

DESARROLLO DEL TRABAJO PRÁCTICO ESTADÍSTICO

LUCAS DELGADO

SEPTEMBER 8, 2023

1 Coleccion de datos

En este trabajo, se realizaron filtrados y análisis de datos de partidos de fútbol correspondientes al año 2022 de la primera división del fútbol argentino. Este proceso se llevó a cabo en un entorno de Python utilizando Jupyter Notebook, lo que permitió la manipulación y visualización de los datos de manera eficiente. La colección de datos original abarcaba el período de 2015 a 2022, pero se focalizó en el año 2022 para obtener información actualizada y específica.

El objetivo principal de este análisis fue calcular el promedio de goles por equipo en los partidos jugados durante el año 2022. Para lograrlo, se utilizaron técnicas y herramientas de programación en Python, lo que facilitó la extracción, filtrado y procesamiento de los datos correspondientes a ese año. Esto aseguró que los resultados obtenidos fueran relevantes y precisos en relación con el período de interés.

2 Descripcion de la muestra

Contiene la información de 2821 partidos de primera división del fútbol argentino agrupando datos de promiedos.com.ar, transfermarkt.de and oddsportal.com en el archivo "afa.2015.2022.spa.csv"

- torneo: nombre del torneo en curso cuando se jugó el partido. promiedos
- fecha: en qué fecha se jugó el partido. promiedos.
- partido: número de partido dentro de la fecha. promiedos.
- equipo_local(visitante): nombre del equipo local(visitante) promiedos.
- goles_local(visitante): número de goles anotados por el equipo local(visitante). promiedos.
- goles_visitante(visitante): porcentaje de posesión del equipo local(visitante). promiedos.
- resultado: resultado del encuentro.
- fecha_encuentro: fecha del encuentro. oddsportal.

Equipo	Goles	Partidos Jugados	Promedio Goles por Partido
Aldosivi	16	27	0.59
Argentinos	30	27	1.11
Arsenal	28	27	1.04
Atl Tucuman	32	27	1.19
Banfield	23	27	0.85
Barracas Central	30	27	1.11
Boca Juniors	34	27	1.26
Central Cba (SdE)	34	27	1.26
Colon	24	27	0.89
Def y Justicia	29	27	1.07
Estudiantes (LP)	28	27	1.04
Gimnasia (LP)	26	27	0.96
Godoy Cruz	25	27	0.93
Huracan	35	27	1.3
Independiente	31	27	1.15
Lanus	22	27	0.81
Newells	26	27	0.96
Patronato	31	27	1.15
Platense	23	27	0.85
Racing Club	41	27	1.52
River Plate	43	27	1.59
Rosario Central	24	27	0.89
San Lorenzo	33	27	1.22
Sarmiento (J)	27	27	1.0
Talleres (C)	28	27	1.04
Tigre	41	27	1.52
Union	28	27	1.04
Velez	30	27	1.11

Figure 1: Tabla de promedio de goles.

Equipo	Goles	Partidos Jugados	Promedio Goles por Partido
Aldosivi	16	27	0.59
Lanus	22	27	0.81
Banfield	23	27	0.85
Platense	23	27	0.85
Rosario Central	24	27	0.89
Colon	24	27	0.89
Godoy Cruz	25	27	0.93
Newells	26	27	0.96
Gimnasia (LP)	26	27	0.96
Sarmiento (J)	27	27	1.0
Talleres (C)	28	27	1.04
Arsenal	28	27	1.04
Estudiantes (LP)	28	27	1.04
Union	28	27	1.04
Def y Justicia	29	27	1.07
Velez	30	27	1.11
Barracas Central	30	27	1.11
Argentinos	30	27	1.11
Patronato	31	27	1.15
Independiente	31	27	1.15
Atl Tucuman	32	27	1.19
San Lorenzo	33	27	1.22
Central Cba (SdE)	34	27	1.26
Boca Juniors	34	27	1.26
Huracan	35	27	1.3
Racing Club	41	27	1.52
Tigre	41	27	1.52
River Plate	43	27	1.59

Figure 2: Tabla de promedio de goles orden ascendente.

3 Grafico de caja

X_{min} : Valor mínimo dentro de los bigotes = 0.59

X_{max} : Valor máximo dentro de los bigotes = 1.59

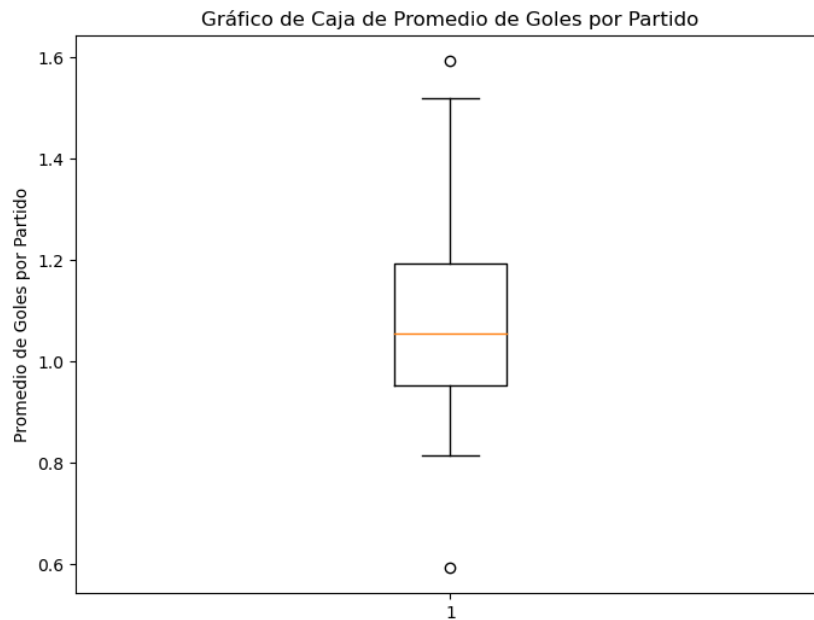
```
In [31]: # Datos para los cuales deseas calcular los cuartiles
data = goles_por_equipo['Promedio Goles por Partido']

# Calcula los cuartiles
q1 = np.percentile(data, 25)
q2 = np.percentile(data, 50) # Este es el cuartil 2 (mediana)
q3 = np.percentile(data, 75)

print("Q1:", q1)
print("Q2 (Mediana):", q2)
print("Q3:", q3)

Q1: 0.9525
Q2 (Mediana): 1.0550000000000002
Q3: 1.1975
```

Figure 3: Calculo cuartiles.



4 Histograma

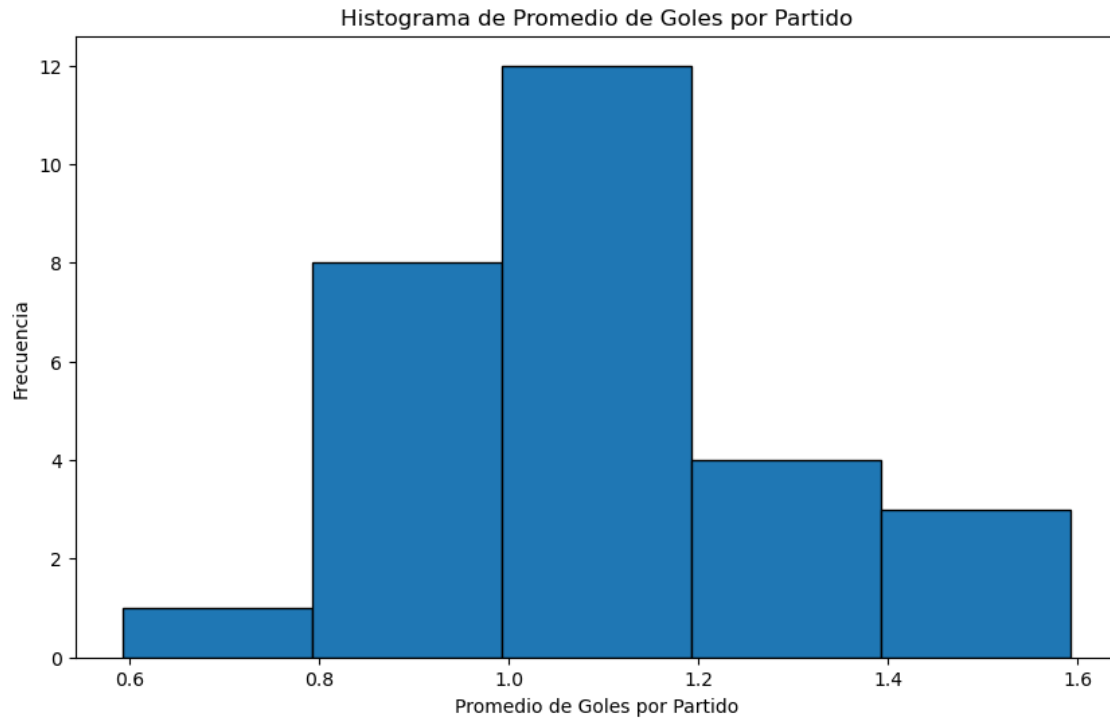
Aquí está la tabla que muestra los límites de los intervalos, la frecuencia absoluta, la frecuencia relativa, la frecuencia acumulada y la frecuencia relativa acumulada del histograma:

$$N_i = \sqrt{n}$$

donde n es el número de observaciones en el conjunto de datos. En nuestro caso, $n = 27$, por lo que:

$$N_i = \sqrt{27} \approx 5$$

Esto significa que aproximadamente necesitaremos 5 intervalos para construir nuestro histograma de manera efectiva.



Intervalo	Límite Inferior	Límite Superior	F abs	F rel	F rel acu	Fr acu
1	0.59	0.79	1	0.04	1	0.04
2	0.79	0.99	8	0.29	9	0.32
3	0.99	1.19	11	0.39	20	0.71
4	1.19	1.39	5	0.18	25	0.89
5	1.39	1.59	3	0.11	28	1.0