

# ANÁLISIS TRIPADVISOR REVIEWS

TECNOLOGÍA DE INTELIGENCIA DE NEGOCIO 2015/2016

JUAN MANUEL BAREA MARTÍNEZ ENRIQUE GÓMEZ MARTÍNEZ

# ÍNDICE

- 1. Introducción
- 2. Fuentes de datos
- 3. Estructura Data Warehouse
- 4. Esquema general del proyecto
- 5. ETL
- 6. Informes
- 7. Cubos OLAP
- 8. Cuadros de mando
- 9. Problemas
- 10. Preguntas



# INTRODUCCIÓN

- El proyecto trata de obtener un estudio de las reviews que realizan los usuarios con la valoración de los hoteles y otros tipos de alojamiento que ofrece TripAdvisor.
- •Se tratará la información y se analizará para realizar Inteligencia de Negocio.



### FUENTES DE DATOS

• Ficheros JSON que se han obtenido de la web <a href="http://times.cs.uiuc.edu/~wang296/Data/">http://times.cs.uiuc.edu/~wang296/Data/</a>

 Habrá en cada archivo JSON las valoraciones respectivas a un hotel.



# FUENTES DE DATOS

Estructura de las reviews de hoteles:

```
"Ratings": {
    "Overall": "2.0"
},
"AuthorLocation": "london england",
"Title": ""Game On"",
"Author": "brenchley",
"ReviewID": "UR5822701",
"Content": "We were booked to stay here for two nights from the 31st of august,
    we arrived at about 7PM only to be told that thay had no parking.
    We were given directions to some public parking but they did not tell us
    that there was a match on in the stadium. we could not find anywhere to park
    so we drove to the best western executive inn by the space needle to see if",
"Date": "September 21, 2006"
},
```

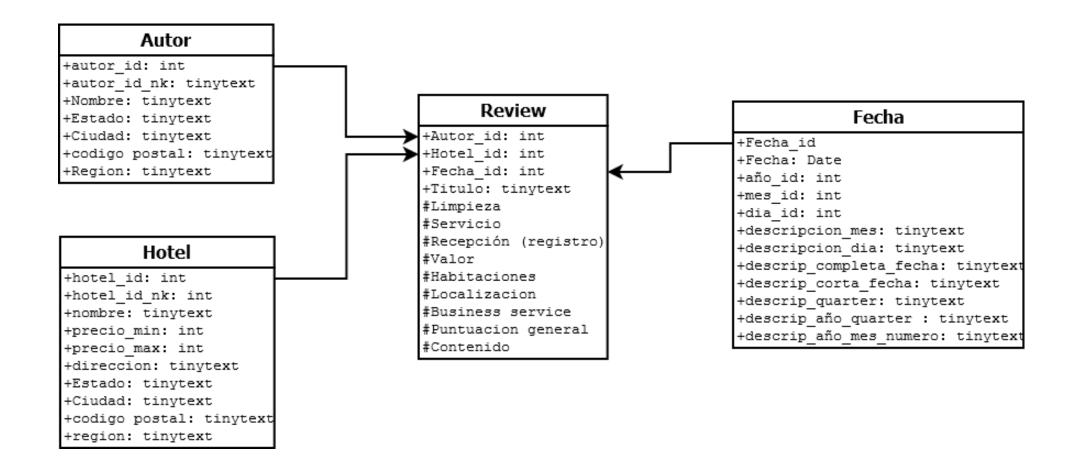


# FUENTES DE DATOS

Descripción del hotel sobre el que se han obtenido las reviews

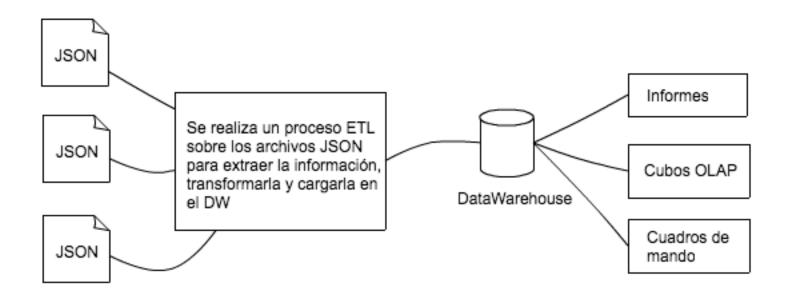


# ESTRUCTURA DATA WAREHOUSE





# ESQUEMA GENERAL DEL PROYECTO





# ETL

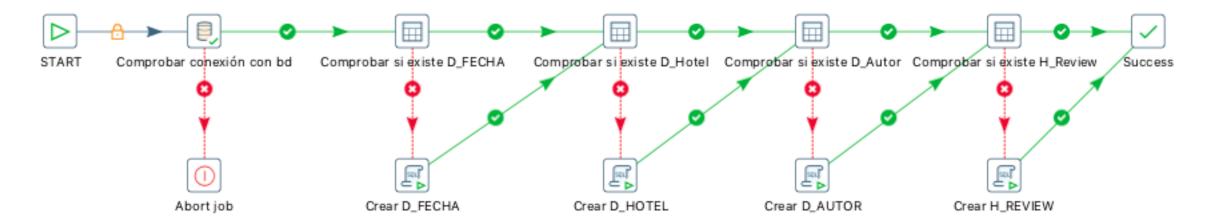
### Trabajo principal





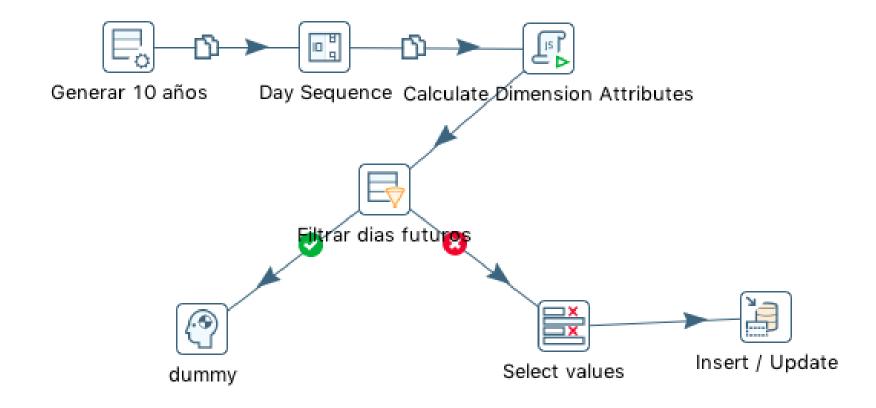
# ETL

### Crear Base de datos





# ETL - CARGA DE FECHA

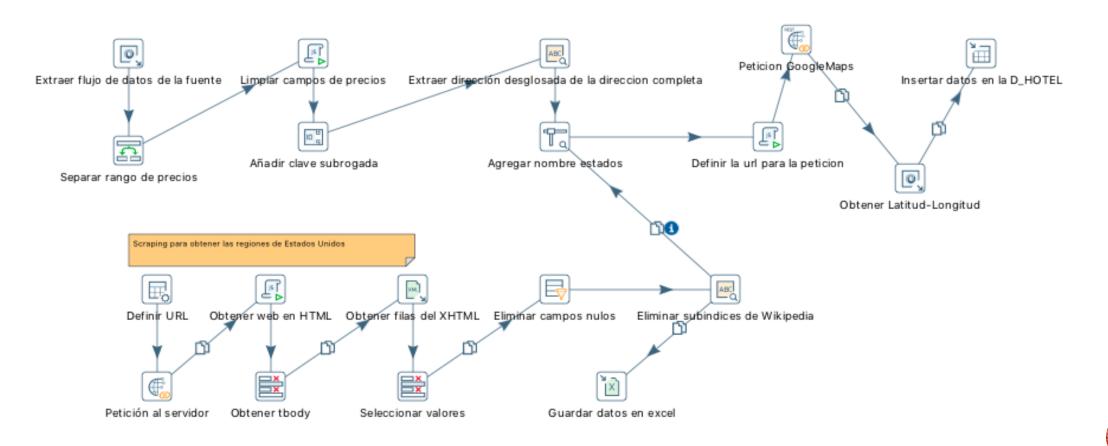


Cargar dimensión tiempo mediante un javascript que carga fechas hasta el día de hoy. Debería ejecutarse cada día

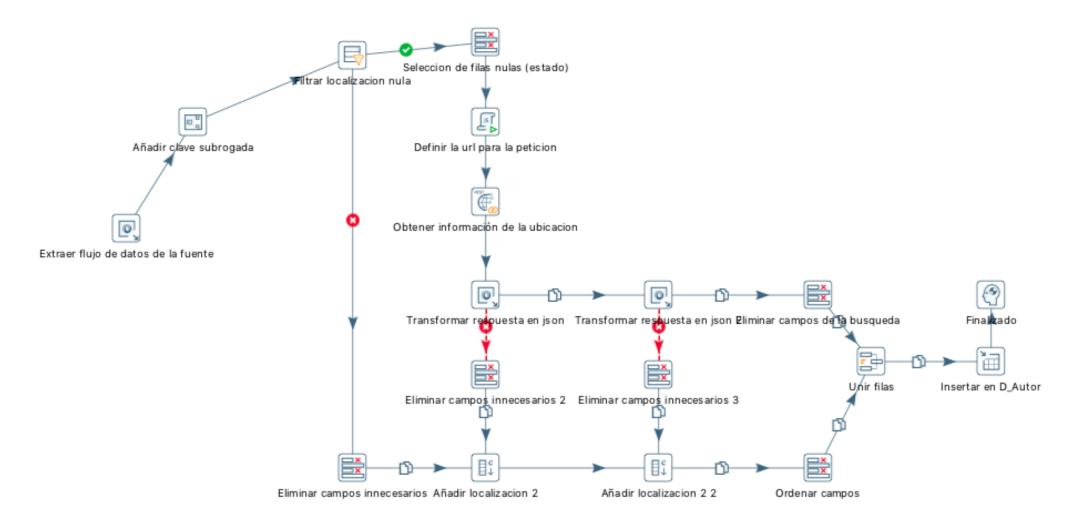


# ETL - CARGA DE HOTEL

Extraer los datos principales de la fuente

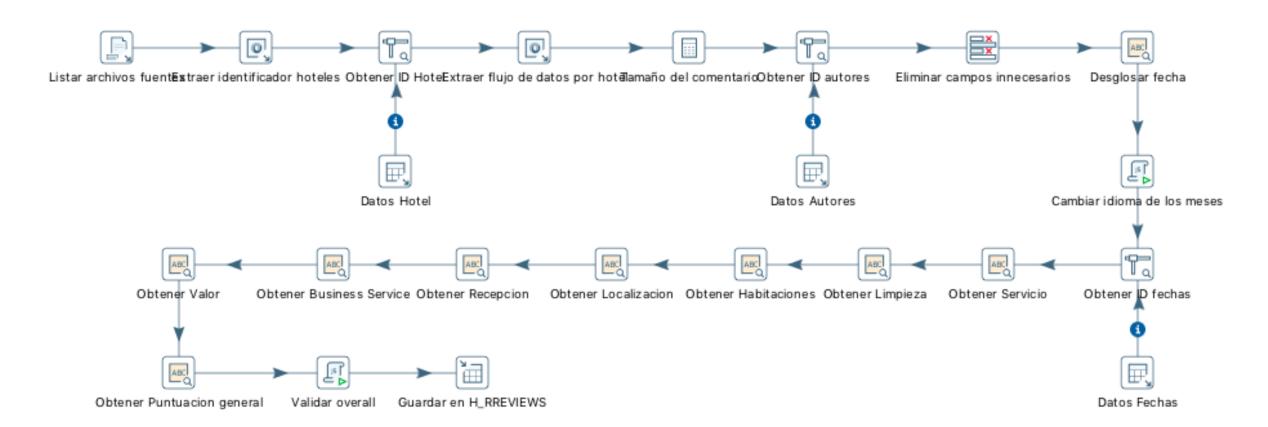


# ETL - CARGA DE AUTOR





# ETL - CARGA REVIEWS





## INFORMES

### Informe puntuación global

Se detalla la puntuación media de las distintas métricas de todos los hoteles.

### Informe de hotel específico

•Se detalla todos los autores que han realizado un review, con la puntuación que han asignado y número de caractéres que tiene el comentario.



# INFORMES



#### **VALORACIÓN DE HOTELES**

Nombre	Servicio	Limpieza	Localización	Habitación	Recepc ión	General
A Victory Inn & Suites Phoenix North	3	4	3	4	0	4
Americas Best Value Im Phoenix/I-10 West	5	4	3	5	0	5
BEST WESTERN Airport Inn	5	4	3	3	0	3
BEST WESTERN Loyal	5	5	5	5	0	5
BEST WESTERN Market Center	4	5	5	4	0	5
BEST WESTERN Phoenix I-17 MetroCenter Inn	3	4	4	4	0	4
BEST WESTERN PLUS Executive Inn	5	0	0	0	0	5
BEST WES TERN PLUS InnSuites Phoenix Hotel & Suites	0	0	0	0	0	5
BEST WESTERN PLUS Pioneer Square Hotel	4	5	5	5	0	5
Christopher's Inn	3	3	4	3	0	4
Comfort Inn & Suites Scattle	1	1	1	1	0	1
Courtyand by Marriott Phoenix North	5	5	4	5	0	4
Days Inn Camelback Phoenix and Conference Center	4	3	4	4	0	4
Days Inn I-17 & Thomas	5	3	2	2	0	2
Embassy Suites Hotel Phoenix-North	3	3	3	2	0	2
Embassy Suites Phoenix Aimort at 24th Street	3	4	2	3	0	2
Grace Inn Phoenix	5	4	4	4	0	4
Hampton Inn Phoenix Midtown (Downtown Area)	4	4	5	4	0	5
Hilton Phoenix Airport	5	5	4	4	0	4
Holiday Inn & Suites Phoenix Airport North	5	4	3	4	0	3



#### BEST WESTERN PLUS Pioneer Square Hotel

AUTOR	ESTADO	CIUDAD	PUNTUACION	CONTENIDO
gowhart32	Massachusetts	Boston	5	710
Nancy W	Wisconsin	Madison	5	675
Janet H	Alaska	Ketchikan	5	293
TimothyFlorida	Florida	N/A	5	513
KarenArmstiong_BC	British Columbia	Armstrong	5	833
Shane33333	N/A	NΆ	1	337
Bekruzn	N/A	NΆ	5	88
Teacherbear	Saskate he wan	N/A	4	406
CandyGnomad	Ontario	Kingston	4	366
idalosandy	klaho	Boise	5	111
CW2S	Maryland	Ellicott City	4	1376
jimmy62_11	Oklahoma	Duncan	5	620



# CUBOS OLAP

### • Un total de 3 consultas:

- Reviews ordenadas por Estados.
- Contenido de la review y puntuación por meses en cada ciudad.
- Contenido de la review y la puntuación por los estados donde residen los usuarios.



# CUBOS OLAP - ESTADOS

ESTADO	NOMBRE	PUNTUACION GENERAL	HABITACIONES	LIMPIEZA	LOCALIZACION	RECEPCION	SERVICIO	VALOR	BUSINESS SERVICE
Arizona	A Victory Inn & Suites Phoenix North	3	2	2	2	0,47	2	3	0,19
	Americas Best Value Inn Phoenix/I-10 West	3	2	1	2	0,00	2	2	0,00
	BEST WESTERN Airport Inn	4	3	3	4	1	3	4	1
	BEST WESTERN PLUS InnSuites Phoenix Hotel & Suites	4	3	4	3	1	4	4	0,22
	BEST WESTERN Phoenix I-17 MetroCenter Inn	4	3	4	3	0,48	3	4	0,39
	Courtyard by Marriott Phoenix North	4	3	3	3	0,26	3	3	0,26
	Days Inn Camelback Phoenix and Conference Center	3	2	2	2	0,00	3	3	0,00
	Days Inn I-17 & Thomas	3	2	2	1	1	2	2	0,46
	Embassy Suites Hotel Phoenix-North	3	3	3	3	1	3	3	1
	Embassy Suites Phoenix Airport at 24th Street	3	3	3	2	0,18	4	3	0,13
	Grace Inn Phoenix	3	2	3	3	1	3	3	1
	Hampton Inn Phoenix Midtown (Downtown Area)	4	3	4	3	1	4	4	1
	Hilton Phoenix Airport	4	3	4	3	1	4	4	1
	Holiday Inn & Suites Phoenix Airport North	4	4	4	3	0,00	4	4	0,00
	Holiday Inn Phoenix Downtown North	4	3	4	3	0,41	4	4	0,41
	Holiday Inn Phoenix-West	3	3	3	2	0,48	3	2	0,35
	Hyatt Regency Phoenix	4	3	4	4	1	4	4	0,34
	La Quinta Inn Phoenix North	4	3	4	3	0,44	4	3	0,44
	La Quinta Inn Phoenix Thomas Road	3	3	3	3	0,47	3	3	0,24
	Lexington Hotel Central Phoenix	4	2	3	2	0,44	2	3	0,35
	Renaissance Phoenix Downtown	5	4	4	4	0,00	4	4	0,00
	Super 8 Phoenix	3	3	3	2	1	3	3	1
California	Christopher's Inn	4	3	3	3	1	4	4	1
Texas	BEST WESTERN Market Center	4	2	3	2	1	2	3	1
Washington	BEST WESTERN Loyal Inn	4	3	4	3	1	4	4	1
	BEST WESTERN PLUS Executive Inn	4	3	4	4	1	4	4	0,43
	REST WESTERN PLUS Pinneer Square Hotel	4	4	4	4	1	4	4	1



# CUBOS OLAP — MESES

Se muestra el contenido y puntuación por ciudad y mes del año

	en	ero	feb	rero	ma	arzo	а	bril	ju	nio	m	вуо
CIUDAD	CONTENIDO MEDIO	PUNTUACION GENERAL	CONTENIDO MEDIO	PUNTUACION (GENERAL								
Phoenix	763	4	715	4	842	4	808	4	975	4	894	4
Calistoga	384	5	641	5	540	5	510	4	425	4	820	4
Dallas	795	3	426	5	475	4	362	5	477	3	304	4
Seattle	775	4	680	4	696	4	1098	4	891	4	901	4



# CUBOS OLAP - USUARIOS

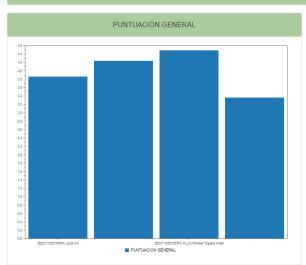
Puntuación general por estado en el que residen los diferentes usuarios.

ESTADO	CONTENIDO	PUNTUACION GENERAL
?stanbul	765	4
Alabama	1614	5
Alaska	5925	4
Alberta	9402	5
Arizona	10931	4
Arkansas	603	4
Bavaria	1671	2
Bremen	2662	3
British Columbia	32505	4
Brussels	1031	4
California	53013	4
Colorado	8901	5
Connecticut	590	4
Delaware	1738	5
District of Columbia	2397	4
Eastern Cape	568	5
England	13144	4
Florida	4307	4
Georgia	2113	4
Guangdong	245	5
Hawaii	713	5
Hong Kong Island	374	5
Idaho	1815	5
Illinois	5718	4
Indiana	5012	5

# CUADROS DE MANDO

#### BUSCA HOTEL CON TRIPADVISOR

Washington ▼



Nombre	Puntuación
BEST WESTERN Loyal Inn	2.8407
BEST WESTERN PLUS Executive Inn	3.1679
BEST WESTERN PLUS Pioneer Square Hotel	3.6266

Nombre	Puntuación
BEST WESTERN Loyal Inn	3.7434
BEST WESTERN PLUS Executive Inn	3.781
BEST WESTERN PLUS Pioneer Square Hotel	4.3433

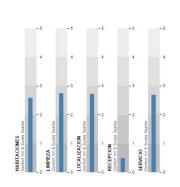
Top localizacion	
Nombre	Puntuación
BEST WESTERN Loyal Inn	3.7434
BEST WESTERN PLUS Executive Inn	3.781
BEST WESTERN PLUS Pioneer Square Hotel	4.3433

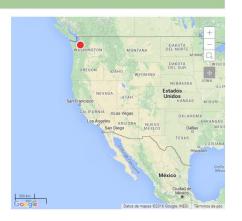
Nombre	Puntuación:
BEST WESTERN Loyal Inn	0.6991
BEST WESTERN PLUS Executive Inn	0.5839
BEST WESTERN PLUS Pioneer Square Hotel	0.7597

Nombre	Puntuación(
BEST WESTERN Loyal Inn	3.5841
BEST WESTERN PLUS Executive Inn	3.8321
BEST WESTERN PLUS Pioneer Square Hotel	4.2532



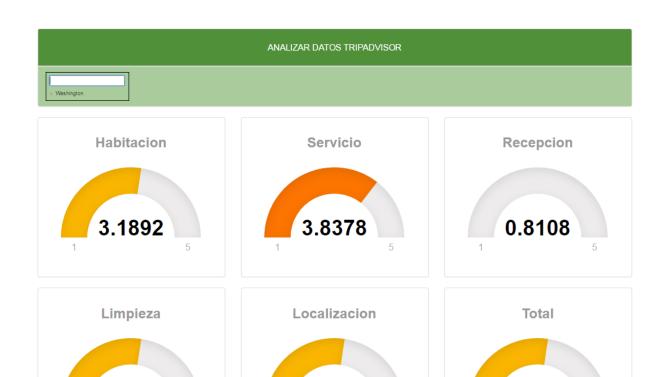
#### Información hotel





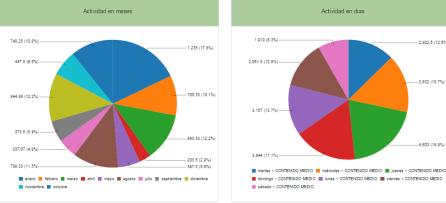


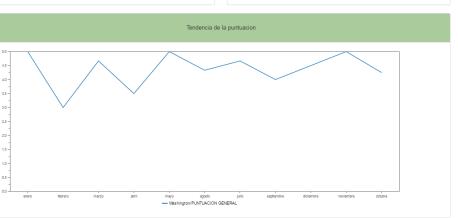
# CUADROS DE MANDO



3.1892

3.1892







2.922.5 (12.6%

\_ 3,632 (15.7%)

4,603 (19.9%)

## **PROBLEMAS**

- Configuración Windows/Linux de Kettle nos ha ocasionado problemas a la hora de manejar errores en el flujo de datos.
- Problemas con el API de Google Maps ya que limita el numero de consultas por día.
- Problemas con el lenguaje xPath para recorrer los nodos de un fichero JSON.
- Limpiar formato HTML al obtener web con JSOUP.
- Añadir métricas en componente mapa de CDE.
- Problemas al crear componente Dial en CDE.



# **PREGUNTAS**



