



Plan de Validación y Tabulación.

Directorio de Empresas y Establecimientos



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)	8
Plan de Validación y Tabulación u_local (Variables Independientes)	9
Plan de Validación y Tabulación u_local (Variables Dependientes)	10
Plan de Validación y Tabulación Otras Variables	12
Valores por defecto y Tratamiento a los datos que presenten inconsistencias	13
Conteos	16
Anexos.....	17
Conclusiones:	18
Recomendaciones:.....	18



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 en cada 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 aaammdd	---	---	3	Validación Automática
3	Expediente	Personas Naturales no tienen expediente.	---	---	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática
4	Acto Jurídico	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática
5	Clase Contribuyente	Verificación Superintendencia de Compañías	Verificación Superintendencia de Compañías	---	3	Validación Automática
		Consultar a la fuente	Reportar los casos desconocidos	---	3	Validación Automática
		Verificar coincidencia con la fuente	Reportar los casos desconocidos	---	3	Validación Automática
		Verificar coincidencia con la fuente	---	---	3	Validación Automática
		Verificar coincidencia con la fuente	---	---	3	Validación Automática
		Verificar coincidencia con la fuente	---	---	3	Validación Automática
6	Estado Funcionamiento	Consultar a la fuente	Verificación Superintendencia de Compañías	Reportar casos desconocidos	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática
		Verificación Superintendencia de Compañías	---	---	3	Validación Automática

	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
	Intervención AGD	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
	Cambio de estado de la empresa	Verificación Superintendencia de Compañías	---	---	3	Validación Automática
	Desconocido Personal Natural	Consultar a la fuente	Reportar los casos desconocidos	---	2	Validación Automática
7	Forma Jurídica	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
	Desconocido	Fuente	---	---	Consultar a la Fuente	---	---	1	Reportar los casos desconocidos (Validación Automática)
8	Unidad Legal Estado 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.	Validación Automática
	Por Revisar	Fuente	---	---	Consultar a la Fuente	---	---	1 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	Nombre Comercial	Validar vs la Fuente SRI	---	---	1	Validación automática
	Mercado Interno	Debe tener valor codificado (00 ó 99)	---	---	3	Validación automática
	Actividad Comercio Exterior	Debe tener valor codificado (01)	---	---	3	Validación automática
2	Exportador	Debe tener valor codificado (02)	---	---	3	Validación automática
	Importador / Exportador	Debe tener valor codificado (03)	---	---	3	Validación automática
3	Número de Unidades Locales	Valores >= 0	Cada unidad local no puede tener el mismo número	---	2	Validación automática

4	Actividad Económica	Descripción = Código CIU 4	Realizar un cruce CPC vs CIU4 casos CENEC	---	1	Validación automática
---	----------------------------	----------------------------	---	-----	---	-----------------------

Plan de Validación y Tabulación Empresa (Variables Dependientes)

# CR UC E	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
5	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
6	Fecha de inicio de actividad	Fecha de inscripción	---	---	Debe existir Fecha de inscripción Fecha cese ≥ Fecha de inscripción.	Fecha de inicio de actividad ≥ fecha de inscripción.	8 números consecutivos. En formato aaaammdd	---	1	Validación automática
7	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
8	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 de reinicio ≥ Fecha de inicio. Fecha reinicio ≥ Fecha de cese	8 números consecutivos. En formato aaaammdd	---	1	Validación automática
9	Fecha de actualización	---	---	---	8 números consecutivos. En formato aaaammdd		---	---	1	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 AD 1	VARIABLE DE COMPARABILIDAD DAD 2	VARIABLE DE COMPARABILIDAD AD 3	Validación 1	Validación 2	Validación 3	Validación 4	Prioridad	Detalle
10	Fecha de cierre	Fecha desde	Fecha de apertura	---	Fecha cierre ≥ fecha de apertura Fecha desde ≥ fecha de 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 de apertura ≥ Fecha de inscripción Fecha de inicio de actividad ≥ 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)

		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	Validación Automática
15	Unidad local estado	1 Abierta	---	---	---				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	---	---	Validación Automática
17	Unidad local tipo	Desconocido		---	Consultar a la Fuente				Reportar los casos desconocidos (Validación Automática)
		Matriz	Número de unidades locales	---	Una sola MATRIZ	---	---	---	Validación Automática
		Auxiliar	Número de unidades locales	---	Número de Unidad Local diferente	---	---	---	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 Público	Empresas Pasivas		Sector público no debe tener ventas (excepto empresas públicas). Cada parroquia debe tener un cantón, y el cantón en su respectiva provincia.	Empresas pasivas: Ventas = 0 Verificar que cada código nuevo que se va a cargar, se encuentre catalogado.	1	Validación Automática
4	Ubicación geográfica	Provincia	Cantón	Parroquia				Validación Automática

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 DIEE
Razón Social	Si la razón social tiene una longitud menor a 3	No aplica	El registro no sube y reportar al DIEE
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 DIEE
Clase Contribuyente	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DIEE
Estado Funcionamiento	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DIEE
Forma Jurídica	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DIEE
Unidad Legal Estado	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DIEE
Obligado a llevar Contabilidad	En caso de no existir un valor, se coloca su valor	-1	Reportar los casos al



	por defecto.		DIEE
Empresa			
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
Actividad Económica	Si la actividad económica presenta errores o no tiene una correspondencia directa	No aplica	Reportar los casos al DIEE
Fecha de inscripción	No pasa su validación	No aplica	Reportar los casos al DIEE
Fecha de inicio de actividad	No pasa su validación	No aplica	Reportar los casos 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	
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	Reportar los casos 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 se debe 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	Reportar los casos al DIEE
Fecha de inscripción	No pasa su validación	No aplica	Reportar los casos 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 la unidad legal	No Aplica	Reportar los casos al DIEE
Unidad local tipo	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos 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>	
CONTEOS	VERIFICACIÓN
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 es la sumatoria de ventas tarifa 0%, ventas tarifa 12% y exportaciones netas.
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 de la base de datos del DIEE, 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 es una información veraz, y los procesos que conllevan a tener esta información serán cada vez 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 DIEE.

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.

Fecha de elaboración: 07 de octubre de 2016		
Elaborado por:	Carolina Mina	
Aprobado por:	Libertad Trujillo	