

# Impala

Marta Verona Almeida

3 de abril de 2019

## 1. Base de datos

Para la realización de esta práctica se ha utilizado una base de datos con información sobre vinos. Ésta ha sido extraída de 1 y consta de 8000 ejemplos con 10 columnas. A continuación se muestran sus nombres y tipos en *Pig*.

1. country: `chararray`.
2. description: `chararray`.
3. designation: `chararray`.
4. points: `int`.
5. price: `float`.
6. province: `chararray`.
7. region\_1: `chararray`.
8. region\_2: `chararray`.
9. variety: `chararray`.
10. winery: `chararray`.

En [2] se encuentra en el enlace a una carpeta en `google drive` que contiene el `csv` correspondiente a esta base de datos.

## 2. Consultas

La tabla creada en `imapala` se ha llamado `crimesdata`. A continuación se muestran las consultas realizadas sobre la misma. Con el objetivo de poder mostrar los resultados obtenidos en cada ejecución, se ha limitado la salida de todas las consultas a 20.

- Proyección.

Obtenemos, el precio y la puntuación de los vinos de la base de datos, mostrando sólo los diez primeros. Para ello se ejecutan las siguientes órdenes:

```
A = FOREACH winemag GENERATE price,points;  
B = LIMIT A 10;  
DUMP B;
```

En la figura 2 se muestra el resultado obtenido.

```
(19.0,83)
(14.0,86)
(13.0,87)
(15.0,87)
(11.0,88)
(24.0,88)
(25.0,90)
(40.0,90)
(33.0,95)
```

Figura 1: Proyección

#### ■ Selección.

Obtenemos toda la información sobre los vinos españoles, mostrando sólo los resultados asociados a los diez primeros. Para ello, se ejecutan los siguientes comandos:

```
A = FILTER winemag BY country == 'Spain';
B = LIMIT A 10;
DUMP B;
```

En la figura 2 puede verse el resultado obtenido.

```
(Spain,This opens with lightly perfumed red-fruit aromas of raspberry and plum, with a bright palate that shows mostly red pl
um, raspberry and cherry flavors. This is a basic, clean wine comprised of 75% Garnacha and 25% Tempranillo.,Castillo de Fuen
dejalón Crianza,85,9.0,Northern Spain,Campo de Borja,,Red Blend,Bodegas Aragonesas)
(Spain,Starts out heavy and roasted, with herbal, earthy berry aromas settling in. Feels prune-y but there's also acidity, whi
le the flavors run herbal and peppery as prune and plum represent the fruit side. Decent but doesn't fully click.,Dominio de
Heredia,84,15.0,Northern Spain,Rioja,,Tempranillo,Altanza)
(Spain,Lemony and citrusy from the get-go, with floral notes to the nose. The palate is sheering and tart, with edgy tannins.
As a whole, it's angular, with tomato and red-fruit flavors. A natural, jacked-up example of Prieto Picudo that's a bit unde
rwhelming.,Paraje de El Santo,84,69.0,Northern Spain,Tierra de León,,Prieto Picudo,Picum)
(Spain,Smells a little too much like berry candy and cough syrup to win you over, and the palate is kind of flat and chunky a
s well as thick with creamy berry fruit. Points of acidic bite give it life, but things are brought back down on the finish w
ith green, creamy oak notes.,Inspiración Valdemar,84,20.0,Northern Spain,Rioja,,Tempranillo Blend,Bodegas Valdemar)
(Spain,Crisp, natural and lightly salinic on the nose, with medium weight, good freshness and solid peach and citrus flavors.
This is a standard but nicely done Rioja rosé with a pithy finish and just the right edge to it. Drink liberally with appeti
zers and you shouldn't go wrong.,87,9.0,Northern Spain,Rioja,,Rosé,Marqués de Cáceres)
(Spain,Lush berry and marzipan aromas are darkened up by toast and graphite notes. This is a major-league specimen from Bierz
o with a full body and manageable tannins. Blackberry, baking spice, mocha and dark-roast-coffee flavors hold steady on the f
inish before hollowing out. Drink through 2024.,Bembibre,92,,Northern Spain,Bierzo,,Mencia,Dominio de Tares)
(Spain,This reserve-level P.X. sets the gold standard for excellence in sweet sherry. The lovely bouquet straddles the line b
etween unadulterated sweetness and impeccable slyness. The flavors of fig, chocolate, caramel, browned butter and cinnamon ar
e amazing. So chewy and thick, but balanced by firm acids. Brilliant.,Cardenal Cisneros Reservas,94,15.0,Andalucía,Jerez,,Ped
ro Ximénez,Sánchez Romate)
(Spain,This smooth, stylish wine is on the way up, as evidenced by the two-point bump it's receiving from its last review, in
December 2004. Today it's round and full, with aromas of hay and popcorn that draw you in. Layered through the middle, with
expressive white fruit, dryness and sweetness all in one. Imported by Pasternak Wine Imports.,Brut Reserva Privada,90,15.0,Ca
talonia,Cava,,Sparkling Blend,Marques de Monistrol)
(Spain,Here's yet another impressive Garnacha-based wine from the traveling Scotsman Norrel Robertson. There's a black-fruit
component that's aided by dried oregano, thyme and briar. The mouth, however, is juicy and ripe, with Syrah-influenced flavor
s of spice and herbs de Provence. Finishes with licorice, pepper and perfect size. The best wine in Robertson's portfolio.,9
1,21.0,Northern Spain,Calatayud,,Red Blend,Papa Luna)
grunt> █
```

Figura 2: Selección

#### ■ Agrupamiento con cálculo sobre él.

Se agrupan los vinos de la base de datos en función de su país y realiza un conteo de cuántos vinos provienen de cada país. Para ello, se ejecutan las siguientes órdenes.

```
A = FOREACH winemag GENERATE country, price;
B = GROUP A BY country;
```

```
C = FOREACH B GENERATE group, COUNT(A) AS NCases;  
DUMP C
```

El resultado obtenido puede verse en la figura 3.



```
(Chile,336)  
(Italy,993)  
(Spain,424)  
(Brazil,1)  
(Canada,10)  
(Cyprus,3)  
(France,815)  
(Greece,52)  
(Israel,32)  
(Mexico,4)  
(Turkey,3)  
(Austria,137)  
(Croatia,7)  
(Georgia,2)  
(Germany,106)  
(Hungary,11)  
(Lebanon,1)  
(Moldova,2)  
(Morocco,3)  
(Romania,2)  
(Uruguay,4)  
(country,0)  
(Bulgaria,3)  
(Portugal,215)  
(Slovenia,4)  
(Argentina,293)  
(Australia,272)  
(Lithuania,1)  
(Macedonia,2)  
(New Zealand,160)  
(South Africa,123)
```

Figura 3: Agrupamiento y cálculo

## Referencias

- [1] <https://www.kaggle.com/zynicide/wine-reviews>
- [2] <https://drive.google.com/file/d/1jTznZZthDLieiqnFT9rYQvmSTjouGn8I/view?usp=sharing>