

Ejercicio 1 (10 puntos):

Dado una empresa que va a realizar una campaña de llamados para mejorar el servicio y ofrecer un nuevo plan de contratación en su afiliación. Como procedimiento van a utilizar un conjunto de datos obtenidos internamente por el equipo comercial que se encuentra disponible en una planilla excel con información insertada por ellos mismos.

1 - En base a la obtención de la información puede identificar alguno de los motivos de falta de calidad de datos obtenidos, fundamente.

Respuesta 1: Los motivos de errores mas frecuentes serían los siguientes:

- Datos de Entrada: Posibles errores humanos de ingresos de datos.
- Datos Externos: Carga errónea de datos proveniente de otras fuentes.
- Errores de carga de los sistemas transaccionales: Errores ocasionados por exportación de datos desde otros sistemas.
- Migraciones: Cargar o unificar fuentes de datos sin controles y/o análisis previo.

2 - Asumiendo que ya identificaron algunos de los motivos y en una primera revisión se encuentra datos faltantes en la columna teléfono, Podrían identificar como se llama este tipo de error y por qué.

Respuesta 2: En general, los tipos de errores comúnmente identificados serían:

- Errores por información incompleta.
- Errores de sintaxis.
- Errores de semántica.

Para el caso en particular de datos faltantes que se menciona en la pregunta, el tipo de error sería “por información incompleta”. Un ejemplo de porque se podría haber dado, es que la carga del dato no fuese obligatoria, o no estuvieses claro que era obligatorio el dato, y el usuario no consideró necesaria la carga, completando los datos definidos como obligatorios, además de que el formulario o base de carga del dato, no tenía ningún tipo de sistemática de seguridad para que el dato sea completado obligatoriamente.

3 - Sabiendo que la base está incompleta, ¿Cómo se denomina la dimensión que determina como característica este faltante?, explique con un ejemplo

Respuesta 3: La dimensión sería “Compleitud”. Estaríamos hablando de por ejemplo la falta de la carga del teléfono, o falte cargar el Departamento yendo a posibles casos de la base de datos a utilizar en este obligatorio.

4 - Explique cuáles son los beneficios de poder transformar estos datos para poder tener una calidad de datos óptima para trabajar.

Respuesta 4: Los beneficios podrían ser la mejora en el desempeño y toma de decisiones, mejorando de esta manera la gestión de las organizaciones. Teniendo una buena calidad de

datos, se beneficia también la efectividad y eficiencia de cualquier organización, generando directamente ganancias ya sea financieras o en términos de gestión. Además de lo anterior, una buena calidad de datos beneficia en representar una realidad verdadera, beneficiando así directamente en la toma de decisiones sobre esa realidad, como ya se mencionó.

Ejercicio 2 (15 puntos):

Desarrolle una expresión regular que valide Pasaportes de Uruguay y Argentina, tomando en cuenta las especificaciones actuales que regulan dicha disposición. Las letras son siempre en mayúsculas.

Uruguay:

- Comienza siempre con UY
- Luego de UY siempre sigue una letra mas
- Se agregan 6 números aleatorios

Entonces seria 3 letras y 6 números siempre

Argentina:

- Comienza siempre con AR
- Luego pueden seguir una o dos letras mas
- Finaliza con 6 o 7 números

Entonces seria 3 o 4 letras y 6 o 7 números

Ejemplos Válidos

- UYA456789
- ARA123456
- ARVZ456321
- ARNN4561234

Ejemplos No Válidos

- UY1234651
- UYAA135123
- YU1111111
- UY123456
- AR123456
- ARA123
- ARTT6555555
- ARUY24124111
- NO165679

Respuesta Ejercicio 2:

REGULAR EXPRESSION 4 matches (0.2ms)

" ^UY[A-Z][0-9]{6}\$|^AR[A-Z]{1,2}[0-9]{6,7}\$ "

TEST STRING

CADENAS DE CARACTERES QUE CUMPLEN CON LA EXPRESIÓN:

UYA456789
ARA123456
ARVZ456321
ARNN4561234

CADENAS DE CARACTERES QUE NO CUMPLEN CON LA EXPRESIÓN:

UY1234651
UYAA135123
YU1111111
UY123456
AR123456
ARA123
ARTT65555555
ARUY24124111
N0165679

Ejercicio 3 (15 puntos):

Dado una columna de datos que se obtienen los teléfonos de varios clientes radicados en varios países, deseamos poder generar una nueva columna donde podamos tener solo el número telefónico sin el código país

Ejemplo:

(598)23446777

La idea es crear una columna nueva llamada numero_telefono donde solo se puede obtener siempre y cuando la columna de datos original contenga estos dos caracteres '(' y ')', si no llega a tener estos dos caracteres debe quedar nula la celda de datos

Ejemplo de como quedaría

phone	numero_telefono
(598)23446777	23446777
(59823446777	Null

Sugerencia usar la función GREL utilizando el if como uno de las funciones más adecuadas, también puede utilizarse otras que ustedes deseen.

Conjunto de datos dado (generar proyecto por Clipboard)

(55)344555685

(235235456234

2352355

(1)12342378

(1234)1245236

125236

12352350)

Respuesta Ejercicio 3:

OpenRefine
Ejercicio 3_Obligatorio
[Permalink](#)

Open...
Export
Help

Facet / Filter
Undo / Redo 3 / 3

Extract...
Apply...

Filter

0. Create project
1. Rename column Column 1 to phone
2. Create new column numero_telefono based on column phone by filling 7 rows with `grel:if(and(contains(value, "("),contains(value,"")), value.partition("[2],"""))`
3. Create new column numero_telefono_expresión_regular based on column phone by filling 3 rows with `grel:match(value, /(.*)([0-9]{1,})\.?.*)/[1]`

7 rows
Extensions Wikibase

Show as: rows records
Show: 5 10 25 50 100 500 1000 rows
« first < previous 1 next > last »

	All	phone	numero_telefono_expresión_regular	numero_telefono
1.	(55)344555685		344555685	
2.	(235235456234			
3.	2352355			
4.	(1)12342378	12342378		12342378
5.	(1234)1245236	1245236		1245236
6.	125236			
7.	12352350)			

Función GREL: `match(value,/(.*)/([0-9]{1,})\.?.*)/[1]`