



Marco de Implementación de Proyectos de Análisis de Datos

DESDE LA DEFINICIÓN HASTA LA TOMA DE DECISION

Contenido

Tablas.....	4
Tabla de Figuras.....	5
Tabla de Anexos.....	6
Capítulo 1. Estado del Arte	7
1.1 Introducción.....	7
1.2 Metodologías utilizadas en la minería de datos.....	8
1.2.1 Metodología SEMMA.....	8
1.2.2 Metodología Catalyst	8
1.2.3 Metodología CRISP-DM	9
1.2.4 Proceso KDD	14
1.3 Antecedentes para la gestión de datos en gobierno.....	16
1.4 Datos abiertos en Panamá.....	17
Capítulo 2. Conceptos Generales de Análisis de Datos	19
2.1 Elementos claves	19
2.2 Enfoque en el tipo de analítica.....	22
2.3 Estrategia de datos	23
2.4 Madurez analítica	25
2.5 Cultura	28
2.6 Mentalidad analítica.....	28
2.7 Elementos para Evolucionar a una Cultura data-driven.....	30
2.8 Roles y Responsabilidades.....	33
2.9 Elementos para Implementar una estructura de analítica de datos.....	36
2.9.1 Enfoque interno.....	36
2.9.2 Enfoque Externo – Atracción del talento	37
2.9.3 Enfoque Externo – Retención del Talento.....	38
2.9.4 Enfoque Externo – Desarrollo del Talento	38
Capítulo 3. Marco de Trabajo para el Análisis de Datos.....	40
3.1 Propósito del marco metodológico para el análisis de datos	40
3.2 Objetivos del marco.....	40
3.3 El público al que está dirigido.....	40
3.4 Estructura del marco	41
3.5 Definiciones	42

3.6 Proceso	43
3.6.1 Alcance.....	43
3.6.2 Dueño	43
3.6.4 Diagrama del proceso.....	44
3.6.5 Salidas	44
Capítulo 4. Planificación y Alineación con Estrategia	45
4.1 Actividades.....	45
4.1.1 Actividad 1: Sensibilización y entendimiento de metas estratégicas	45
4.1.2 Actividad 2: Desarrollo de Estrategia	47
4.2 Conceptos relacionados.....	50
4.2.1 Definición de Casos de Uso de la Organización.....	50
4.2.2 Priorización de Principales casos de Uso.....	52
4.2.3 Cargas de Trabajo	52
Capítulo 5. Ejecución del Ciclo de Descubrimiento	53
5.1 Actividades.....	53
5.1.1 Actividad 1 - Definición del Estudio o problema a solucionar.....	53
5.1.2 Actividad 2 – Preparación de Data	54
5.1.3 Actividad 3 – Búsqueda y manejo de valores atípicos	56
5.1.4 Actividad 4 – Evaluación y manejo de datos nulos o faltantes.	57
5.1.5 Actividad 5 – Escalamiento y Estandarización de la data.....	58
5.1.6 Actividad 6 – Aplicación del modelo Estadístico	59
5.2 Conceptos Relacionados.....	60
5.2.1 Underfitting y Overfitting	60
5.2.2 Reducción de errores y sesgos	60
5.2.3 Proceso de Limpieza	61
5.2.4 Separación de datos de Entrenamiento- Validación- Prueba	63
5.2.5 Definición y aplicación de Modelos sobre data de prueba.	64
5.2.6 Seleccionar modelo de prueba	64
5.2.7 Eficiencia del modelo.....	65
Capítulo 6. Comunicación de Resultados	66
6.1 Actividades.....	66
6.1.1 Actividad 1 – Preparar comunicación de resultados.....	66
6.2 Conceptos relacionados	68

6.2.1 Enfoques para la comunicación.....	68
6.2.2 ¿Cómo encontrar resultados rápidamente?	69
6.2.3 Herramientas de comunicación de resultados.....	70
6.2.4 Recomendaciones generales de presentación de resultados	71
Capítulo 7. Implementación de Acciones	74
7.1 Actividades.....	74
7.1.1 Actividad 1 – Implementar las acciones	74
7.1.2 Actividad 2 – Monitorear las acciones implementadas	75
7.2 Conceptos relacionados	75
Anexos	76
Ejemplo de cronograma de proyecto.....	76
Plan de proyecto.....	76
Formato para entrevista a personal de alto nivel de institución receptora	77
Formulario creación de caso de uso.....	78
Formulario de conjunto de datos	79
Matriz priorización.....	81
Referencias	83

Tablas

Tabla 1. Cuadro comparativo de enfoques de la data a través del tiempo	19
Tabla 2. Enfoque moderno de cultura organizacional	28
Tabla 3. Mentalidad analítica	29
Tabla 4. Roles en la Gestión Analítica.....	35
Tabla 5. Actividad 1. Sensibilización y entendimiento de metas estratégicas.....	47
Tabla 6. Actividad 2, Líder de implementación analítica.....	50
Tabla 7. Actividad 1, Ejecución del ciclo de Descubrimiento	54
Tabla 8. Actividad 2, Ejecución del ciclo de Descubrimiento	55
Tabla 9. Actividad 3, Ejecución del ciclo de Descubrimiento	56
Tabla 10. Actividad 4, Ejecución del ciclo de Descubrimiento	57
Tabla 11. Actividad 5, Ejecución del ciclo de Descubrimiento	58
Tabla 12. Actividad 6, Ejecución del ciclo de Descubrimiento	60
Tabla 13. Actividad 1, Comunicación de resultados.....	67
Tabla 14. Actividad 1, Implementar las acciones	74
Tabla 15. Actividad 2, Monitorear las acciones implementadas.....	75

Tabla de Figuras

Figura 1. Modelo Metodología CRISP-DM.....	9
Figura 2. Encuesta de metodología más utilizada para minería de datos	10
Figura 3. Niveles de CRISP-DM	10
Figura 4. CRISP-DM Fase Comprensión del negocio	11
Figura 5. CRISP-DM Fase Comprensión de los datos.....	11
Figura 6. CRISP-DM Fase Preparación de los datos.....	12
Figura 7. CRISP-DM Fase Modelado	12
Figura 8. CRISP-DM Fase Evaluación	13
Figura 9. CRISP-DM Fase Implantación.....	13
Figura 10. Etapas del proceso KDD.....	14
Figura 11. Elementos claves para el análisis de datos.....	19
Figura 12. Cadena de valor para a generación de Insights.....	21
Figura 13. Enfoque en el tipo de analítica	22
Figura 14. Enfoque pragmático	22
Figura 15. Formas de abordar la data	24
Figura 16. Nivel de Madurez de la Analítica de Datos.....	25
Figura 17. Madurez analítica, ¿cómo está capturada la data?	26
Figura 18. Madurez analítica, ¿qué es capturado?	27
Figura 19. Madurez analítica, otros.....	27
Figura 20. Sistemas que influyen la cultura.....	29
Figura 21. Elementos para evolucionar a una cultura data-driven	30
Figura 22. Proceso de Analítica de Datos	44
Figura 23. Subproceso de Planificación y Alineación con la Estrategia.....	45
Figura 24. Definición de Casos de Uso	51
Figura 25. Agrupación de cargas de trabajo.....	52
Figura 26. Subproceso Ejecución del Ciclo de Descubrimiento	53
Figura 27. Subproceso Comunicación de Resultados.....	66
Figura 28. Enfoque estratégico vs Enfoque Táctico	68
Figura 29. Enfoque estratégico y Enfoque Táctico	68
Figura 30. Cómo encontrar grupos en la data.....	69
Figura 31. Enfoque correcto de visualización.....	70
Figura 32. Uso correcto de colores en gráficos	71
Figura 33. Uso correcto de ejes en gráficos	72
Figura 34. Tableros no cargados.....	72
Figura 35. Ejemplo de cómo relacionar información para facilitar su recuerdo.....	72
Figura 36. Ejemplo Panel de Ventas	73
Figura 37. Herramientas BI	73
Figura 38. Subproceso Implementación de Acciones	74

Tabla de Anexos

Anexos 1. Ejemplo de Cronograma de Proyecto	76
Anexos 2. Plan de Proyecto	76
Anexos 3. Formato para entrevista a personal	77
Anexos 4. Formato para entrevista a personal	77
Anexos 5. Formulario de creación de casos	78
Anexos 6. Formulario de conjunto de datos	79
Anexos 7. Matriz de priorización	82

Capítulo 1. Estado del Arte

1.1 Introducción

Muchos países de Latinoamérica han centrado sus esfuerzos en la gestión y diseño de estrategias (Borzacchiello & Craglia, 2012) orientadas a fortalecer la gobernanza (Gonzalez-Zapata & Heeks, 2015), a través del uso estratégico de los datos en todos los estamentos públicos que pertenecen a su estructura de políticas públicas (Ruijter, Grimmelikhuijsen, & Meijer, 2017; Zhang, Puron-Cid, & Gil-Garcia, 2015). Según la OECD, se han establecido algunas recomendaciones para un mejor acceso y un uso más eficaz de la información del sector público. Public Sector Information (PSI) (OECD, 2008) que hacen referencia a qué es información dentro del sector público. Se incluyen aquí toda información tanto de productos como servicios, que es generada, creada, recolectada, procesada, preservada, mantenida, diseminada o financiada por una institución pública o gubernamental.

En (Ubaldi, 2013) se presentan estudios sobre políticas de gobernanza pública preparados para su uso dentro de la OCDE. Estos documentos abordan cuestiones como las pensiones, el envejecimiento, la gestión, la remuneración y la descentralización en el sector público. Sugieren una metodología que comprende un marco analítico para iniciativas de Datos de Gobierno Abierto (OGD) (que se aplicarán al análisis ex post y ex ante de iniciativas) y un conjunto de datos relacionados que se recopilarán en los países de la OCDE.

El artículo (Matheus et al., 2012) presenta los datos de gobierno abierto en los gobiernos municipales, estatales y nacionales de Brasil. Y muestran que el uso de los datos abiertos ha generado transparencia, mejoras en el control social y participación ciudadana. Este análisis se llevó a cabo con un estudio exploratorio verificando los ocho principios de los OGD y las cinco estrellas de datos abiertos vinculados creados por Tim Berners-Lee que miden el nivel de los datos abiertos en sitios web.

Una investigación publicada en (Linders, 2013) demuestra cuales técnicas actualmente en el área de datos abiertos del gobierno han proporcionado contribuciones significativas para hacer realidad la integración de entrega de ayuda. También investigan como las agencias de ayuda aprovechan los recursos de datos abiertos para ellos mejorar, sin embargo, el gran porcentaje de dichos recursos no se han explorado y aprovechado con eficacia.

En Panamá se ha iniciado el diseño de nuevas estrategias que permiten al sector gubernamental, una adecuada gestión de los datos públicos de las diversas instituciones, desde la Autoridad Nacional para la Innovación Gubernamental (AIG, 2020).

Estos proyectos se han formulado basándose en las necesidades fundamentales del país, buscando darle un valor público a los datos que se generan de forma masiva. Como punto de inicio, se ha desarrollado un diagnóstico en donde se han detectado algunas necesidades como: facilitar la interoperabilidad de los datos, mejorar la cultura en los experto de TI para aprovechamiento de los datos que se generan desde diversas instituciones del estado,

instituciones públicas con estrategias con la necesidad de nuevas estrategias, mejora de una política y marco de trabajo para el depuramiento, aprovechamiento y explotación de los datos como política de estado, facilitar la relación entre las tecnologías existentes, el perfil de las personas y la cultura de trabajo centrada en datos. Todos estos elementos nos han llevado a definir una estrategia que busca diseñar un ecosistema para la gestión y uso estratégico de datos públicos a través de un modelo que fortalezca la gobernanza de Panamá.

En este documento se presentan las fases y elementos que se han fundamentado y propuesto como parte de ese ecosistema, la detección de actores involucrados, y sobre todo el marco de trabajo a seguir para facilitar la integración de diversos tipos de datos de otros sectores del estado.

1.2 Metodologías utilizadas en la minería de datos

En los inicios del año 1996, el modelo KDD (Knowledge Discovery in Databases) constituyó el primer modelo aceptado en la comunidad científica que estableció las etapas principales de un proyecto de explotación de información. Formalmente el modelo establece que la minería de datos es la etapa dentro del proceso en la cual se realiza la extracción de patrones a partir de los datos. Sin embargo, actualmente, en la comunidad científica y en la literatura, el término KDD y minería de datos se utilizan indistintamente para hacer referencia al proceso completo de descubrimiento de conocimiento (Juan et al., n.d.).

Algunos modelos profundizan en mayor detalle sobre las tareas y actividades a ejecutar en cada etapa del proceso de minería de datos (como CRISP-DM), mientras que otros proveen sólo una guía general del trabajo a realizar en cada fase (como el proceso KDD o SEMMA).

A continuación, presentamos las metodologías más aceptadas en la industria.

1.2.1 Metodología SEMMA

La metodología SEMMA se encuentra enfocada especialmente en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema que se está abordando. Fue propuesta especialmente para trabajar con el software de minería de datos de la compañía SAS. Este producto organiza sus herramientas (llamadas “nodos”) en base a las distintas fases que componen la metodología. Es decir, el software proporciona un conjunto de herramientas especiales para la etapa de muestreo, otras para la etapa de exploración, y así sucesivamente. Sin embargo, el usuario podría hacer uso del mismo siguiendo cualquier otra metodología de minería de datos (como CRISP-DM por ejemplo).

1.2.2 Metodología Catalyst

La metodología Catalyst, conocida como P3TQ (Product, Place, Price, Time, Quantity), fue propuesta por Dorian Pyle en el año 2003. Esta metodología plantea la formulación de dos

modelos: el Modelo de Negocio y el Modelo de Explotación de Información. La metodología Catalyst, en sus dos modelos, está compuesta por una serie de pasos llamados “boxes”. El concepto es que luego de llevar a cabo una acción, se deben evaluar los resultados y determinar cuál es el próximo paso (box) a seguir. La secuencia y la interacción entre los distintos pasos permiten una flexibilidad muy grande, y una amplia variedad de caminos posibles.

1.2.3 Metodología CRISP-DM

Estructura el proceso en seis fases (figura 1): Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implantación [5].

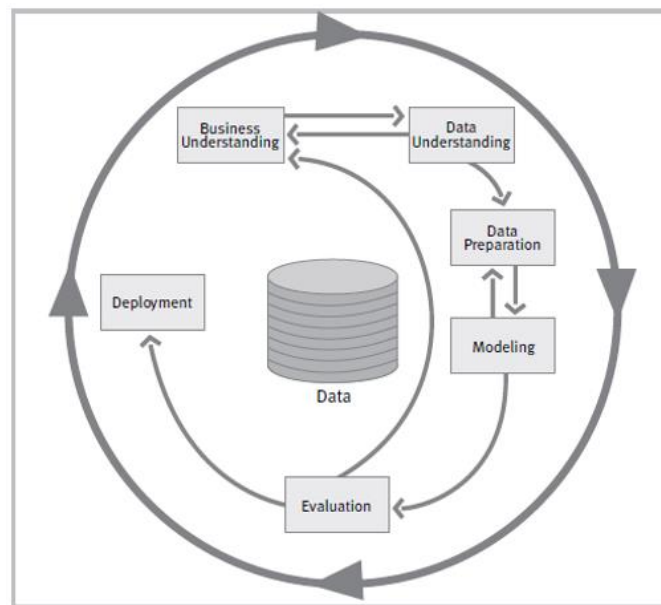


Figura 1. Modelo Metodología CRISP-DM

La sucesión de fases no es necesariamente rígida. Cada fase es descompuesta en varias tareas generales de segundo nivel; es decir, aquellas tareas que... (explicar que es una tarea de segundo nivel). Las tareas generales se proyectan a tareas específicas, pero en ningún momento se propone como realizarlas. Es decir, CRISP-DM establece un conjunto de tareas y actividades para cada fase del proyecto, pero no especifica cómo ejecutarlas.

CRISP-DM, como se puede ver en la figura 2 es actualmente la guía de referencia más utilizada en el desarrollo de proyectos de Data Mining.

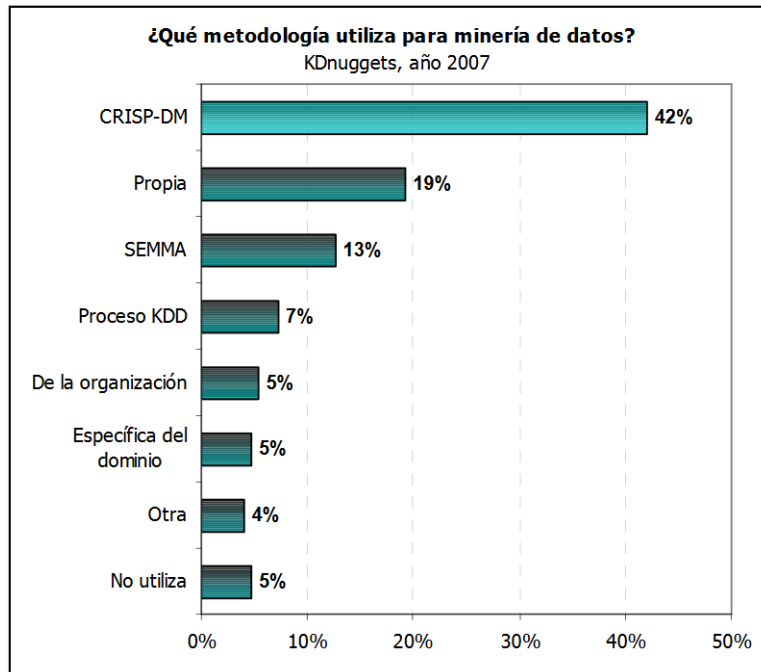


Figura 2. Encuesta de metodología más utilizada para minería de datos

CRISP-DM se divide en 4 niveles de abstracción organizados jerárquicamente en tareas que van desde el nivel más general hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases (figura 3) (Gallardo Arancibia, 2013).

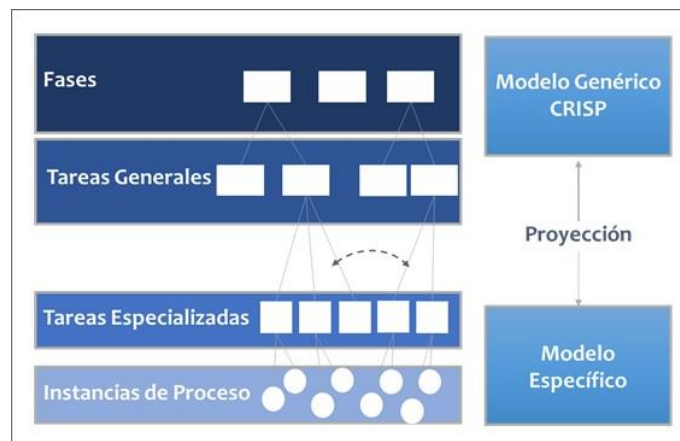


Figura 3. Niveles de CRISP-DM

Cada una de las fases en las que se divide CRISP-DM se describe a continuación (Figura 4).

Fase de comprensión del negocio o problema: Probablemente sea la más importante y aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, para convertirlos en objetivos técnicos y un plan de proyecto. Sin poder comprender estos objetivos, ningún algoritmo, por sofisticado que sea, permitirá obtener resultados fiables.

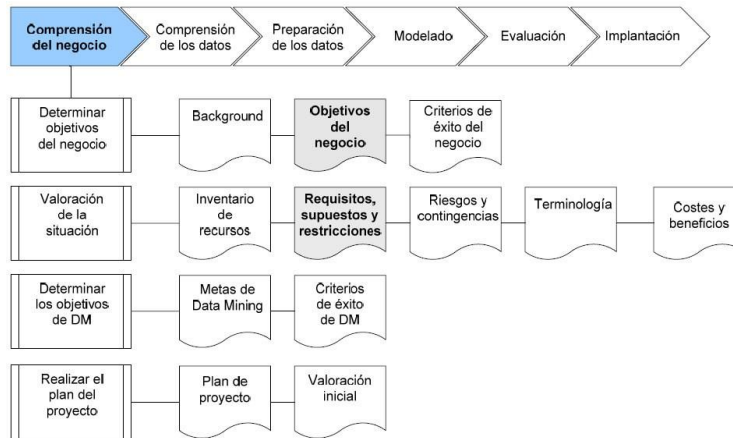


Figura 4. CRISP-DM Fase Comprensión del negocio

Fase de comprensión de los datos: comprende la recogida inicial de datos, con el objetivo de establecer un primer contacto con el problema, familiarizarse con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis (Figura 5).

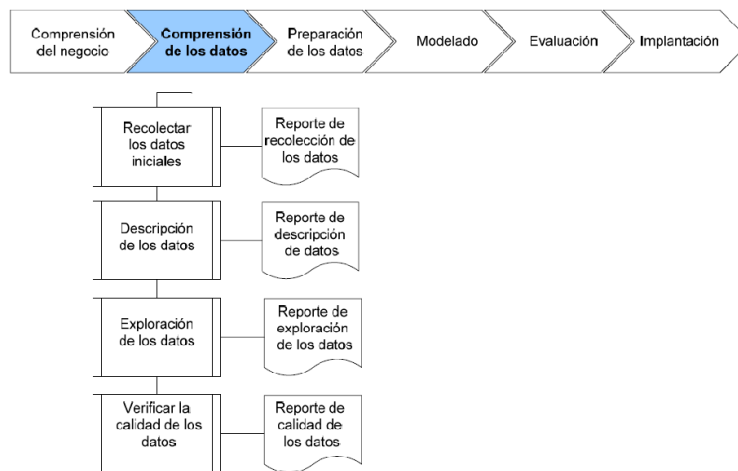


Figura 5. CRISP-DM Fase Comprensión de los datos

Fase de preparación de datos: una vez realizada la recogida de datos inicial, se prepara para adaptarlos a las técnicas de Data Mining que se utilizan posteriormente, como técnicas de visualización de datos, búsqueda de relaciones entre variables u otras. medidas para la exploración de datos. La preparación de datos incluye las tareas generales de seleccionar datos a los que se aplicará una técnica de modelado particular, limpieza de datos, generación de variables adicionales, integración de diferentes fuentes de datos y cambios de formato (Figura 6).

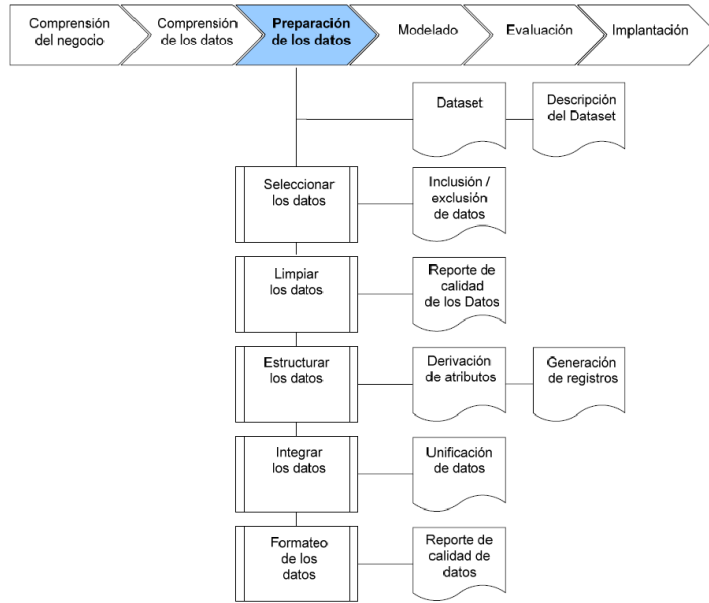


Figura 6. CRISP-DM Fase Preparación de los datos

Fase de modelado: En esta fase de CRISP-DM, se seleccionan las técnicas de modelado más adecuadas. Se inicia con la construcción del modelo en la herramienta seleccionada, aquí se genera el resultado del modelo aplicado. Luego se evalúa el mismo aplicando las pruebas de evaluación y se compara su resultado en base a los criterios de evaluación definidos (Figura 7).

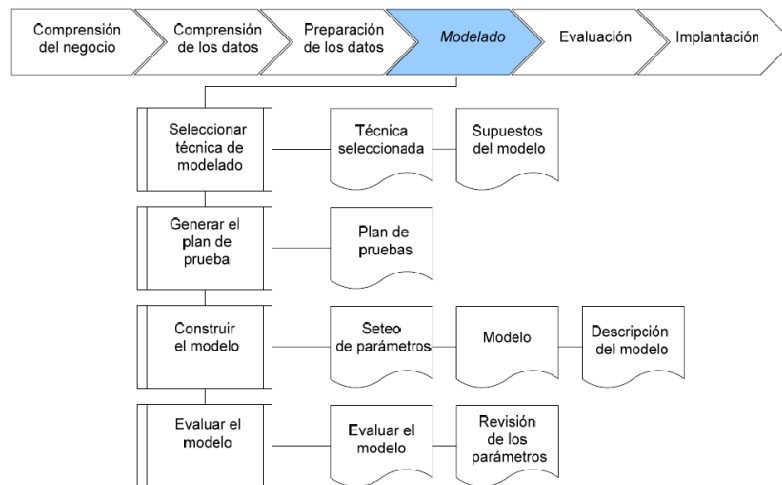


Figura 7. CRISP-DM Fase Modelado

Fase de evaluación: En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. También se debe considerar que la confiabilidad calculada para el modelo se aplica solo a los datos sobre los cuales se realizó el análisis (Figura 8).

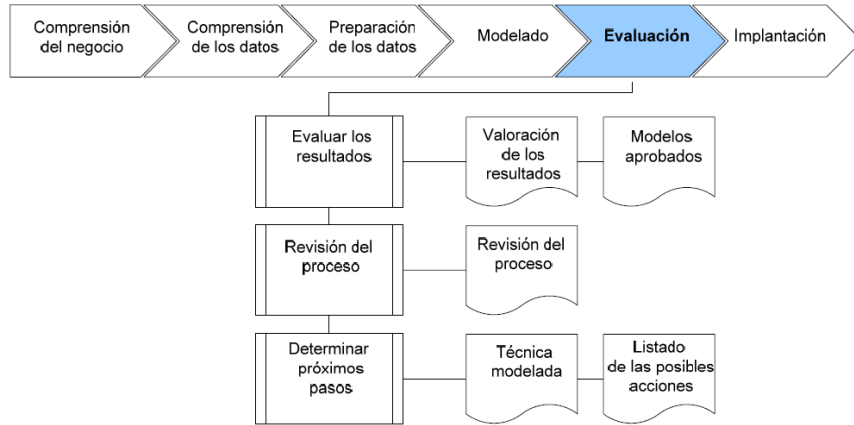


Figura 8. CRISP-DM Fase Evaluación

Fase de implementación: una vez construido y validado el modelo, el conocimiento obtenido se transforma en acciones dentro del proceso de negocio, ya sea que el analista recomiende acciones basadas en la observación del modelo y sus resultados, o aplicando el modelo a diferentes conjuntos de datos o como parte del proceso, como en el análisis de riesgo crediticio, detección de fraude, etc. (Figura 9).

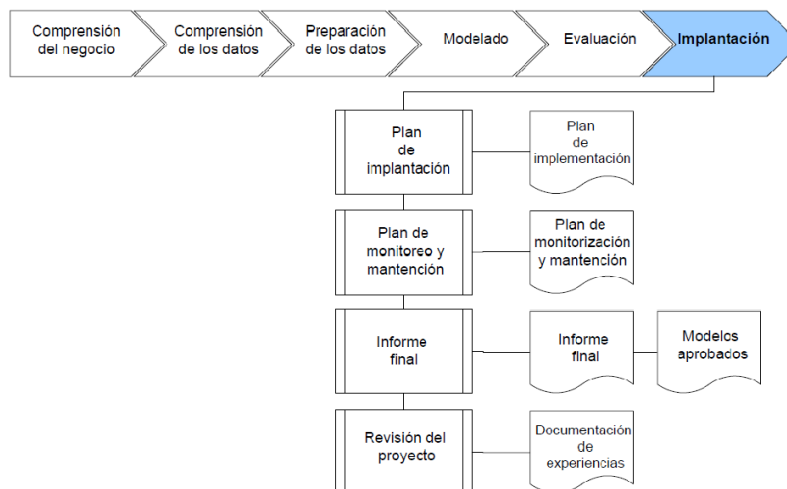


Figura 9. CRISP-DM Fase Implantación

1.2.4 Proceso KDD

El concepto de KDD se ha desarrollado, y continúa desarrollándose, desde la intersección de la investigación de áreas tales como bases de datos, aprendizaje automático, reconocimiento de patrones, estadística, teoría de la información, inteligencia artificial, razonamiento con incertidumbre, visualización de datos y computación de altas prestaciones (Riquelme et al., 2006) . El descubrimiento de conocimiento en bases de datos se caracteriza por ser un proceso compuesto por tres pasos operativos (la figura 10) muestra los pasos operativos del proceso KDD) cuyos nombres surgieron fueron los siguientes: *paso de preprocesamiento*, que comprende las funciones que se relacionan la captura, organización y procesamiento de datos, cuya finalidad es preparar los datos para los algoritmos de la siguiente etapa; *paso de Data Mining* ,que realiza la búsqueda efectiva de conocimientos útiles en el contexto propuesto para la aplicación de KDD; *paso de pos-procesamiento*, que cubre el tratamiento de los conocimientos obtenidos en el paso anterior (Nazareno Pereira Boente Fabiano Saldanha Gomes de Oliveira José Luiz Dos Anjos Rosa, n.d.).

El paso de pos-procesamiento no siempre es necesario, aunque tiene como objetivo evaluar la utilidad del conocimiento descubierto. Como puede verse, en general, la complejidad del proceso KDD radica en la dificultad de percibir e interpretar correctamente los innumerables hechos observados durante un proceso y en la dificultad de juzgar dinámicamente tales interpretaciones para decidir qué acciones deben realizarse en cada caso.

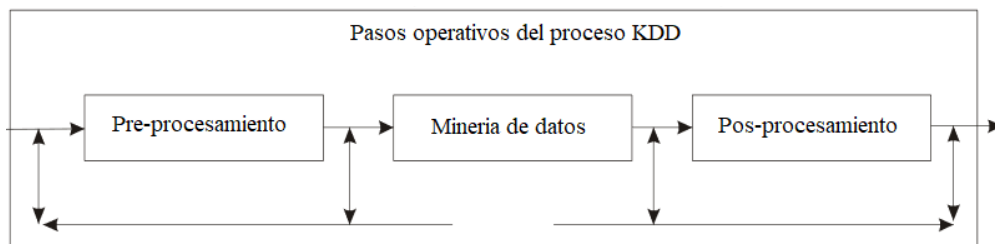


Figura 10. Etapas del proceso KDD

Las principales funciones del Preprocesamiento de Datos son las siguientes:

- **Selección de Datos:** También conocida como Reducción de Datos, identifica qué información, entre las bases de datos existentes, debe ser efectivamente considerada durante el proceso KDD.
- **Limpieza de Datos:** Implica cualquier tipo de tratamiento relacionado con los datos seleccionados con el fin de asegurar la calidad de los hechos que representa.
- **Codificación de datos:** los datos se codifican de manera que puedan tener una forma que se pueda utilizar eficazmente como entrada para los algoritmos de minería de datos.

- **Enriquecimiento de datos:** Consiste en obtener de alguna manera más información que se puede agregar a los registros existentes, enriqueciendo así los datos para que sean capaces de brindar más información para llevar a cabo el proceso de descubrimiento de conocimiento en bases de datos.

La elección de la técnica a utilizar en el proceso de Data Mining depende exclusivamente del tipo de tarea KDD a realizar de manera efectiva, de acuerdo con la siguiente clasificación:

- **Descubrimiento de Asociación:** Abarca la búsqueda de ítems que ocurren frecuentemente desde simultáneamente en varias transacciones de la base de datos.
- **Clasificación:** Consiste en descubrir una función que mapea un conjunto de registros a un conjunto de etiquetas categóricas predefinidas, llamadas clases.
- **Regresión:** Consiste en un mapeo efectivo de los registros de una determinada base de datos en busca de valores reales. Esta tarea está muy cerca de la tarea de clasificación, presentada anteriormente.
- **Agrupación:** El objetivo separa los registros de una base de datos en subconjuntos de agrupaciones para que los elementos de una agrupación compartan propiedades comunes que los distinguen de los elementos de otras agrupaciones.
- **Resumen:** Busca identificar características comunes entre conjuntos de datos.
- **Detección de Desviaciones:** Consiste en localizar registros de una base de datos cuyas características no cumplen con los estándares considerados normales en el contexto actual.
- **Secuencia de descubrimiento:** Se utiliza para descubrir que existen posibles cambios estacionales en la cantidad en el envío de ciertas piezas.

Finalmente, la etapa de pos-procesamiento abarca el tratamiento de los conocimientos adquiridos a través del proceso de Data Mining. A menudo, este tipo de tratamiento se vuelve innecesario, ya que este paso apunta a la facilidad de interpretación y evaluación, por parte del hombre, en relación con la utilidad del conocimiento descubierto. Entre las principales funciones de la etapa de pos-procesamiento se encuentran la elaboración y organización, que puede incluir la simplificación de gráficos, esquemas y otro tipo de informes demostrativos, además de convertir la forma de representación de los conocimientos obtenidos.

1.3 Antecedentes para la gestión de datos en gobierno

Presentamos algunos trabajos relacionados que tienen como objetivo orientar el proceso de gestión de datos en el sector gubernamental desde un marco o modelo.

En el trabajo propuesto (Kum et al., 2015) sintetizan los diferentes marcos para el uso de BIG DATA con la metodología o proceso KDD en conjunto, utilizando dicho proceso en los datos abiertos de la gobernanza del sistema de bienestar infantil, además de actualización del proceso KDD que se puede implementar con los frameworks para su correcto uso con los BIG DATA.

En (Vetrò et al., 2016) se señalan ciertas limitaciones para los problemas actuales que contienen la difusión de datos de gobierno abierto y un marco de indicadores, como propuesta, para medir la calidad de los datos que se exponen o difunden. La validación de este marco propuesto se basa en comparar la divulgación centralizada y descentralizada de datos de gobernanza abierta. Como resultado de las mediciones, se obtuvieron métricas de guía para mejorar las debilidades observadas en la liberación descentralizada de los datos.

El artículo (Sieber & Johnson, 2015) propone cuatro modelos que relacionan ciudadanos y gobierno, estos incluyen el status quo de publicación de datos gubernamentales, el gobierno como tal actuando como activista de los datos abiertos, como rastreador de problemas cívicos y datos abiertos participativos. Finalmente, se enfatiza que el futuro de los datos abiertos estará relacionado con la ética y la economía que existe entre la ciudadanía y el gobierno.

En (Magalhaes et al., 2013) proponen un marco que ayuda a desarrollar una terminología coherente para describir, comparar y discutir diferentes enfoques a los servicios, ya que las terminologías utilizadas en el campo de los datos de gobernanza no son muy claras, lo que lleva a confundir a las personas.

El trabajo propuesto en (Kalampokis et al., 2011) contempla un modelo que contiene dos dimensiones, complejidad organizativa y tecnológica y valor agregado para los consumidores de estos datos, con el fin de cubrir los problemas que existen a la hora de publicar datos abiertos, referidos a la lineamientos, marcos y evaluación, es decir, un procedimiento a seguir. Cabe señalar que este modelo será utilizado por profesionales para construir hojas de ruta y para realizar evaluaciones comparativas.

En (Janssen & van den Hoven, 2015) se propone un marco denominado BOLD (Big and Open Linked Data), dicho marco tiene la capacidad de brindar oportunidades para el análisis del comportamiento de las personas, aumentar el control y reducir la privacidad. Sin embargo, también se puede utilizar para crear un gobierno abierto y transparente. Uno de los pilares fundamentales en el uso de este marco es la privacidad y la transparencia, que se puede ajustar con seguridad.

En (Ruijter et al., 2020) se desarrolla un marco analítico para el estudio de la política de datos abiertos del gobierno, basado en teorías de respuestas de las instituciones. Esto en sí mismo establece que pueden existir diferentes estrategias por parte de las instituciones, individualmente, que muestran sus datos de manera abierta y que sus respuestas pueden depender de sus antecedentes. La aplicación de este marco se implementó en Francia y Holanda, obteniendo como resultados diferentes niveles de transparencia, además de un concepto más comprensible sobre las prácticas de datos abiertos con relación a las instituciones y su transparencia con el gobierno.

El autor (Ubaldi, 2013) propone un marco analítico de datos abiertos de gobierno que permite obtener evidencia clara sobre el desarrollo de un conjunto de métricas para una evaluación de impacto y creación de valor entre los países que implementan la difusión de datos abiertos y marco, es decir, proponen un marco metodológico para medir el impacto y valor acumulado de iniciativas en el área de Datos Abiertos de Gobierno (DAG).

Como poder apreciar, DAG forma parte en gran medida de iniciativas y proyectos relacionados a la gestión de datos gubernamentales. A continuación, se describe brevemente el proyecto de Datos Abiertos de Panamá (DAP).

1.4 Datos abiertos en Panamá

La Autoridad Nacional de Transparencia y Acceso a la Información (ANTAI) y la Autoridad Nacional de Innovación Gubernamental (AIG) son las entidades que lideran el nivel nacional de acceso y publicación de datos abiertos (<https://www.datosabiertos.gob.pa/>). El acceso a la información pública que genera el Estado promueve la participación ciudadana a través del uso y reutilización de esta información en aras de la transparencia e innovación de los servicios públicos, la interoperabilidad del Gobierno y la prestación de las restricciones a las donaciones establecidas por la Ley (Apertura, 2016).

Los datos generados por las instituciones públicas pueden ponerse a disposición de la sociedad en la calidad de Datos Abiertos, aumentando su potencial de uso, reutilización y redistribución, generando impacto en el sentido de:

- Fomento de la participación ciudadana y fortalecimiento de la democracia.
- Incrementar la transparencia y la rentabilidad del dinero.
- Impulsar el crecimiento económico, fortalecer la competitividad y promover la innovación.
- Fomentar una mayor eficiencia gubernamental, lo que imposibilita diseñar políticas públicas basadas en las necesidades de la comunidad y mejorar los servicios públicos, apoyando objetivos de desarrollo, generación de conocimiento y buen gobierno.

- Emitir datos en el marco de la privacidad y protección de la información personal.
- Permitir la interoperabilidad entre organizaciones y sistemas. La capacidad de integrar componentes esenciales para construir sistemas complejos, grandes y ricos en información.
- Fomentar la investigación y periodicidad de datos, lo que posibilita la recuperación y análisis de grandes cantidades de datos utilizando software especializado para brindar información integral a la comunidad a través de artículos, infografías y visualizaciones de datos, entre otros.

Plan institucional como estrategia para la apertura de datos

- **Delegación:** Es necesario en primera instancia designar al líder de datos abiertos, quien será el responsable institucional de apertura de datos. La designación, por regla general, se lleva a cabo por la autoridad de la institución.
- **Conformación:** El equipo de trabajo actuará como un consejo institucional para implementar la política de datos abiertos en toda la institución.
- **Generación:** Es importante ordenar la información con el cuidado de la institución. Para ello, se crea el inventario de datos institucionales que incluye toda la información que se produce a partir de las bases de datos y otros sistemas de manejo de datos con información estadística, financiera, investigativa, administrativa, legal, geolocalización, proyecto, entre otros.
- **Publicación:** Una vez listado el inventario, el siguiente paso es decidir en qué orden se publicarán los datos abiertos, tanto en el sitio web de la institución como en el portal nacional, tomando como guía el potencial de generar las mismas cosas. impacto en la comunidad.

Capítulo 2. Conceptos Generales de Análisis de Datos

Análisis de datos, es un término utilizado para representar el enfoque cuantitativo para la toma de decisiones en las organizaciones. En el sentido amplio es un proceso, en el cual, un grupo de personas ayuda a la organización a tomar mejores decisiones (objetivo) a través del análisis de la información (actividad).

A través de tiempo, el enfoque de la data ha cambiado en las organizaciones. En la tabla 1, podemos ver y comparar el enfoque.

Antes	Ahora
Es caro generar data	La data se genera continuamente en todas partes
El reto de la data es almacenarla	El reto es transformar la data en resultados
Solo se usa data estructurada	La data no-estructurada se incrementa y es valiosa
La data es administrada de forma aislada	El valor de la data está en conectar para crear sinergias
La data es una herramienta para optimizar procesos	La data es un activo intangible para crear valor

Tabla 1. Cuadro comparativo de enfoques de la data a través del tiempo

2.1 Elementos claves

En términos generales, el análisis de datos conlleva 4 elementos claves para que pueda ser desarrollado de forma efectiva (Figura 11):

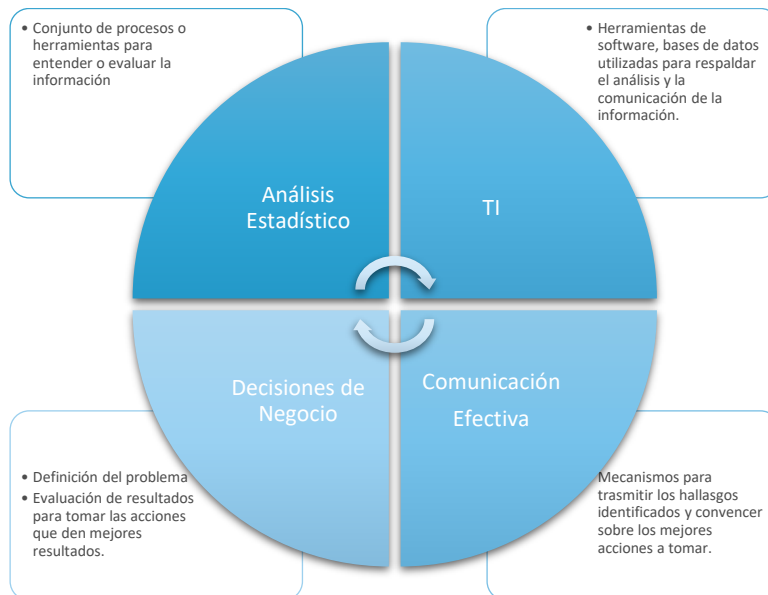


Figura 11. Elementos claves para el análisis de datos

Adicional a estos elementos claves, existen otras 2 consideraciones importantes, que son los datos que vamos a trabajar, y por supuesto, las personas y procesos que van a ejecutar la analítica de datos, que veremos identificados en la cadena de valor para la generación de insights.

El análisis de datos debe servirnos para tomar la mejor decisión, en tiempo y forma. El propósito del análisis de datos es entender el problema que queremos resolver para definir acciones concretas¹. El análisis de datos no es solo el software o la estadística. De hecho, estas son las herramientas que soportan nuestro proceso de análisis, cuyo fin es crear valor mediante la toma de decisiones más objetivas. En este sentido nuestra línea de pensamiento debe responder las siguientes preguntas:

- ¿Cómo la data puede facilitarnos recomendaciones accionables?
- ¿Cuáles son las herramientas que necesitamos para analizar la data?
- ¿Qué significan los “outputs” de las herramientas analíticas para tomar decisiones?
- ¿Cuáles son las conclusiones relevantes del análisis de datos?
- ¿Cómo hacemos esas conclusiones accionables para mejorar la gestión gubernamental?

Siguiendo esta línea, estos 4 elementos clave se pueden descomponer en subelementos, cada uno con un objetivo en particular, que combinados deben llevar al objetivo final que es la mejor toma de decisiones para solventar situaciones particulares. De forma amplia, extraer valor de los datos conlleva los elementos contemplados en la cadena de valor para la generación² de “insights” mostrado en la figura 12:

¹ Nicolas Glady Nicolas Glady. Accenture Strategic Business Analytics chair holder and Center for Business Director, ESSEC Business School.

² Adaptación propia de Achieving Business Impact with data, A comprehensive perspective on the insight value chain. Digital Mckinsey, 2018.



Figura 12. Cadena de valor para a generación de Insights

Precisamente, esta agrupación en aspectos técnicos y aspectos organizacionales nos genera nuestra cascada de implementación para nuestro proceso de análisis de datos y captura de insights.

2.2 Enfoque en el tipo de analítica

Antes se describía a la analítica de datos como algo que iba evolucionando, en el que se le hallaba más valor en base a la complejidad, iniciando con la analítica descriptiva, luego el diagnóstico analítico, la analítica predictiva y la analítica prescriptiva.

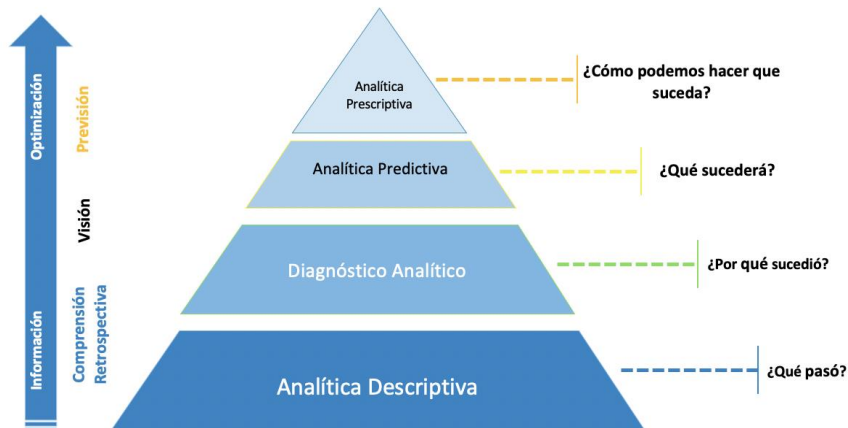


Figura 13. Enfoque en el tipo de analítica

No necesariamente mientras más complejo sea el algoritmo que utilizamos o la interrelación de información vamos a tener mejores resultados, eso depende del caso y del tema institucional que queremos resolver. Por esa razón se utiliza más el enfoque pragmático.

Enfoque pragmático

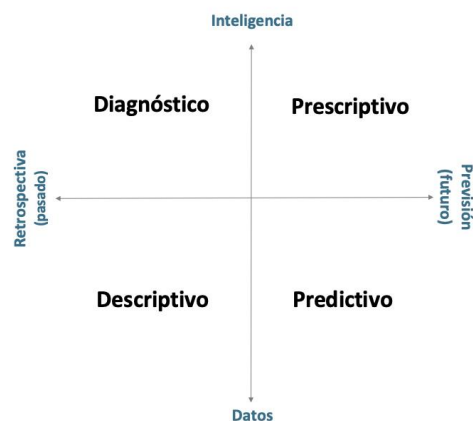


Figura 14. Enfoque pragmático

En este enfoque tenemos dos ejes:

- Va de la información o los datos hacia el uso de uso o aprovechamiento de la misma llamada “inteligencia”
- En sentido horizontal va del pasado hacia el futuro o prevención.

Este cuadro muestra que:

- Con más inteligencia y más retrospectivas se genera el diagnóstico se ve (¿por qué ocurrieron las cosas?)
- Con inteligencia más datos en el pasado se ve (¿qué ocurrió?, ¿cuántas personas fueron?, números)
- En cuanto al predictivo, pero hacia la información o datos en el futuro se ve (¿qué pasará?, ¿cuánto será mañana?, el próximo año)
- El prescriptivo tiene que ver con la información que nos da el sistema para nosotros tomar mejores decisiones.

2.3 Estrategia de datos

La estrategia de datos se trata de que hacemos con la data, no de que le hacemos a la data. El enfoque moderno de la estrategia de datos es “facilitar” el cumplimiento de las líneas estratégicas de la organización. El proceso de la estrategia de datos es iterativo, se empieza por entender la estrategia.

Hay que traducir la estrategia en objetivos

- Traducir o Segregar tu estrategia para hacerla tangible, alcanzable y medible.
- Estos son tus objetivos organizacionales.
- Esto proveerá una la guía para diferentes elementos de la estrategia de datos.

Toda organización debe contar con una estrategia de Datos, que de enfoque y resultados.

- Existen diferentes tipos de datos, así como el uso que podemos darle.
- Existen diferentes formas de abordar la data, dependiendo de los objetivos de negocio.

Existen diferentes formas de abordar la data, dependiendo de los objetivos de negocio



Figura 15. Formas de abordar la data

Estas indican que las posibles estrategias en el uso de la información son básicamente las mismas que otros grupos literarios tienen, el primero es:

- **Descubrir patrones**, lo llaman descubrir lo invisible, en cuanto al análisis de información, se utiliza la data para encontrar tendencias en grupos económicos, actividades económicas o en personas.
- **Targetting** tiene que ver con reducir el enfoque, separación o segmentación, se definen mecanismos de segmentación.
- **Personalización**, se ajusta lo que uno ofrece a sus grupos económicos de manera personalizado, grupos de edades, por sexo. Son grupos específicos.
- **Contexto**, es cuando se pone a disposición de las personas información para que tomen decisiones.

2.4 Madurez analítica

Identificar la capacidad analítica de la institución es importante para determinar las brechas y así establecer las acciones apropiadas, de tal forma que el proceso de analítica cumpla los objetivos para el cual está diseñado.

La madurez se puede identificar en función de la disponibilidad de los datos para ser transformados en información útil, evaluando la información que es capturada, que mecanismos son utilizados para su registro y almacenamiento y qué políticas de seguridad y documentación rigen la institución.

Categoría	Nivel 1	Nivel 2	Nivel 3	Nivel 4
¿Cómo está capturado?	No es posible reutilizarse o analizarse.	Es posible analizarse de forma complicada o mediante mecanismos especializados.	Es posible analizarse en la fuente.	Permite la comunicación interna y externa, de forma segura.
¿Qué es capturado?	Información irrelevante, con baja calidad, frecuencia no apropiada.	Algo relevante, falta de información, zonas geográficas muy amplias.	Relevante y se puede resolver, debilidades en la información histórica, algunos errores en atributos, oportunidades en la granularidad.	Se tiene data suficiente para generar análisis de calidad, información en tiempo real, granularidad suficiente.
Privacidad y Documentación	Falta de políticas y metadata.	Diccionario de Datos existente.	Uso de información con aprobaciones en sitio, se cuenta con metadata.	Software para uso de datos de forma protegida, metadata más comentarios y sesgos en la data.

Figura 16. Nivel de Madurez de la Analítica de Datos

En el Marco de Madurez Analítica hay tres elementos que se deben identificar:

1. **¿Cómo esta capturada la data?** Se ve si estamos combinando elementos de gobierno con elementos relacionados a la estrategia de analítica. En este punto vemos tres puntos principales:
 - **Accesibilidad:** en el nivel 1 se mantiene la data solo donde es capturada, en el nivel 2 es accesible por medio de software especializado, en el nivel 3 se pueden utilizar

formatos y en el nivel 4 se puede acceder en la mayoría de los dispositivos con formatos estándar.

- **Almacenaje:** en el nivel 1 se maneja en papel, en el nivel 2 PDF e imágenes, en el nivel 3 se puede tener archivos de texto y en el nivel 4 base de datos.
- **Integración:** se puede ver que en el nivel 1 se mantiene la fuente, en el nivel 2 se puede ver si es exportado ocasionalmente, en el nivel 3 si se tiene un DWH centralizado en tiempo real y si se maneja e incluye data externa que está integrada a la base de datos se ve en el nivel 4.

Categoría	Área	Nivel 1	Nivel 2	Nivel 3	Nivel 4
Cómo está capturada la data	Accesibilidad	Solo dónde es capturada	Accesible con software especializados.	Accesible en formatos (CSV, JSON, XML)	Cualquier dispositivo, en formato estándar y API
	Almacenaje	Papel	Pdf o imágenes	Archivos de Texto	Bases de datos
	Integración	Permanece en la fuente	Exportada ocasionalmente y de forma Ad-hoc	DWH centralizado, tiempo real.	Data externa también integrada

Figura 17. Madurez analítica, ¿cómo está capturada la data?

2. **¿Qué es capturado?** Aquí vemos la información de lo que se captura, en que hay 5 áreas de evaluación.

- Relevancia y suficiencia en función de los problemas que queremos resolver, cada uno de los que tengan definidos dentro de la estrategia.
- Calidad, vemos específicamente lo que se quiere resolver ya que cuando hablamos de “Gobierno de datos” debe ser más detallado y profundo en base al problema que queremos solucionar.
- Frecuencia de la información.
- Amplitud y profundidad (Granularidad).
- Historia.

Categoría	Área	Nivel 1	Nivel 2	Nivel 3	Nivel 4
Qué es capturado	Relevancia y Suficiencia	Data irrelevante para los problemas que quieres resolver.	Alguna data es relevante, pero insuficiente porque hay información faltante.	Relevante y suficiente para resolver el problema. Insuficiente para resolverlo bien.	Se tiene toda la data relevante y suficiente.
	Calidad	Falta de filas e información importante.	Falta de columnas o atributos.	Algunos errores tipográficos.	Sin faltantes de información
	Frecuencia	Una vez, y nunca más.	Anual	mensual	Tiempo real
	Amplitud y Profundidad. Granularidad	A nivel provincia.	A nivel distrito.	A nivel individual, persona u hogar.	A nivel de incidentes.
	Historia	Sin Historia	Se almacena, pero se sobrescribe con nueva información	Se almacena y la nueva es incluida, vía "timestamp" pudiendo usar valores antiguos.	La nueva data se incluye mapeada.

Figura 18. Madurez analítica, ¿qué es capturado?

3. Otros elementos, en este vemos 2 áreas

- Privacidad, que tengamos información clara con los elementos de privacidad.
- Documentación, se ve que al aumentar los niveles van incrementando los detalles de información.

Categoría	Área	Nivel 1	Nivel 2	Nivel 3	Nivel 4
Otros	Privacidad	Sin políticas de privacidad	Se puede utilizar información "non-PII"	Aprobaciones ad hoc para el uso de información PII	Software especializado que permite el uso de información, protegiendo los datos el individuo.
	Documentación	Sin registros de metadata ni documentación existente.	Existencia de Diccionario de datos.	Diccionario de datos más metadata	Diccionario de datos más metadata, incluyendo comentarios, asunciones y sesgos.

Figura 19. Madurez analítica, otros

Una vez que se tiene identificado cuales son las estrategias que debemos seguir y los objetivos que debemos alcanzar, hay que reconocer en qué nivel de madurez analítica se encuentra la organización, complementando dos elementos, desde la perspectiva de la data y disponibilidad de la información.

2.5 Cultura

La cultura es la **mayor barrera** para la implementación de cualquiera estrategia, precisamente porque genera un cambio en la forma de cómo realizamos las cosas. Al ser la estrategia un proceso para “mover” la institución, conlleva naturalmente elementos que son diferentes, precisamente porque se busca un resultado diferente.

Adicionalmente, La forma de funcionar cambió, **las instituciones exitosas deben** evolucionar de funcionar como islas a **potenciar la capacidad analítica interna** (procesos, eficiencia, rentabilidad) **y externa** (conocimiento del cliente, competencia).

Enfoque	Anticuoado	Moderno	Cómo
Mind-set	Reactivo	Proactivo	Aprovechar la Información
Estructura	Silos/ Islas	Convergente	Centrada en el cliente
Talento	Tradicional	Enfoque en resultados	Procesos e Indicadores diseñados acorde a la estrategia
Estilo de Liderazgo	Adaptativo	Ágil y analíticamente orientado	Decisiones basadas en hechos

Tabla 2. Enfoque moderno de cultura organizacional

2.6 Mentalidad analítica

La Analítica es tanto una forma de pensar, una mentalidad. La mentalidad de estar siempre prestando atención en lugar de correr en piloto automático.

Esta mentalidad parte de investigar, tratar de entender que es lo que pasa, querer investigar, comprender y luego se trata de entender por medio de un modelo, posiblemente sea detallado o simple. Luego se evalúa, asumiendo que, si la hipótesis que se planteó, que podía resolverse mediante un modelo, se cumple.

Esta mentalidad tiene que ver con estar siempre pendiente a lo que ocurre, se debe traducir en las acciones que nos llevarían a una cultura para tomar las decisiones basadas en datos por medio de observaciones cualitativas o cuantitativos.



Investigar

En lugar de simplemente reaccionar a lo que percibe que es su entorno, recopila datos comprendiendo los límites y los límites de esos datos.



Entender

Lo introduces en un modelo. Puede ser un modelo muy detallado o un modelo simple sobre cómo evolucionan las situaciones con el tiempo si toma las opciones A, B o C, o alguna combinación de estas, y luego prueba esa hipótesis.



Evaluar

Tiene el ciclo de retroalimentación continua para decir si lo que está haciendo tiene sentido y también vigila su entorno porque lo que pudo haber tenido sentido hace un año o un mes puede que ya no lo tenga.

Tabla 3. Mentalidad analítica

Para lograr cualquier implementación hacia una cultura data-driven (o cultura basada en datos), de forma exitosa, se requiere influir en los elementos de la cultura organizacional.

La forma inicial de trabajar la estrategia es modificando los sistemas de gestión, es decir los procesos, que es el primer escalón, para luego ir facilitando la orientación a la gestión por resultados; y esta orientación es específicamente a decisiones basadas en datos. Definitivamente no es un cambio sencillo y conlleva la utilización de diferentes herramientas denominadas “Elementos para Evolucionar a una Cultura data-driven”.

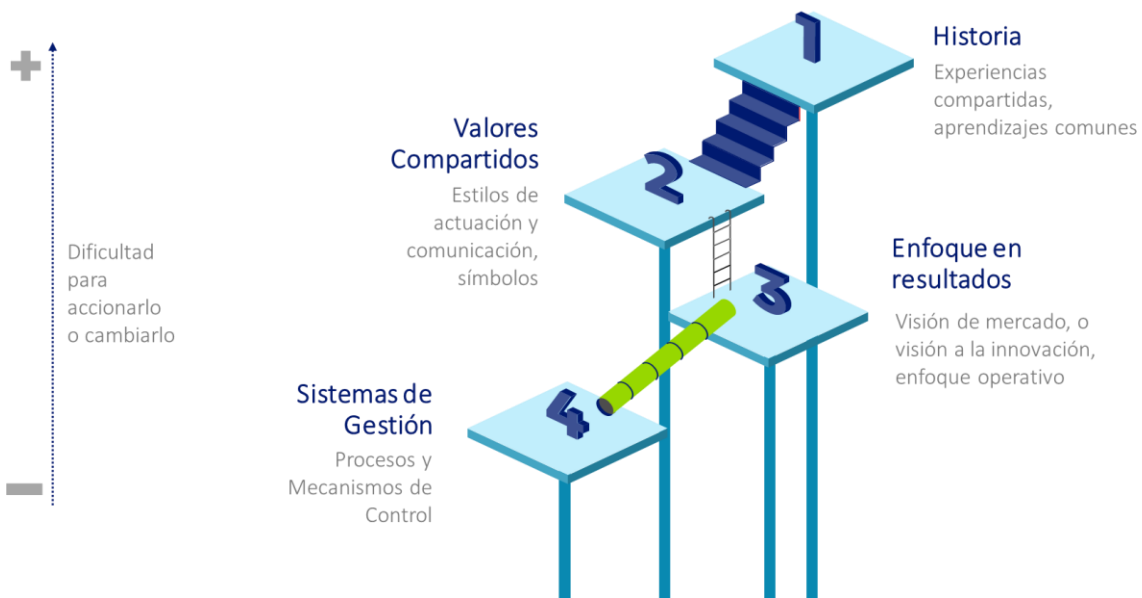


Figura 20. Sistemas que influyen la cultura

2.7 Elementos para Evolucionar a una Cultura data-driven



Figura 21. Elementos para evolucionar a una cultura data-driven

Estos seis elementos tratan de integrar que cosas debemos hacer para fortalecer la cultura enfocada en la gestión por resultados. Son herramientas que nos ayudarán a evolucionar poco a poco hacia una cultura de decisiones basada en datos.

Liderazgo

Los líderes deben cambiar el “mindset” de la organización para adoptar la data y la forma como se toman las decisiones.

- Iniciar el uso de la data activamente. Antes de emitir un criterio o al enviar un correo, indicar que indicador les hizo tomar la decisión.
- En reuniones y consultas, preguntar primero por la data, antes de tomar la decisión.
- Comunicar continuamente indicadores claves de rendimiento (KPI por sus siglas en inglés) por correos, presentaciones, reuniones.
- Tomar espacios para entrenamientos en habilidades para el manejo de la data.
- Mostrar continuamente métricas en la web interna.

Confianza

Crear una sola fuente de confianza que la compañía permite, tener acceso a ella de forma efectiva, disminuye barreras para su uso y adopción.

- Eliminar cualquier “él dice – ella dice” cuándo la data está en el mismo punto, todos miran a la misma métrica.
- Reducir el tiempo de búsqueda de data.
- Empoderar colaboradores que realmente conocen el negocio y la data, para inculcar en otros que la información es confiable.
- Expandir descubrimientos entre departamentos, es decir, compartir resultados y las razones (explicadas con datos) por la cual toman decisiones específicas.
- Reducir el trabajo de los administradores de información, colocando la data en un solo lugar. De tal forma que se reduzca el procesamiento o la necesidad de administrar múltiples fuentes de información.

Compromiso

Una cultura data-driven descansa en compromisos de largo plazo

- Asegurar que la analítica está enlazada a los esfuerzos críticos, indicando y comunicando con claridad que proyectos se ejecutan con data y que proyectos se tratan de resolver.
- Mantener la data al frente del coaching de los empleados. Cuando se conversa del desempeño, resaltar los indicadores que se deben alcanzar, las métricas que se van a buscar.
- Siempre, las metas deben estar presente para monitorear el desempeño. Mostrar los resultados públicamente, que toda la institución esté clara que es lo que se busca por los esfuerzos y el trabajo realizado.
- Invertir en tecnología que tenga la data presente, que facilite su uso e integración para generación de data y KPI's.

Métricas

Asegúrate que tu organización está bien equipada para cambiar a una cultura data-driven

- Escoger los KPIS correctos para medir el éxito.
- Revisar la información histórica para establecer metas realistas.
- Identificar constantemente nueva tecnología para acceder y analizar información.
- Asignar administradores de datos responsables de la exactitud de la data y proveer acceso a la organización.

Conocimiento de la Data

Desarrolla un plan para impulsar el conocimiento de la data

- Asegurar que los colaboradores comprenden el significado de la data.
- Asegurar que también pueden leer tablas y gráficos
- Saben cómo usar visualizaciones con data.
- Son capaces de generar conclusiones de la lectura de data.
- Son capaces de reconocer cuándo la data es mal usada o promueve sesgos.

Entrenamiento

Entrena a los colaboradores para comprender que la data es solo el comienzo.

- Transformar data en acciones es la verdadera meta.
- Que la data está disponible para tomar las decisiones que se necesitan tomar.
- Estar seguro de que el equipo entiende la data.
- Tener claridad si ¿Contratamos a un científico o lo desarrollamos?
- Tenemos las habilidades necesarias para hacer un buen “storytelling”

2.8 Roles y Responsabilidades

En los temas de analítica de datos existen diversos puntos de vista en cuanto los nombres que reciben los roles, así como a las funciones que conlleva cada uno. De forma práctica, se consideran los siguientes roles, cómo necesarios en cualquier equipo de Implementación Analítica.

- a. Líder de Implementación Analítica
- b. Especialista en Análisis de Negocio
- c. Especialista en Análisis Computacional
- d. Especialista en Análisis Estadístico
- e. Especialista en Herramientas de Visualización
- f. Administrador de Proyecto
- g. Especialista en Implementación de Estructuras de Datos

Rara vez un experto encaja en una sola categoría, pero comprender que existen diferentes funciones de ciencias de datos es relevante. Es importante tener en cuenta que incluso los profesionales con este conjunto de habilidades hipotéticas suelen tener sus puntos fuertes principales, que deben tenerse en cuenta al distribuir roles dentro de un equipo. En la mayoría de los casos, la adquisición de talentos implicará una mayor formación dependiendo de sus antecedentes.

Pero las personas y sus roles son dos cosas diferentes. Dependiendo del tamaño de la organización, el equipo asignado o la dimensión del proyecto, una persona puede combinar varios roles. Por lo tanto, para esta sección se va a ignorar cuántos expertos reales puede tener y esbozar los roles en sí. Obviamente, muchos conjuntos de habilidades entre roles se pueden combinar.

Rol	Otros nombres	Función principal
Líder de Implementación Analítica	Director de Análisis, Director de Datos, Chief Analytics Officer (CAO), Chief Data Officer (CDO)	<ul style="list-style-type: none"> ▪ Liderar la implementación Estratégica de Datos. ▪ Supervisar la implementación Táctica. ▪ “Traducir negocios”, cerrar la brecha entre la ciencia de datos y la experiencia en dominios que actúa tanto como un visionario y un líder técnico. ▪ Integrar el entendimiento del negocio con la utilización de herramientas y procesos estadísticas y computacionales, para mejorar la toma de decisiones. ▪ Habilidades preferidas: ciencia y análisis de datos, habilidades de programación, experiencia

Rol	Otros nombres	Función principal
		<p>en el dominio, liderazgo y habilidades de gestión de equipo, capacidad de negociación, visión integradora.</p>
Especialista en Análisis de Negocio	Analista de negocios, Periodistas de datos	<ul style="list-style-type: none"> ▪ Convertir las expectativas de negocio en análisis de datos. ▪ Traducir necesidades del negocio a preguntas que la analítica debe responder. ▪ Explicar al negocio la visión analítica de los problemas a solucionar. ▪ Habilidades preferidas: visualización de datos, inteligencia empresarial, SQL
Especialista en Análisis Computacional	Científico de datos con orientación Computacional, Analista de datos (Junior, Senior).	<ul style="list-style-type: none"> ▪ Limpiar, ordenar y preparar la data. ▪ Asegurar que los datos recopilados sean relevantes y exhaustivos, al mismo tiempo que interpretar los resultados de análisis. ▪ Convertir resultados analíticos en elementos visuales. ▪ Aplicar modelos estadísticos con diferentes niveles de complejidad, dependiendo de la experiencia. ▪ Aplicar modelos de procesamiento de imágenes y/o lenguaje natural, entre otros, dependiendo de la experiencia. ▪ Habilidades preferidas: R, Python, JavaScript, C/C++, SQL.
Especialista en Análisis Estadístico	Científico de datos con orientación Estadística, Ing. De Estadística.	<ul style="list-style-type: none"> ▪ Diseñar herramientas estadísticas de captura de información. ▪ Ejecutar experimentos y análisis para obtener información ▪ Aplicar modelos estadísticas para reducir errores y sesgos. ▪ Especialista en métodos estadísticos ▪ Conocimientos, R, Python, Matlab, SQL. Estadística Descriptiva, Estadística Inferencial, Modelos de Optimización, Cálculo, Álgebra Vectorial, Machine Learning, Modelos de Regresión.
Especialista en Herramientas de Visualización	Ingeniero de visualización de aplicaciones/datos	<ul style="list-style-type: none"> ▪ Desarrollar modelos interactivos de visualización. ▪ Visualizar información geoespacial. ▪ Aplicar instrumentos apropiados de visualización acorde al público objetivo. ▪ Nota: dependiendo del nivel de complejidad analítica, en ocasiones es desarrollado por Especialistas de Front-End, o programadores

Rol	Otros nombres	Función principal
		<p>WEB o con experiencia en herramientas de visualización.</p> <ul style="list-style-type: none"> Habilidades preferidas: programación, D3, JavaScript (para visualización), Tableau, POWER BI, paquetes de visualización de Python, SQL, noSQL.
Administrador de Proyectos		<ul style="list-style-type: none"> Ser responsable de manera global en un proyecto para que el inicio, la planificación, el diseño, la ejecución, la revisión, el control y el cierre de dicho proyecto sean exitosos. Tomar decisiones en tiempo y forma para que el proyecto finalice en tiempo y forma.
Especialista en Implementación de Estructuras de Datos	Arquitecto de datos, Ingeniero de datos	<ul style="list-style-type: none"> Definir la arquitectura de la base de datos, centralizar los datos y garantizar la integridad en diferentes orígenes. Construir tuberías de datos y sistemas de almacenamiento personalizados. Diseñar infraestructuras para que los datos no solo se recopilen, sino que también sean fáciles de manejar. Mantener los accesos a los datos, considerando las normas de seguridad. Nota: Para grandes sistemas distribuidos y grandes conjuntos de datos, el arquitecto también se encarga del rendimiento. Habilidades preferidas: SQL, noSQL, XML, Hive, Pig, Hadoop, Spark, Administración de Base de Datos, MongoDB, Scala, Nube.

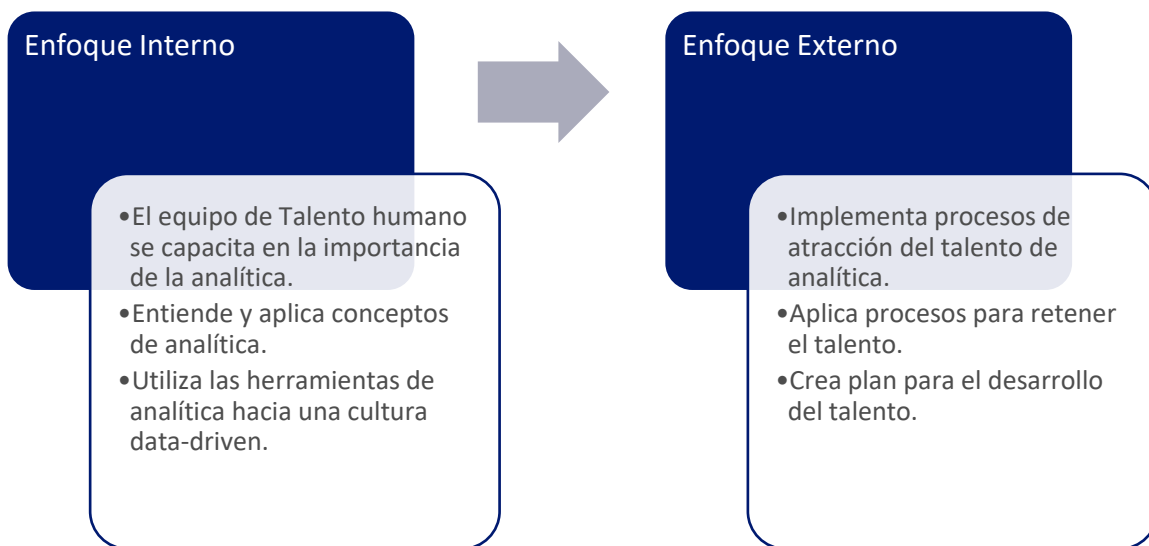
Tabla 4. Roles en la Gestión Analítica

2.9 Elementos para Implementar una estructura de analítica de datos

Los elementos de transformación digital, así como la evolución de organizaciones más enfocadas en la innovación y en el desarrollo de conceptos de analítica de datos, influye en los elementos del manejo del talento humano.

En cualquier tipo de organización el enfoque va alineado al retorno de la inversión, un retorno económico en entidades privadas o el retorno medido en impacto obtenido de acuerdo a los objetivos y metas trazadas, para entidades públicas y sin ánimo de lucro.

Para facilitar la implementación del proceso de analítica en las instituciones gubernamentales, se sugieren dos mecanismos de implementación:



2.9.1 Enfoque interno

Al hablar de cultura data-driven, un elemento importante, es que el equipo de talento humano adopte, comprenda e implemente el concepto cultural, con lo cual se recomiendan las siguientes acciones:

- El equipo de Talento es parte del proceso: Durante la implementación del proceso de analítica, el equipo de Recursos Humanos es parte del proyecto, aportando ideas para su implementación y ejecutando tareas en conjunto para facilitar el éxito del proyecto, así como el cambio cultural.
- El equipo de Talento debe estar capacitado y sensibilizado en la Analítica de datos: De tal forma que tenga un claro entendimiento de lo que se busca y lo que involucra,

de tal forma que tenga una visión amplia de los perfiles requeridos, así como el desarrollo interno que debe recibir el equipo.

- El equipo de Talento aplica la analítica: Lo ideal en etapas más avanzadas del proceso, es la implementación de indicadores en las áreas de Talento Humano o Recursos Humanos, de tal forma que vayan también evolucionando a una cultura data-driven. Indicadores tradicionales como rotación de personal, tiempos de reclutamiento pueden ser implementados de forma sencilla y con revisión periódica para facilitar el asentamiento de la cultura data-driven.
- El equipo de Talento debe recibir apoyo técnico: Se debe recibir constante respaldo y retroalimentación de los líderes implementadores del proceso analítico, desde los perfiles de puesto que se buscan, las descripciones de puestos, los planes a futuro para ir desarrollando las destrezas en el equipo, en conjunto con los planes de capacitación.

2.9.2 Enfoque Externo – Atracción del talento

En todas las organizaciones existen procesos para atraer el talento de acuerdo con el perfil que se requiere, las posiciones en analítica de datos son de alta demanda, con lo cual las herramientas que desarrollamos para encontrar e incluir en la organización profesionales con perfiles de analítica las describimos a continuación.

Partimos de la premisa, que los perfiles de analítica por su formación les gustan los elementos modernos (juegos de video, gadgets útiles, conocer temas diferentes), principalmente orientados a la tecnología, les gusta la estructura, investigar y tienen relativa facilidad para aprender cosas nuevas, no necesariamente son buenos comunicadores (es parte de los elementos a desarrollar), deben tener interés por realizar actividades que tengan impacto social.

- Crear el “Branding” o imagen de analítica: Al ser un perfil competitivo en el mercado laboral, se requieren crear una imagen de la organización en los medios o redes sociales, de tal forma que haya cierto conocimiento de lo que se realiza. Publicar información sobre los proyectos, retos y desafíos que tiene el proceso de analítica para crear el interés en formar parte del equipo.
- Mantener una imagen acorde: Al contactar posibles candidatos, el proceso de envío de información sea lo más “digital” posible, que se mantenga esa imagen de innovación. Los formularios, pruebas iniciales y cualquier otro mecanismo con una imagen de alto profesionalismo y herramientas tecnológicas apropiadas.
- Proceso de Selección basado en pruebas y casos: En los casos que el candidato cumpla los requerimientos básicos, es preferible que se realice un filtro a través de una prueba o un caso, en donde se aplique y evalúe el conocimiento requerido. En la práctica, es más importante la capacidad de saber dónde y cómo buscar

información para resolver un problema que conocer la respuesta de memoria (a menos que sea algo muy específico). La siguiente prueba debe ser en las instalaciones de la entidad contratante, dónde se de otro caso o prueba que involucre la facilidad de aplicar el conocimiento a la realidad.

- Oferta acorde al mercado laboral: Se debe participar de las encuestas salariales, de tal forma que se tenga una propuesta salarial dentro escalas apropiadas para incluir en la organización talento de alto nivel, con interés de desarrollarse en la organización en el largo plazo.
- Oferta de permanencia laboral: En este tipo de posiciones se compite con entidades privadas (nacionales o multinacionales) en la cual se ofrecen posiciones de carácter indefinido con lo cual, para atraer talento de alto nivel, se requiere que dentro de la estructura de analítica contar con condiciones similares a las del mercado laboral para ser competitivos.

2.9.3 Enfoque Externo – Retención del Talento

En el sector tecnológico la rotación es de las más altas de la industria, con lo cual se deben identificar mecanismos para retener el talento que tiene buen potencial y desempeño en la organización por el mayor tiempo posible.

- Ajustes en la cultura y metodologías de trabajo: Las nuevas generaciones valoran la posibilidad de poder hacer las cosas a su manera, con eficacia y eficiencia. El mercado laboral ha cambiado, y es común encontrar horarios laborales que combinan la flexibilidad y la metodología fija, para que el colaborador de el mejor esfuerzo posible.
- Incluir elementos de comunicación y buen clima: Crear ambientes laborales que faciliten la comunicación entre los miembros del equipo, al igual que permitir dar la opinión por igual a todos los colaboradores, con respeto y bien enfocadas, crea la posibilidad de escuchar más ideas, y que los colaboradores se sientan parte de la organización.
- Vender la misión: Definitivamente no somos Apple o Google, con lo cual se deben crear mecanismos de atracción y retención del talento en el tiempo. Con lo cual nuestra gran misión que es generar valor en la ciudadanía debe ser una gran palanca para conservar en el equipo personas con gran interés generar de impacto social positivo.

2.9.4 Enfoque Externo – Desarrollo del Talento

Este es un sector que todavía está en crecimiento, con múltiples posibles líneas de especialización, es importante estar constantemente fortaleciendo las capacidades del equipo.

- Definición de planes de carrera claros: Establecer un plan de desarrollo, dónde haya una serie de módulos o conocimientos que se deben adquirir. Anualmente los que están faltantes de algún tema deben ir avanzando en el plan de tal forma que vayan incrementando sus competencias en analítica de datos.
- Opción en capacitaciones innovadoras locales o extranjeras: Participar en cursos, seminarios o diplomados, con la intención de que se replique ese conocimiento con los demás miembros del equipo, y en caso de superar un monto definido, establecer acuerdos de permanecer en la organización por un tiempo mínimo.
- Competencia libre de proyectos: Realizar concursos entre diferentes instituciones para que los colaboradores, puedan investigar, proponer, desarrollar e implementar proyectos de analítica de alto impacto.

Capítulo 3. Marco de Trabajo para el Análisis de Datos

3.1 Propósito del marco metodológico para el análisis de datos

Este marco tiene como propósito establecer un modelo de proceso comprensible para la función de análisis de datos; estableciendo las actividades, roles y entregables principales que permitan la creación de una cultura de toma de decisiones basados en datos y alineados a las líneas estratégicas de la institución.

3.2 Objetivos del marco

Este Marco tiene como objetivos:

- Proporcionar un lenguaje y metodología común para las funciones de análisis de datos en la AIG y que permita ser un modelo de recomendación para las demás instituciones de gobierno.
- Establecer los diferentes roles y responsabilidades relacionados al análisis de datos.
- Aclarar el alcance y los límites de las funciones de análisis de datos.
- Proporcionar guía, métodos y técnicas que faciliten las funciones de análisis de datos.

3.3 El público al que está dirigido

Este marco está dirigido a:

- Profesionales responsables del análisis de datos o personal de tecnología que mantenga relación con estas funciones.
- Proveedores u organismos externos que presten servicios a la AIG con relación al análisis de datos.
- Directivos o personal de alto perfil que mantengan interés o estén relacionados a la generación de conocimiento como activo gubernamental.

3.4 Estructura del marco

En este documento pretende explicar de forma sencilla y práctica como ejecutar las actividades de un proyecto de análisis de datos, por lo que se ha estructurado en capítulos de la siguiente manera:

- **Capítulo 1. Estado del Arte:** el cual se resume el resultado de la investigación realizada para el entendimiento de la situación actual y avances en relación con la minería de datos en el sector gubernamental. Este capítulo contiene las metodologías más populares, casos similares y el avance a nivel nacional.
- **Capítulo 2. Conceptos Generales de Análisis de Datos.** En este capítulo se presentan conceptos generales a conocer para un mejor entendimiento del análisis de datos, principalmente desde el enfoque estratégico y cultural.
- **Capítulo 3. Marco de Trabajo para el Análisis de Datos.** En este capítulo se realizar una presentación general del marco, su objetivo, propósito y estructura de proceso. En los siguientes capítulos se irá desarrollando de manera más detallada cada una de las actividades planteadas en el marco y sus conceptos relacionados.
- **Capítulo 4. Planificación y alineación con Estrategia, Capítulo 5. Ejecución del Ciclo de Descubrimiento, Capítulo 6. Comunicación de Resultados y Capítulo 7- Implementación de Acciones.** Se explican de manera detallada las actividades a realizar, sus entradas, salidas y roles responsables. Además, se incorporan en cada capítulo los conceptos relacionados de valor para el entendimiento de las actividades.
- **Anexos.** Se anexa ejemplos y plantillas que pueden ser utilizadas para la ejecución de las actividades planteadas en el marco de trabajo.

3.5 Definiciones

- Estrategia: es un plan que especifica un conjunto de pasos que tienen como objetivo la implementación de un determinado objetivo. Requiere de seleccionar y priorizar objetivos, definir un plan y dar seguimiento a su ejecución. (que significado)
- Estrategia de datos: es un plan de ruta que ayuda a definir qué hacer con los datos de una compañía o entidad. Los datos deben ser controlados en base a una estrategia ya que esto facilita su acceso y manejo. (ARIMETRICS)
- Ingeniero de datos: es el encargado de brindarle la información que se utilizará, mantienen los accesos a los datos, diseñan infraestructuras para que los datos no solo se recopilen, sino que también sean sencillas. (Discover the new)
- Analista de datos: es aquel que por medio de la interpretación de los datos puede establecer estrategias dentro de la empresa, con la ayuda de gran cantidad de datos. (Unir Reviste)
- Científico de datos: se dedica a analizar e interpretar las bases de datos de una empresa. (Herrington)
- Datos geoespaciales: se relaciona con el conjunto de datos temáticas de información geográfica, que se usan para optimizar la mayoría de las aplicaciones, para mejorar la interoperabilidad e incrementar el uso de los datos territoriales. (IDECA)
- Almacenamiento de datos: uso de medios de grabación para conservar los datos utilizando PC. Los sistemas de almacenamiento son de gran ayuda a cualquier empresa para planear sus estrategias en los procesos de negocio. (KIO Networks)

3.6 Proceso

Para la definición del marco de trabajo planteado se ha utilizado los aportes realizados por diferentes modelos metodológicos existentes, siendo los más influyente en este marco el modelo planteado por la metodología CRISP-DM (CRoss-Industry Standard Process for Data Mining) y KDD (Knowledge Discovery in Databases).

Adicional, se ha planteado en formato proceso con la visión que estas actividades pasen de ser manejadas como proyectos esporádicos a ser una gestión recurrente sobre una cultura data-driven.

Al ser un marco metodológico, funciona como una guía de procesos adaptativa que puede ser aplicada en diferentes contextos o estudios. No necesariamente tiene una explicación para todas las posibles situaciones técnicas o la aplicación de todos los posibles modelos estadísticos. Sin embargo, integra las diversas líneas de conocimiento, estratégico, funcional, estadístico y computacional, de tal forma que personas con conocimiento en cualquiera de estas áreas funcionales pueden poner en práctica la metodología de análisis de datos en forma conjunta. A la vez, que facilita la implementación del proceso de forma organizada y estructurada.

3.6.1 Alcance

Este proceso tiene como alcance la ejecución de las actividades desde la planeación y alineación estratégica hasta la implementación de las acciones. Detalla las actividades y roles necesarios para la ejecución entre diferentes instituciones, siendo la AIG la institución rectora, encargada de generar el conocimiento y asegurar se cumplan las actividades definidas.

3.6.2 Dueño

El dueño del proceso debe tener altas cualidades de liderazgo, administración de las comunicaciones, cambio y facilitación de grupos; además de un alto conocimiento funcional sobre los temas relacionados a su proceso. Es el responsable final del diseño, buen funcionamiento, documentación y divulgación de este con todos los miembros del equipo e interesados.

Se plantea como dueño de este proceso o marco de trabajo a la AIG, posición Coordinador de Planes y Programas de la Dirección de Innovación Gubernamental.

3.6.3 Entradas

Se entiende por entradas cualquier material, servicio o información que necesita el proceso para ejecutar las actividades definidas y generar una salida. Las entradas son los inputs del proceso. El proceso recibe las entradas y las transforma en salidas o outputs.

Las entradas definidas para este proceso son:

- Estrategia Gubernamental
- Estrategia de la Institución Rectora
- Estrategia de la AIG

3.6.4 Diagrama del proceso

Este diagrama nos permite tener una visualización macro de los subprocesos a ejecutar, siendo los subprocesos una agrupación de actividades. Luego para cada uno de los subprocesos se detallarán las actividades y pasos.



Figura 22. Proceso de Analítica de Datos

3.6.5 Salidas

Se entiende por salida a el resultado final del proceso. Son los productos, servicios o información producidos para el cliente.

Para este proceso se definen como salidas:

- Acciones implementadas en base a los datos analizados
- Conocimiento generado
- Reportes automatizados

Capítulo 4. Planificación y Alineación con Estrategia

La estrategia es importante porque nos ayuda alcanzar los objetivos principales, de manera eficiente, ahorrando tiempo y recursos. Es el proceso mediante el cual organizamos nuestros recursos y dirigimos nuestros esfuerzos para alcanzar las metas prioritarias.

4.1 Actividades



Figura 23. Subproceso de Planificación y Alineación con la Estrategia

4.1.1 Actividad 1: Sensibilización y entendimiento de metas estratégicas

Definimos 2 grupos: Equipo Implementador de gestión analítica (quién guiará en el proyecto, y en la mayoría de los casos será el equipo AIG) llamaremos equipo implementador o simplemente “el implementador”. Institución dónde se implementará el proceso de Implementación analítica que llamaremos Institución receptora o simplemente “la receptora”.

Responsable: Líder de la Implementación Analítica

ENTRADAS:			
<ul style="list-style-type: none"> Alcance esperado del proyecto de gestión analítica 			
EJECUTOR	Pasos	TIPO	SLA
Líder de la Implementación Analítica	1. Verifique alcance esperado del proyecto, e identifique las actividades claves requeridas para su ejecución.	Actividad Manual	n/a
Líder de la Implementación Analítica	2. Defina cronograma propuesto de implementación de gestión analítica. Ver ejemplo de cronograma en Anexo.	Actividad Manual	n/a
Líder de la Implementación Analítica	3. Envíe a Líder de la Institución receptora el cronograma de implementación propuesto.	Actividad Manual	n/a

Líder de la Implementación Analítica	<p>4. Solicite conformación de equipo por parte de la Institución receptora, con al menos los siguientes roles:</p> <ul style="list-style-type: none"> ▪ Patrocinador: Preferiblemente el despacho superior o en su defecto una posición relevante dentro de la organización. ▪ Líder Técnico: director o responsable principal del departamento de Tecnología o en su defecto una posición relevante dentro del área de tecnología con conocimiento de los repositorios de datos y estructura. ▪ Líder de Implementación: Persona con conocimiento de proyectos que funcionará como Administrador de proyectos de la Institución receptora. ▪ Otros: De acuerdo con el alcance se debe solicitar la inclusión de expertos de áreas funcionales, así como de programadores o administradores de base de datos. 	Actividad Manual	n/a
Líder de la Implementación Analítica	<p>5. Solicite a Líder de Institución receptora una reunión para definir: alcance, posibles fechas de reuniones con personal de alto nivel, enfoque estratégico, realización de reunión de sensibilización al equipo de la Institución receptora, revisar cronograma, identificar equipo del proyecto, listado de conjunto de datos existentes en la Institución y fechas de inicio.</p>	Actividad Manual	n/a
	<p>6. Elabore Plan de Proyecto en base a la reunión sostenida. <i>Ver formato en Anexo.</i></p>	Actividad Manual	n/a
Líder de Institución Receptora	<p>7. Revise Plan de proyecto. Importante que valide y confirme la fecha de inicio y equipo responsable de administrar el proyecto para la institución receptora.</p>	Actividad Manual	n/a
Líder de la Implementación Analítica	<p>8. Realice reunión de inicio de proyecto con el equipo, presentando lo detallado en el plan de proyecto.</p>	Actividad Manual	n/a
Líder de la Implementación Analítica	<p>9. Realice las Entrevistas de acuerdo con el formato de preguntas (ver formato en Anexos), con cada personal de alto nivel de la institución receptora definido, tratando de identificar lo siguiente:</p> <ul style="list-style-type: none"> ▪ Estrategia global de la institución ▪ Qué se espera se resuelva con datos ▪Cuál es la visión del uso de los datos. 	Actividad Manual	n/a
	<p>10. Realice documento resumen de entrevistas, detallando la información más relevante para la construcción de las entrevistas. En no más de 3 páginas en Microsoft Word o similar.</p>	Actividad Manual	n/a

Líder de la Implementación Analítica	11. Realice reunión de sensibilización al equipo del proyecto en la Institución receptora, que contenga o se comparta al menos la siguiente información <ul style="list-style-type: none"> Elementos de una estrategia de datos Importancia de la estrategia de datos Próximos pasos a seguir para implementar la Estrategia de uso de la información en la institución receptora. 	Actividad Manual	n/a
Líder de la Implementación Analítica	12. Prepare y proponga los lineamientos estratégicos para el uso de la data en la institución receptora, colocándolo en el formulario de resumen de lineamientos del uso estratégico de la data.	Actividad Manual	n/a
Líder de Institución receptora	13. Apruebe lineamientos estratégicos para el uso de la data en la institución receptora, que es la base para la construcción del plan uso de la data.	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> Cronograma de Implementación Analítica Resumen de Entrevistas Formulario resumen de lineamientos estratégicos Listado de conjunto de datos existentes en institución receptora 			

Tabla 5. Actividad 1. Sensibilización y entendimiento de metas estratégicas

4.1.2 Actividad 2: Desarrollo de Estrategia

Responsable: Líder de la Implementación Analítica

ENTRADAS:			
<ul style="list-style-type: none"> Cronograma de Implementación Analítica Resumen de Entrevistas Formulario resumen de lineamientos estratégicos Listado de conjunto de datos existentes en institución receptora 			
EJECUTOR	Pasos	TIPO	SLA
Líder de la Implementación Analítica	1. Defina los Objetivos: Del formulario resumen de lineamientos estratégicos, identifica en conjunto con el equipo del proyecto cuáles son los objetivos que se va a buscar con la implementación de la gestión. Deben ser Objetivos Cuantitativos, de lo que se requiere lograr. En primeras iteraciones puede ser hasta la utilización de una cantidad específica de conjunto de datos.	Actividad Manual	n/a

<p>Líder de la Implementación Analítica</p>	<p>2. Defina los casos de Uso</p> <ul style="list-style-type: none"> ▪ La situación ideal es que se parta de los objetivos y se identifiquen los casos de uso, en función de las necesidades de la organización o tratando de solucionar un problema particular. ▪ La otra vía es que se tenga data recolectada y se parta el análisis en función de lo que se tiene, que puede aumentar el riesgo de no vinculación con la estrategia. Con lo cual se debe clarificar el problema que se solucionará. <p>Puede haber 2 fuentes posibles:</p> <table border="1" data-bbox="532 636 1105 850"> <thead> <tr> <th>Si el caso de uso parte de ...</th> <th>Entonces...</th> </tr> </thead> <tbody> <tr> <td>Tratar de Solucionar un problema</td> <td>Continúe con el paso 3.</td> </tr> <tr> <td>Utilizar un conjunto de datos existentes</td> <td>Continúe con el paso 4.</td> </tr> </tbody> </table>	Si el caso de uso parte de ...	Entonces...	Tratar de Solucionar un problema	Continúe con el paso 3.	Utilizar un conjunto de datos existentes	Continúe con el paso 4.	<p>Actividad Manual</p>	<p>n/a</p>
Si el caso de uso parte de ...	Entonces...								
Tratar de Solucionar un problema	Continúe con el paso 3.								
Utilizar un conjunto de datos existentes	Continúe con el paso 4.								
<p>Líder de la Implementación Analítica</p>	<p>3. Complete el formulario de Caso de Uso. En conjunto con el equipo del proyecto, se identifica una línea estratégica que se quiera solucionar, y en base a ello se describe los elementos cómo KPI's , Datos requeridos, Tecnologías, competencias, entre otros elementos para desarrollar el caso de uso.</p>	<p>Actividad Manual</p>	<p>n/a</p>						
<p>Equipo de proyecto de Institución receptora</p>	<p>4. Caracterice conjuntos de datos existente: Se debe completar el <i>formulario descripción de conjunto de datos</i>, con el fin de conocer la información contenida, así como posibilidades de explotación y relación con los objetivos estratégicos.</p>	<p>Actividad Manual</p>	<p>n/a</p>						
<p>Líder de la Implementación Analítica</p>	<p>5. Complete matriz de priorización de datos: Enumerar en la matriz de priorización los casos de uso y los conjuntos de datos.</p>	<p>Actividad Manual</p>	<p>n/a</p>						
<p>Equipo de proyecto de Institución receptora</p>	<p>6. En conjunto con el Equipo de Implementación Analítica, pondere cada iniciativa (casos de uso y conjunto de datos) en función a los elementos de priorización, en el siguiente orden:</p> <ul style="list-style-type: none"> ▪ Lo primero es revisar las ponderaciones asignadas por categoría (estratégica, técnica, implementación). ▪ Luego se valora cada iniciativa desde una perspectiva estratégica, solamente. ▪ Las que no tengan impacto estratégico se deben descartar. Considerar que algunas iniciativas pueden ser requeridas para otras iniciativas, con 	<p>Actividad Manual</p>	<p>n/a</p>						

	<p>lo cual en la parte estratégica se debe tomar en cuenta el encadenamiento.</p> <ul style="list-style-type: none"> ▪ La evaluación técnica y de implementación se hace en conjunto. ▪ Al final se ordenan en función de las que tienen mayor puntuación. Luego se incluyen 2 elementos, si hay iniciativas dependientes y el tiempo de implementación. ▪ De esta forma se genera nuestra matriz priorizada. 		
Equipo de proyecto de Institución receptora	<p>7. En conjunto con el equipo de implementación analítica, seleccione las iniciativas prioritarias.</p> <ul style="list-style-type: none"> ▪ Con el listado organizado por el valor resultante de la ponderación, se seleccionan las 3 o 4 iniciativas con mayor ponderación, catalogándolas como prioritarias. ▪ Las iniciativas prioritarias se agrupan con aquellas iniciativas que requieren construcción previa. (Es decir las iniciativas bases de las cuales las iniciativas prioritarias son dependientes). Creado las agrupaciones prioritarias. ▪ A las agrupaciones prioritarias son las que se estimarán esfuerzos y recursos, requeridos para su implementación. 	Actividad Manual	n/a
Equipo de proyecto de Institución receptora	<p>8. En conjunto con el equipo de implementación analítica, defina los requerimientos de data. Iniciando por responder las siguientes preguntas:</p> <ul style="list-style-type: none"> ▪ Quién es tu audiencia objetivo ▪ Cuáles son tus objetivos de negocio ▪ Cómo está la data. 	Actividad Manual	n/a
Equipo de proyecto de Institución receptora	<p>9. En conjunto con el equipo de implementación analítica, entender los datos a ser analizados. Se completa el catálogo de datos (sino se ha completado).</p> <p>Entender cómo es generada y cómo es usada.</p> <ul style="list-style-type: none"> ▪ Por Fuente ▪ Entidad ▪ Dueño de la data 	Actividad Manual	n/a
Líder de la Implementación Analítica	10. Identificar brechas de data o información existente.	Actividad Manual	n/a
Líder de la Implementación Analítica	11. Identificar brechas de software, hardware o arquitectura tecnológica para implementar las soluciones y procesos analíticos.	Actividad Manual	n/a

Líder de la Implementación Analítica	12. En conjunto con el equipo de la institución receptora, se forma más apropiada de organizar los recursos para cerrar las brechas de tecnológicas y de data. Una opción puede ser agrupar los esfuerzos en cargas de trabajo por tipo de solución analítica o caso de uso, que es más recomendable para primeras iteraciones. Otro mecanismo es definir cargas de trabajo por proceso, por ejemplo, todos los procesos de captura se consideran una carga de trabajo, todos los procesos de limpieza pueden ser una carga de trabajo. También se pueden definir cargas de trabajo por tipo de funcionalidad, como por ejemplo cargas de trabajo de aprendizaje supervisado.	Actividad Manual	n/a
Líder de la Implementación Analítica	13. En conjunto con el equipo de proyecto de la institución receptora, ordenar las cargas de trabajo por horizonte de tiempo, definiendo un cronograma de implementación con fechas y responsables. Preferible planificar para 3 o 4 horizontes de tiempo, de 2-3 meses cada uno.	Actividad Manual	n/a
Líder de la Implementación Analítica	14. Colocar los horizontes de tiempo en un plan de trabajo, con su requerimiento de recursos y responsables. Este documento se presenta con el Plan de Trabajo de Implementación Analítica.	Actividad Manual	n/a
Líder de Institución receptora	15. Aprobar Plan de Trabajo de Implementación Analítica.	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> ▪ Plan de Trabajo de Implementación Analítica ▪ Formulario de Requerimientos de Data ▪ Formulario de Creación de casos de uso 			

Tabla 6. Actividad 2, Líder de implementación analítica

4.2 Conceptos relacionados

4.2.1 Definición de Casos de Uso de la Organización

En el marco de proyectos vinculados a la Ciencia de Datos, los casos de uso son una descripción de las actividades necesarias, para generar conocimiento a través de la analítica.

Un caso de uso debe ser lo suficientemente sólido como para brindar un nivel claro de comprensión del objetivo que se está tratando de lograr, por lo que será requisito la asociación del caso de uso a uno o más objetivos.

Para este fin es necesario el uso de técnicas que permitan obtener la información suficiente para definir los casos de uso. Son altamente recomendadas las siguientes:

1. Entrevista guiada.
2. Observación.
3. Revisión documental y fuentes de información.
4. Mesas de Trabajo (Workshop)
5. Tormenta de ideas.

En el ámbito gubernamental, dado el contexto del proyecto y los perfiles de los integrantes que conforman el equipo, se ha seleccionado la entrevista guiada como la técnica más idónea para obtener la información necesaria a fin de modelar los casos de usos.

El proceso de entrevista debe incluir parte del equipo de proyecto, tomando en consideración los roles dentro del proyecto.

El proceso de entrevista es llevado a cabo de manera individual, guiado por un cuestionario de preguntas abiertas, el cual da apertura al intercambio de información e ideas, que permiten también conocer las expectativas en relación con el resultado esperado del proyecto.

Una vez obtenida la información pertinente, los casos de uso pueden ser construidos, con base en el enfoque de cargas de trabajo estratégicas, es decir la suma de cargas de trabajo en función del proceso donde varias cargas de trabajo son comunes dado que comparten esfuerzos y experiencias.



Figura 24. Definición de Casos de Uso

Los casos de uso deben ser documentados en lenguaje natural, en un formato legible para el equipo de proyecto. Tal documento deberá dar cobertura a los siguientes puntos:

1. Objetivo general
2. Objetivos específicos
3. Fuentes (personas, documentos, entre otros)
4. Actores del flujo
5. Descripción del flujo básico
6. Condiciones iniciales
7. Condiciones esperadas una vez se implemente el caso de uso
8. Oportunidades y desafíos

4.2.2 Priorización de Principales casos de Uso

Una vez definidos y documentados los casos de uso, deben ser priorizados para su ejecución y desarrollo.

Una opción es trabajar cada caso de uso cómo un proyecto; otra forma, es en función de las cargas de trabajo que tiene cada caso de uso.

Una herramienta recomendada para llevar a cabo la priorización es la matriz de **Horizontes de Desarrollo**, que permite combinar la prioridad de los objetivos con la carga técnica de trabajo, dando como resultado la determinación de casos de uso priorizados para su ejecución.

La priorización de casos de uso puede involucrar varias iteraciones, hasta alcanzar una prioridad coherente, considerando además de la carga de trabajo técnico, otros criterios como:

1. Contribución a la estrategia
2. Dependencias entre casos de usos para su implementación.
3. Riesgos asociados a la implementación del caso de uso.

Con las prioridades definidas, es tarea estructurar un **roadmap** para abordar el desarrollo del proyecto, es decir de los horizontes de desarrollo definidos previamente, en una línea de tiempo prevista.

4.2.3 Cargas de Trabajo

La mejor manera de agrupar las cargas de trabajo es por medio de procesos, como se ve en el cuadro puede haber cargas de trabajo de captura de información, todas las cargas de trabajo de transacciones desde el móvil puede ser una carga de trabajo de captura de información, puede haber cargas de captura de información desde la actividad web. Se puede tener diferentes casos de uso, pero se van agrupando por procesos o dentro del proceso un área de conocimiento; depende de la estructura que se tiene y de las personas que son parte de esta.

1. Es más eficiente agrupar en función del proceso.
2. Se comparten esfuerzos y se genera mayor experiencia.
3. Los equipos se enfocan en el desarrollo específico.



Figura 25. Agrupación de cargas de trabajo

Capítulo 5. Ejecución del Ciclo de Descubrimiento

5.1 Actividades



Figura 26. Subproceso Ejecución del Ciclo de Descubrimiento

5.1.1 Actividad 1 - Definición del Estudio o problema a solucionar

ENTRADAS:			
<ul style="list-style-type: none"> ▪ Documento de Caso de Uso ▪ Resumen de información estratégica ▪ Matriz de priorización de iniciativas 			
EJECUTOR	Pasos	TIPO	SLA
Especialista en Análisis de Negocio	1. Revise en conjunto con el Especialista en Análisis Estadístico y el Especialista en Análisis Computacional, el documento de Caso de Uso.	Actividad Manual	n/a
Especialista en Análisis de Negocio	2. Proponga las preguntas que deben ser contestadas para dar solución al caso de negocio.	Actividad Manual	n/a
Especialista en Análisis Estadístico	3. Realice el “Análisis Exploratorio” en dónde identifica, características de la data. Con el siguiente enfoque: <ul style="list-style-type: none"> ▪ Entienda los atributos (cantidades), tipos (numéricos, categóricos, continuos, discontinuos, etc.). ▪ Tamaño de la muestra. ▪ Identifique valores faltantes, nulos, o en blancos. Lo recomendable es un máximo de 5% de datos faltantes, o en casos extremos hasta 10%. En cualquier caso, dentro del plan se debe sugerir el plan a seguir para manejar los datos en blanco. ▪ Visualice medias, cuartiles, máximos, mínimos. ▪ Identifique valores atípicos. En casos de duda se debe considerar datos atípicos, aquellos inferiores a $Q1 - 1.5 * RIC$, o superiores a $Q3 + 1.5 * RIC$. Entendiendo Q1 y Q3 como primer y tercer cuartil, respectivamente y RIC cómo el rango Inter-cuartil, que no es más que la diferencia de $Q3 - Q1$. ▪ Revise la fuente de Captura y posibles sesgos. Entre los posibles sesgos, Muestras no aleatorias o no 	Actividad Manual	n/a

	<p>estratificadas acorde a la población, respuestas voluntarias, entre otros.</p> <ul style="list-style-type: none"> ▪ Clasifique la variable de “salida”, independiente o respuesta. ▪ Escriba conclusiones sobre la información encontrada. ▪ Durante el proceso, se pueden correr múltiples análisis exploratorios, a medida que se va realizando el proceso de preparación o limpieza. 		
Especialista en Análisis Estadístico	4. Compare las preguntas de negocio, con la data a analizar, de tal forma que define cómo responder a las preguntas de negocio.	Actividad Manual	n/a
Especialista en Análisis Estadístico	<p>5. Evalúe en conjunto con el Especialista en Análisis Computacional, el camino a seguir para la ejecución del proceso de Análisis de datos, y lo documentarán en el Plan de Análisis de Datos, que conllevará fechas, responsables y herramientas a utilizar. qué usualmente llevarán los siguientes elementos, muchas veces iterativos:</p> <ul style="list-style-type: none"> ▪ Organizar los datos acordes a su tipo (numéricos, categóricos, textos, fechas). ▪ Identificación y manejo de valores atípicos. ▪ Identificación y manejo de campos vacíos. ▪ Escalamiento y/o estandarización de datos. ▪ Separación de data en grupos de entrenamiento, validación y prueba. ▪ Selección de modelo analítico a ser aplicado. ▪ Revisión de resultados. ▪ Verificación de errores. 	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> ▪ Plan de Trabajo de Análisis 			

Tabla 7. Actividad 1, Ejecución del ciclo de Descubrimiento

5.1.2 Actividad 2 – Preparación de Data

ENTRADAS:			
<ul style="list-style-type: none"> ▪ Base de datos o conjunto de datos 			
EJECUTOR	Pasos	TIPO	SLA
Especialista de Análisis Computacional	1. Exporte la base o el conjunto de datos al sistema de preferencia (R o Python preferiblemente).	Actividad Manual	n/a
Especialista de Análisis Computacional	2. Revise cada atributo y verifica si contiene el tipo correcto (entero, número real, fecha, categórico, etc.) dependiendo del software que está utilizando.	Actividad Manual	n/a

Especialista de Análisis Computacional	3. Ejecute el proceso de “tidyng” u ordenamiento de la data, consiste en que cada registro tenga su propia fila y cada atributo tenga su propia columna. Ver Ejemplo	Actividad Manual	n/a
Especialista de Análisis Computacional	4. Verifique cada atributo o variable, para identificar si no tienen diferencias de tipografía que afecte el proceso. Por ejemplo, Sí, Si, si, sl; que pudiese reconocerlo como valores diferentes.	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> Base o conjunto de datos ordenado. 			

Tabla 8. Actividad 2, Ejecución del ciclo de Descubrimiento

Ejemplo de Data ordenada

tabla1 - Ordenado

#> # A tibble: 6 x 4

```
#> country      year  cases  population
#> <chr>        <int> <int>    <int>
#> 1 Afghanistan 1999    745    19987071
#> 2 Afghanistan 2000   2666   20595360
#> 3 Brazil       1999  37737  172006362
#> 4 Brazil       2000  80488  174504898
#> 5 China        1999 212258 1272915272
#> 6 China        2000 213766 1280428583
```

tabla2- No ordenado

#> # A tibble: 12 x 4

```
#> country      year  type      count
#> <chr>        <int> <chr>    <int>
#> 1 Afghanistan 1999  cases      745
#> 2 Afghanistan 1999  population 19987071
#> 3 Afghanistan 2000  cases      2666
#> 4 Afghanistan 2000  population 20595360
#> 5 Brazil       1999  cases      37737
#> 6 Brazil       1999  population 172006362
```

En tabla2 – No ordenado, ver la variable Afganistán- 1999 tiene 2 filas, “Cases” y “Population” y en la tabla ordenada, ambos se convierten en columnas.

5.1.3 Actividad 3 – Búsqueda y manejo de valores atípicos

ENTRADAS:			
<ul style="list-style-type: none"> Base de datos o conjunto de datos ordenados. 			
EJECUTOR	Pasos	TIPO	SLA
Especialista de Análisis Computacional	1. Corra nuevamente el análisis exploratorio para cada variable cuantitativa.	Actividad Manual	n/a
Especialista de Análisis Computacional	2. Corrobore la existencia de valores atípicos. <ul style="list-style-type: none"> En casos de duda se debe considerar datos atípicos, aquellos inferiores a $Q1 - 1.5 * RIC$, o superiores a $Q3 + 1.5 * RIC$. Entendiendo $Q1$ y $Q3$ como primer y tercer cuartil, respectivamente y RIC cómo el rango Inter-cuartil, que no es más que la diferencia de $Q3 - Q1$. Otro mecanismo es incluyendo en el análisis exploratorio, gráficas como “boxplot” (diagrama de caja) o histogramas. 	Actividad Manual	n/a
Especialista de Análisis Computacional	3. Investigue con el Especialista de Análisis de Negocio, las causas de valores atípicos. Por ejemplo, puede haber valores atípicos debido a: <ul style="list-style-type: none"> Ocurrencia en condiciones normales Errores de captura Situaciones extremas 	Actividad Manual	n/a
Especialista de Análisis Computacional	4. Defina con el Especialista de Análisis Estadístico la recomendación de cómo abordar el valor atípico, al aplicar el modelo estadístico. Algunos ejemplos: <ul style="list-style-type: none"> Ocurrencia en condiciones normales, se deben considerar en el modelo. Errores de captura, se deben eliminar en el modelo. Situaciones extremas, pueden considerarse como análisis de escenarios. Un escenario sin considerar la situación extrema, otro escenario dónde se considera la situación extrema. Con esto se estima el promedio ponderado, en función de las probabilidades de ocurrencia de cada escenario. 	Actividad Manual	n/a
Especialista de Análisis Computacional	5. Ajuste el conjunto de datos de acuerdo a las consideraciones del punto anterior.	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> Base o conjunto de datos ordenado y valores atípicos evaluados. 			

Tabla 9. Actividad 3, Ejecución del ciclo de Descubrimiento

5.1.4 Actividad 4 – Evaluación y manejo de datos nulos o faltantes.

ENTRADAS:			
<ul style="list-style-type: none"> Base de datos o conjunto de datos ordenados y valores atípicos evaluados 			
EJECUTOR	Pasos	TIPO	SLA
Especialista de Análisis Computacional	1. Identifique valores faltantes, nulos, o en blancos. Lo recomendable es un máximo de 5% de datos faltantes, o en casos extremos hasta 10%.	Actividad Manual	n/a
Especialista de Análisis Computacional	2. Indique al especialista de análisis estadístico, el nivel de datos faltantes y definen mejor mecanismo para manejarlos. Algunos ejemplos son: <ul style="list-style-type: none"> Datos faltantes “correctos”, fecha de fallecimiento a personas que están vivas. Datos faltantes por fallas técnicas, falló eléctrico al momento que se registraba un dato. Datos faltantes aleatorios, algunos clientes completan, otros no lo completan. Datos faltantes con sesgo, casos en donde grupos específicos no registran sus ingresos en los formularios. 	Actividad Manual	n/a
Especialista de Análisis Computacional	3. Ajuste el conjunto de datos, de acuerdo con el mecanismo para manejar datos en blanco del punto anterior, entre las opciones pueden ser: <ul style="list-style-type: none"> Crear una variable categórica adicional que identifica campos específicos en blanco, cuándo hay dudas de posibles sesgos o correlaciones. Cargar a cada campo en blanco el valor promedio del atributo. Realizar una regresión para asignar el valor en función de otras variables, existentes en la tabla. Obviamente puede generar correlaciones entre variables y no considera la varianza. Otro método es realizar una regresión incluyendo alguna aleatoriedad en el modelo. 	Actividad Manual	n/a
Especialista de Análisis Computacional	4. Ajuste el modelo de acuerdo a las consideraciones del punto anterior.	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> Base o conjunto de datos ordenado y valores atípicos evaluados. 			

Tabla 10. Actividad 4, Ejecución del ciclo de Descubrimiento

5.1.5 Actividad 5 – Escalamiento y Estandarización de la data

ENTRADAS:			
<ul style="list-style-type: none"> Base o conjunto de datos ordenado y valores atípicos evaluados. 			
EJECUTOR	Pasos	TIPO	SLA
Especialista de Análisis Computacional	1. Identifique las variables cuantitativas y verifica si requiere un proceso de Escalamiento o Estandarización.	Actividad Manual	n/a
Especialista de Análisis Computacional	2. Define con Especialista en Análisis estadístico el proceso de transformación adecuado, para mitigar el riesgo de que variables con escalas diferentes sesguen el resultado. Por ejemplo, una variable cuya escala va de 0 a 1000, “puede” influenciar más el resultado que una variable de 0 a 1. En este caso es apropiado “escalar” o “estandarizar” ambas variables. <ul style="list-style-type: none"> Escalamiento: Coloca todas las variables en la misma escala, por ejemplo, valores de 0 a 1. Ajustando cada atributo con su valor mínimo igual a cero y su valor máximo igual a 1. Calculando cada valor del atributo $(\text{Valor } n - \text{Valor Min}) / \text{Valor Max}$ Estandarización: Transforma o normaliza el valor de cada atributo en función del promedio y la desviación estándar, creando valores entre -1 y 1. Siendo el promedio igual a cero. $(\text{Valor } n - \text{Promedio}) / \text{Desviación Estándar}$ 	Actividad Manual	n/a
Especialista de Análisis Computacional	3. Aplica proceso definido en el punto anterior, que en algunos casos pueden ser realizados por el algoritmo utilizado para el modelado, en otros casos se debe aplicar mediante funciones específicas, como “scale” en R.	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> Base o conjunto de datos ordenado, valores atípicos evaluados, valores “escalados”. 			

Tabla 11. Actividad 5, Ejecución del ciclo de Descubrimiento

5.1.6 Actividad 6 – Aplicación del modelo Estadístico

ENTRADAS:											
<ul style="list-style-type: none"> Base o conjunto de datos ordenado, valores atípicos evaluados, valores “escalados”. Plan de Trabajo de Análisis 											
EJECUTOR	Pasos	TIPO	SLA								
Especialista de Análisis Computacional	1. Separe la data entre grupos de entrenamiento, validación y prueba. Usualmente puede ser proporción 70%- 15%-15% o 60%- 20% -20%.	Actividad Manual	n/a								
Especialista de Análisis Estadístico	2. Proponga la forma de aplicar el modelo, así como el orden de ser aplicado. Puede haber casos que haya diferentes modelos que se construyen uno sobre otro, dependiendo de la información que se busca.	Actividad Manual	n/a								
Especialista de Análisis Estadístico	3. Seleccione el modelo Estadístico a ser aplicado. En la mayoría de los casos es más de un modelo que se aplica sobre la data de entrenamiento, de tal forma que haya formas de comparar y seleccionar el que mejor predice la variable objetivo.	Actividad Manual	n/a								
Especialista de Análisis Estadístico	4. Aplique, en conjunto con el Especialista de Análisis Computacional el o los modelos estadísticos, sobre la data de prueba. Este proceso es iterativo.	Actividad Manual	n/a								
Especialista de Análisis Estadístico	5. Verifique los resultados del modelo sobre la data de validación. <table border="1" data-bbox="534 1136 1105 1388"> <thead> <tr> <th>Si el resultado es</th> <th>Entonces...</th> </tr> </thead> <tbody> <tr> <td>...</td> <td></td> </tr> <tr> <td>Satisfactorio</td> <td>Continúe con el paso 6.</td> </tr> <tr> <td>No Satisfactorio</td> <td>Ajusta el modelo o implementa un modelo diferente y Continúe con el paso 4.</td> </tr> </tbody> </table>	Si el resultado es	Entonces...	...		Satisfactorio	Continúe con el paso 6.	No Satisfactorio	Ajusta el modelo o implementa un modelo diferente y Continúe con el paso 4.	Actividad Manual	n/a
Si el resultado es	Entonces...										
...											
Satisfactorio	Continúe con el paso 6.										
No Satisfactorio	Ajusta el modelo o implementa un modelo diferente y Continúe con el paso 4.										
Especialista de Análisis Estadístico	6. Aplique el modelo sobre la data de prueba e identifica la probabilidad del modelo de predecir el resultado en nuevos conjuntos de datos. <table border="1" data-bbox="534 1518 1024 1835"> <thead> <tr> <th>Si el resultado es</th> <th>Entonces...</th> </tr> </thead> <tbody> <tr> <td>...</td> <td></td> </tr> <tr> <td>Satisfactorio</td> <td>Continúe con el paso 7.</td> </tr> <tr> <td>No Satisfactorio</td> <td>Ajusta el modelo o implementa un modelo diferente y Continúe con el paso 4.</td> </tr> </tbody> </table>	Si el resultado es	Entonces...	...		Satisfactorio	Continúe con el paso 7.	No Satisfactorio	Ajusta el modelo o implementa un modelo diferente y Continúe con el paso 4.	Actividad Manual	n/a
Si el resultado es	Entonces...										
...											
Satisfactorio	Continúe con el paso 7.										
No Satisfactorio	Ajusta el modelo o implementa un modelo diferente y Continúe con el paso 4.										

Especialista de Análisis Estadístico	7. Documente el modelo seleccionado y emite conclusiones sobre las preguntas de negocio. Y recomienda la forma de repetir el análisis a futuro y que elementos pueden ser sujetos del proceso de persistencia del modelo.	Actividad Manual	n/a
Especialista de Análisis Estadístico	8. Explique a Especialista de Análisis de Negocio, los resultados del análisis.		
Salidas			
<ul style="list-style-type: none"> ▪ Documentación del modelo aplicado. ▪ Conclusiones o resultados del análisis realizado. 			

Tabla 12. Actividad 6, Ejecución del ciclo de Descubrimiento

5.2 Conceptos Relacionados

5.2.1 Underfitting y Overfitting

Overfitting se traduce como “sobreajuste” y el Underfitting como “subajuste”.

Esto hace referencia al fallo o sesgo generado en nuestro modelo, que genera una baja probabilidad de predecir el resultado con otra muestra del mismo grupo poblacional (sobreajuste).

Hay diferentes maneras de prevenir el “Overfitting”, una de ellas es a través de la reducción de cantidad de atributos es muy grande en relación a la cantidad de elementos en el conjunto de dato.

Cuando la cantidad de atributos es muy grande se puede utilizar los modelos como “Lasso Regression” o el “Elastic Net” que ayudan a reducir la cantidad de variables, seleccionando las variables más representativas, de tal manera que solo se tienen ciertas variables al momento de diseñar el modelo y así se reduce el riesgo del “Overfitting”.

Si el modelo entrenado con el conjunto de train tiene un 90% de aciertos y con el conjunto de test tiene un porcentaje muy bajo esto es claramente un problema de Overfitting.

5.2.2 Reducción de errores y sesgos

Un conjunto de datos sesgados puede derivar en decisiones incorrectas. Los sesgos pueden ser producidos por falta de exhaustividad en la recolección de datos, presunciones, prejuicios e inclusive variables presentes en la data pueden llevar al sesgo. A continuación, algunas consideraciones a tener en cuenta para evitar la influencia de sesgos en el tratamiento de los datos:

- Considerar el contexto del que salen los datos y la metodología empleada en recogerlos.
- Establecer una hipótesis a validar como punto de partida, ya que, de lo contrario, cualquier conclusión resultara válida.
- Usar datos transversales que incluyan minorías y todos los segmentos representativos de la población.
- Buscar datos de calidad.
- Evitar cuestionarios con preguntas que induzcan respuestas.
- Cruzar distintas variables siempre para comprender cómo es la relación entre ellas, y poder inferir adecuadamente.

5.2.3 Proceso de Limpieza

Como parte de la fase de descubrimiento, el proceso de limpieza de datos viene dado por la acción de corregir y/o eliminar registros de datos erróneos de una fuente de datos, es decir, este proceso permite identificar datos erróneos, incompletos, no pertinentes, entre otros.

El proceso de limpieza puede involucrar varias iteraciones, hasta alcanzar la adecuación de los datos, cumpliendo el siguiente flujo:

- **Análisis Exploratorio de Datos:** Comprende una inspección manual, a fin de determinar en las muestras de datos los errores, datos a eliminar, entre otros. En este punto se pueden incorporar herramientas de estadística descriptiva que apoyen tal inspección; como la media, mediana, histogramas, gráficos de 2 dimensiones, cuartiles.
- **Transformación y mapeo:** Es recomendable procurar dos niveles de mapeo de datos para su transformación. Un primer nivel dirigido a una única fuente de datos, a fin de prepararlos para su integración con los datos de otras fuentes. También un segundo nivel dirigido a transformar los datos que provengan de diversas fuentes de datos. Mapeo es el proceso de establecer correspondencias de una base de datos a otra.
- **Verificación:** se aplican varias iteraciones de pruebas para verificar la efectividad de la transformación y detectar posibles nuevos errores.

Manejo de tipos de datos (Fechas, texto, datos, categóricos).

- **Datos de texto:** los datos de texto se utilizan para especificar categorías. Un ejemplo, columnas que pueden tener los valores de texto “masculino” y “femenino” para los dos valores de género.

- **Datos geoespaciales:** Es el registro digital con combinaciones de valores de atributos y debe estar georreferenciado y contener un atributo que lo relacione con el tiempo para fines de comparación. Como ejemplo de este tipo de datos se pueden citar ciudades, ríos, montañas.

- **Datos de red:**

Aquellos que tienen origen en:

- + Información sobre clicks en vínculos y elementos
- + Búsquedas en Google
- + RRSS (fuentes de datos de Twitter, publicaciones en Facebook, otras RRSS)
- + Contenido Web (páginas, imágenes, enlaces, etc.)

- **Datos cualitativos:** son típicamente descriptivos y expresan en forma de palabras o textos, producto generalmente de preguntas ó encuestas abiertas, y que permiten comprender un contexto determinado. El flujo para abordar este tipo de datos pasa generalmente por una selección de los datos para tomar los que realmente son de provecho, para disponerlos, presentarlos y organizarlos de forma ordenada a fin de hacer posible la extracción de conclusiones.

- **Datos cuantitativos:** Este tipo de datos comprende datos **cuantitativos continuos y discretos**. Los datos cuantitativos continuos (no se restringen a valores enteros) se miden en lugar de contarse. Además, tienen entre sus características que pueden dividirse. Por ejemplo: Medir la altura de una persona, en metros, centímetros, es decir, los datos son continuos. La edad se puede dividir en años, meses y hasta días. Los datos **cuantitativos discretos** son valores enteros, por valores completos. Se cuentan, no se miden. Por ejemplo: Número de hijos, adultos o mascotas en su familia

Manejo de campos en blanco y nulos. Definición de proceso de “rellenar datos blancos”.

En base al análisis de los datos, se identifican campos vacíos. Nunca los campos están completos al 100%.

Existen diferentes metodologías para ello (desde regresiones para estimar el valor, hasta la creación de campos adicionales para identificarlos), pero si superan más del 5% de toda la base pudiese dudarse de la calidad de la base de datos.

Por ende, entender la causa de información faltante, es importante para evitar sesgos en el modelo analítico.

Identificación de valores atípicos. Manejo de Valores atípicos.

Es importante entender los valores atípicos, comprender si son producto de un error o valores que se generaron en condiciones especiales.

Existen diferentes tipos de modelos para abordar esto, una forma es generar un modelo sin los valores atípicos, otro modelo con los valores atípicos y estimar probabilidades de ocurrencia de cada escenario para estimar un “modelo” esperado.

Escalamiento y estandarización de datos.

Dependiendo del tipo de variables, se puede escoger uno u otro. **Escalar**, quiere decir colocar todas las variables en una escala de 0 a 1.

Y **estandarizar**, de transformar las variables a distribución normal, dónde cero es la media.

Esto con el fin de no tener coeficientes dispares en el modelo analítico.

Es cómo hablar de ingresos anuales en términos de miles de dólares y edad en términos de unidades, estas distorsiones pueden crear sesgos en el modelo.

5.2.4 Separación de datos de Entrenamiento- Validación- Prueba

Durante la construcción de modelos hay 3 conjuntos de datos fundamentales a procurar:

- **Conjunto de datos de entrenamiento:** son los datos que entrenan los modelos
- **Conjunto de datos de validación:** selecciona el mejor de los modelos entrenados
- **Conjunto de datos de prueba:** ofrece el error real cometido con el modelo seleccionado

La distribución del 100% de los datos recolectados debe seguir en lo posible la siguiente distribución:

- 60% para entrenar
- 20% para validación
- 20% para pruebas

El conjunto de prueba deberá tener muestras diversas en lo posible y una cantidad de

muestras suficientes para poder comprobar los resultados una vez entrenado el modelo.



5.2.5 Definición y aplicación de Modelos sobre data de prueba.

Una vez segmentada la data, es momento construir los modelos. Un modelo por lo general es una fórmula de carácter predictivo, cuyo objetivo puede ser buscar resultados, comportamientos, entre otros.

Algunas opciones de algoritmos que se pueden aplicar son:

- **Regresión lineal:** es un algoritmo de aprendizaje supervisado que se utiliza en Machine Learning y en estadística. En su versión más sencilla, nos indicará la tendencia de un conjunto de datos continuos, en relación a otra variable. Es decir, nos ayudará a identificar la dependencia de una variable “dependiente” en relación a una o más variables independientes.
- **Regresión logística:** es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores.

5.2.6 Seleccionar modelo de prueba

Sobre la misma data de prueba, se deben definir o aplicar al menos 2 modelos diferentes. Una vez “entrenado” los modelos, es momento de seleccionar el modelo final, es decir, aquel modelo que tiene una mayor probabilidad de predecir el resultado en el conjunto de validación.

La selección del modelo deberá hacerse tomando como referencia un valor previamente definido al que llamaremos el criterio de evaluación. Este criterio debe estar directamente alineado con los objetivos de mayor nivel del negocio, y también directamente relacionada con el tipo de problema al que nos enfrentamos:

- Los problemas de regresión utilizan ciertas métricas de evaluación, como el error cuadrático medio.
- Los problemas de clasificación usan métricas de evaluación como precisión, exactitud y recuerdo.

5.2.7 Eficiencia del modelo

En este punto se evalúa el modelo finalmente seleccionado con un conjunto de datos que no ha sido utilizado ni en el entrenamiento ni en la validación de dicho modelo, es decir, se evalúa el modelo con un conjunto de datos de pruebas.

Esto evaluará la capacidad real del modelo tomando como referencia dos tipos de métricas de aprendizaje supervisado:

- Precisión, para modelos de clasificación.
- Error, para modelos de regresión.

Capítulo 6. Comunicación de Resultados

Los resultados del modelo analítico deben ser comunicados de forma simple, directo y fácil de recordar.

La comunicación debe ser capaz de expresar las ideas complejas de manera simple y adaptada para la audiencia correcta. La comunicación debe ser orientada a la acción, es decir, con los elementos relevantes del análisis, se debe lograr interpretar y hacer conclusiones que solucionen lo que se quiere resolver.

6.1 Actividades



Figura 27. Subproceso Comunicación de Resultados

6.1.1 Actividad 1 – Preparar comunicación de resultados

ENTRADAS:									
<ul style="list-style-type: none"> Documentación del modelo aplicado. Conclusiones o resultados del análisis realizado. 									
EJECUTOR	Pasos	TIPO	SLA						
Especialista de Análisis de Negocio	1. Revise los resultados o conclusiones del análisis de datos. Compare los resultados con las preguntas iniciales a responder.	Actividad Manual	n/a						
Especialista de Análisis de Negocio	2. Verifique si las conclusiones responden a las preguntas que se necesitaban responder.	Actividad Manual	n/a						
	<table border="1"> <tr> <th>Si las conclusiones responden las preguntas de forma ...</th> <th>Entonces...</th> </tr> <tr> <td>Satisfactoria</td> <td>Continúe con el paso 3.</td> </tr> <tr> <td>No Satisfactoria</td> <td>Converse con el Especialista de Análisis Estadístico y vaya a la actividad 6</td> </tr> </table>			Si las conclusiones responden las preguntas de forma ...	Entonces...	Satisfactoria	Continúe con el paso 3.	No Satisfactoria	Converse con el Especialista de Análisis Estadístico y vaya a la actividad 6
	Si las conclusiones responden las preguntas de forma ...			Entonces...					
Satisfactoria	Continúe con el paso 3.								
No Satisfactoria	Converse con el Especialista de Análisis Estadístico y vaya a la actividad 6								

		del ciclo de descubrimiento.		
Especialista de Análisis de Negocio	3.	Inicie una presentación (preferiblemente herramientas similares a power point, keynote o prezi) para colocar los resultados del análisis.	Actividad Manual	n/a
Especialista de Análisis de Negocio	4.	Tenga claro su público objetivo, así como el objetivo de la comunicación. Redacte los objetivos que busca y las preguntas que quiere responder en la presentación.	Actividad Manual	n/a
Especialista de Análisis de Negocio	5.	Cree una agenda de los puntos que incluirá en la presentación. <ul style="list-style-type: none"> ▪ Agenda ▪ Objetivos ▪ Problema a resolver ▪ Descubrimientos o información relevante. ▪ Acciones sugeridas. ▪ Próximos pasos. 	Actividad Manual	n/a
Especialista de Análisis de Negocio	6.	Coloque un máximo de 3 conceptos, ideas o mensajes por diapositiva.	Actividad Manual	n/a
Especialista de Análisis de Negocio	7.	Identifique el problema a resolver, de la forma más cuantitativa posible.	Actividad Manual	n/a
Especialista de Análisis de Negocio	8.	Explique el enfoque utilizado para la solución, resumiendo los pasos realizados, en no más de 3 conceptos claves.	Actividad Manual	n/a
Especialista de Análisis de Negocio	9.	Explique los resultados obtenidos de forma sencilla, con la utilización mínima posible de conceptos complejos técnicos, matemáticos o estadísticos. De tal forma que la presentación sea explicada por sí misma, que no requiera tener al presentador cerca para poder entenderla.	Actividad Manual	n/a
Especialista de Análisis de Negocio	10.	Sugiera los próximos pasos en cuanto a la información encontrada, que deben ir en relación a replicaciones del modelo, automatizaciones y acciones en la institución para solucionar el problema planteado.	Actividad Manual	n/a
Salidas				
<ul style="list-style-type: none"> ▪ Presentación de los resultados del análisis. 				

Tabla 13. Actividad 1, Comunicación de resultados

6.2 Conceptos relacionados

6.2.1 Enfoques para la comunicación

Cuando se habla de comunicación de resultados debemos tener claro que el objetivo de la analítica es lograr los objetivos de la organización. En este caso nos estamos enmarcando en el “enfoque estratégico” de la comunicación, donde su objetivo es reportar descubrimientos y detectar patrones o tendencias.

Si se estuviera en el enfoque táctico de la comunicación, lo que se requiere es monitorear la ejecución, dar seguimiento a las acciones, conocer que está pasando.

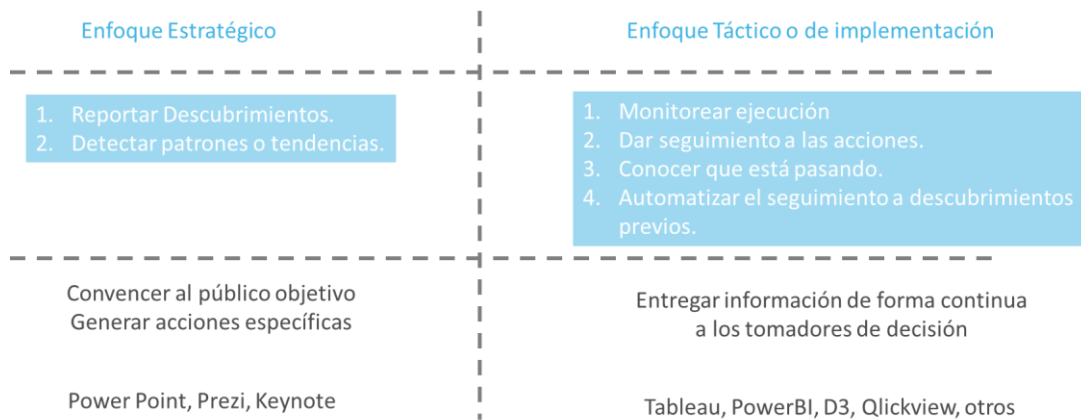


Figura 28. Enfoque estratégico vs Enfoque Táctico

Es importante aclarar que no hay enfoque más importante que el otro, son parte del mismo ecosistema que se utilizan en momentos diferentes, tienen objetivos diferentes y son dos tipos de comunicaciones diferentes.

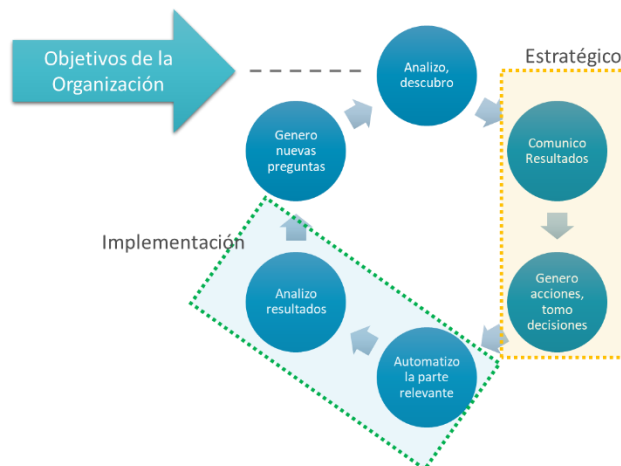


Figura 29. Enfoque estratégico y Enfoque Táctico

La comunicación se debe al Análisis estratégico de datos

- Debe ser capaz de expresar ideas complejas en términos simples.
- Debe adaptarse a la audiencia, de tal forma que el mensaje sea más efectivo.
- Debe estar orientada a la acción

En la comunicación de nuestro análisis estratégico, nos enfocamos en:

- Cómo obtener rápidamente estadísticas relevantes.
- Cómo interpretar esos resultados y hacer conclusiones accionables, en el contexto del caso de uso que queremos resolver.

6.2.2 ¿Cómo encontrar resultados rápidamente?

Se deben identificar grupos: cuales son los grupos que tenemos en la data. Estos pueden darse por medio de la clusterización o la clasificación.

¿Cómo encontrar grupos en la data?

Tenemos 4 estrategias básicas para el análisis de la información

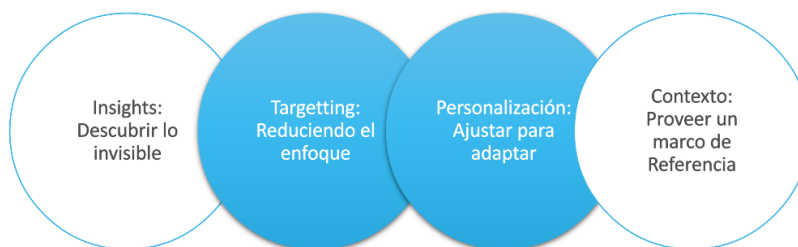


Figura 30. *Cómo encontrar grupos en la data*

Encontrar grupos es básicamente encontrar el balance perfecto entre los parecidos y los diferentes, para aprovechar las economías de escala, para tratar igual a los que se parecen y al mismo tiempo tratar diferente a los que son diferentes para que las acciones sean más efectivas.

¿Hay correlaciones? Hay factores que generan otros eventos, aunque muchas veces encontrar causalidad es muy complejo, con que encontremos correlación entre 2 eventos es suficiente para predecir probabilidad de ocurrencias.

6.2.3 Herramientas de comunicación de resultados

Al momento de explicar los resultados, se explica “libre” de modelo.

- Lo primero, cuál es el poder de explicación del modelo, puede ser medido por la correlación o la matriz de confusión. De no ser así no sirve de nada.
- Lo segundo, cuáles son los factores significantes. Debemos identificar los significantes, de cuáles no. (p-value).
- El “signo” del efecto (correlación positiva o negativa).

Al momento de presentar los resultados no debemos incluir elementos o conceptos estadísticos en la explicación a menos que sea a la unidad analítica.

Se debe solo reportar lo significativo de una manera visual basado solo en los factores relevantes, priorizando dichos factores y analizar si impactan de forma positiva o negativa.

Para mostrar resultados debemos formular un cuadro sencillo, colocando lo que se quiere destacar como un ranking, no siempre la mejor manera de presentar es con gráfica, mientras más simple y sencilla es mejor.

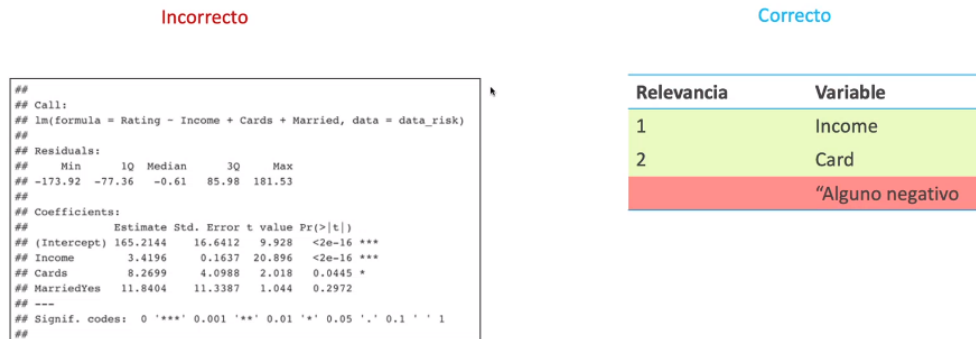


Figura 31. Enfoque correcto de visualización

¿Es posible predecir algún evento?

Al momento de predecir se pueden clasificar datos por medio de análisis de supervivencia o dimensionar capacidades.

Un método de análisis efectivo es, Survival analyses, es muy útil cuando se quiere predecir eventos futuros. ¿Cuándo pasará?

6.2.4 Recomendaciones generales de presentación de resultados

Definir el Color Adecuado

Dicen que una imagen vale más que mil palabras, sin embargo, también hay que asegurarse que el gráfico sea fácil de entender y directo al grano.

Algunos errores comunes son mostrar graficas de pie o con dos o tres dimensiones, porque a nivel visual es difícil identificar dentro de un pie el grado que tiene un dato en la porción del gráfico.

El gráfico debe ser muy sencillo, fácil de entender, sin mucha variabilidad de colores, debido a que en la audiencia puede haber personar que tengan algún nivel de daltonismo por esta razón se debe buscar que paleta de colores es la más apropiada.

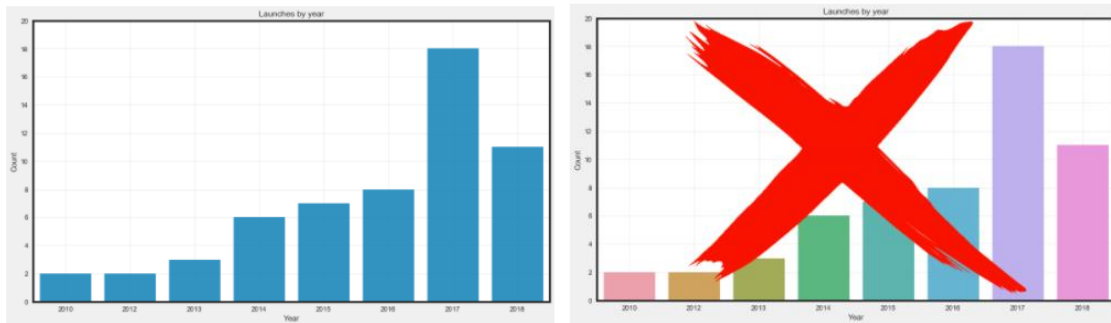


Figura 32. Uso correcto de colores en gráficos

- Daltonismo: Rojo y verde son los más comunes (pero no los únicos)
- Buscar información y simuladores en línea
- Paletas accesibles para colores daltónicos
- Fuentes fáciles de leer y etiquetas
- Utilizar letras Sans-serif son fáciles de leer
- Colocar siempre etiquetas (Título, Eje x, Eje y, leyenda)

Ejes

Los ejes deben hablar sobre el resultado de la información.

Un ejemplo claro podemos verlo en la *imagen 21*, se ve 6,000,000 y 7,066,000, por utilizar un eje incorrecto se ve que uno tiene muy poca cantidad y el otro bastante, cuando en realidad tienen casi lo mismo.



Figura 33. Uso correcto de ejes en gráficos

Tableros

Es importante tener tableros donde puedas ver varios resultados, como los que hay en un auto, pero es importante que estén separados y que sean entendibles e independientes.



Figura 34. Tableros no cargados

Comparto mi información de una forma fácil de recordar y accionar

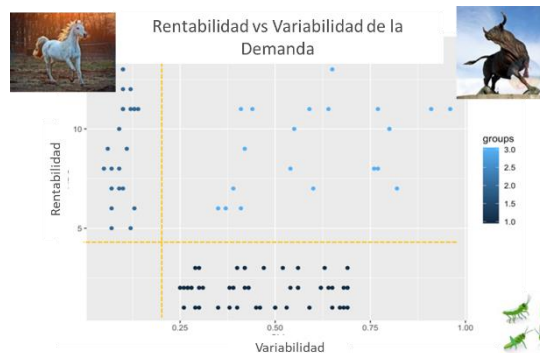


Figura 35. Ejemplo de cómo relacionar información para facilitar su recuerdo

1. Utiliza elementos que relacionen la descripción.
2. De una forma que se genere un mensaje claro y entendible.
3. Utilizar nombre y analogías, fácil de recordar.

Herramientas de un Científico de Datos

- Los paneles agrupan toda la información en solo lugar para la recopilación de ideas y actuar sobre ellas.
- Por ejemplo, en este panel, cualquier vendedor puede ver cómo van progresando las ventas en el trimestre, pero también como está la progresión con los trimestres anteriores, además de realizar un seguimiento de las transacciones y oportunidades, así como el recuento de clientes. Todo en un mismo lugar.

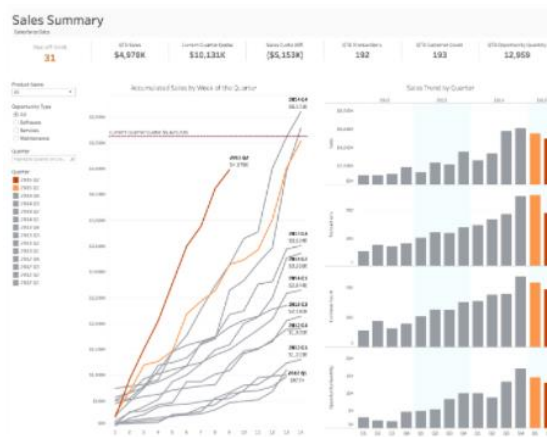


Figura 36. Ejemplo Panel de Ventas

Herramientas BI

Estas herramientas son muy útiles para apoyarse en la comunicación estratégica, en la que debemos estar mostrando información continuamente.

Las herramientas de Business Intelligence permiten limpiar, explorar y visualizar datos y construir paneles sin necesidad de tener conocimientos de programación, también puedes usar utilizando Python, R o incluso JavaScript.



Figura 37. Herramientas BI

Capítulo 7. Implementación de Acciones

7.1 Actividades



Figura 38. Subproceso Implementación de Acciones

7.1.1 Actividad 1 – Implementar las acciones

ENTRADAS:			
<ul style="list-style-type: none"> Presentación de resultados del análisis. 			
EJECUTOR	Pasos	TIPO	SLA
Líder de la Implementación Analítica	1. Recomienda a la Institución receptora las acciones a ejecutar. En conjunto con el líder de la Institución receptora define los siguientes pasos.	Actividad Manual	n/a
Líder de la Implementación Analítica	2. Defina en conjunto con el líder de la Institución receptora, el plan de trabajo para la ejecución de las acciones.	Actividad Manual	n/a
Líder de la Implementación Analítica	3. Proponga los medidores de éxito de las acciones implementadas, de tal forma que se tomen	Actividad Manual	n/a
Líder de Institución receptora	4. Ejecuta las acciones acordadas con el Líder de Implementación analítica.	Actividad Manual	n/a
Salidas			
<ul style="list-style-type: none"> Plan de Trabajo para la ejecución de acciones. 			

Tabla 14. Actividad 1, Implementar las acciones

7.1.2 Actividad 2 – Monitorear las acciones implementadas

ENTRADAS:											
<ul style="list-style-type: none"> Plan de Trabajo para la ejecución de acciones. 											
EJECUTOR	Pasos	TIPO	SLA								
Líder de la Implementación Analítica	1. Evalúa en conjunto con la institución receptora, los avances en la implementación de acciones.	Actividad Manual	n/a								
Líder de la Implementación Analítica	2. Identifica si las acciones implementadas generan el resultado esperado. <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>Si el resultado es</th> <th>Entonces...</th> </tr> </thead> <tbody> <tr> <td>...</td> <td></td> </tr> <tr> <td>Satisfactoria</td> <td>Continúe con el paso 3.</td> </tr> <tr> <td>No Satisfactoria</td> <td>Vaya al paso 1 de la actividad 1, del proceso de implementación de acciones.</td> </tr> </tbody> </table>	Si el resultado es	Entonces...	...		Satisfactoria	Continúe con el paso 3.	No Satisfactoria	Vaya al paso 1 de la actividad 1, del proceso de implementación de acciones.	Actividad Manual	n/a
Si el resultado es	Entonces...										
...											
Satisfactoria	Continúe con el paso 3.										
No Satisfactoria	Vaya al paso 1 de la actividad 1, del proceso de implementación de acciones.										
Líder de la Implementación Analítica	3. Concluye la intervención del análisis, generando un reporte de pos-implementación, en dónde se pueden incluir automatizaciones o el desarrollo de nuevos reportes, o inclusive la generación de estudios a mayor profundidad.	Actividad Manual	n/a								
Salidas											
<ul style="list-style-type: none"> Reporte de Resultados post- implementación de acciones. 											

Tabla 15. Actividad 2, Monitorear las acciones implementadas

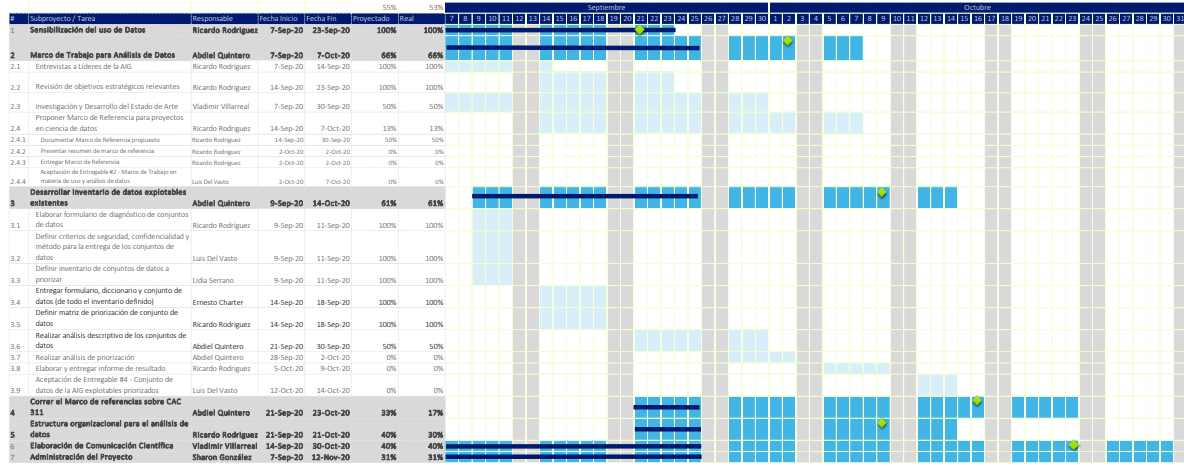
7.2 Conceptos relacionados

Siempre que se aplica una acción como resultado de un proceso analítico debemos considerar lo siguiente:

- Establecer indicadores de medición, que permitan identificar el éxito de las iniciativas ejecutadas, y que sean fácil de medir o monitorear.
- Definir un periodo de tiempo apropiado para evidenciar el resultado de la iniciativa ejecutada.
- Siempre en caso de no lograr los objetivos se debe ver que impidió alcanzar el resultado o si hubo algún supuesto que no se logró comprobar.
- Importante siempre tener en cuenta que la generación de acciones es un proceso iterativo que debe continuamente sufrir ajustes para lograr el resultado.

Anexos

Ejemplo de cronograma de proyecto



Anexos 1. Ejemplo de Cronograma de Proyecto

Plan de proyecto



Formato%20Plan%20de%20Proyecto.docx

Autoridad Nacional para la Innovación Gubernamental			AIG	
Documento que resume las características y objetivos del proyecto. Autoriza formalmente la existencia del proyecto y confiere al líder la autoridad para asignar los recursos de la organización en la atención y cumplimiento de las actividades en tiempo, alcance y calidad.				
Datos Generales del Proyecto				
Proyecto				
Institución Receptora		Fecha Inicio		
Patrocinador		Fecha de Fin		
Líder Institución Receptora		Líder AIG		
Caso de Negocio				
Necesidad o Problema	Objetivo(s) Esperado(s)		Medida(s) de Éxito	
Planteamiento o Solución				
Descripción				
Alcance				
Fuera de Alcance				
Supuestos				
Restricciones				
Cronograma de Trabajo e Hitos				
Equipo				
Rol	Nombre	Responsabilidades		
Patrocinador		Responsable de proporcionar recursos y apoyo para el proyecto y facilitar su éxito.		
Líder Institución Receptora		Encargado de proporcionar la visión general para la implementación de un proyecto. Responsable de que se cumplan las medidas de éxito de este.		
Líder AIG		Persona con conocimiento completo de los requerimientos funcionales del proyecto. Encargado de administrar el detalle cada una de las actividades de índole funcional definidas dentro del proyecto, garantizar el cumplimiento de estas.		
Líder Técnico		Director o responsable principal del departamento de Tecnología o en su defecto una posición relevante dentro del área de tecnología con conocimiento de los repositorios de datos y estructura.		
Participantes				
Entregable(s)				
Administración de Riesgos				
Riesgo	Impacto	Probabilidad	Estrategia	Responsable
No atención o atrasos en las tareas del proyecto por parte de los participantes	Alto	Medio	Mitigar. Asegurar la disponibilidad de los recursos o reasignación de nuevos en el caso de ser necesario.	
Información incompleta o errada para el análisis	Alto	Baja	Mitigar. Validar la información proporcionada en las entrevistas con el líder funcional.	
Comunicaciones				
Comunicación	Frecuencia		Audiencia	
Informe de Seguimiento	Semanal		Todos los miembros del proyecto	
Presentación resultado final	Al final de la consultoría		Patrocinador y Líder del Proyecto	
Gestión de los Cambios				
Se considera un cambio todo aquello que modifique las limitaciones iniciales del proyecto, las cuales deben estar claramente definidas en este plan del proyecto, siendo lo más habitual la modificación del alcance (solicitud de incluir nuevos entregables), impactos en tiempos o costos.				
No se consideran cambios aquellas modificaciones que se deban hacer para corregir un entregable, lo cual se considera como el resultado de una mala gestión de la calidad.				
Todo solicitud de gestión de cambio debe ser presentada de manera formal por el Líder del Proyecto, analizada en conjunto con el Líder AIG y aprobada por el Patrocinador del Proyecto.				
Aprobación				
Aprobado por (Firma Patrocinador)			Fecha	

Anexos 2. Plan de Proyecto

Formato para entrevista a personal de alto nivel de institución receptora

Entrevista Autoridades

1. ¿Cuál es la estrategia Global de la Autoridad?

Escriba su respuesta

2. ¿Cuál es la visión que quiere alcanzar? ¿Por qué?

Escriba su respuesta

3. ¿Cuáles son las principales oportunidades que ha percibido que pueden impactar en el éxito de su estrategia?

Escriba su respuesta

4. ¿Qué espera que se pueda lograr con los datos? ¿Cuál es su perspectiva a nivel de la institución?

Escriba su respuesta

Anexos 3. Formato para entrevista a personal

5. ¿Cuáles son las principales barreras internas a la organización para lograr un uso de datos de forma exitosa?

Escriba su respuesta

6. ¿Cuáles ve cómo las principales situaciones a resolver con los datos?

Escriba su respuesta

7. ¿Qué proyectos o iniciativas paralelas atienden a la digitalización institucional?

Escriba su respuesta

8. ¿Cómo visualiza el uso de la data a nivel de su institución?

Escriba su respuesta

Anexos 4. Formato para entrevista a personal

Formulario creación de caso de uso



Formulario%20Caso
%20de%20Uso.docx

Formulario de Creación de Caso de Uso



- 1. Enlace al objetivo estratégico**
¿Cuál es el objetivo estratégico que apoyará este caso de uso de datos?
- 2. Preguntas objetivas y comerciales**
¿Cuál es el objetivo de este caso de uso de datos? ¿Puede articular alguna pregunta que este caso de uso de datos le ayudará a responder?
- 3. Medidas de Éxito (KPIs)**
¿Cómo se medirá el éxito? ¿Qué métricas afectará esta iniciativa? ¿Cuáles serán los resultados clave?
- 4. Propietarios de casos de uso**
¿Quién será el propietario o patrocinador de este caso de uso de datos?
- 5. Usuarios y clientes de datos**
¿Quiénes serán los usuarios o clientes de datos? ¿Quiénes serán los consumidores de las percepciones?
- 6. Datos requeridos**
¿Qué datos se requerirán? (interno - externo, nuevo - existente, diversidad de datos, estructurado - no estructurado)
- 7. Gobernanza de datos**
¿Qué retos de gobernanza de datos, privacidad de datos, acceso a datos, propiedad de datos y seguridad existente?
- 8. Análisis y Analítica de datos**
¿Cómo se convertirán los datos en conocimientos? ¿Cuáles son los enfoques analíticos propuestos?
- 9. Tecnología**
¿Cuáles son los desafíos y requisitos tecnológicos? (recopilación de datos, almacenamiento de datos, procesamiento de datos, salida de datos)
- 10. Habilidades y capacidad**
¿Cuáles son los desafíos en cuanto a habilidades, capacidades, capacidad y recursos? ¿Quién entregará? Interno - subcontratado - híbrido.

Anexos 5. Formulario de creación de casos

11. Implementación y Gestión del cambio

¿Cuáles son los desafíos y requisitos de implementación y gestión del cambio?

12. Notas Extras

Formulario de conjunto de datos



Form Diagnostico
Conjunto de Datos.pdf

1. Institución origen de la información *

2. Relación del tamaño de estas bases de datos Vs tamaño del total de la institución origen *

3. Sistema o proceso de origen de la información *

- Sitio web de la institución
- Registro en plataforma de la institución (captura manual)
- Proceso de extracción y/o transformación de data
-

Otras

Anexos 6. Formulario de conjunto de datos

4. Detalle el proceso de origen de la información indicado en la pregunta anterior *

5. Indique el tiempo de historia de la información (en años) *

6. ¿La información ha tenido cambios en la historia? *

- Sí
 No

7. Detalle los cambios que ha presentado *

8. Favor indique el tamaño de la Base de Datos *

9. ¿Se ha realizado análisis a esta data previamente? *

- No, es primera vez que se analizará
 Sí, realizamos reportes
 Sí, mantenemos un Datawarehouse
 Sí, mantenemos un Datawarehouse y realizamos reportes

10. Favor especifique *

11. Según su conocimiento, evalúe el nivel de calidad de los datos de esta BD *

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Muy Baja

Muy Alta

12. ¿Cuál es la razón principal para su calificación? *

13. Según su conocimiento, evalúe el nivel de confiabilidad de los datos de esta BD *

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Muy Baja Muy Alta

14. ¿Cuál es la razón principal para su calificación? *

15. Comentarios adicionales

Matriz priorización



Formato%20Matriz%
20Priorización.xlsx

Conjunto de Datos	Casos de usos	Elementos Estratégicos (40%)			Elementos Técnicos (20%)			Elementos de Implementación (30%)				Total
		Impacto ciudadano	Impacto Operacional Interno	Alineamiento Estratégico	Disponibilidad de la Data	Capacidades técnicas del equipo	Disponibilidad de herramientas tecnológicas	Costo Beneficio	Dificultad Cambio Cultural	Potencial de Replicar un piloto exitoso	Nivel de uso de la herramienta	
		20%	10%	10%	15%	10%	5%	10%	10%	5%	5%	
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00
		25	25	25	25	25	25	0	100	25	25	30.00

Anexos 7. Matriz de priorización

Referencias

- Borzacchiello, M., & Craglia, M. (2012). The impact on innovation of open access to spatial environmental information: A research strategy. *International Journal of Technology Management*, 60, 114-129.
- Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data. Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32, 441-452.
- Ruijter, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, 34, 45-52.
- Zhang, J., Puron-Cid, G., & Gil-Garcia, J. (2015). Creating public value through Open government: Perspectives, experiences and applications. *Information Polity*, 33, 97-101.
- Autoridad Nacional para la Innovación Gubernamental. República de Panamá. Proyectos. Accedido el 13 de septiembre de 2020. En <https://innovacion.gob.pa/proyectos/?csrt=9139383259801222398>
- Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", OECD Working Papers on Public Governance, No. 22, OECD Publishing, Paris, <https://doi.org/10.1787/5k46bj4f03s7-en>.
- Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? In *Government Information Quarterly* (Vol. 32, Issue 4, pp. 363–368). Elsevier Ltd. <https://doi.org/10.1016/j.giq.2015.11.007>
- Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). Open government data: A stage model. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6846 LNCS, 235–246. https://doi.org/10.1007/978-3-642-22878-0_20
- Kum, H. C., Joy Stewart, C., Rose, R. A., & Duncan, D. F. (2015). Using big data for evidence based governance in child welfare. *Children and Youth Services Review*, 58, 127–136. <https://doi.org/10.1016/j.childyouth.2015.09.014>
- Linders, D. (2013). Towards open development: Leveraging open data to improve the planning and coordination of international aid. *Government Information Quarterly*, 30(4), 426–434. <https://doi.org/10.1016/j.giq.2013.04.001>
- Magalhaes, G., Roseira, C., & Strover, S. (2013). Open government data intermediaries: A terminology framework. *ACM International Conference Proceeding Series*, 330–333. <https://doi.org/10.1145/2591888.2591947>
- Matheus, R., Ribeiro, M. M., & Vaz, J. C. (2012). New perspectives for electronic government in brazil: The adoption of open government data in national and

subnational governments of Brazil. *ACM International Conference Proceeding Series*, 22–29. <https://doi.org/10.1145/2463728.2463734>

- Ruijter, E., Détienne, F., Baker, M., Groff, J., & Meijer, A. J. (2020). The Politics of Open Government Data: Understanding Organizational Responses to Pressure for More Transparency. *American Review of Public Administration*, 50(3), 260–274. <https://doi.org/10.1177/0275074019888065>
- Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, 32(3), 308–315. <https://doi.org/10.1016/j.giq.2015.05.003>
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325–337. <https://doi.org/10.1016/j.giq.2016.02.001>
- OECD (2008), —Recommendation of the Council for enhanced access and more effective use of Public Sector Information, www.oecd.org/internet/ieconomy/40826024.pdf.
- Apertura, U. E. De. (2016). *Gobierno Apertura y Publicación de Datos Abiertos de Guía Panamá de Implementación*.
- Gallardo Arancibia, J. A. (2013). *Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM*. 84, 487–492. <http://ir.obihiro.ac.jp/dspace/handle/10322/3933>
- Juan, I., Moine, M., Ana, D., Haedo, S., & Gordillo, D. S. (n.d.). *Estudio comparativo de metodologías para minería de datos*.
- Nazareno Pereira Boente Fabiano Saldanha Gomes de Oliveira José Luiz Dos Anjos Rosa, A. (n.d.). *Utilização de Ferramentas de KDD para Integração de Aprendizagem e Tecnologia em Busca da Gestão Estratégica do Conhecimento na Empresa*.
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. In *Revista Iberoamericana de Inteligencia Artificial*. No (Vol. 29). <http://www.aepia.org>