

2023

Data Analytics



Rocío Bottero

CODER HOUSE

Índice:

Presentación del proyecto:.....	2
Introducción:.....	3
Hipótesis:.....	4
Objetivos:.....	5
Usuario final:.....	5
Base de datos:.....	7
Diagrama de la Base de datos:.....	10
Descripción de las tablas:.....	11
Transformación de datos en Power BI:.....	13
Tabla calendario:.....	20
Tabla medidas:.....	21
Solapas versión inicial - mockup:.....	22
Solapas versión final:.....	26
Navegación entre solapas:.....	30
Conclusión:.....	32



Presentación del proyecto:

Un viaje por la calidad y variedad del café

En este proyecto, nos embarcaremos en un apasionante viaje a través de la calidad y variedad que ofrece esta apreciada bebida en distintos países productores alrededor del mundo. A lo largo de nuestro recorrido, exploraremos el aroma, cuerpo, sabor y otros aspectos clave para comprender qué naciones se destacan en la producción de café de alta calidad y qué comprenden las distintas variedades, como así las más producidas.



Nuestro objetivo principal es analizar y comparar diferentes parámetros relacionados con el café, utilizando como base un dataset gratuito obtenido de la reconocida plataforma Kaggle. Este conjunto de datos, titulado "Coffee Quality Data CQI", nos brindará una sólida base para investigar y descubrir los secretos que se esconden detrás de cada taza de café.

Al estudiar la información proporcionada en el dataset, seremos capaces de determinar qué tipo de café es más producido en cada país, así como establecer un ranking general de la calidad de cada uno. Además, investigaremos las empresas líderes en la producción de café, exploraremos la durabilidad y longevidad de los diferentes cafés, y analizaremos otros aspectos relevantes que influyen en la excelencia y reconocimiento de esta popular bebida.

Este proyecto nos brinda la oportunidad de adentrarnos en el apasionante mundo del café, así como de comprender las dinámicas y desafíos que enfrenta la industria cafetera a nivel global. A través del análisis de datos y la investigación complementaria, esperamos obtener resultados sólidos que nos permitan identificar los países destacados en la producción de café de alta calidad, y comprender los factores clave que contribuyen a su excelencia.



Introducción:

En este proyecto, nos introduciremos en el gran mundo del café para explorar la calidad y variedad que ofrecen distintos países productores. El objetivo principal es analizar y comparar en base a diferentes parámetros claves, como aroma, cuerpo, sabor y más, con el fin de comprender qué naciones se destacan en la producción de esta apreciada bebida.

Utilizaremos un dataset gratuito obtenido de la página Kaggle, el cual nos proporcionará información valiosa sobre la calidad del café:

<https://www.kaggle.com/datasets/fatihb/coffee-quality-data-cqi>

DATASET:  db_cafe.xlsx



Hipótesis:

Partiendo de estos datos, buscaremos descubrir qué tipos de café son más producidos en cada país, establecer un ranking general de la calidad de cada uno, identificar las empresas líderes en la producción, y explorar la durabilidad y longevidad de los diferentes cafés. A través de este análisis, esperamos obtener una visión más completa y detallada de los países que se destacan en la producción de café de alta calidad, así como de los factores que influyen en su excelencia y reconocimiento a nivel mundial.



Objetivos:

Nuestro objetivo principal es brindar una comprensión más profunda de la calidad y variedad del café en distintos países, así como de los factores que influyen en su excelencia. A través del análisis de datos, estableceremos un ranking de calidad, identificaremos empresas destacadas y exploraremos aspectos relacionados con la durabilidad y longevidad del café.



Usuario final:

Como apasionada del café y entusiasta del viaje, el usuario final de mi proyecto será tanto consumidores de café ansiosos por descubrir nuevas experiencias, como también empresas y productores del sector cafetero que buscan mejorar su oferta y estrategia de mercado.

Usuario final: Consumidores y amantes del café.

Nivel de aplicación: Operativo y táctico.

El análisis estará dirigido a brindar información útil y relevante a los consumidores de café y a aquellos que aprecian esta bebida. El enfoque operativo se centra en proporcionar información detallada sobre las características, sabores y perfiles de diferentes tipos de café, así como comparar empresas y países productores destacados. El nivel táctico implica ofrecer consejos y recomendaciones para disfrutar del café de manera óptima, como técnicas de preparación, maridajes, etc.

Usuario final: Empresas y productores de café.

Nivel de aplicación: Operativo, táctico y estratégico.

En este caso, el análisis estaría dirigido a las empresas y productores de café que deseen mejorar su oferta y estrategia de mercado. A nivel operativo, se podría proporcionar información detallada sobre las preferencias de los consumidores, las tendencias del mercado y las características clave del café en diferentes países. A nivel táctico, se podría ofrecer recomendaciones sobre la selección de granos y posicionamiento de las empresas. A nivel estratégico, se podría realizar un análisis competitivo, identificar oportunidades de mercado, evaluar la expansión a nuevos mercados y diseñar estrategias a largo plazo.



Base de datos:

```
create database Cafe
```

```
go
```

```
use Cafe
```

```
go
```

```
create table Paises(
```

```
idPais int identity(1,1),
```

```
Pais varchar(50),
```

```
constraint pk_pais primary key (idPais))
```

```
create table Regiones(
```

```
idRegion int identity(1,1),
```

```
Region varchar(50),
```

```
idPaisRegion int,
```

```
constraint fk_region_pais foreign key(idPaisRegion) references Paises(idPais),
```

```
constraint pk_region primary key(idRegion))
```

```
create table Granjas(
```

```
idGranja int identity(1,1),
```

```
Granja varchar(50),
```

```
idRegionGranja int,
```

```
constraint fk_granja_region foreign key(idRegionGranja) references Regiones(idRegion),
```

```
constraint pk_granja primary key(idGranja))
```

```
create table ProcesosDeMiel(
```

```
idProceso int identity(1,1),
```

```
NombreProceso varchar(50),
```

```
constraint pk_miel primary key(idProceso))
```




```
create table Valoraciones(  
  idValoracion int identity(1,1),  
  PuntajeGeneral float,  
  PuntajeTotal float,  
  constraint pk_valoracion primary key(idValoracion))
```

```
create table Caracteristicas(  
  idCaracteristica int identity(1,1),  
  Aroma float,  
  Sabor float,  
  Regusto float,  
  Acidez float,  
  Cuerpo float,  
  Balance float,  
  Uniformidad float,  
  Dulzura float,  
  Defectos float,  
  Color varchar(25),  
  constraint pk_caracteristicas primary key(idCaracteristica))
```

```
create table Variedades(  
  idVariedad int identity(1,1),  
  Variedad varchar(50),  
  idCaractVariedad int,  
  idMielVariedad int,  
  constraint fk_variedad_caracteristicas foreign key(idCaractVariedad) references  
  Caracteristicas(idCaracteristica),  
  constraint fk_variedad_proceso foreign key(idMielVariedad) references  
  ProcesosDeMiel(idProceso),
```



```
constraint pk_variedad primary key(idVariedad))
```

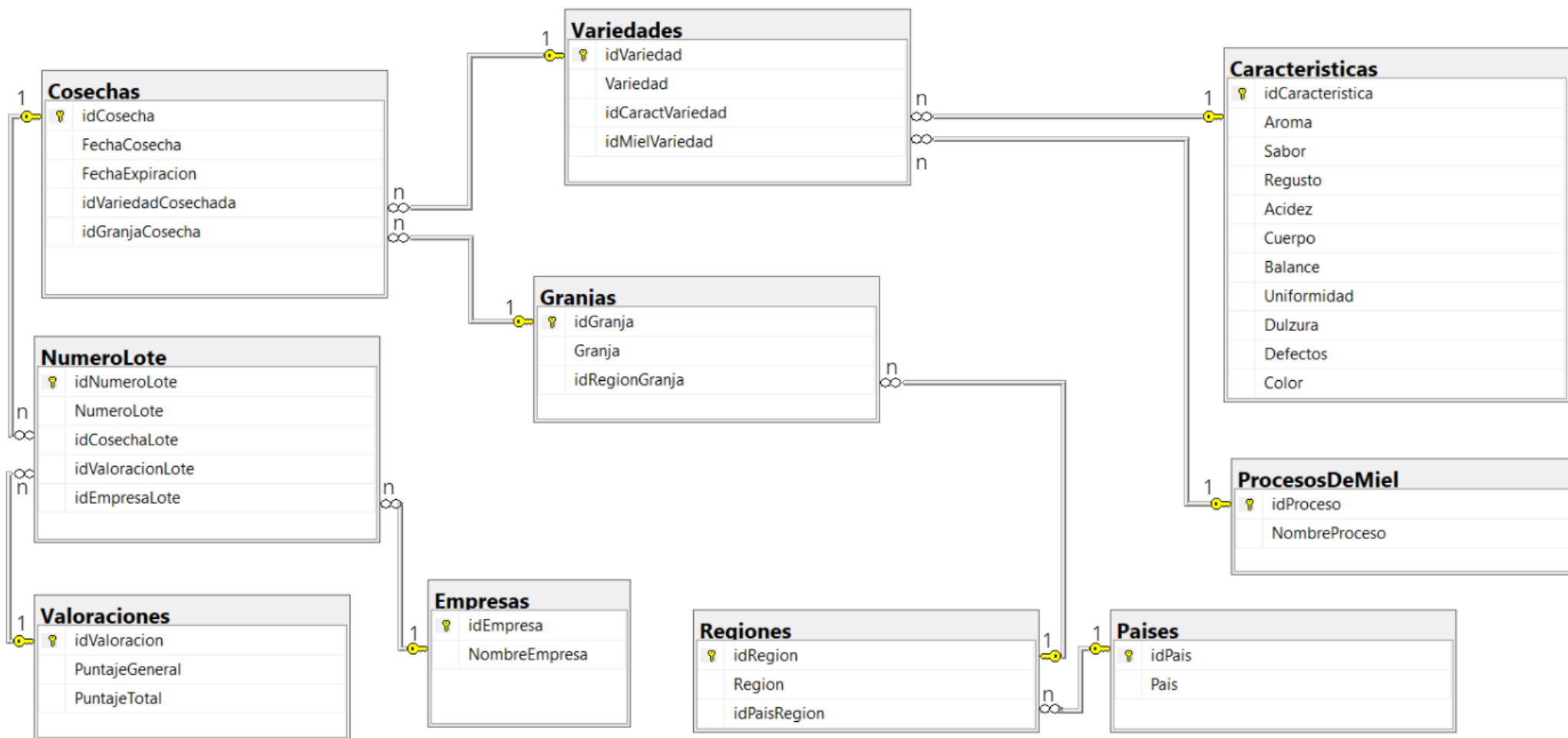
```
create table Cosechas(  
idCosecha int identity(1,1),  
FechaCosecha date,  
FechaExpiracion date,  
idVariedadCosechada int,  
idGranjaCosecha int,  
constraint fk_variedad_cosecha foreign key(idVariedadCosechada) references  
Variedades(idVariedad),  
constraint fk_granja_cosecha foreign key(idGranjaCosecha) references Granjas(idGranja),  
constraint pk_cosecha primary key (idCosecha))
```

```
create table Empresas(  
idEmpresa int identity(1,1),  
NombreEmpresa varchar(50),  
constraint pk_empresa primary key(idEmpresa))
```

```
create table NumeroLote(  
idNumeroLote int identity(1,1),  
NumeroLote varchar(30),  
idCosechaLote int,  
idValoracionLote int,  
idEmpresaLote int,  
constraint fk_valoracion_lote foreign key(idValoracionLote) references  
Valoraciones(idValoracion),  
constraint fk_cosecha_lote foreign key(idCosechaLote) references Cosechas(idCosecha),  
constraint fk_empresa_lote foreign key(idEmpresaLote) references Empresas(idEmpresa),  
constraint pk_n_lote primary key(idNumeroLote))
```



Diagrama de la Base de datos:



Descripción de las tablas:

Países: Almacena información sobre los países.

Columnas:

- idPais (entero, clave primaria)
- Pais (cadena de texto)

Regiones: Almacena información sobre las regiones.

Columnas:

- idRegion (entero, clave primaria)
- Region (cadena de texto)
- idPaisRegion (entero, clave foránea hacia la tabla Países)

Granjas: Almacena información sobre las granjas.

Columnas:

- idGranja (entero, clave primaria)
- Granja (cadena de texto)
- idRegionGranja (entero, clave foránea hacia la tabla Regiones)

ProcesosDeMiel: Almacena información sobre los procesos de miel.

Columnas:

- idProceso (entero, clave primaria)
- NombreProceso (cadena de texto)

Valoraciones: Almacena información sobre las valoraciones de los cafés.

Columnas:

- idValoracion (entero, clave primaria)
- PuntajeGeneral (número decimal)
- PuntajeTotal (número decimal)

Características: Almacena información sobre las características de los cafés.

Columnas:

- idCaracteristica (entero, clave primaria)
- Aroma (número decimal)
- Sabor (número decimal)
- Regusto (número decimal)



- Acidez (número decimal)
- Cuerpo (número decimal)
- Balance (número decimal)
- Uniformidad (número decimal)
- Dulzura (número decimal)
- Defectos (número decimal)
- Color (cadena de texto)

Variedades: Almacena información sobre las variedades.

Columnas:

- idVariedad (entero, clave primaria)
- Variedad (cadena de texto)
- idCaractVariedad (entero, clave foránea hacia la tabla Caracteristicas)
- idMielVariedad (entero, clave foránea hacia la tabla ProcesosDeMiel)

Cosechas: Almacena información sobre las cosechas.

Columnas:

- idCosecha (entero, clave primaria)
- FechaCosecha (fecha)
- FechaExpiracion (fecha)
- idVariedadCosechada (entero, clave foránea hacia la tabla Variedades)
- idGranjaCosecha (entero, clave foránea hacia la tabla Granjas)

Empresas: Almacena información sobre las empresas que se dedican a la producción de café.

Columnas:

- idEmpresa (entero, clave primaria)
- NombreEmpresa (cadena de texto)

NumeroLote: Almacena información sobre los números de lote.

Columnas:

- idNumeroLote (entero, clave primaria)
- NumeroLote (cadena de texto)
- idEmpresaLote (entero, clave foránea hacia la tabla Empresas)
- idCosechaLote (entero, clave foránea hacia la tabla Cosechas)
- idValoracionLote (entero, clave foránea hacia la tabla Valoracion)



Transformación de datos en Power BI:

Usar la primera fila como encabezado

Power BI Query Editor interface showing a data table and the 'Configuración de la consulta' (Query Settings) pane.

Query Formula: = Origen{{Item="Variedad",Kind="Sheet"}}[Data]

Column1	Column2
ID	Variedad
1	Castillo
2	Gesha
3	Java
4	Gesha
5	Red Bourbon
6	Gesha
7	Gesha
8	SL34+Gesha
9	SL34
10	Bourbon
11	Ethiopian Heirlooms
12	Gesha
13	Gesha
14	Gesha
15	Caturra
16	SL34
17	Wolishalo,Kurume,Dega
18	Gesha
19	Gesha
20	Bourbon
21	Gesha
22	Typica
23	Catimor
24	Castillo Paraguaycito
25	SL34
26	Typica
27	Gesha

Configuración de la consulta

- PROPIEDADES**
 - Nombre: Variedad
 - Todas las propiedades
- PASOS APLICADOS**
 - Origen
 - Navegación**
 - Encabezados promovidos
 - Tipo cambiado

ión de perfiles de columnas basada en las 1000 primeras filas

VISTA PREVIA DESCARGADA A LAS 20:45



Cambiar tipo de datos

			= Table.TransformColumnTypes("#Encabezados promovidos",{{"ID", Int64.Type}, {"Variedad", type text}})		Configuración de la consulta	
			ID		A ^B C Variedad	
1		1	Castillo			
2		2	Gesha			
3		3	Java			
4		4	Gesha			
5		5	Red Bourbon			
6		6	Gesha			
7		7	Gesha			
8		8	SL34+Gesha			
9		9	SL34			
10		10	Bourbon			
11		11	Ethiopian Heirlooms			
12		12	Gesha			
13		13	Gesha			
14		14	Gesha			
15		15	Caturra			
16		16	SL34			
17		17	Wolishalo,Kurume,Dega			
18		18	Gesha			
19		19	Gesha			
20		20	Bourbon			
21		21	Gesha			
22		22	Typica			
23		23	Catimor			
24		24	Castillo Paraguaycito			
25		25	SL34			
26		26	Typica			
27		27	Gesha			
28		28	Gesha			

ón de perfiles de columnas basada en las 1000 primeras filas

PROPIEDADES

Nombre

 Variedad

Todas las propiedades

PASOS APLICADOS

- Origen
- Navegación
- Encabezados promovidos
- Tipo cambiado

VISTA PREVIA DESCARGADA A LAS 20:45



Consultas [10]

Características

Cosecha

Empresas

Granjas

NroLote

Países

ProcesoDeMiel

Región

Valoración

Variedad

fx

= Table.TransformColumns(#"Tipo cambiado",{{"Color", Text.Upper, type text}})

	1.2 Balance	1.2 Uniformidad	1 ² ₃ Dulzura	1 ² ₃ Defectos	1 ² ₃ Categoría	AB _C Color
1	8,25	8,42	10	10	0	GREEN
2	7,92	8,25	10	10	0	BLUE-GREEN
3	7,92	8,17	10	10	0	YELLOWISH
4	8,17	8,08	10	10	0	GREEN
5	7,92	7,92	10	10	0	YELLOW-GREEN
6	7,83	8,17	10	10	0	GREEN
7	7,83	8,25	10	10	0	GREEN
8	7,92	8,08	10	10	0	YELLOW GREEN
9	7,92	8	10	10	0	GREENISH
10	8	8	10	10	0	GREENISH
11	7,92	7,92	10	10	0	GREENISH
12	8,25	8,17	10	10	0	BROWNISH
13	8	8	10	10	0	GREEN
14	8	8	10	10	0	YELLOW- GREEN
15	8,08	8	10	10	0	GREEN
16	8	7,92	10	10	0	GREEN
17	7,75	7,92	10	10	0	GREEN
18	7,92	7,92	10	10	0	YELLOW-GREEN
19	7,83	8	10	10	0	BROWNISH
20	8,17	7,75	10	10	0	GREENISH
21	7,75	8	10	10	0	YELLOWISH
22	8	7,83	10	10	0	GREEN
23	8,08	8	10	10	0	BROWISH-GREEN
24	7,75	7,83	10	10	0	BROWNISH
25	8	7,92	10	10	0	BLUIH-GREEN
26	8	7,83	10	10	0	2 GREEN
27	7,83	7,92	10	10	0	GREEN
28						

Configuración de la consulta

PROPIEDADES

Nombre

Características

Todas las propiedades

PASOS APLICADOS

Origen

Navegación

Encabezados promovidos

Tipo cambiado

✕ Texto en mayúsculas

12 COLUMNAS, 207 FILAS

Generación de perfiles de columnas basada en las 1000 primeras filas

VISTA PREVIA DESCARGADA A LAS 21:30

Buscar

23:29

24/7/2023

Bottero Rocío 17 | 33

Crear las relaciones entre las tablas

The screenshot displays a data modeling interface with a central canvas containing several tables. Each table is represented by a card with a title, a list of fields, and a 'Contrair' button. The tables and their fields are:

- Pais**: ID, País de origen
- Valoración**: ID, Puntaje general, Puntaje total
- Variedad**: ID, Variedad
- ProcesoDeMiel**: ID, Proceso de miel
- NroLote**: ID, Número de lote
- Características**: Acidez, Aroma, Balance, Categoría, Color, Cuerpo, Defectos, Dulzura
- Granjas**: ID, Nombre de la granja
- Cosecha**: Año de cosecha, Expiration, Fecha de cosecha
- Empresas**: Empresa, ID
- Región**: ID, Región

The interface includes a sidebar on the right with tabs for 'Datos', 'Propiedades', and 'Relaciones'. At the bottom, there is a bar with 'Todas las tablas' and a zoom slider set to 100%.



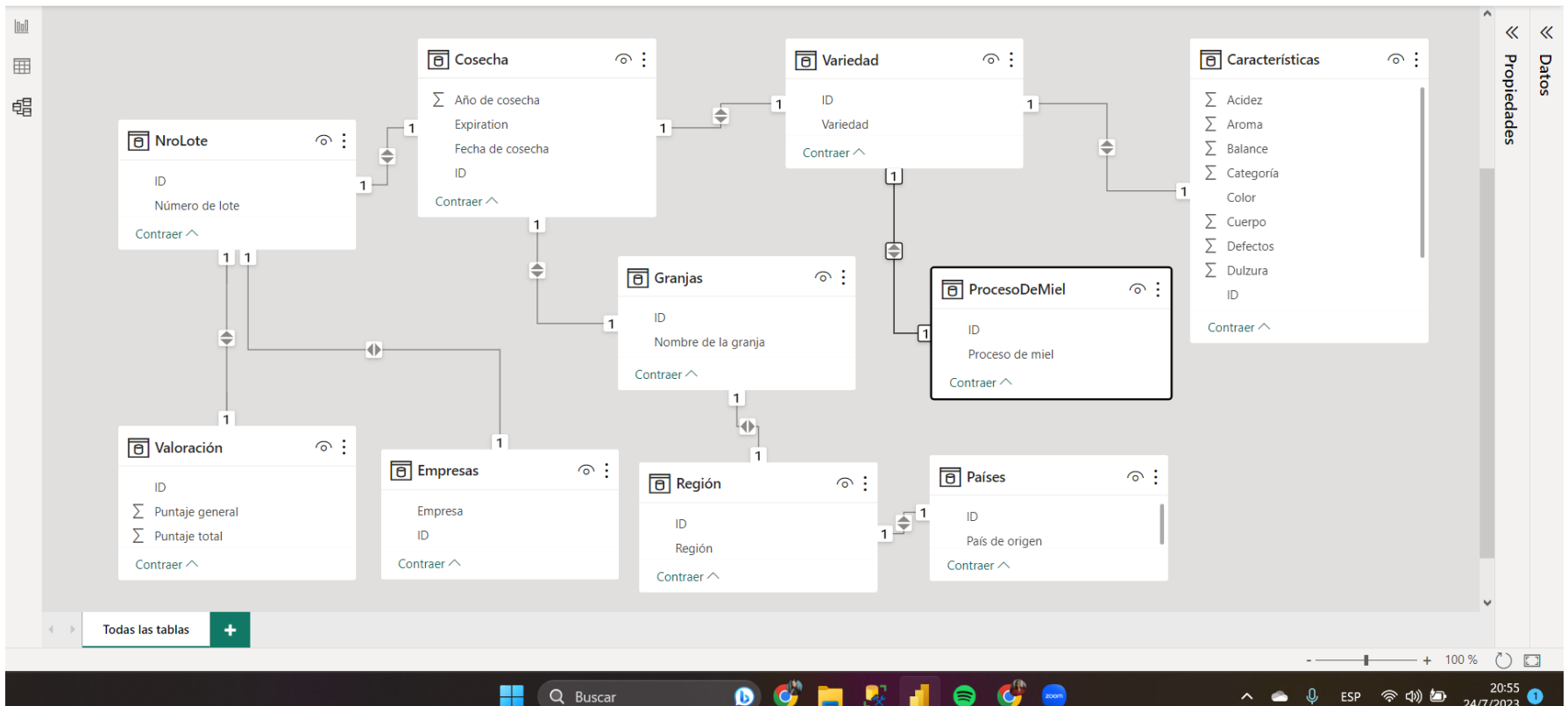


Tabla calendario:

Calendario = CALENDAR(MIN('Cosecha'[Fecha de cosecha]),TODAY())

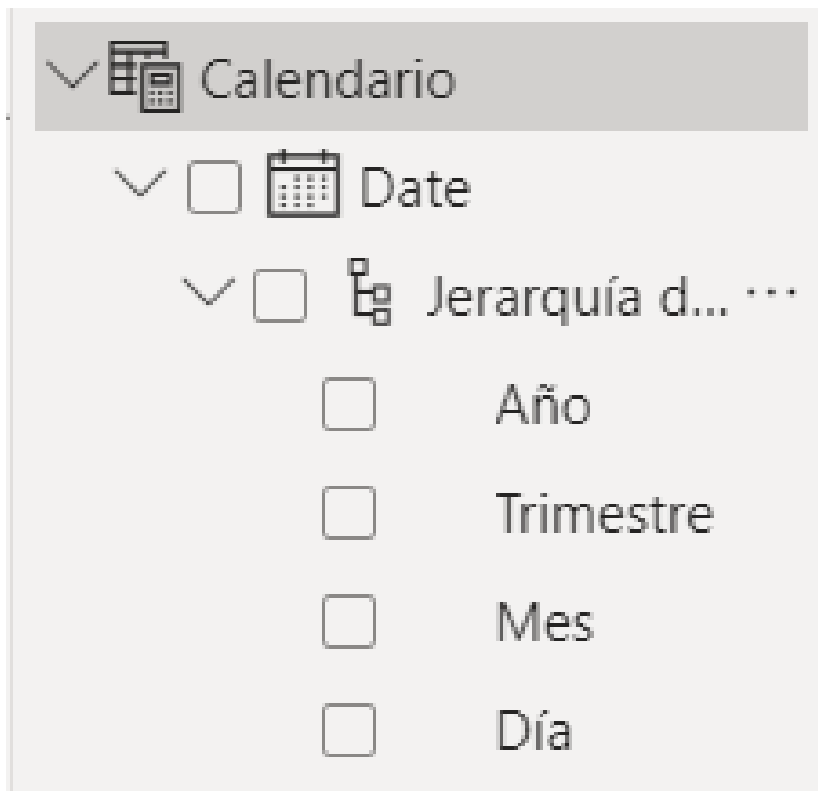
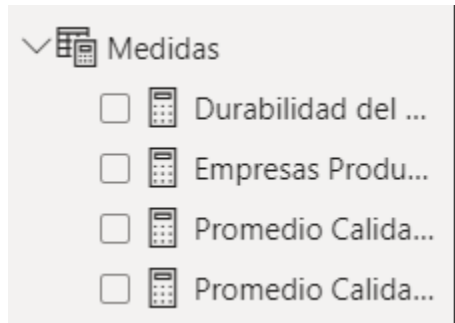


Tabla medidas:



Promedio Calidad Café General = `AVERAGE('Valoración'[Puntaje general])`

Promedio Calidad Café Total =
`CALCULATE(
| DIVIDE(
| SUM('Valoración'[Puntaje total]),
| COUNT('Valoración'[Puntaje total])
|)
)`

Empresas Productoras de Café = `DISTINCTCOUNT('Empresas'[Empresa])` |

Durabilidad Café =
`CALCULATE(
| SUMX('Cosecha',
| DATEDIFF(Cosecha[Fecha de cosecha].[Año], Cosecha[Expiration].[Año],YEAR)))`



Solapas versión inicial - mockup:

Menú: Esta solapa contiene la navegación y las opciones para acceder a diferentes análisis y secciones del informe.

Un viaje por la calidad y variedad del café



Análisis demográfico

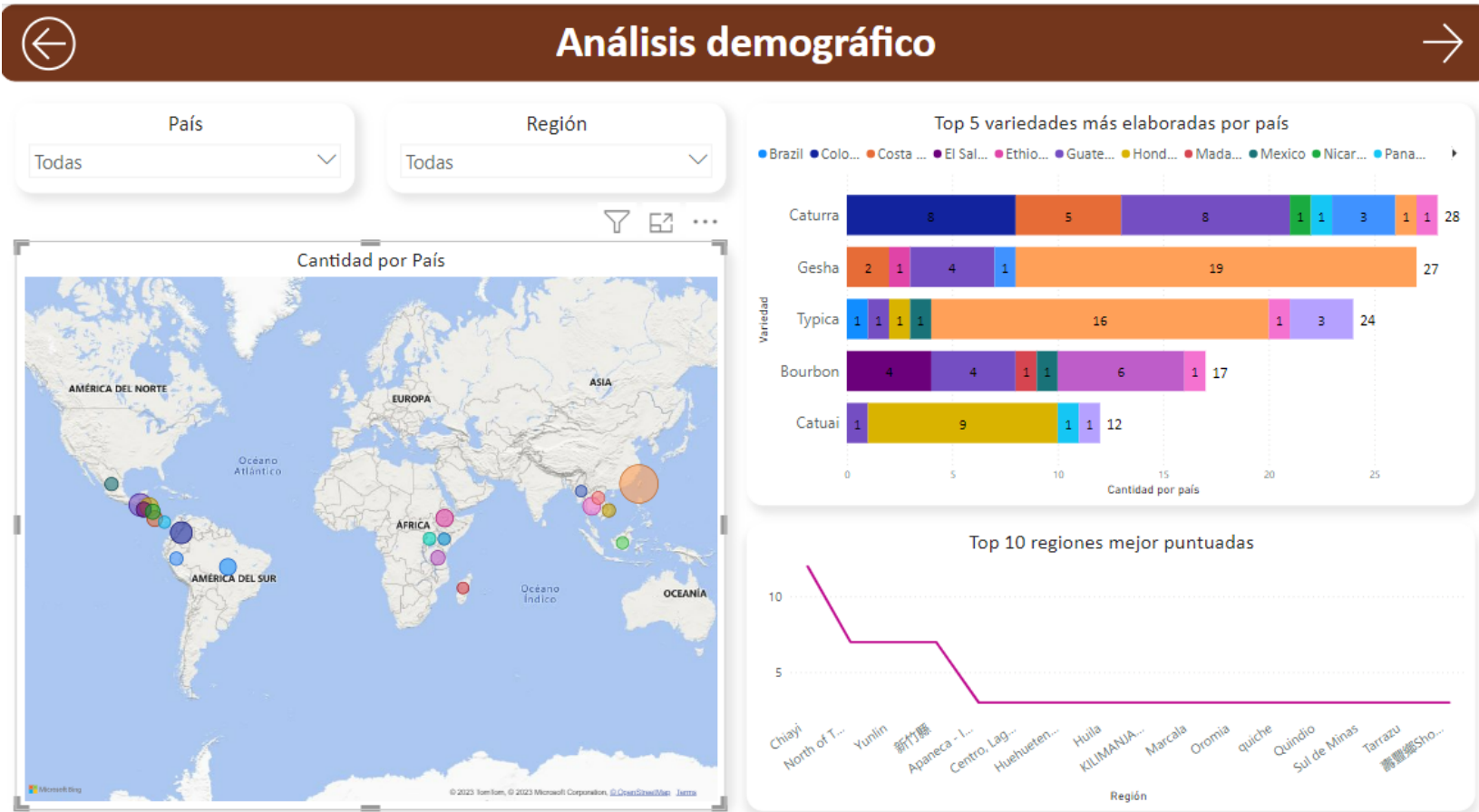


Análisis por variedad

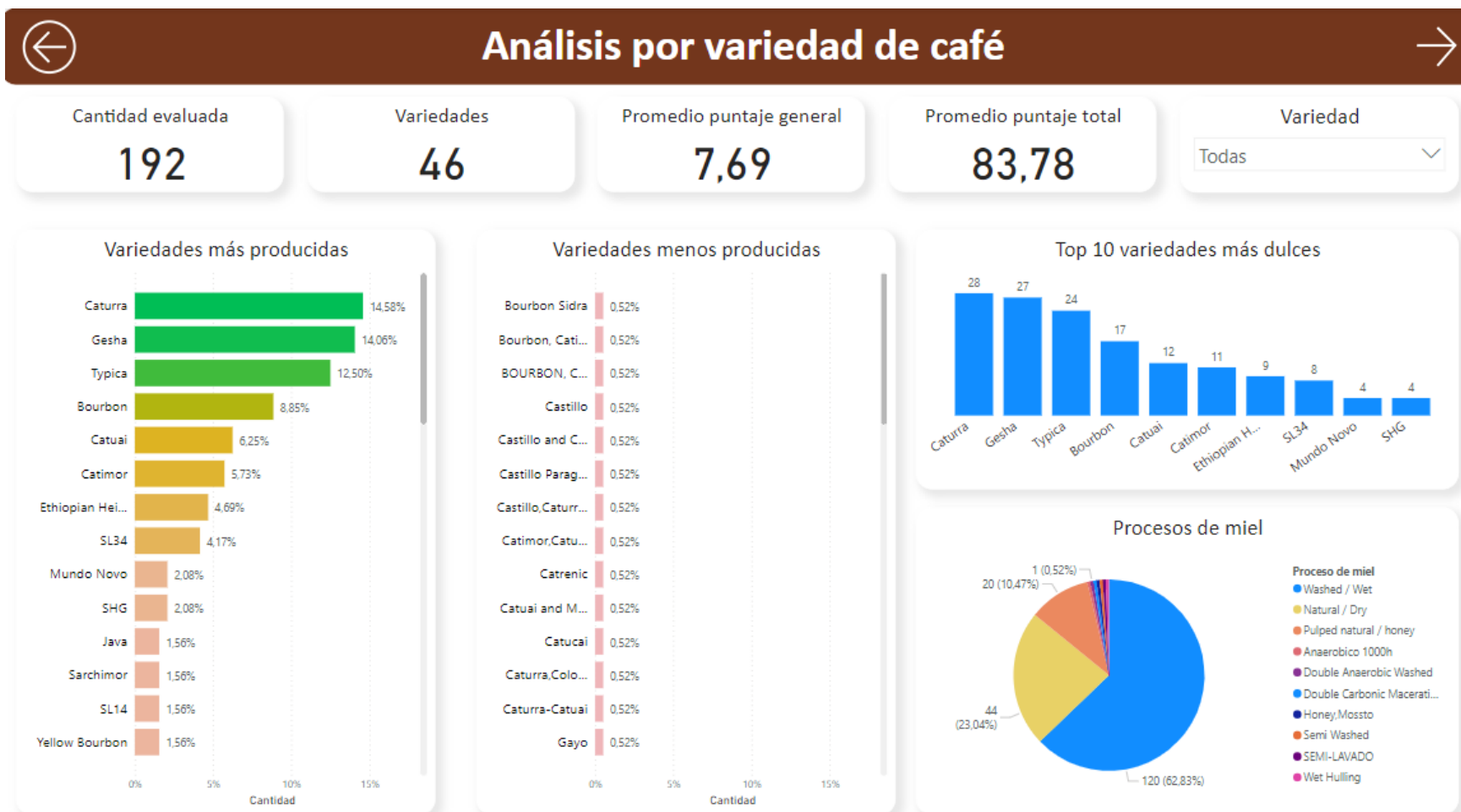
PROYECTO DATA ANALYTICS - ROCÍO BOTTERO



Análisis demográfico: Proporciona un análisis general que abarca países y regiones, ofreciendo una visión amplia de los datos demográficos.



Análisis por variedad: Enfocado en analizar el café en función de sus variedades, esta sección explora distintos aspectos relevantes de cada tipo de café.





Análisis por variedad de café

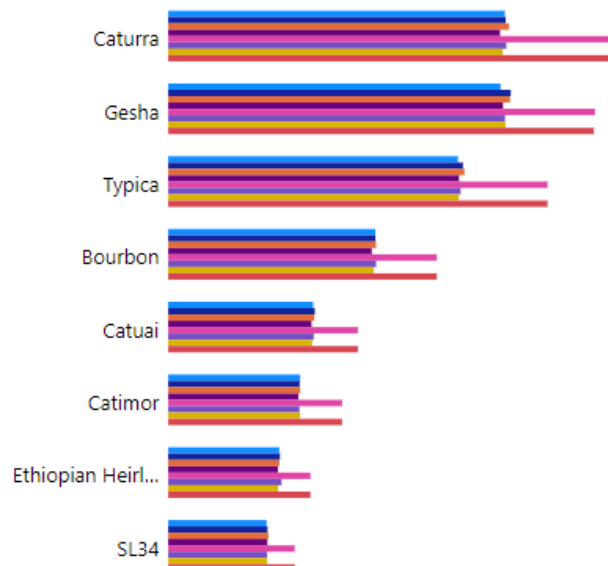
Cantidad evaluada

192

Variedad

Todas

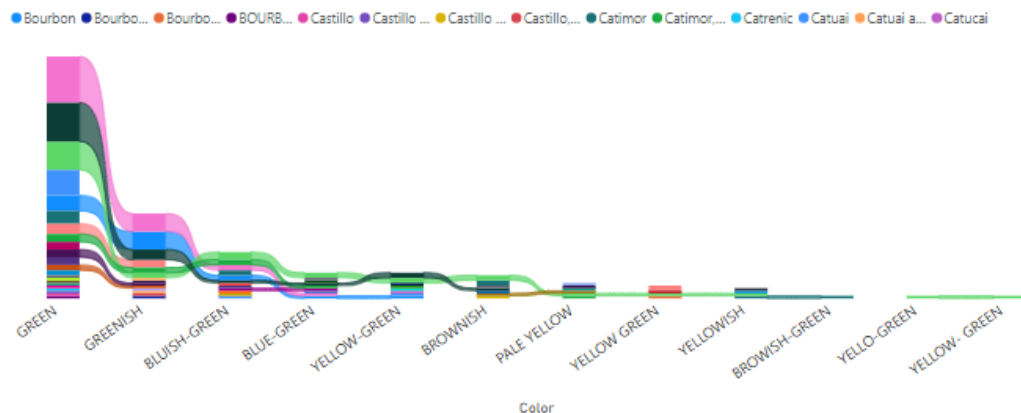
Características por variedad



Cantidad de variedades producidas por Empresa



Variedad de café por color



Solapas versión final:

Menú: La primera solapa de este dashboard funciona como un centro de navegación, facilitando la exploración de las diversas secciones. Además, ofrece acceso directo a la documentación completa del proyecto a través de un botón ubicado en la parte inferior derecha de la pestaña.

Última actualización: 08/08/2023

Un viaje por la calidad y variedad del café



Análisis demográfico

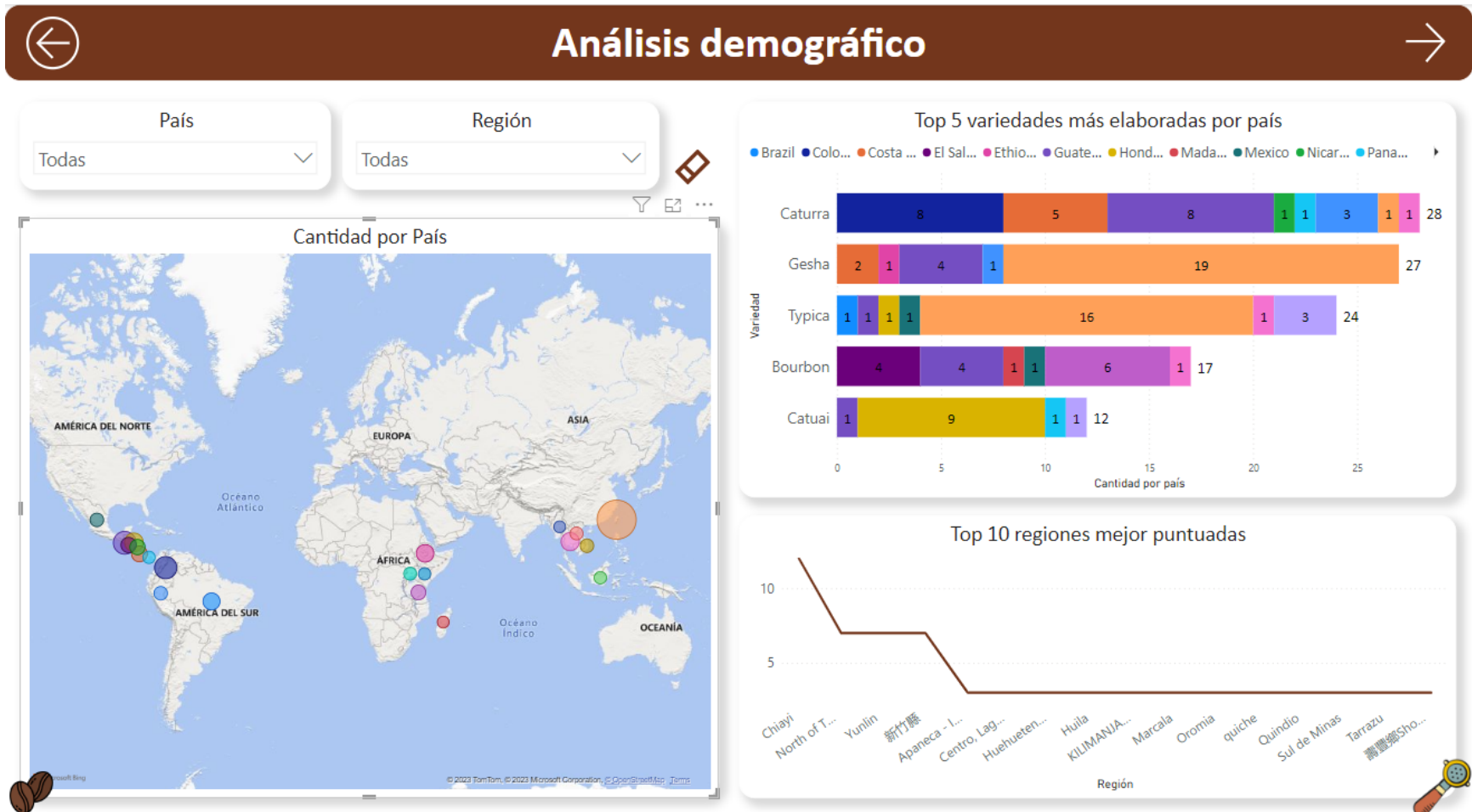


Análisis por variedad

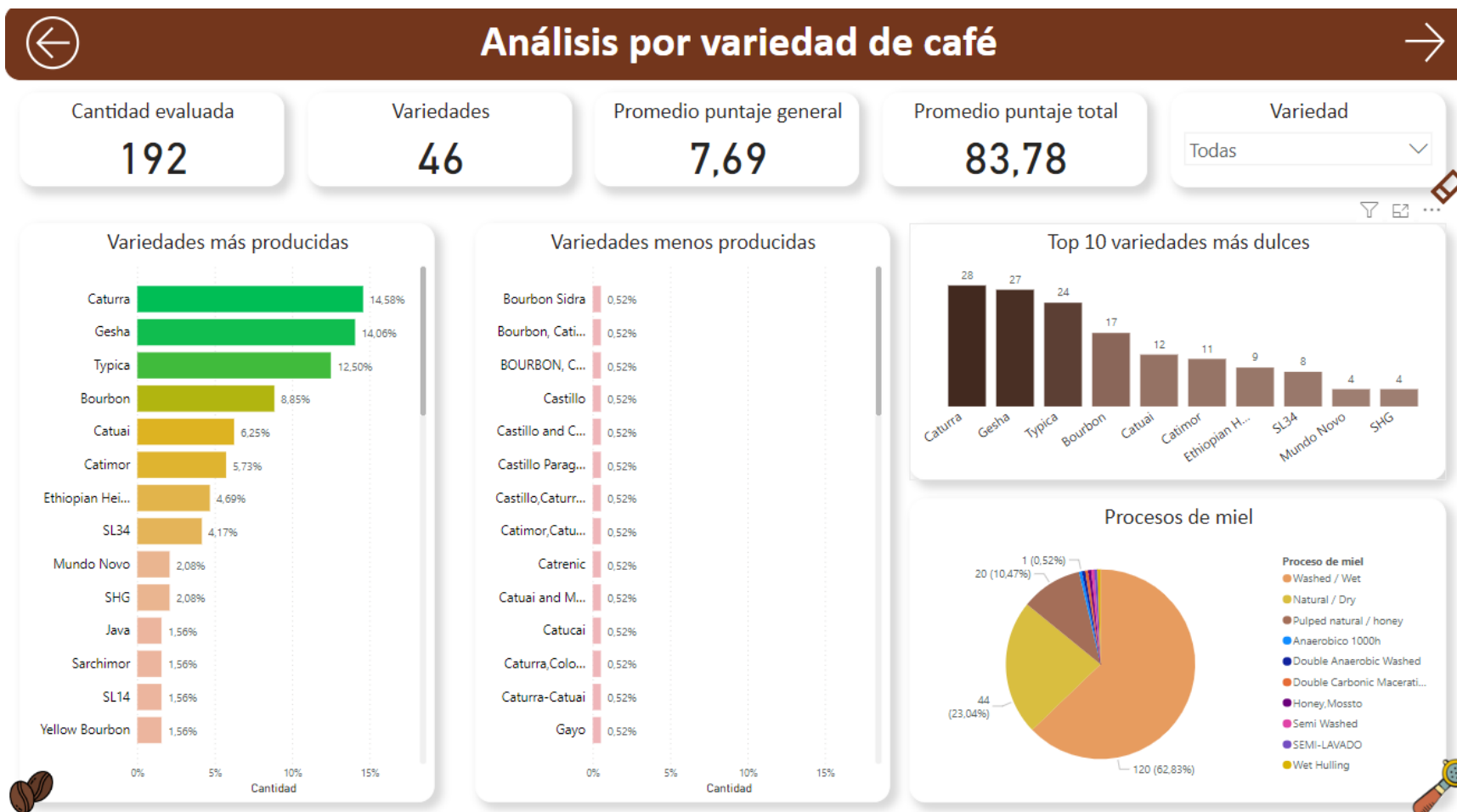
PROYECTO DATA ANALYTICS - ROCÍO BOTTERO



Análisis demográfico: Esta sección permite un análisis detallado de las diversas variedades de café en relación a los países y regiones donde se producen.



Análisis por variedad: Esta sección se centra en un análisis del café según sus distintas variedades, explorando aspectos clave que caracterizan a cada tipo de café.





Análisis por variedad de café



Cantidad evaluada

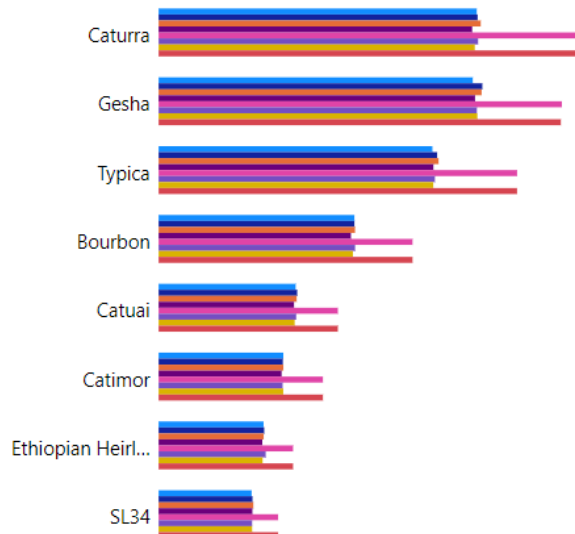
192

Variedad

Todas



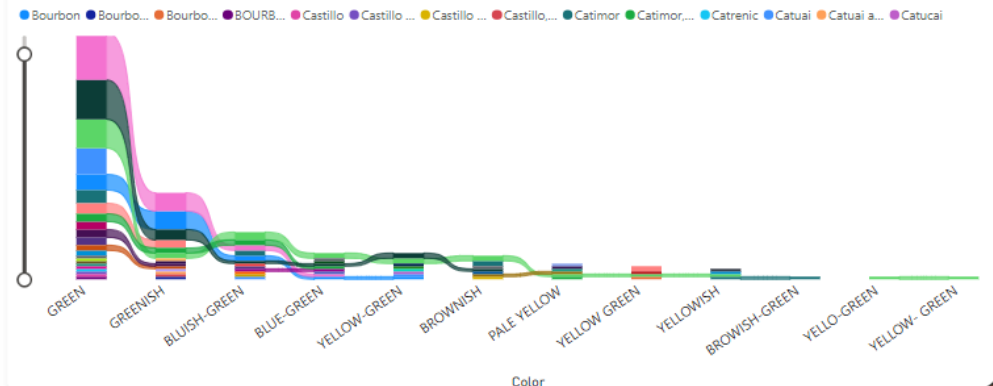
Características por variedad



Cantidad de variedades producidas por Empresa



Variedad de café por color



Navegación entre solapas:

En todas las solapas tenemos botones que permiten manejarnos dentro del dashboard:

- Botón para ir a la pestaña de análisis demográfico



Análisis demográfico

Ir a Análisis demográfico



Análisis demográfico

- Botón para ir a la pestaña nº1 de análisis por variedad



Análisis por variedad

Ir a análisis por variedad



Análisis por variedad



- Botón en pestaña menú que nos permite ir a la documentación completa del proyecto.



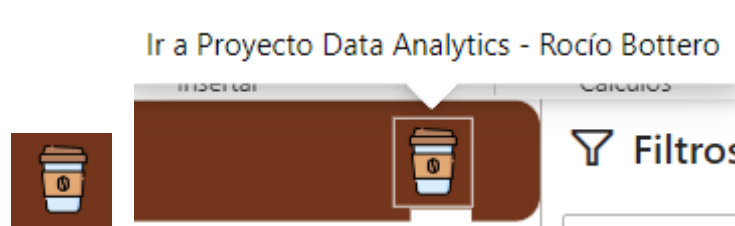
- Botón de navegación para volver al menú.



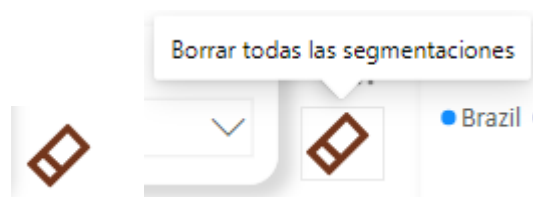
- Botón de navegación para ir a la página siguiente.



- Botón en la pestaña final que nos permite ir a la documentación completa del proyecto.



- Botón para borrar todos los filtros/segmentaciones en la página



Conclusión:

En este proyecto, se ha explorado el gran mundo del café a través de un análisis detallado de su calidad y variedad en distintos países productores. Mediante el examen de parámetros esenciales como aroma, cuerpo, color y sabor, hemos logrado comparar y evaluar naciones destacadas en la producción de esta apreciada bebida.

En el análisis demográfico, hemos podido determinar la distribución de la cantidad de café por país, brindando una imagen clara de los mayores productores. Además, identificamos las cinco variedades más elaboradas en cada país, lo que nos otorga una visión profunda de las preferencias regionales. También, al destacar las diez regiones mejor puntuadas, hemos logrado resaltar los lugares donde la calidad del café es especialmente alta.

En el análisis por variedad, encontramos que la variedad más producida es un punto de referencia crucial en la industria del café, ya que establece tendencias y preferencias globales. Identificamos las variedades más dulces, revelando una dimensión importante del sabor en el mundo del café. La exploración de los procesos de miel más utilizados nos da una comprensión esencial de las técnicas más populares en la producción de café. Evaluar las características por variedad nos proporciona una rica descripción de las sutilezas que distinguen cada tipo de café. Además, cuantificar la cantidad de variedades producidas por empresas nos brinda una visión sobre la diversidad y alcance de las compañías cafetaleras.

En conclusión, estos análisis han respaldado nuestra hipótesis original sobre la variedad y calidad del café en diferentes contextos. A través de un enfoque multifacético, hemos iluminado tanto las preferencias regionales como las características esenciales de cada tipo de café. Este proyecto ha enriquecido nuestra apreciación del café, una bebida que trasciende fronteras y culturas, y que es tan diversa como el mundo que lo produce.

