

## **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 .....	14
Conteos .....	17
Anexos.....	18
Validación del RUC:.....	18
Tipos de validaciones a realizar:.....	18
Informe Limpieza de Datos.....	18
Ver Anexo 3 (Informes Limpieza de Datos).....	18
Plan de Inconsistencias con cruce de Variables.....	¡Error! Marcador no definido.
Conclusiones: .....	19
Recomendaciones:.....	19

## 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 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 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:**

<b>Prioridad</b>	<b>Descripción</b>
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 las publicadas.

Baja: Variables que no tienen relación con la Publicación.

## Plan de Validación y Tabulación u\_legal (Variables Independientes)

GRUPO	VARIABLES INICIAL	Validación 1	Validación 2	Validación 3	Propiedad	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
		Verificar coincidencia con la fuente	Reportar los casos desconocidos	---	2	Validación Automática
4	Clase Contribuyente	Desconocido	---	---	2	Validación Automática
		Especial	Verificar coincidencia con la fuente	---	2	Validación Automática
		Otros	Verificar coincidencia con la fuente	---	2	Validación Automática
		RISE	Verificar coincidencia con la fuente	---	2	Validación Automática
5	Forma Jurídica	Desconocido	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)

GRUPO	VARIABLE/INICIAL	VARIABLE DE COMPARABILIDAD 1	VARIABLE DE COMPARABILIDAD 2	VARIABLE DE COMPARABILIDAD 3	Validación 1	Validación 2	Validación 3	Error# and	Detalle	
7	<b>Unidad Legal Estado</b>	<b>Desconocido</b>	Fuente	---	---	Consultar a la Fuente	---	1	Reportar los casos desconocidos (Validación Automática)	
		<b>Activa</b>	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	
		<b>Pasiva</b>	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	
		<b>Suspensión definitiva</b>	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	
		<b>Cerrado</b>	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	
	<b>Por Revisar</b>	Fuente	---	---	Consultar a la Fuente	---	---	1	Reportar los casos desconocidos (Validación Automática)	
8	<b>Obligado a llevar Contabilidad</b>	Desconocido	---	---	Reportar los casos desconocidos (3er dígito RUC < 6 ) =PERSONA NATURAL	---	---	2	Validación Automática	
9	<b>Razón Social</b>	RUC	---	---	---	(3er dígito RUC ≥ 6) = EMPRESAS	---	1	Validación Automática	

## Plan de Validación y Tabulación Empresa (Variables Independientes)

GRUPO	VARIABLE INICIAL	Validación 1	Validación 2	Validación 3	Prioridad	Detalle
1	Nombre Comercial	Validar vs la Fuente SRI	---	---	1	Validación automática
	Actividad Comercio Exterior	Mercado Interno	Debe tener valor codificado (00 ó 99)	---	3	Validación automática
		Importador / Exportador	Debe tener valor codificado (01) / Debe tener valor codificado (02)	---	---	3
3	Número de Unidad Local	Importador / Exportador	Debe tener valor codificado (03)	---	3	Validación automática
			Valores >= 0	---	2	Validación automática
4	Actividad Económica	Descripción = Código CIU 4	---	---	1	Validación automática

## Plan de Validación y Tabulación Empresa (Variables Dependientes)

GR UC	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	Propiedad	Detalle
5	Fecha de inscripción	Fecha de inicio de actividad	---	---	Debe existir Fecha de inicio.	8 números consecutivos. En formato aaaaammdd		---	1	Validación automática
6	Fecha de inicio de actividad	Fecha de inscripción	---	---	Debe existir Fecha de inscripción	8 números consecutivos. En formato aaaaammdd		---	1	Validación automática
7	Fecha de cese de actividad	Fecha de inscripción	Fecha de inicio de actividad	---	8 números consecutivos. En formato aaaaammdd			---	1	Validación automática
8	Fecha de reinicio de actividad	Fecha de inscripción	Fecha de inicio de actividad	Fecha de cese de actividad	8 números consecutivos. En formato aaaaammdd			---	1	Validación automática
9	Fecha de actualización	---	---	---	8 números consecutivos. En formato aaaaammdd		---	---	1	Validación automática

## Plan de Validación y Tabulación u\_local (Variables Independientes)

CORRE	VARIABLE INICIAL	Validación 1	Validación 2	Validación 3	Prunidad	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 de unidad local.	Validar que no exista el número cero en el identificador.	1	Validación Automática
3	Nombre Unidad local	> 3 dígitos	---	---	2	Validación Automática
4	Actividad Económica	Descripción = Código CIU 4		---	1	Validación Automática

## Plan de Validación y Tabulación u\_Local (Variables Dependientes)

4	GRUPO	VARIABLES INICIALES	VARIABLE DE COMPARABILIDAD AD1	VARIABLE DE COMPARABILIDAD AD2	VARIABLE DE COMPARABILIDAD AD3	Validación 1	Validación 2	Validación 3	Validación 4	Prioridad	Detalle
5		Fecha de cierre	Fecha desde	Fecha de apertura	---	8 números consecutivos. En formato aaaaammdd	---	---	---	3	Validación Automática
11		Fecha de apertura	Fecha de inicio de actividad	Fecha de inscripción	---	8 números consecutivos. En formato aaaaammdd	---	---	---	3	Validación Automática
12		Fecha de inscripción	Fecha de inicio de actividad	---	---	8 números consecutivos. En formato aaaaammdd	---	---	---	3	Validación Automática
13		Fecha de actualización	---	---	---	8 números consecutivos. En formato aaaaammdd	---	---	---	3	Validación Automática
14		Fecha de reinicio de actividad	Fecha de inscripción	Fecha de inicio de actividad	Fecha de cese de actividad	8 números consecutivos. En formato aaaaammdd	---	---	---	3	Debe existir sólo para los registros que presentan fecha de cese. Validación Automática
15	Unidad local estado	1 Abierta	Fecha de apertura	---	---	Existir: fecha apertura	Existir: fecha de cierre	---	---	3	Por lo menos una U local abierta
		4 Cerrada (Suspensión definitiva de sus actividades)	Fecha de inicio de actividad	Fecha de cierre	---	Existir: fecha de apertura	Existir: fecha de cierre	---	---	1	

17	Unidad local tipo	Desconocido		Fuente	Número de unidades locales	Fuente	Número de unidades locales	Consultar a la Fuente	Número de Unidades Local diferente	Reportar los casos desconocidos (Validación Automática)
		Matriz	Audiliar							
								Consultar a la Fuente		1
								Una sola MATRIZ		1
										1

## Plan de Validación y Tabulación Otras Variables

GRUPO	VARIABLE DE IDENTIFICACIÓN	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)	Comprobar en cuanto a unidades locales que la suma de sus empleados sea igual a la que reporta la empresa.	1	Validación Automática
2	Empleados	Empleados Hombres afiliados	Empleados Mujeres afiliadas	Total empleados afiliados y afiliadas	Empresas SPA = 2 (PAH + PAM) (PA personal afiliado)		1	Validación Automática
3	Remuneraciones	Remuneraciones Hombres	Remuneraciones Mujeres	Total remuneraciones	Empresas DR = 2 (RH + RM) (R remuneraciones)	No deben existir valores negativos	1	Validación Automática
4	Ventas	Empresas del sector Público			Sector público no debe tener ventas (excepto empresas públicas).	Los casos que no se pudo recuperar, entregar un listado de los casos restantes al equipo de Análisis.	1	Validación Automática
5	Ubicación geográfica	Provincia	Cantón	Parroquia	Si ilegal_tipo=1 & forma_inst=7 & obligado='N' → forma_inst=2	Si ilegal_tipo=1 & forma_inst=7 & obligado='S' → forma_inst=3	1	Validación Automática
6	Forma institucional	Forma institucional	Unidad legal tipo	Obligado a llevar contabilidad	Si ilegal_tipo='RIS' & forma_inst<>1    RIS=1 → clase_contrib = 'OTR'	Si ilegal_tipo='OTR' & RIS=1 & forma_inst=1 → clase_contrib = 'RIS'	1	Validación Automática
7	Clase contribuyente	Clase contribuyente	Declaración RISE	Forma institucional			1	Validación Automática

8	Estratos ventas y empleo	Estrato ventas o empleo	Ventas o empleo		Si tiene ventas-0 debe tener estrato ventas. Si tiene empleo debe tener estrato empleo		1	Validación Automática
9	Tamaño	Estrato ventas o empleo			Si tiene estrato de ventas o estrato empleo debe tener tamaño		1	Validación Automática
10	Situación	Tamaño			Si tiene tamaño debe tener situación		1	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.

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

por defecto.

### Empresa

<b>Nombre Comercial</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Actividad Comercio Exterior</b>	En caso de no existir un valor, se coloca su valor catalogado	Desconocido	
<b>Actividad Económica</b>	Si la actividad económica presenta errores o no tiene una correspondencia directa	No aplica	Reportar los casos al DIEE
<b>Fecha de inscripción</b>	No pasa su validación	No aplica	Reportar los casos al DIEE
<b>Fecha de inicio de actividad</b>	No pasa su validación	No aplica	Reportar los casos al DIEE
<b>Fecha de cese de actividad</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Fecha de reinicio de actividad</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Fecha de actualización</b>	En caso de no existir un valor, se coloca la fecha de corte de la información.	Ejm: 20171231	

### Unidad Local

<b>Numero Ruc</b>	No presenta mínimo 13 dígitos	No aplica	El registro no sube y reportar al DIEE
<b>Número de unidad local</b>	Existen unidades locales con el mismo número	No aplica	Reportar los casos al DIEE
<b>Nombre Unidad local</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Actividad Económica</b>	Si la actividad económica presenta errores o no tiene una correspondencia directa	No aplica	El registro se debe reportar al DIEE
<b>Actividad Económica Secundaria</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Producto Elaborado</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	

<b>Producto Comercializado</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Producto Ofertado</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Materia Prima</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Fecha de cierre</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Fecha de apertura</b>	No pasa su validación	No aplica	Reportar los casos al DICE
<b>Fecha de inscripción</b>	No pasa su validación	No aplica	Reportar los casos al DICE
<b>Fecha de actualización</b>	En caso de no existir un valor, se coloca la fecha de corte de la información.	Ejm: 20171231	
<b>Fecha de reinicio de actividad</b>	En caso de no existir un valor, se coloca su valor por defecto.	-1	
<b>Unidad local estado</b>	En los casos de que difieran los estados de la unidad local, con la unidad legal	No Aplica	Reportar los casos al DICE
<b>Unidad local tipo</b>	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
<b>Clase contribuyente</b>	NA	No aplica
<b>Obligado a llevar contabilidad</b>	NA	No aplica
<b>Actividad comercio exterior</b>	5	No aplica

## 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.

<b><u>Conteos a realizar</u></b>	
<b>CONTEOS</b>	<b>VERIFICACIÓN</b>
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%, exportaciones netas, ingresos por comisiones, ingresos por arrendamientos e ingresos por

	servicios de construcción.
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)

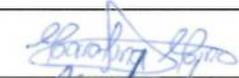
## Conclusiones:

El documento ha recopilado los procesos de validación, limpieza y conteos de las variables de 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 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 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.

<b>Fecha de elaboración:</b> 05 de octubre de 2018		
<b>Elaborado por:</b>	Carolina Mina	
<b>Aprobado por:</b>	Libertad Trujillo	