



# STEAM

## Store Games



Grupo 11  
CC5205 Minería de Datos  
9 de junio del 2023

Equipo 11 [José Arriagada · Ignacio Muñoz · Vicente Guzmán · Nicolás González · Javier Atenas]



# IDEA GENERAL DEL PROYECTO

Estudiar el dataset para entender las preferencias en el mercado de los videojuegos, ver cuáles son los más vendidos y cuáles tienen las mejores valoraciones. Y así con esto, descubrir patrones y tendencias, identificar oportunidades de negocio en este mercado y comprender mejor el comportamiento de los clientes, las preferencias de estos y sus necesidades.



# PREGUNTAS

- ¿Sería posible predecir el precio de un nuevo juego de la tienda si solo tengo sus características?
- ¿Existen grupos importantes de juegos que se comporten de manera similar en cuanto a precio de acuerdo a características del juego (promedio de tiempo jugado, valoración) o en cuanto al género o tags que posee? ¿Hay alguna relación entre estas?
- ¿Hay características específicas de los juegos que permitan tener mejor o peor valoración del público?
- ¿Sería posible conocer la valoración (aproximada) de un nuevo juego que entra al mercado considerando sus características?



# PROPUESTA EXPERIMENTAL

- Las preguntas 1, 3 y 4 son de carácter predictivo.
- Para la pregunta 2, se usará clustering para visualizar si es que las características

# RESULTADO PRELIMINAR

```
1 from sklearn.tree import DecisionTreeRegressor
2 from sklearn.model_selection import train_test_split
3 from sklearn.model_selection import cross_val_score
4 from sklearn.metrics import mean_squared_error
5 from sklearn.metrics import accuracy_score
6 from sklearn.metrics import classification_report
7
8 clf = DecisionTreeRegressor()
9 y = X.price
10 Xwp = X.copy()
11 Xwp = Xwp.drop("price", axis=1)
12 clf.fit(Xwp, y)
13 X_train, X_test, y_train, y_test = train_test_split(Xwp,
14 y, test_size=0.2, random_state=42)
15 regressor = DecisionTreeRegressor()
16 regressor.fit(X_train, y_train)
17
18 y_pred = regressor.predict(X_test)
19 mse = mean_squared_error(y_test, y_pred)
20 print("Mean Squared Error (MSE):", mse)
```

MSE: 131.67

```
1 # Valores de ejemplo para los otros atributos
2 english = 1
3 required_age = 18
4 achievements = 50
5 positive_ratings = 1000
6 negative_ratings = 200
7 average_playtime = 120
8 median_playtime = 90
9
10 # Preparar los valores de los atributos para la
    predicción
11 input_data = [[english, required_age, achieveme
    nts, positive_ratings, negative_ratings, averag
    e_playtime, median_playtime]]
12
13 # Realizar la predicción
14 predicted_price = regressor.predict(input_data)
15
16 # Imprimir el valor predicho
17 print("Predicted Price:", predicted_price)
```

Predicted Price: 49.99