



Informe de Modificación de variables

DIEE 2019

Marzo, 2021

www.ecuadorencifras.gob.ec



sembramos
Futuro

Lenín





01.

Informe de modificación de variables de la base DICE 2019.





Antecedentes

La unidad de gestión de estadísticas económicas en base a registros administrativos - GEERA y en función al calendario establecido anualmente, calcula y publica los resultado del Directorio de Empresas y Establecimientos DIEE, operación estadística que se construye a partir de registros administrativos, que se obtienen de dos entidades técnicas, autónomas, con personería jurídica y derecho público, como lo son: el Servicio de Rentas Internas (SRI) y el Instituto Ecuatoriano de Seguridad Social (IESS).

En la construcción del DIEE, se toma en cuenta estándares mínimos necesarios de acuerdo a experiencias y recomendaciones internacionales, desde su captación hasta llegar a su difusión, etapa que también cuenta con normas de publicación; adicional a ello también se considera las necesidades de usuarios de la operación estadística.

En GEERA anualmente se genera una serie de documentación para publicar los resultados del DIEE, documentación que se elabora en función a normas establecidas institucionalmente; esta serie de documentos son elaborados por el personal técnico de la unidad, con ello se da a conocer los resultados del DIEE.

Por ello, el presente documento constituye una relatoría de los cambios realizados en la base de datos pública del DIEE 2019, debido al análisis de riesgo de re identificación. Con esta modificación se garantiza la privacidad de la información a publicar del Directorio de Empresas y Establecimientos del periodo 2019 - DIEE2019.

Desarrollo

Se analizó el cambio de las siguientes variables públicas:

Tabla 1. Análisis de riesgo de reidentificación para cada variable – DIEE 2019

VARIABLE	PROBABILIDAD DE IDENTIFICACIÓN			TOTAL		NIVEL RIESGO
	100%	50%	(50%-20%)	COUNT	%	
id_empresa *	67.363	3.048	66	70.477	100%	5
ventas_totales	8.724	304	158	9.186	13,03%	5
ventas_nacionales	8.673	304	157	9.134	12,96%	5
remuneraciones	11.577	4.514	4.196	20.287	28,79%	5
remuneraciones_hombres	8.981	3.168	3.037	15.186	21,55%	5
remuneraciones_mujeres	7.433	2.544	2.138	12.115	17,19%	5
fecha_inicio_actividad	3.106	2.234	6.239	11.579	16,43%	5
fecha_reinicio_actividad	1.173	1.394	5.410	7.977	11,32%	5
fecha_inscripcion	1.556	1.558	6.228	9.342	13,26%	5
fecha_suspension_definitiva	1.728	1.276	2.044	5.048	7,16%	4
fecha_cese_actividad	1.251	932	1.291	3.474	4,93%	3
fecha_actualizacion	616	556	1.361	2.533	3,59%	3



- Las variables ventas totales, ventas nacionales, ventas tarifa 0, ventas tarifa 12, ventas servicios construcción, ventas arrendamientos, ventas comisiones y exportaciones netas, además de remuneraciones, remuneraciones hombres y remuneraciones mujeres, estas variables se expresan en miles de dólares.
- Las variables, fecha inicio de actividad, fecha reinicio actividad, fecha inscripción, fecha suspensión definitiva, fecha cese actividad y fecha actualización.

1. Sintaxis.

Se elaboró una sintaxis en Spss, con la cual se expresó ventas y remuneraciones en miles de dólares además se incluye el cambio de las fechas las cuales se expresan según el año y trimestre ("2012 I Trimestre").

Anexo: Sintaxis cambios BDD

Ventas

ventas_totales	ventas_nacionales	exportaciones_netas	ventas_tarifa0	ventas_tarifa12	ventas_servicios_construccion	ventas_arrendamientos
-	-	-	-	-	-	-
-	-	-	-	-	-	-
0	0	0	0	0	0	-
0	0	0	0	0	0	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
0	0	0	0	0	0	-
0	0	0	0	0	0	-
221	221	0	6	215	-	-
230	230	0	6	224	-	-
256	256	0	9	248	-	-
257	257	0	13	244	-	-
224	224	0	13	210	-	-
203	203	0	0	203	-	-
196	196	0	11	185	-	-
195	195	0	0	195	-	-

Fechas

fecha_inicio_actividad	fecha_cese_actividad	fecha_reinicio_actividad	fecha_inscripcion	fecha_actualizacion	fecha_suspension_definitiva
2001 I Trimestre		2012 II Trimestre	2001 I Trimestre	2012 II Trimestre	
2001 I Trimestre		2012 II Trimestre	2001 I Trimestre	2012 II Trimestre	
2001 I Trimestre		2012 II Trimestre	2001 I Trimestre	2012 II Trimestre	
2001 I Trimestre		2012 II Trimestre	2001 I Trimestre	2012 II Trimestre	
1995 III Trimestre			1983 II Trimestre		
1983 I Trimestre		2013 IV Trimestre	1983 II Trimestre		
1983 I Trimestre		2013 IV Trimestre	1983 II Trimestre		
1983 I Trimestre		2013 IV Trimestre	1983 II Trimestre		
1983 I Trimestre		2013 IV Trimestre	1983 II Trimestre	2014 I Trimestre	
1983 I Trimestre		2013 IV Trimestre	1983 II Trimestre	2014 I Trimestre	2017 IV Trimestre
1983 I Trimestre		2018 III Trimestre	1983 II Trimestre	2018 III Trimestre	2017 IV Trimestre
1983 I Trimestre		2018 III Trimestre	1983 II Trimestre	2018 III Trimestre	2019 II Trimestre
1977 II Trimestre	1981 IV Trimestre	2000 I Trimestre	1981 IV Trimestre	2008 III Trimestre	
1977 II Trimestre	1981 IV Trimestre	2000 I Trimestre	1981 IV Trimestre	2008 III Trimestre	1981 IV Trimestre
1977 II Trimestre	1981 IV Trimestre	2000 I Trimestre	1981 IV Trimestre	2008 III Trimestre	1981 IV Trimestre
1996 I Trimestre			2001 IV Trimestre		
2001 III Trimestre			2001 IV Trimestre		2013 II Trimestre
2001 III Trimestre			2001 IV Trimestre		2013 II Trimestre
2001 III Trimestre			2001 IV Trimestre	2012 IV Trimestre	2013 II Trimestre
2016 I Trimestre			2016 I Trimestre		
2016 I Trimestre			2016 I Trimestre		
2012 III Trimestre			2012 III Trimestre		
2012 III Trimestre			2012 III Trimestre		
2012 III Trimestre			2012 III Trimestre		
2012 III Trimestre			2012 III Trimestre		



Remuneraciones

remuneraciones	remuneraciones_hombres	remuneraciones_mujeres
-	-	-
-	-	-
18	14	4
25	15	10
-	-	-
-	-	-
-	-	-
8	2	5
12	5	8
30	15	16
33	15	17
30	14	16
26	12	14
28	16	12
30	18	12
35	22	12
34	21	12
-	-	-
-	-	-
-	-	-
-	-	-
2	1	1
8	4	4
8	4	4
0	4	4

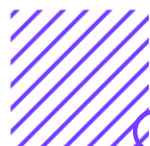
2. Validaciones en R

Se incluye un script con validaciones para la base de datos modificada, cuenta con validaciones para ventas, remuneraciones y fechas, se creó un pequeño catalogo para poder realizar las comparaciones de fechas y trimestres, además en lo que respecta a ventas y remuneraciones se utilizó las validaciones de la malla I solo se modificó el valor máximo y mínimo ya que este debía estar expresado en miles de dólares.

Anexo: Malla_I_base_modificada
diee_catologo

3. Tiempos

ACTIVIDAD	TIEMPO	RESULTADO
Modificar las variables con la sintaxis	20 minutos	Base de datos con las variables expresadas en miles de dólares y las fechas con año y trimestre.
Corre el script de validación con las variables modificadas	20 minutos en cargar la base a R 20 minutos en correr las validaciones	Base de datos validada



Conclusión:

- Al final se realizó las pruebas con 500 casos y también con la base completa de esta forma se obtuvo el correcto funcionamiento de la sintaxis y el script en R. Se obtiene una base de datos con variables que no permitan identificar a las empresas.

Anexos

Excel - Comprobación totales

Elaborado por:	Jacqueline Imbaquingo	
Revisado por:	Libertad Trujillo	
Fecha:	22/03/2021	



INEC | Buenas cifras,
mejores vidas



@ecuadorencifras



@ecuadorencifras



@InecEcuador



t.me/euadorencifras



INEC/Ecuador



INECEcuador