



**UNIVERSIDAD
DE LA FRONTERA**



DEPARTAMENTO DE
CS. COMPUTACIÓN E INFORMÁTICA

Hito 2 - Ing. de Datos

Historia de Hearthstone

Rodrigo Valenzuela
Armin Rodríguez
Pablo Nahuelpán
Nicolás Hidalgo
Octubre, 2022



Mejoras: Contexto

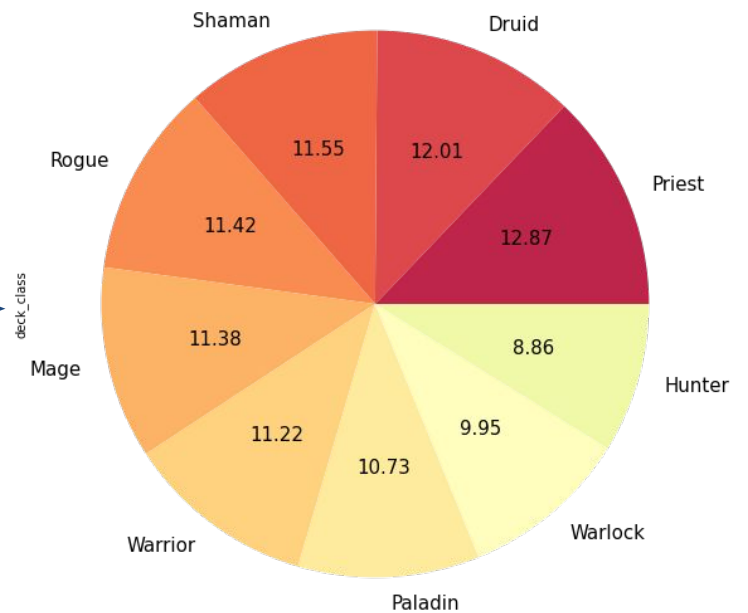
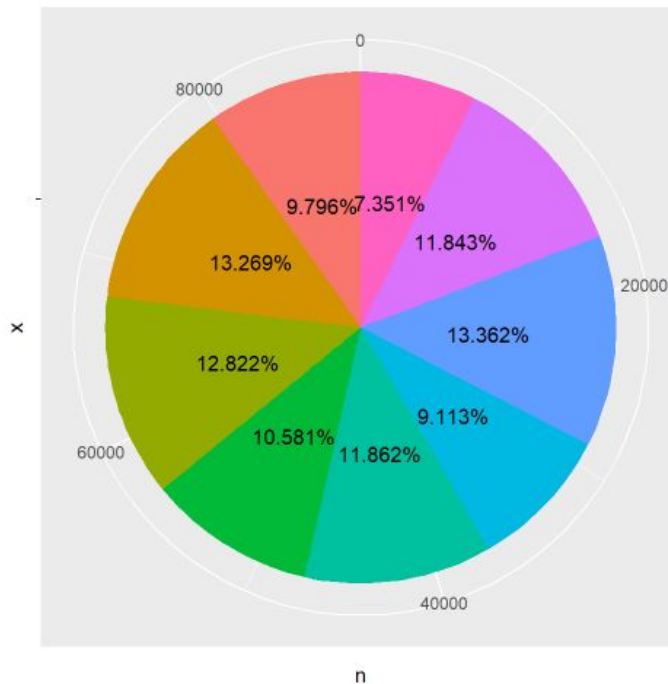
- En la versión original del hito, el contexto dado en la introducción solo habla sobre la existencia del juego de donde provienen los datos, sin definir un objetivo u dirección clara.
- Se acotó la información del juego, para darle espacio a la dirección que toma el EDA y el objetivo que se busca cumplir.

Mejoras: Exploración de datos

- En la versión original de la exploración de datos, las descripciones acompañando la información presentada, tenía un foco más centrado en explicar que significa la información por sobre el valor que esta trae al contexto.
- Además, el formato de ciertos gráficos no facilitaban la navegación de los datos, utilizando colores que hostigan la visión y dificultan la lectura, además de reducir el espacio del gráfico con información que puede ser presentada de mejor manera (leyendas).

Mejoras: Legibilidad

Por medio del uso de paletas de colores más amigables a la visión, y la integración de la información contenida en la leyenda del gráfico, la nueva versión mantiene el foco en el gráfico, y le permite el espacio para presentar sus datos sin ser aglomerados.



Problemas identificados

1. Problema de clasificación binaria basado en la variable "deck_format".
2. Encontrar una característica o atributos de un mazo que puedan realizar una mejor clasificación de un formato.
3. ¿Cuáles serían los atributos que puedan permitir hacer una mejor predicción del formato de un mazo?

Propuesta experimental

- Cómo sabemos que un mazo se compone de atributos que hacen referencia a cartas del mazo, vamos a agrupar cartas que se juegan en un formato y en el otro para así lograr hacer una clasificación de un mazo respecto a las cartas que utiliza, para aquello utilizaremos una técnica de clustering.
- Vamos a clasificar con 3 modelos árbol de decisión, knn y naive bayes para identificar qué modelo sirve para clasificar el formato de un mazo.
- Analizaremos la correlación de los atributos no relacionados a las cartas de un mazo con el formato de este para así detectar cuáles son los atributos nos permiten hacer una mejor clasificación.

Experimento preliminar

Predicción a través de árbol binario

	precision	recall	f1-score	support
S	0.99	0.99	0.99	37916
W	0.99	1.00	0.99	43466
accuracy			0.99	81382
macro avg	0.99	0.99	0.99	81382
weighted avg	0.99	0.99	0.99	81382

Matriz de confusión:

```
[[37677  239]
 [ 198 43268]]
```

Predicción por Navie Bayes

	precision	recall	f1-score	support
S	0.47	1.00	0.64	75829
W	0.65	0.00	0.00	86934
accuracy			0.47	162763
macro avg	0.56	0.50	0.32	162763
weighted avg	0.56	0.47	0.30	162763

Matriz de confusión:

```
[[75804  25]
 [86888  46]]
```

Experimento preliminar

Predicción por Regresión logística

	precision	recall	f1-score	support
S	0.49	0.92	0.64	75571
W	0.72	0.17	0.28	87192
accuracy			0.52	162763
macro avg	0.61	0.55	0.46	162763
weighted avg	0.61	0.52	0.45	162763

Matriz de confusión:

```
[[69690  5881]
 [72000 15192]]
```




UNIVERSIDAD
DE LA FRONTERA



DEPARTAMENTO DE
CS. COMPUTACIÓN E INFORMÁTICA

Hito 2 - Ing. de Datos

Historia de Hearthstone

Rodrigo Valenzuela
Armin Rodríguez
Pablo Nahuelpán
Nicolás Hidalgo
Octubre, 2022

