# Migración Personas

## Análisis de viabilidad

### Descripción del Proyecto

Se desea realizar una comprobación de unos registros del personal indicado por la empresa.

### Antecedentes del Proyecto

Tras un virus se corrompió el fichero que almacenaba los datos del personal. Se desea automatizar el proceso de comprobación de los datos del fichero.

### Recursos Humanos necesarios

Se ha estimado que se necesitará contar con una plantilla compuesta por 1 trabajador, que tendrá asignadas, en función de su puesto de trabajo, ciertas tareas imprescindibles para el correcto desarrollo de la actividad.

En la siguiente tabla se resume la plantilla necesaria desglosada por los distintos puestos o cargos. También se hace mención al régimen de la seguridad social al cual estarán afiliados y el sueldo líquido mensual que recibirán.

|  |  |  |  |
| --- | --- | --- | --- |
| Puesto o Cargo | Nº Trabajadores | Regimen Seguridad Social | Sueldo Líquido Mensual |
| Administrador/Programador | 1 | Cuenta Ajena | 1.500,00 € |

Antes de terminar con este punto, únicamente comentar que el sueldo líquido mensual es el importe que el trabajador cobrará una vez descontadas la seguridad social y el IRPF. Es decir, lo que el trabajador ingresará en su cuenta bancaria. Nos hemos decantado por indicar el sueldo líquido mensual (sueldo neto), en lugar del sueldo bruto, por considerar, que esta es la variable que más le interesa conocer a todo trabajador. Esta claro que el coste del trabajador para la empresa será superior a este sueldo líquido, este coste es el que se ha tenido en cuenta para el cálculo de los costes mensuales de la empresa.

## Requisitos

### Fichero de entrada

Según convención con el cliente, se proveerá del fichero de entrada que tendrá como nombre “personas.txt”, y estará disponible localmente por el programa.

Nota: El fichero de entrada tendrá que ser convertido a la codificación UTF-8 siempre que no se provea de esa manera.

#### Formato

Cada línea de datos será un registro diferente. El carácter límite del registro será el salto de línea.

Se espera que cada línea disponga de 7 campos, si una línea no comprende este número de campos, será considerada incorrecta y será excenta para las siguiente comprobaciones.

Para considerar una línea correcta, aparte del corte anterior, tendrá que tener el siguiente formato y condiciones:

Campo 1 -> Nombre: Tipo texto, cuyo texto ha de tener caracteres válidos según la norma UTF-8.

Campo 2 -> Apellido1: Tipo texto, cuyo texto ha de tener caracteres válidos según la norma UTF-8.

Campo 3 -> Población: Tipo texto, cuyo texto ha de tener caracteres válidos según la norma UTF-8.

Campo 4 -> Edad: Numérica, debe estar comprendida entre los valores 18 y 99.

Campo 5 -> Mail: Tipo texto, tendrá que ser validado su formato como dirección de correo electrónico correcta.

Campo 5 -> Nif/Cif: Tipo texto, tendrá que ser validado su formato como Nif/Cif correcto.

Campo 7 -> Categoría: Tipo texto, cuyo texto ha de tener caracteres válidos según la norma UTF-8.

Estos campos estará separados por el carácter “,”.

### Fichero de salida

Se solicita se dispongan cuatro ficheros de salida.

#### Fichero de salida con filas correctas

Este fichero mostrará los registros correctos tras el análisis del fichero de entrada. Cada registro de este fichero tendrá el mismo formato que tenía en el fichero de entrada.

El nombre del fichero de salida será “personas\_correctas.txt”.

#### Fichero de salida con filas incorrectas

Este fichero mostrará los registros incorrectos tras el análisis del fichero de entrada. Cada registro de este fichero contendrá el primer fallo del registro analizado seguido de los datos del registro con el mismo formato de entrada, por ejemplo:

Error => campo1, campo2, campo3, etc…

El nombre del fichero de salida será “personas\_error.txt”.

#### Fichero de salida con filas duplicadas

Este fichero mostrará los registros duplicados que se encuentren después del análisis.

Se tomará como registro duplicado aquellos registros que compartan el campo DNI. Estarán exentos de comprobación de duplicados aquellos registros cuyo número de campos no sea el esperado.

Se mostrarán en el fichero tanto como si son registros correctos como si no. Que aparezcan los registros en este fichero no implica que aparezcan en los ficheros de salida de filas incorrectas o correctas.

El formato de salida de cada registro de este fichero será la siguiente:

* Errores, siempre que los tenga
* Campo X y Valor

Por ejemplo:

UTF8 => nombre: valor, apellido1: valor, …

El nombre del fichero de salida será “personas\_repetidas.txt”.

#### Fichero de salida con estadísticas

Este fichero tiene que mostrar las estadísticas relativas al procesado del fichero de entrada. Estas estadísticas tienen que incluir los siguientes datos:

* Tiempo de proceso de análisis del fichero. Excluyendo la creación y el llenado de los ficheros de salida.
* Número de registros analizados.
* Número de registros correctos.
* Número de registros incorrectos.

El nombre del fichero de salida será “estadisticas.txt”.