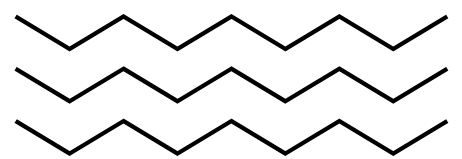
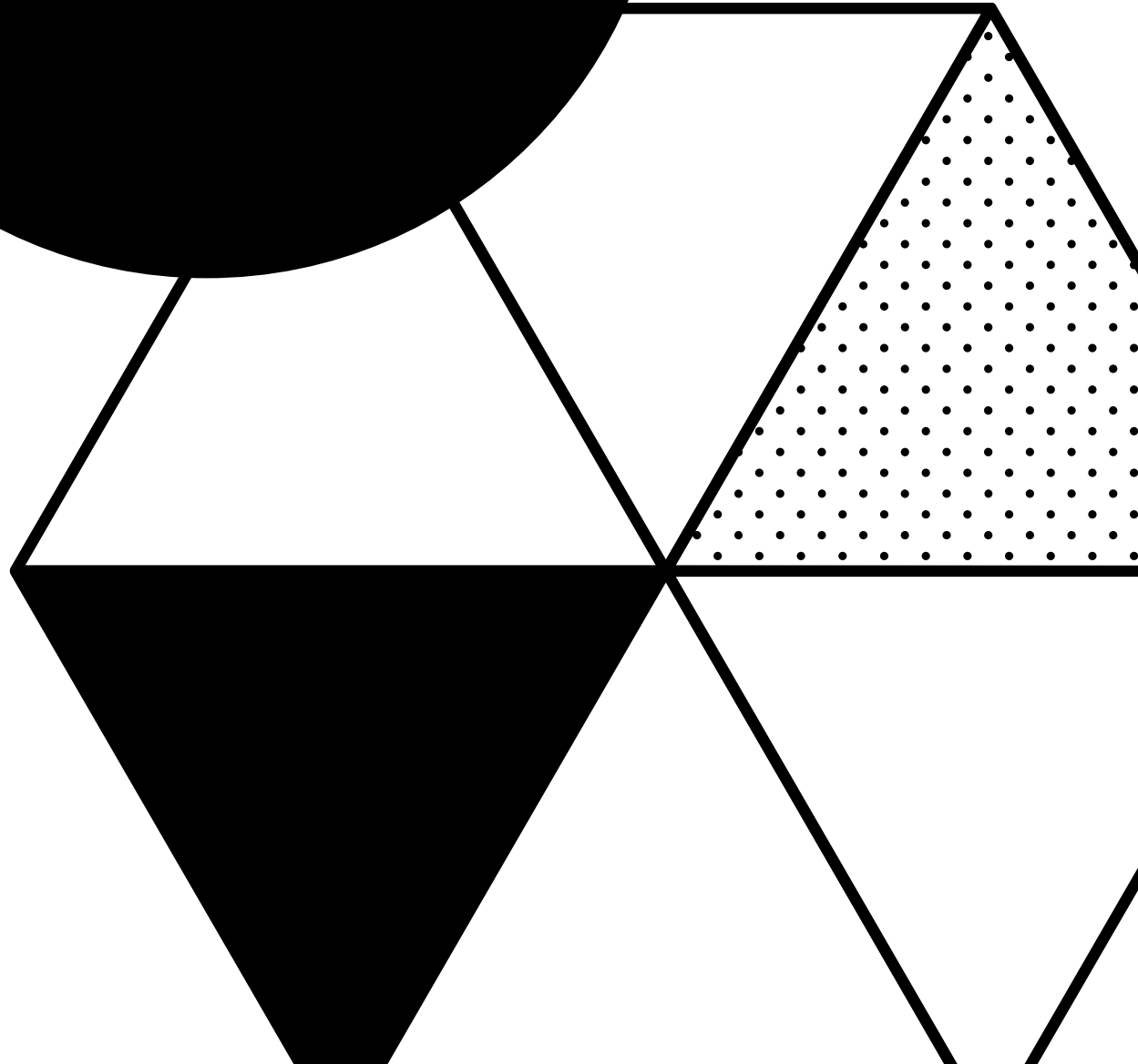
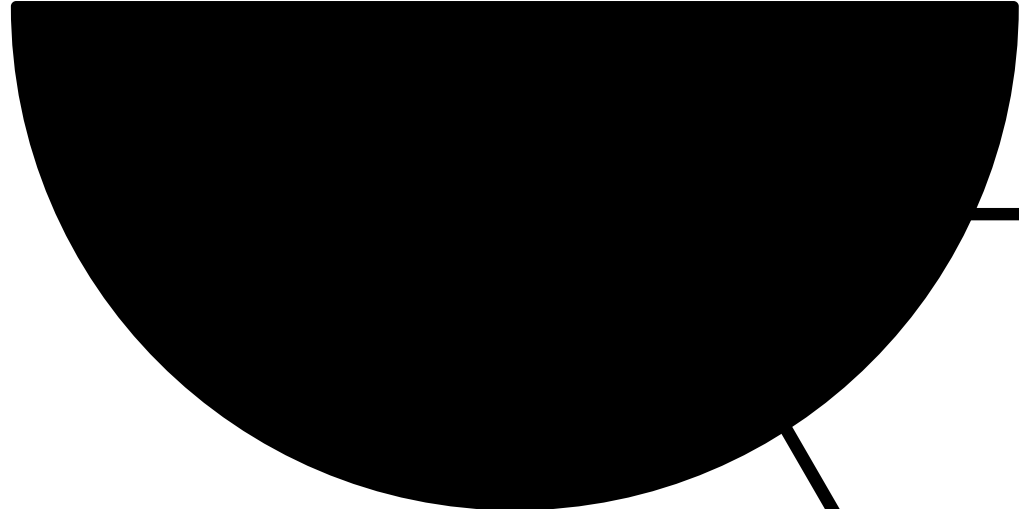
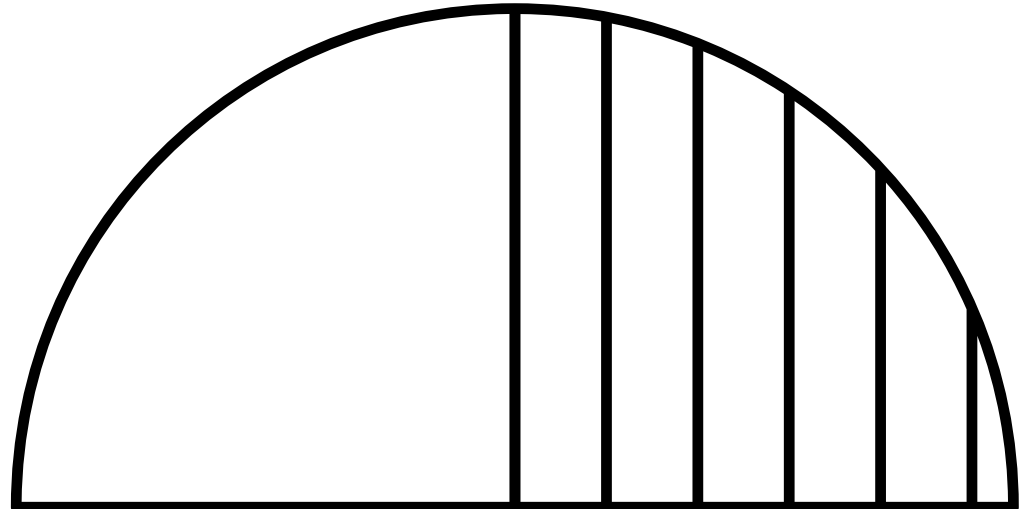
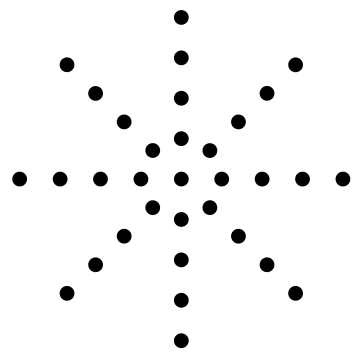
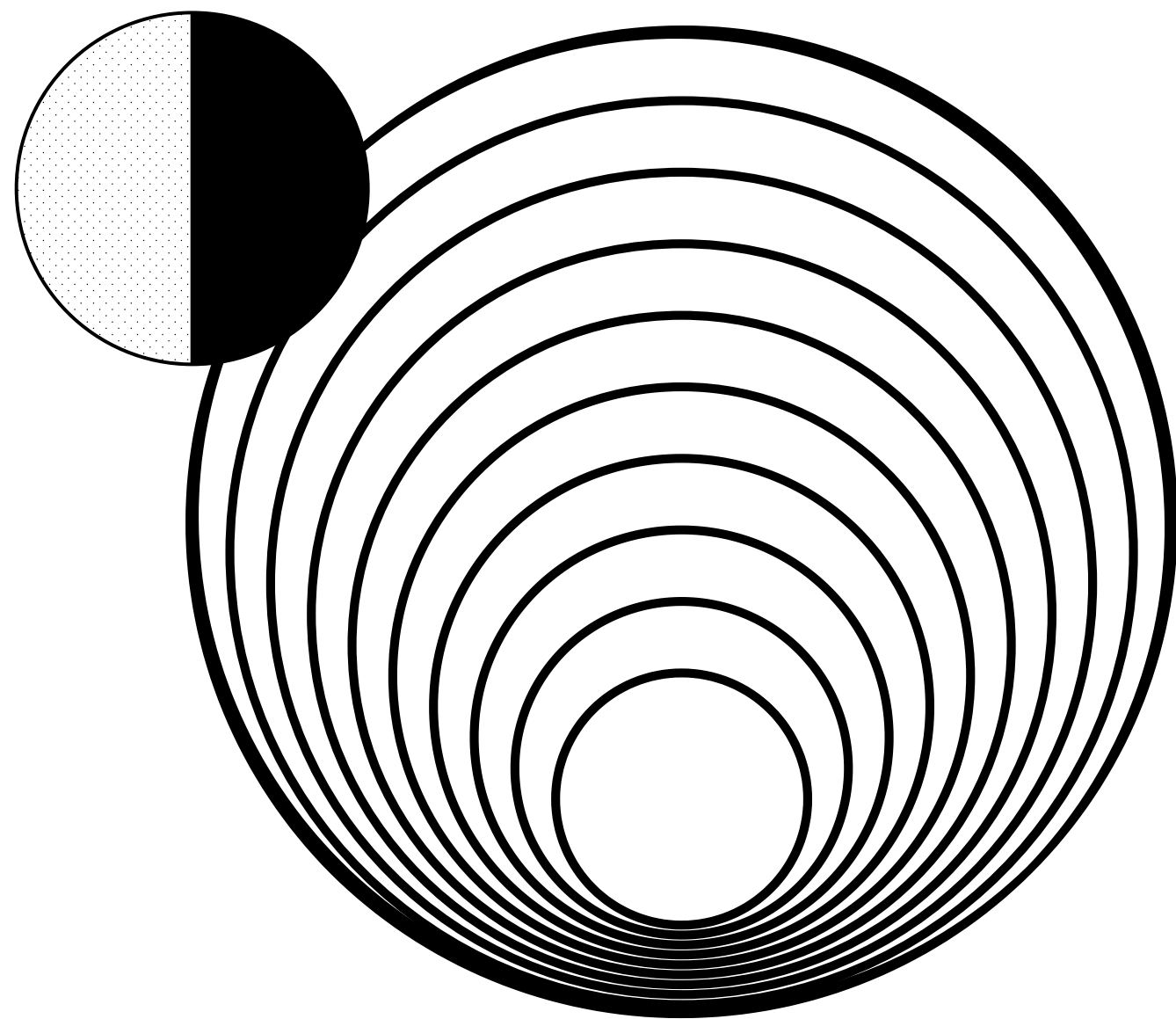


ANÀLISIS DEL DATASET NIKE VS ADIDAS



Judit Panisello Lozano
1605512



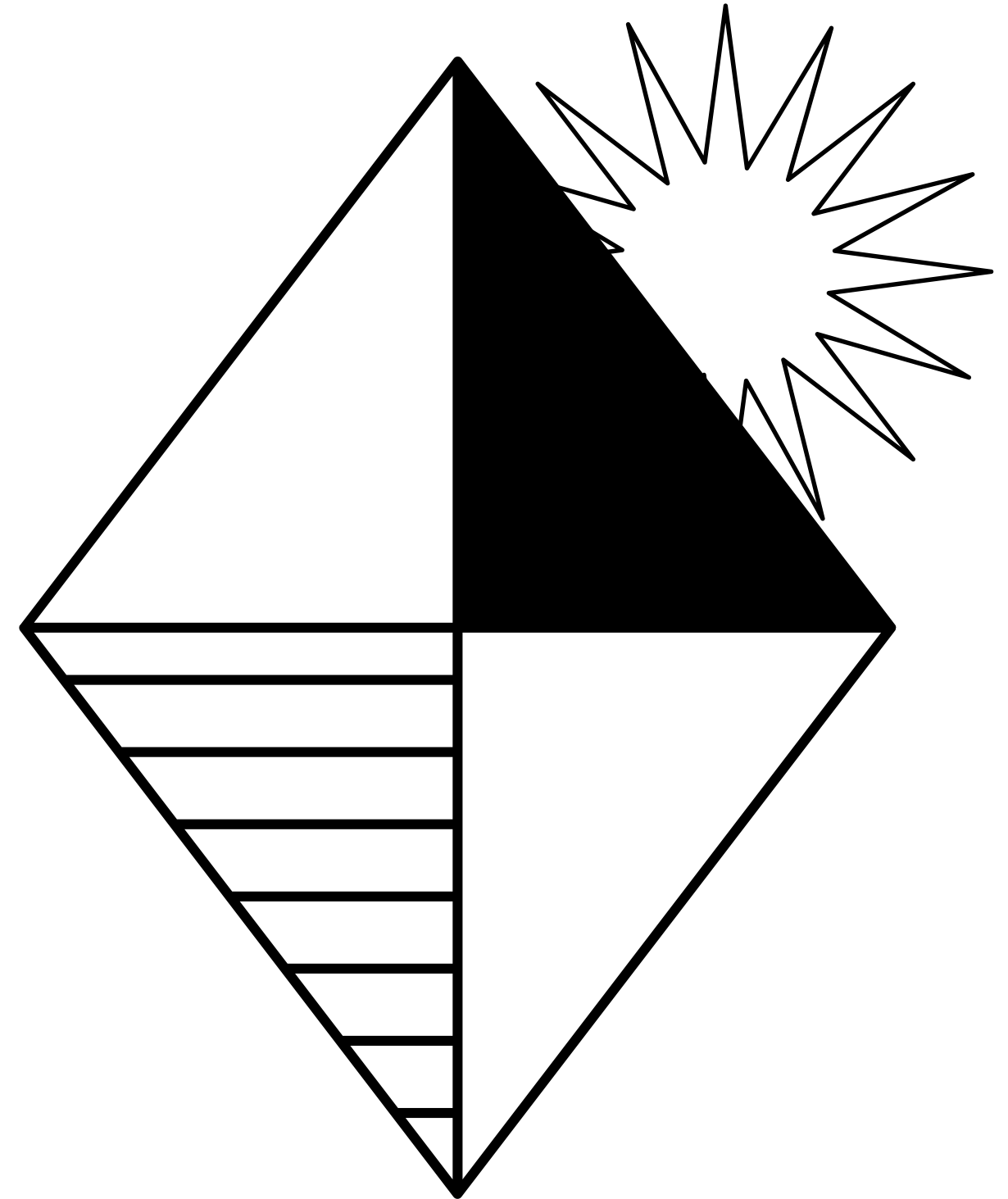


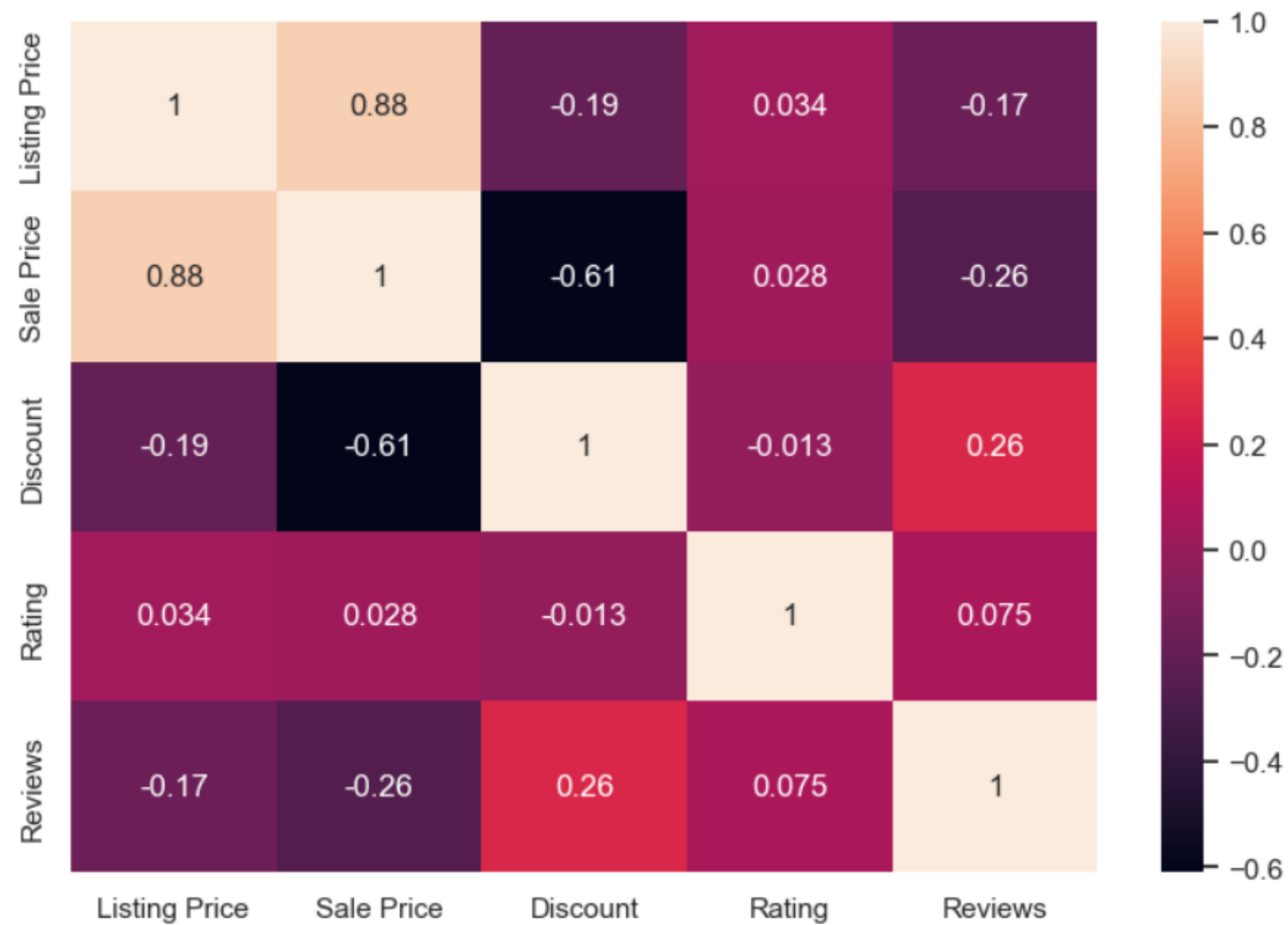
Índex

1. Introducció
2. EDA
3. Processament de Dades
4. Primera Implementació de Models
5. Models de Classificació Binària
6. Conclusions

Introducció

Link al github:
<https://github.com/JuditPanisello/CasKaggleNikeVsAdidas>





EDA

Informació de les 2 marques principals Vs Informació Total de marques principals

Adidas	2625	Adidas CORE / NEO	1111
Nike	643	Adidas ORIGINALS	908
		Nike	643
		Adidas SPORT PERFORMANCE	606

Processament de Dades

01

Completar Columnas

02

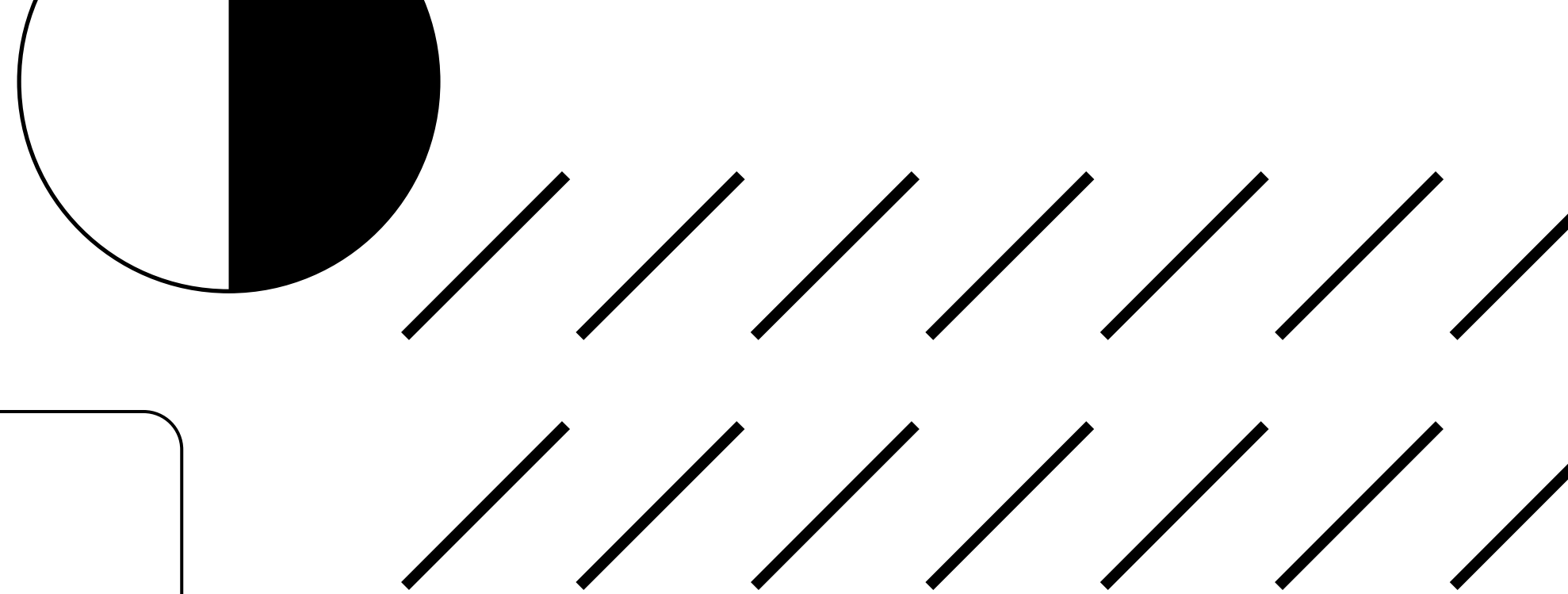
Eliminació d'atributs

03

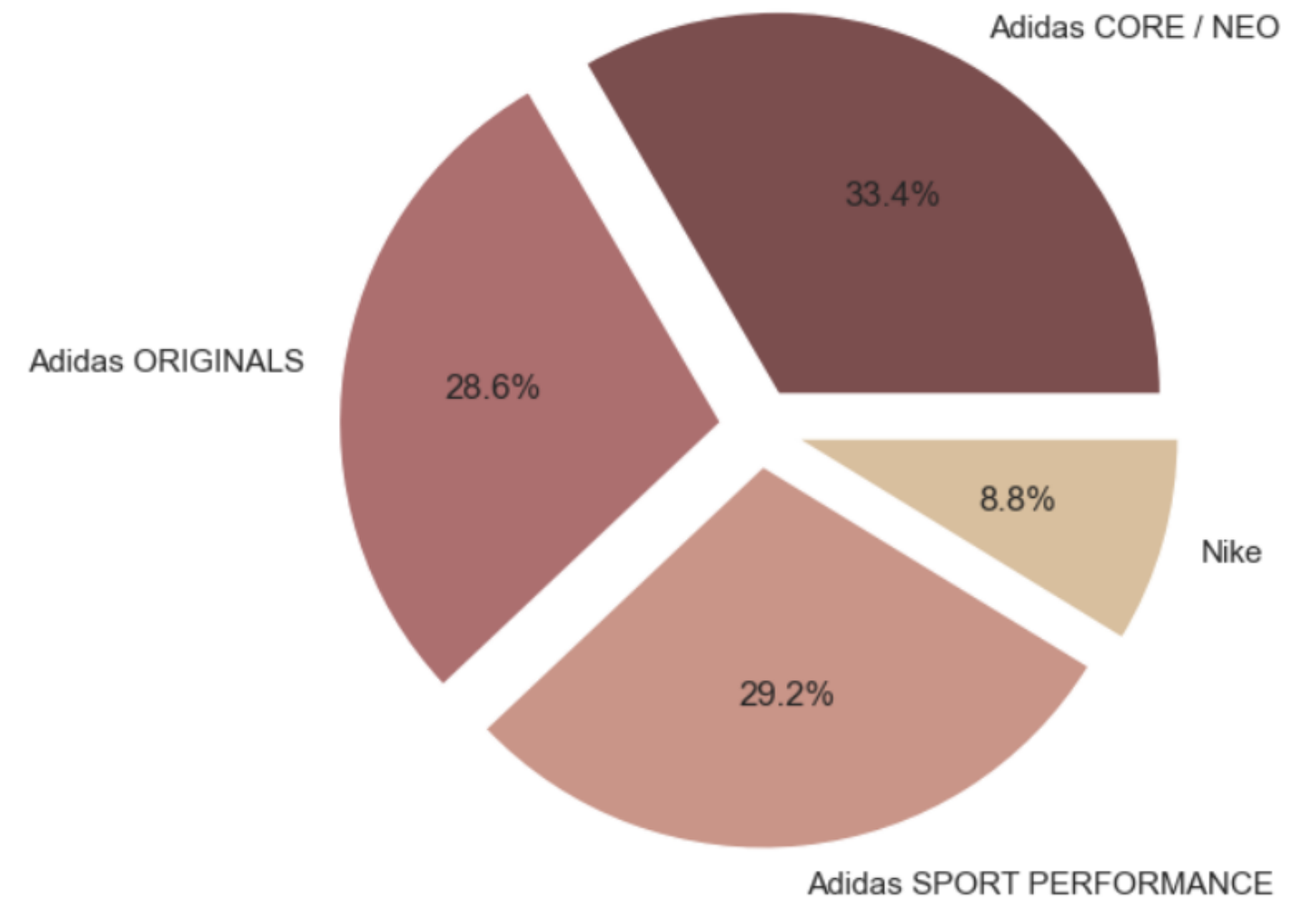
Estandardització

04

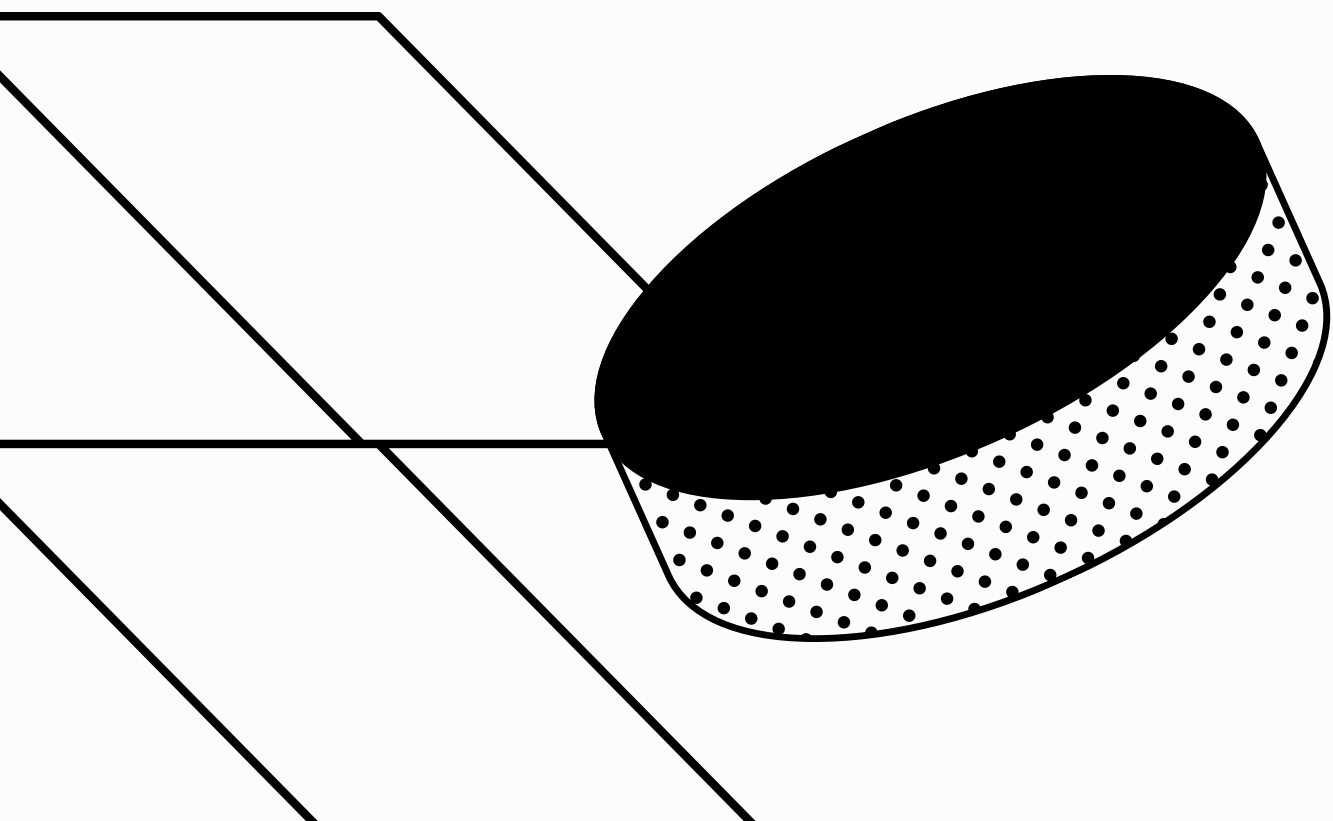
Balanceig de dades



Discounts offered by Brands



Prmiera Implementació dels Models



LogisticRegression	
Accuracy	Roc
0.689	0.808

SVM	
Accuracy	Roc
0.717	0.861

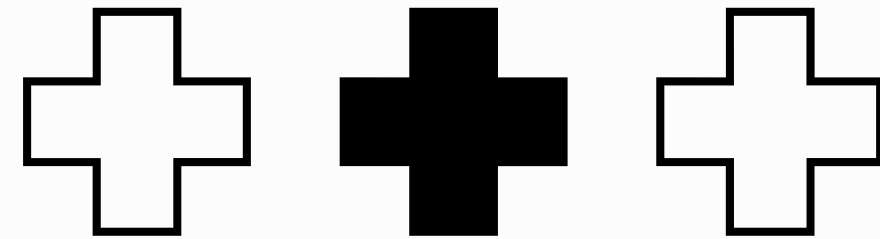
K-Nearest	
Accuracy	Roc
0.692	0.855

Decision Tree's	
Accuracy	Roc
0.667	0.771

Random Forest	
Accuracy	Roc
0.739	0.893

XGBoost	
Accuracy	Roc
0.755	0.903

Models de Clasificació Binària



RandomForest amb Oversampling

Accuracy: 0.969
Roc: 0.994
Cross validation: 0.980
Temps execució: 1 segons

RandomForest amb resampling

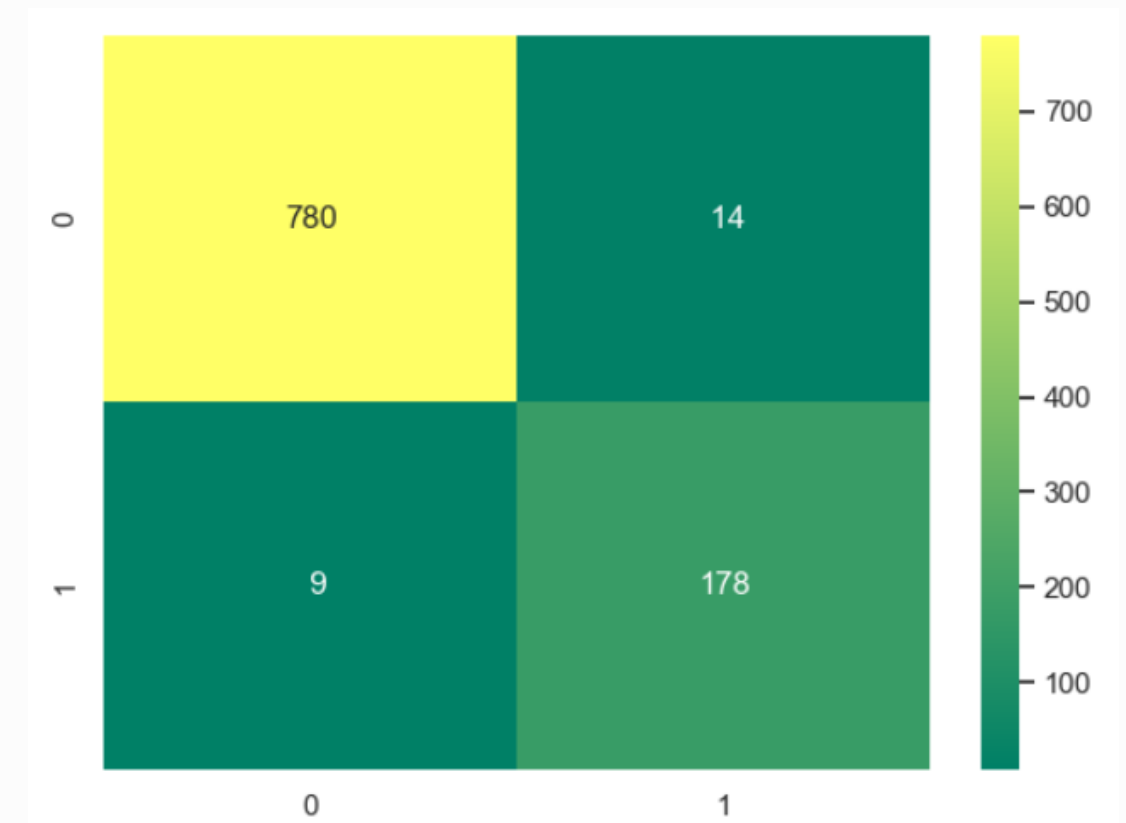
Accuracy: 0.963
Roc: 0.992
Cross validation: 0.967
Temps execució: 1.136 segons

XGBoost amb cerca d'hiperparametres

Accuracy: 0.978
Roc: 0.997
Cross validation: 0.969
Temps execució: 191.130 segons

XGBoost balancejat

Accuracy: 0.977
Roc: 0.996
Cross validation: 0.966
Temps execució: 0.379 segons



Conclusions

Després de fer aquesta anàlisi podem deduir que les millors formes de classificació són el RandomForest i el XGBoosti que la millor elecció per tractar i classificar el dataset ha sigut classificant només les marques Nike i Adidas i no cada submarca del dataset.

