

Entrega TP3

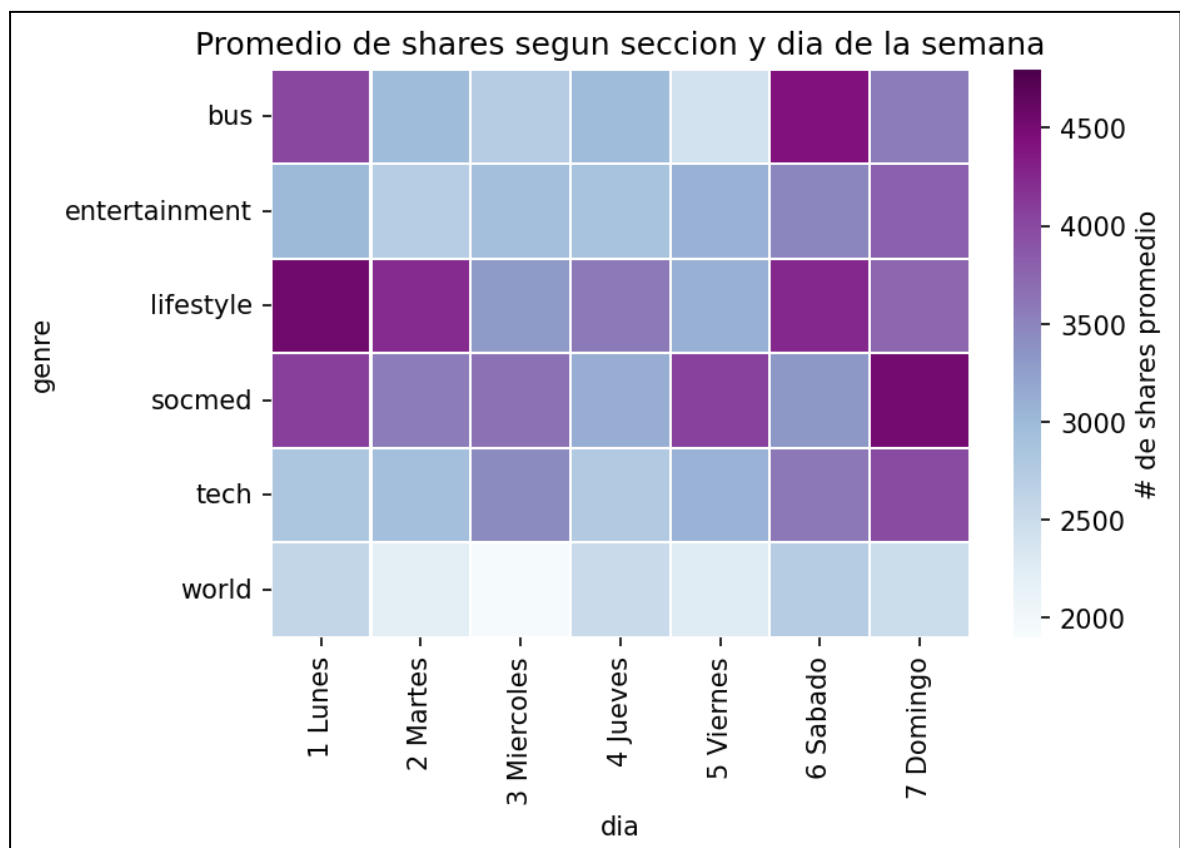
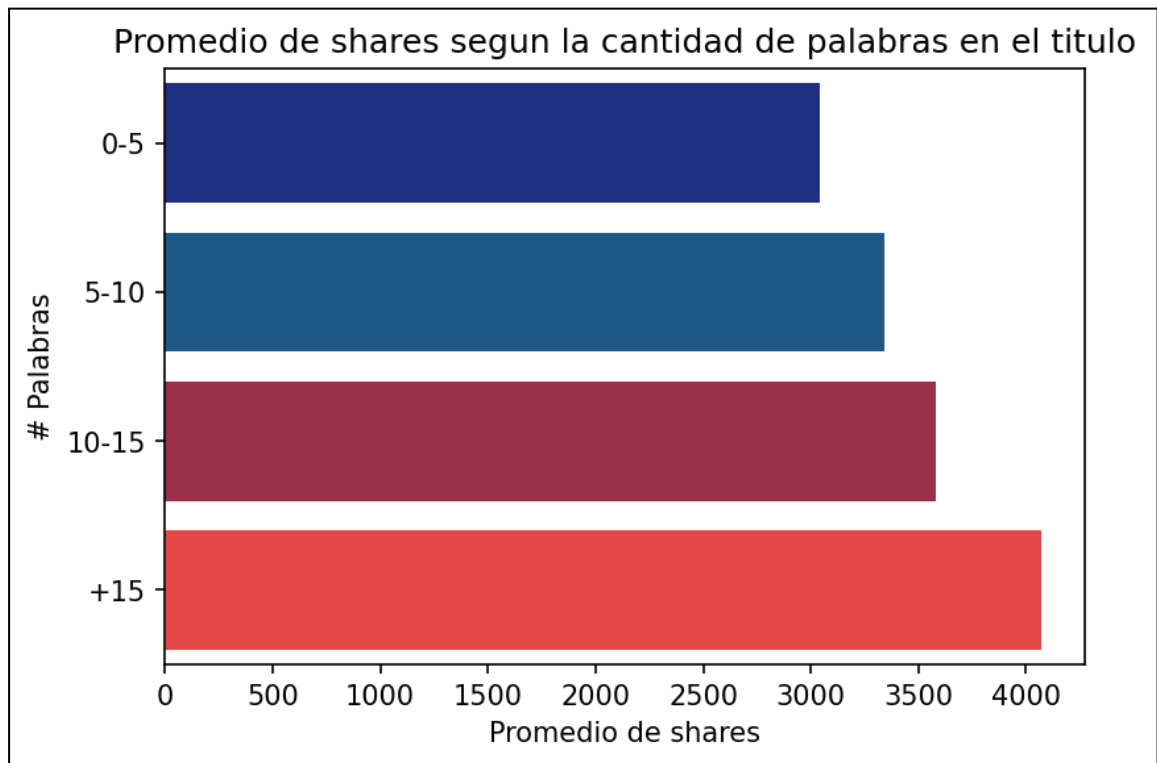
Padrón: 101589

Nombre completo: Juan Manuel Alvarez

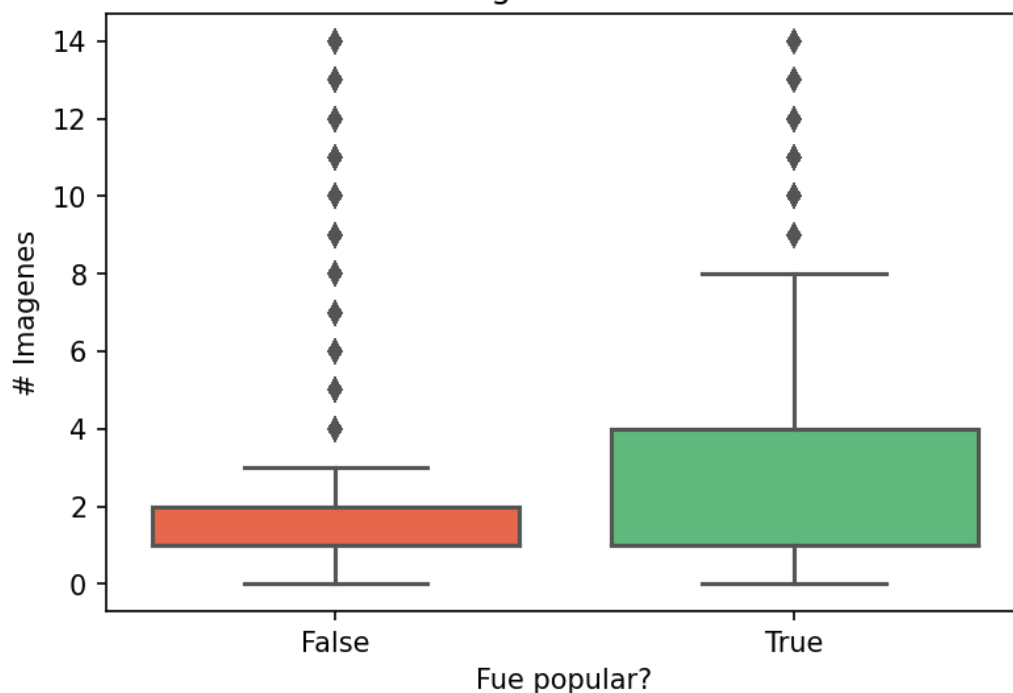
Corrector: Ignacio Brusati

Primera parte - Análisis exploratorio

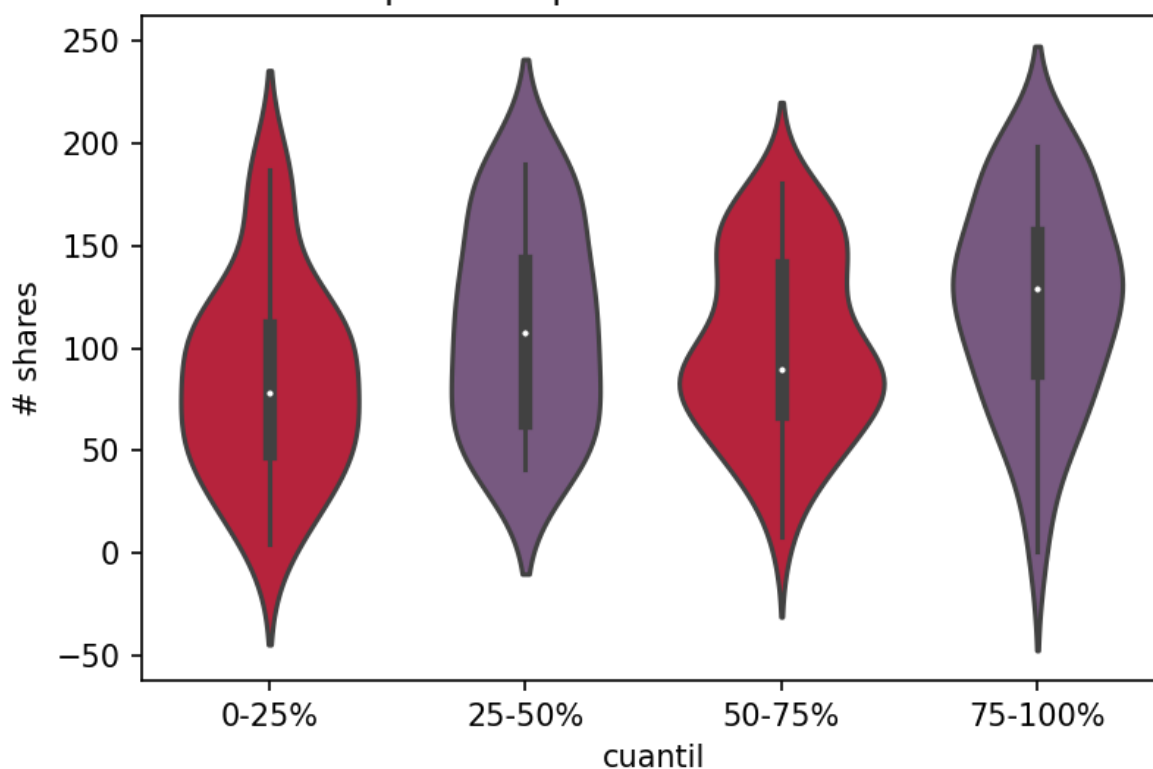
Link al notebook : [Análisis Exploratorio](#)



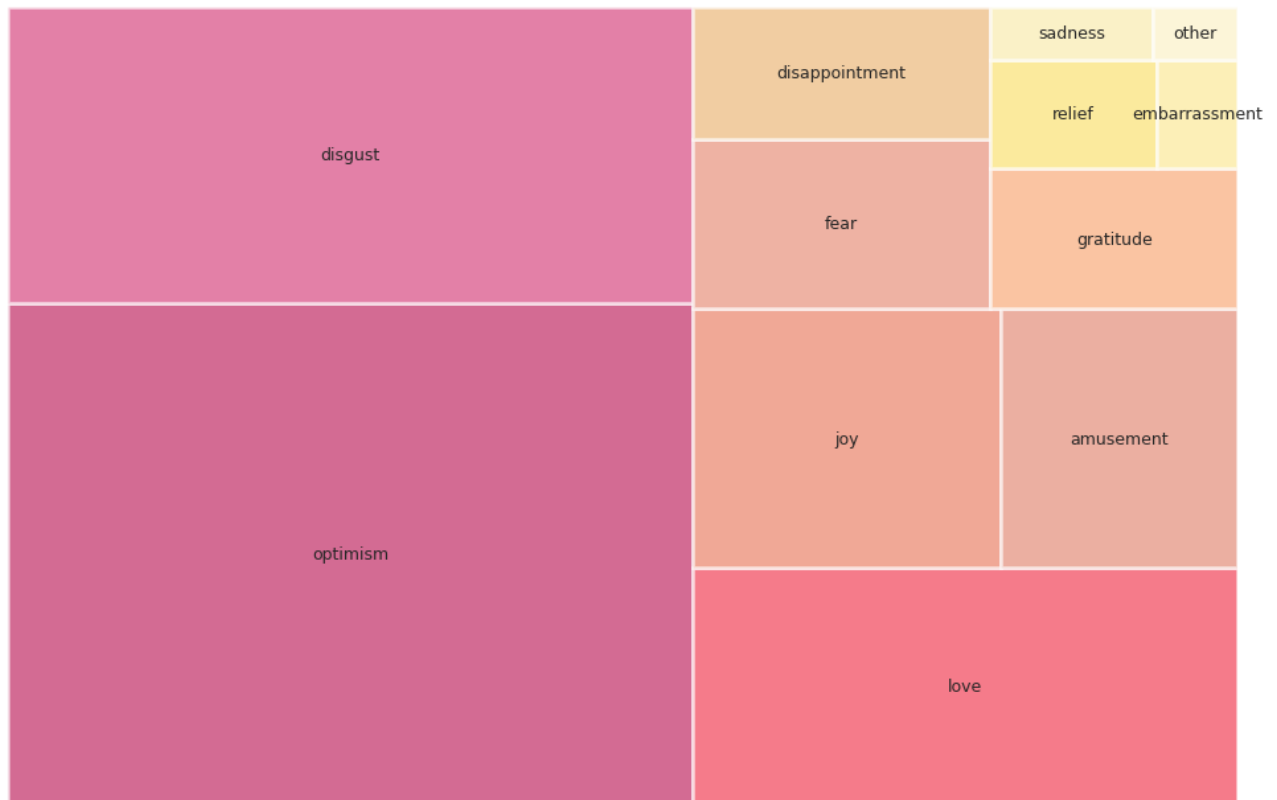
Distribucion de la cantidad de imagenes observando si el articulo fue popular



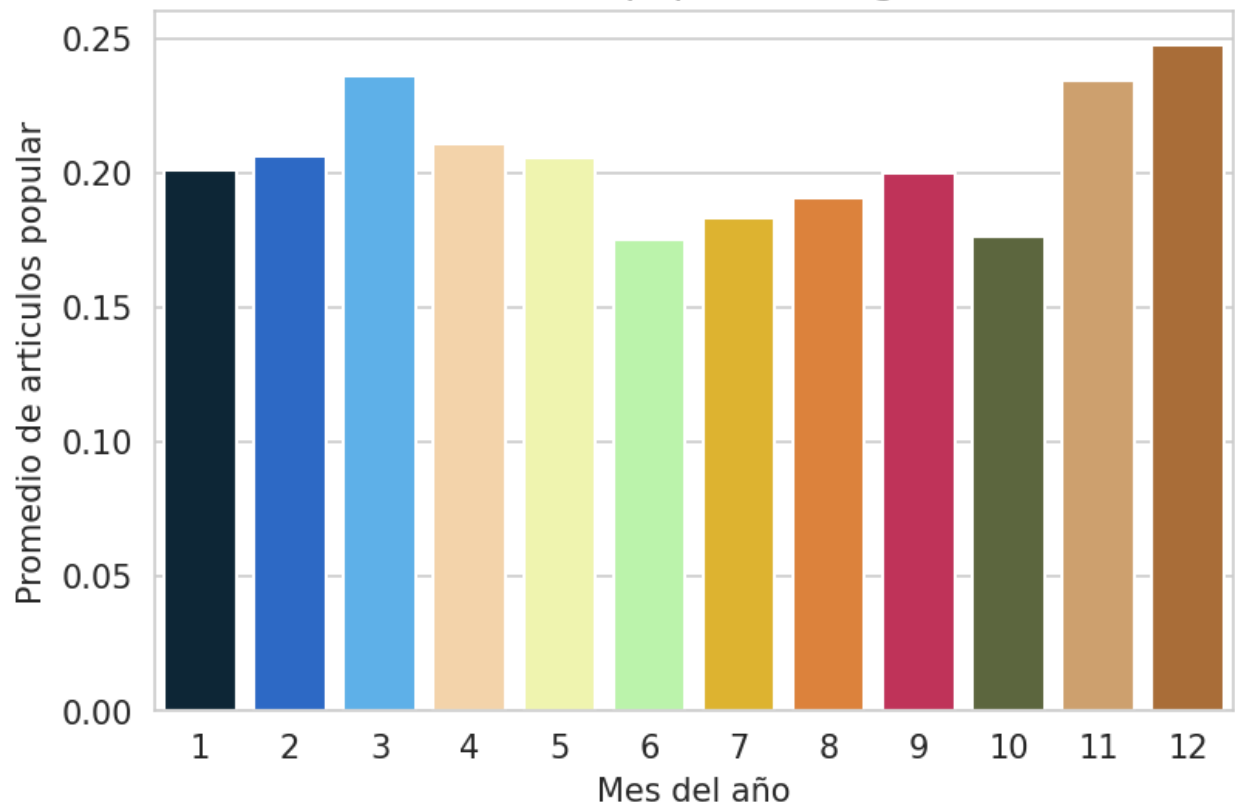
Separados en 4 cuantiles segun la cantidad de palabras, distribucion de los shares de cada cuantil a los articulos que no superan los 200 shares



Proporcion de las emociones de los articulos etiquetados



Promedio de articulos populares segun mes del año



Segunda Parte - Machine learning baseline

Link al notebook : [Regresión Logística](#)

¿Cuál es el mejor score de validación obtenido?

- 0.6845

Al predecir con este modelo para test, ¿Cuál es el score obtenido?

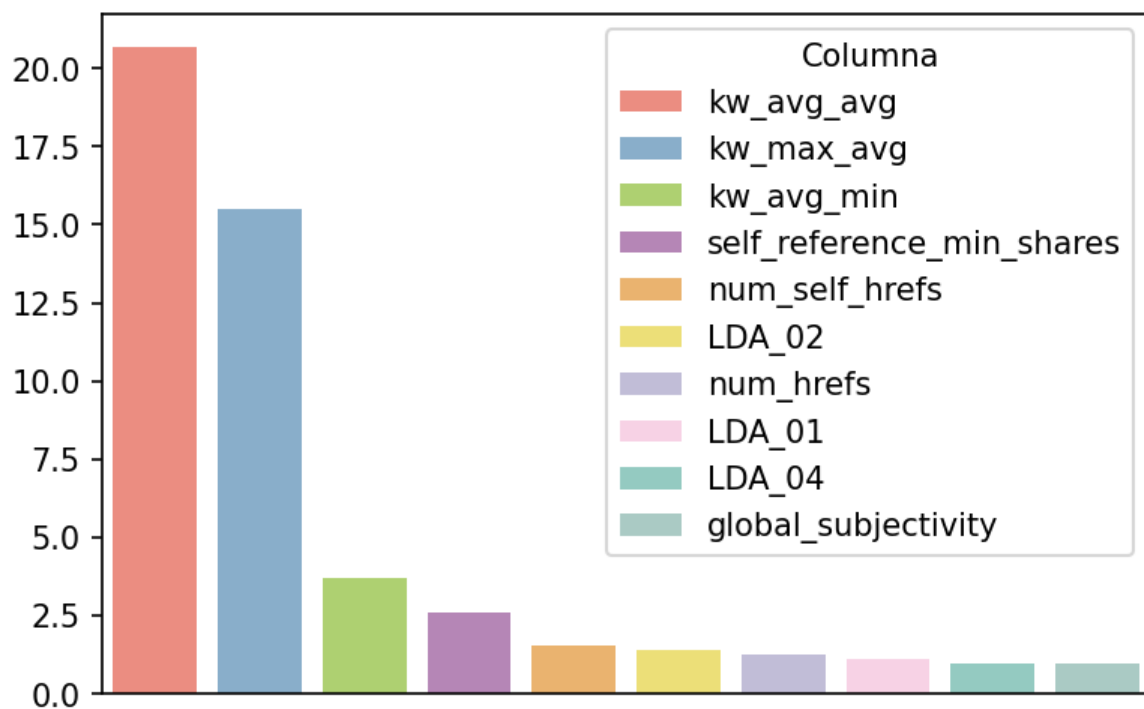
- 0.7100

CSV con el url y la probabilidad del modelo que sea popular

- [\[Link\]](#)

¿Qué features son los más importantes para predecir con el mejor modelo? Graficar

Coeficientes de las 10 features mas determinantes del dataframe



Tercera Parte - Machine learning

Link al notebook : [Random Forest](#)

Link al notebook : [XGBoost](#)

¿Cuál es el mejor de los dos?

- El score de validación fue superior en Random Forest [0.7341], aunque no hay diferencias significativas entre ambos

CSV con el url y la probabilidad del modelo que sea popular

- [CSV Random Forest](#)
- [CSV XGBoost](#)

Puntos Extra - Keras

Entrenar una red neuronal con Keras que sea reproducible, usando al menos 40 features y un score en validación superior a 0,6. Debe ser un modelo por separado a los propuestos, no necesita búsqueda de hiper parámetros ni cumplir otra condición. ¿Cuál es su score en validación y en test?

Link al notebook: [Keras](#)

Puntos Extra - Importancia features Parte III

Graficar la importancia de features para algún modelo de la parte III. ¿Qué tanto se parece a los features importantes de la parte II?

Link al notebook : [Random Forest](#)

