	PROCEDIMIENTO DE ANALÍTICA DE DATOS			
	Proceso: Gestión del Conocimiento	Versión: 4	Código: P5.GC	Fecha Aprobación: 28/03/2022

1. OBJETIVO:

Establecer la ruta de trabajo para poner al servicio de los equipos de la Comisión de la Verdad los datos, análisis y herramientas técnicas orientadas al procesamiento de información estructurada y no estructurada, proveniente de fuentes primarias y secundarias. El cumplimiento de este procedimiento garantiza, por un lado, brindar los insumos cuantitativos a los equipos de investigación que den soporte suficiente para el análisis, contrastación y verificación de hipótesis; y, por otro lado, validar la imparcialidad, transparencia, auditabilidad y rigurosidad en la aplicación de las metodologías de análisis.

2. ALCANCE:


El procedimiento describe el flujo de trabajo para la producción de analítica de datos en la Comisión de la Verdad, desde la identificación de las necesidades, hasta que se producen, socializan y resguardan los reportes y ejercicios elaborados. Esto aplica para todos los niveles de la entidad responsables de los procesos misionales.

3. GENERALIDADES:

La analítica de datos se sirve de las fuentes propias y externas disponibles en las múltiples herramientas del Sistema de Información Misional. Son, a su vez, fuentes estructuradas, semiestructuradas y no estructuradas, que abarcan no sólo las dimensiones directamente relacionadas con el conflicto, como violencia, seguridad, defensa, convivencia, justicia, actividad política y asociativa, cultivos ilícitos, entre otros; sino, también, otras dimensiones que permiten descubrir los determinantes e impactos del conflicto, sus patrones y contextos explicativos. En este último segmento se cuentan, por ejemplo, datos económicos, culturales, sociodemográficos y ambientales. El análisis de información comprende los ámbitos cuantitativos y cualitativos, con la rigurosidad, métodos y validación/contrastación de información siguiendo los lineamientos del Decreto Ley 588 de 2017 y los Lineamientos Metodológicos de la Comisión de la Verdad.

El tratamiento de estas fuentes, así como la divulgación de los resultados de los análisis, exige el uso combinado de técnicas estadísticas y matemáticas, por un lado, y soluciones tecnológicas, por otro lado. Se debe recurrir a métodos de investigación y, a su vez, implementar tecnologías de información para la exploración de los datos y el descubrimiento de patrones multivariados, temporales, espaciales y textuales en los mismos. Se aplican herramientas que permitan aplicar técnicas sobre datos estructurados y no estructurados de forma ágil, incremental, fiable, reproducible y amigable para los analistas de la entidad, cuantitativa y cualitativamente.

La principal responsabilidad en el análisis de la información cuantitativa y cualitativa está en cabeza de quienes lideran los equipos de investigación, en los cuales hay personas analistas de datos cualitativos y cuantitativos; y, la responsabilidad del SIM es brindar el apoyo y soporte técnico en materia de verificación de la integridad, validez y confiabilidad de la información, así como la disposición de herramientas tecnológicas para el acceso, disposición y uso de los datos e información recolectada. En el SIM se encuentran los equipos encargados de desarrollo de software, soporte técnico para la operación de las distintas herramientas, visualizaciones, mapas u otros ejercicios estadísticos, consecución y curaduría de información, catalogación y procesamiento de información,

	PROCEDIMIENTO DE ANALÍTICA DE DATOS			
	Proceso: Gestión del Conocimiento	Versión: 4	Código: P5.GC	Fecha Aprobación: 28/03/2022


junto a las herramientas para el acopio de información y disposición para el uso de los equipos y legado de la Comisión.

El relacionamiento con el equipo del SIM se hará a través de los líderes de los equipos.


Los seminarios realizados en el marco de la aplicación de este procedimiento, están conformados por equipo directivo, de coordinación y profesionales. La Directora de Conocimiento y el Coordinador del SIM participarán en el seminario de revisión de pares de resultados con el fin de asegurar que el profesional cumpla con los estándares establecidos en la metodología para la investigación de analítica de datos. Los profesionales de la Dirección de Conocimiento, Dirección de Territorios, Dirección de Diálogo Social o Dirección de Pueblos Étnicos y los profesionales del SIM participan de los seminarios para garantizar la rigurosidad e imparcialidad de la investigación, así como los principios establecidos en la metodología para la investigación de analítica de datos. Asimismo, los profesionales de las direcciones y del sistema mencionados anteriormente, que cuenten con habilidades cuantitativas, son los encargados de ejecutar los proyectos dentro del Git.

4. DESCRIPCIÓN DE ACTIVIDADES:

No	Descripción de la actividad	Responsable	Registro (evidencia de la actividad, si aplica)
1	Plantear una pregunta de investigación que pueda ser respondida con métodos cuantitativos, a partir de la identificación de una necesidad de análisis de datos prevista por cualquier área misional de la entidad (equipos de investigación, Grupo de Informe Final, Pleno de Comisionadas y Comisionados, Coordinaciones o Direcciones) ¿El equipo cuenta con un Profesional para realizar análisis cuantitativo? Si: Continuar a la actividad 3 No: Continuar a la actividad 2	Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social	Acta de reunión o Comité
2	Dirigir solicitud de asistencia técnica al Sistema de Información Misional-SIM para el apoyo en la formulación de un proyecto analítico.	Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social	Correo electrónico
3	Definir y notificar al equipo solicitante, cual es el profesional designado para asistencia técnica en la formulación de un proyecto analítico.	Coordinador del Sistema De Información Misional- SIM	Correo electrónico
4	Delimitar la pregunta de investigación con el analista cuantitativo en materia de población, periodo y objeto de estudio. Ver Metodología para la investigación de analítica de datos.	Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social Profesionales de la Coordinación del SIM	Correo electrónico Acta de reunión o comité

	PROCEDIMIENTO DE ANALÍTICA DE DATOS			
	Proceso: Gestión del Conocimiento	Versión: 4	Código: P5.GC	Fecha Aprobación: 28/03/2022

5	<p>Construir una propuesta articulada de trabajo que defina:</p> <ol style="list-style-type: none"> 1. Descripción de la necesidad 2. Objeto 3. Alcance 4. Cronograma <p>Esta propuesta es definida por el equipo que realizó la solicitud y el Profesional analista cuantitativo.</p>	<p>Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social</p> <p>Profesionales de la Coordinación del SIM</p>	Formato Propuesta de Analítica
6	<p>Registrar y desarrollar las actividades de la propuesta de Analítica en el repositorio Git del SIM.</p> <p>El registro y desarrollo de las actividades debe seguir la metodología de analítica de datos.</p> <p>Ver Metodología para la investigación de analítica de datos.</p>	<p>Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social</p> <p>Profesionales de la Coordinación del SIM</p>	Repositorio Git
7	<p>Participar en seminarios periódicos de revisión de pares para el seguimiento técnico de los avances de los proyectos.</p>	<p>Coordinador SIM</p> <p>Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social</p> <p>Profesionales de la Coordinación del SIM</p>	Acta de reunión o Comité
8	<p>Entregar reportes de avance de los ejercicios de Datos, Estadísticas y Analítica, Visualizaciones de datos: gráficas, mapas, etc.</p>	<p>Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social</p> <p>Profesionales de la Coordinación del SIM</p>	Acta de reunión o Comité
9	<p>Cargar permanentemente la documentación del proyecto en el Git.</p> <p>Según lo establecido en el Manual de repositorio Gitlab de Analítica</p>	<p>Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social</p> <p>Profesionales de la Coordinación del SIM</p>	Repositorio Git
10	<p>Elaborar, en el repositorio Git, el Documento del informe de resultado de Analítica de datos.</p> <p>Siguiendo las directrices establecidas en el Manual de repositorio Gitlab de Analítica</p>	<p>Analistas cuantitativos de los equipos de trabajo misionales</p> <p>Profesionales de la Coordinación del SIM</p>	Informe de resultado de Analítica de datos
11 P.C	<p>Participar del seminario de revisión de pares del proyecto finalizado para la evaluación de la rigurosidad, neutralidad y transparencia de los métodos empleados y resultados obtenidos.</p>	<p>Director(a) de Conocimiento</p> <p>Profesionales de la Coordinación del SIM</p>	Acta de Reunión o Comité

	PROCEDIMIENTO DE ANALÍTICA DE DATOS			
	Proceso: Gestión del Conocimiento	Versión: 4	Código: P5.GC	Fecha Aprobación: 28/03/2022

	¿La investigación es rigurosa, neutral y equilibrada en sus resultados? Sí: Continuar a la actividad 12 No: Reajustar la propuesta de analítica y regresar a la actividad 5. Nota: Cuando el proyecto presentado sea un capítulo del informe final se debe realizar Auditoría de Datos. Ver instructivo Auditoría de Datos del Informe Final	Profesional Dirección de Conocimiento Coordinador del SIM	
12	Cerrar el proyecto en el Git conforme a los lineamientos establecidos en la metodología de análisis de datos. Ver Metodología para la investigación de analítica de datos. Ver Manual de repositorio Gitlab de Analítica	Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social Profesionales de la Coordinación del SIM	Repositorio Git
13	Difundir el Reporte de analítica con los equipos interesados.	Profesionales de la Dirección de Conocimiento Dirección de Territorios Dirección de Pueblos Étnicos Dirección de Diálogo Social Profesionales de la Coordinación del SIM	Correo electrónico
Fin			


5. RESULTADO FINAL:

Reportes de Analítica, ejercicios estadísticos, script, mapas, visualizaciones, documentación técnica u otros, validados, socializados, publicados y cargados en la plataforma tecnológica SIM.

6. DEFINICIONES:

Analítica de datos: La analítica de datos en la Comisión de la Verdad es un procedimiento orientado al óptimo aprovechamiento de los grandes volúmenes de información misional para el entendimiento de los patrones, las dinámicas, y los contextos explicativos, abordando desde esta perspectiva las principales preguntas e hipótesis de investigación (ver Decreto Ley 588 de 2017). Esto exige un adecuado tratamiento de la información propia de la Comisión, sus fuentes primarias, así como una estratégica gestión de fuentes externas, pues se trata de entender el conflicto armado en relación con otros fenómenos. Son, a su vez, fuentes estructuradas, semiestructuradas y no estructuradas, que abarcan no sólo las dimensiones directamente relacionadas con el conflicto, como violencia, seguridad, defensa, convivencia, justicia, actividad política y asociativa, cultivos ilícitos, entre otros; sino, también, otras dimensiones que permiten descubrir los determinantes e impactos del conflicto, sus patrones y contextos explicativos. En este último segmento se cuentan, por ejemplo, datos económicos, culturales, sociodemográficos y ambientales. Analítica de datos es un proceso que lideran y desarrollan los equipos de investigación, con apoyo del SIM.

Fuentes Propias: Las fuentes propias corresponden a las entrevistas a víctimas, testigos o familiares; las entrevistas a profundidad; los diagnósticos comunitarios; las historias de vida; las entrevistas a terceros civiles; las entrevistas colectivas; las entrevistas a sujetos colectivos; y las entrevistas a actores armados.

	PROCEDIMIENTO DE ANALÍTICA DE DATOS			
	Proceso: Gestión del Conocimiento	Versión: 4	Código: P5.GC	Fecha Aprobación: 28/03/2022

Fuentes Externas: Las fuentes de información externas a la Entidad, se trata de fuentes oficiales administradas por organizaciones del Gobierno Nacional y entes territoriales, así como de fuentes no oficiales en propiedad de operadores de cooperación internacional, instituciones educativas, centros de pensamiento y organizaciones de la sociedad civil, entre otras.

Herramientas de análisis estadístico: La Comisión cuenta con licencias del software para análisis estadístico Stata, para procesamiento y análisis geográfico ArcGIS, y para visualización de datos Tableau. Más todavía, se ha configurado un ambiente de soluciones y lenguajes para estos propósitos, que complementan las dos anteriores, como RStudio (R), Jupyter Notebook (Python), ObservableHQ (Javascript, D3), QGIS, GeoDa, Gephi, Timeline Storyteller. Así como aplicaciones para la edición de textos de contenido científico y técnico como Texmaker (LaTeX) y JabRef. (Baldwin, 2016; Bilbro, Ojeda y Bengfort, 2019; Colin y Trivedi, 2009; Pimpler, 2017; Silge y Robinson, 2017; Yang, 2019).

Datos estructurados: son las piezas de información cuyo formato es conocido y sigue un modelo. Se trata, en la mayoría de los casos, de conjuntos de datos como tablas, celdas, geometrías, redes o series de tiempo. En los cuales, se identifica claramente la definición de los ítems (como unidades básicas de observación, o “individuos” de una muestra poblacional), las celdas, las posiciones o las unidades de tiempo asociadas en los valores contenidos en los conjuntos de datos; así como el posible contenido de los atributos (o dominio de los campos), que pueden ser categóricos u ordenados. Con frecuencia, su representación lógica es la de archivos planos o bases de datos relacionales.


Datos no estructurados: son piezas de información para las cuales el analista desconoce el formato de su contenido. Se refiere comúnmente a conjuntos de archivos de texto, imagen o audio para los que se desconoce el alcance y modelo preciso de su contenido, más allá de metadatos como su origen, fechas de modificación, extensión, espacio en almacenamiento, entre otros.

Código de programación de analítica de datos: es el documento que contiene las instrucciones, escritas en lenguaje codificado de programación, para ejecutar las diversas funciones que permiten el preprocesamiento, la visualización y/o la aplicación de técnicas exploratorias y confirmatorias de análisis cuantitativo, espacial o textual sobre los datos.

Repositorio Git: es un repositorio en el que se almacena la documentación relacionada con el proyecto de investigación cuantitativo. Tiene control de versionamiento, por lo que se tiene total trazabilidad de lo que sucede en el mismo y es posible volver a puntos anteriores del trabajo, eliminando el riesgo de pérdida de información.

7. DOCUMENTOS DE REFERENCIA:

- Manual de repositorio Gitlab de Analítica
- Metodología para la investigación de analítica de datos.
- Procedimiento de Recepción de Informes, Casos y Fuentes Documentales Externas para el Esclarecimiento.
- Lineamientos Metodológicos de la Comisión de la Verdad.
- Baldwin, D. (2016). *Mastering Tableau*. Packt Publishing Ltd.
- Bilbro, R., Ojeda, T., & Bengfort, B. (2019). *Applied text analysis with Python*.
- Colin, C. A., & Trivedi, P. K. (2009). *Microeconometrics using STATA*. Lakeway Drive, TX: Stata Press Books.

	PROCEDIMIENTO DE ANALÍTICA DE DATOS			
	Proceso: Gestión del Conocimiento	Versión: 4	Código: P5.GC	Fecha Aprobación: 28/03/2022

- Pimpler, E. (2017). *Spatial Analytics with ArcGIS*. Packt Publishing Ltd.
- Presidencia de la República (2017). Decreto Ley 588 de 2017, "Por el cual se organiza la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la no Repetición"
- Procedimiento de captura y procesamiento del Sistema de Información Misional
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly Media, Inc.
- Yang, X. S. (2019). *Introduction to Algorithms for Data Mining and Machine Learning*. Academic Press.

8. RELACIÓN DE FORMATOS:

CODIGO	NOMBRE DEL FORMATO
F1.P5.GC	Propuesta de analítica
F1.I1.P5.GC	Formato de Auditoría de Datos
F2.P1.DESE	Formato Acta de reunión o Comité

9. ANEXOS:

No Aplica

10. CONTROL DE CAMBIOS:

Fecha	Versión	Descripción del Cambio
16/07/2021	V3	En la actividad número 11 se incluye nota " Cuando el proyecto presentado sea un capítulo del informe final se debe realizar Auditoría de Datos. Ver instructivo Auditoría de Datos del Informe Final" con el fin de definir el flujo para esta actividad fundamental para el informe final. También se incluye el formato de Auditoría de Datos.
22/12/2020	V2	Se incluye el Manual de Repositorio de Gitlab, con el fin de establecer las características básicas que permitan realizar el cargue y mantener el orden de los proyectos de analítica en el repositorio establecido para este fin, incluyendo los pasos que permitan garantizar la replicabilidad de los diferentes estudios realizados por el equipo de analítica. La implementación de las directrices relacionadas en el Manual, son nuevas y complementan el flujo de actividades del procedimiento, por lo tanto afecta la versión del mismo.
06/03/2020	V1	Se modifica el nombre del formato F1.P5.GC Formato Propuesta de Analítica: Es importante porque no todos los proyectos que se realizan son para "predecir" existen otros ejercicios de estadística, como estadística descriptiva, que no son predictivos. En cuanto al procedimiento se ajusta la redacción del objetivo, y en general, todo el documento se ajusta a la luz de las nuevas disposiciones emitidas por el pleno en relación a la configuración de equipos y metodologías de trabajo de cara a la escritura del informe final.

VALIDACIÓN	NOMBRE	CARGO	FECHA
Elaboró	Camilo Andrés Avellaneda / Andrea del Pilar Gonzalez/ José Giovanni Batativa/ Daihana Lisbeth Flórez	Analistas Coordinación del Sistema de Información Misional	28/03/2022
Revisó	Diana Britto	Directora de Conocimiento	28/03/2022
Aprobó	Diana Britto	Directora de Conocimiento	28/03/2022