

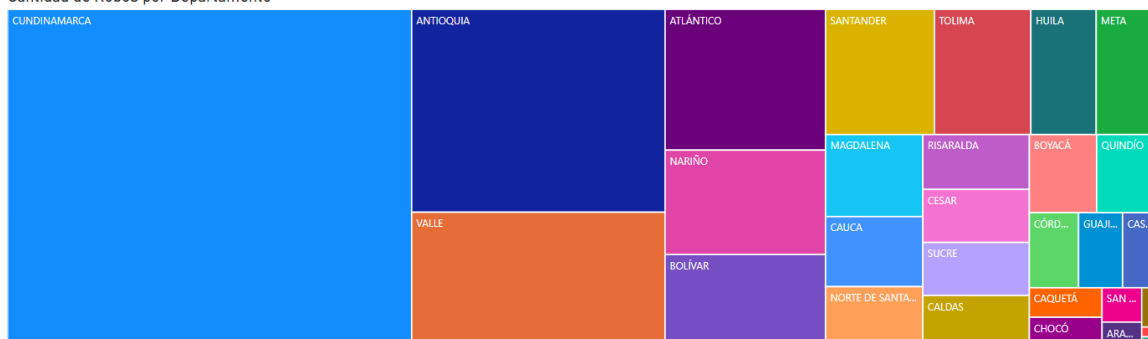
Prueba Científico de Datos

Fecha:

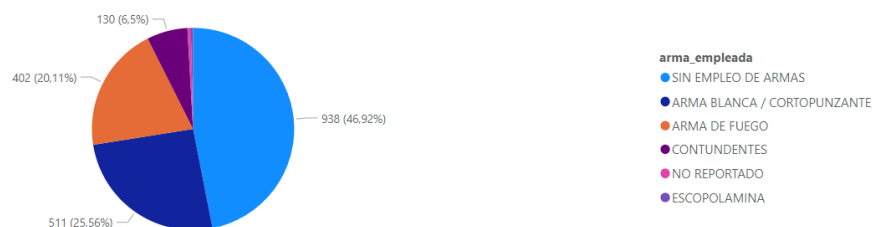
Nombre del candidato: CRISTIAN CARMONA

RETO 2. Power BI:

Cantidad de Robos por Departamento

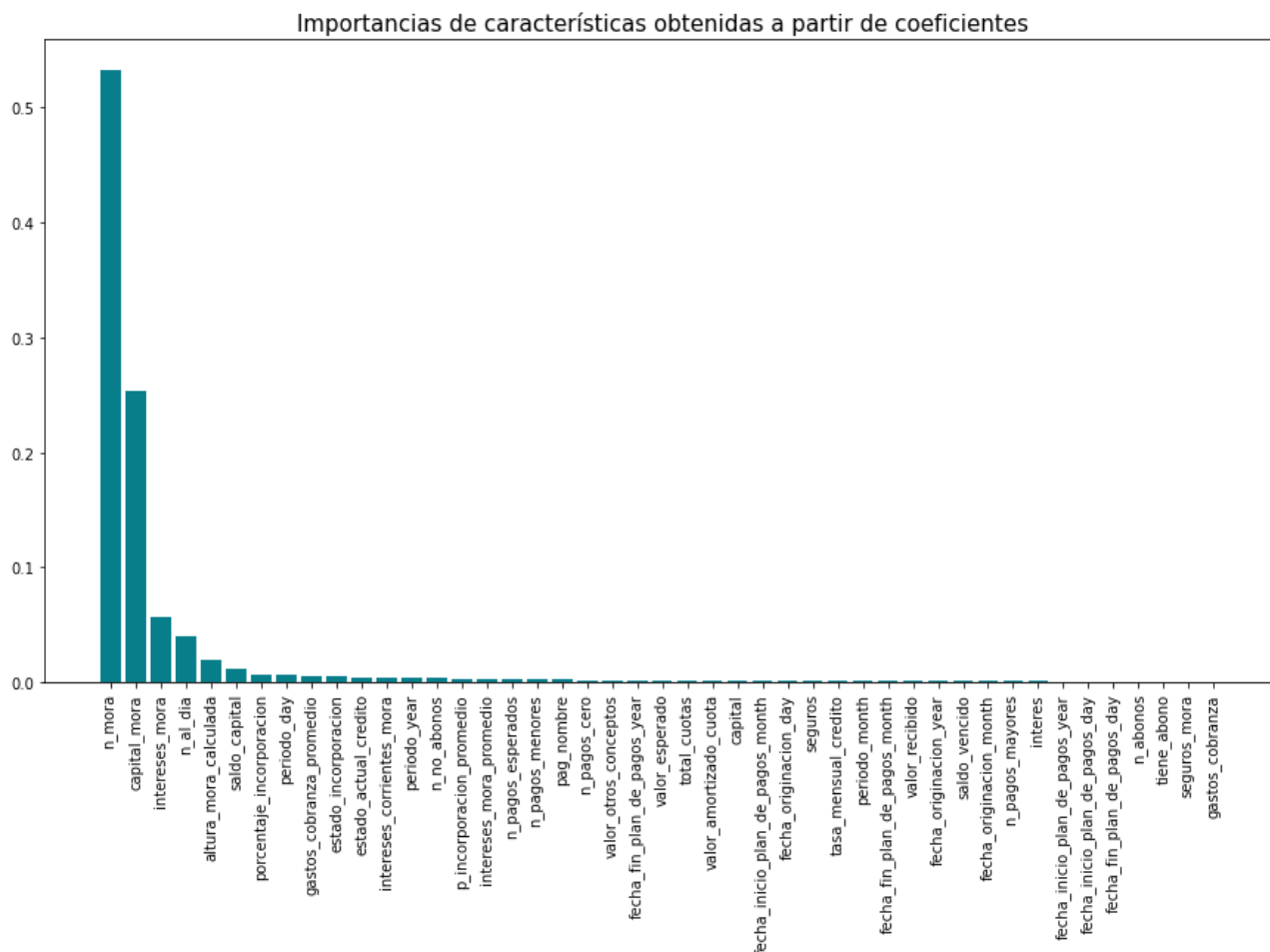


Cantidad de Robos por Arma



RETO 3. Python:

1. Utilizando el archivo remitido 'set_datos' estime un modelo predictivo que permita conocer:
 - Cuáles son las variables que más peso tienen sobre la probabilidad de que un cliente pague o no pague.



- Métricas de desempeño del modelo.

	Model	Training time	True Positives	True Negatives	False Positives	False Negatives	Accuracy	Kappa	Sensitivity	Specificity	Precision	f_1 Score
0	Decision Tree	0.136029	3994.0	4119.0	14.0	83.0	0.988185	0.976366	0.979642	0.996613	0.996613	0.988054
1	Random Forest	0.117019	3958.0	3903.0	230.0	119.0	0.957491	0.914993	0.970812	0.944350	0.944350	0.957398
2	AdaBoost	3.705842	4030.0	4120.0	13.0	47.0	0.992692	0.985382	0.988472	0.996855	0.996855	0.992646
3	Naive Bayes	0.041010	3509.0	1759.0	2374.0	568.0	0.641657	0.285424	0.860682	0.425599	0.425599	0.569557
4	QDA	0.095021	4006.0	3776.0	357.0	71.0	0.947868	0.895782	0.982585	0.913622	0.913622	0.946850

- Matriz de confusión.



- ¿Cómo llevarías el pipeline de ML de este modelo a producción? Solo comentar los pasos a seguir.
1. Definir un flujo de extracción de nuevos datos
 2. Pasar los datos por un proceso de limpieza
 3. Versionar la data con DVC
 4. Dividirla en sets de entrenamiento y testeo
 5. Desarrollar un pipeline de reentrenamiento
 6. Evaluar el modelo
 7. Empaquetar en formato pkl o onnx
 8. Desarrollar el API (transformar el modelo en servicio)
 9. Empaquetar el API con docker
 10. Monitorear modelo y desarrollar triggers de reentrenamiento.