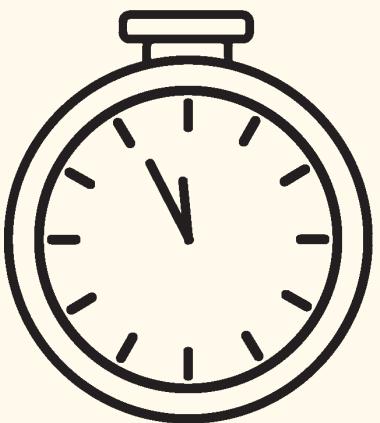


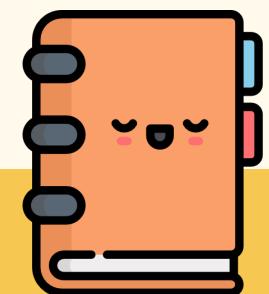
# PROYECTO FINAL ANALÍTICA DE DATOS NBA - SALARIES

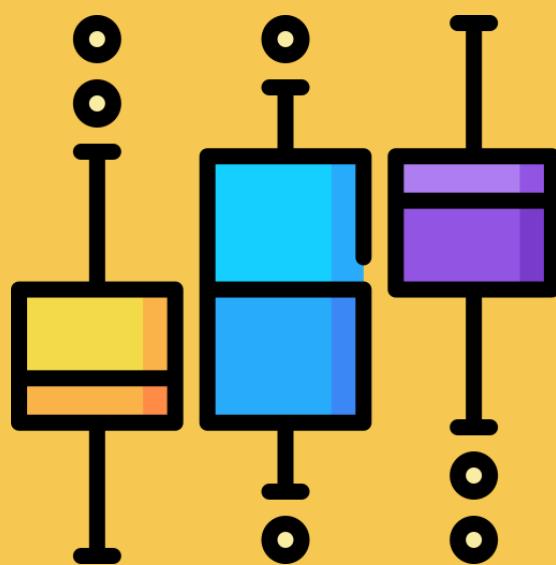
Juan Manuel Imbachí Güengue  
Juan Diego Carvajal Castaño



1. Limpieza y EDA
2. Modelos predictivos
3. Cambio de Representación del dataset
4. Caracterización de jugadores

# AGENDA





## DATOS FALTANTES

Existen datos faltantes para las variables: TS%, 3PAr, FTr, TOV%

Registros Eliminados: 2

11

Total Eliminados

## DATOS ATÍPICOS

Se eliminaron los valores atípicos más extremos

Registros Eliminados: 9

## DATOS REPETIDOS

Key Felder (3 registros)

Registros Eliminados: 0

## PROBLEMAS DE ESCALA

- Valores de distinto orden
- Diferencias significativas
  - Promedios
  - Máximos y Mínimos

## DISTRIBUCIÓN DE LOS DATOS

- Variación significativa de percentiles

## PROBLEMAS DE TIPO

- Variables categóricas:
  - 'Tm'
  - 'NBA\_Country'
  - 'Player'

# MODELOS PREDICTIVOS

- PROTOCOLO DE EVALUACIÓN
- SELECCIÓN DE VARIABLES
- ENTRENAMIENTOS
- REGRESIÓN LINEAL
- RED NEURONAL
- K-NN
- RANKING

## PROTOCOLO DE EVALUACIÓN

Training 70%

Test  
30%

## ENTRENAMIENTOS



- Baseline
- Todas las variables
- Forward
- Backward

## SELECCIÓN DE VARIABLES

Forward Stepwise: 10 variables

Backward Stepwise: 15 variables

## MODELOS

- Regresión Lineal
- Red Neuronal
- K-NN

# REGRESIÓN LINEAL

Baseline (Promedio)

- R2 adj: -0.2050

## TODAS LAS VARIABLES

- R2 adj: 0.3329

## FORWARD

- R2 adj: 0.4396



## BACKWARD

- R2 adj: 0.3946



# RED NEURONAL

Baseline (Promedio)

- R2 adj: -0.2050

## TODAS LAS VARIABLES

- R2 adj: 0.3656

## FORWARD

- R2 adj: 0.3841



## BACKWARD

- R2 adj: 0.3139

## OPTIMIZACIÓN DE PARÁMETROS

- GridSearchCV

# K-NEAREST NEIGHBORS

Baseline (Promedio)

- R2 adj: -0.2050

## TODAS LAS VARIABLES

- R2 adj: 0.1626



## FORWARD

- R2 adj: 0.1582

## BACKWARD

- R2 adj: 0.1604

## OPTIMIZACIÓN DE PARÁMETROS

- Variación de K entre 1 y 31

# RANKING

## Resultados

R2 adj Regresión Lineal

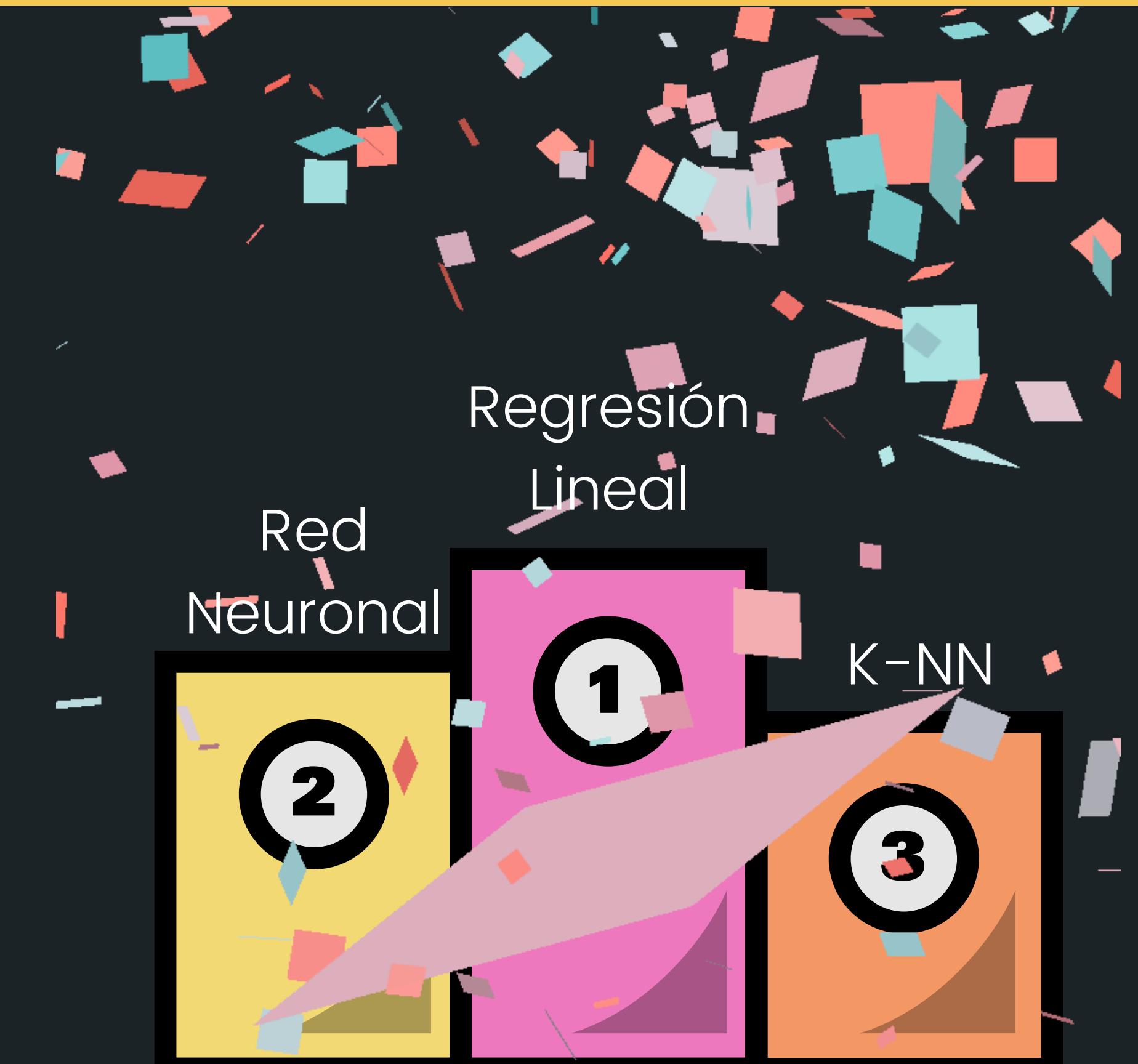
- 0.4396

R2 adj Redes Neuronales

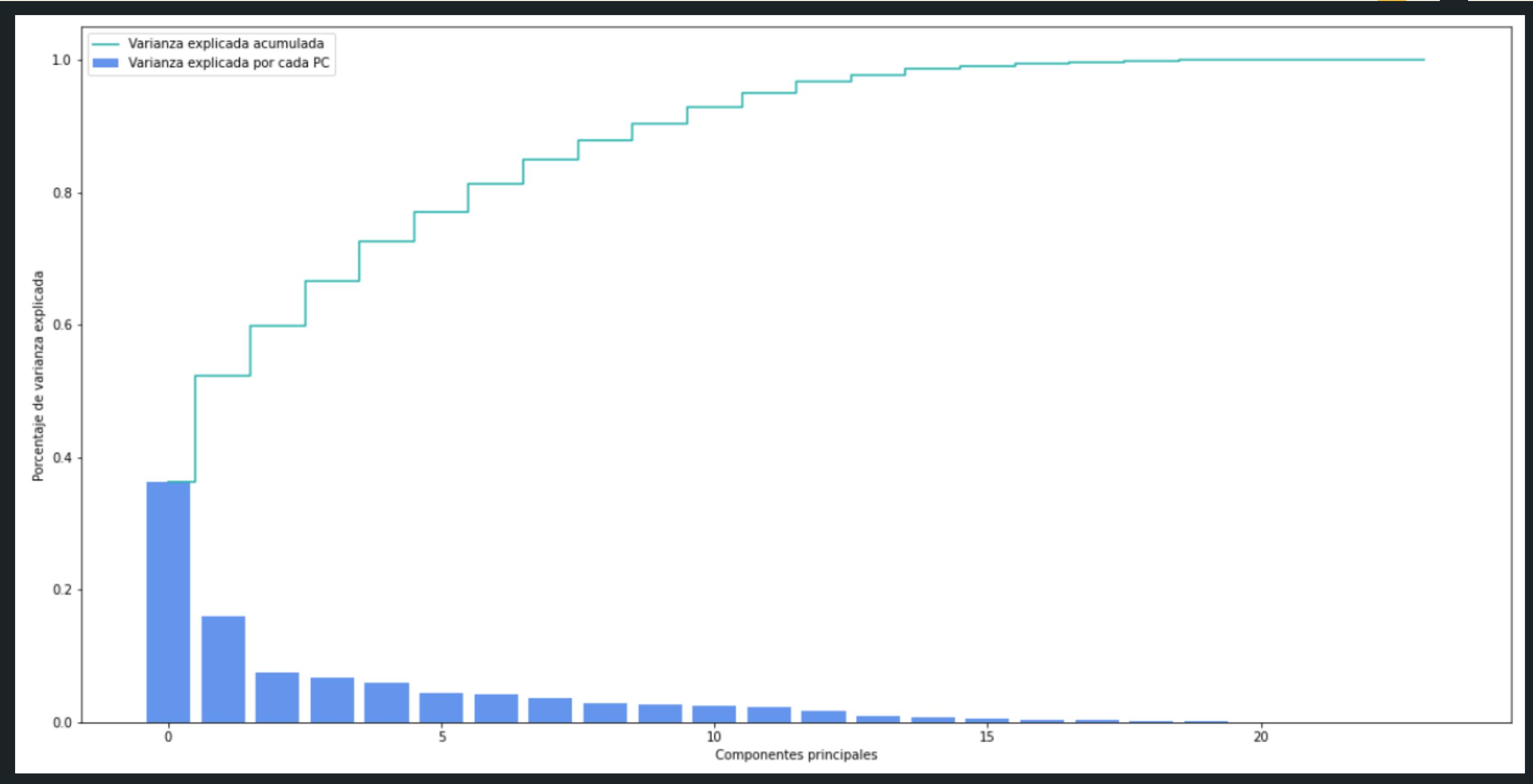
- 0.3841

R2 adj K-Nearest Neighbors

- 0.1626



# CAMBIO DE REPRESENTACIÓN

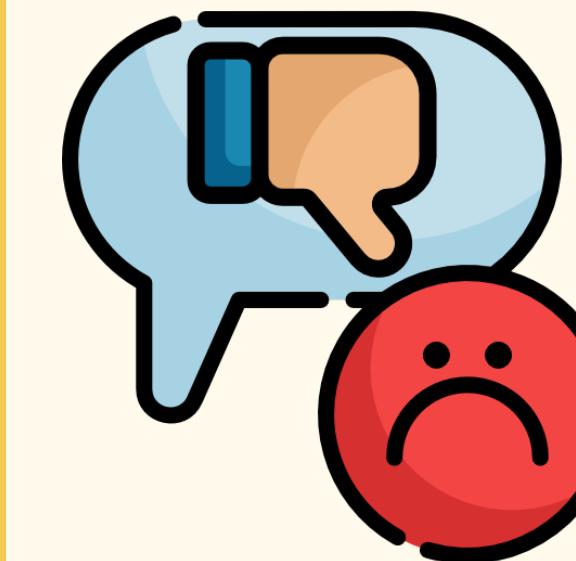


	Cluster0	Cluster1	Cluster2
Salary	2469600.34	9373811.45	7301616.41
NBA_DraftNumber	41.80	22.11	25.53
Age	25.33	27.09	26.10
G	30.02	65.97	50.75
MP	429.36	1764.66	1050.15
PER	8.10	14.55	17.10
TS%	0.47	0.56	0.58
3PAr	0.45	0.41	0.09
FTr	0.19	0.22	0.36
ORB%	3.10	2.86	10.20
DRB%	11.78	13.11	22.55
TRB%	7.43	7.98	16.37
AST%	11.25	16.73	9.41
STL%	1.51	1.65	1.38
BLK%	1.14	1.21	3.29
TOV%	12.93	12.20	14.38
USG%	16.85	20.09	18.70
OWS	0.00	1.91	1.77
DWS	0.35	1.66	1.42
WS	0.35	3.57	3.20
WS/48	0.01	0.10	0.13
OBPM	-3.56	0.42	-1.28
DBPM	-1.41	-0.47	1.01
BPM	-4.96	-0.06	-0.27
VORP	-0.15	0.98	0.86

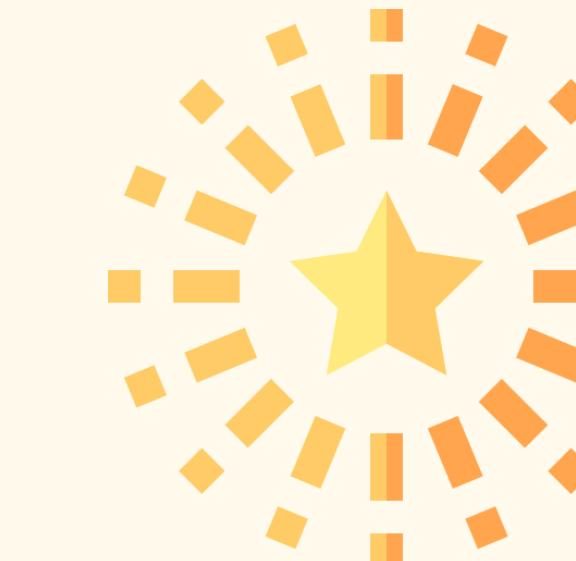
# K-MEANS

OPTIMIZACIÓN DE PARÁMETROS

CALINSKI-HARABASZ



CLUSTER 0



CLUSTER 1



CLUSTER 2



# **¡GRACIAS!**

**¿Preguntas?**