

Plan de Validación y Tabulación.

Directorio de Empresas y Establecimientos

INEC – 2014/ 12/ 20

Contenido

Introducción.....	3
Variables involucradas	4
Plan de Validación y Tabulación u_legal (Variables Independientes)	5
Plan de Validación y Tabulación u_legal (Variables Dependientes)	6
Plan de Validación y Tabulación Empresa (Variables Independientes)	7
Plan de Validación y Tabulación Empresa (Variables Dependientes).....	9
Plan de Validación y Tabulación u_local (Variables Independientes)	11
Plan de Validación y Tabulación u_local (Variables Dependientes)	12
Plan de Validación y Tabulación Otras Variables	14
Limpieza de datos	15
Valores por defecto y	20
Tratamiento a los datos que presenten inconsistencias	20
Conteos	23
Anexos.....	25
Validación del RUC:	25
Tipos de validaciones a realizar:	25
Informe Limpieza de Datos	25
Ver Anexo 3 (Informes Limpieza de Datos).....	25
Plan de Inconsistencias con cruce de Variables	25
Conclusiones:.....	26
Recomendaciones:	26

Introducción

El Directorio de Empresas y Establecimientos maneja una gran cantidad de variables, las cuales deben ser validadas, de tal manera que se tenga información confiable y veraz al momento de elaborar las publicaciones anuales de un nuevo Directorio.

Existen variables que dependen de otras para su validación y variables independientes, el software que se utiliza para la administración de la base de datos puede realizar validaciones automáticas pero también tenemos validaciones manuales, todas éstas con el objetivo de obtener información de calidad.

El presente documento contiene el plan de validación y tabulación de las variables que existen en el DIEE, basado en la experiencia obtenida a través del trabajo de la primera publicación, también se incluyen documentos de la limpieza de datos detallados en una matriz a tomar en cuenta al momento de proceder con el arreglo de los mismos, análisis de inconsistencias en la información o algún tipo de novedad adicional que se dé en el manejo de la base de datos

Variables involucradas

Existe una gran cantidad de variables, estas después de su validación deben ser subidas a la base de datos del Directorio de Empresas y Establecimientos, por lo que, se ha creado una matriz de priorización de validación de datos por cada variable de análisis.

Matriz de prioridad:

Prioridad	Descripción
1	Alta
2	Media
3	Baja

Valoración de la matriz de prioridad:

Alta: Variables que están directamente ligadas a la Publicación

Media: Variables que no están en la Publicación pero si tienen relación directa con la publicadas.

Baja: Variables que no tienen relación con la Publicación.

Plan de Validación y Tabulación u_legal (Variables Independientes)

# CRUCE	VARIABLE INICIAL	Validación 1	Validación 2	Validación 3	Prioridad	Detalle	
1	Numero Ruc	=13 dígitos	Validación del RUC (ANEXO 1)	---	1	Validación Automática	
2	Fecha Desde (FECHA APERTURA)	8 números consecutivos. En formato aaaammdd	---	---	3	Validación Automática	
3	Expediente	Personas Naturales no tienen expediente.	---	---	3	Validación Automática	
4	Acto Jurídico	Constitución	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Escisión	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Fusión	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		No requiere	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Desconocido	Consultar a la fuente	Verificación Superintendencia de Compañías	---	3	Validación Automática
5	Clase Contribuyente	Desconocido	Verificar coincidencia con la fuente	Reportar los casos desconocidos	---	3	Validación Automática
		Especial	Verificar coincidencia con la fuente	---	---	3	Validación Automática
		Otros	Verificar coincidencia con la fuente	---	---	3	Validación Automática
		Rise	Verificar coincidencia con la fuente	---	---	3	Validación Automática
6	Estado Funcionamiento	Desconocido	Consultar a la fuente	Verificación Superintendencia de Compañías	Reportar casos desconocidos	3	Validación Automática
		Inactiva	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		En disolución	Verificación Superintendencia de Compañías	---	---	3	Validación Automática

Dirección de Estadísticas Económicas

		En reactivación	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		En liquidación	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Cancelación	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Intervenida AGD	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Cambio de estado de la empresa	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
7	Forma Jurídica	Desconocido	Consultar a la fuente	Reportar los casos desconocidos	---	2	Validación Automática
		Personal Natural	3er dígito RUC < 6 PERSONA NATURAL	---	---	2	Validación Automática
		Sociedades	3er dígito RUC ≥ 6 EMPRESAS	---	---	2	Validación Automática

Plan de Validación y Tabulación u_legal (Variables Dependientes)

# CRUCE	VARIABLE INICIAL	VARIABLE DE COMPARABILIDAD 1	VARIABLE DE COMPARABILIDAD 2	VARIABLE DE COMPARABILIDAD 3	Validación 1	Validación 2	Validación 3	Prioridad	Detalle	
8	Unidad Legal Estado	Desconocido	Fuente	---	---	Consultar a la Fuente	---	---	1	Reportar los casos desconocidos (Validación Automática)
		Activa	Fecha de inscripción	Fecha de inicio actividad	Fecha de cese actividad	Debe existir: fecha de inscripción y fecha de inicio de actividad	Si ha cerrado debe existir: fecha de cese de actividad	Si ha cerrado debe existir: fecha de reinicio de actividad	1	Validación Automática
		Pasiva	Fecha de inscripción	Fecha de inicio actividad	Fecha de cese actividad	Debe existir: fecha de inscripción	Debe existir: fecha de inicio de actividad	Debe existir: fecha de cese de actividad.	1	Validación Automática
		Suspensión definitiva	Fecha de inscripción	Fecha de inicio actividad	Fecha de cese actividad	Debe existir: fecha de inscripción	Debe existir: fecha de inicio de actividad	Debe existir: fecha de cese de actividad.	1	Validación Automática

		Cerrado	Fecha de inscripción	Fecha de inicio actividad	Fecha de cese actividad	Debe existir: fecha de inscripción	Debe existir: fecha de inicio de actividad	Debe existir: fecha de cese de actividad.	1	Validación Automática
		Por Revisar	Fuente	---	---	Consultar a la Fuente	---	---	1	Reportar los casos desconocidos (Validación Automática)
9	Obligado a llevar Contabilidad		Estrato Ventas	---	---	Ventas del año (a la publicación) > \$ 100.000	---	---	2	Validación Automática
10	Razón Social		RUC	---	---	(3er dígito RUC < 6) =PERSONA NATURAL	(3er dígito RUC ≥ 6) = EMPRESAS	---	1	Validación Automática

Plan de Validación y Tabulación Empresa (Variables Independientes)

# CRUCE	VARIABLE INICIAL	Validación 1	Validación 2	Validación 3	Prioridad	Detalle
1	Numero Ruc	= 13 dígitos	Validación del RUC (ANEXO 1)	---	1	Validación automática
2	Nombre Comercial	Validar vs la Fuente SRI	---	---	1	Validación automática
3	Actividad Comercio Exterior	Sin Actividad	Debe tener valor codificado (00)	---	3	Validación automática
		Importador	Debe tener valor codificado (01)	---	3	Validación automática
		Exportador	Debe tener valor codificado (02)	---	3	Validación automática

Dirección de Estadísticas Económicas

		Importador / Exportador	Debe tener valor codificado (03)	---	---	3	Validación automática
		Desconocido	Validar con la Fuente	Debe tener valor codificado (99)		3	Validación automática
4	Número de Unidades Locales		Valores > 0	Cada unidad local no puede tener el mismo número	---	2	Validación automática
5	Actividad Económica		Descripción = Código CIU 4	Realizar un cruce CPC vs CIU4 casos CENEC	---	1	Validación automática
6	Sitio Web		Estructura de una página web (www.pagina.com)	---	---	3	Validación automática
7	Tipo de Unidad Legal	Personal Natural	3er dígito RUC < 6 PERSONA NATURAL	---	---	2	Validación automática
		Personal Jurídica	(3er dígito RUC ≥ 6) EMPRESAS	---	---	2	Validación automática
		Desconocido	Consultar a la fuente	Reportar los casos desconocidos	---	---	2

Plan de Validación y Tabulación Empresa (Variables Dependientes)

# CRUCE	VARIABLE INICIAL	VARIABLE DE COMPARABILIDAD AD 1	VARIABLE DE COMPARABILIDAD AD 2	VARIABLE DE COMPARABILIDAD AD 3	Validación 1	Validación 2	Validación 3	Validación 4	Prioridad	Detalle
8	Razón Social	RUC	---	---	(3er dígito RUC < 6) =PERSONA NATURAL	(3er dígito RUC ≥ 6) = EMPRESAS	---	1	1	Validación Automática
9	Fecha de inscripción	Fecha de inicio de actividad	---	---	Debe existir Fecha de inicio.	Fecha de inscripción ≤ fecha de inicio de actividad.	8 números consecutivos. En formato aaaammdd	---	1	Validación automática
10	Fecha de inicio de actividad	Fecha de inscripción	---	---	Debe existir Fecha de inscripción	Fecha de inicio de actividad ≥ fecha de inscripción.	8 números consecutivos. En formato aaaammdd	---	1	Validación automática
11	Fecha de cese de actividad	Fecha de inscripción	Fecha de inicio de actividad	---	Fecha cese ≥ Fecha de inscripción. Fecha cese ≥ Fecha de inscripción.	Estado empresa = PASIVO.	8 números consecutivos. En formato aaaammdd	---	1	Validación automática
12	Fecha de reinicio de actividad	Fecha de inscripción	Fecha de inicio de actividad	Fecha de cese de actividad	Cuando la empresa nuevamente cesa en actividad, la fecha de reinicio de actividad pasa a estar en blanco	Fecha de reinicio ≥ Fecha de inscripción. Fecha reinicio ≥ Fecha de inicio. Fecha reinicio ≥ Fecha de cese	8 números consecutivos. En formato aaaammdd	---	1	Validación automática
13	Fecha de actualización	---	---	---	8 números consecutivos. En formato aaaammdd		---	---	1	Validación automática
14	Estado Empresa	Desconocido	Fuente	---	---	Consultar a la Fuente		---	1	Reportar los casos desconocidos (Validación Automática)
		Activa	Fecha de inscripción	Fecha de inicio de actividad	---	Debe existir: Fecha de inscripción y Fecha de inicio de actividad	Si ha cerrado debe existir: Fecha de cese de actividad	Si ha cerrado debe existir: Fecha de reinicio de actividad	---	1

Dirección de Estadísticas Económicas

		Pasiva	Fecha de inscripción	Fecha de inicio de actividad	Fecha de cese de actividad	Debe existir: Fecha de inscripción	Debe existir: Fecha de inicio de actividad	Debe existir: fecha de cese de actividad.	---	1	Validación Automática
		Suspensión definitiva	Fecha de inscripción	Fecha de inicio de actividad	---	Debe existir: fecha de inscripción	Debe existir: Fecha de inicio de actividad	---	---	1	Validación Automática
		Cerrado	Fecha de inscripción	Fecha de inicio de actividad	Fecha de cese de actividad	Debe existir: fecha de inscripción	Debe existir: Fecha de inicio de actividad	Debe existir: fecha de cese de actividad.	---	1	Validación Automática
		Por Revisar	Fuente	---	---	Consultar a la Fuente	---	---	---	1	Reportar los casos desconocidos (Validación Automática)
15	id_empleado_clase_ ins1		---	---	---	Tener asignado su rango correcto	---	---	---	1	Validación Automática
16	id_empleado_clase_ ins2		---	---	---	Tener asignado su rango correcto	---	---	---	1	Validación Automática
17	id_ventas_clase_ ins1		Obligado a llevar contabilidad	---	---	Tener asignado su rango correcto	Obligado a llevar contabilidad= SI; asignado estratos altos (Estrato IV o V)	---	---	1	Validación Automática
18	id_ventas_clase_ ins2		Obligado a llevar contabilidad	---	---	Tener asignado su rango correcto	Obligado a llevar contabilidad= SI; asignado estratos altos (Estrato IV o V)	---	---	1	Validación Automática
19	Obligado a llevar contabilidad		Estrato Ventas	---	---	Ventas del año anterior (a la publicación) > \$ 100.000	---	---	---	2	Validación Automática

Plan de Validación y Tabulación u_local (Variables Independientes)

# CRUCE	VARIABLE INICIAL	Validación 1	Validación 2	Validación 3	Prioridad	Detalle
1	Numero Ruc	= 13 dígitos	Validación del RUC (ANEXO 1)	---	1	Validación Automática
2	Número de unidad local	Al menos una unidad local registrada.	No pueden existir unidades locales con el mismo número identificador.	Validar que no exista el número cero en el identificador.	1	Validación Automática
3	Nombre Unidad local	Comprobar la información con la fuente	---	---	2	Validación Automática
4	Actividad Económica	Descripción = Código CIU 4	Realizar un cruce CPC vs CIU4 casos CENEC	---	1	Validación Automática
5	Actividad Económica Secundaria	Descripción = Código CIU 4	Realizar un cruce CPC vs CIU4 casos CENEC	---	3	Validación Automática

Plan de Validación y Tabulación u_local (Variables Dependientes)

# CRUCE	VARIABLE INICIAL	VARIABLE DE COMPARABILIDAD 1	VARIABLE DE COMPARABILIDAD 2	VARIABLE DE COMPARABILIDAD 3	Validación 1	Validación 2	Validación 3	Validación 4	Prioridad	Detalle
10	Fecha de cierre	Fecha desde	Fecha de apertura	---	Fecha cierre ≥ fecha desde Fecha cierre ≥ fecha apertura	U Local = CERRADA.	---	8 números consecutivos. En formato aaaammdd	1	Validación Automática
11	Fecha de apertura	Fecha de inicio de actividad	Fecha de inscripción	---	Fecha apertura ≥ Fecha de inicio de actividad. Fecha apertura ≥ Fecha de inscripción	Fecha desde ≤ Fecha de inscripción.	---	8 números consecutivos. En formato aaaammdd	1	Validación Automática
12	Fecha de inscripción	Fecha de inicio de actividad	---	---	Fecha de inicio de actividad ≤ Fecha de inscripción.	8 números consecutivos. En formato aaaammdd	---	---	1	Validación Automática
13	Fecha de actualización	---	---	---	8 números consecutivos. En formato aaaammdd	---	---	---	1	Validación Automática
14	Fecha de reinicio de actividad	Fecha de inscripción	Fecha de inicio de actividad	Fecha de cese de actividad	En caso de que exista un campo vacío, se coloca (-1).	Existe nuevamente fecha de cese la fecha de reinicio de actividad se debe poner en blanco	Fecha reinicio ≥ Fecha de inscripción. Fecha reinicio ≥ Fecha de inicio. Fecha reinicio ≥ Fecha de cese	8 números consecutivos. En formato aaaammdd	1	Debe existir sólo para los registros que presenten fecha de cese y estén activos (Validación Automática)

15	Unidad local estado	1 Abierta	Fecha de apertura	---	---	Existir fecha apertura	Si ha cerrado debe existir: fecha de cese de actividad	Si ha abierto debe existir: fecha de reinicio de actividad	Po lo menos una U legal abierta	1	Validación Automática
		4 Cerrada (Suspensión definitiva de sus actividades)	Fecha de inicio de actividad	Fecha de cierre	---	Existir: fecha de apertura	Existir: fecha de cierre	---	---	1	Validación Automática
16	Obligado a llevar contabilidad		Estrato Ventas	---	---	Ventas del año anterior (a la publicación) > \$ 100.000	---	---	---	2	Validación Automática
17	Unidad local tipo	Desconocido	Fuente	---	---	Consultar a la Fuente	---	---	---	1	Reportar los casos desconocidos (Validación Automática)
		Matriz	Número de unidades locales	---	---	Una sola MATRIZ	---	---	---	1	Validación Automática
		Auxiliar	Número de unidades locales	---	---	Número de Unidad Local diferente	---	---	---	1	Validación Automática

Plan de Validación y Tabulación Otras Variables

# CRUCE	VARIABLE INICIAL	VARIABLE DE COMPARABILIDAD 1	VARIABLE DE COMPARABILIDAD 2	VARIABLE DE COMPARABILIDAD 3	Validación 1	Validación 2	Prioridad	Detalle
1	Teléfono	Provincia			El código telefónico debe corresponder a su respectiva provincia. (EMPRESA)		1	Validación Automática
2	Empleados	Empleados Hombres afiliados	Empleados Mujeres afiliadas	Total empleados afiliados y afiliadas	Empresas $\Sigma PA = \Sigma (PAH + PAM)$ (PA personal afiliado)	Comprobar en cuanto a unidades locales que la suma de sus empleados sea igual a la que reporta la empresa.	1	Validación Automática
3	Ventas	Empresas del sector Publico	Empresas Pasivas		Sector público no debe tener ventas (excepto empresas públicas).	Empresas pasivas: Ventas = 0	1	Validación Automática
4	Ubicación geográfica	Provincia	Cantón	Parroquia	Cada parroquia debe tener un cantón, y el cantón en su respectiva provincia.	Verificar que cada código nuevo que se va a cargar, se encuentre catalogado.		Validación Automática

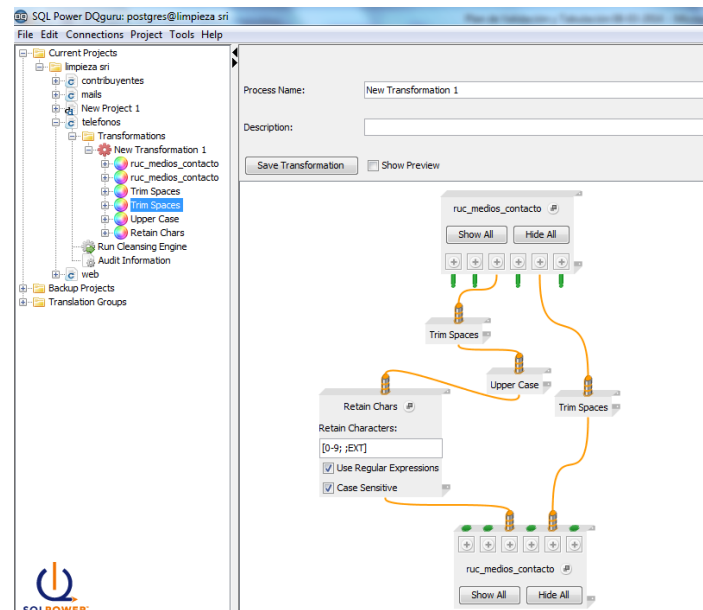
Limpeza de datos

Una vez obtenida la información de las diferentes fuentes, el procedimiento para la limpieza de datos a seguir es el siguiente:

- Congelamiento de la base nueva
- Partir de esta nueva base con la respectiva limpieza, y validación de datos.
- Cargar nueva información a la base del DIEE.

Para el proceso de limpieza de la información se utiliza la herramienta SQL Power DQguru, con la cual se eliminan todos los caracteres extraños o no permitidos en un campo determinado, es importante mencionar que esta herramienta, a diferencia de otras, no solamente realiza un análisis de la información que necesita ser evaluada sino que ataca directamente a la misma eliminando la información innecesaria o errónea. Este proceso se lo realiza especialmente a la tabla de medios de comunicación donde se tiene gran cantidad de caracteres innecesarios que no son permitidos por ejemplo en los números telefónicos. En el resto de tablas también se realiza una limpieza para eliminar los espacios en blanco principalmente.

En el siguiente gráfico se muestra un ejemplo:



A continuación se destacan los procedimientos que se debe realizar a toda la información que se la va a dar una limpieza de datos:

- Colocar los datos en sus respectivas catalogaciones:

- Todos los datos que contengan las palabras www. ó http. Se colocará en tipo web.
- Todos los datos que contengan el símbolo @. Se colocará en tipo correo electrónico.
- Colocar todo en mayúsculas con excepción de datos pertenecientes a correos y páginas web.

Números telefónicos:

Esta limpieza de datos se lo realizará a todos los campos que contengan información del número de teléfono de la siguiente manera:

- Se debe borrar todo dato que no tenga por lo menos un número y no esté compuesto solo por cero.
- Se debe borrar los espacios en blanco.
- Se debe borrar los paréntesis para abrir "(" y cerrar ")"
- Se debe borrar la palabra "REF"
- Se debe borrar el signo punto ".", slash "/", asterisco "*", guión "-", mas "+", llaves "{" y "}", dos puntos ":"
- Se debe borrar toda letra con excepción de datos que contengan la palabra "EXT"
- Se debe borrar los datos de números que terminen en 000000 cuya longitud sea menor a 10
- Cambio de tipo de medio de contacto.
 - Se debe cambiar a tipo celular a los datos cuya longitud es 8 y comienzan con 9 u 8.
 - Se debe cambiar a tipo celular a los datos cuya longitud es 9 y comienzan con 9.
 - Se debe cambiar a tipo celular a los datos cuya longitud es 9 y comienzan con 09 o 08.
 - Se debe cambiar a tipo celular a los datos cuya longitud es 10 y comienzan con 09.
- Los números que tienen solo 6 dígitos, se guardan con tipo de contacto = 9 y quedan pendientes a ser revisados.

- Se debe aumentar el código provincial a los números que solo tengan 7 dígitos.
- Se debe borrar todo dato cuya longitud sea menor a 6 dígitos.
- Se debe aumentar el cero “0” a los datos tengan 8 dígitos e inicien con 2, 3, 4, 5, 6 o 7.
- Se debe aumentar el número 2 luego del código provincial de teléfono para los números de 8 dígitos que coincidían sus dos primeros dígitos con los dígitos de código provincial de teléfono.
- Se debe borrar los números cuya longitud sea de 6 dígitos y empiece con cero.
- Cuando el número comienza con 593 y la longitud es mayor a 8 se borran los 3 dígitos iniciales.
- Datos que presentan la palabra EXT
 - Se debe borrar los datos en donde la palabra EXT esté en una posición menor a 7.
 - Se debe borrar los datos en donde la palabra EXT esté en una posición mayor a 10.
- Se debe aumentar el código provincial a los números con 7 dígitos.
- Se debe aumentar el número 2 luego del código provincial de teléfono para los números de 8 dígitos que coincidían sus dos primeros dígitos con los dígitos de código provincial de teléfono.
- Para los datos cuya longitud es igual a 8, comienzan en cero y no concuerdan sus dos primeros dígitos con el código de teléfono provincial. Se debe colocar en el campo estado el valor de cero “0”, que signifique que la información está inactiva.
- Se debe borrar los datos de números que terminen en 000000 y empiecen con 09990, 0990 y 0910.
- NOTA: esta validación también aplica a números telefónicos de FAX

Celulares:

- Se debe borrar todo dato que no tenga por lo menos un número y no esté compuesto solo por cero.
- Se debe borrar los espacios en blanco.
- Se debe borrar los signos punto “.”, slash “/”, asterisco “*”, guión “-”, mas “+”, llaves “{” y “}”, dos puntos “:”
- Se debe borrar toda letra.
- Se debe borrar los datos que tengan longitud mayor a 10.

- Se debe borrar los datos que tengan longitud menor a 8.
- Se debe aumentar el “09” a los datos de 8 dígitos que comiencen en 9 u 8.
- Se debe aumentar el “09” a los datos de 9 dígitos que comiencen en 0.
- Se debe aumentar todos los datos de 9 dígitos que no comiencen en cero.
- Se debe borrar todos los datos de 10 dígitos que no comiencen en cero.
- Para los datos cuya longitud es igual a 8, se debe colocar en el campo estado el valor de cero “0”, que signifique que la información está inactiva.

Correo Electrónico:

Esta limpieza de datos se la debe realizar de la siguiente manera:

Se debe borrar todo dato que por lo menos no contenga el símbolo “@”

- Se debe borrar todo espacio en blanco.
- Se debe borrar todo dato cuya longitud sea menor a 6.
- Se debe borrar todo dato que no contenga por lo menos una letra.

Web:

Esta limpieza de datos se la debe realizar de la siguiente manera:

Se debe borrar todo espacio en blanco.

- Se debe borrar todo dato cuya longitud sea menor a 3.
- Se debe borrar todo dato que no contenga por lo menos una letra.

Nombre de Contacto:

Esta limpieza de datos se la debe realizar en los datos de los nombres de contactos de las empresas de la siguiente manera:

- Se debe borrar todo número.
- Se debe verificar que se tenga una longitud mínima de 3.
- Se debe borrar los espacios en blanco al principio y al final de los datos.

Detalle geografía:

Se debe validar los campos: provincia, cantón, parroquia y nombre.

- Para la provincia se validará: Que exista por lo menos 3 caracteres y contenga letras de la A-Z se incluye la letra Ñ.
- En caracteres especiales se permite los siguientes: (. \ / Á Ê Í Ó Ú , - ())
- Para cantón parroquia y nombre se admite los anteriores y también números.
- Se debe revisar la base de datos y verificar que no existan símbolos extraños estos campos (como por ejemplo BA♦OS DE AGUA SANTA).

Razón social y Nombre comercial:

- Se verificará con el SRI que la razón social y nombre comercial estén correctamente escritos.
- Se debe verificar que se tenga una longitud mínima de 3.
- Además se debe permitir los caracteres que contenga Ñ, tildes, comillas, @, &, puntos, números, signo de suma +, \, /, #, À,È,Ì,Ò,Ù,Ã,Ë,Ï,Ö,Û , !, _ , °.
- Reemplazar los caracteres extraños correspondientes a problemas de la letra Ñ (Ã), letra Í (Ã), apostrofe (Â´).

Valores por defecto y Tratamiento a los datos que presenten inconsistencias

Después del proceso de limpieza, existirán datos que se borren de la base y muchos campos van a quedar en blanco, para estos casos tenemos valores por defecto que vendrán a llenar estos espacios en blanco.

Pero también tendremos variables que si presentan inconsistencias no se deberá incluir al registro dentro de la Base de Datos.

Variable	Inconsistencia	Valor por defecto	Nota
Unidad Legal			
Numero Ruc	No presenta mínimo 13 dígitos	No aplica	El registro no sube y reportar al DICE
Razón Social	Si la razón social tiene una longitud menor a 3	No aplica	El registro no sube y reportar al DICE
Fecha Desde	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Expediente	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Acto Jurídico	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DICE
Clase Contribuyente	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DICE
Estado Funcionamiento	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DICE
Forma Jurídica	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DICE PN o SC
Unidad Legal Estado	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	El registro no sube y reportar al DICE

Obligado a llevar Contabilidad	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Empresa			
Numero Ruc	No presenta mínimo 13 dígitos	No aplica	El registro no sube y reportar al DIEE
Razón Social	Si la razón social tiene una longitud menor a 3	No aplica	El registro no sube y reportar al DIEE
Nombre Comercial	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Actividad Comercio Exterior	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DIEE
Número de Unidades Locales	En caso de tener unidades locales con el mismo número	No aplica	El registro no sube y reportar al DIEE
Actividad Económica	Si la actividad económica presenta errores o no tiene una correspondencia directa	No aplica	El registro no sube y reportar al DIEE
Sitio Web	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Tipo de Unidad Legal	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	El registro no sube y reportar al DIEE
Fecha de inscripción	No pasa su validación	No aplica	El registro no sube y reportar al DIEE
Fecha de inicio de actividad	No pasa su validación	No aplica	El registro no sube y reportar al DIEE
Fecha de cese de actividad	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Fecha de reinicio de actividad	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Fecha de actualización	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Estado Empresa	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	El registro no sube y reportar al DIEE

Obligado a llevar Contabilidad	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Unidad Local			
Numero Ruc	No presenta mínimo 13 dígitos	No aplica	El registro no sube y reportar al DIEE
Número de unidad local	Existen unidades locales con el mismo número	No aplica	El registro no sube y reportar al DIEE
Nombre Unidad local	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Actividad Económica	Si la actividad económica presenta errores o no tiene una correspondencia directa	No aplica	El registro no sube y reportar al DIEE
Actividad Económica Secundaria	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Producto Elaborado	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Producto Comercializado	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Producto Ofertado	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Materia Prima	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Fecha de cierre	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Fecha de apertura	No pasa su validación	No aplica	El registro no sube y reportar al DIEE
Fecha de inscripción	No pasa su validación	No aplica	El registro no sube y reportar al DIEE
Fecha de actualización	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Fecha de reinicio de actividad	En caso de no existir un valor, se coloca su valor por defecto.	-1	

Unidad local estado	En los casos de que difieran los estados de la unidad local, con el estado de la empresa y la unidad legal no debe subirse el dato	No Aplica	El registro no sube y reportar al DIEE
Obligado a llevar contabilidad	En caso de no existir un valor, se coloca su valor por defecto.	-1	
Unidad local tipo	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar al DIEE

Conteos

Los conteos se los debe realizar después de todas las validaciones y limpiezas de datos para poder saber qué cantidad de datos han subido con éxito y así poder ver si existe coherencia en la cantidad de información entre tablas de la base de datos.

<u>Conteos a realizar</u>	
Contar el número de unidades legales	El número de unidades legales debe ser igual al número de empresas
Contar el número de empresas	
Contar el número de unidades locales	
Contar el número de unidades locales que son matriz	El número de unidades locales matriz debe ser igual al número de empresas y de unidades legales.
Contar las empresas por estado	Reporte
Contar las unidades legales que tienen el campo de geografía = 0 o = null	Resultado debe ser 0
Contar las unidades locales que tienen el campo de geografía = 0 o = null	Resultado debe ser 0

Empleados	El valor de empleados H , M y Total de empresa debe ser igual a la sumatoria de H, M y Total de unidades locales.
Ventas	El valor de ventas de empresa debe ser igual a la sumatoria de ventas de las unidades locales.
El conteo de unidades locales sin dirección debe ser cero.	
El conteo de unidades legales sin dirección debe ser cero, si hay unidades legales sin dirección colocar la dirección de la unidad local matriz.	
Contar las empresas que tienen el campo de numero_unidades_locales = 0 o = null	Resultado debe ser 0
Contar las empresas que no tienen actividad_economica	Resultado debe ser 0
Contar las empresas que tienen el campo obligado_llevar_contabilidad vacio o null	Resultado debe ser 0
Todas las unidades locales deben contar con número de ruc y número de establecimiento	
Todas las unidades legales y empresas deben contar con número de ruc	

Anexos

En los anexos vamos a poder encontrar documentos que contienen información de cómo se procedió con el trabajo de limpieza de datos, validación de campos e información referente y complementaria a este documento, para la obtención de una información de calidad, es por eso que es importante revisar los anexos y a partir de ellos trabajar con las bases de datos.

Validación del RUC:

Ver Anexo 1 (Validación del RUC)

Tipos de validaciones a realizar:

Ver Anexo 2 (Plan de validaciones automáticas y manuales)

Informe Limpieza de Datos

Ver Anexo 3 (Informes Limpieza de Datos)

Plan de Inconsistencias con cruce de Variables

Ver Anexo 4 (Plan de Inconsistencias con cruce de Variables)

Conclusiones:

El documento ha recopilado los procesos de validación, limpieza y conteos de las variables que van a entrar a la base de datos del DICE, validaciones que han nacido a partir de la experiencia y el trabajo que se viene realizando en el directorio de empresas, podemos concluir que las validaciones a seguir, y todos los procesos que se detallan en el presente documento serán de gran ayuda para los futuros trabajos que vamos a tener en el Directorio de Empresas, de esta manera la información que el Directorio publica será una información veraz, y los procesos que conlleva tener esta información cada vez serán más ágiles y automáticos.

Recomendaciones:

Se recomienda tener siempre en cuenta todas las validaciones que tiene el documento, aquí se puede encontrar como proceder de manera correcta al momento del trabajo en la base de datos antes de subir la información a la base del DICE.

Si en el futuro se generan nuevas validaciones, se recomienda documentarlas, para poder agregarlas a este documento y así tener siempre un Plan de Validación y Tabulación actualizado.