

# Programa de Especialización en Econometría Aplicada

## Centro de Formación Continua

### Modelos Variable conteo

### Clase 3

Edinson Tolentino  
MSc Economics  
email: [edinson.tolentino@gmail.com](mailto:edinson.tolentino@gmail.com)

Twitter: @edutoleraymondi

Universidad Nacional de Ingeniería



- 1 Caso Estudio
  - Data
  - Variables
  - Manejo de data y variables
  - Respondiendo Preguntas

## Analisis de caso



Se debe plantear preguntas que deberán analizarse con información, **no se puede responder sobre la base de tener información**

Las preguntas:

- **¿Qué equipos son los que presentan mayor cantidad de goles durante un torneo?**
- **¿Qué factores estan asociados con el numero de goles de un equipo que desempeña de manera local?**

Donde comenzar, la respuesta esta orientado a buscar la mejor información que me permita responder estas dos preguntas planteadas

La respuesta debe estar orientada en usar los conocimientos estadísticos vistos durante la clase

# Analisis de caso: Data



## Analisis de equipos de la Premier League (England)

Pagina de la Premier League para descargar información:

**Football-Data.co.uk**  
*Results / Odds / Tipsters*

**bet365** Apostar en fútbol

[Regístrate](#)  
 Juega de forma responsable 18+

Network Sites ▼

Updated: 15/06/23
 [BeGambleAware](#)
[Follow](#)

[Home](#)
[Free Bets](#)
[Livescores](#)
[Books on Betting](#)
[Casino](#)
[Poker](#)
[Tennis](#)
[Contact](#)
[Like](#)

**bet365** Apostar en fútbol

[Regístrate](#)  
 Juega de forma responsable 18+

**WORLD'S FAVOURITE BOOKMAKER**  
**bet365**

**OTHERS**  
**BoyleSports**  
**William Hill**  
**Betfred**  
**Betway**

**bet365**  
 Descargar información

**Data Files: England**  
 Last updated: 28/05/23

Registering with any of the advertised bookmakers on [Football-Data](#) will help keep access to the historical results & betting odds data files FREE.  
  
 Below you will find download links to all available CSV data files to use for quantitative testing of betting systems in spreadsheet applications like Excel. League tables, head2head statistics and information on goalscorers, first scorers and top scorers can now be accessed through the [Livescore](#) service. Latest betting odds are available through the [Odds Comparison](#).  
  
 You are free experiment with the data yourselves, but if you are looking for a bespoke Excel application that has been designed specifically to work with Football-Data's files, visit [BetGPS](#) for an exceptional data analysis workbook. Like all of Football-Data's files, it free to download.

**Notes.txt**  
 (text file key to the data files and data source acknowledgements)

**Season 2022/2023**  
[Premier League](#) (FT & HT results; match stats; match, total goals & AH odds)  
[Championship](#) (FT & HT results; match stats; match, total goals & AH odds)  
[League 1](#) (FT & HT results; match stats; match, total goals & AH odds)  
[League 2](#) (FT & HT results; match stats; match, total goals & AH odds)  
[Conference](#) (FT & HT results; match, total goals & AH odds)

**SITE RESOURCES**  
[Historical Data](#)  
[Learn to Bet](#)  
[Free Bets](#)  
[Books](#)  
[Other Sites](#)

**BETTING SYSTEMS**  
[Football Ratings](#)  
[Wisdom of Crowds](#)  
[Contrarian Betting](#)  
[Pinnacle Odds Drop](#)  
**NEW**

**BETTING ARTICLES**  
[Football-Data](#)  
[Pinnacle Sportsbook](#)

**BET CALCULATORS**

## Analisis de caso: variables



Lo primero es poder analizar el diccionario de variables que se tiene en la base de datos (aqui algunas variables)

Variables	Descripcion
FTHG and HG	Full Time Home Team Goals
HS	Home Team Shots
AS	Away Team Shots
HY	Home Team Yellow Cards
AY	Away Team Yellow Cards

## Analisis de caso: carga de datos



Es importe poder cargar la información de excel, csv, etc a un programa que permita realizar un analisis de manera sistematica, rapida y sencilla los analisis

- Python al ser un software de codigo abierto, permite trabajar información sobre la información
- Librias, es importante instalarlas para poder trabajar en python

```
: # Librerias para manipulacion de datos
import pandas # manipulacion de datos
import os # sistema operativo
import numpy # manipulacion de variables
import sympy # operacion math

# Librerias de Graficos
import seaborn
import matplotlib.pyplot as plt

# librerias de distribucion estadistica
from scipy.stats import norm # distribucion normal
```

Carga de datos de un archivo csv hacia python  
, usando la libreria **pandas.read**

```
data = pandas.read_csv("E0.csv")
```

# Análisis de caso: manipulación de datos



```
[6]: # Observo cuantas filas y columna tiene la data
data.shape
```

```
[6]: (380, 106)
```

```
[5]: # Mostrar Los primeros cinco datos
data.head().T
```

```
[5]:
```

	0	1	2	3	4
Div	E0	E0	E0	E0	E0
Date	13/08/2021	14/08/2021	14/08/2021	14/08/2021	14/08/2021
Time	20:00	12:30	15:00	15:00	15:00
HomeTeam	Brentford	Man United	Burnley	Chelsea	Everton
AwayTeam	Arsenal	Leeds	Brighton	Crystal Palace	Southampton
...	...	...	...	...	...
PCAHA	2.13	1.77	2.14	1.81	1.88
MaxCAHH	2.05	2.19	1.82	2.16	2.08
MaxCAHA	2.17	1.93	2.19	1.93	1.9
AvgCAHH	1.8	2.1	1.79	2.06	2.03
AvgCAHA	2.09	1.79	2.12	1.82	1.86

106 rows × 5 columns

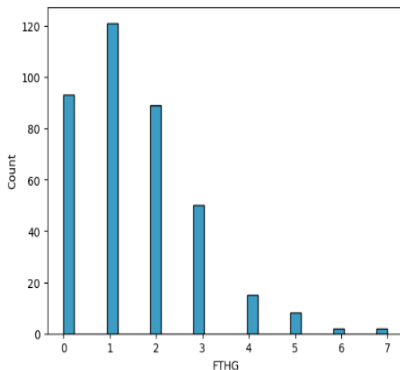
- Manipulación de variables, utilizando principales comandos
- El comando **shape** te muestra las dimensiones de la data

# Analisis de caso: manipulacion de datos



```
[9]: # Grafico de histograma
seaborn.histplot(data, x='FTHG', bins=30)

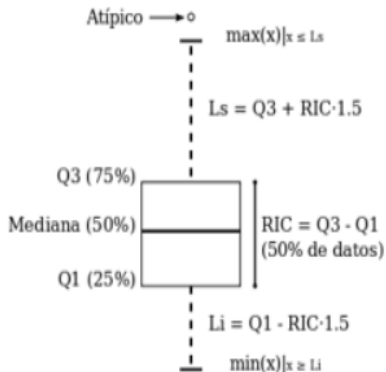
[9]: <AxesSubplot: xlabel='FTHG', ylabel='Count'>
```



- La principal variable a utilizar es FTHG, el cual luego de revisar la definición de la misma permite entender que mide el numero de goles que ha realizado un equipo local (home)
- El gráfico es un grafico de histograma de los valores de dicha variable (se pudo mejorar considerando un grafico de barras)
- **¿Qué puede concluir?**



# Analisis de caso: manipulacion de datos



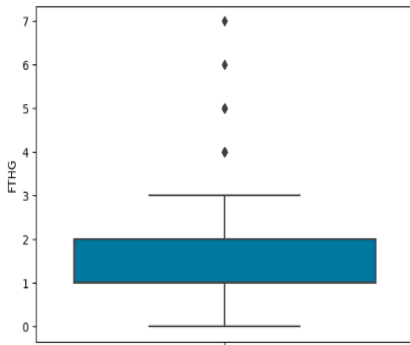
- Siguiendo con el analisis de nuestra principal variable, se tiene el grafico de caja, el cual presenta como indica el gráfico un analisis de percentiles , promedio y mediana. Asimismo datos atipicos que puede tener la variable en cuestion

# Analisis de caso: manipulacion de datos



```
# Grafico de Caja  
seaborn.boxplot(data, y='FTHG')
```

```
<AxesSubplot: ylabel='FTHG'>
```



- La principal variable a utilizar es FTHG, el cual luego de revisar la definición de la misma permite entender que mide el numero de goles que ha realizado un equipo local (home)
- El gráfico es un grafico de caja de los valores de dicha variable
- **¿Qué puede concluir?**

# Analisis de caso: pregunta 1



```
[11]: # ¿Quienes son esos outliers?
      cuantil = data['FTHG'].quantile(0.95)
      data_outliers = data[data['FTHG'] > cuantil]

      print()
      print("Quienes son los quipos con mas goles (mayor de 95 %)")
      print()
      print(data_outliers['HomeTeam'].unique())
```

Quienes son los quipos con mas goles (mayor de 95 %)

['Man United' 'Man City' 'Chelsea' 'Liverpool' 'Tottenham' 'Arsenal']

- La principal variable a utilizar es FTHG, el cual luego de revisar la definición de la misma permite entender que mide el numero de goles que ha realizado un equipo local (home)
- La pregunta es poder analizar en detalle el numero de goles que ha realizado cada equipo local (home team) pero si se realiza en excel se tendria la respuesta de manera tardia
- Para poder responder la pregunta, se observo previamente el grafico de caja, y se concluye que los equipos que tienen mayores goles , son datos atipicos



## Análisis de caso: pregunta II

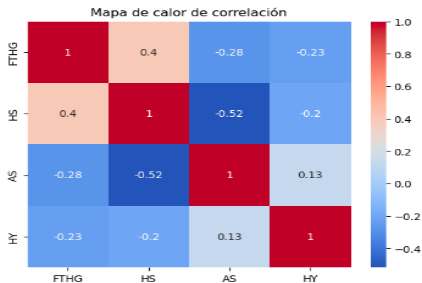


¿Qué factores están asociados con el número de goles de un equipo que desempeña de manera local?

```
# Correlación (mapa)
# Calcular la matriz de correlación
correlation_matrix = data[base_variables].corr()

# Crear el mapa de calor de correlación
seaborn.heatmap(correlation_matrix, annot=True, cmap='coolwarm')

# Mostrar el gráfico
plt.title('Mapa de calor de correlación')
plt.show()
```



- Se observa un análisis de asociación o correlación de las variables con la variables FTHG

- FTHG and HG : Full Time Home Team Goals
- HS : Home Team Shots
- AS : Away Team Shots
- HY : Home Team Yellow Cards