

# Sistemas de Información - Practica 1



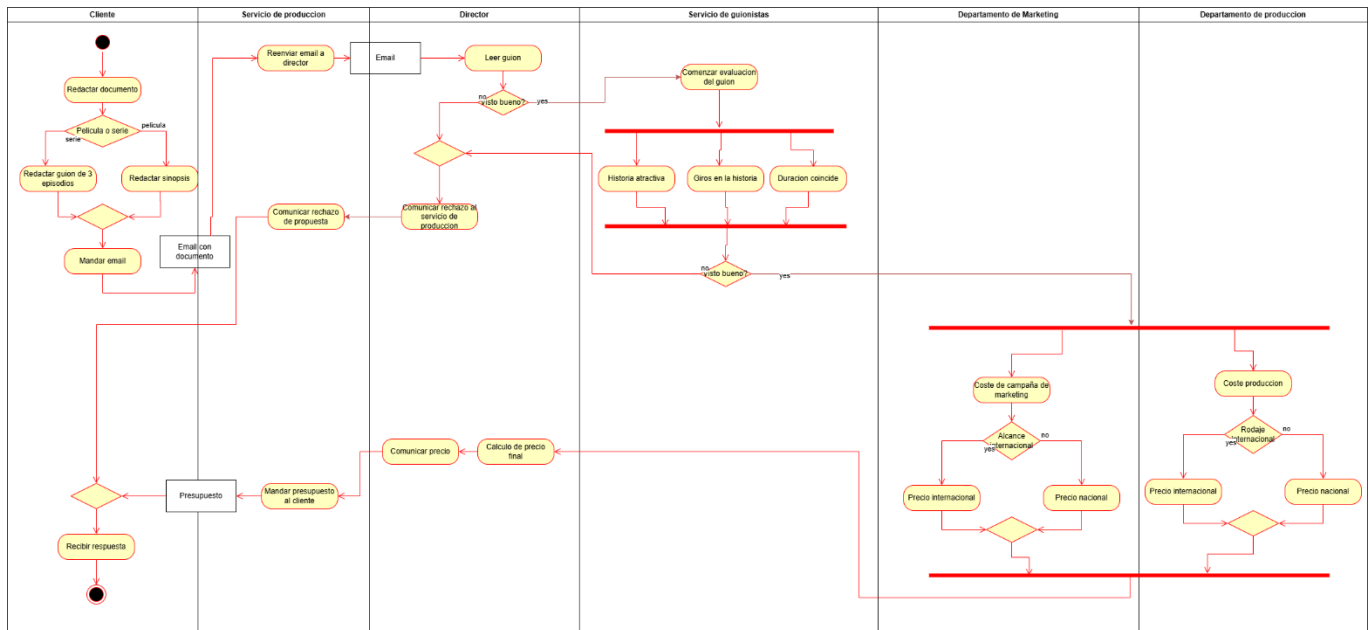
**URJC**



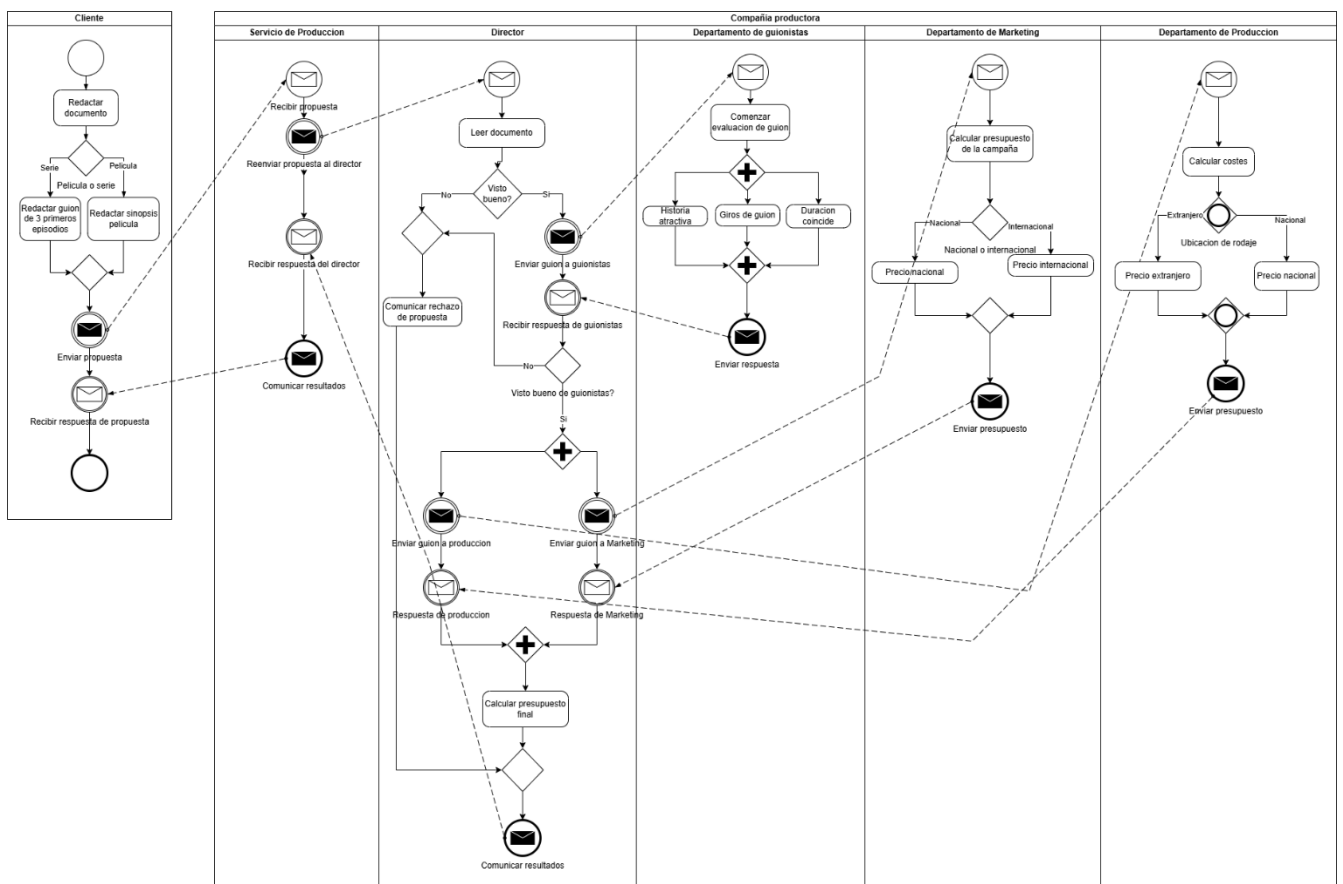
Escuela Técnica Superior  
de Ingeniería Informática

Fernando López Berrocal

## UML



## BPMN



## Ejercicio 2

En la primera parte de este ejercicio hay que pasar los datos contenidos en un archivo csv a una base de datos MySQL. Para ello se ha desarrollado un script de Python usando la librería Pandas para manejar los datos leyéndolos del csv e insertándolos en la base de datos creada.

Antes de insertar los datos, han tenido que limpiarse y formatearse. Primero de todo, había valores "missing" en los datos, filas con columnas con valores nulos. Estos valores se leen en Python como "numpy.nan" y se han pasado a "None", porque de esta manera la BBDD los reconoce como "NULL".

El formato de los datos también hubo que ajustarlo. La mayoría se detectaban como "Strings". De estos "Strings" algunos se han formateado a "VARCHAR" en MySQL, pero la columna "date\_added" se ha formateado a "DATE" y la columna "duration" se ha pasado a un "SMALLINT" significando número de temporadas en el caso de las series y minutos en las películas.

Por último, también se detectaron 3 filas con valores mal puestos, ya que el valor otorgado a la columna "rating" realmente pertenecía a la columna "duration" y se quedaría la columna "rating" como "missing". Estas filas tienen los siguientes id: "s5542", "s5795", "s5813". Se han corregido manualmente estas filas

En cuanto a las preguntas del ejercicio:

a) Número de muestras (valores distintos de missing).

El número de muestras es 101.377. Habiendo 8807 filas y 12 columnas cada una, debería haber 105.684 muestras. Por lo tanto, hay 4307 valores missing.

b) Media y desviación estándar de la columna duración.

No tienen sentido los cálculos porque algunos datos hacen referencia a minutos de duración en las películas y otros al número de temporadas en las series

c) Valor mínimo y valor máximo de las columnas duración y año

Valor mínimo y máximo de duración no tiene sentido por el motivo anterior. El valor mínimo y máximo de año es 1925 y 2021 respectivamente.

## Ejercicio 3

Este ejercicio se ha realizado usando queries a la BBDD, aunque también se podría haber usado “pandas.DataFrame” en Python.

### Ejercicio 3.a

Agrupacion por contenido

	type	Numero de Observaciones	Numero de Valores Missing	Media	Varianza	Minimo	Maximo	Mediana
0	Movie	6131	0	99.5650	800.165527	3	312	98.0
1	TV Show	2676	0	1.7649	2.504167	1	17	1.0

### Ejercicio 3.b

Agrupacion por temporadas (Series)

	type	Numero de Observaciones	Numero de Valores Missing	Media	Varianza	Minimo	Maximo	Mediana
0	<=2 Temporadas	2218	0	1.1916	0.154898	1	2	1.0
1	>2 Temporadas	458	0	4.5415	4.580157	3	17	4.0

### Ejercicio 3.c

Agrupacion por duracion (Peliculas)

	type	Numero de Observaciones	Numero de Valores Missing	Media	Varianza	Minimo	Maximo	Mediana
0	<=90 min	1993	0	71.0993	365.348886	3	90	76.0
1	>90 min	4138	0	113.2750	431.358878	91	312	108.0

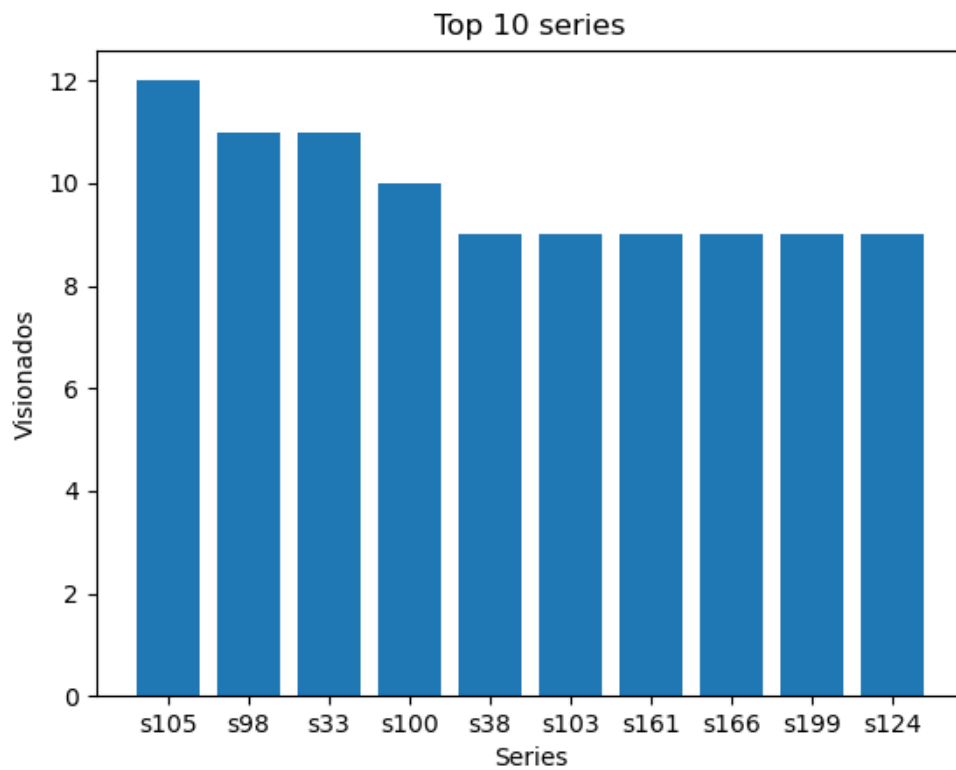
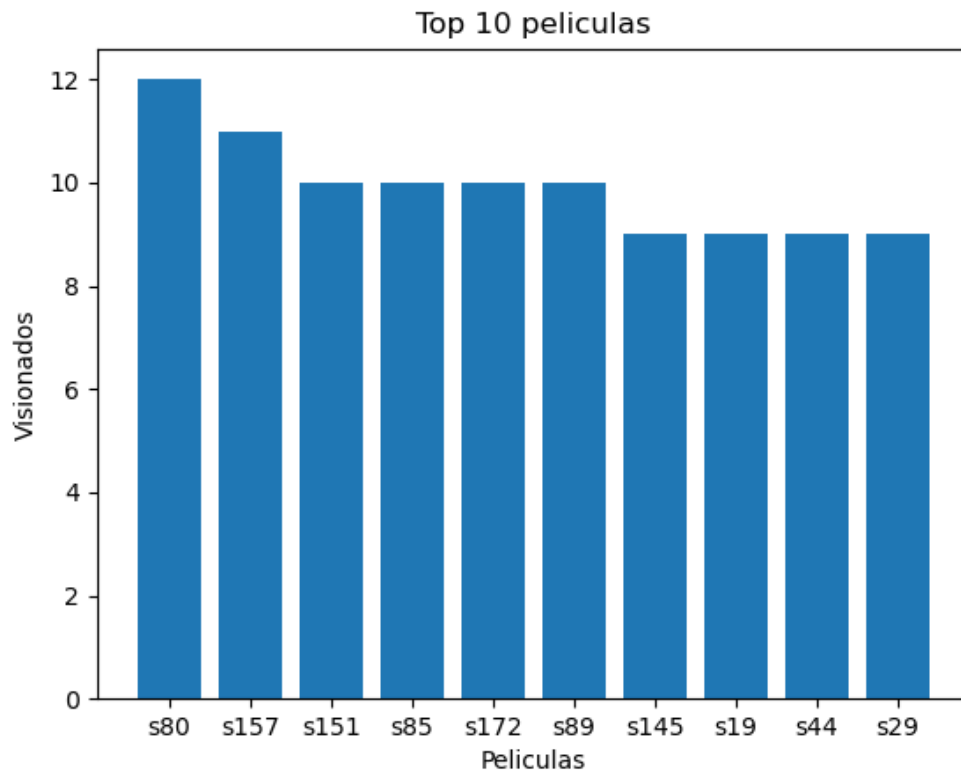
## Ejercicio 4

En este ejercicio primero se ha creado una tabla de usuarios con las columnas: “uid”, “name” y “last\_login”. Los datos se han generado aleatoriamente, entre 20-30 entradas con nombres aleatorios de una lista y última fecha de login aleatoria entre 01/01/2000 y 31/12/2025.

Después se ha creado la tabla “visionados” que relaciona las series/películas vistas por un usuario y su valoración. Para aleatorizar los visionados, cada usuario ha visto entre 1 y 200 series/películas entre las primeras 200. La valoración es aleatoria entre 1 y 10.

## Ejercicio 5

Este ejercicio también se ha resuelto usando queries de SQL. Para las 2 primeras preguntas se ha usado la librería “matplotlib” para generar los gráficos de barras.



#### Ejercicio 5.a

	show_id	title
0	s80	Tughlaq Durbar (Telugu)
1	s157	Letters to Juliet
2	s151	In Too Deep
3	s85	Omo Ghetto: the Saga
4	s172	Same Kind of Different as Me
5	s89	Blood Brothers: Malcolm X & Muhammad Ali
6	s145	House Party
7	s19	Intrusion
8	s44	Jaws 3
9	s29	Dark Skies

#### Ejercicio 5.b

	show_id	title
0	s105	Tayo the Little Bus
1	s98	Kid Cosmic
2	s33	Sex Education
3	s100	On the Verge
4	s38	Angry Birds
5	s103	Countdown: Inspiration4 Mission to Space
6	s161	Major Dad
7	s166	Oldsters
8	s199	King of Boys: The Return of the King
9	s124	Luv Kushh

#### Ejercicio 5.c

	Agrupacion	Media visionados
0	>90 min	6.8302
1	<=90 min	5.8571

#### Ejercicio 5.d

	Agrupacion	Media visionados
0	<=2 Temporadas	6.2679
1	>2 Temporadas	6.5294