



Big (data) Experts

Reporte de Producción de Cereales en Argentina

Trabajo Final de Certificación
Juan Pablo Avila

Índice de Contenidos

1. Introducción.....	página 3
2. Desarrollo de las Tablas.....	página 4
2. Desarrollo de las Tablas.....	página 4
3. Transformaciones de las Tablas.....	página 5
4. Diagrama Entidad – Relación.....	página 8
5. Columnas y medidas calculadas.....	página 9
6. Análisis del Dashboard.....	página 12
7. Conclusiones del reporte.....	página 15

Análisis de Datos de Producción de Cereales en la Argentina

1. Introducción

La agricultura en Argentina es una de las bases de su economía. La producción de alimentos agropecuarios es, tradicionalmente, uno de los puntales de las exportaciones argentinas, principalmente la producción de granos (cereales y oleaginosas) y la cadena de la soja en conjunto (porotos, semillas, aceite, pellets alimenticios, harina y biodiésel), uno de los principales encadenamientos productivos del país. Argentina es uno de los líderes en el mercado mundial de granos, aceites y subproductos. La agricultura y ganadería en Argentina son intensivas en capital y en 2018 el sector representaba el 6,14 % del PIB. Hacia julio de 2016, el sector agrario empleaba, junto a la silvicultura, la caza y la pesca, a 337.196 personas, sobre una fuerza laboral de 17,47 millones de personas, lo que representaba a menos del 2% del total. Al 2018, el 8,13% de la población argentina vivía en áreas rurales, uno de los porcentajes más bajos del mundo. El Ministerio de Agricultura, Ganadería y Pesca es la organización del gobierno nacional a cargo de la supervisión de la agricultura.

Las oleaginosas (soja y girasol) llegaron a ser importante ya que su precio internacional aumentó durante el siglo XX. De los aproximadamente 52.000.000 t (cincuenta y dos millones de toneladas) producidas anualmente, alrededor del 92% son soja, y el 7% son semillas de girasol. La superficie total cultivada de semillas oleaginosas es de alrededor de 41.000 km² (cuarenta y un mil kilómetros cuadrados).

El cultivo de las oleaginosas en Argentina se ha destacado desde el siglo XX, cuando el país era el exportador principal de lino (linaza) del mundo. El colapso de ese mercado en la década de 1930 y cualidades del cultivo del suelo, sin embargo, pusieron fin a la posición dominante del país en el sector.

El país es uno de los mayores productores mundiales de semillas de girasol: en 2010, fue el tercer productor mundial, con 2.200.000 t (dos millones doscientas mil toneladas).

Análisis

El objetivo de este análisis y gestión de datos es entender el desarrollo y evolución de las cosechas a lo largo de las últimas décadas en suelo argentino. Particularmente en el espectro que abarca a la avena, el girasol, y el mijo, podremos observar y analizar con detenimiento una comparativa específica de estos cultivos, determinando similitudes y diferencias.

Destinatarios

Si bien este trabajo fue solicitado directamente por el Ministerio de Agricultura, Ganadería y Pesca, quienes estén interesados en obtener un panorama más detallado sobre la actividad de siembra y cosecha de estos productos, puede encontrar información útil para conocer más sobre las áreas de cultivo en las distintas provincias, tendencias de crecimiento y rendimiento de la producción, ya sea a fines informativos o con alguna finalidad de inversión a futuro.

2. Desarrollo de las Tablas

La información con el que se realizó este trabajo fue obtenida desde el portal de Datos del Ministerio de Agricultura, Ganadería y Pesca (<https://datos.magyp.gob.ar/>). En este sitio, se obtuvieron 3 datasets específicos para cada cultivo, los cuales se integraron en una sola grupo de datos. A partir de aquí se desarrollaron las siguientes Tablas:

TABLA PRODUCCION_CEREALES: Tabla principal con la información concreta de las producciones

Columna	Tipo de Dato	Descripción
cultivo_nombre	Texto	Nombre del cultivo
campania	Texto	Campaña de cosecha
provincia_id	Número Entero (FK)	Id de Provincia (Foreign Key de la Tabla Provincias)
departamento_id	Número Entero (FK)	Id de Departamento (Foreign Key de la Tabla Departamentos)
superficie_semrada_ha	Número Entero	Superficie sembrada en Hectáreas
superficie_cosechada_ha	Número Entero	Superficie cosechada en Hectáreas
producción_tm	Número Entero	Producción de cultivo en Toneladas métricas
rendimiento_kgxha	Número Entero	Rendimiento de la producción (Kg. Cosechados por hectárea)
anio_ok	Date (yyyy)	Fecha de la producción (expresada sólo en formato anual)

TABLA PROVINCIAS: Tabla generada para detallar las provincias

Columna	Tipo de Dato	Descripción
provincia_nombre	Texto	Nombre de la provincia
provincia_id	Número Entero (PK)	Id de registro, Primary Key de la tabla

Reporte de Producción de Cereales en Argentina

Juan Pablo Avila

TABLA DEPARTAMENTOS: Tabla generada para detallar los departamentos

Columna	Tipo de Dato	Descripción
departamento_nombre	Texto	Nombre de la provincia
departamento_id	Número Entero (PK)	Id de registro, Primary Key de la tabla
provincia_id	Número Entero (FK)	Id de registro de la provincia a la que pertenece el departamento, Foreign Key de la tabla Provincias

TABLA CALENDARIO: Tabla generada para especificar el rango de tiempo del dataset

Columna	Tipo de Dato	Descripción
Date	Date (Short Date)	Dato de fecha

3. Transformaciones de las Tablas

Antes de traer nuestros datos a Power BI, realizamos algunos cambios previos en Excel (los datos originales se obtuvieron en formato .xlsx). Se editaron los departamentos que no tenían nombre ni código. Se decidió normalizar esos datos con la descripción "Otros Departamentos", y se les asignó un código único para esas filas.

Una vez importados los Datasets a nuestro tablero de Power BI, realizamos las siguientes modificaciones en el editor de Power Query:

✓ Tabla Produccion_cereales

Se combinaron las tablas de origen (datasets de cada cultivo que se obtuvieron por separado).

```
= Table.Combine({avena_serie_1923_2020, girasol_serie_1969_2019, tabla_mijo_1935_2019})
```

Se eliminaron las columnas 'provincia_nombre' y 'departamento_nombre', ya que se genera una tabla específica con esos datos.

```
= Table.RemoveColumns(Source,{"provincia_nombre", "departamento_nombre"})
```

Reporte de Producción de Cereales en Argentina

Juan Pablo Avila

Tomando como ejemplo la columna 'anio', se genera una nueva columna de tipo Date para reemplazarla con el tipo de dato adecuado. Luego se elimina la columna 'anio'.

```
= Table.AddColumn(#"Removed Columns", "anio_ok", each Date.From(Text.From([anio], "es-AR")), type date)
```

```
= Table.RemoveColumns(#"Inserted Date",{"anio"})
```

Finalmente, se aplica un recorte (trim) de los datos de la columna 'campania', porque tenían espacios iniciales.

```
= Table.TransformColumns(#"Removed Columns1",{{"campania", Text.Trim, type text}})
```

```
= Table.SelectRows(#"Texto recortado", each true)
```

✓ Tabla Provincias

Para generar esta tabla, elegimos duplicar la tabla 'Produccion_cereales' y eliminar las columnas innecesarias, dejando sólo las columnas 'provincia_nombre' y 'provincia_id'.

```
= Table.RemoveColumns(Source,{"cultivo_nombre", "anio", "campania", "departamento_nombre", "departamento_id", "superficie_sebrada_ha", "superficie cosechada_ha", "produccion_tm", "rendimiento_kgxa"})
```

Se eliminan las filas duplicadas, para obtener una tabla de datos únicos de provincias.

```
= Table.Distinct(#"Removed Columns", {"provincia_id"})
```

✓ Tabla Departamentos

Procedemos de la misma manera para generar la tabla de Departamentos, elegimos duplicar la tabla 'Produccion_cereales' y eliminar las columnas innecesarias, dejando sólo las columnas 'departamento_nombre', 'departamento_id' y 'provincia_id'.

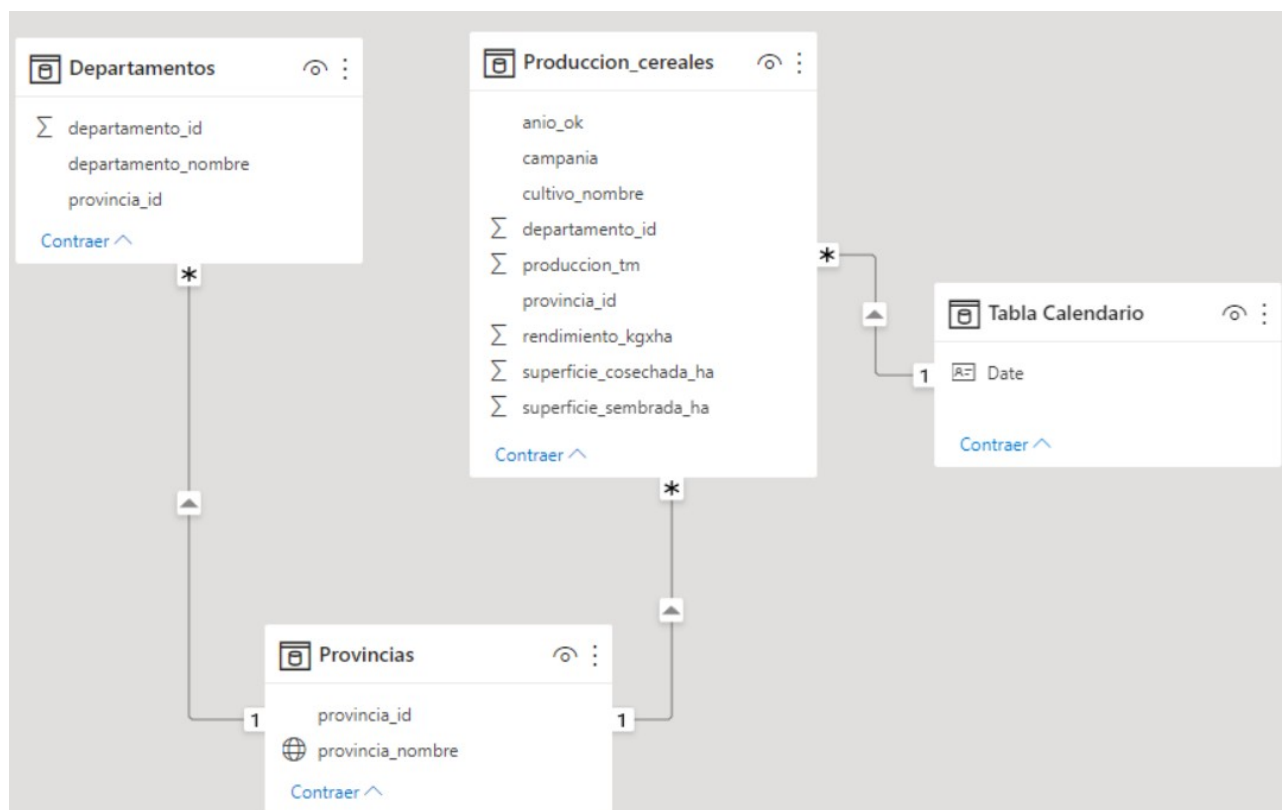
```
= Table.RemoveColumns(Source,{"cultivo_nombre", "anio", "campania", "provincia_nombre", "superficie_semada_ha", "superficie_cosechada_ha",  
"produccion_tm", "rendimiento_kgxha"})
```

Se eliminan las filas duplicadas, para obtener una tabla de datos únicos de departamentos.

```
= Table.Distinct("#Removed Columns", {"departamento_nombre"})
```

4. Diagrama Entidad – Relación

Realizamos las relaciones necesarias entre las tablas, conectando los Id's de provincias, departamentos y tabla calendario a la tabla principal:



5. Columnas y medidas calculadas

Creamos una Tabla Calendario con la siguiente medida calculada:

```
1 Tabla Calendario = // generamos tabla calendario basada en las fechas del dataset
2 CALENDAR(
3     MIN(Produccion_cereales[anio_ok]),
4     MAX(Produccion_cereales[anio_ok])
5 )
```

Esta medida nos entrega un rango de fechas (expresado en años) en base a los datos del dataset.

En base a los requerimiento del trabajo, se generaron las siguientes medidas calculadas (las cuales se encuentran agrupadas en una nueva tabla llamada 'Tabla de Medidas':

Max % y Min %: Estas simples medidas sólo fueron creadas para poder mostrar los valores en porcentaje en la vista de Medidor de Superficie sembrada vs la cosechada. Utilizamos este método ya que trabajamos con datos decimales y no podían visualizarse correctamente.

```
1 Max % = 1      1 Min % = 0
```

Porcentaje de Superficie Sembrada y Cosechada: calcula el porcentaje de superficie que fue cosechada, en base a la superficie sembrada.

```
1 Porcentaje de Superficie Sembrada y Cosechada =
2 // determinamos el % de cosecha Vs siembra
3 DIVIDE(
4     SUM(Produccion_cereales[superficie_cosechada_ha]),
5     SUM(Produccion_cereales[superficie_sembrada_ha]),
6     0
7 )
```

Promedio Rendimiento por Hectárea: calculamos el promedio de rendimiento (Kg. Por Ha.) utilizando la función AVERAGE.

```
1 Promedio Rendimiento por Hectárea =
2 //aplicamos la función AVERAGE para promediar los valores de rendimiento
3 AVERAGE(
4     Produccion_cereales[rendimiento_kgxha]
5 )
```

Reporte de Producción de Cereales en Argentina

Juan Pablo Avila

Promedio Superficie cosechada avena: calculamos el promedio de la tabla, filtrando por el nombre del cultivo.

```
1 Promedio superficie cosechada avena =  
2 //calculamos el promedio filtrando por el nombre del cultivo  
3 CALCULATE(  
4     AVERAGE(Produccion_cereales[superficie_cosechada_ha])  
5     ,Produccion_cereales[cultivo_nombre] == "avena"  
6 )
```

Promedio Superficie cosechada girasol: calculamos el promedio de la tabla, filtrando por el nombre del cultivo.

```
1 Promedio superficie cosechada girasol =  
2 //calculamos el promedio filtrando por el nombre del cultivo  
3 CALCULATE(  
4     AVERAGE(Produccion_cereales[superficie_cosechada_ha])  
5     ,Produccion_cereales[cultivo_nombre] == "girasol"  
6 )
```

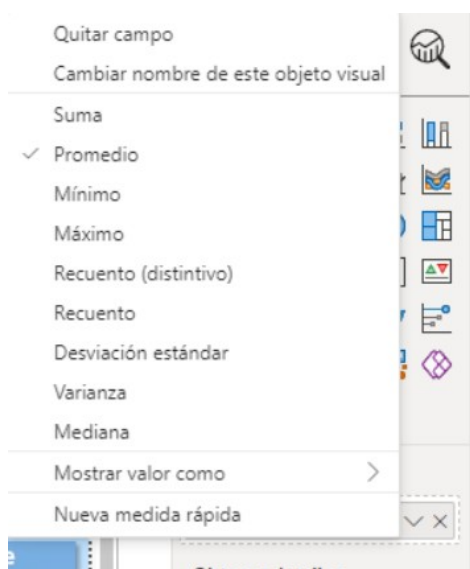
Promedio Superficie cosechada mijo: calculamos el promedio de la tabla, filtrando por el nombre del cultivo.

```
1 Promedio superficie cosechada mijo =  
2 //calculamos el promedio filtrando por el nombre del cultivo  
3 CALCULATE(  
4     AVERAGE(Produccion_cereales[superficie_cosechada_ha])  
5     ,Produccion_cereales[cultivo_nombre] == "mijo"  
6 )
```

Promedio Superficie Sembrada: calculamos la superficie promedio de siembra general.

```
1 Promedio Superficie Sembrada =  
2 // determinamos el promedio de superficie sembrada de los cultivos en Has  
3 AVERAGE(Produccion_cereales[superficie_sembrada_ha])
```

A fines de prueba, el **promedio de superficie cosechada** general no se hizo con una medida calculada, sino que se utilizó la función integrada de Power Bi de mostrar valores promediados de las tablas, directamente en la visualización:



6. Análisis del Dashboard

Al momento de la creación del dashboard se decidió minimizar la cantidad de solapas, y posibilitar el filtrado de los datos de cada una de ellas por cultivo (a través de botones interactivos).

Este dashboard tiene 3 solapas:

- **Portada**

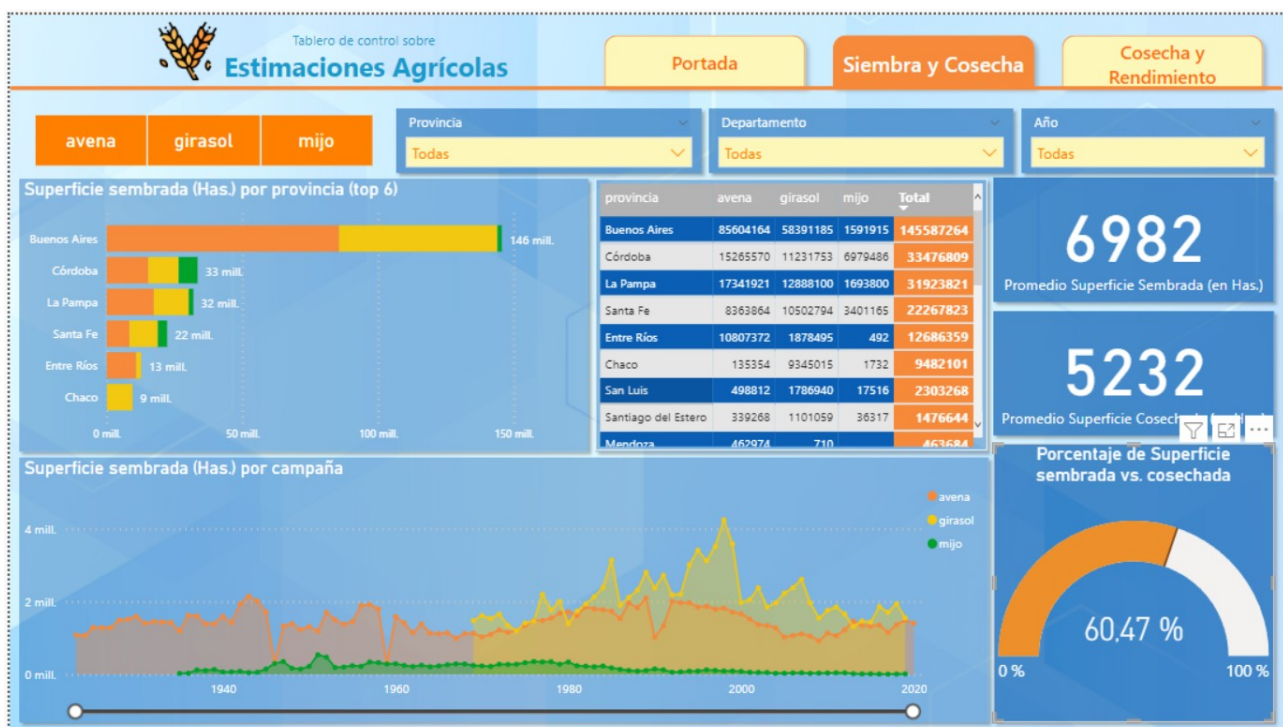
La portada cuenta con los títulos descriptivos del trabajo realizado, y los controles para poder navegar en las distintas solapas del Tablero.



- *Siembra y Cosecha*

Esta solapa se enfoca principalmente en los datos de Siembra de los 3 cultivos analizados, y en la comparativa con los datos de cosecha de los mismos. Indicamos promedios de superficie, comparativas porcentuales y gráficas de análisis geográfico y temporal. También se generaron los segmentadores para: Tipo de cultivo, Provincia (y departamento), y Campaña.

Al momento de realizar la gráfica de **superficies sembradas por campaña**, se encontraron muchas inconsistencias en la columna "campania", por lo que se decidió utilizar la columna de año para mostrar información más consistente.



- *Cosecha y Rendimiento*

En esta solapa se hace foco principalmente sobre los datos de cosecha y rendimientos anuales de los cultivos. Se especifican datos individuales de cada uno de ellos y un Top 10 de las provincias con mayores cosechas. Además se generaron los segmentadores para: Tipo de cultivo, Provincia (y departamento), y Campaña.



Reporte de Producción de Cereales en Argentina

Juan Pablo Avila

7. Conclusiones del reporte

En base al análisis de los datos gestionados y expuestos en el tablero, podemos llegar a las siguientes conclusiones:

- ✓ Las provincias que agrupan la mayor producción de estos cultivos son: Buenos Aires, La Pampa, Córdoba y Santa Fe, marcando una gran brecha con respecto al resto de las provincias. Esto tiene lógica ya que éstas áreas son las más propicias para el desarrollo de dicha actividad.
- ✓ Si bien la provincia de Chaco ingresa en el top 6 en cuanto a su superficie sembrada, notamos que sólo concentra la producción de Girasol en gran escala a partir del año 1969, y prácticamente es el único de estos cultivos registrado de 1994 en adelante.
- ✓ La producción de mijo tuvo su auge durante las décadas de los '50, hasta casi los '90. A partir de ahí se denota una baja muy grande en su producción nacional (promediando las 10,000 Has. sembradas en los últimos 5 años).
- ✓ El cultivo de avena es el único que tiene un registro constante en el tiempo (abarcando desde la década del '20 hasta la actualidad), sobre todo contrastando con los datos del girasol. Queda pendiente de un posterior análisis determinar si la producción de girasol no comenzó hasta fines de los '60, o estamos en falta de los registros más antiguos.
- ✓ Analizando los porcentajes de la superficie cosechada en contraste con la sembrada, vemos una gran diferencia entre los 3 cultivos, donde la avena cosecha cerca del 34% de lo sembrado, el mijo aumenta ese porcentaje a más del 60%, y finalmente el girasol tiene el mayor rendimiento, llegando a un promedio de 95%.
- ✓ Los rendimientos anuales de la avena (expresados en Kg. por Ha.) están en crecimiento constante desde la década del '70 aproximadamente. En cuanto al girasol, su crecimiento ha sido sostenido durante todo el registro obtenido. En ambos casos podemos observar datos de años "excepcionales" donde hubo grandes caídas en su rendimiento, probablemente ligadas a condiciones climáticas de esos años.