

Área de Tecnologías
Aplicadas



Ciencia de Datos y Python

Presentado por: Andrés Ibáñez



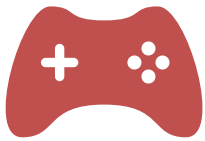
¿Quién Soy?

- Nombre: Andrés
- Idiomas:
 - Inglés (C1)
 - Español (Nativo)
 - Portugués (Básico)
- *Skills:*
 - *Python*
 - *Data Science*
 - *SQL*
 - *Linux*
 - *Windows y Windows Server*
 - *Git y GitHub*
 - *Critical Thinking and Problem Solving*
 - *IT Desk Support*
- Me apasiona el aprendizaje, la informática, los puzles criptográficos, los deportes de tuerca y los videojuegos.

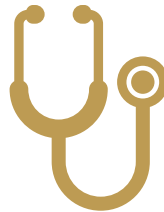
¿Qué es la ciencia de datos?

- Práctica que consiste en la utilización de volúmenes de datos.
- Su objetivo es de ayudar a la toma de decisiones.
- Enfoque multi disciplinario.
- Es una de las carreras más demandadas en el último periodo.
- Ayuda a la toma de decisiones estratégica
- Se utiliza en distintas áreas.

¿Para qué se utiliza?



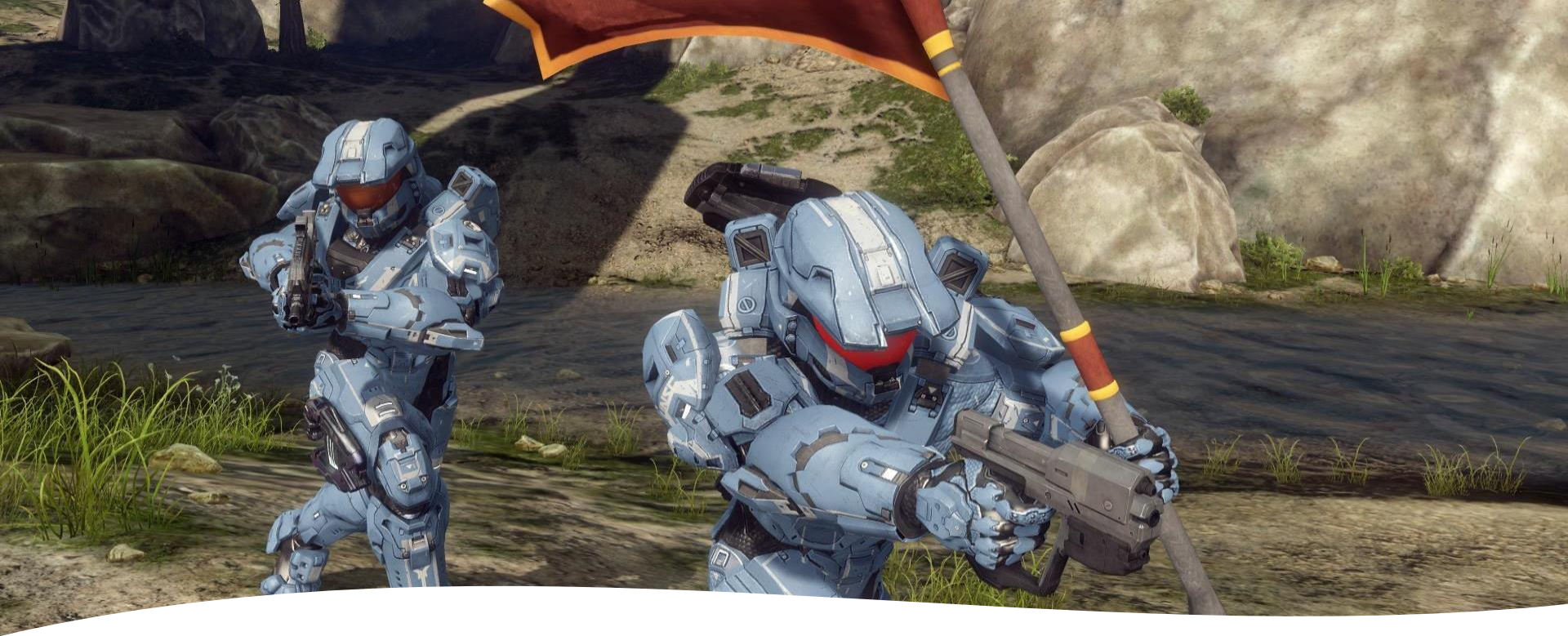
Videojuegos



Medicina



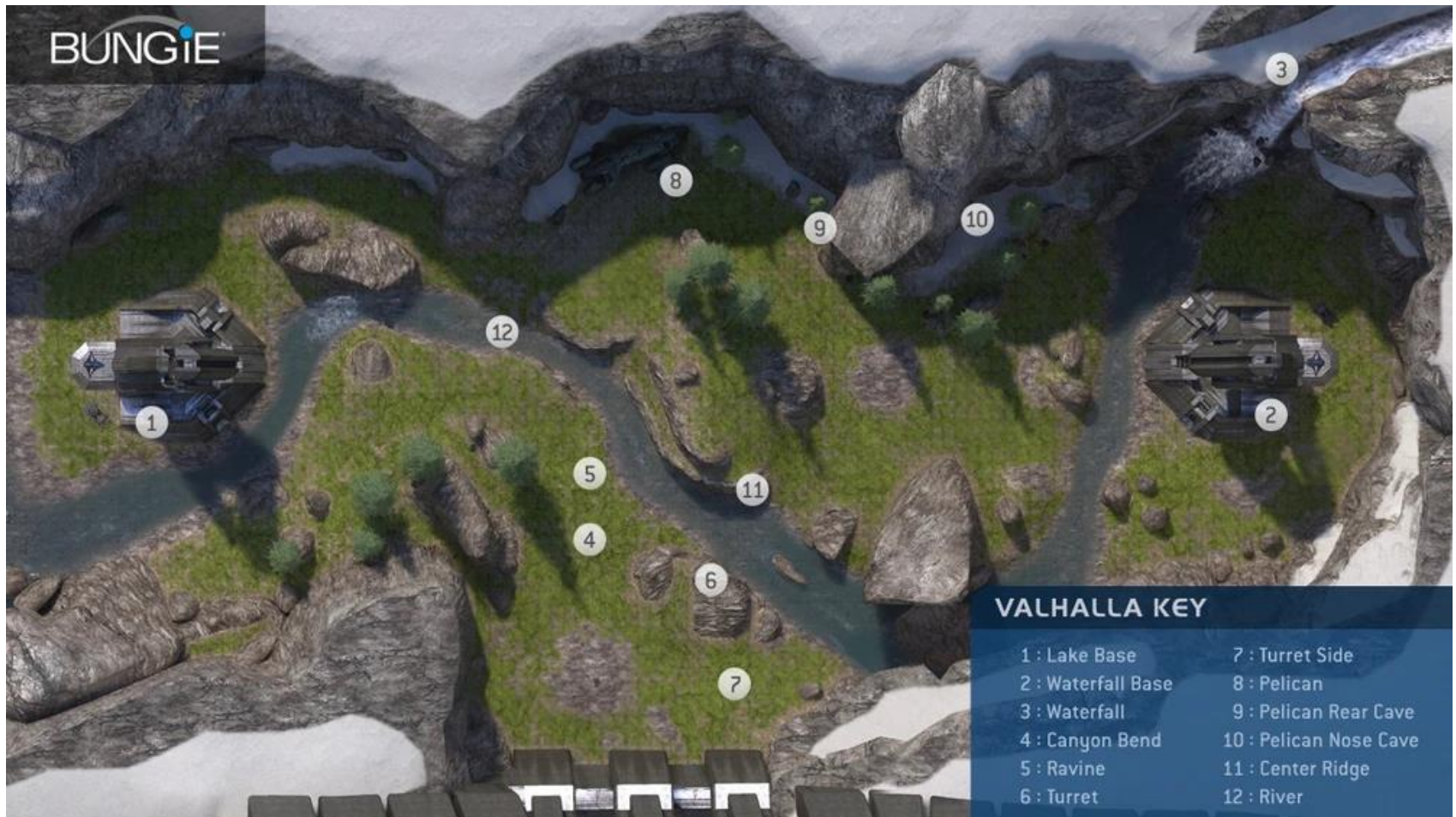
Deportes de
motor



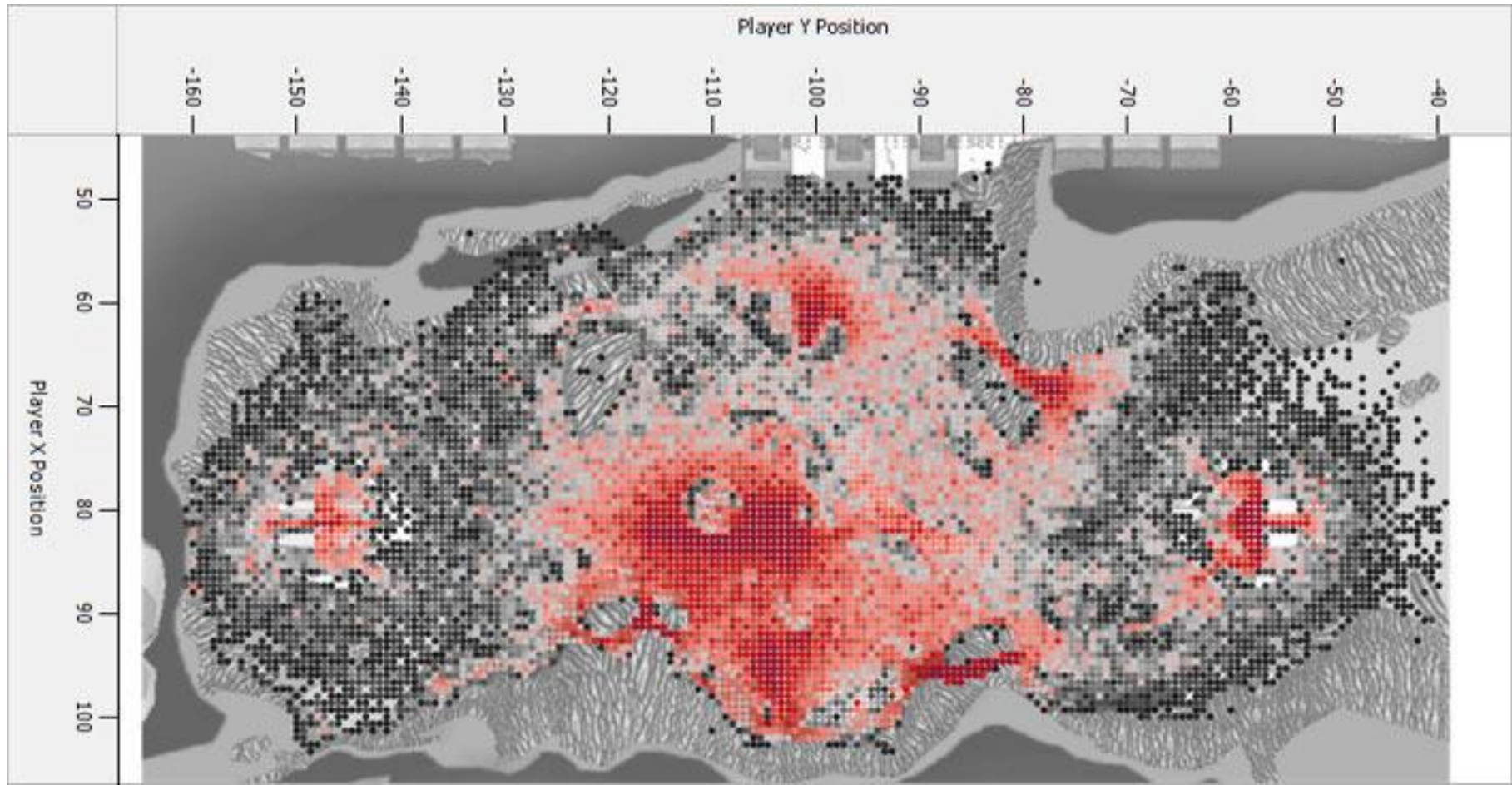
El caso de Halo 3

- Nacimiento del gaming en línea a gran escala.
- Problemas y más problemas.
- Hizo que el nacimiento de una idea brillante surgiera de las oficinas de Bungie.

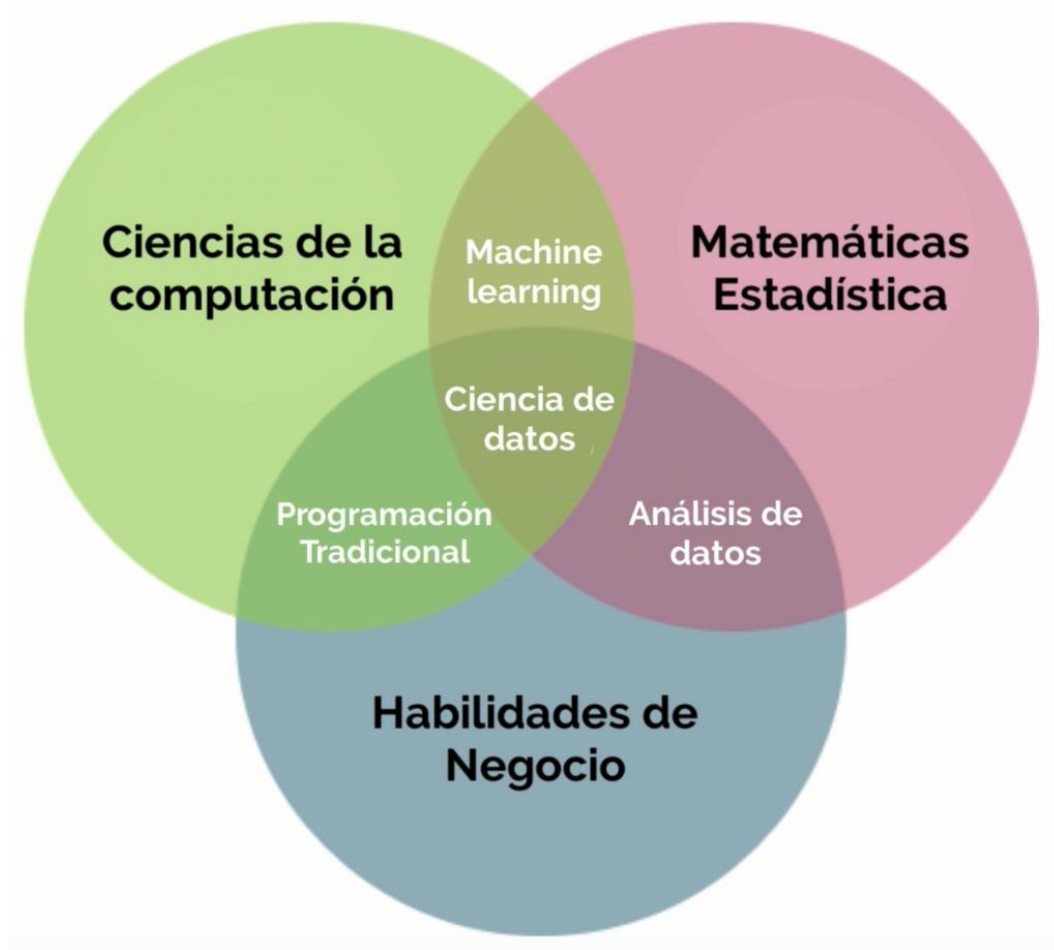
Conociendo Valhalla



El uso de heatmaps para la solución del problema.



La Triada del Conocimiento de la Ciencia de Datos



La Ciencia de Datos Antes de Python

Weka (/ˈwɛkə/)

- Nace en 1999 como software de minería de datos y aprendizaje automático.
- Permite trabajar a través de una interfaz de usuario amigable (*para la época*)
- Escrito en Java.
- Software libre.



R

- Nacido en 1993 con la premisa de ser un software de visualización estadística.
- Utilizado en *machine learning*, minería de datos e inferencias estadísticas.
- Escrito en C, Fortran y por sí mismo.
- Software libre.



¿Sus problemas?



```
gapminder <- readr::read_csv("https://raw.githubusercontent.com/cle  
View(gapminder)  
  
a <- min(gapminder$anio)  
b <- max(gapminder$anio)  
fin_i <- seq(a, b, 5) # seq(inicio, fin, salto)  
fin_f  
for (i in fin_i) {  
  temporal <- subset(gapminder, anio=i) # Creo una base temporal fi  
  temporal$continente <- factor(temporal$continente,  
                                levels = c("Americas", "Europa", "Af  
                                "Asia", "Oceania"))  
  temporal$continente <- c("green", "yellow", "blue", "red", "red2")  
  
  plot(x = temporal$pib_per_capita, # Vector numérico con las  
       y = temporal$esperanza_de_vida, # Vector numérico con  
       main = "Gráfico de dispersión", # Título del gráfico  
       xlab = "PIB per capita", # Nombre del eje X  
       ylab = "Esperanza de vida", # Nombre del eje Y  
       pch = 19, # Tipo de punto  
       xlim = c(240, 114000), # Rango del eje X  
       ylim = c(20, 90), # Rango del eje Y  
       col = temporal$continente)  
  temporal$pais <- factor(temporal$pais, levels = temporal$pais)  
  text(temporal$pib_per_capita, temporal$esperanza_de_vida, labels  
  abline(lm(temporal$pib_per_capita ~ temporal$esperanza_de_vida)  
        lty = 1, # Tipo de línea  
        lwd = 2, # Grueso de línea  
        col = temporal$continente) # Color de línea  
  Sys.sleep(1.5)  
}
```



Weka Explorer

fy Cluster Associate Select attributes Visualize

Open URL... Open DB... Generate... Undo Edit... Save...

Apply

Attributes: 10
Sum of weights: 286

Invert Pattern

Remove

Selected attribute

Name: age
Missing: 0 (0%)
Distinct: 6
Type: Nominal
Unique: 1 (0%)

No.	Label	Count	Weight
1	10-19	0	0.0
2	20-29	1	1.0
3	30-39	36	36.0
4	40-49	90	90.0
5	50-59	96	96.0

Class: Class (Nom)

Visualize All

Log

Python y la ciencia de datos

- Nació como un Proyecto de hobby de Guido Van Rossum para mantenerse ocupado en sus tiempos libres.
- Su uso en DS comenzó en 2007. (Nacimiento de pandas)
- Lidera las encuestas de preferencia en DS, ML y DM.
- Multiparadigma.
- Posee múltiples librerías para cualquier tarea que se le quiera dar.
- Fácil de leer y aprender.



Las Tres Librerías Más Utilizadas para la Ciencia de Datos



Matplot Lib
(Visualización de datos)



Entrenamiento de Modelos
(Machine Learning)



Manipulación de Datos
(Data Mining)

Recomendaciones

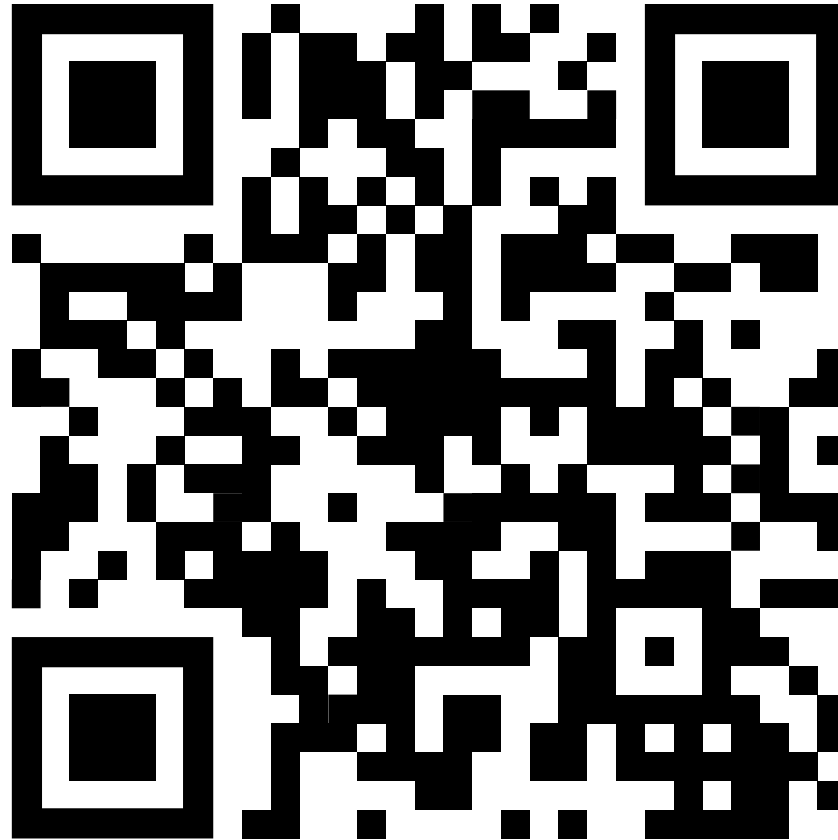


Curso DS Kaggle



Las mejores herramientas para DS

¡Muchas gracias por su atención!



¡Conecta conmigo en LinkedIn para conocer mis proyectos actuales y futuros!