

SEM309 – 10160 SEMINARIO DE PRÁCTICA EN CIENCIA DE DATOS

TRABAJO PRÁCTICO N° 3

INFORME TÉCNICO – BASE DE DATOS RELACIONAL

MARIANO NATIELLO (VLDC000632)

OCTUBRE 2024

1. ÍNDICE	1
2. CONTEXTO	2
3. MODELO DE BASE DE DATOS RELACIONAL	2
4. PREPARACIÓN DE DATOS	10
5. CONSULTAS SQL	10
6. ANÁLISIS MULTIDIMENSIONAL	14
7. CONCLUSIONES	22
8. REFERENCIA BIBLIOGRÁFICA	23

2. CONTEXTO

Considerando el dataset brindado como parte del Trabajo Práctico, el cual contiene varias reseñas de vinos que hemos trabajado anteriormente, se llevó a cabo tanto el diseño de una base de datos, su implementación, como así también la elaboración y ejecución de consultas sobre el conjunto de datos, desde aquellas más simples hasta consultas multidimensionales.

3. MODELO DE BASE DE DATOS RELACIONAL

Previo a una explicación del modelo de base de datos relacional construido, definimos a continuación ciertos conceptos principales:

Base de datos: en su libro, Reinosa define una base de datos como un conjunto de datos estructurados y definidos a través de un proceso específico, que busca evitar la redundancia, y que se almacenará en algún medio de almacenamiento masivo, como un disco. Por su lado, Elmasri en su libro “Fundamentos de Sistemas de Bases de datos”, define a la misma como una colección de datos relacionados, considerando a los datos como los hechos conocidos que se pueden grabar y que tienen un significado implícito. Para el presente informe, utilizamos la base definida con el nombre “WinemagData”.

Sistemas de Gestión de Base de datos: estos sistemas, por su sigla SGBD, son software diseñados especialmente para gestionar bases de datos, permitiendo crear, modificar y consultar información estructurada. Es nuestro caso, utilizamos MySQL como SGBD y MySQL Workbench como herramienta para facilitar la interacción con el mencionado SGBD.

Entidades: es un concepto abstracto que representa objetos del mundo real, pudiendo ser un objeto físico, como una persona o un producto o bien, una abstracción, como un “pedido” o una “venta”. En nuestro modelo, “Taster” representa al catador que realizó la reseña de los vinos.

Atributos: describen las características de dichas entidades. Siguiendo el caso anterior, se definen como atributos de la entidad Taster a “TasterName” o “TasterTwitterHandle”.

Relaciones: es una representación en dos dimensiones, o de doble entrada, constituida por filas, o tuplas y columnas o atributos. En otras palabras, dentro del diseño de la base de datos las relaciones establecen conexiones entre entidades reflejando justamente “como se relacionan entre sí”.

Tabla: es la implementación física de una entidad de la base de datos. En otras palabras, donde se almacenan los datos. Por ejemplo en nuestra base de datos “WinemagData”, la tabla Country almacena todos los países que están involucrados en las reseñas.

Fila (o Registro): cada fila representa una instancia única de la entidad definida por la tabla. En el ejemplo anterior, cada fila de la tabla “Country” representa un país diferente.

Columna (o Atributo): cada columna representa una característica o propiedad de la entidad. La tabla Country contiene columnas como “CodCountry”, “Country”.

Clave primaria o Primary Key (PK): es un atributo o conjunto de atributos que identifica de manera única cada registro (o fila) dentro de una tabla. Esto significa que cada valor de clave primaria es único en toda la tabla y no se repite. Según lo expresado por Reinosa en su libro “Bases de datos”, estas claves asumen una gran importancia en el diseño de la base de datos, dado que, si no se determinan bien los datos almacenados, no será posible identificar de manera inequívoca y provocarán la repetición de tuplas, que generarán problemas demasiado costosos para resolver en un sistema en producción.

Clave foránea o Foreign Key (FK): es también un atributo o combinación de atributos que permite la combinación de datos de las distintas relaciones del sistema.

Al igual que la clave primaria, es muy importante su correcta definición. Por ejemplo, en nuestro modelo, para la tabla Province, la clave foránea es CodCountry, estableciendo este campo una relación entre la tabla Province y la tabla Country al referenciar la columna CodCountry en Country, siendo esta su clave primaria. Al ser una clave foránea, CodCountry en la tabla Province permite únicamente valores que existen en la columna CodCountry de la tabla Country asegurando esto que cada provincia este vinculada a un país válido.

Por lo expresado anteriormente, toda clave foránea de una entidad requiere de la existencia de una clave primaria en otra entidad.

Cardinalidad: describe la relación numérica entre las entidades de dos tablas, indicando el numero de registros de una tabla que pueden estar asociados con registros de otra tabla. Existen tres tipos de cardinalidad:

- 1) Uno a Uno (1:1) en donde cada registro de la primera tabla se relaciona con un único registro en la segunda tabla y viceversa.
- 2) Uno a Muchos (1:N), donde un registro de la primera tabla puede estar relacionado con múltiples registros en la segunda tabla, pero cada registro en la segunda tabla se relaciona con un solo registro en la primera tabla.
- 3) Muchos a Muchos (N:N), donde varios registros en la primera corresponden a varios registros en la segunda tabla y viceversa.

Redundancia: se refiere a la duplicación innecesaria de datos, en donde la misma puede ocurrir cuando los mismos datos se repiten en varias filas sin necesidad, lo que puede provocar problemas de eficiencia y precisión en la base de datos.

Integridad Referencial: es la regla que asegura que las claves foráneas en una tabla tienen valores coincidentes en las claves primarias de la tabla relacionada, manteniendo consistencia entre tablas relacionadas.

Si vemos las tablas Country y Province, la integridad referencial se aplica en la definición de la clave foránea CodCountry en la tabla Province que hace referencia a la clave primaria CodCountry de la tabla Country.

Imaginemos en el ejemplo anterior, si en lugar de utilizar CodCountry, incluimos el nombre completo del país en cada registro de la tabla Province. Esto ocasionaría que la información del país se repitiera en cada registro de “provincia” que pertenezca al mismo país. Por ejemplo, Province podría tener estos registros:

- CodProvince=1, ProvinceName="Córdoba", CountryName="Argentina"
- CodProvince=2, ProvinceName="Mendoza", CountryName="Argentina"
- CodProvince=2, ProvinceName="Neuquén", CountryName="Argentina"

En este caso, el nombre del país “Argentina” se repite en cada registro de la tabla Province. En su lugar, usamos una clave foránea (CodCountry) permitiendo que Country almacene el nombre del país una sola vez y en la tabla Province solo sería necesario almacenar el código (CodCountry), reduciendo de esta manera la redundancia y optimizando el almacenamiento.

Normalización: es el proceso de organizar una base de datos con el objetivo de reducir la redundancia y mejorar la integridad de los datos. La normalización implica dividir tablas en estructuras mas pequeñas y definir relaciones entre ellas.

Habitualmente, se consideran las siguientes tres formas normales:

1. Primera Forma Normal (1NF), la cual requiere que cada columna tenga solo un valor y que cada registro sea único.
2. Segunda Forma Normal (2NF), la cual asegura que todos los atributos no clave dependan completamente de la clave primaria. En el caso que cada fila tenga varias dependencias funcionales parciales, será necesario dividir la tabla en dos nuevas tablas para lograr la Segunda Forma Normal.

3. Tercera Forma Normal (3NF), en esta forma cada tabla almacena solo la información que le corresponde sin redundancia, estando los datos más organizados, ocupando menos espacio y siendo más fáciles de actualizar.

Ahora bien, al partir del dataset recibido como documentación complementaria, el cual contiene unas 130.000 reseñas de vino, se consideraron algunos campos cuyo análisis determino la decisión de convertir estos campos (columnas) en una tabla separada dentro de la base de datos.

Esta decisión, se basó en factores claves tales como:

- a) Evitar la redundancia de datos al dividir los datos en varias tablas evitamos la duplicación de información.
- b) Mejorar la consistencia e integridad de los datos, con relaciones entre tablas mediante claves foráneas permitiendo definir reglas de integridad referencial, asegurando por ejemplo que cada provincia incluida en Region (CodProvince) exista en la tabla principal (Province).
- c) Facilitar la Normalización, lo cual al dividir datos en varias tablas de forma que se reduzcan las dependencias y redundancias, ayuda a organizar la base de datos para mejorar la eficiencia del almacenamiento y asegurar la integridad de los datos.
- d) Mejorar el rendimiento al utilizar tablas mas pequeñas, el motor de la base de datos puede optimizar mejor las operaciones y reducir la carga, especialmente al indexar claves primarias y foráneas para agilizar la recuperación de datos.
- e) Facilitar el mantenimiento y la actualización de los datos, al tener cada entidad su propia tabla (Country, Region, Province, Taster) es más fácil hacer actualizaciones o eliminar registros sin que esto afecte a la integridad de otros datos.

A continuación, se detalla el modelo diseñado para la base de datos WinemagData, el cual se adjunta como anexo:

TABLAS

Tabla	WineReview	
Descripción	Es la tabla principal donde se registran las reseñas de los vinos. Cada reseña está vinculada a diversas características del vino, como su país de origen, región, variedad, bodega y la persona que hizo la reseña (catador).	
Columnas	WineReview_id	Clave primaria de la tabla
	Country_id	Clave foránea que conecta con la tabla Country
	Province_id	Clave foránea que conecta con la tabla Province
	Region_id	Clave foránea que conecta con la tabla Region
	Taster_id	Clave foránea que conecta con la tabla Taster
	Variety_id	Clave foránea que conecta con la tabla Variety
	Winery_id	Clave foránea que conecta con la tabla Winery
	DescriptionName	Indica las descripciones, incluyendo notas de cata detalladas, que un sommelier utilizaría para evaluar el vino
	DesignationName	Indica el viñedo específico dentro de la bodega
	Points	Indica la puntuación según la calidad percibida del vino
	Price	Indica el precio de una botella de vino.
	Title	Indica el título de la reseña.

Tabla	Country	
Descripción	Almacena información sobre los países de origen del vino.	
Columnas	Country_id	Clave primaria de la tabla
	CountryName	Nombre del país

Relación	La columna Country_id en la tabla WineReview es una clave foránea que indica el país de origen del vino.
-----------------	--

Tabla	Province	
Descripción	Almacena información sobre la provincia que junto con la columna de país (Country_id) proporciona información geográfica.	
Columnas	Province_id	Clave primaria de la tabla
	ProvinceName	Nombre de la provincia
	Country_id	Clave foránea que conecta con Country_id de la tabla Country (Country.Country_id) indicando a que país pertenece la provincia.
Relación	La columna Province_id en la tabla WineReview hace referencia a la provincia específica de origen del vino.	

Tabla	Region	
Descripción	Almacena información sobre las áreas vinícolas específicas permitiendo identificar las características particulares de los vinos de esas regiones.	
Columnas	Region_id	Clave primaria de la tabla
	RegionName1	Nombre de la región
	RegionName2	Nombre de la región complementaria.
	Province_id	Clave foránea que conecta con Province_id de la tabla Province (Province.Province_id) indicando a que provincia pertenece la región.
Relación	La columna Region_id en la tabla WineReview hace referencia a la región específica de origen del vino.	

Tabla	Taster	
Descripción	Almacena información sobre el catador que realizó la reseña.	
Columnas	Taster_id	Clave primaria de la tabla
	TasterName	Nombre del catador

	TasterTwitterHandle	Usuario en Twitter del catador (opcional).
Relación	La columna Taster_id en la tabla WineReview es una clave foránea que se conecta con Taster.Taster_id indicando quién hizo la reseña.	

Tabla	Variety	
Descripción	Almacena información sobre la variedad de uva utilizada para la elaboración del vino.	
Columnas	Variety_id	Clave primaria de la tabla
	VarietyName	Nombre de la variedad de la uva
Relación	La columna Variety_id en la tabla WineReview indica la variedad de la uva utilizada en el vino que se hizo la reseña.	

Tabla	Winery	
Descripción	Almacena información sobre la bodega que produjo el vino.	
Columnas	Winery_id	Clave primaria de la tabla
	WineryName	Nombre de la bodega
Relación	La columna Winery_id en la tabla WineReview indica la bodega que produjo el vino del cual se hizo la reseña.	

RELACIONES ENTRE TABLAS

- WineReview como tabla central está relacionada con otras tabla a través de claves foráneas (Country_id, Province_id, Region_id, Taster_id, Variety_id y Winery_id)
- Las tablas Country, Province y Region están relacionadas jerárquicamente de la siguiente manera:
 - Province está relacionada con Country a través de Country_id.
 - Region está relacionada con Province a través de Province_id.

- La tabla principal WineReview está relacionada con las tablas Winery y Variety para detallar el tipo de vino y la bodega que lo produce.
- También esta tabla principal (WineReview) tiene una relación con la tabla Taster para indicar el catador que hizo la reseña del vino.

4. PREPARACIÓN DE DATOS

Considerando los datos brindados en el dataset como parte del Trabajo Práctico N° 3, a modo de resumen se detallan las operaciones realizadas:

- a) Se creó la base de datos WineMagData utilizando MySQL Workbench.
- b) Se crearon las diferentes tablas “maestras” y se insertaron los correspondientes datos utilizando el script “1. SemCD_TP3.sql”.
- c) Se creó una tabla temporal con el nombre “TmpWineData”.
- d) Se insertaron los datos en dicha tabla temporal utilizando el script “1.1 Script_TmpWineData”
- e) Se creó una tabla temporal con el nombre TmpDesignation, insertando los datos en dicha tabla a través del script “1.2 Inserta_Designation”.
- f) Se actualizó el campo Designation de la tabla TmpWineData utilizando los datos de la tabla TmpDesignation, respetando que el campo identificador de dicha tabla coincida con el campo id de la tabla TmpWineData.
- g) Finalmente, utilizando los datos incluidos en la tabla TmpWineData, se insertaron los registros en la tabla principal WineReview, respetando los códigos (y relaciones) de esta tabla con aquellas definidas anteriormente, como por ejemplo, Country, Province, Taster, entre otras.

5. CONSULTAS SQL

a) Permite obtener la cantidad de vinos por país.

```
SELECT c.CountryName as Pais, COUNT(*) AS cantidad_vinos
FROM WineReview wr      -- tabla principal desde donde obtenemos las reseñas
JOIN Country c          -- realiza join con la tabla desde donde obtenemos los países
on wr.Country_id=c.Country_id -- relaciona las dos tablas por el campo Country_id
GROUP BY c.CountryName  -- agrupa por el nombre del país
ORDER BY cantidad_vinos DESC; -- ordena por la cantidad de vinos descendente
```

Pais	cantidad_vinos
US	48608
France	19535
Italy	17295
Spain	5873
Portugal	5053
Chile	4026
Argentina	3412
Austria	2920
Australia	2120
Germany	1900
South Africa	1239
New Zealand	1230
Israel	444
Greece	413
Canada	223
Hungary	129
Bulgaria	129
Romania	100
Uruguay	92
Turkey	80
Georgia	76
Slovenia	75
Mexico	66
England	63
Croatia	62
Moldova	55
Brazil	48
Lebanon	32
Morocco	24
Peru	16
Ukraine	13
Czech Republic	11
Macedonia	11
Cyprus	10
Serbia	8
India	8
Switzerland	6
Luxembourg	5
Armenia	2
Slovakia	1
Bosnia and Herzegovina	1

China	1
Egypt	1

b) Variedades de uva más comunes para una región en particular. En este caso consideramos la región “California”.

```
SELECT vr.VarietyName AS variedad, COUNT(*) AS cantidad
FROM WineReview wr -- tabla principal con reseña
JOIN Variety vr ON wr.Variety_id = vr.Variety_id -- join con tabla Variety
JOIN Region r ON wr.Region_id = r.Region_id -- join con tabla Region
WHERE r.RegionName1 = 'California' -- filtra por región igual a California
GROUP BY vr.VarietyName -- agrupa por la variedad
ORDER BY cantidad DESC; -- Ordena de mayor a menor cantidad
```

Variedad	Cantidad
Chardonnay	343
Red Blend	316
Cabernet Sauvignon	309
Pinot Noir	206
Merlot	173
Pinot Grigio	133
Zinfandel	130
Sauvignon Blanc	128
White Blend	98
Sparkling Blend	95
Moscato	56
Syrah	51
Ros	46
Petite Sirah	40
Riesling	29
Rh ne-style Red Blend	26
Viognier	23
Shiraz	17
Bordeaux-style Red Blend	13
Pinot Gris	12
Champagne Blend	9
Malbec	8
Meritage	7
Grenache	7
Orange Muscat	7
Barbera	6
Sangiovese	6
Verdelho	6
Muscat	5

Port	5
Cabernet Franc	4
Semillon-Sauvignon Blanc	4
Black Muscat	4
Tempranillo	4
Muscat Canelli	4
Chenin Blanc-Viognier	4
Symphony	3
Rosé	3
Tannat	3
Petit Verdot	3
Vermentino	2
Chenin Blanc	2
Claret	2
Black Monukka	2
Cabernet Merlot	2
G-S-M	2
Cabernet Sauvignon-Shiraz	1
Dolcetto	1
Grenache Blanc	1
Arneis	1
Rosato	1
Madeira Blend	1
Viognier-Chardonnay	1
Mourv dre	1
Tinta Madeira	1
Syrah-Petite Sirah	1
Torront s	1
Rh ne-style White Blend	1
Chardonnay-Viognier	1
Sauvignon Blanc-Semillon	1
Merlot-Cabernet Sauvignon	1
Aglianico	1
Fum Blanc	1
Montepulciano	1
Grenache-Syrah	1
Johannisberg Riesling	1
Fiano	1
Cabernet Sauvignon-Merlot	1
Tempranillo Blend	1

c) Precio promedio de la botella según la calificación obtenida

```

SELECT wr.points AS calificacion, -- selecciona el puntaje obtenido
format(AVG(wr.Price),2) AS precio_promedio -- calcula el precio promedio
FROM WineReview wr
GROUP BY wr.points
ORDER BY calificacion; -- ordena por calificación

```

calificacion	precio_promedio
80	163.22
81	167.79
82	182.57
83	174.33
84	180.94
85	186.32
86	206.33
87	232.31
88	267.72
89	295.96
90	344.20
91	403.33
92	473.63
93	576.14
94	749.19
95	1,010.27
96	1,427.03
97	1,861.57
98	2,164.31
99	2,411.52
100	4,859.47

6. ANÁLISIS MULTIDIMENSIONAL

d) Permite conocer la cantidad de reseñas publicadas. Es decir solo aquellas con puntaje 80 o mayor

```

SELECT count(*) as CantResPublicadas FROM
WineReview AS wr
WHERE WR.POINTS >= 80;

```

CantResPublicadas	^
115472	

e) Esta consulta extrae 10 reseñas con puntaje igual a 100, incluyendo información como el país, provincia, región y catador.

```

SELECT
    wr.WineReview_id,          -- ID único de la reseña
    c.CountryName,             -- Nombre del país
    p.ProvinceName,            -- Nombre de la provincia
    r.RegionName1,             -- Nombre de la región
    ta.TasterName,              -- Nombre del catador
    ta.TasterTwitterHandle,    -- Twitter del catador (si existe)
    v.VarietyName,              -- Nombre de la variedad de uva
    w.WineryName,               -- Nombre de la bodega
    wr.DescriptionName,         -- Descripción del vino
    wr.DesignationName,         -- Designación del vino
    wr.Points,                  -- Puntos asignados al vino
    wr.Price,                   -- Precio del vino
    wr.Title                    -- Título de la reseña
FROM
    WineReview AS wr
LEFT JOIN
    Country AS c ON wr.Country_id = c.Country_id
LEFT JOIN
    Province AS p ON wr.Province_id = p.Province_id
LEFT JOIN
    Region AS r ON wr.Region_id = r.Region_id
LEFT JOIN
    Taster AS ta ON wr.Taster_id = ta.Taster_id
LEFT JOIN
    Variety AS v ON wr.Variety_id = v.Variety_id

```


LEFT JOIN

Winery AS w ON wr.Winery_id = w.Winery_id

WHERE WR.POINTS >= 100 -- solo considera aquellos con puntaje igual a 100

limit 10; -- considera solo 10 registros

WineRev...	CountryName	ProvinceNa...	RegionName1	TasterName	TasterTwitterHan...	VarietyName	WineryName	DescriptionName	Designation...	Points	Price	Title
59464	France	Bordeaux	NULL	Roger Voss	@vossroger	Bordeaux-style Red Bl...	Ch teau Lafite Rothschild	Almost black in color, this stunning...		100.00	15000.00	Ch teau Lafite Rothschild
64951	France	Bordeaux	NULL	Roger Voss	@vossroger	Bordeaux-style White B...	Ch teau Haut-Brion	Full of ripe fruit, opulent and concen...		100.00	8480.00	Ch teau Haut-Brion 20
83700	US	California	Napa Valley	NULL	NULL	Cabernet Sauvignon	Cardinale	Tasted in a flight of great and famou...		100.00	2000.00	Cardinale 2006 Caber
88132	France	Champagne	NULL	Roger Voss	@vossroger	Champagne Blend	Louis Roederer	This latest incarnation of the famou...	Cristal Vinta...	100.00	2500.00	Louis Roederer 2008 (
21933	Italy	Tuscany	NULL	NULL	NULL	Merlot	Tenuta dell'Ornellaia	A perfect wine from a classic vintag...	Masseto	100.00	4600.00	Tenuta dell'Ornellaia 2
378	Australia	Victoria	NULL	Joe Czerwi...	@JoeCz	Muscat	Chambers Rosewood Vi...	This wine contains some material o...	Rare	100.00	3500.00	Chambers Rosewood
23515	Portugal	Douro	NULL	Roger Voss	@vossroger	Portuguese Red	Casa Ferreirinha	This is the latest release of what ha...	Barca-Velha	100.00	4500.00	Casa Ferreirinha 2008
7299	Italy	Tuscany	NULL	NULL	NULL	Prugnolo Gentile	Avignonesi	Thick as molasses and dark as cara...	Occhio di Pe...	100.00	2100.00	Avignonesi 1995 Occh
60486	US	Washington	Columbia Valley...	Paul Gregutt	@paulgwine	Syrah	Charles Smith	In 2005 Charles Smith introduced th...	Royal City	100.00	800.00	Charles Smith 2006 R
65234	US	Washington	Walla Walla Vall...	Paul Gregutt	@paulgwine	Syrah	Cayuse	Initially a rather subdued Frog as if i...		100.00	800.00	Cayuse 2008 Bionic Fi

f) Calcula el puntaje promedio de los vinos para cada bodega

SELECT
 w.WineryName as Bodega, -- Nombre de la bodega
 FORMAT(AVG(wr.Points), 2) AS PuntajePromedio -- Puntaje promedio reseñas

FROM

WineReview AS wr

JOIN

Winery AS w ON wr.Winery_id = w.Winery_id

GROUP BY

w.WineryName; -- Agrupa los resultados por bodega

Bodega	PuntajePromedio
Nicosia	87.70
Quinta dos Avidagos	87.63
Rainstorm	88.00
St. Julian	85.50
Sweet Cheeks	86.81
Tandem	88.40
Terre di Giurfo	87.63
Trimbach	91.75
Heinz Eifel	87.71
Jean-Baptiste Adam	89.22
Kirkland Signature	86.07
Leon Beyer	90.05
Louis M. Martini	88.76
Masseria Setteporte	87.50
Mirassou	85.56
Richard B cking	89.00
Felix Lavaque	85.33
Gaucho Andino	84.00
Pradorey	86.23
Qui vremont	85.67
Acrobat	88.21
Baglio di Pianetto	87.27
Bianchi	85.67
Canicatt	86.67
Castello di Amorosa	89.65
Stemmari	86.71
Clarksburg Wine C...	86.57
Domaine de la Ma...	87.14

g) Devuelve los 10 catadores que tienen más de 8000 reseñas ordenados por mayor a menor cantidad de reseñas

```

SELECT
    ta.TasterName as Catador,           -- Nombre del catador
    COUNT(wr.WineReview_id) AS TotalResenas -- Total de reseñas del catador
FROM
    WineReview AS wr
JOIN
    Taster AS ta ON wr.Taster_id = ta.Taster_id
GROUP BY
    ta.TasterName                       -- Agrupa los resultados por catador
HAVING
    TotalResenas > 8000                 -- solo catadores con más de 8000 reseñas
order by TotalResenas desc             -- ordena el resultado de mayor a menor
limit 10;                             -- visualiza solo 10 registros que cumplen con la condición

```



Result Grid		Filter Rows:
Catador	TotalResenas	
Roger Voss	22634	
Michael Schachner	13491	
Kerin O Keefe	9198	
Paul Gregutt	8524	
Virginie Boone	8394	

h) Lista las regiones de un país específico (por ejemplo Argentina)

```

SELECT
    r.RegionName1           -- Nombre de la región
FROM
    Region AS r
JOIN
    Province AS p ON r.Province_id = p.Province_id -- join con tabla de las provincias
JOIN
    Country AS c ON p.Country_id = c.Country_id
WHERE
    c.CountryName = 'Argentina' -- filtra para el país Argentina
order by r.RegionName1; -- ordena alfabeticamente

```

Result Grid   Filter Rows		
	RegionName1	
	Agrelo	
	Alto Valle del R o Negro	
	Alto Valle del Río Negro	
	Altos de Mendoza	
	Argentina	
	Buenos Aires	
	Cafayate	
	Calchaqu Valley	
	Catamarca	
	Famatina Valley	
	Jujuy	
	La Consulta	
	La Rioja	
	Luj n de Cuyo	
	Maip	
	Maipú	
	Medrano	
	Mendoza	
	Neuqu n	
	Patagonia	
	Pedernal Valley	
	Perdriel	
	R o Negro Valley	
	Salta	
	San Carlos	
	San Juan	
	San Rafael	
	Tulum Valley	
	Tupungato	
	Uco Valley	
	Valle de Uco	
	Vista Flores	
	Zonda Valley	

- i) **Selecciona información detallada de las bodegas, países, regiones y provincias para los 10 vinos con mayor puntaje.**

```

SELECT
w.WineryName AS Bodega,           -- nombre de la bodega
c.CountryName AS País,            -- nombre del país
r.RegionName1 AS Región,          -- nombre de la región principal
p.ProvinceName AS Provincia,      -- nombre de la provincia
wr.Title AS Título,               -- título de la reseña del vino
    FORMAT(wr.Points, 2) AS Puntaje, -- puntos asignados al vino
    FORMAT(wr.Price, 2) AS Precio   -- precio de una botella de vino

```

```

FROM
    WineReview AS wr
JOIN
    Winery AS w ON wr.Winery_id = w.Winery_id
JOIN
    Country AS c ON wr.Country_id = c.Country_id
JOIN
    Province AS p ON wr.Province_id = p.Province_id
JOIN
    Region AS r ON wr.Region_id = r.Region_id
ORDER BY
    wr.Points DESC          -- ordena por puntaje de mayor a menor
LIMIT 10;                 -- limita los resultados a los 10 vinos con mayor puntaje

```

Bodega	País	Región	Provincia	Título	Puntaje	Precio
Verit	US	Sonoma County	California	Verit 2007 La Muse Red (Sonoma County)	100.00	4,500.00
Cardinale	US	Napa Valley	California	Cardinale 2006 Cabernet Sauvignon (Napa Vall...	100.00	2,000.00
Cayuse	US	Walla Walla Valley (WA)	Washington	Cayuse 2008 Bionic Frog Syrah (Walla Walla Va...	100.00	800.00
Charles Smith	US	Columbia Valley (WA)	Washington	Charles Smith 2006 Royal City Syrah (Columbia...	100.00	800.00
Alpha Omega	US	Napa Valley	California	Alpha Omega 2012 ERA Red (Napa Valley)	99.00	3,000.00
Alpha Omega	US	Atlas Peak	California	Alpha Omega 2012 Stagecoach Vineyard Cabe...	99.00	2,500.00
Failla	US	Sonoma Coast	California	Failla 2010 Estate Vineyard Chardonnay (Sono...	99.00	440.00
David Arthur	US	Napa Valley	California	David Arthur 2009 Elevation 1147 Estate Caber...	99.00	1,500.00
Trefethen	US	Oak Knoll District	California	Trefethen 2005 Reserve Cabernet Sauvignon (...)	99.00	1,000.00
Cayuse	US	Walla Walla Valley (OR)	Oregon	Cayuse 2009 En Chamberlin Vineyard Syrah (...)	99.00	750.00

j) Muestra las regiones con al menos 100 bodegas

```

SELECT
    r.RegionName1 AS Región,          -- nombre de la región
    p.ProvinceName AS Provincia,      -- nombre de la provincia
    c.CountryName AS País,           -- nombre del país
    COUNT(DISTINCT w.Winery_id) AS TotalBodegas -- total bodegas en la región
FROM
    Region AS r
JOIN
    Province AS p ON r.Province_id = p.Province_id
JOIN
    Country AS c ON p.Country_id = c.Country_id
JOIN
    WineReview AS wr ON wr.Region_id = r.Region_id
JOIN
    Winery AS w ON wr.Winery_id = w.Winery_id
GROUP BY
    r.RegionName1, p.ProvinceName, c.CountryName -- Agrupa por región, provincia
    y país
HAVING
    TotalBodegas >= 100              -- filtra para mostrar regiones con 100 o más bodegas
order by totalbodegas desc;

```

Región	Provincia	País	TotalBodegas
Napa Valley	California	US	897
California	California	US	697
Russian River Valley	California	US	503
Columbia Valley (WA)	Washington	US	426
Paso Robles	California	US	350
Sonoma County	California	US	331
Sonoma Coast	California	US	320
Willamette Valley	Oregon	US	303
Santa Barbara County	California	US	240
Central Coast	California	US	232
Carneros	California	US	221
Lodi	California	US	201
Dry Creek Valley	California	US	192
Alexander Valley	California	US	177
Yakima Valley	Washington	US	168
Sta. Rita Hills	California	US	155
Walla Walla Valley (...)	Washington	US	152
Sonoma Valley	California	US	147
North Coast	California	US	138
Santa Maria Valley	California	US	137
Santa Lucia Highlands	California	US	134
Rutherford	California	US	131
Santa Ynez Valley	California	US	130
Oregon	Oregon	US	120
Red Mountain	Washington	US	119
Monterey County	California	US	118
Washington	Washington	US	113

k) Muestra los vinos con precios superiores al promedio

```

SELECT
  wr.Title AS Título,           -- título de la reseña
  w.WineryName AS Bodega,      -- nombre de la bodega
  c.CountryName AS País,       -- país del vino
  v.VarietyName AS Variedad,   -- variedad de uva
  wr.Points AS Puntaje,        -- puntos asignados
  wr.Price AS Precio           -- precio de la botella de vino
FROM
  WineReview AS wr
JOIN
  Winery AS w ON wr.Winery_id = w.Winery_id
JOIN
  Country AS c ON wr.Country_id = c.Country_id
JOIN
  Variety AS v ON wr.Variety_id = v.Variety_id

```


WHERE

-- filtra vinos cuyo precio está por encima del promedio

wr.Price > (SELECT AVG(Price) FROM WineReview)

ORDER BY

wr.Price DESC;

-- ordena de mayor a menor precio

Título	Bodega	País	Variedad	Puntaje	Precio
Ch teau les Ormes Sorbet 2013 M doc	Ch teau les Ormes Sorbet	France	Bordeaux-style Red Blend	88.00	33000.00
Domaine du Comte Liger-Belair 2010 La Roma...	Domaine du Comte Liger-Belair	France	Pinot Noir	96.00	25000.00
Blair 2013 Roger Rose Vineyard Chardonnay (A...	Blair	US	Chardonnay	91.00	20130.00
Domaine du Comte Liger-Belair 2005 La Roma...	Domaine du Comte Liger-Belair	France	Pinot Noir	96.00	20000.00
Ch teau P trus 2011 Pomerol	Ch teau P trus	France	Bordeaux-style Red Blend	97.00	20000.00
Ch teau Margaux 2009 Margaux	Ch teau Margaux	France	Bordeaux-style Red Blend	98.00	19000.00
Ch teau Cheval Blanc 2010 Saint- milion	Ch teau Cheval Blanc	France	Bordeaux-style Red Blend	100.00	15000.00
Ch teau Lafite Rothschild 2010 Pauillac	Ch teau Lafite Rothschild	France	Bordeaux-style Red Blend	100.00	15000.00
Ch teau Mouton Rothschild 2009 Pauillac	Ch teau Mouton Rothschild	France	Bordeaux-style Red Blend	96.00	13000.00
Ch teau Haut-Brion 2009 Pessac-L ognan	Ch teau Haut-Brion	France	Bordeaux-style Red Blend	96.00	12000.00
Domaine du Comte Liger-Belair 2006 La Roma...	Domaine du Comte Liger-Belair	France	Pinot Noir	94.00	11250.00
Emmerich Knoll 2013 Ried Loibenberg Smarag...	Emmerich Knoll	Austria	Gr ner Veltliner	94.00	11000.00
Ch teau La Mission Haut-Brion 2009 Pessac-L...	Ch teau La Mission Haut-Brion	France	Bordeaux-style Red Blend	97.00	11000.00
Ch teau La Mission Haut-Brion 2009 Pessac-L...	Ch teau La Mission Haut-Brion	France	Bordeaux-style White Bl...	94.00	10000.00
W. & J. Graham's NV 90-year Old Tawny (Port)	W. & J. Graham's	Portu...	Port	97.00	10000.00
Kopke 1935 Colheita White (Port)	Kopke	Portu...	Port	94.00	9800.00
Domaine Jacques Prieur 2014 Musigny	Domaine Jacques Prieur	France	Pinot Noir	95.00	9730.00
Ch teau Haut-Brion 2013 Pessac-L ognan	Ch teau Haut-Brion	France	Bordeaux-style White Bl...	97.00	9320.00
Biondi Santi 2006 Riserva (Brunello di Montalci...	Biondi Santi	Italy	Sangiovese Grosso	94.00	9000.00
Olivier Leflaive 2014 Montrachet	Olivier Leflaive	France	Chardonnay	97.00	8860.00
Penfolds 2010 Grange Shiraz (South Australia)	Penfolds	Austr...	Shiraz	99.00	8500.00
Penfolds 2008 Grange Shiraz (South Australia)	Penfolds	Austr...	Shiraz	98.00	8500.00
Ch teau Ausone 2007 Saint- milion	Ch teau Ausone	France	Bordeaux-style Red Blend	95.00	8500.00
Ch teau Haut-Brion 2014 Pessac-L ognan	Ch teau Haut-Brion	France	Bordeaux-style White Bl...	100.00	8480.00
Henschke 2010 Hill of Grace Shiraz (Eden Valley)	Henschke	Austr...	Shiraz	96.00	8200.00
Ch teau Haut-Brion 2007 Pessac-L ognan	Ch teau Haut-Brion	France	Bordeaux-style White Bl...	98.00	8000.00
Krug 2002 Clos du Mesnil Brut Blanc de Blancs...	Krug	France	Chardonnay	99.00	8000.00
Domaine Pell 2014 Morogues Ros (Menetou-S...	Domaine Pell	France	Ros	87.00	8000.00
Biondi Santi 2007 Riserva (Brunello di Montalci...	Biondi Santi	Italy	Sangiovese	94.00	8000.00
Ch teau d'Yquem 2005 Sauternes	Ch teau d'Yquem	France	Bordeaux-style White Bl...	97.00	8000.00
Burmester 1963 Colheita (Port)	Burmester	Portu...	Port	87.00	7900.00
Henschke 2009 Hill of Grace Shiraz (Eden Valley)	Henschke	Austr...	Shiraz	91.00	7800.00
Robert Weil 2012 Kiedrich Gr fenberg Trockenb...	Robert Weil	Germ...	Riesling	95.00	7750.00
Robert Weil 2013 Kiedrich Gr fenberg Trockenb...	Robert Weil	Germ...	Riesling	95.00	7750.00

I) Selecciona 10 reseñas en donde aparecen ciertas notas de cata en el campo Description y el puntaje es mayor a 95

SELECT wr.WineReview_id, w.WineryName, wr.DescriptionName,

wr.DesignationName, wr.points, wr.price, wr.title

FROM WineReview wr

join Winery w

-- join con la tabla Winery para obtener la bodega

on wr.Winery_id=w.Winery_id

-- une ambas tablas por el campo Winery_id

WHERE

wr.points>95 -- considera solo puntaje mayor a 95

-- cita algunas notas de cata para considerar en la búsqueda

AND (wr.DescriptionName LIKE '%tasting notes%'

OR wr.DescriptionName LIKE '%fruity%'

OR wr.DescriptionName LIKE '%floral%'

OR wr.DescriptionName LIKE '%earthy%'

OR wr.DescriptionName LIKE '%spicy%'

```

OR wr.DescriptionName LIKE '%toasted%'
OR wr.DescriptionName LIKE '%acidic%'
OR wr.DescriptionName LIKE '%sweet%'
OR wr.DescriptionName LIKE '%complex%')
order by wr.price desc      -- ordena por precio mayor a menor
limit 10;                  -- selecciona 10 registros

```

WineRev...	WineryName	DescriptionName	Designation...	points	price	title
88967	Domaine du Comt...	A superb wine from a great year, this is powerful and structured, with...		96.00	25000.00	Domaine du Comte Liger...
1591	Ch teau Margaux	A massive wine for Margaux, packed with tannins and ripe fruit. It has...		98.00	19000.00	Ch teau Margaux 2009 M...
41708	Penfolds	This has all the size and weight you've come to expect from Australia'...	Grange	98.00	8500.00	Penfolds 2008 Grange Shi...
94479	Krug	Pure Chardonnay from the walled vineyard in the heart of Le Mesnil, t...	Clos du Mes...	99.00	8000.00	Krug 2002 Clos du Mesnil...
102215	Ch teau d'Yquem	This isn't sweet, but just so wonderfully rich. It's the concentration of...		97.00	8000.00	Ch teau d'Yquem 2005 S...
380	Robert Weil	Dusty, saffron-spiced earthiness is juxtaposed against intense mango...	Kiedrich GrÃ...	97.00	7750.00	Robert Weil 2014 Kiedrich...
29966	Ch teau La Missio...	This is finely structured, a tense wine with great fruits and an intense...		96.00	7670.00	Ch teau La Mission Haut...
44611	Bouchard P re &...	From arguably the finest white wine vineyard in Burgundy, this is a po...		98.00	7570.00	Bouchard P re & Fils 2007...
12050	Ch teau Lafite Rot...	The wine shows the power typical of a Lafite but within the context of...		96.00	6850.00	Ch teau Lafite Rothschild...
64691	Robert Weil	Honey and smoke mix seductively on this piercingly bright yet intense...	Kiedrich GrÃ...	96.00	6720.00	Robert Weil 2010 Kiedrich...

7. CONCLUSIONES

Si tomamos en cuenta los resultados obtenidos en las consultas detalladas anteriormente, principalmente aquellas que involucran más de una dimensión, como región, variedad, precio es viable concluir lo siguiente:

- Las variedades de vino más populares de cada región o país pueden ayudar a identificar las preferencias del mercado local o internacional.
- El análisis de la relación precio-puntuación permite evaluar si realmente existe una correlación significativa entre el precio y la calidad percibida. Con esto, podríamos concluir que ciertos tipos de vinos o regiones ofrecen una mejor precio-calidad, lo cual puede ser útil para los consumidores y los productores.
- Importancia de las notas de cata. Si consideramos notas de cata particulares y su relación con el puntaje obtenido (y precio), es posible imaginar una potencial preferencia del consumidor en aquellos vinos cuya notas de cata respondan a ciertas características.

A nivel de posibles áreas para investigaciones futuras, considero que la aplicación de lenguaje natural (NLP) en las notas de cata, tal como lo mencionamos, permitiría identificar patrones en las descripciones de notas de cata y clasificar vinos según sabor y aroma.

Además de lo anterior, el análisis de reseñas de vino con técnicas avanzadas, que involucren también herramientas de análisis de datos, inteligencia artificial y modelos predictivos, sin duda que contribuirán significativamente en entender mejor el comportamiento del mercado.

8. REFERENCIA BIBLIOGRÁFICA

Material de Lectura N° 3 de la materia Seminario de práctica en Ciencia de Datos de la carrera Licenciatura en Ciencias de Datos.

Enunciado del Trabajo Práctico N° 3 de la materia Seminario de práctica en Ciencia de Datos de la carrera Licenciatura en Ciencias de Datos.

Elmasri, Ramez (2007). Fundamentos de Sistemas de Bases de Datos. Madrid, Pearson Educación S.A.

Reinosa, Enrique José (2012), Bases de datos, Buenos Aires, Alfaomega Grupo Editor Argentino.

Utilización de MySQL Workbench.