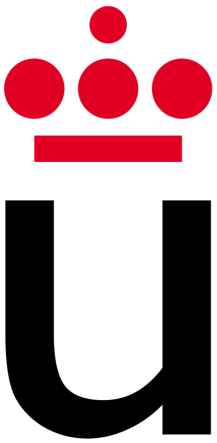


Universidad Rey Juan Carlos

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

GRADO EN INGENIERÍA DE LA CIBERSEGURIDAD

PRÁCTICA 1

SISTEMAS DE LA INFORMACIÓN
CURSO 2021-2022

Redactado por: Yeneva Miranda Basilio

Gabriel Izquierdo González

Mario Ruano Díaz

ÍNDICE

REPOSITORIO DE GITHUB **3**

EJERCICIO 1 **4**

EJERCICIO 2 **6**

EJERCICIO 3 **6**

EJERCICIO 4 **7**

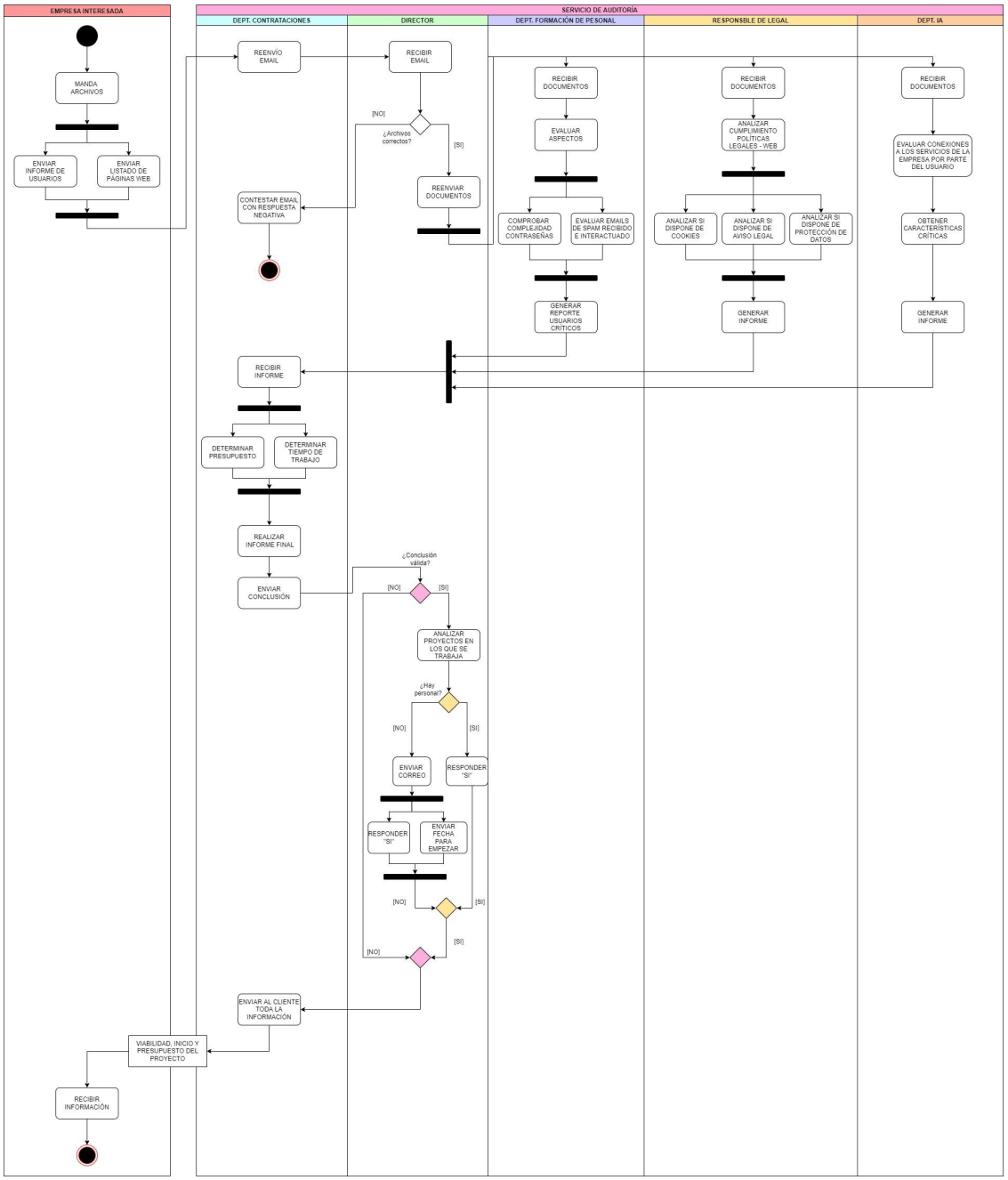
REPOSITORIO DE GITHUB

https://github.com/Mari0RD/Sistemas_Informacion

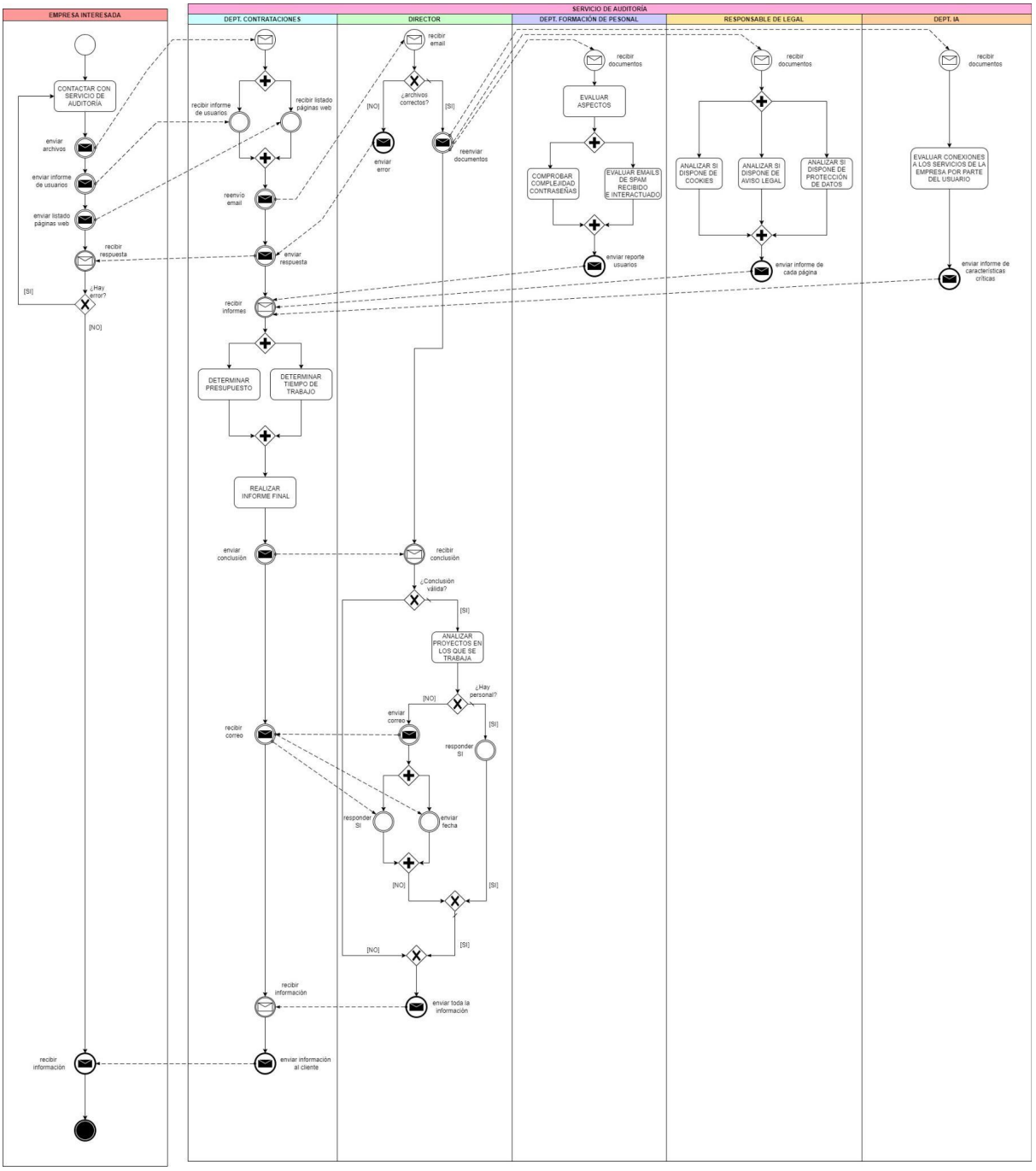
EJERCICIO 1

En este primer ejercicio, el grupo deberá desarrollar el modelado del proceso de negocio descrito anteriormente usando las dos notaciones vistas en teoría: Business Process Modeling Notation y Unified Modeling Language.

UML (Unified Modeling Language)



BPMN (Business Process Model and Notation)

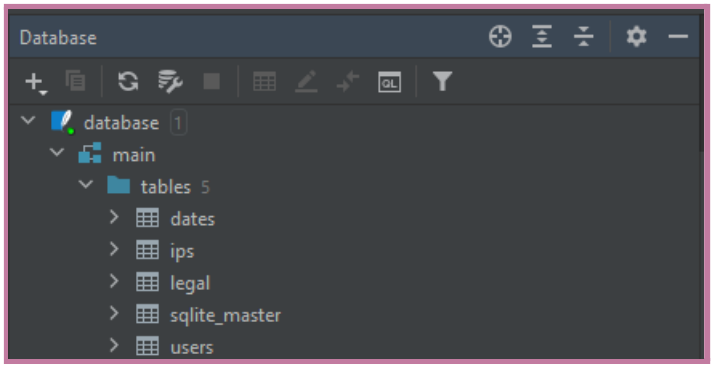


EJERCICIO 2

El objetivo de este ejercicio será el de desarrollar un sencillo sistema ETL. No es necesario desarrollar las fases de extracción ya que disponemos del archivo JSON. Debemos diseñar las tablas en la base de datos y desarrollar los códigos necesarios para leer los datos del fichero JSON y almacenarlos en la base de datos. Después, será necesario leer los datos desde la BBDD (usando diferentes consultas) y se almacenarán los resultados en un DataFrame para poder manipularlos. En este ejercicio, para el correcto desarrollo del sistema MIS, será necesario calcular los siguientes valores:

- Número de muestras (valores distintos de missing).
- Media y desviación estándar del total de fechas que se ha iniciado sesión.
- Media y desviación estándar del total de IPs que se han detectado.
- Media y desviación estándar del número de emails recibidos.
- Valor mínimo y valor máximo del total de fechas que se ha iniciado sesión.
- Valor mínimo y valor máximo del número de emails recibidos.

En el fichero **database.py** hemos creado la base de datos con las tablas USERS, IPS, FECHAS y LEGAL para posteriormente cargar en ellas los valores contenidos en users.json y legal.json.



Los valores presentados en la imagen se han obtenido de la ejecución del fichero **ejercicio2.py** aplicando a los dataframes sobre las distintas consultas las funciones **mean()** para la media, **std()** para la varianza, **min()** para el mínimo y **max()** para el máximo.

```
Numero de muestras:  227
Media del total de fechas que se ha iniciado sesion:  9.866666666666667
Desviacion estandar del total de fechas que se ha iniciado sesion:  6.055680338972335
Media del total de IPs que se han detectado:  9.6
Desviacion del total de IPs que se han detectado:  6.152415521465408
Media del numero de emails recibidos:  247.86666666666667
Desviacion estandar del numero de emails recibidos:  141.44274663167653
Valor minimo del total de fechas que se ha iniciado sesion:  1
Valor maximo del total de fechas que se ha iniciado sesion:  20
Valor minimo del numero de email recibidos:  20
Valor maximo del numero de email recibidos:  493
```

EJERCICIO 3

Hay datos que nos interesa analizar basándonos en agrupaciones, para darle un sentido a nuestro análisis en base a esa agrupación. De una manera más específica, vamos a trabajar con las siguientes agrupaciones:

- **Por tipo de permisos de usuario (0 equivalente a usuario y 1 equivalente a administrador)**

- **Número de emails recibidos: estableceremos dos rangos diferentes, el primero aquellos contenidos que tengan menos de 200 correos, y aquellos que tengan igual o más de 200 correos.**

En este caso deberemos calcular la siguiente información para la variable dentro del email de phishing:
- **Número de observaciones**
- **Número de valores ausentes (missing)**
- **Mediana**
- **Media**
- **Varianza**
- **Valores máximo y mínimo**

En el archivo **ejercicio3.py** hemos creado distintos dataframes que nos permitirán realizar la agrupación por permisos de usuario o de administrador y de más de 200 correos o de menos de 200.
Para ello hemos utilizado las siguiente sentencias:

```
SELECT phishingEmails  FROM users WHERE users.permissions = 1 GROUP  BY username
SELECT phishingEmails  FROM users WHERE users.permissions = 0 GROUP  BY username
SELECT phishingEmails  FROM users WHERE users.totalEmails < 200 GROUP  BY username
SELECT phishingEmails  FROM users WHERE users.totalEmails >= 200 GROUP  BY username
```

Las funciones usadas han sido **mean()** para la media, **std()** para la varianza, **min()** para el mínimo y **max()** para el máximo.

```
<--- PERMISOS DE ADMINISTRADOR (1) --->
Numero de observaciones:  2003
Numero de valores ausentes:  0
Mediana:  138
Media:  143.07142857142858
Varianza:  12490.84065934066
Minimo:  1
Maximo:  372
```

```
<--- PERMISOS DE USUARIO (0) --->
Numero de observaciones:  1277
Numero de valores ausentes:  0
Mediana:  41
Media:  79.8125
Varianza:  9881.229166666666
Minimo:  0
Maximo:  382
```

```
<--- CANTIDAD CORREOS MENOR A 200 --->
Numero de observaciones:  699
Numero de valores ausentes:  0
Mediana:  41
Media:  58.25
Varianza:  1945.1136363636363
Minimo:  1
Maximo:  133
```

```
<--- CANTIDAD DE CORREOS MAYOR O IGUAL A 200 --->
Numero de observaciones:  2581
Numero de valores ausentes:  0
Mediana:  134.5
Media:  143.38888888888889
Varianza:  15699.545751633985
Minimo:  0
Maximo:  382
```

EJERCICIO 4

Por último, se programaran las diferentes funciones del MIS. En concreto, se deben generar gráficos sencillos para obtener los siguientes datos:

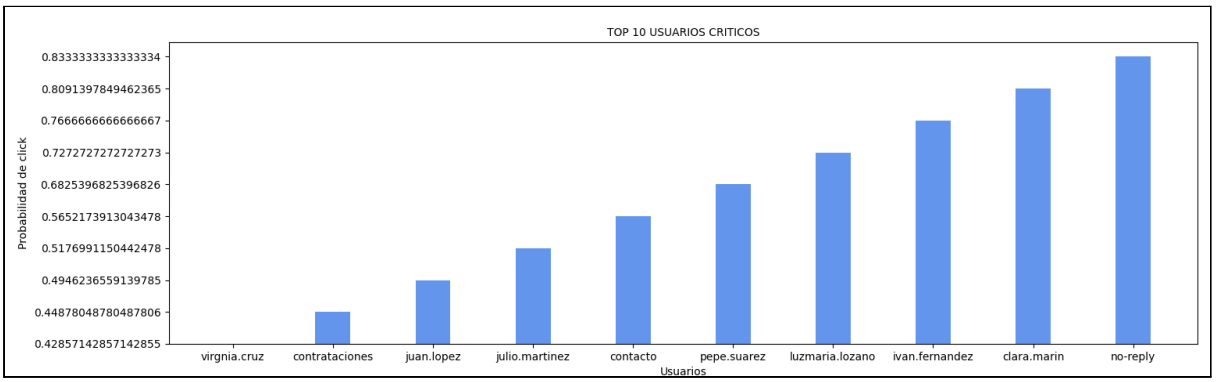
- **Mostrar los 10 usuarios más críticos (un usuario crítico es aquel usuario que tiene la contraseña débil y además tiene mayor probabilidad de pulsar en un correo de spam), representadas en un gráfico de barras.**

Para este apartado hemos introducido manualmente los Hashes de las contraseñas en "https://crackstation.net/" para ver que usuarios son los más críticos. Posteriormente hemos creado un archivo "**passwords.txt**" con las contraseñas en texto plano, para poder, mediante código, comparar sus hashes y determinar cuales son los usuarios con contraseñas vulnerables.

Hash	Type	Result
6aa2ab5334956e999997539cd9035ae0	Unknown	Not found.
eace8d19d31a304e95c409ab819079a7	Unknown	Not found.
5f4dcc3b5aa765d61d8327deb882cf99	md5	password
3bf1114a986ba87ed28fc1b5884fc2f8	md5	shadow
a3ff7b065d6e0919200ce26642c87b60	Unknown	Not found.
058755b89fbec19a8f13a48d395cd42f	Unknown	Not found.
276f8db0b86edaa7fc805516c852c889	md5	baseball
84d961568a65073a3bcf0eb216b2a576	md5	superman
0acf4539a14b3aa27deeb4cbdf6e989f	md5	michael
37386d73866a7ecfa68464c681d7f366	Unknown	Not found.
6522005376b658a8ae437f2553e8fa7a	Unknown	Not found.
1660fe5c81c4ce64a2611494c439e1ba	md5	jennifer
d16d377af76c99d27093abc22244b342	md5	jordan
d073536bba0de057994e2be46bb64a04	Unknown	Not found.
eb0a191797624dd3a48fa681d3061212	md5	master
607693c5fef9a92c91e914d274b988c6	Unknown	Not found.
bf787578e2c6e374529ba85b218322c1	Unknown	Not found.
e90988574b3f2078b73d98d291db85f8	Unknown	Not found.
62e2f762af3518bf3a7b14d4414e3fb7	Unknown	Not found.

714ab9fbdad5c5da1b5d34fe1a093b79	md5	pug
5b628ede9539b225d683ce4a9cb797d8	Unknown	Not found.
8e7910d5ba9a04d791aba89ab75d3adb	Unknown	Not found.
d8578edf8458ce06fbc5bb76a58c5ca4	md5	qwerty
cf3bc4839182dafee429b4ae7a415ec9	Unknown	Not found.
55ca44dabad13a109b038fe2f694b7dd	Unknown	Not found.
4297f44b13955235245b2497399d7a93	md5	123123
37b4e2d82900d5e94b8da524fbeb33c0	md5	football
d0763edaa9d9bd2a9516280e9044d885	md5	monkey
ff91144bcc200fe870bdf2136ac59930	Unknown	Not found.
a152e841783914146e4bcd4f39100686	md5	asdfgh

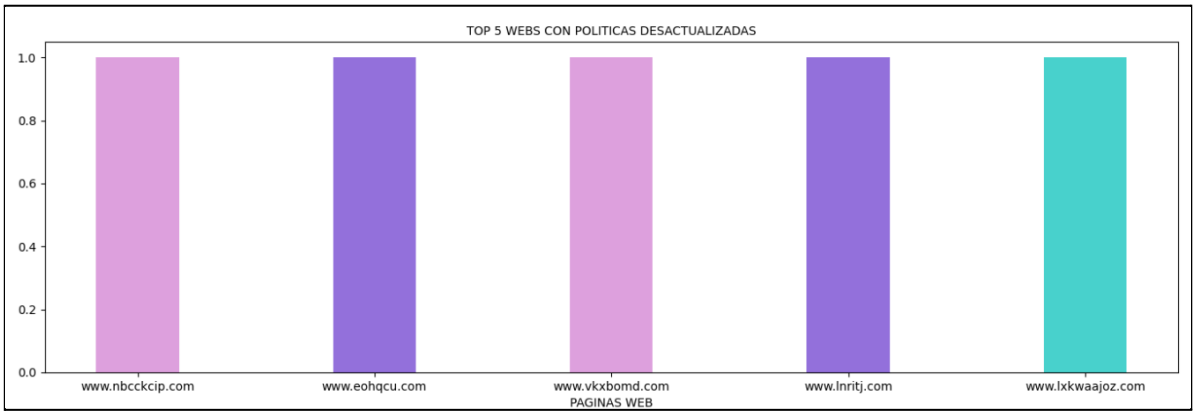
Los usuarios críticos los hemos obtenido mirando aquellos que ya presentaban una contraseña vulnerable y además una alta probabilidad de clicar correos de spam.




```
<--- TOP 10 USUARIOS CRITICOS --->
USUARIO: no-reply CON PROBABILIDAD DE: 0.8333333333333334
USUARIO: clara.marin CON PROBABILIDAD DE: 0.8091397849462365
USUARIO: ivan.fernandez CON PROBABILIDAD DE: 0.7666666666666667
USUARIO: luzmaria.lozano CON PROBABILIDAD DE: 0.7272727272727273
USUARIO: pepe.suarez CON PROBABILIDAD DE: 0.6825396825396826
USUARIO: contacto CON PROBABILIDAD DE: 0.5652173913043478
USUARIO: julio.martinez CON PROBABILIDAD DE: 0.5176991150442478
USUARIO: juan.lopez CON PROBABILIDAD DE: 0.4946236559139785
USUARIO: contrataciones CON PROBABILIDAD DE: 0.44878048780487806
USUARIO: virgnia.cruz CON PROBABILIDAD DE: 0.42857142857142855
```

• Mostrar las 5 páginas web que tienen más políticas (cookies, protección de datos o aviso legal) desactualizadas, representadas en un gráfico de barras según las políticas.

Hemos considerado que las páginas que tienen las políticas más desactualizadas son aquellas que tengan más "0s" en las columnas, cookies, protección de datos y aviso legal de la tabla legal. Para ello hemos sumado todas las columnas de las correspondientes urls, las hemos ordenado (**.sort_values()**) y nos hemos quedado con las 5 menores (**.head(5)**)




```
<--- TOP 5 WEBS CON POLITICAS DESACTUALIZADAS --->
                 url  cookies  warning  dataProtection  result
0    www.nbcckcip.com        0        1               0       1
17     www.eohqcu.com        0        0               1       1
14    www.vkxbomd.com        0        1               0       1
11     www.lnritj.com        0        0               1       1
6   www.lxkwaajoz.com        1        0               0       1
```

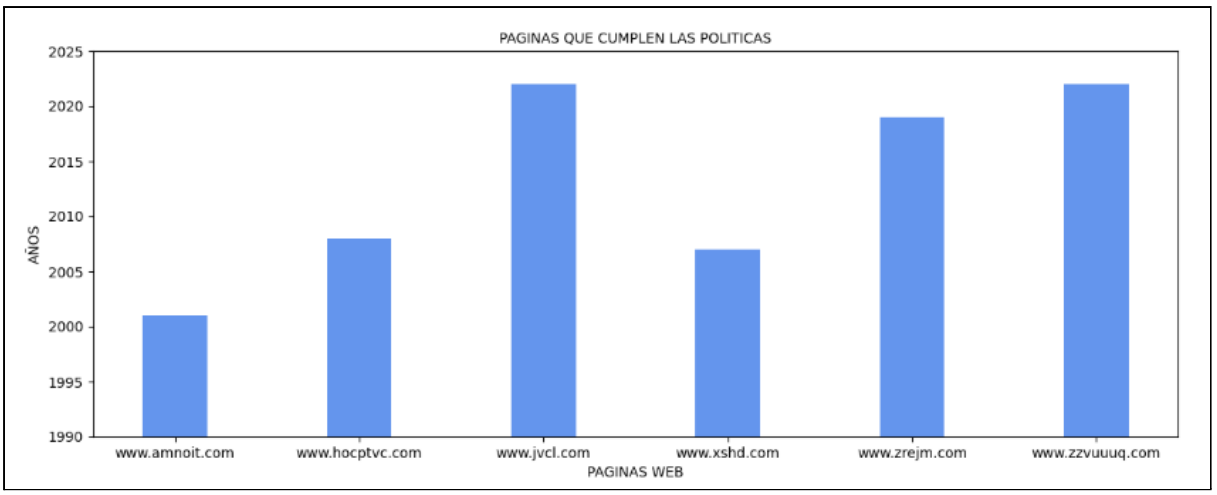
• Mostrar según el año de creación las webs que cumplen todas las políticas de privacidad, frente a las que no cumplen la política de privacidad.

CUMPLEN LA POLÍTICA DE PRIVACIDAD

(Hemos supuesto que cumplen la política de privacidad cuando TODAS sus columnas están a 1.)

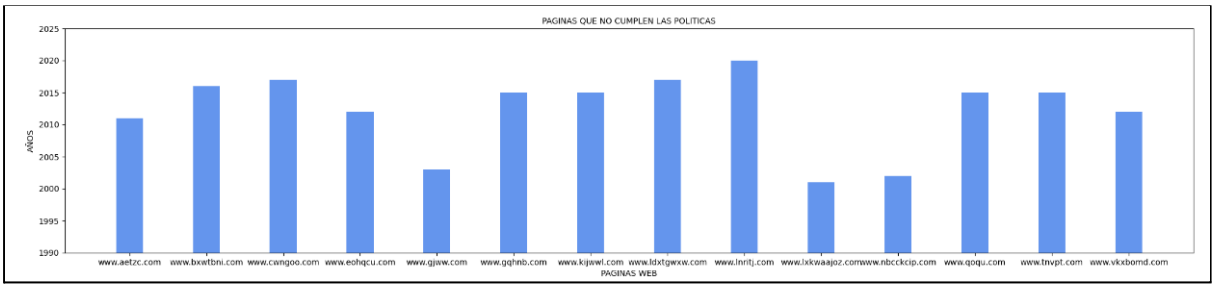
Al igual que en el apartado 2 de este mismo ejercicio, hemos sumado las columnas de los años de creación de las webs correspondientes.

A diferencia del anterior en este nos hemos quedado con aquellas urls donde sus columnas sumasen 3, es decir, que todas sus columnas estuviesen a 1.



NO CUMPLEN LA POLÍTICA DE PRIVACIDAD

(Hemos supuesto que no cumplen la política de privacidad cuando NO TODAS sus columnas están a 1.) Para obtener los resultados de este apartado hemos elegido aquellas que, por tanto, no suman 3 sus columnas.



- **Mostrar el número de contraseñas comprometidas y contraseñas no comprometidas.**

Para la realización del primer apartado hemos utilizado la solución de este apartado. Para calcular el número de contraseñas comprometidas y no comprometidas comprobamos que usuarios tenían la contraseña vulnerable y si es así la sumamos a un contador. Posteriormente al total de contraseñas le restamos este contador para conocer el número de contraseñas no comprometidas.