

Prototipo Super Básico recomendador de noticias

Este prototipo busca orientar de forma muy básica y describir algunas ideas que servirán de inspiración para el desarrollo de un producto mas robusto de un recomendador de noticias para clientes corporativos.

Paquetes Utiles

```
import pandas as pd
pd.options.mode.chained_assignment = None # default='warn'
import re
import json
import numpy as np

from ImpalaHelper.Impala_Helper import Helper

cache = {
    "connStr" : "DSN=impala-prod" , # Cambiar de ser necesario
    "db" : "proceso_cap_analit_y_gob_de_inf"
}

hp = Helper(cache)
```

Impala Helper Instanciado. Versión: 1.26.20220608

Importación información entregada

```
clientes = pd.read_csv("C:/Users/jucadiaz/Datos/Dataton_2022/clientes_prototipo.csv", sep = ',')
clientes.dtypes
clientes.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	Unnamed: 0	nit	nombre	cod_ciiu	desc_ciiu_division	desc_ciiu_grupo	desc_ciiuu_clase	subsec	nit_ciiuu
0	0	900215071	BANCO DE LAS MICROFINANZAS BANCAMIA SA	6412	ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO ...	INTERMEDIACION MONETARIA	BANCOS COMERCIALES	BANCOS	64
1	1	900378212	BANCO W SA	6412	ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO ...	INTERMEDIACION MONETARIA	BANCOS COMERCIALES	BANCOS	64
2	2	860051894	BANCO FINANDINA SA FINANDINA ESTABLECIMIENTO B...	6412	ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO ...	INTERMEDIACION MONETARIA	BANCOS COMERCIALES	BANCOS	64
3	3	860034594	SCOTIABANK COLPATRIA SA	6412	ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO ...	INTERMEDIACION MONETARIA	BANCOS COMERCIALES	BANCOS	64
4	4	900628110	BANCO SANTANDER DE NEGOCIOS COLOMBIA SA	6412	ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO ...	INTERMEDIACION MONETARIA	BANCOS COMERCIALES	BANCOS	64

```
clientes["nombre"] = clientes["nombre"].apply(lambda x: re.sub("\\ssa\\b|\\s\\tda?\\b", "",x.lower()))
clientes["nombre"]
#clientes.shape
```

```
0          banco de las microfinanzas bancamia
1                      banco w
2      banco finandina finandina establecimiento banca
3                      scotiabank colpatría
4      banco santander de negocios colombia
5                      citibank colombia citibank
6                      banco bbva colombia
7                      credibanco
8                      banco agrario de colombia
9                      banco serfinanza
10                     banco de bogota
11                     banco falabella
12      banco cooperativo coopcentral
13                     banco davivienda
14      banco interamericano de desarrollo bid s
15                     mufg bank
16                     banco de la republica
17                     banco gnb sudameris
18      itau corpbanca colombia
19                     banco popular
20                     mibanco
21      corporacion andina de fomento
22                     banco coomeva
23      banco comercial av villas
Name: nombre, dtype: object
```

```
np.unique(clientes[["desc_ciiu_division", "desc_ciiu_grupo","desc_ciiuu_clase"]].values)
```

```
array(['ACTIVIDADES DE SERVICIOS FINANCIEROS, EXCEPTO LAS DE SEGUROS Y DE PENSIONES',
      'BANCOS COMERCIALES', 'INTERMEDIACION MONETARIA'], dtype=object)
```

Idea Inspiracional #1:

- 1. tener una mejor limpieza de los nombre de negocio de cliente para que queden los nombres más comerciales, sería más pertinente
- 2. hacer una busqueda de las descripciones del codigo ciuu dentro de la noticia para determinar similitud con el sector

```
clie_news = pd.read_csv("C:/Users/jucadiaz/Datos/Dataton_2022/clientes_noticias_prototipo.csv", sep = ',')
clie_news.head()
#clie_news.dtypes
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	nit	news_url_absolute	news_init_date	news_final_date	new_id
0	900378212	https://www.bluradio.com/economia/precio-dolar...	2022-07-30	2022-08-14	new86016
1	900378212	https://www.semana.com/economia/macroeconomia/...	2022-07-30	2022-08-14	new49153
2	900378212	https://elcomercio.pe/respuestas/que/gustavo-p...	2022-07-30	2022-08-14	new38914
3	860051135	https://www.lanacion.com.ar/estados-unidos/cua...	2022-07-30	2022-08-14	new67592
4	860051135	https://www.lanacion.com.ar/estados-unidos/com...	2022-07-30	2022-08-14	new45065

```
data_news = pd.read_csv("C:/Users/jucadiaz/Datos/Dataton_2022/noticias_prototipo.csv", sep = ',')
data_news.head()
#data_news.dtypes
#data_news.columns
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	news_url_absolute	news_init_date	news_final_date	news_title	news_text_content	new_id
0	https://www.bluradio.com/economia/precio-dolar...	2022-07-30	2022-08-14	Precio dolar hoy: la cotizacion de la divisa a...	Este martes, 2 de agosto, el dolar alcanzo un ...	new86016
1	https://www.semana.com/economia/macroeconomia/...	2022-07-30	2022-08-14	Es cierto: El presidente Gustavo Petro quiere ...	No hay nada mas permanente que lo temporal y a...	new49153
2	https://elcomercio.pe/respuestas/que/gustavo-p...	2022-07-30	2022-08-14	Gustavo Petro: ¿Que dice el informe final de s...	El equipo de empalme del gobierno del presiden...	new38914
3	https://www.lanacion.com.ar/estados-unidos/cua...	2022-07-30	2022-08-14	Cuanto cuesta y como enviar dinero de Estados ...	¿Buscas opciones para el envio de dinero de Es...	new67592
4	https://www.lanacion.com.ar/estados-unidos/com...	2022-07-30	2022-08-14	Como enviar dinero de Estados Unidos a otros p...	¿Buscas opciones para el envio de dinero de Es...	new45065

```
data_news[['news_title', 'news_text_content']] = data_news[['news_title', 'news_text_content']].astype(str)
```

```
with open("C:/Users/jucadiaz/GIT/Dataton-BC/2022/src/data/auxiliary/diccionario_categorias.json") as file:
    dict_categoria = json.load(file)

dict_categoria
```

```
{'alianzas': ['aliado',
'complemento',
'coalicion',
'confederacion',
'federacion',
'union',
'liga',
'asociacion',
'pacto',
'convenio',
'convencion',
'acuerdo',
'trato',
'concierto',
'conexion',
'parentesco',
'enlace',
'lazo',
'vinculo'],
'innovacion': ['innovacion',
'creacion',
'novedad',
'invento',
'descubrimiento',
'adelanto',
'desarrollo',
```

```
'nuevo',
'perfeccionamiento',
'mejora',
'optimizacion',
'mejoria',
'actualizacion',
'modernizacion',
'novedoso'],
'macroeconomicas': ['macroeconomia',
'economia',
'interes',
'activo',
'ahorro',
'alza',
'amortizacion',
'auditoria',
'banco',
'banco de la republica',
'beneficiario',
'capital',
'canje',
'deposito',
'ipc',
'dolar',
'trm',
'euro',
'inflacion',
'tasa'],
'regulaciones': ['regular',
'ley',
'reglamentacion',
'ordenamiento',
'ordenacion',
'regularizacion',
'organizacion',
'normalizacion',
'reajuste',
'ajuste',
'regla',
'norma',
'precepto',
'decreto',
'dictamen',
'disposicion',
'mandato',
'orden',
'obligacion',
'estatuto',
'legislacion',
'constitucion'],
'reputacion': ['reputacion',
'conocido',
'prestigio',
'estima',
'fama',
'renombre',
'notoriedad',
'celebridad',
'nombradia',
'popularidad',
'imagen'],
'sostenibilidad': ['sostenible',
'ambiente',
'entorno',
'clima',
'habitat',
'ambiental',
'ambientalista',
'ecolÃ³gico',
'ambientalismo',
'biodiversidad',
'bioenergia',
'biomasa',
'calentamiento global',
'cambio climatico',
'contaminacion',
'decibelio',
```

```
'deforestacion',
'desarrollo sostenible',
'ecologia',
'erosion',
'lluvia',
'medio ambiente',
'reciclaje',
'recurso',
'recurso natural',
'residuos',
'reutilizacion']}]}
```

Ideas Inspiracionales #2:

1. tener un diccionario más pertinente y con mayores términos podríamos encontrar mayores términos
2. tener ponderaciones de las palabras
3. utilizar stemming o lemmatization de palabras

Conteos de palabras: cliente, sector, ciu

```
data_result = pd.DataFrame()

column_select = ['nit', 'new_id', 'news_url_absolute', 'news_init_date',
'news_final_date', 'news_title', 'count_name_tittle',
'count_name_text', 'count_sect_tittle', 'count_sect_text', 'count_ciu_tittle', 'count_ciu_text'
]

for j, i in enumerate(clientes["nit"]):
    clie_news_filter = clie_news[clie_news["nit"]==i][["new_id", "nit"]]
    data_filter = pd.merge(data_news, clie_news_filter, on= "new_id")
    nombre = clientes[clientes["nit"]==i][["nombre"].iloc[0]]
    sector = clientes[clientes["nit"]==i][["subsec"].iloc[0]]
    ciu = clientes[clientes["nit"]==i][["desc_ciu_grupo"].iloc[0]]

    data_filter["count_name_tittle"] = data_filter["news_title"].apply(lambda x: len(re.findall(nombre, x, re.IGNORECASE)))
    data_filter["count_name_text"] = data_filter["news_text_content"].apply(lambda x: len(re.findall(nombre, x, re.IGNORECASE)))
    data_filter["count_sect_tittle"] = data_filter["news_title"].apply(lambda x: len(re.findall(sector, x, re.IGNORECASE)))
    data_filter["count_sect_text"] = data_filter["news_text_content"].apply(lambda x: len(re.findall(sector, x, re.IGNORECASE)))
    data_filter["count_ciu_tittle"] = data_filter["news_title"].apply(lambda x: len(re.findall(ciu, x, re.IGNORECASE)))
    data_filter["count_ciu_text"] = data_filter["news_text_content"].apply(lambda x: len(re.findall(ciu, x, re.IGNORECASE)))

    data_result = pd.concat([data_result, data_filter[column_select]], ignore_index=True)
    print("Termine de contar el cliente:" + str(i) + " avance " + str((j + 1)/len(clientes["nit"])))
```

```
Termine de contar el cliente:900215071 avance 0.041666666666666664
Termine de contar el cliente:900378212 avance 0.08333333333333333
Termine de contar el cliente:860051894 avance 0.125
Termine de contar el cliente:860034594 avance 0.16666666666666666
Termine de contar el cliente:900628110 avance 0.20833333333333334
Termine de contar el cliente:860051135 avance 0.25
Termine de contar el cliente:860003020 avance 0.29166666666666667
Termine de contar el cliente:860032909 avance 0.3333333333333333
Termine de contar el cliente:800037800 avance 0.375
Termine de contar el cliente:860043186 avance 0.41666666666666667
Termine de contar el cliente:860002964 avance 0.4583333333333333
Termine de contar el cliente:900047981 avance 0.5
Termine de contar el cliente:890203088 avance 0.5416666666666666
Termine de contar el cliente:860034313 avance 0.5833333333333334
Termine de contar el cliente:800091063 avance 0.625
Termine de contar el cliente:900381786 avance 0.6666666666666666
Termine de contar el cliente:860005216 avance 0.7083333333333334
Termine de contar el cliente:860050750 avance 0.75
Termine de contar el cliente:890903937 avance 0.7916666666666666
Termine de contar el cliente:860007738 avance 0.8333333333333334
Termine de contar el cliente:860025971 avance 0.875
Termine de contar el cliente:860054041 avance 0.9166666666666666
Termine de contar el cliente:900406150 avance 0.9583333333333334
Termine de contar el cliente:860035827 avance 1.0
```

```
data_result.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	nit	new_id	news_url_absolute	news_init_date	news_final_date	news_title	count_name_title	co
0	900215071	new96404	https://www.eltiempo.com/economia/finanzas-per...	2022-07-30	2022-08-14	?En que consiste la banca grupal que comienza ...	0	0
1	900215071	new70000	https://www.laopinion.com.co/innovaciones/entr...	2022-07-15	2022-07-30	Entregan conectividad gratuita a 500 jovenes e...	0	0
2	900215071	new43447	https://www.bbva.com/es/sostenibilidad/por-pri...	2022-07-15	2022-07-30	Por primera vez una entidad privada mide la po...	0	0
3	900215071	new12728	https://www.laopinion.com.co/economia/42-de-lo...	2022-07-15	2022-07-30	42% de los emprendedores solo tiene educacion ...	0	0
4	900215071	new45496	https://www.lapatria.com/economia/el-172-de-lo...	2022-07-15	2022-07-30	El 17,2% de los clientes de Bancamia estan en ...	0	0

Conteos de palabras: cliente, sector, ciuu

```
data_news2 = data_news.copy()

for i in dict_categoria:
    regular_exp = "\\b" + '(s)?(es)?\\b|\\b'.join(list(dict_categoria[i])) + "(s)?(es)?\\b"
    nom_colum1 = "count_" + i + "_tittle"
    nom_colum2 = "count_" + i + "_text"

    data_news2[nom_colum1] = data_news2["news_title"].apply(lambda x: len(re.findall(regular_exp, x, re.IGNORECASE)))
    data_news2[nom_colum2] = data_news2["news_text_content"].apply(lambda x: len(re.findall(regular_exp, x, re.IGNORECASE)))
```

```
regular_exp
```

```
'\\bregular(s)?(es)?\\b|\\bley(s)?(es)?\\b|\\breglamentacion(s)?(es)?\\b|\\bbordenamiento(s)?(es)?\\b|\\bbordenacion(s)?(es)?\\b|\\bbreglarizacion(s)?(es)?\\b|\\bborganizacion(s)?(es)?\\b|\\bbnormalizacion(s)?(es)?\\b|\\bbreaajuste(s)?(es)?\\b|\\bbajuste(s)?(es)?\\b|\\bbregla(s)?(es)?\\b|\\bbnorma(s)?(es)?\\b|\\bbprecepto(s)?(es)?\\b|\\bbdecreto(s)?(es)?\\b|\\bbdictamen(s)?(es)?\\b|\\bbdisposicion(s)?(es)?\\b|\\bbmandato(s)?(es)?\\b|\\bborden(s)?(es)?\\b|\\bbobligacion(s)?(es)?\\b|\\bbestatuto(s)?(es)?\\b|\\bblegislacion(s)?(es)?\\b|\\bbconstitucion(s)?(es)?\\b'
```

```
columnas2 = ['new_id',
             'count_macroeconomicas_tittle', 'count_macroeconomicas_text', 'count_sostenibilidad_tittle',
             'count_sostenibilidad_text', 'count_innovacion_tittle',
             'count_innovacion_text', 'count_regulaciones_tittle',
```

```
        'count_regulaciones_text', 'count_alianzas_tittle',
        'count_alianzas_text', 'count_reputacion_tittle',
        'count_reputacion_text']
data_news2[columnas2]
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	new_id	count_macroeconomicas_tittle	count_macroeconomicas_text	count_sostenibilidad_tittle	count_sostenibilidad_text	count_innovaci
0	new86016	1	7	0	0	0
1	new49153	0	5	0	0	0
2	new38914	0	0	0	1	0
3	new67592	0	18	0	1	0
4	new45065	0	18	0	1	0
...
773	new94186	0	2	0	0	0
774	new16364	1	3	0	0	0
775	new18413	0	2574	0	159	0
776	new36846	0	0	0	0	0
777	new49138	0	3	0	0	0

778 rows × 13 columns

```
data_result2 = pd.merge(clie_news[["nit", "new_id"]], data_news2[columnas2], on = "new_id", how = "left")
data_result_final_aux = pd.merge(data_result, data_result2, on = ["nit", "new_id"])
data_result_final = pd.merge(clientes[["nit", "nombre"]], data_result_final_aux, on = "nit")
data_result_final["fuente"] = data_result_final["news_url_absolute"].apply(lambda x: str(re.findall('https?:\/\/.+?/', x,
re.IGNORECASE)[0]))
data_result_final.head()
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	nit	nombre	new_id	news_url_absolute	news_init_date	news_final_date	news_title	count_i
0	900215071	banco de las microfinanzas bancamia	new96404	https://www.eltiempo.com/economia/finanzas-per...	2022-07-30	2022-08-14	?En que consiste la banca grupal que comienza ...	0
1	900215071	banco de las microfinanzas bancamia	new70000	https://www.laopinion.com.co/innovaciones/entr...	2022-07-15	2022-07-30	Entregan conectividad gratuita a 500 jovenes e...	0
2	900215071	banco de las microfinanzas bancamia	new43447	https://www.bbva.com/es/sostenibilidad/por-pri...	2022-07-15	2022-07-30	Por primera vez una entidad privada mide la po...	0

	nit	nombre	new_id	news_url_absolute	news_init_date	news_final_date	news_title	count_i
3	900215071	banco de las microfinanzas bancamia	new12728	https://www.laopinion.com.co/economia/42-de-lo...	2022-07-15	2022-07-30	42% de los emprendedores solo tiene educacion ...	0
4	900215071	banco de las microfinanzas bancamia	new45496	https://www.lapatia.com/economia/el-172-de-lo...	2022-07-15	2022-07-30	El 17,2% de los clientes de Bancamia estan en ...	0

5 rows × 26 columns

Ideas Inspiracionales #3:

1. revisar fuentes que no sean colombianas por sus extensiones de URL
2. revisar categorizaciones por la URL de la noticia

```
with open("C:/Users/jucadiaz/opcionesServer.json") as file:
    opt_srvr = json.load(file)

tablename = 'proceso_cap_analit_y_gob_de_inf.Dataton22_Base_conteo_prototipo'

hp.fromPandasDF(data_result_final , tablename , serverOpts = opt_srvr)
```

```
c:\users\jucadiaz\virtual_env\test_python3_5\lib\site-packages\paramiko\transport.py:33: CryptographyDeprecationWarning: Python
3.5 support will be dropped in the next release of cryptography. Please upgrade your Python.
  from cryptography.hazmat.backends import default_backend
c:\users\jucadiaz\virtual_env\test_python3_5\lib\site-packages\pysftp\__init__.py:61: UserWarning: Failed to load HostKeys from
C:\Users\jucadiaz\.ssh\known_hosts. You will need to explicitly load HostKeys (cnopts.hostkeys.load(filename)) or
disableHostKey checking (cnopts.hostkeys = None).
  warnings.warn(wmsg, UserWarning)
```

```
SFTP Instanciado
Enviando file.parq
Archivo Enviado, 0.18s
HDFS Instanciado, usando conexión remota. Versión: 1.2.20220118
Ejecutando Comando: kinit jucadiaz
Comando ejecutado correctamente
Ejecutando Comando: hdfs dfs -rm -r -skipTrash /user/jucadiaz/proceso_cap_analit_y_gob_de_inf.Dataton22_Base_conteo_prototipo/
Comando ejecutado correctamente
Ejecutando Comando: hdfs dfs -mkdir /user/jucadiaz/proceso_cap_analit_y_gob_de_inf.Dataton22_Base_conteo_prototipo/
Comando ejecutado correctamente
Ejecutando Comando: hdfs dfs -put file.parq
/user/jucadiaz/proceso_cap_analit_y_gob_de_inf.Dataton22_Base_conteo_prototipo/file.parq
Comando ejecutado correctamente
Archivo file.parq Eliminado, 0.03s
Cerrando SFTP
Transcurrido: 1663797432, Tiempo de Refresco = 300
Refrescando Conexión
Se ha cargado el DataFrame satisfactoriamente
```

```
rutaArchivo = 'C:/Users/jucadiaz/GIT/Dataton-BC/2022/src/modelo_recomendacion_SQL.sql'

hp.executeFile( rutaArchivo )
```

```
Ejecutando Archivo: C:/Users/jucadiaz/GIT/Dataton-BC/2022/src/modelo_recomendacion_SQL.sql
Transcurrido: 3765, Tiempo de Refresco = 300
Refrescando Conexión
Cerrando Conexión
Duración de Archivo (s): 74
```

Fase de indicadores categóricos

creo unos indicadores claves para la categorización y participación de acuerdo a los conteos

```
query = """select *
from proceso.Dataton22_tabla_Categorias"""

query1 = hp.getDataFrame(query)
query1.head(10)

.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	nit	nombre	new_id	news_url_absolute	news_init_date	news_final_date	news_title	co
0	900215071	banco de las microfinanzas bancamia	new96404	https://www.eltiempo.com/economia/finanzas-per...	2022-07-30	2022-08-14	?En que consiste la banca grupal que comienza ...	0
1	900215071	banco de las microfinanzas bancamia	new70000	https://www.laopinion.com.co/innovaciones/entr...	2022-07-15	2022-07-30	Entregan conectividad gratuita a 500 jovenes e...	0
2	900215071	banco de las microfinanzas bancamia	new43447	https://www.bbva.com/es/sostenibilidad/por-pri...	2022-07-15	2022-07-30	Por primera vez una entidad privada mide la po...	0
3	900215071	banco de las microfinanzas bancamia	new12728	https://www.laopinion.com.co/economia/42-de-lo...	2022-07-15	2022-07-30	42% de los emprendedores solo tiene educacion ...	0
4	900215071	banco de las microfinanzas bancamia	new45496	https://www.lapatria.com/economia/el-172-de-lo...	2022-07-15	2022-07-30	El 17,2% de los clientes de Bancamia estan en ...	0
5	900215071	banco de las microfinanzas bancamia	new43451	https://www.semana.com/economia/emprendimiento...	2022-07-15	2022-07-30	Los retos de emprender bajo el umbral de pobre...	0
6	900215071	banco de las microfinanzas bancamia	new26558	https://www.semana.com/tecnologia/articulo/fin...	2022-07-15	2022-07-30	?Finanzas abiertas en Colombia?	0
7	900215071	banco de las microfinanzas bancamia	new32707	https://www.portafolio.co/economia/finanzas/ta...	2022-07-15	2022-07-30	Tasas de captacion en los bancos superaron el ...	0
8	900215071	banco de las microfinanzas bancamia	new38856	https://www.pulzo.com/economia/lulo-bank-quita...	2022-07-15	2022-07-30	Primer banco en Colombia que devolvera el 4x1....	0
9	900215071	banco de las microfinanzas bancamia	new73672	https://cambiocolombia.com/articulo/economia/h...	2022-07-15	2022-07-30	?Hace compras en dolares con Visa y Mastercard...	0

10 rows × 41 columns

Entregable 1: Categorización

```
query = """select *
from proceso_cap_analit_y_gob_de_inf.Dataton22_entregable_1"""

query2 = hp.getDataFrame(query)
query2.head(15)
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	nombre_equipo	nit	new_id	participacion	categoria
0	dataton_owners	800037800	new27713	No_aplica	Descartable
1	dataton_owners	900215071	new49138	No_aplica	Descartable
2	dataton_owners	860032909	new89165	No_aplica	Descartable
3	dataton_owners	860035827	new18781	No_aplica	Descartable
4	dataton_owners	890203088	new58708	Cliente	Macroeconomia
5	dataton_owners	860051135	new19378	Sector	Macroeconomia
6	dataton_owners	900406150	new92490	No_aplica	Descartable
7	dataton_owners	900378212	new90858	No_aplica	Descartable
8	dataton_owners	860051135	new55361	No_aplica	Descartable
9	dataton_owners	860002964	new18271	Cliente	Macroeconomia
10	dataton_owners	860005216	new38520	No_aplica	Descartable
11	dataton_owners	860034594	new72522	Cliente	Macroeconomia
12	dataton_owners	860034313	new36193	No_aplica	Descartable
13	dataton_owners	900378212	new27434	No_aplica	Descartable
14	dataton_owners	860005216	new99978	No_aplica	Descartable

```
query2.to_csv("C:/Users/jucadiaz/GIT/Dataton-BC/2022/src/data/output/categorizacion.csv", index = False)
```

Armando un Score de recomendación

```
query = """
select
nit,
nombre,
new_id,
fuente,
news_title,
participacion,
categoria,
rank_fiabilidad_fuente,
ind_reciente,
peso_participacion,
peso_categoria,
score1,
peso_intra_categoria,
recomendacion
from proceso.Dataton22_tabla_recomendacion
where recomendacion <= 3
order by nit, recomendacion
limit 300
"""

query3 = hp.getDataFrame(query)
query3.head(10)
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	nit	nombre	new_id	fuelle	news_title	participacion	categoria	rank_fiabilidad_fuent
0	800037800	banco agrario de colombia	new71462	https://www.larepublica.co/	?Esta o no descapitalizado el Banco Agrario? E...	Cliente	Macroeconomia	0.994819
1	800037800	banco agrario de colombia	new69416	https://confidencialcolombia.com/	Listo el convenio FEDEGAN - Banco Agrario de C...	Cliente	Macroeconomia	0.694301
2	800037800	banco agrario de colombia	new31939	https://www.contextoganadero.com/	Agendese en la Puerta de Oro el campo en Agroexpo	Cliente	Innovacion	0.000000
3	800091063	banco interamericano de desarrollo bid s	new80450	https://www.cronista.com/	Una a una, las fuentes de dolares que contabil...	Sector	Macroeconomia	0.829016
4	800091063	banco interamericano de desarrollo bid s	new76362	https://elpais.com/	La ministra de Economia de Argentina presenta ...	Sector	Macroeconomia	0.829016
5	800091063	banco interamericano de desarrollo bid s	new97666	https://www.computerweekly.com/	Se mantiene crecimiento de las fintech en Peru	Sector	Macroeconomia	0.544041
6	860002964	banco de bogota	new10748	https://www.eltiempo.com/	El Grupo Aval obtuvo ganancias por \$ 675.500 m...	Cliente	Macroeconomia	0.979275
7	860002964	banco de bogota	new26473	https://www.portafolio.co/	Grupo Aval gano \$2,4 billones en el primer sem...	Cliente	Macroeconomia	0.968912
8	860002964	banco de bogota	new18271	https://www.portafolio.co/	App Banca Movil del Banco de Bogota fue la mej...	Cliente	Macroeconomia	0.968912
9	860003020	banco bbva colombia	new76861	https://www.semana.com/	La inflacion de los alimentos, ? amenaza de ham...	Sector	Macroeconomia	1.000000

Ideas Inspiracionales #4:

1. el sistema de score no es fiable, realizar una mejor version con modelos de ML
2. realizar un sistema de recomendacion mas robusto a traves de algoritmos de ML y modelos no supervisados

Entregable 2: Recomendación

```
query = """select *
from proceso_cap_analit_y_gob_de_inf.Dataton22_entregable_2"""
```

```
query4 = hp.getDataFrame(query)
query4.head(15)
```

Transcurrido: 319, Tiempo de Refresco = 300
Refrescando Conexión
Cerrando Conexión

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	nombre_equipo	nit	new_id	participacion	categoria	recomendacion
0	dataton_owners	860003020	new87797	Sector	Macroeconomia	6
1	dataton_owners	860034594	new60234	Cliente	Alianzas	27
2	dataton_owners	860003020	new76861	Sector	Macroeconomia	1
3	dataton_owners	860005216	new50679	Sector	Macroeconomia	35
4	dataton_owners	860034594	new43432	Cliente	Macroeconomia	4
5	dataton_owners	860050750	new59581	Sector	Macroeconomia	3
6	dataton_owners	900047981	new45294	Sector	Macroeconomia	1
7	dataton_owners	860005216	new77341	Sector	Sostenibilidad	36
8	dataton_owners	900628110	new59440	Sector	Macroeconomia	2
9	dataton_owners	900628110	new43958	Sector	Macroeconomia	5
10	dataton_owners	900378212	new88832	Sector	Macroeconomia	19
11	dataton_owners	860002964	new96093	Cliente	Macroeconomia	11
12	dataton_owners	860034594	new67975	Cliente	Regulaciones	30
13	dataton_owners	860005216	new90275	Cliente	Alianzas	29
14	dataton_owners	860005216	new32242	Sector	Macroeconomia	31

```
query4.to_csv("C:/Users/jucadiaz/GIT/Dataton-BC/2022/src/data/output/recomendacion.csv", index = False)
```

preguntas

Muchas Gracias