



Escola
Técnica
Superior
de Enxeñaría



Visualización Avanzada

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

PRÁCTICA 3.4: PRODUCCIÓN DE VINO EN CALIFORNIA

■ Hugo Vázquez Docampo

FECHA DE ENTREGA: 17 de octubre de 2022

ÍNDICE

1. Introducción	2
2. Motivación, preguntas a resolver y origen de los datos	2
A. Fuente de datos	2
3. Diccionario de datos	2
4. Preparación y limpieza de los datos	3
5. Elaboración del dashboard 1	4
6. Elaboración del dashboard 2	4
7. Análisis de los resultados observados	5
8. Imágenes de los Dashboards resultantes	6
9. Conclusiones	6

1. INTRODUCCIÓN

Se incluye en este informe un análisis de un conjunto de datos de libre elección sobre la **producción de vino en California** a lo largo de los años. En él, se incluye un análisis de las cuestiones que se desean resolver; así como una motivación para la selección de este conjunto de datos. Además, se incorpora una breve descripción de los datos mostrados. A mayores, se introduce un apartado en el cual se relatan los pasos a seguir para realizar la limpieza de los datos. A continuación, se lleva a cabo el análisis de los mismos mediante la realización de dos Dashboards. Por último, se finaliza con una conclusión de los resultados observados.

2. MOTIVACIÓN, PREGUNTAS A RESOLVER Y ORIGEN DE LOS DATOS

El vino de California es uno de los más importantes a nivel mundial. Representa casi el 90 por ciento de toda la producción de vino estadounidense. Por si sola, es el cuarto mayor productor de vino del mundo.

Por ello, se busca mostrar la gran importancia de este Estado en el mundo de la viticultura respondiendo a las siguientes cuestiones:

1. ¿Cómo de importante es para el estado de California el cultivo de uvas para la producción de vino?
2. ¿De cuánto territorio dispone este país para llevar a cabo esta actividad?
3. ¿Qué beneficios aporta a la economía del territorio?
4. ¿Cuáles son los condados con mayor influencia en el sector?
5. ¿Cómo ha ido evolucionando esta tradición a lo largo de los años?

A. Fuente de datos

Los datos utilizados para la realización de este análisis se han obtenido de las siguientes fuentes: California Wine Production de Kaggle y California State Association of Counties. El primero de los conjuntos de datos ha sido el principal precursor del análisis a realizar dado que los campos que incluye otorga un amplio abanico de posibilidades para el estudio.

Por otra parte, es de interés incluir el segundo de los Datasets dado que con el se pueden analizar variables de interés tales como la superficie total del estado. Estos datos recolectados servirán como contraste para el estudio llevado a cabo.

3. DICCIONARIO DE DATOS

Se recoge a continuación una tabla con el diccionario de datos asociado a las columnas de los conjuntos de datos utilizados para el análisis.

Campo	Descripción	Tipo de dato
Year	Año de producción	Numérico
CountyCode	Código de condado	Numérico
County	Nombre del condado	Texto
HarvestedAcres	Hectáreas cosechadas	Numérico
Yield	Rendimiento de la cosecha (toneladas/hectárea)	Numérico
Production	Producción en toneladas	Numérico
Price	Precio en dólares por tonelada	Moneda
Unit	Unidades (toneladas)	Numérico
Value	Valor en dólares de la ganancia total	Moneda
TotalAcres	Hectáreas totales del condado (calculadas como millas ² * 259)	Numérico

4. PREPARACIÓN Y LIMPIEZA DE LOS DATOS

Para comenzar, se importan los datos a Excel desde los datasets que se van a emplear. Esto se lleva a cabo utilizando la opción de datos>obtener datos>de un archivo>y escogiendo csv y libro de excel en función del datasets que se incorpora.

En segundo lugar, y antes de realizar la limpieza, se opta por combinar ambos conjuntos de datos. Para ello, se hace uso de la función **BuscarV**. Analizando los campos comunes a ambos datasets, observamos como es posible realizar la unificación aprovechándose del campo **county**. Se opta por utilizar esta opción en lugar de cualquier otra estudiada en la asignatura por varios motivos. El hecho de usar esta función viene dado porque solo se desea emplear una de las columnas del segundo conjunto de datos. Por ello, se emplea el siguiente comando:

=BUSCARV([@County];'County Government'!A:I;6;1).

Se incluye un **1 (true)** al final para filtrar por coincidencias no exactas. Esta decisión surge a raíz de analizar los valores de **county** para ambas tablas. En la hoja de **California_Wine_Production_1980_**, las ubicaciones formadas por dos palabras como Contra Costa aparecen unidas sin espacio (ContraCosta), mientras que en la hoja de **County Government** se escriben por separado. De este modo, la columna tomada de la otra hoja es la asociada a **SquareMiles**, con la que posteriormente se obtendrán las hectáreas totales.

A continuación, se procede a llevar a cabo la preparación de los datos para la elaboración de los Dashboards seleccionados. Para ello, se han desarrollado las siguientes acciones:

- **Campo unit:** la mayoría de los datos se encuentran medidos en toneladas (**tons**). Sin embargo, existen discrepancias en la forma de escribirlas (observables en las opciones de filtrado de la columna de unidades) y, por tanto, se opta por unificar el formato. Para ello, se selecciona la columna y se hace uso de la herramienta buscar y reemplazar, sustituyendo TON por Tons.
En relación con el campo de unidades, se encuentra una única fila con otra magnitud (acre). Como realizando una búsqueda en internet no se han encontrado equivalencias entre las unidades, se ha optado por eliminar dicha fila de datos.
- **Campos nulos:** a continuación, se opta por realizar un análisis de aquellas columnas con valores nulos. Así, después de examinar los datos, se observa como para los condados de San Mateo, Trinidad y Nevada son varios los campos nulos. Tras tratar de inferir o calcular los valores que faltan, se deduce que no existe forma de obtenerlos y se escoge eliminar las entradas asociadas a estos territorios.

Por último, para poder comenzar a elaborar las tablas dinámicas y las gráficas necesarias, se procede a incorporar los datos al **modelo de datos** desde el menú de Power Pivot, en la opción de agregar al modelo de datos.

A continuación, ya es posible comenzar a elaborar los Dashboards creando la primera de las tablas dinámicas desde en menú de Insertar >Tabla dinámica y escogiendo una nueva hoja de cálculo.

5. ELABORACIÓN DEL DASHBOARD 1

- **Gráfico de tarjeta para la producción total y las hectáreas cultivadas:** para llevar a cabo estos gráficos se han creado nuevas tablas dinámicas; colocando los campos **Production** y **HarvestedAcres** respectivamente en el sumatorio de valores. Para poder pasarlo al Dashboard sin errores, se ha realizado una referencia a las celdas de la tabla dinámica en la propia hoja donde estas se contruyeron y, a continuación, se ha creado una referencia a esta última en la página asociada al Dashboard. Por último se le ha aplicado el formato deseado al gráfico.
- **Gráfico de tabla para la producción y la ganancia totales por condados:** debido al gran número de condados existentes, para una mejor visualización, se ha decidido aplicar un filtro de valor seleccionando así las cinco mejores entradas en función de la producción total. Una vez realizado esto, se procede a ordenar los datos de mayor a menos según este mismo criterio. A continuación, sobre el campo de producción total, se escoge la opción de mostrar valores como porcentaje del total general. Para concluir con el gráfico, se aplica un formato condicional de barra de datos y escala de color a las columnas de producción y ganancia respectivamente.
- **Gráfico de columnas para el porcentaje de territorio cultivado por condado:** para elaborar esta gráfica se genera una tabla dinámica con los valores de superficie total, cultivada y los condados. Una vez realizado este paso, se crea una tabla auxiliar, haciendo referencia a las celdas de dicha tabla. Después, se le añadirá una nueva columna relativa a dividir la superficie total entre la aprovechada para el cultivo y pasar este valor a porcentaje. Así, con el nombre de los condados y el valor calculado, es posible elaborar el gráfico deseado.
- **Gráfico de segmentación de datos para los condados:** para realizar este último gráfico, se pulsa sobre una de las tablas dinámicas elaboradas para este primer Dashboard. Desde la opción de Analizar Tabla Dinámica, se escoge Insertar Segmentación de datos. Entonces se seleccionará el campo buscado y, tras darle formato, una vez pulsado sobre la segmentación creada, se selecciona la opción de Conexiones de Informes, para enlazar así el gráfico de segmentación de datos con las tablas dinámicas requeridas.

6. ELABORACIÓN DEL DASHBOARD 2

- **Gráficos de tarjeta para la producción total y el total de ingresos:** para desarrollar estos gráficos se siguen los pasos explicados en la sección anterior.
- **Gráfico de barras de rendimiento por condado:** para elaborar esta gráfica, se crea una nueva tabla dinámica a partir de la anterior, incluyendo esta vez los campos de condados y rendimiento. Tras haberla generado, se ordenan los datos de forma ascendente en función del rendimiento. Por último, se genera la figura y se le aplica el formato deseado.
- **Gráfico de segmentación de datos para los años:** para realizar este último gráfico, se siguen los pasos recogidos en la sección anterior.
- **Gráfico de tabla para la producción y el precio de los 10 últimos años:** a la hora de elaborar este gráfico se sigue el mismo proceso que para generar el asociado al Dashboard

anterior; filtrando esta vez por los 10 últimos años y sin aplicar ningún criterio de ordenación.

- **Gráfico combinado para la producción y el precio de los 10 últimos años:** haciendo uso del gráfico de tabla anterior, se procede a elaborar este nuevo gráfico; aplicándole el formato escogido.

7. ANÁLISIS DE LOS RESULTADOS OBSERVADOS

Para justificar los resultados observados, se utiliza como punto de partida las preguntas planteadas al inicio del informe. Así, empleando como material de apoyo los dashboard elaborados, se recogen los siguientes resultados:

1. ¿Cómo de importante es para el estado de California el cultivo de uvas para la producción de vino?

Para dar respuesta a esta cuestión se parte de la información presente en los dos Dashboards elaborados. En el primero de ellos, se observa cómo la producción y el cultivo de la vid es una tarea común en la mayoría de condados. Los territorios en los que se gestiona esta actividad ocupan un porcentaje considerable de estos territorios; llegando a sobrepasar el 20 % en alguno. Además, se observa como la producción y las ganancias totales a lo largo de los años alcanzan un monto a tener en cuenta en el mundo de la vinicultura; destacando algunas zonas como Fresno o San Joaquín.

Por otra parte, atendiendo a la información recogida en el segundo de los dashboard, se observa cómo, con el paso de los años, la producción y las ganancias aumenta a medida que pasa el tiempo; sugiriendo que esta actividad se ha ido mejorando con los años y que ha ganado importancia en la sociedad.

2. ¿De cuánto territorio dispone este país para llevar a cabo esta actividad?

En el primero de los Dashboards, es posible ver las hectáreas cultivadas, tanto de forma global, como filtrando por condados; llegando a ocupar una parte a tener en cuenta del territorio californiano. Además, se incluye otra información de interés cómo el porcentaje de territorio dedicado a este cultivo según el condado.

3. ¿Qué beneficios aporta a la economía del territorio?

En el segundo dashboard es posible visualizar las ganancias totales y por condado que esta práctica ha proporcionado. Como se observa, el cultivo de la uva proporciona unos ingresos considerables. Estos montos han ido aumentando con los años por motivos que, como se puede deducir a partir de gráfica de producción y precio por año, viene movida por el aumento de estos dos valores.

4. ¿Cuáles son los condados con mayor influencia en el sector?

A nivel económico, se observa en el Dashboard 1 como destacan los condados de Fresno, San Joaquín o Madera por su alta producción y ganancias. No obstante, en el dashboard 2 se observa como a nivel de rendimiento destacan los condados de Tulare o Kings por encima del resto.

5. ¿Cómo ha ido evolucionando esta tradición a lo largo de los años?

En líneas generales, se aprecia una tendencia creciente a nivel de producción y coste, visible en el Dashboard 2.

8. IMÁGENES DE LOS DASHBOARDS RESULTANTES

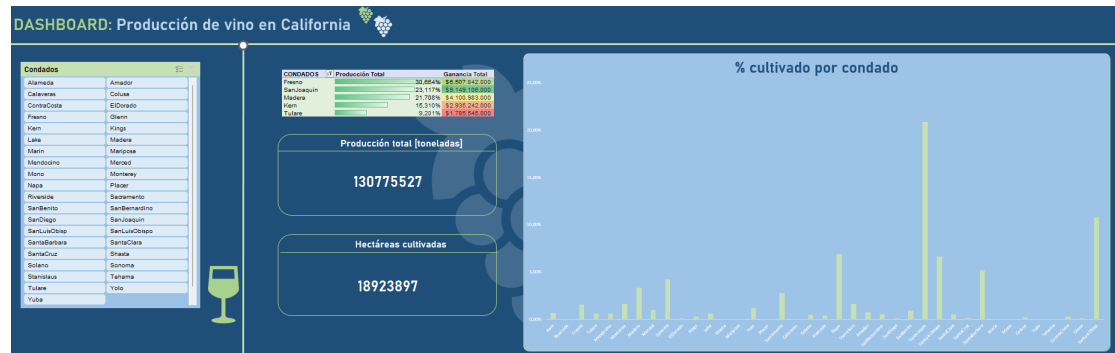


Figura 1: Dashboard 1

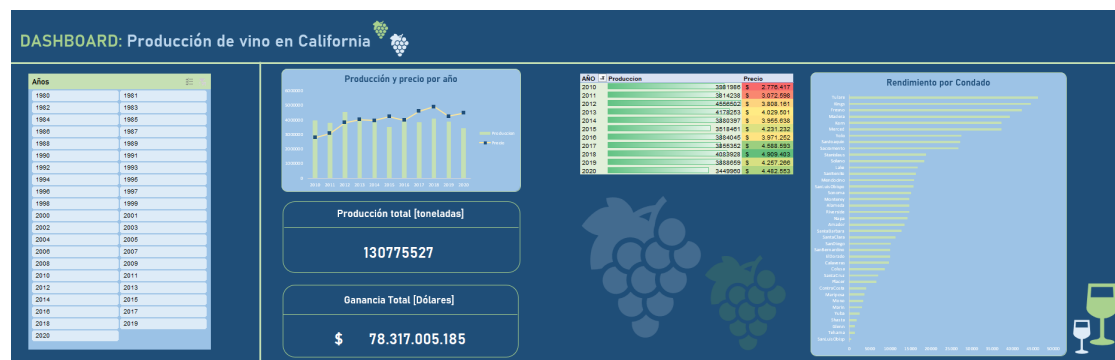


Figura 2: Dashboard 2

9. CONCLUSIONES

Atendiendo a los resultados de los análisis realizados, se observa cómo el cultivo de la uva es una práctica importante en el mercado laboral californiano que, pese a no ser su principal actividad económica, aporta beneficios considerables a este estado. Además, tal y como se mencionó con anterioridad, es posible concluir que cada vez es mayor la producción y el precio de venta; aumentando cada vez más los aportes de esta disciplina.

REFERENCIAS

- [1] https://es.wikipedia.org/wiki/Vino_de_California
- [2] <https://www.counties.org/data-and-research>
- [3] <https://www.kaggle.com/datasets/jarredpriester/california-wine-production-19802020>