



CRISIS ALIMENTARIA



— PROYECTO INDIVIDUAL

Del 04/08/200 al 30/08/200

Elsa Toribio Herrero

Git hub: @ElsaTH

THE BRIDGE
Data Science

VISIÓN

GENERAL

Entregar un delivery forma parte del Bootcamp como Data Science. Dicho delivery tiene como objetivo practicar sobre los conocimientos adquiridos en la primera parte de este Bootcamp.

El tema a tratar es la gran crisis económica que ESPAÑA sufrió en 2008 y cómo afectó este hecho a la producción de alimentos a nivel nacional. Tomo como referencia los datos oficiales de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO). Es una agencia especializada de las Naciones Unidas que lidera los esfuerzos internacionales para vencer el hambre.

Su objetivo es lograr la seguridad alimentaria para todos y asegurar de que las personas tengan acceso regular a suficientes alimentos de alta calidad para llevar una vida activa y saludable. Con más de 194 estados miembros, la FAO trabaja en más de 130 países en todo el mundo.

METAS

El objetivo de este proyecto es poner en práctica conceptos que he adquirido durante la primera parte del Bootcamp sobre la minería de datos:

- Cargar el conjuntos de datos.
- Procesarlos utilizando diferentes bibliotecas de Python
- Visualizar de forma gráfica el resultado de los análisis realizados.

Todo ello para obtener la conclusión sobre mi hipótesis.

Mi meta sería poder realizar hasta la opción A incluida, e intentar realizar algún apartado de la opción A+.

ESPECIFICACIONES

Para hacer los objetivos y hacer la entrega lo mejor posible, vamos a especificar todo lo necesario para hacerlo posible:

SOFTWARE

- Visual Studio Code versión 1.48.2
- Python 3.8.3 64 bits
- Google Slide
- Google doc
- Photoshop

HARDWARE

- MacBook Pro (15 pulgadas, mediados 2009)
- Procesador 2,8 GHz Intel Core 2 Duo
- Memoria 8 GB 1067 MHz DDR3
- Gráficos NVIDIA GeForce 9400M 256 MB

REQUISITOS

Librerías importadas e instaladas para el proyecto:

- | | |
|--------------|------------|
| • Pandas | • requests |
| • Numpy | • Json |
| • Seaborn | • os |
| • Matplotlib | • sys |
| • Plotly | • Flask |



PASOS DEL PROYECTO

CRISIS ALIMENTARIA

PASOS DEL PROYECTO

1. INVESTIGAR EL CONTEXTO Y DEFINIR LA HIPÓTESIS DEL PROYECTO

La investigación la realicé sobre la FAO. La Organización de las Naciones Unidas para la Agricultura y la Alimentación.

He definido las siguientes hipótesis:

La crisis económica de España en 2008 tuvo un impacto negativo en la producción de alimentos a nivel nacional.

Durante este proyecto analizaré los datos para ver si la hipótesis es o no correcta..

2. OBTENER DATOS

Para trabajar en el proyecto lo primero que hemos de conseguir son los datos.

Parto inicialmente de un dataset obtenido del enlace de kaggle: <https://www.kaggle.com/roshansharma/world-food-production/data>

Para seguir comparando y contrastando información, sigo con la obtención de datos.

Después de buscar opciones comprobé que la página oficial de la FAO.org era la mejor alternativa debido a que muchas otras eran de pago. <http://www.fao.org/faostat/en/#data>

PASOS DEL PROYECTO

3. LIMPIEZA DE DATOS

La Eliminación y limpieza de datos del resto de países quedándome sólo con los datos de España.

CREACIÓN DE VARIOS DATAFRAME:

- **df**: es el obtenido del dataset de kaggle
- **df2**: producción de cultivos en España
- **df3**: Relación entre la producción y los precios por materias primas
- **df4**: Relación entre producción y precios por agrupación de cultivo, agricultura, cereales y comida

Utilizo diferentes dataframes para poder demostrar mi hipótesis. Son datos que no puedo unificar en un sólo, ya que contienen información que analiza datos sobre el tema a tratar, agrupándolos de diferente manera y llamando a las materias primas con nombres distintos.

DATA

CLEANING:

Para todos ellos aplicó la misma práctica:

- Comprobar los tipos de columna y eliminar las que no son necesarias.
- Comprobar los valores de NaN y eliminarlos.
- Comprobando si hay duplicados en los datos y eliminarlos.

PASOS DEL PROYECTO

4. VISUALIZACIÓN DE DATOS

Para poder analizar los datos he utilizado la visualización de datos con tablas, gráficos.

1. Analicé de manera global todos los años para ver la evolución que había en España de la producción de materias primas.
2. Analicé cinco años antes de la crisis y cinco años después de la crisis la evolución de producción.
3. Comprobando cuáles eran los años claves que influían en la decisión de demostrar mi hipótesis decidí obtener las gráficas para realizar la comparativa. Esos años fueron 2007 un año antes de la crisis, 2008 el año donde España empezó la crisis y 2012, el año donde España tocó fondo y empezó su recuperación.

5. CONCLUSIONES

Después de analizar los datos y visualizarlos, he escrito la conclusión de mi hipótesis.

De una manera muy clara y analizando los datos con diferentes dataset y en agrupaciones por cultivo, agricultura, comida y cereales, comprobamos que hay un descenso del precio en cuanto a las materias de producción. Y que claramente demostrado que en 2008 no es que hubiera una crisis, sino que empezó una gran crisis que no terminaría hasta cuatro años después en 2012.

+6.5%

Trends

More info

78%

Shares

More info

95%

Performance

More

PREGUNTAS SOBRE EL PROYECTO

CRISIS ALIMENTARIA

PREGUNTAS SOBRE EL PROYECTO

OPCIÓN C

Como parte de los pasos obligatorios del proyecto de la opción C, responderé a las siguientes preguntas:

- **¿Fue posible demostrar la hipótesis? ¿Por qué?**
Sí, con la visualización y el análisis de los datos pude responder la hipótesis con una conclusión.
- **¿Qué puede concluir sobre su estudio de datos?**
Los cereales son los mayores perjudicados en esta crisis tanto a nivel de producción como a nivel de precio.

Que España estuvo sumergida en una crisis durante 5 años.

Que la recuperación de la crisis comenzaría a ver su fin a lo largo de 2013.

- **¿Qué cambiaría si tuviera que hacer otro proyecto de EDA?**
Optimización de tiempos y mejora en todos los pasos del procesos con una supervisión guiada por alguien con mayor experiencia, debido a la falta de conocimiento y experiencia en la materia.

Los tiempos se han demorado y el resultado final no ha sido 100% el que esperaba.

- **¿Qué aprendiste haciendo este proyecto?**
Durante este proyecto puse en práctica los diferentes pasos de un proyecto EDA, trabajando con las diferentes herramientas que nos han enseñado tanto nuestros docentes : Gabriel Vázquez Torres y Clara Piniella Martinez durante esta primera parte del Bootcamp de Data Science en The Bridge.

PREGUNTAS SOBRE EL PROYECTO

OPCIÓN C

Otra de las cosas que también he conseguido, es enfrentarme sola a todas aquellas pequeñas trabas que van saliendo mientras realizas un proyecto. Algunas me han llevado demasiado tiempo y otras las he conseguido resolver de una manera rápida y eficaz.

He practicado mucho con este proyecto no sólo a buscar en google información, también con el tema de visualización de datos. No conformándome cómo salían las gráficas sino modificando todos los estilos de visualización (colores, formas, tamaños...) para que en la presentación del proyecto todo tuviese un mismo estilo, jerarquía y armonía.

He realizado varios gráfico dinámicos y no todos los que deseaba, porque era invertir demasiado tiempo en algo que no podría integrar en la presentación y mostrarlos en su esencia(o al menos no he encontrado la manera).

OPCIÓN B

- **¿Cuáles son las columnas con mayor correlación?**
El mapa de color es una forma visual muy útil para para conocer las variables y sus relaciones. A primera vista llama mucho la atención que los años elegidos tienen una correlación que roza el 1. Parece haber una correlación significativa; en realidad es tan fuerte que podría indicar multicolinealidad, es decir, que básicamente ofrecen la misma información.

PREGUNTAS SOBRE EL PROYECTO

OPCIÓN A

- Investigación para guardar cada parcela en archivos locales

Cómo guardar una parcela en un archivo PNG

Puede guardar fácilmente una figura, por ejemplo, en un archivo png haciendo uso de plt.savefig(). El único argumento que debe pasar a esta función es el nombre del archivo

Ejemplo:

```
# Save Figure
plt.savefig("foo.png")

# Save Transparent Figure
plt.savefig("foo.png", transparent=True)
```

Con solo ejecutar esta línea de código, guardará el gráfico que ha realizado en un archivo de imagen en lugar de mostrarlo.

Cómo guardar una parcela en un archivo PDF

Si desea guardar múltiples trazados en un archivo pdf, desea hacer uso del backend pdf, que puede importar fácilmente:

Ejemplo:

```
# Import PdfPages
from matplotlib.backends.backend_pdf import PdfPages

# Initialize the pdf file
pp = PdfPages('multipage.pdf')
```

PREGUNTAS SOBRE EL PROYECTO

OPCIÓN A

```
# Save the figure to the file
pp.savefig()

# Close the file
pp.close()
```

Cómo guardar una parcela en un archivo html

Si desea guardar múltiples trazados en formato html, el ejemplo muestra los pasos a seguir:

Ejemplo:

```
import plotly
# create a simple plot
bar = plotly.graph_objs.Bar(x=['giraffes', 'orangutans',
                              'monkeys'],
                           y=[20, 14, 23])
layout = plotly.graph_objs.Layout()
fig = plotly.graph_objs.Figure([bar], layout)
# convert it to JSON
fig_json = fig.to_json()
# a simple HTML template
template = """<html>
<head>
  <script
src="https://cdn.plot.ly/plotly-latest.min.js"></script>
</head>
<body>
  <div id='divPlotly'></div>
  <script>
    var plotly_data = {}
    Plotly.react('divPlotly', plotly_data.data,
plotly_data.layout);
  </script>
</body>
</html>"""
```


PREGUNTAS SOBRE EL PROYECTO

OPCIÓN A

```
# write the JSON to the HTML template
with open('new_plot.html', 'w') as f:
    f.write(template.format(fig_json))
```

Todas mis gráficas están guardadas como png en archivos locales en la carpeta de resources.

- **¿Hay valores atípicos o algunos datos raros?**

Comprobación de outliers de la producción de la agricultura, los cereales, el cultivo y la comida que se ha producido en España en los años:

- 2007 un año antes de la crisis
- 2008 año en la cual empezó la crisis en España
- 2012 año donde España tocó fondo y empezó a recuperar

Se observa cómo la producción del conjunto de materias primas ha ido bajando desde que empezó la crisis y que en el año 2007 y 2008 hay casos atípicos que tienen características diferentes de las demás. Este tipo de casos no pueden ser caracterizados categóricamente como benéficos o problemáticos sino que deben ser contemplados en el contexto del análisis y debe evaluarse el tipo de información que pueden proporcionar. Sin embargo en 2012 el outliers es más cercano, estando por debajo de la media y no lo consideraría tan atípico a simple vista.

PREGUNTAS SOBRE EL PROYECTO

OPCIÓN A

Comprobamos outliers en el precio-producción en España en los años críticos;

- 2007 un año antes de la crisis
- 2008 año en la cual empezó la crisis en España
- 2012 año donde España tocó fondo y empezó a recuperar

En esta gráfica observamos que la media de los datos es más homogénea en cuanto a las materias producidas frente al precio y que en los tres años se producen outliers por encima de la media.

CONCLUSIÓN SOBRE LA REALIZACIÓN DE MI PROYECTO:

He tenido algo de dificultad a la hora de enfocar mi proyecto debido a la inexperiencia. Y también a la hora de decidir cómo demostrar mi hipótesis, ya que había tantos datos que era difícil centrarte en unos pocos. No obstante me gustaría seguir trabajando en el proyecto y mejorando todo aquello que no ha quedado bien definido y aumentado dicho proyecto con los nuevos conocimientos que vaya a adquirir en esta segunda etapa del Bootcamp.

A hand holding a globe is the background image. Overlaid on the globe is a white graphic consisting of a vertical line with five horizontal bars of increasing length, resembling a staircase or a stylized letter 'I'. Below this graphic is the word 'GRACIAS' in a bold, white, sans-serif font. The exclamation mark at the beginning and the tilde at the end of the word are replaced by the same stylized graphic.

!GRACIAS!

Elsa Toribio Herrero The Bridge Acelerator | Data Science