SMART

Messi vs Ronaldo Data Analytics Challenge

Alejandro Salazar Loza A01665123
Emiliano Torres Sandoval A01666136

The challenge involves processing and analyzing two datasets containing the performance of Lionel Messi and Cristiano Ronaldo in various international football competitions. The main goal is to apply data analytics techniques, including statistical analysis, visualization, text mining, and clustering to extract meaningful insights from both structured and unstructured data.

Emiliano Torres Sandoval

S	M	A	R	T
Compare Messi and Ronaldo's goal efficiency across competitions.	Use metrics like total goals, appearances, and goal/match ratios.	Achievable with pandas, matplotlib, scikit-learn.	Relevant as it uses real-world sports data to apply analytics concepts.	Completed on time.

Alejandro Salazar Loza

S

Identify and compare the statistical outliers in Messi and Ronaldo's goal and appearance distributions across international competitions, to assess whether either player demonstrates exceptionally high or low performance in specific tournaments.

M

Use boxplots to detect outlier competitions and quantify them by z-scores or IQR thresholds; compare the number and nature of these outliers between both players.

A

This was accomplished using seaborn boxplots and pandas filtering; the datasets are clean and limited in size, making the task feasible with standard Python tools.

R

Understanding outlier behavior reveals not just average performance, but peak or underwhelming outputs—crucial for evaluating consistency and clutch moments in a player's career.

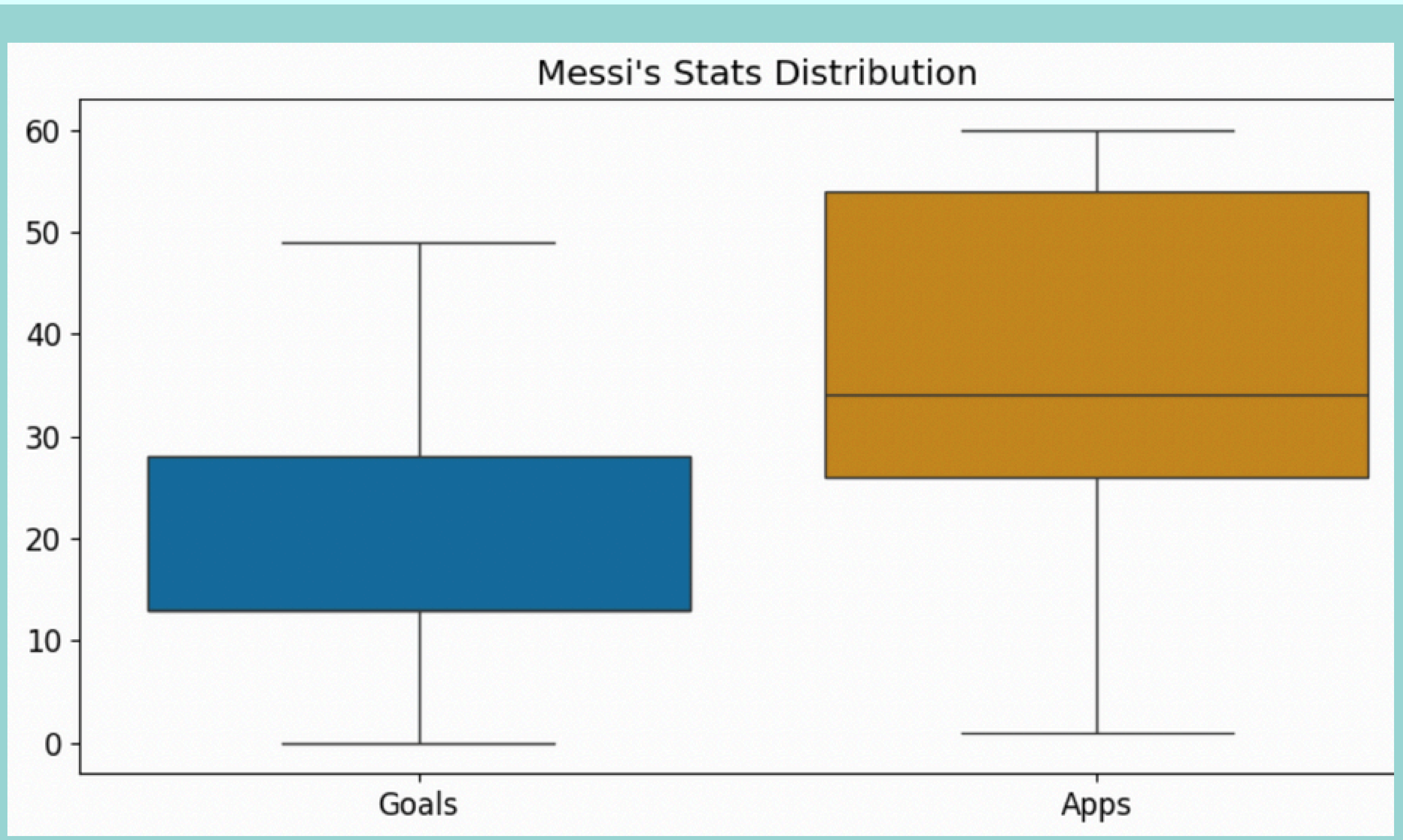
T

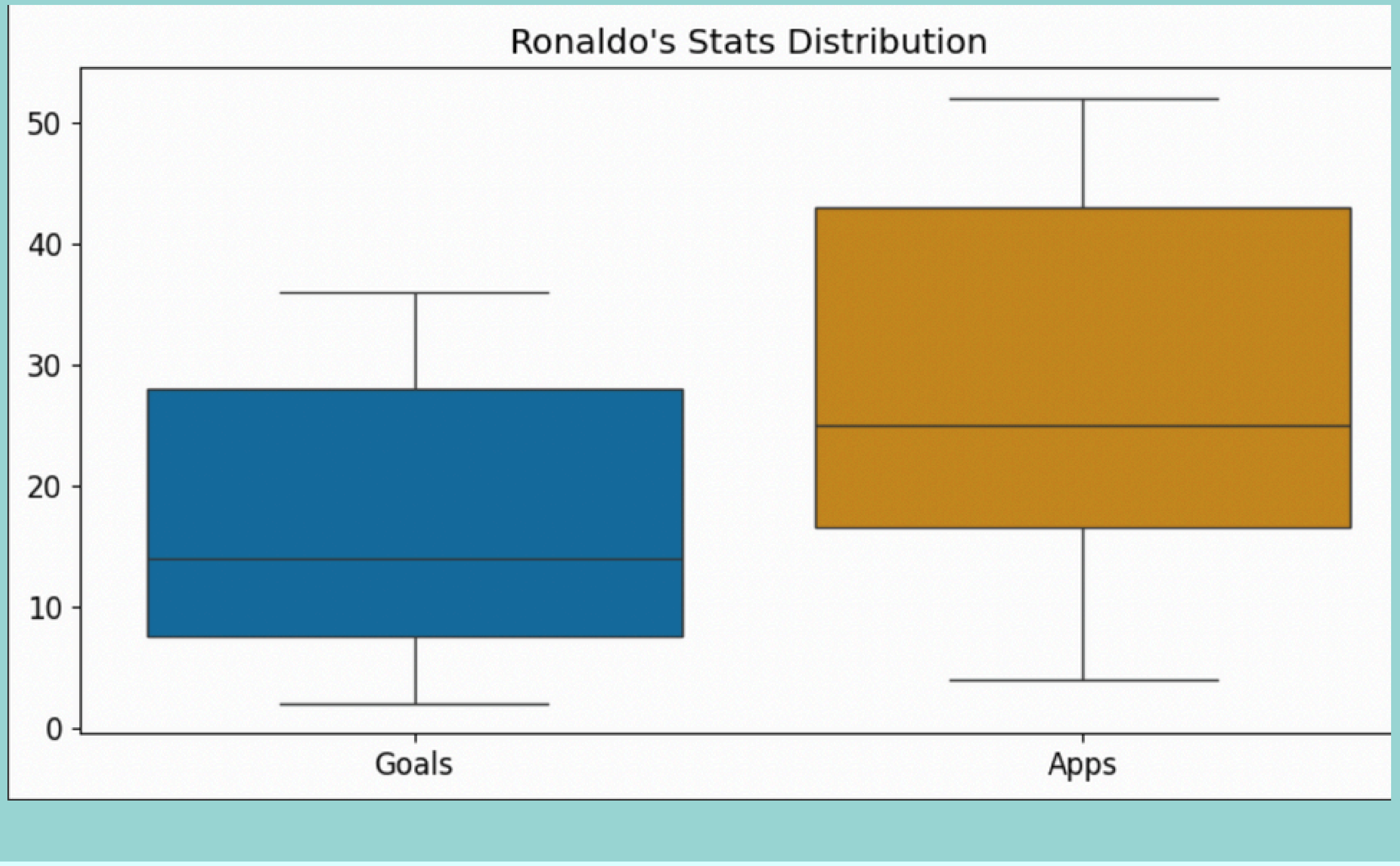
The analysis and interpretation of outlier data were completed within one working session during the challenge timeline.

BOXPLOTS

Messi is highly efficient in friendlies (49 goals / 54 games)

Ronaldo dominates in Euro qualifiers (36/39)

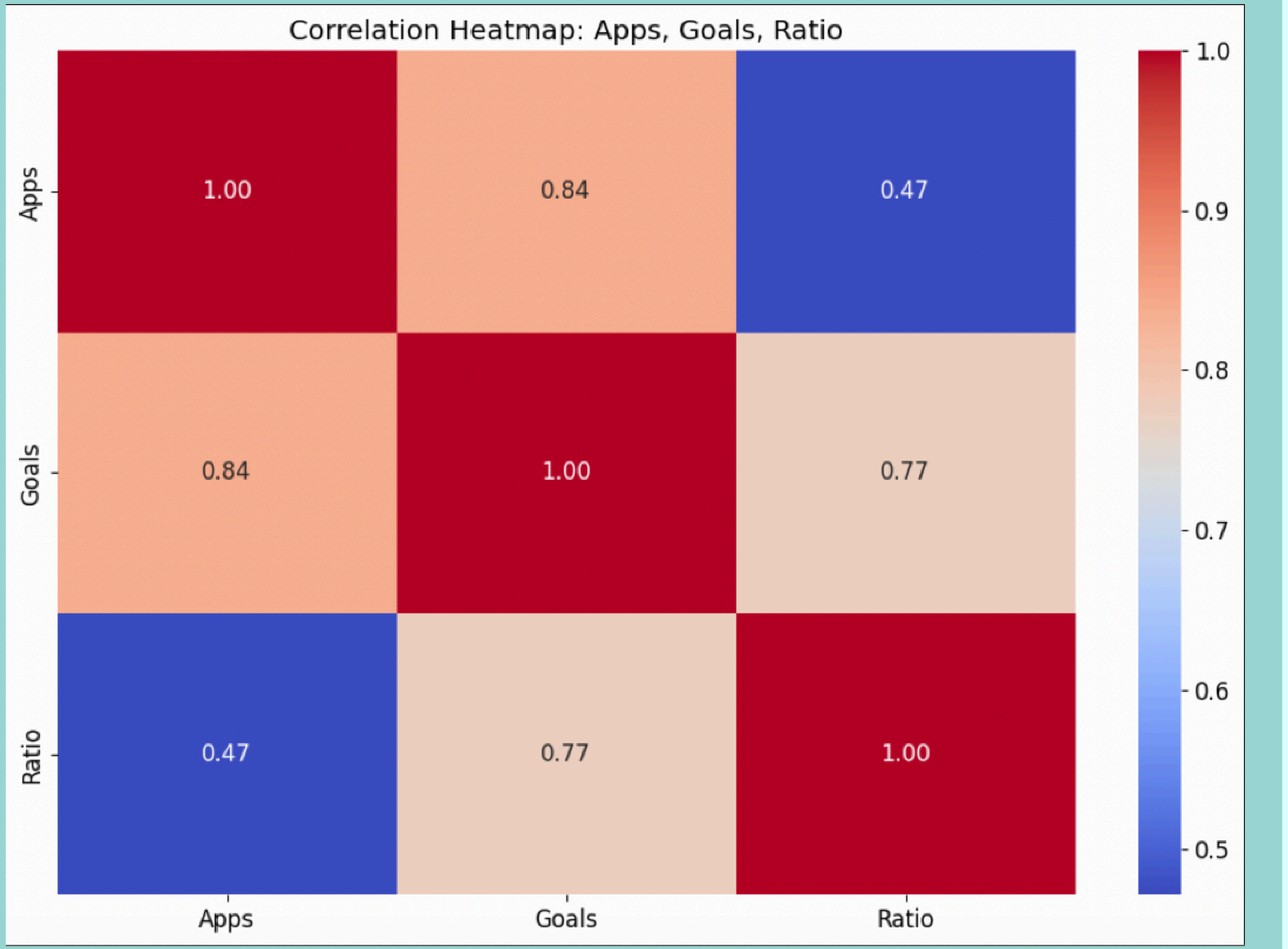




HEAT MAP

Strong positive correlation between Goals and Apps

Ratio is a better metric for fair comparison



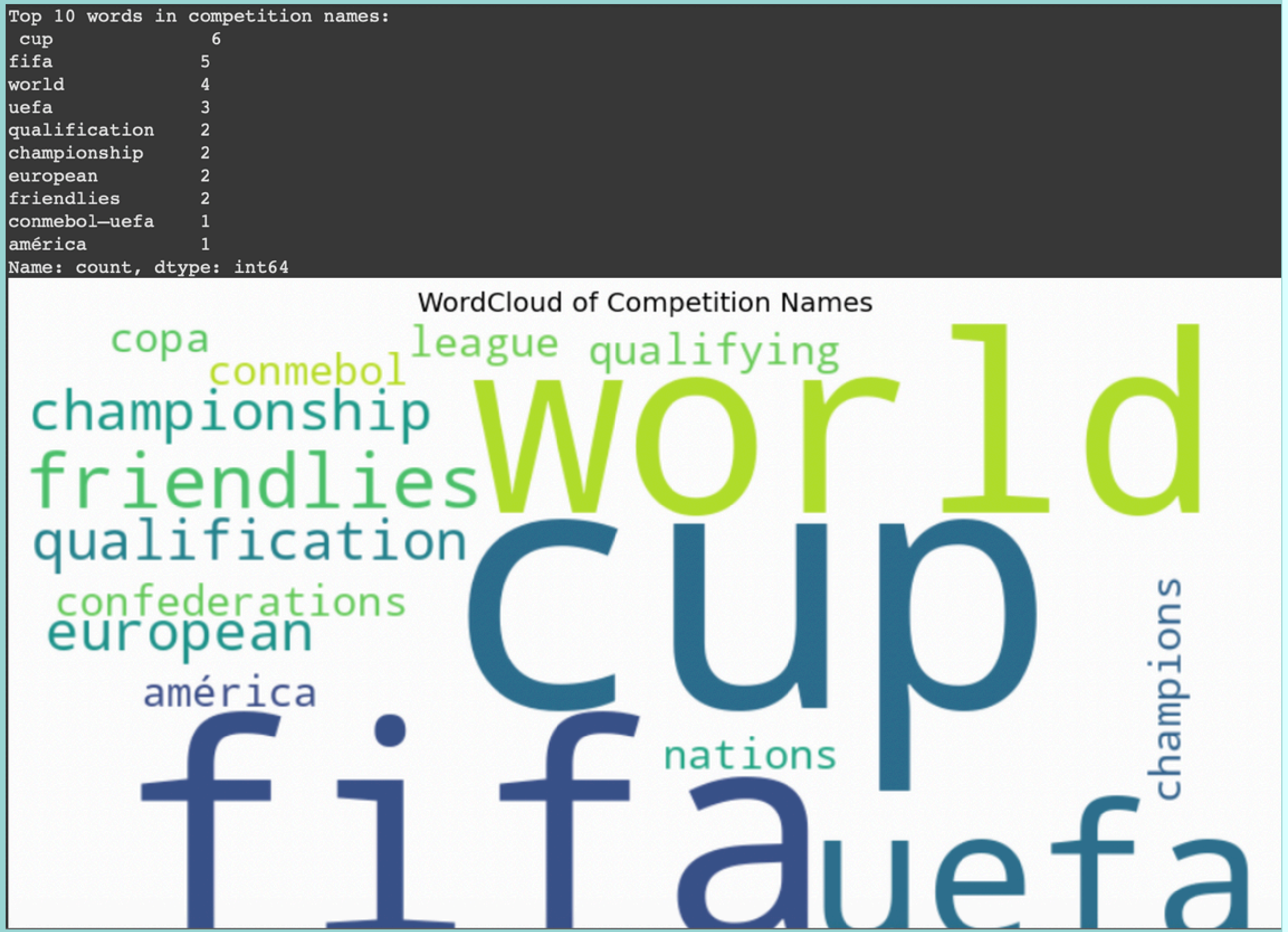
Correlation Heatmap: Apps, Goals, Ratio

	Apps	Goals	Ratio
Apps	1.00	0.84	0.47
Goals	0.84	1.00	0.77
Ratio	0.47	0.77	1.00

WORD CLOUD

"World", "qualification", "cup", "european" are dominant terms

Indicates performance is mostly assessed in competitive tournaments



```
Top 10 words in competition names:
 cup              6
fifa             5
world            4
uefa             3
qualification    2
championship     2
european         2
friendlies       2
conmebol-uefa    1
américa          1
Name: count, dtype: int64
```

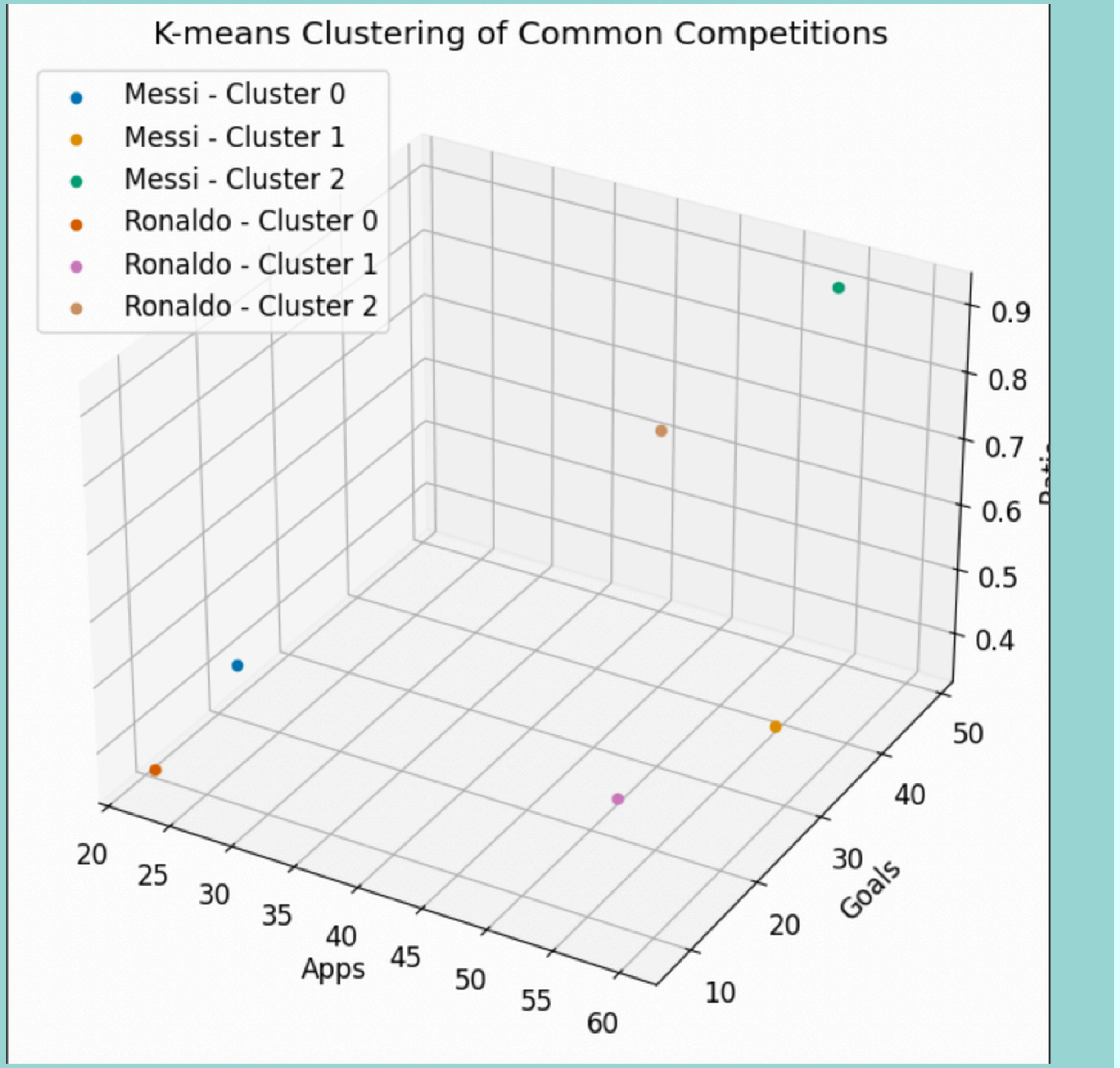
K-MEANS

Cluster 0: High appearances, moderate ratios

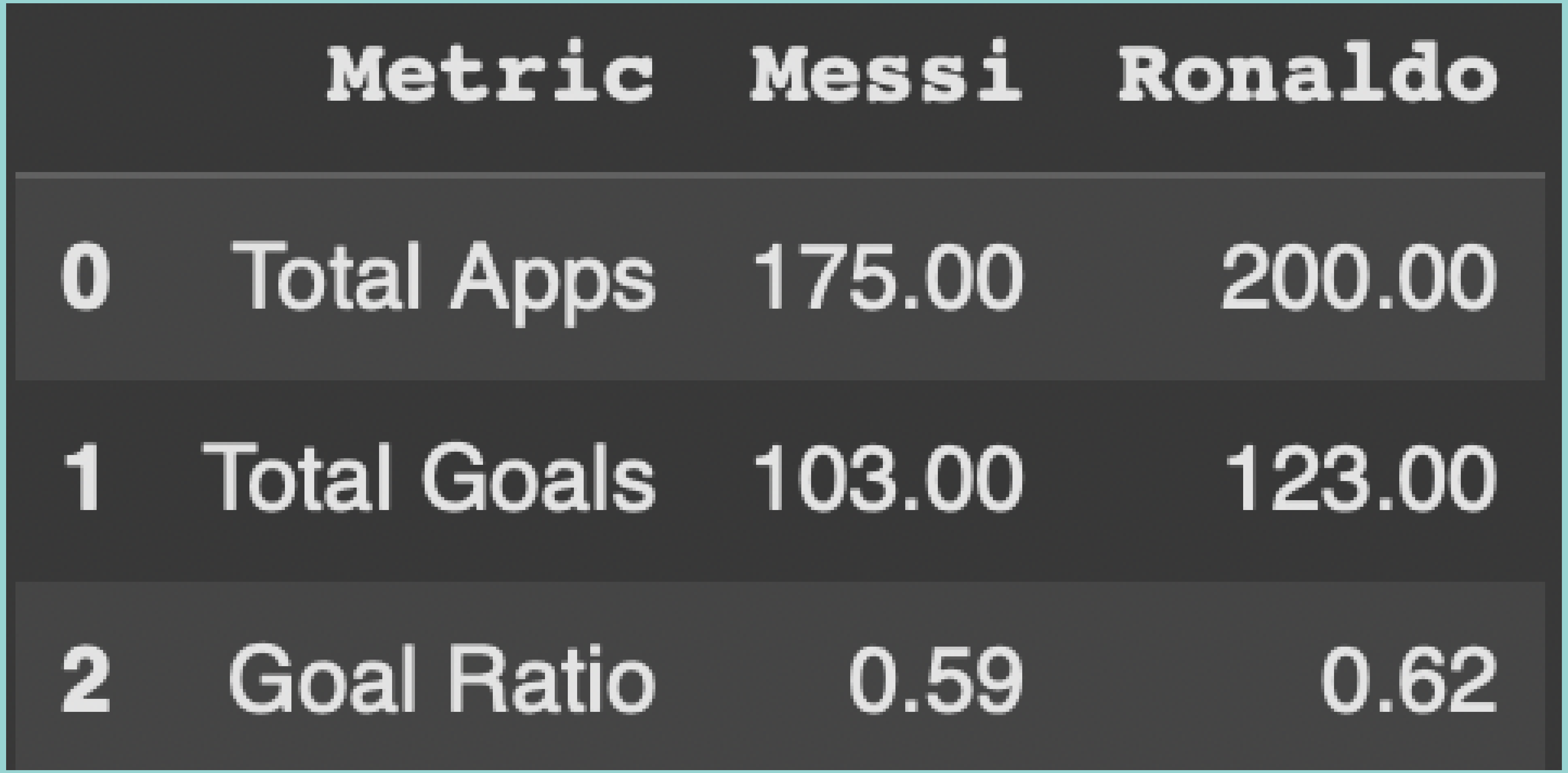
Cluster 1: Low appearances, higher efficiency

Messi tends to have high scoring in fewer games at the Copa America

Ronaldo performs consistently in long qualifiers



SUMMARY-TABLE



	Metric	Messi	Ronaldo
0	Total Apps	175.00	200.00
1	Total Goals	103.00	123.00
2	Goal Ratio	0.59	0.62

Conclusions

Emiliano Torres Sandoval

Ronaldo has more goals and matches than Messi, but their goal-per-match ratio is very close. Messi stands out in friendlies, while Ronaldo is most efficient in qualifiers. Boxplots and heatmaps helped us spot these patterns, and text mining showed that most matches were in big tournaments like World Cups and qualifiers.

With K-means clustering, we discovered two main types of competitions: ones with high efficiency in few matches, and others with many matches but lower efficiency.

Alejandro Salazar Loza

This analysis showed that Messi and Ronaldo have similar overall performance metrics, but differ in specific competitions. While both players maintain high scoring rates, their efficiency varies depending on the tournament.

The word frequency analysis highlighted differences in the types of competitions they played, reflecting their association with different football confederations. K-means clustering helped group shared competitions by performance, revealing where each player stood out.

Overall, combining statistics, text mining, and clustering gave a clearer and more balanced comparison of their international careers.