



Aprovechamiento de parados en encuestas de hogares para el aseguramiento de la calidad

Subdepartamento de Estadísticas Estructurales del Trabajo
Subdirección Técnica

Equipo de Ciencia de Datos, Unidad de Gobierno de Datos

INSTITUTO NACIONAL DE ESTADÍSTICAS

Noviembre 2024

ine.gob.cl



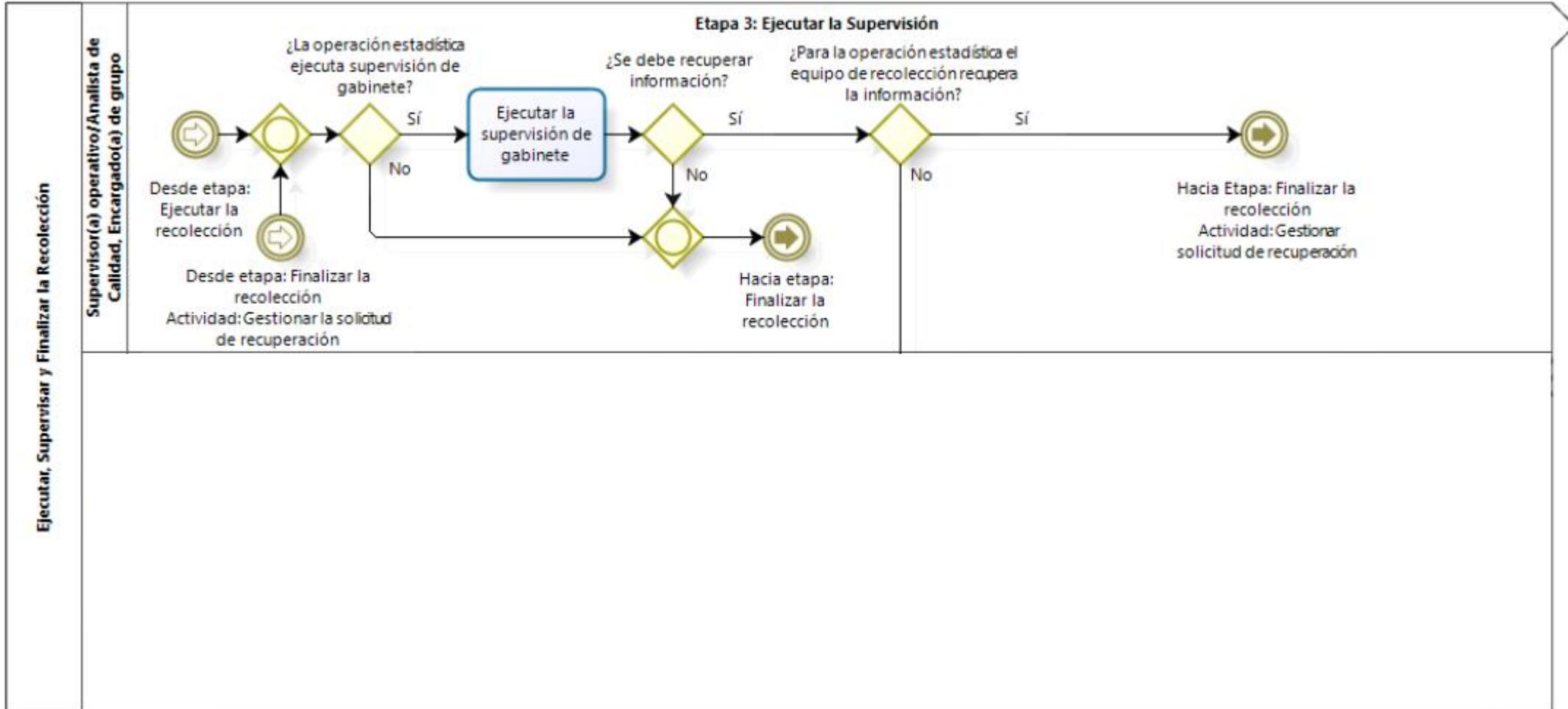


1. Contexto
2. ENUT: Usos de los Paradata: Recolección y procesamiento
3. Ciencia de datos: solución tecnológica
4. Ciencia de datos: casos de uso
5. Reflexiones finales

1.

Contexto

Ejecución de la Supervisión - Línea de negocio



- Las actividades de Supervisión recaen en los equipos de Trabajo de Campo.
- ¿Sería posible utilizar alguna herramienta adicional para optimizar la supervisión a los equipos de recolección?

¿Qué son los paradata?

- Información **recolectada durante el proceso de aplicación** de un instrumento (Callegaro, 2013). Esto, más allá de las respuestas de los informantes.
- Entre algunos de los elementos a identificar está el **orden en que se responde** una pregunta, la **fecha y hora** de registro y también, el momento en que se **inicia o sincronizan las respuestas** para enviar la información hacia los servidores donde se alojan las respuestas (Olson & Parkhurst, 2013).

INTERVIEW



Data export



OUTPUT



Instrumentos de recolección - II ENUT

Cuestionario del Hogar

Caracteriza al hogar entrevistado y a las distintas personas que forman parte de él. Pregunta por atributos como sexo, edad, relaciones de parentesco, nivel educacional, ocupación, ingresos, personas en situación de dependencia, etc.

Se aplica a un informante idóneo de 18 años o más

Formulario del Cuestionario del Hogar de la II Encuesta Nacional sobre Uso del Tiempo 2023. Incluye secciones para: INFORMACIÓN GENERAL, INFORMACIÓN DEL ENCUESTADOR, INFORMACIÓN DE LA VIVIENDA, PERÍODOS DE REFERENCIA ADJUDICADOS, OBSERVACIONES DE LA VIVIENDA, e IDENTIFICACIÓN DEL INFORMANTE IDÓNEO.

Cuestionario sobre Uso del Tiempo

Busca identificar las actividades que los distintos integrantes del hogar realizaron en un “período de referencia” determinado.

Se aplica a todas las personas de 12 años o más.

Formularios del Cuestionario sobre Uso del Tiempo de la II Encuesta Nacional sobre Uso del Tiempo 2023. Incluye secciones para: INFORMACIÓN GENERAL, INFORMACIÓN DEL ENCUESTADOR, INFORMACIÓN DE LA VIVIENDA, PERÍODOS DE REFERENCIA ADJUDICADOS, OBSERVACIONES DE LA VIVIENDA, e IDENTIFICACIÓN DEL INFORMANTE IDÓNEO.



Survey Solutions

Paradatos de *Survey Solutions*

interview_id	order	event	responsible	role	timestamp_utc	tz_offset	parameters
53fb5b4bb4b04626a010b08236bcbfd1	41	AnswerSet	999999999-0	1	2023-09-02 18:10:39	-04:00:00	h1 6
53fb5b4bb4b04626a010b08236bcbfd1	42	AnswerSet	999999999-0	1	2023-09-02 18:11:19	-04:00:00	h2 1
53fb5b4bb4b04626a010b08236bcbfd1	43	AnswerSet	999999999-0	1	2023-09-02 18:11:31	-04:00:00	hogar 1
53fb5b4bb4b04626a010b08236bcbfd1	44	AnswerSet	999999999-0	1	2023-09-02 18:12:04	-04:00:00	nombre Arnold
53fb5b4bb4b04626a010b08236bcbfd1	45	AnswerSet	999999999-0	1	2023-09-02 18:12:30	-04:00:00	nombre Arnold Blanco
53fb5b4bb4b04626a010b08236bcbfd1	46	AnswerSet	999999999-0	1	2023-09-02 18:12:43	-04:00:00	nombre Arnold Blanco Claudia
53fb5b4bb4b04626a010b08236bcbfd1	47	AnswerSet	999999999-0	1	2023-09-02 18:12:50	-04:00:00	nombre Arnold Blanco Claudia marcos joseua
53fb5b4bb4b04626a010b08236bcbfd1	48	AnswerSet	999999999-0	1	2023-09-02 18:12:58	-04:00:00	informante_idoneo 4
53fb5b4bb4b04626a010b08236bcbfd1	49	AnswerRemoved	999999999-0	1	2023-09-02 18:12:58	-04:00:00	informante_idoneo
53fb5b4bb4b04626a010b08236bcbfd1	50	AnswerSet	999999999-0	1	2023-09-02 18:13:11	-04:00:00	nombre Arnold Blanco Claudia David
53fb5b4bb4b04626a010b08236bcbfd1	51	AnswerSet	999999999-0	1	2023-09-02 18:13:26	-04:00:00	nombre Arnold Blanco Claudia David Evelyn
53fb5b4bb4b04626a010b08236bcbfd1	52	AnswerSet	999999999-0	1	2023-09-02 18:13:38	-04:00:00	nombre Arnold Blanco Claudia David Evelyn Francisco
53fb5b4bb4b04626a010b08236bcbfd1	53	AnswerSet	999999999-0	1	2023-09-02 18:13:41	-04:00:00	informante_idoneo 1
53fb5b4bb4b04626a010b08236bcbfd1	54	AnswerSet	999999999-0	1	2023-09-02 18:13:58	-04:00:00	c2 3 2

- Dentro de *parameters* podemos encontrar la información de respuesta y también el número del roster que estamos considerando.

2.

Usos de los Paradata: Recolección y procesamiento II ENUT

Cuestionarios sobre Uso del Tiempo (CUT) en días distintos

El **CUT** se debe aplicar en su completitud en un día, por los días de referencia entrevistados (día de semana y fin de semana). Esto para resguardar consistencia en el cuestionario.

Realizamos **revisiones por medio de Paradata** durante los procesos de pago por encuesta:

1. Filtramos solo las preguntas asociadas al CUT.
2. Procesamos la hora con el fin de tener el uso horario de Chile.
3. Realizamos condicionales para comprobar que el CUT se entrevistó en un día.
4. Luego, realizamos una revisión manual de los paradata, caso a caso, evaluando la magnitud de respuestas en cada día en un CUT.



Duración de entrevistas

Para calcular los tiempos contamos con distintas reglas:

1. Restar la diferencia de tiempo de una interacción de una fila a otra y alojarlo en la fila superior.
2. Imputar a cero los tiempos de interacciones que van de un día a otros en un cuestionario.
3. Dado que fue posible cambiar la fecha y hora de la DMC, también imputamos a 0 las diferencias de tiempo que son negativas.

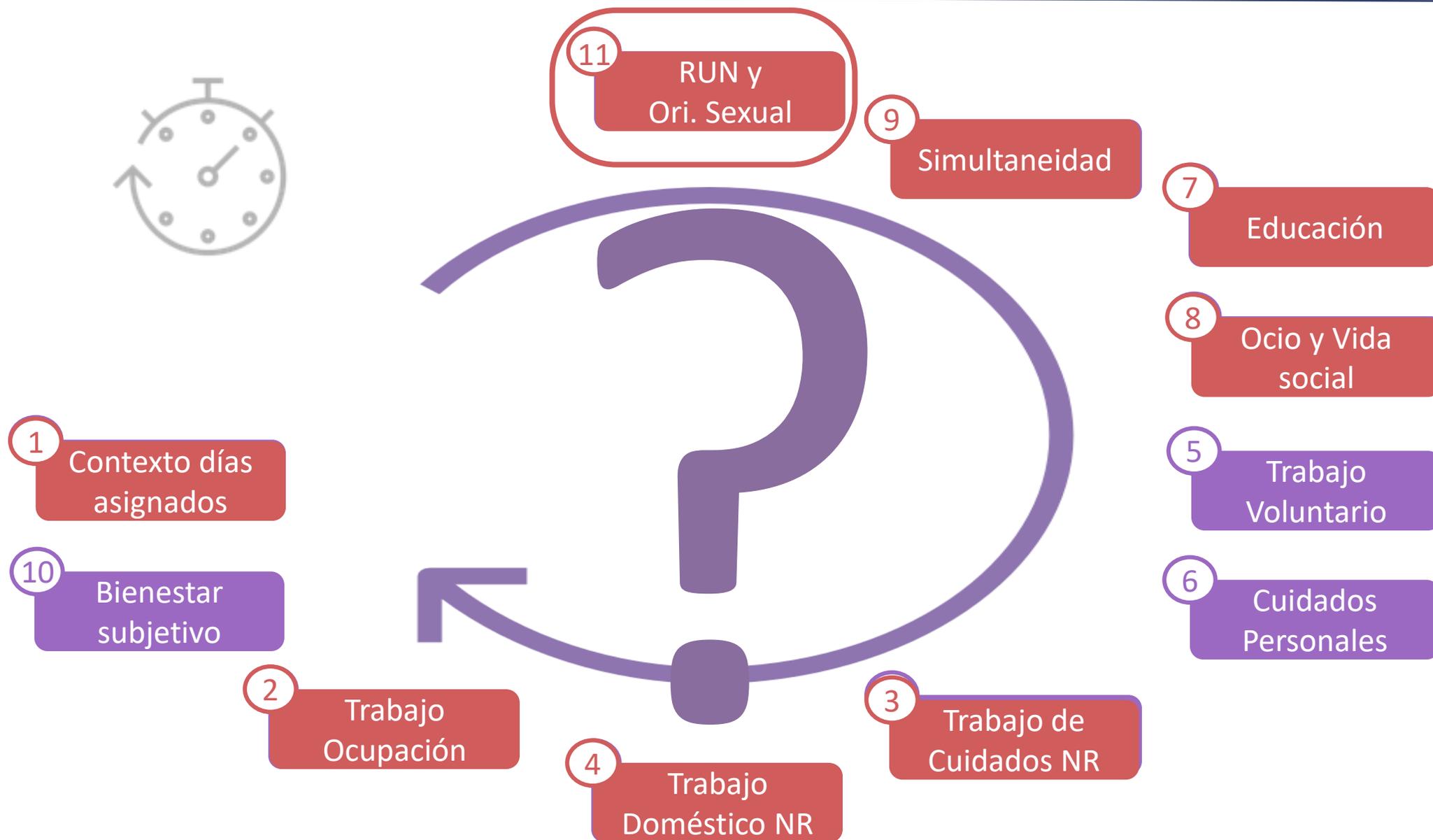
Instrumento	Duración de entrevista		Cuestionarios uso incorrecto de botones
	Paradatos	Manual/Botones	
Cuestionario sobre Uso del Tiempo	19 min.	23 min.	36%

Fuente: elaboración propia.

Nota: Datos preliminares

Por último, se sumaron todas las interacciones restantes, por el identificador único de cada encuesta, obteniendo los tiempos de duración.

Flujos erráticos CUT



Reingresos de información

hora_real	parameters
2023-12-12 18:22:02	tv2_p_ds 2 1
2023-12-12 18:22:12	tv2_p_fds 1 1
2023-12-12 18:22:16	tv2_t_fds 04:00 1
2023-12-12 18:22:38	tv2_t_fds 24:00 1
2023-12-12 18:22:57	tv2_t_fds Su nieta se viene a quedar todos los sabados 1
2023-12-15 17:46:49	tv2_t_fds 12:00 1
2023-12-15 17:47:23	tv2_t_fds El aclara 1ue que la cuida todo el día y la niña se queda a dormir en su casa 1



- 1) Registra 4 hrs. De trabajo voluntario el fin de semana.
- 2) Segundos después modifica las horas registradas a 24 hrs.
- 3) 3 días después modifica el dato a 12 hrs. de trabajo voluntario.

- ¿Por qué modifica el dato 3 días después?
- No conocemos la historia detrás de estas modificaciones.

3.

Ciencia de datos: solución tecnológica

Motivación

interview_id	order	event	responsible	role	timestamp_utc	tz_offset	parameters
53fb5b4bb4b04626a010b08236bcbfd1	41	AnswerSet	999999999-0	1	2023-09-02 18:10:39	-04:00:00	h1 6
53fb5b4bb4b04626a010b08236bcbfd1	42	AnswerSet	999999999-0	1	2023-09-02 18:11:19	-04:00:00	h2 1
53fb5b4bb4b04626a010b08236bcbfd1	43	AnswerSet	999999999-0	1	2023-09-02 18:11:31	-04:00:00	hogar 1
53fb5b4bb4b04626a010b08236bcbfd1	44	AnswerSet	999999999-0	1	2023-09-02 18:12:04	-04:00:00	nombre Arnold
53fb5b4bb4b04626a010b08236bcbfd1	45	AnswerSet	999999999-0	1	2023-09-02 18:12:30	-04:00:00	nombre Arnold Blanco
53fb5b4bb4b04626a010b08236bcbfd1	46	AnswerSet	999999999-0	1	2023-09-02 18:12:43	-04:00:00	nombre Arnold Blanco Claudia
53fb5b4bb4b04626a010b08236bcbfd1	47	AnswerSet	999999999-0	1	2023-09-02 18:12:50	-04:00:00	nombre Arnold Blanco Claudia marcos joseua
53fb5b4bb4b04626a010b08236bcbfd1	48	AnswerSet	999999999-0	1	2023-09-02 18:12:58	-04:00:00	informante_idoneo 4
53fb5b4bb4b04626a010b08236bcbfd1	49	AnswerRemoved	999999999-0	1	2023-09-02 18:12:58	-04:00:00	informante_idoneo
53fb5b4bb4b04626a010b08236bcbfd1	50	AnswerSet	999999999-0	1	2023-09-02 18:13:11	-04:00:00	nombre Arnold Blanco Claudia David
53fb5b4bb4b04626a010b08236bcbfd1	51	AnswerSet	999999999-0	1	2023-09-02 18:13:26	-04:00:00	nombre Arnold Blanco Claudia David Evelyn
53fb5b4bb4b04626a010b08236bcbfd1	52	AnswerSet	999999999-0	1	2023-09-02 18:13:38	-04:00:00	nombre Arnold Blanco Claudia David Evelyn Francisco
53fb5b4bb4b04626a010b08236bcbfd1	53	AnswerSet	999999999-0	1	2023-09-02 18:13:41	-04:00:00	informante_idoneo 1
53fb5b4bb4b04626a010b08236bcbfd1	54	AnswerSet	999999999-0	1	2023-09-02 18:13:58	-04:00:00	c2 3 2

Retos para la construcción del servicio

- Manejo de grandes volúmenes de datos
 - Millones de datos
- Actualización **constante y rápida** de la información
- Balance entre **estandarización y flexibilidad**
- Actualmente en producción
 - Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC)
 - Encuesta Nacional de Participación Cultural y Comportamiento Lector (ENPCCL)
 - Piloto Encuesta de Microemprendimiento (EME)

Estrategia utilizada

- Se ejecuta 3 veces por semana

Survey Solutions



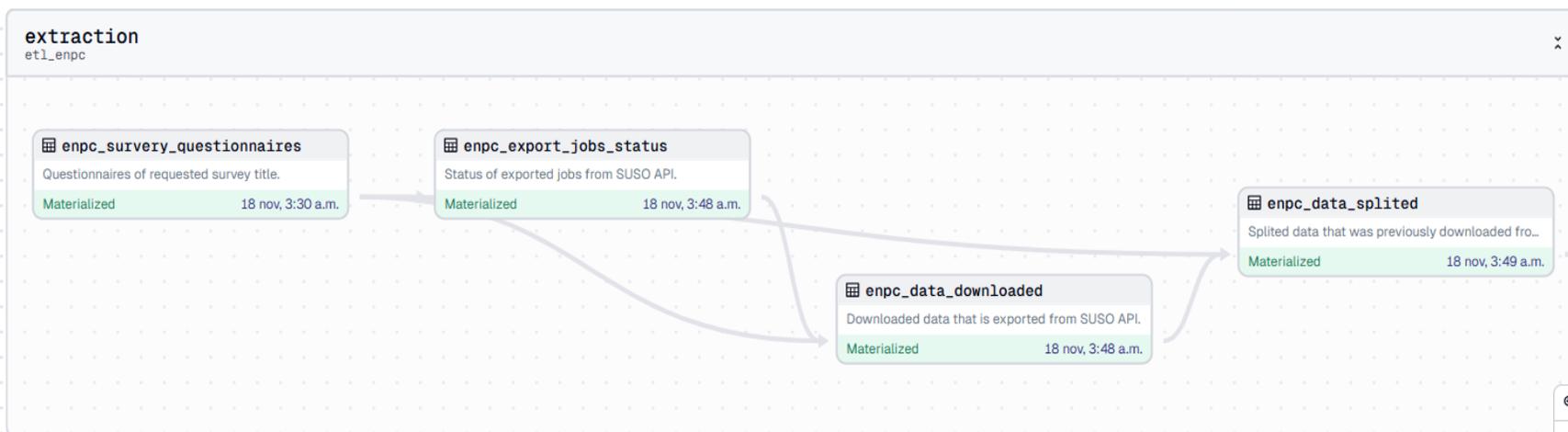
MINIO



trino



Streamlit



Producto entregado

Login

General

Revisión encuestador

Revisión encargado de grupo

Revisión modelo

Revisión RPH

Cerrar Sesión

Dashboard Paradata ENUSC 2024

Análisis del comportamiento de registro de respuestas en las entrevistas

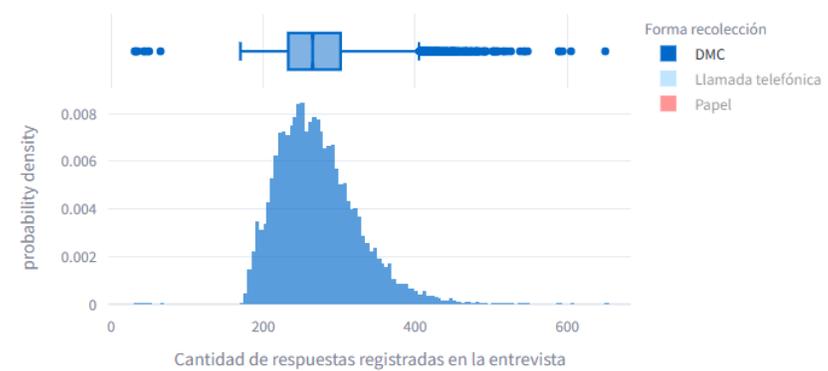
Seleccione columnas para filtrar:

reg_cdf

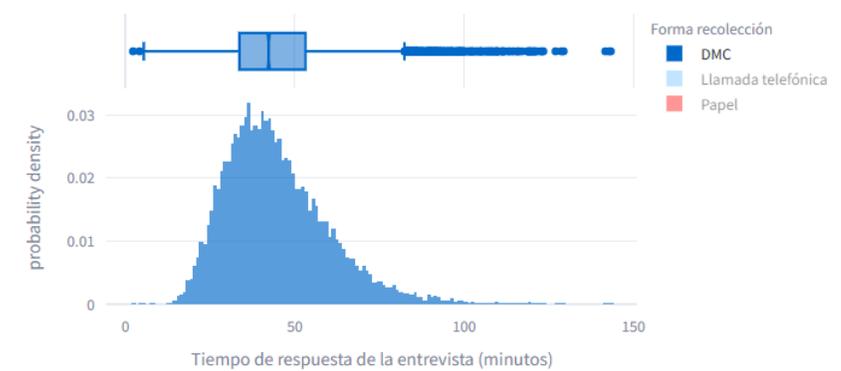
Seleccione uno o más valores para reg_cdf

11

Cantidad de respuestas por entrevista



Tiempo de respuesta de entrevistas



¿Qué estamos reportando?

Ilustración 56. Selección Kish en el DMC

— RPH - IDENTIFICACIÓN DE HOGARES Y REGISTRO DE PERSONAS EN EL HOGAR

IDENTIFICACIÓN DE HOGARES

+ REGISTROS DE PERSONAS

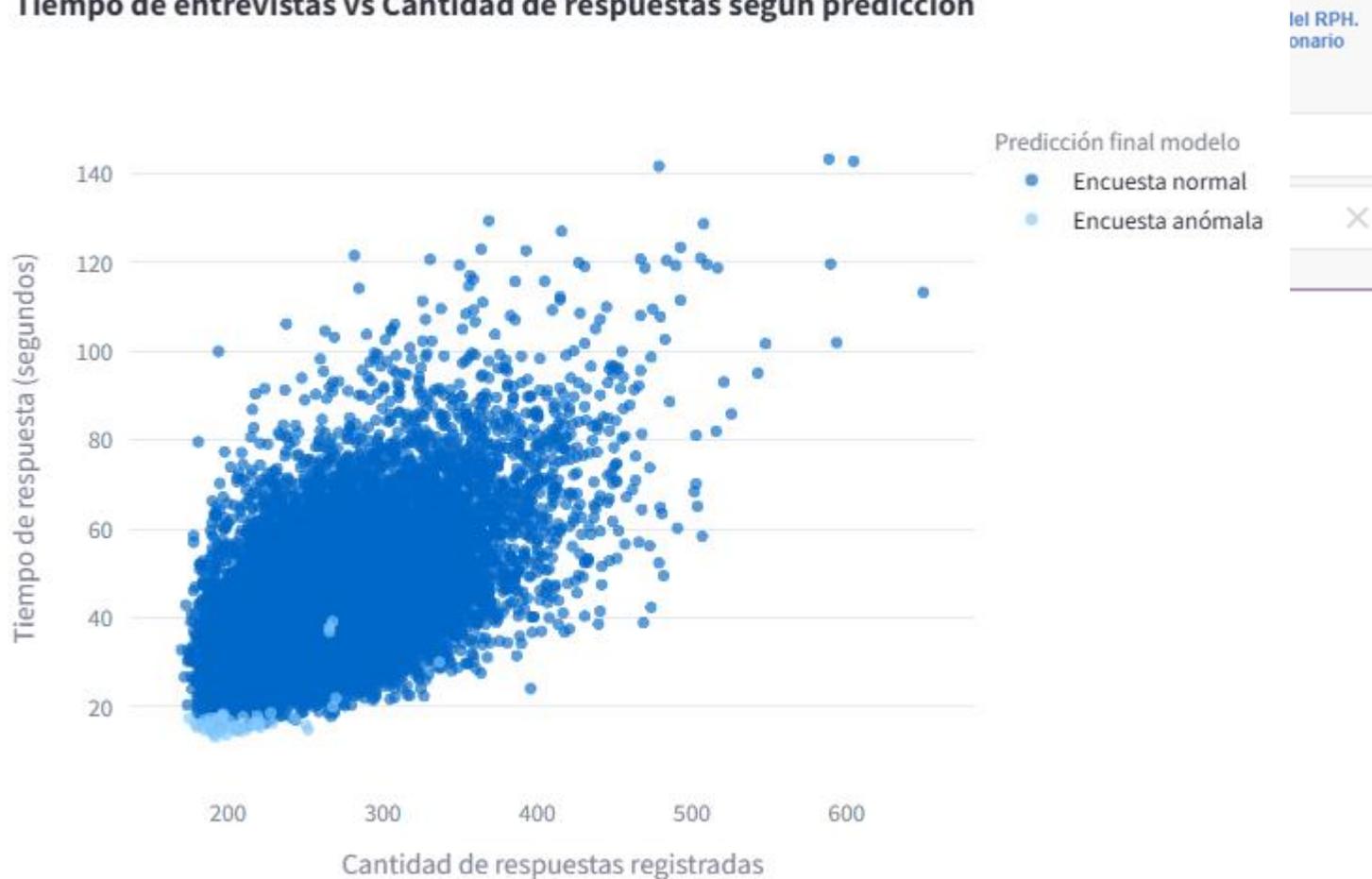
SELECCIÓN KISH

+ REGISTRO DE PERSONAS INFORMANTE KISH

La persona informante kish corresponde a papa de 31 años.

[rph_kish_disponible]
¿La persona informante Kish está disponible para continuar con la entrevista?

Tiempo de entrevistas vs Cantidad de respuestas según predicción



4.

Ciencia de datos: casos de uso

¡Veamos un ejemplo!

Login

General

Revision encuestador

Revision encargado de grupo

Revision modelo

Revision RPH

Cerrar Sesión

Dashboard Paradatos ENUSC 2024

Análisis del comportamiento de registro de respuestas en las entrevistas [?]

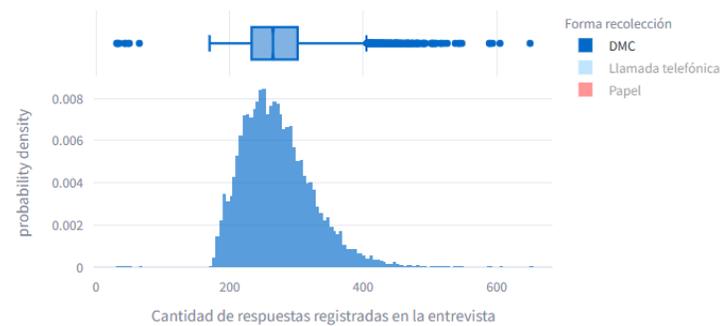
Seleccione columnas para filtrar:

reg_cdf ×

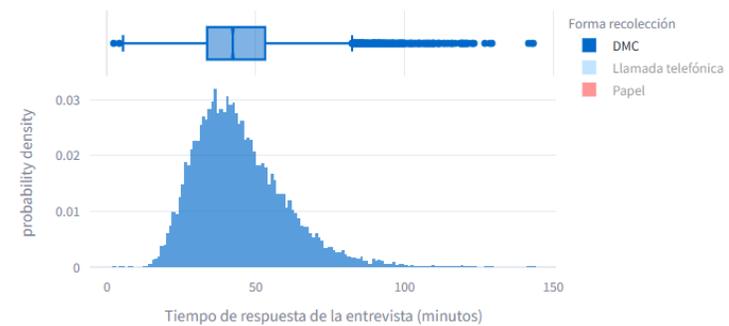
Seleccione uno o más valores para reg_cdf

11 ×

Cantidad de respuestas por entrevista



Tiempo de respuesta de entrevistas



Cantidad media de respuestas de encuestadores para las preguntas: *integrantes en el hogar, asistencias a actividades e infraestructuras*



Seleccione una región para ver el promedio de respuestas de los encuestadores asociados.

Región:

Biobío ✕ ▼

CDF encuestador

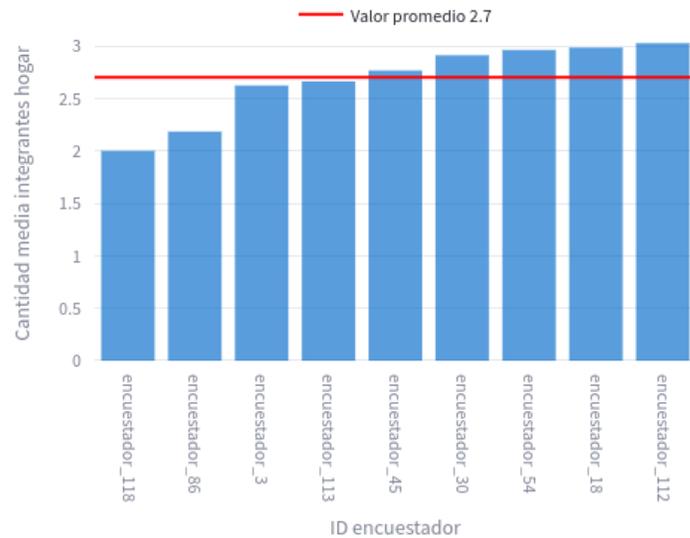
11 ✕ ▼

CDF encargado de grupo:

11 ✕ ▼

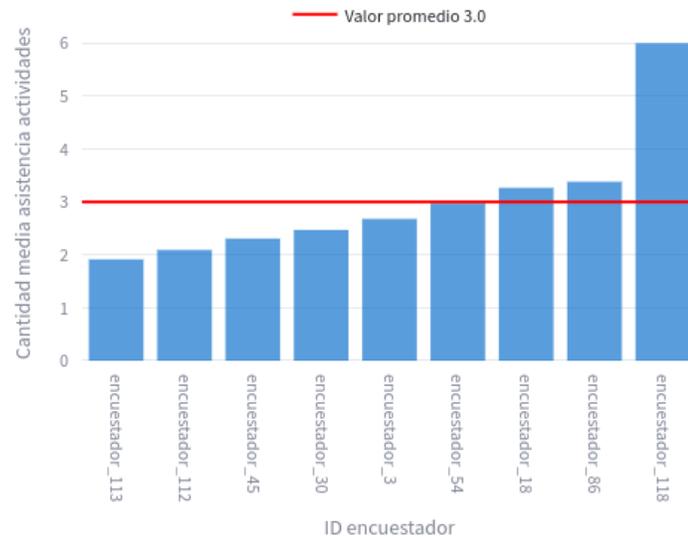
Cantidad media de integrantes en el hogar

Biobío



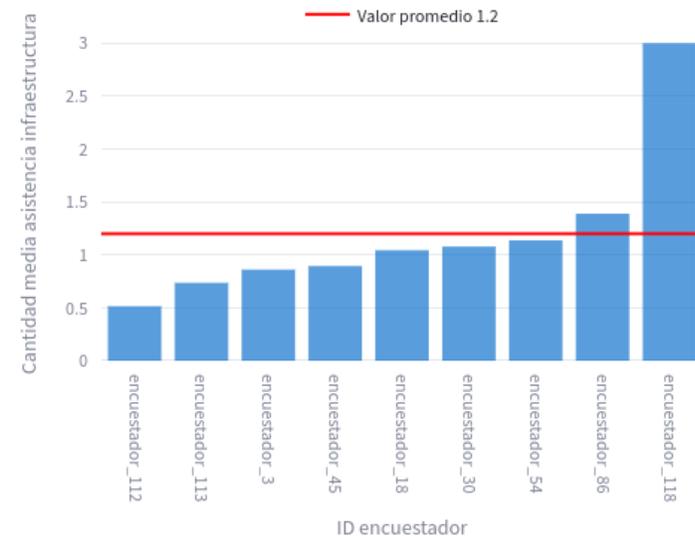
Cantidad media de asistencia a actividades

Biobío

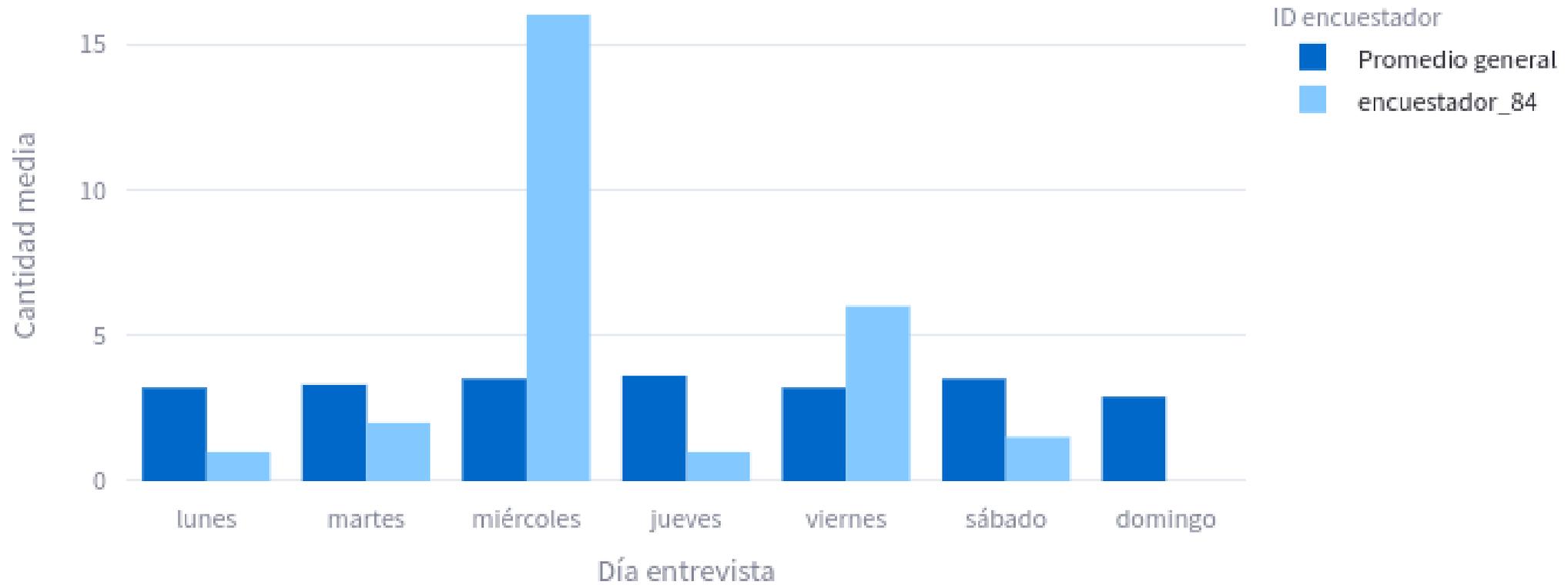


Cantidad media de asistencia a infraestructuras

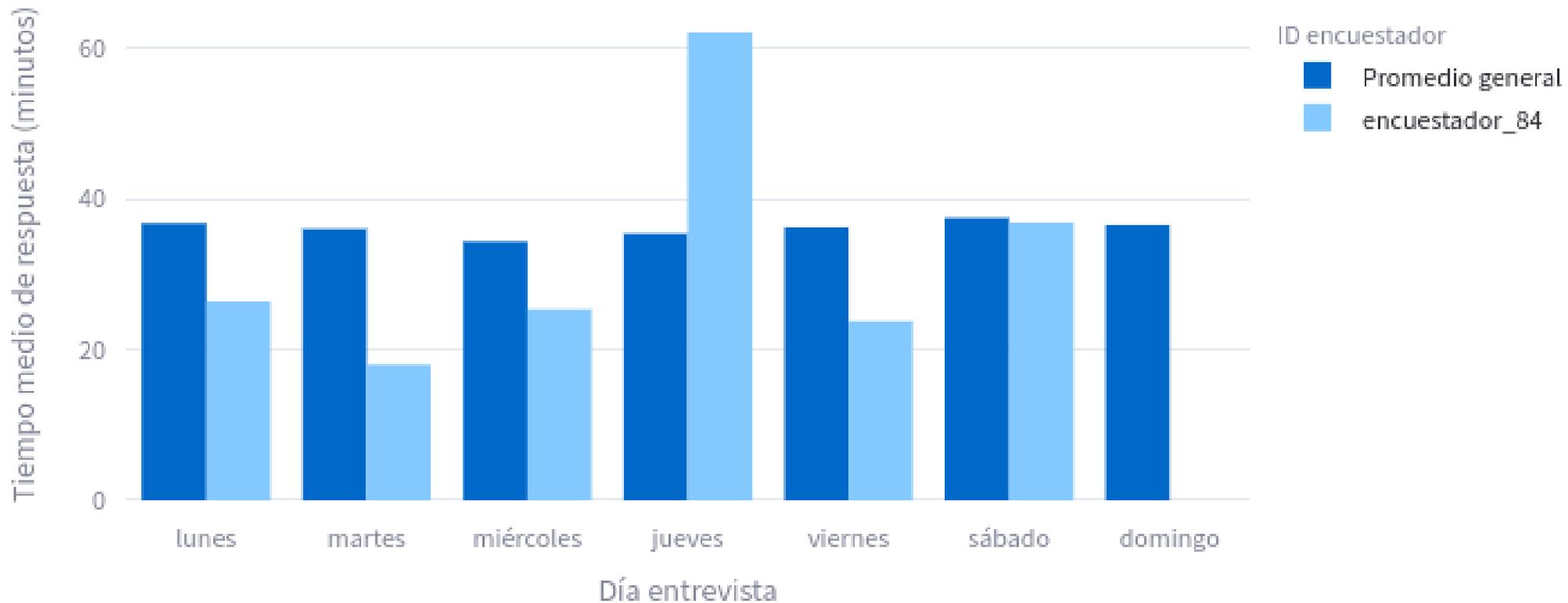
Biobío



Cantidad media de entrevistas realizadas diariamente



Tiempo medio de entrevistas realizadas diariamente



Revisión RPH 2024

Cantidad de modificaciones:



Tabla con cantidad de modificaciones a las preguntas de *rph_edad* y *rphsexo* ?

	interview_id	folio	comuna	pregunta_rph	indice_integrante_con_modificaciones	n_modificaciones_respuesta_rph	ih1_ultimo_registro	ih2_ultimo_registro	n_visita_ultimo_registro	
0	7120ef02bf324a18aad9d19a9accf81e	11788	12101	rph_edad	5	23	7	2	6	
1	7120ef02bf324a18aad9d19a9accf81e	11788	12101	rph_edad	3	17	7	2	6	
<input checked="" type="checkbox"/>	2	7120ef02bf324a18aad9d19a9accf81e	11788	12101	rph_edad	6	17	7	2	6
3	7120ef02bf324a18aad9d19a9accf81e	11788	12101	rph_edad	2	16	7	2	6	

Mostrando 1 folios

Desglose de información ingresada por el encuestador para las encuestas seleccionadas

Mostrando información para entrevistas seleccionadas con folio(s): ['11788'] (Total: 1)

	interview_id	event	responsable	role	timestamp_tz	pregunta_rph	respuestas	indice_integrante_con_modificaciones	folio	comuna	n_visita_ultimo_registro
0	7120ef02bf324a18aad9d19a9accf81e	AnswerSet	encuestador_83	1	2024-09-09 15:03:21	rph_nombres	1	None	11788	12101	
1	7120ef02bf324a18aad9d19a9accf81e	AnswerSet	encuestador_83	1	2024-09-09 15:04:18	rph_nombres	Claudio	None	11788	12101	
2	7120ef02bf324a18aad9d19a9accf81e	AnswerSet	encuestador_83	1	2024-09-09 15:04:32	rph_nombres	Claudio ELI	None	11788	12101	
3	7120ef02bf324a18aad9d19a9accf81e	AnswerSet	encuestador_83	1	2024-09-09 15:04:38	rph_nombres	Claudio ELI rodrigo	None	11788	12101	
4	7120ef02bf324a18aad9d19a9accf81e	AnswerSet	encuestador_83	1	2024-09-09 15:04:47	rph_nombres	Claudio ELI rodrigo rudi	None	11788	12101	
5	7120ef02bf324a18aad9d19a9accf81e	AnswerSet	encuestador_83	1	2024-09-09 15:05:03	rph_nombres	Claudio ELI rodrigo rudi g	None	11788	12101	
6	7120ef02bf324a18aad9d19a9accf81e	AnswerSet	encuestador_83	1	2024-09-09 15:05:21	rph_nombres	Claudio ELI rodrigo rudi g	None	11788	12101	

Desafíos de corto plazo para ciencia de datos

- Escalamiento ordenado del servicio
- Delegación de la continuidad operativa

ENUSC 

ENPCCL 

EME 

- Interoperabilidad con el resto de los sistemas del INE



5.

Consideraciones finales

Reflexiones finales

- El mayor potencial de los parados es detectar problemas en "**tiempo real**" y no a posteriori
- Todos los proyectos tienen **particularidades**, pero hay espacio para la **estandarización**
 - Viabilidad del proyecto
- **Ajustar los diseños de cuestionario y proceso de trabajo de campo**
 - En base a los hallazgos en base a parados durante la recolección y procesamiento de datos.
- **Estandarización del uso de Parados en Trabajo de Campo**
 - Modelamiento de comportamientos de personas encuestadoras
 - Automatización de la identificación de comportamiento anómalos
 - Optimización de recursos de supervisión al focalizar supervisiones



Aprovechamiento de parados en encuestas de hogares para el aseguramiento de la calidad

Subdepartamento de Estadísticas Estructurales del Trabajo
Subdirección Técnica

Equipo de Ciencia de Datos, Unidad de Gobierno de Datos

INSTITUTO NACIONAL DE ESTADÍSTICAS

Noviembre 2024

ine.gob.cl

