

1 Práctica 1: Recopilación, estructuración y análisis de datos

El objetivo de esta primera práctica es el de profundizar en algunos de los conceptos vistos en las sesiones de teoría. La situación que se plantea es la siguiente: acabas de entrar en una gran compañía que proporciona servicio de auditorías. Tu labor dentro de la compañía será analizar algunos procesos del modelo de negocio y desarrollar un pequeño sistema de información. En la empresa tienen definidos una gran variedad de procesos, pero para esta práctica nos interesa el proceso mediante el cual se decide si trabajar con una empresa o no. Este proceso de negocio por el cual se evalúa si trabajar con una empresa o no comienza cuando la empresa interesada en los servicios de auditoría, manda diferentes archivos a través del departamento de contrataciones, estos archivos son 1) informe de usuarios con el log de sesiones, hash de sus contraseñas, permisos, datos personales, etc 2) el listado de páginas webs que tiene la empresa. En ese momento, ese email se reenvía al director del servicio de auditorías que se encargará de verificar si los archivos que se han enviado son correctos, y decidir si continua el análisis de la empresa adelante o no. Si el director no da su visto bueno, el departamento de contrataciones contestará al email del cliente con la respuesta negativa. En caso contrario, será el propio director el que mandará los documentos a los diferentes departamentos, para que realicen un análisis de la empresa:

- **Departamento de formación de personal.** La compañía tiene un departamento encargado de analizar los perfiles de los usuarios de la empresa. En este punto se evalúan varios aspectos: 1) complejidad de las contraseñas (mediante el hash), y 2) emails de spam que ha recibido e interactuado. Este departamento realizará un reporte de usuarios críticos y lo adjuntarán al departamento de contrataciones.
- El segundo departamento que recibe los documentos será el servicio **responsable de legal**, tiene como tarea verificar el cumplimiento de todas las políticas legales en los diferentes servicios webs. En concreto, analizará si las webs disponen de cookies, aviso legal y protección de datos. Completará el archivo enviado con los datos por cada página y este informe será recibido por el departamento de contrataciones.
- Por último, el **departamento de inteligencia artificial** evaluará las conexiones a los servicios de la empresa por parte de los usuarios y reportará el número estimado sin profundizar de características críticas de la empresa. Una vez tenga este número de características críticas, adjuntará al departamento de contrataciones un informe.
- El **departamento de contrataciones**, además de establecer la comunicación inicial con el cliente, se encargará de determinar el presupuesto y tiempo de trabajo necesario, basado en el número de características/vulnerabilidades encontradas en el sistema del cliente.

Cuando el departamento de contrataciones hayan finalizado su informe final, le mandarán su conclusión al director. Si el director recibe un “Sí” por parte del departamento de contrataciones, analizará los proyectos dónde se está trabajando para poder determinar si se tiene personal para comenzar a trabajar con la empresa, en caso de no tener personal remitirá un correo al departamento de contrataciones para que respondan con un “Sí” y la fecha en la que se puede comenzar a realizar el proyecto, en caso de tener personal responderá con un “Sí” mostrando que se puede comenzar cuanto

antes y en caso de ver en el informe poca cantidad de vulnerabilidades/características reportará un “No” debido a que será un trabajo breve con el cual no quiere dedicar tiempo. Finalmente, el departamento de contratación le comunicará al cliente toda la información: si es viable su proyecto, el inicio del proyecto y el presupuesto estimado. La primera tarea dentro de esta compañía será la de modelar, usando BPMN y UML, este proceso de negocio que se acaba de describir. La segunda tarea que te han encomendado es la de desarrollar, una primera versión de un sistema de información gerencial o MIS (Management Information System). Para ello, se te proporcionan los archivos JSON extraído de sus base de datos. Tu labor será la de desarrollar, usando el lenguaje de programación Python y una base de datos SQLite, este pequeño MIS que genere la información que se detalla en los diferentes ejercicios de este enunciado. Además, se pondrá en práctica el uso de librerías pensadas para realizar un análisis de datos que permita generar información interesante para un potencial usuario final, cliente o, simplemente, alguien interesado en recibir la información que se obtiene tras el filtrado y análisis de los datos proporcionados. Concretamente, utilizaremos la librería Pandas para ayudarnos a realizar esta tarea. Es importante que el alumnado realice esta primera práctica con éxito y de forma adecuada, ya que la segunda práctica se basará en gran medida en los resultados, o avances alcanzados en esta práctica. La práctica se realizará en **grupos de 3 personas**. La práctica sólo debe ser entregada por un integrante.

2 Conjunto de datos

Para esta práctica, utilizaremos el archivo de material adicional, el cual contiene diferentes archivos JSON.

- **legal.json** contiene el análisis realizado por el departamento de legal de cada una de las webs.
- **users.json** contiene la información de cada uno de los usuarios de la empresa.

3 Ejercicio 1 [3 puntos]

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 (1.5 punto) y Unified Modeling Language (1.5 punto)

4 Ejercicio 2 [2 puntos]

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.

5 Ejercicio 3 [2.5 puntos]

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

6 Ejercicio 4 [2.5 puntos]

Por último, se programarán 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.
- Mostrar las 5 páginas web con que tienen más políticas (cookies, protección de datos o aviso legal) desactualizadas, representadas en un gráfico de barras según las políticas.
- Mostrar la media de conexiones de usuarios con contraseña vulnerable, frente a los que no son vulnerables.
- 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.
- Mostrar el número de contraseñas comprometidas y contraseñas no comprometidas.

7 GitHub

Será de uso obligatorio la creación de un repositorio público de GitHub para la realización de las prácticas con los miembros del grupo. Es recomendable el uso correcto de GitHub, se valorará negativamente la incorporación de todos los datos en un solo commit.

8 Normas de Entrega

La entrega de la práctica consistirá en un archivo comprimido con los siguientes ficheros:

- **Carpeta src** del proyecto de PyCharm.
- Archivo SQLite con la base de datos creada.
- Un solo documento con la memoria completa en formato PDF en la que se muestren los ejercicios resueltos.
- La memoria debe incluir el nombre y apellidos de los integrantes del grupo y el enlace al repositorio de GitHub.
- Cualquier memoria entregada fuera de plazo o de forma corrupta será considerada como un 0.

Ante cualquier duda durante la resolución de la práctica, escribir a isaac.lozano@urjc.es vía mail. En caso de no poderse resolver la duda vía mail, se puede concertar una tutoría, siempre y cuando se concierte en un período de **hasta 48 horas antes** de la fecha de entrega de la práctica.

La fecha límite para entregar esta práctica será el 28 de Marzo a las 23:55 y se realizará por la plataforma Aula Virtual.

9 Peso de la práctica

La evaluación de esta práctica supondrá un 40% de la nota de prácticas.