



Plan de Validación y Procesamiento

DIEE 2020

Octubre, 2021





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.

El software que sirve como herramienta para la Validación de variables, se emplea para la realización de validaciones automáticas, pero también dentro del proceso es indispensable la realización de validaciones manuales, todo esto con el objetivo de obtener información de mejor calidad.

Para la validación, es importante considerar que existen variables que dependen de otras para ser validadas y otras que son independientes, es decir, que no requieren de ninguna variable adicional para su validación.

El presente documento contiene el plan de validación y tabulación de las variables que existen en el DICE, 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 unidad legal (Variables Independientes)

# CRUCE	VARIABLE INICIAL	Validación 1	Validación 2	Prioridad	Detalle	
1	Expediente	Personas Naturales no tienen expediente.	---	3	Validación Automática	
2	Clase Contribuyente	Desconocido	Reportar los casos desconocidos	---	1	Validación Automática
		Especial	Debe tener valor codificado (1)	---	1	Validación Automática
		Otros	Debe tener valor codificado (2)	---	1	Validación Automática
		RISE	Debe tener valor codificado (3)	---	1	Validación Automática
3	Unidad legal tipo	Desconocido	Consultar a la fuente	Reportar los casos desconocidos	1	Validación Automática
		Personal Natural	Verificar coincidencia con la fuente	---	1	Validación Automática
		Persona jurídica	Verificar coincidencia con la fuente	---	1	Validación Automática
4	Obligado a llevar Contabilidad	Desconocido	Reportar los casos desconocidos	---	1	Validación Automática



Plan de Validación y Tabulación unidad 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
www.ecuadorencifras.gob.ec	Unidad Legal Estado	Desconocido	Fuente	---	---	Consultar a la Fuente	---	---	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
		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
		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
		Cerrado	Fecha de inscripción	Fecha de inicio actividad	Fecha de cese actividad	Debe existir: fecha de inscripción	Debe existir: fecha de inicio de	Debe existir: fecha de cese de	1





							actividad	actividad.		
2	Obligado a llevar Contabilidad	Ventas	---	---	Ventas del año (a la publicación) > \$ 100.000	---	---	---		1

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	
2	Actividad Comercio Exterior	Mercado Interno	Debe tener valor codificado (4)	---	---	3	Validación automática
		Importador	Debe tener valor codificado (2)	---	---	3	Validación automática
		Exportador	Debe tener valor codificado (1)	---	---	3	Validación automática
		Importador / Exportador	Debe tener valor codificado (3)	---	---	3	Validación automática
3	Número de Unidades Locales	Valores >= 0	Cada unidad local no puede tener el mismo número	---	2	Validación automática	





# CRUCE	VARIABLE INICIAL	Validación 1	Validación 2	Validación 3	Prioridad	Detalle
4	Actividad Económica	Descripción a Código CIU 4	---	---	1	Validación automática

Plan de Validación y Tabulación unidad local (Variables Independientes)

# CRUCE	VARIABLE INICIAL	Validación 1	Validación 2	Validación 3	Prioridad	Detalle
1	Número de unidad local	Al menos una unidad local registrada.	No pueden existir unidades locales con el mismo número de unidad local.	Validar que no exista el número cero en el identificador.	1	Validación Automática
2	Nombre Unidad local	> 3 dígitos	---	---	1	Validación Automática
3	Actividad Económica	Descripción a Código CIU4	---	---	1	Validación Automática





Plan de Validación y Tabulación unidad local (Variables Dependientes)

# CRUCE	VARIABLE INICIAL	VARIABLE DE COMPARABILIDAD 1	VARIABLE DE COMPARABILIDAD 2	Validación 1	Validación 2	Validación 3	Prioridad	Detalle	
4	Unidad local estado	1 Abierta	Fecha de apertura	---	Existir fecha apertura	Por lo menos una unidad local abierta	3		
		2 Cerrada (Suspensión definitiva de sus actividades)	Fecha de inicio de actividad	Fecha de cierre	Existir: fecha de apertura	Existir: fecha de cierre	---	1	
5	Unidad local tipo	Desconocido	Fuente	---	Consultar a la Fuente	---	---	Reportar los casos desconocidos (Validación Automática)	
		Matriz	Número de unidades locales	---	Una sola MATRIZ	---	---	1	
		Auxiliar	Número de unidades locales	---	Número de Unidad Local diferente	---	---	1	





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.		1	
2	Empleo registrado	Empleados Hombres	Empleados Mujeres	Total empleo	Empresas $\Sigma ER = \Sigma (ERH + ERM)$ (ER empleo registrado)		1	
3	Plazas de empleo registrado	Plazas Hombres	Plazas Mujeres	Total plazas de empleo registrado	Empresas $\Sigma PE = \Sigma (PEH + PEM)$ (PE plazas de empleo registrado)	Comprobar en cuanto a unidades locales que la suma de plazas de empleo registrado sea igual a la que reporta la empresa.	1	
4	Remuneraciones	Remuneraciones Hombres	Remuneraciones Mujeres	Total remuneraciones	Empresas $\Sigma R = \Sigma (RH + RM)$ (R remuneraciones)	No deben existir valores negativos	1	
5	Ubicación geográfica	Provincia	Cantón	Parroquia	Si está incompleta o termina en 99,	Los casos que no se pudo recuperar,	1	





					tomar geografía del establecimiento matriz, para empresas con establecimiento único.	entregar un listado de los casos restantes al equipo de Análisis.		
6	Forma institucional	Forma institucional	Unidad legal tipo	Obligado a llevar contabilidad	Si ulegal_tipo=1 & forma_inst=7 & obligado=1 → forma_inst=2	Si ulegal_tipo=1 & forma_inst=7 & obligado=2 → forma_inst=3	1	Validación Automática
7	Clase contribuyente	Clase contribuyente	Declaración RISE	Forma institucional	Si (clase_contrib=3 & forma_inst<>1) RISE=1 → clase_contrib = 2	Si clase_contrib=2 & RISE=1 & forma_inst=1 → clase_contrib = 3	1	Validación Automática
8	Estratos ventas, empleo y plazas	Estrato ventas o empleo y plazas	Ventas o empleo y plazas		Si tiene ventas>0 o RISE= 1, debe tener estrato ventas. Si tiene empleo o plazas debe tener estrato empleo o plazas		1	Validación Automática
9	Tamaño	Estrato ventas o empleo y plazas			Si tiene estrato de ventas o estrato empleo o estrato plazas debe tener tamaño		1	Validación Automática
10	Situación	Tamaño			Si tiene tamaño debe tener situación		1	Validación Automática





					siempre y cuando no tenga codigo_seccion en: X, T, U, W, Z			
11	Rangos por edades de plazas de empleo registrado	1.Rango de 15 a 17 años 2.Rango de 18 a 29 años 3.Rango de 30 a 44 años	4.Rango de 45 a 64 años 5.Rango de mayor o igual a 65 años	Total de plazas de empleo registrado	Empresas $\Sigma PE = \Sigma (RP1 + RP2 + RP3 + RP4 + RP5)$ (PE es plazas de empleo y RP es rango de edad por plazas)	No deben existir valores negativos	1	
12	Rangos por edades de empleo registrado	1.Rango de 15 a 17 años 2.Rango de 18 a 29 años 3.Rango de 30 a 44 años	4.Rango de 45 a 64 años 5.Rango de mayor o igual a 65 años	Total de empleo registrado	Empresas $\Sigma ER = \Sigma (RE1 + RE2 + RE3 + RE4 + RE5)$ (ER es empleo registrado y RE es rango de edad por empleo)	No deben existir valores negativos	1	





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			
Razón Social	Si la razón social tiene una longitud menor a 3	No aplica	El registro no sube y reportar al DICE
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	
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	
Forma Jurídica	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	
Unidad Legal Estado	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	Reportar los casos al DICE
Obligado a llevar Contabilidad	En caso de no existir un valor, se coloca su valor por defecto.	-1	Reportar los casos al DICE
Empresa			
Nombre Comercial	En caso de no existir un valor o se tiene la palabra <i>null</i> , se coloca su valor por defecto.	-1	
Actividad Comercio Exterior	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	
Actividad	Si la actividad	Se asigna el valor	



Variable	Inconsistencia	Valor por defecto	Nota
Económica	económica presenta errores o no tiene una correspondencia directa	Z000000	
Fecha de inscripción	No pasa su validación	No aplica	
Fecha de inicio de actividad	No pasa su validación	No aplica	
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 la fecha de corte de la información.	Ejm: 20171231	
Unidad Local			
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	Se asigna el valor Z000000	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	





Variable	Inconsistencia	Valor por defecto	Nota
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	
Fecha de inscripción	No pasa su validación	No aplica	
Fecha de actualización	En caso de no existir un valor, se coloca la fecha de corte de la información.	Ejm: 20171231	
Fecha de reinicio de actividad	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 los casos al DICE

Valores por defecto para empresas del Ministerio de Educación

Variable	Valor por defecto	Descripción
Clase contribuyente	5	No aplica
Obligado a llevar contabilidad	3	No aplica
Actividad comercio exterior	5	No aplica
Forma institucional	7	Institución Pública

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.

Conteos a realizar	
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.



Conteos a realizar	
CONTEOS	VERIFICACIÓN
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
Plazas de empleo registrado	El valor de plazas de empleo registrado hombres, mujeres y total de empresa debe ser igual a la sumatoria de las plazas de empleo registrado hombres, mujeres 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 vacío 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.

Tipos de validaciones a realizar

Ver Anexo 1 (Plan de validaciones automáticas y manuales)

Informe Limpieza de Datos

Ver Anexo 2 (Informes Limpieza de Datos)





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, 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: 13 de octubre de 2021	
Elaborado por: Carolina Mina	Aprobado por: Libertad Trujillo