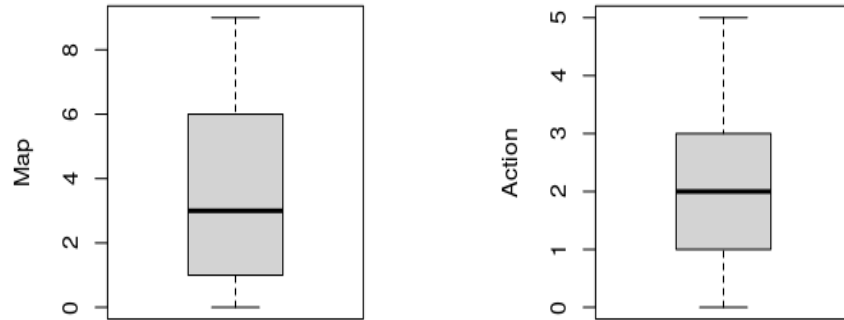
El primer dataset, tras haber sido filtrados los registros erróneos, consta de 47828 filas correspondientes a los diferentes acciones de unos drones en una serie de mundos virtuales. En cada registro se detallan los siguientes atributos:

- *year*: identifica el curso académico en el que se realizó dicha acción.
- *group*: grupo de prácticas que ha progradado al dron que acomete la acción.
- *date*: fecha en la que se lleva a cabo la acción.
- *map*: mundo virtual en el que se ha realizado la acción.
- *action*: indica el tipo de acción realizada.

En la Tabla 4.1 se presentan los primeros seis registros del dataset. Además, en la Tabla 4.2 puede apreciarse un resumen de los datos que tenemos.

La Figuras 4.1a y 4.1b muestran, respectivamente, los gráficos de caja y bigotes de las variables *map* y *action*.

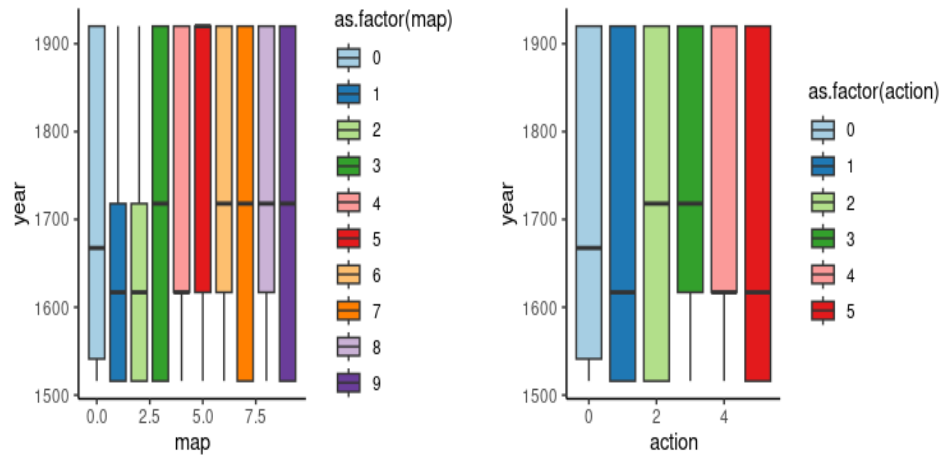
Además, podemos ver la distribución de los diferentes niveles del factor *map* y de los diferentes niveles del factor *action* en las Figuras 4.2a y 4.2b.



(a) Gráfico de caja y bigotes de la variable *map*

(b) Gráfico de caja y bigotes de la variable *action*

Figura 4.1: Diagramas de caja y bigotes de las variables *map* y *action*



(a) Distribución de los diferentes niveles del factor *map*

(b) Distribución de los diferentes niveles del factor *action*

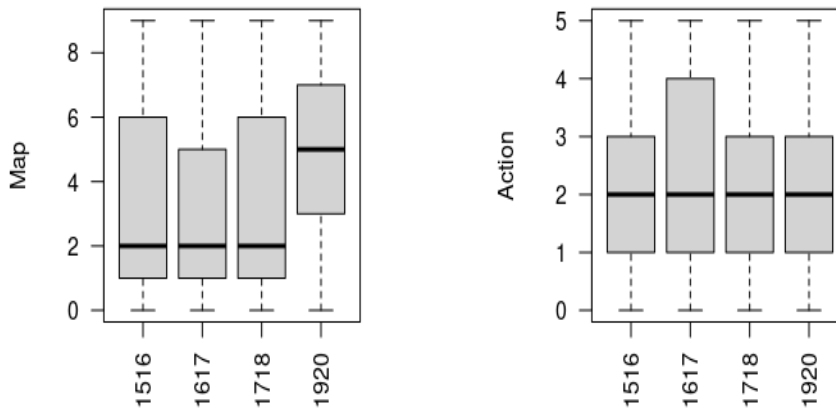
Figura 4.2: Distribuciones para diferentes niveles de los factores

	year	group	date	map	action
1	1516	Achernar	17/10/2015 19:41:45	0	0
2	1516	Bellatrix	17/10/2015 19:41:45	0	0
3	1516	Cerastes	17/10/2015 19:41:45	0	0
4	1516	Denebola	17/10/2015 19:41:45	0	0
5	1516	Elnath	17/10/2015 19:41:45	0	0
6	1516	Furud	17/10/2015 19:41:45	0	0

Cuadro 4.1: Muestra del primer dataset

year	group	date	map	action
Min. :1516	Length:47828	Length:47828	Min. :0.000	Min. :0.000
1st Qu.:1516	Class :character	Class :character	1st Qu.:1.000	1st Qu.:1.000
Median :1617	Mode :character	Mode :character	Median :3.000	Median :2.000
Mean :1700			Mean :3.834	Mean :2.325
3rd Qu.:1920			3rd Qu.:6.000	3rd Qu.:3.000
Max. :1920			Max. :9.000	Max. :5.000

Cuadro 4.2: Resumen del primer dataset que tenemos



(a) Distribución de la variable *map* en función de *year* (b) Distribución de la variable *action* en función de *year*

Figura 4.3: Distribuciones de las variables *map* y *action* dependiendo del valor de *year*

Por último, también podemos ver la distribución de las variables *map* y *action* en función del año (variable *year*) en las Figuras 4.3a y 4.3b.

El segundo dataset, tras haber sido eliminados algunas columnas que no eran interesantes para nuestro estudio, consta de 118 filas correspondientes

a las diferentes calificaciones de los equipos en las dos prácticas realizadas en la asignatura [Desarrollo Basado en Agentes](#). En cada registro se detallan los siguientes atributos:

- *Group*: grupo de prácticas que ha progradado al dron.
- *Team*: cadena de texto que identifica el curso académico en el que se realizó dicha acción, la práctica realizada y el grupo de prácticas conjuntamente.
- *Size*: tamaño del grupo de prácticas.
- *Year*: identifica el curso académico en el que se realizó dicha acción.
- *Grade*: calificación obtenida por el grupo de prácticas.

En la Tabla 4.3 se presentan los primeros seis registros del dataset. Además, en la Tabla 4.4 puede apreciarse un resumen de los datos que tenemos.

	Group	Team		Size	Year	Grade
1	G1	DBA 1819 P3 GL		4	1819	10,00
2	G2	DBA 1920 P3 GJ		4	1920	4,01
3	G3	DBA 1819 P2 GH		4	1819	7,96
4	G4	DBA 1920 P2 GE		4	1920	8,95
5	G5	DBA 1920 P3 GK		4	1920	4,51
6	G6	DBA 1415 P3 G6		6	1415	7,20

Cuadro 4.3: Muestra del segundo dataset

Group	Team	Size	Year	Grade
Length:118	Length:118	Min. :3.000	Min. :1314	Length:118
Class :character	Class :character	1st Qu.:4.000	1st Qu.:1516	Class :character
Mode :character	Mode :character	Median :5.000	Median :1718	Mode :character
		Mean :4.831	Mean :1662	
		3rd Qu.:6.000	3rd Qu.:1819	
		Max. :6.000	Max. :1920	

Cuadro 4.4: Resumen del segundo dataset que tenemos

4.1.2. Introducción

En este estudio inicial se desarrollará un modelo estadístico para determinar el efecto de los parámetros *map* y *action* (dos variables explicativas) en la variable respuesta *year*.

La relevancia de cada una de las variables en el modelo se determinará por el test *two way ANOVA* con un 5 % de nivel de significancia y se empleará la técnica de los *mínimos cuadrados* para estimar los coeficientes del modelo considerado.

4.1.3. Two way ANOVA

Un resumen de los resultados obtenidos al realizar el test two way ANOVA se muestra en la Tabla 4.5. Puede observarse que la variable *map* es significativa al nivel 0, que la variable *action* es significativa al nivel 0.01 y que la variable *map:action* (el término de interacción) no es significativa. Así pues, puede concluirse que el dataset es homogéneo, es decir, las combinaciones *map:action* son estadísticamente iguales en todos los años considerados.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
map	1	62918413.58	62918413.58	2689.15	0.0000
action	1	101244.69	101244.69	4.33	0.0375
map:action	1	8797.03	8797.03	0.38	0.5398
Residuals	47824	1118946173.13	23397.17		

Cuadro 4.5: Resultados del test two way ANOVA

La notación escalar del modelo ajustado al aplicar el test tiene la siguiente estructura:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \cdot x_2 + \epsilon \quad (4.1)$$

donde β_0 es el intercepto, β_1 y β_2 son los coeficientes de los efectos principales, β_3 es el coeficiente del término de interacción, x_1 y x_2 son los parámetros sometidos a investigación (en este caso, x_1 representa el parámetro mapa y x_2 representa la acción), y representa el año y ϵ es el *término error*.

La Tabla 4.6 muestra los valores de los coeficientes de la fórmula que se han obtenido tras ajustar el modelo de regresión a los datos.

	x
(Intercept)	1655.03
map	12.31
action	-1.51
map:action	0.12

Cuadro 4.6: Coeficientes del modelo

La Figura 4.4 muestra la interacción entre los parámetros mapa y acción.

Así pues, puede observarse que todas las líneas de la gráfica siguen más o menos el mismo patrón, lo que evidencia que no hay una gran interacción entre ambos.

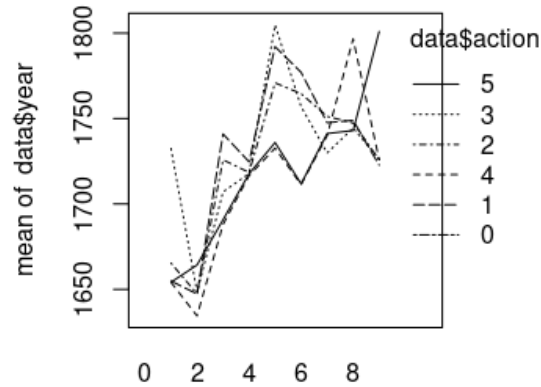


Figura 4.4: Interacción entre las variables *map* y *action*

La Figura 4.5 muestra que no se violan las suposiciones que hemos realizado sobre el modelo. La media y la varianza de los residuos no parece que varíe respecto de los valores ajustados. Como consecuencia, concluiré que podemos suponer la homocedasticidad. Además, si nos fijamos en el *Normal Q-Q plot*, puede observarse que los residuos son gaussianos.

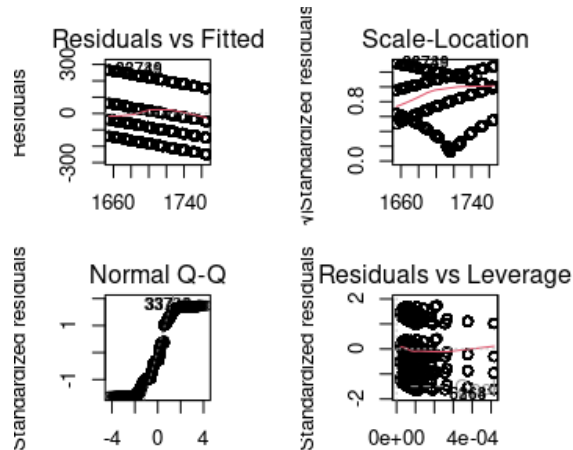


Figura 4.5: Gráficas diagnósticas del modelo ANOVA

4.1.4. Segmentación de los datos

Se ha realizado una segmentación de los registros del primer dataset en función de la calificación obtenida por los grupos que han realizado las acciones en los mapas.

Para ello, se ha sustituido la columna *Team* del segundo dataset por el nombre del grupo en cada caso y se ha eliminado la columna *Group*. Después de realizar este proceso el dataset tiene la forma que se muestra en la Tabla 4.7

	group	size	year	grade
1	Lesath	4	1819	10,00
2	Jabbah	4	1920	4,01
3	Haldus	4	1819	7,96
4	Elnath	4	1920	8,95
5	Keid	4	1920	4,51
6	Furud	6	1415	7,20

Cuadro 4.7: Dataset 2 tras ser preparado para hacer join

A continuación, se transformarán los valores de la columna *grade* a tipo `double` y se agruparán los datos de este dataset por grupo y año, calculando la calificación media de las prácticas para cada par grupo-año. El resultado de realizar esta operación puede apreciarse en la Tabla 4.8.

	group	year	size	mean_grade
1	Achernar	1314	6	9.40
2	Achernar	1415	5	9.00
3	Achernar	1516	5	8.20
4	Achernar	1617	5	10.00
5	Achernar	1718	5	8.33
6	Bellatrix	1516	5	8.20

Cuadro 4.8: Dataset 2 tras agrupar por grupo y año

Ahora, se juntan ambos datasets usando la función `inner_join` por grupo y año. El resultado de esta operación puede observarse en la Tabla 4.9.

Por último, se separarán los datos de esta última tabla en función de las calificaciones obtenidas por los diferentes grupos:

- **Suspense:** calificación mayor o igual que 0 y menor que 5.
- **Aprobado:** calificación mayor o igual que 5 y menor que 7.
- **Notable:** calificación mayor o igual que 7 y menor que 9.

	group	year	size	mean_grade	date	map	action
1	Achernar	1516	5	8.2	17/10/2015 19:41:45	0	0
2	Achernar	1516	5	8.2	22/10/2015 17:29:21	1	1
3	Achernar	1516	5	8.2	22/10/2015 17:29:22	1	2
4	Achernar	1516	5	8.2	22/10/2015 17:29:39	1	3
5	Achernar	1516	5	8.2	22/10/2015 17:34:09	1	1
6	Achernar	1516	5	8.2	22/10/2015 17:34:10	1	2

Cuadro 4.9: Unión de ambos datasets

- **Sobresaliente:** calificación mayor o igual que 9 y menor que 10.
- **Matrícula de Honor:** calificación igual a 10.

Una muestra de los datasets generados tras la segmentación puede apreciarse en las tablas 4.10, 4.11, 4.12, 4.13 y 4.14.

	group	year	size	mean_grade	date	map	action
1	Girtab	1617	6	4.67	15/11/2016 7:31:31	1	1
2	Girtab	1617	6	4.67	15/11/2016 7:31:53	1	1
3	Girtab	1617	6	4.67	15/11/2016 12:50:03	1	1
4	Girtab	1617	6	4.67	15/11/2016 12:50:28	1	1
5	Girtab	1617	6	4.67	15/11/2016 13:27:27	1	1
6	Girtab	1617	6	4.67	15/11/2016 13:27:58	1	1

Cuadro 4.10: Muestra del dataset de los grupos **suspensos**. Consta de 2397 registros con 7 campos cada uno.

	group	year	size	mean_grade	date	map	action
1	Bellatrix	1920	4	6.835	05/11/2019 10:24:51	1	1
2	Bellatrix	1920	4	6.835	05/11/2019 10:24:51	1	2
3	Bellatrix	1920	4	6.835	05/11/2019 10:25:04	1	1
4	Bellatrix	1920	4	6.835	05/11/2019 10:25:26	1	2
5	Bellatrix	1920	4	6.835	05/11/2019 10:25:39	1	1
6	Bellatrix	1920	4	6.835	05/11/2019 10:25:45	1	2

Cuadro 4.11: Muestra del dataset de los grupos **aprobados**. Consta de 7053 registros con 7 campos cada uno.

	group	year	size	mean_grade	date	map	action
1	Achernar	1516	5	8.2	17/10/2015 19:41:45	0	0
2	Achernar	1516	5	8.2	22/10/2015 17:29:21	1	1
3	Achernar	1516	5	8.2	22/10/2015 17:29:22	1	2
4	Achernar	1516	5	8.2	22/10/2015 17:29:39	1	3
5	Achernar	1516	5	8.2	22/10/2015 17:34:09	1	1
6	Achernar	1516	5	8.2	22/10/2015 17:34:10	1	2

Cuadro 4.12: Muestra del dataset de los grupos con **notable**. Consta de 18840 registros con 7 campos cada uno.

	group	year	size	mean_grade	date	map	action
1	Bellatrix	1617	5	9.72	07/11/2016 11:13:31	1	1
2	Bellatrix	1617	5	9.72	07/11/2016 11:36:59	1	1
3	Bellatrix	1617	5	9.72	07/11/2016 11:41:07	1	1
4	Bellatrix	1617	5	9.72	07/11/2016 11:42:22	1	1
5	Bellatrix	1617	5	9.72	07/11/2016 11:45:37	1	1
6	Bellatrix	1617	5	9.72	07/11/2016 11:49:23	1	1

Cuadro 4.13: Muestra del dataset de los grupos con **sobresaliente**. Consta de 17891 registros con 7 campos cada uno.

4.2. Extracción de procesos ocultos en los datasets

4.3. Análisis de los procesos por distintas categorías

4.4. Análisis de los motivos en los cambios de comportamiento

4.5. Análisis de si hay cambios inducidos por los éxitos de los demás grupos

	group	year	size	mean_grade	date	map	action
1	Achernar	1617	5	10	09/11/2016 21:23:22	1	1
2	Achernar	1617	5	10	09/11/2016 21:25:11	1	1
3	Achernar	1617	5	10	10/11/2016 0:21:37	1	1
4	Achernar	1617	5	10	10/11/2016 0:22:50	1	1
5	Achernar	1617	5	10	10/11/2016 0:23:12	1	1
6	Achernar	1617	5	10	10/11/2016 0:25:53	1	1

Cuadro 4.14: Muestra del dataset de los grupos con **matrícula de honor**.
Consta de 1646 registros con 7 campos cada uno.

