

Mejores prácticas en la gestión de datos de ERT

Mayo 2023



Elaboración

Adrián Calleros, Instituto de Políticas para el Transporte y el Desarrollo

Daniel Bustillos Camargo, Instituto de Políticas para el Transporte y el Desarrollo

Miguel Iván Quintana Nucamendi, Instituto de Políticas para el Transporte y el Desarrollo

Revisión

Santiago Fernandez Reyes, Instituto de Políticas para el Transporte y el Desarrollo

Jorge Ramos, Instituto de Políticas para el Transporte y el Desarrollo

Eloy González Madrazo, Instituto de Políticas para el Transporte y el Desarrollo

Gonzalo Peón Carballo, Instituto de Políticas para el Transporte y el Desarrollo

Pilar Avendaño, Laboratorio de Innovación del Banco Interamericano de Desarrollo

Tatiana Virviescas, Laboratorio de Innovación del Banco Interamericano de Desarrollo

Contribuciones

Andrea Cabrera, Instituto de Políticas para el Transporte y el Desarrollo

Diseño

Brenda Martínez Sandoval, Instituto de Políticas para el Transporte y el Desarrollo

Índice de abreviaturas

API	Application Program Interface
AVL	Automatic Vehicle Location
AWS	Amazon Web Services
CSV	Comma-Separated Value
ERT	Empresa de Red de Transporte
ETL	Extract Transform and Load
GBFS	General Bikeshare Feed Specification
GPS	Global Positioning System
GTFS	General Transit Feed Specification
GUI	Graphical User Interface
HTML	HyperText Markup Language
IAAS	Infrastructure as a Service
IDE	Integrated Development Environment
IoT	Internet Of Things
IPK	Índice de Pasajeros por Kilómetro
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
MDS	Micromobility Data Specification
N/A	Not Applicable
NA	Not Available
NaN	Not a Number
NDA	Non-Disclosure Agreement
PAAS	Platform as a Service
PMI	Plan de Movilidad Institucional
QA	Quality Assurance
SMART	Specific, Measurable, Achievable, Realistic, and Timely
USD	Dólar Estadounidense
XMLS	Extensible Markup Language

Resumen ejecutivo

Las Empresas de Redes de Transporte (ERT) son compañías que brindan servicios de movilidad a través de aplicaciones digitales, conectando a las personas usuarias con una amplia gama de servicios de movilidad en tiempo real. Este ecosistema abarca desde empresas de taxi por aplicación, conocidas como "ride hailing", hasta empresas de micromovilidad que ofrecen bicicletas compartidas y autos compartidos, además de aquellas que proveen tecnología para el transporte público, entre otras.

Al ofrecer sus servicios a través de aplicaciones móviles, las ERT generan información valiosa sobre el uso y los patrones de viaje de sus personas usuarias. La recopilación de estos datos es fundamental para su modelo de negocios, ya que les permite mejorar su eficiencia y aumentar sus ingresos. Por otro lado, esta información permite a las ciudades planificar nuevos servicios de transporte y comprender los patrones de viaje de la población.

El objetivo general de este documento es facilitar la gestión de los datos generados por las ERT en las etapas de planeación, monitoreo y evaluación de proyectos de movilidad. Esto se logra a través del desarrollo de una estrategia de evaluación, la gestión legal del uso de la información y el uso de herramientas y tecnologías necesarias para su análisis. Está dirigido a empresas, organizaciones, ciudades y otros actores interesados en llevar a cabo proyectos de movilidad basados en el uso de datos e información. Con el propósito de facilitar este proceso, se describen las actividades necesarias para aprovechar la información, abarcando desde la identificación y solicitud de datos relevantes hasta las necesidades técnicas y de capacidades. Además, se describen experiencias de tres proyectos piloto del programa Ideamos. Un programa gestionado por el Instituto de Políticas para el Transporte y Desarrollo en México (ITDP por sus siglas en inglés), en colaboración con el Banco Interamericano de Desarrollo (BID) en su división de transporte y el BID Lab, un programa que impulsa un ecosistema de movilidad sostenible e incluyente, enfocado en las personas y su derecho a moverse de manera segura. Para ello, generó entre enero del 2020 y abril del 2023 pilotos innovadores en colaboración con Empresas de Redes de Transporte (ERT), autoridades, iniciativa privada y organizaciones de la sociedad civil, con ayuda de la tecnología y uso de datos.

Planeación y actividades previas al acceso a la información

- **Estrategia de evaluación:** este plan debe incluir los indicadores o KPI, métodos de recolección de datos, fechas de recolección y la metodología para analizar la información, esto con el fin de medir el cumplimiento de los objetivos. Para ello, es necesario el desarrollo de un **Plan de recolección de datos** que contenga las acciones necesarias para obtener la información y para acceder a ella, así como un **Plan de análisis de datos** en el que se deberán incluir los métodos y técnicas para transformar los datos obtenidos en los indicadores de medición de impacto.
- **Convenios legales para el uso de la información:** Para establecer medidas legales y garantizar el acceso y uso adecuado de la información, se sugiere contar con asesoría legal especializada. Se deben cumplir las normas aplicables en materia de privacidad de los datos, y establecer políticas y protocolos de privacidad para proteger la información de las personas usuarias.

Necesidades para el uso de la información

- **Información:** incluyen medidas para asegurar el aprovechamiento eficiente de la información como requerimientos de interoperabilidad y de calidad de la información. Estos elementos son fundamentales para aprovechar al máximo los datos contribuyendo a la toma de decisiones con información confiable.
- **Tecnología:** El uso de formatos abiertos y estándares de datos como GTFS (*General Transit Feed Specification*), GBFS (*General Bikeshare Feed Specification*) y MDS (*Micromobility Data Specification*) es clave para lograr interoperabilidad y facilitar el intercambio de información. Además, se requiere software especializado y herramientas para el procesamiento, análisis y visualización de los datos. Es fundamental implementar medidas de seguridad tecnológica, como actualización de software, seguridad de red, encriptación de datos y prácticas de seguridad, para proteger la integridad y privacidad de la información.

- **Hardware:** según las necesidades de cada proyecto, se requiere hardware potente y rápido, como servidores de cómputo, para manejar grandes volúmenes de datos y realizar operaciones complejas de manera eficiente. Estos servidores deben ser confiables y tener capacidades de recuperación ante fallas. La elección de la implementación puede variar según las necesidades y recursos disponibles, incluyendo servidores locales, cómputo en la nube o cómputo en la nube sin servidor.

Transferencia, procesamiento y análisis de la información

El uso de datos implica una serie de acciones de transferencia, procesamiento y análisis de información, para ello es recomendable realizar procesos automatizados como un *Data Pipeline*, que puede mejorar la eficiencia y calidad del servicio. Se destacan los pasos principales a continuación:

- **Procesamiento y transferencia de datos:** implica el acceso eficiente y seguro a la información, utilizando técnicas y tecnologías para transmitir datos de las ERT a servidores para su procesamiento, donde se emplean algoritmos y herramientas para transformar datos crudos en información procesable, incluyendo la limpieza, combinación de fuentes y generación de resultados útiles.
- **Exploración de la información:** consiste en buscar, recopilar y examinar datos con el objetivo de encontrar patrones, tendencias o relaciones significativas. Este paso es esencial para descubrir información relevante y valiosa dentro de conjuntos de datos grandes y complejos.
- **Análisis de datos:** tiene como objetivo extraer información valiosa, identificar patrones, tendencias o insights que puedan ser utilizados para tomar decisiones informadas. Esto implica el uso de técnicas estadísticas, algoritmos de aprendizaje automático o visualización de datos para obtener conocimientos significativos.

Recomendaciones clave y conclusiones

La implementación del uso de datos en colaboración con gobiernos y ERT ha brindado aprendizajes significativos en el desarrollo de proyectos y toma de decisiones informadas en el ámbito de la movilidad. A continuación se presentan algunas de las lecciones aprendidas más relevantes identificadas:

- **Conocimiento detallado de los datos:** Comprender la metodología y significado de los datos es crucial. Los análisis exploratorios y estadísticos son fundamentales para garantizar la validez de los análisis, y el uso de herramientas estadísticas adecuadas mejora la precisión y relevancia de los resultados para la toma de decisiones.
- **Equipo técnico capacitado:** La gestión y procesamiento de datos requiere de un equipo técnico especializado en herramientas y tecnologías de gestión de datos. Un equipo bien estructurado puede minimizar errores, optimizar recursos y obtener un mayor valor de la información.
- **Adaptación a la disponibilidad de la información:** Es esencial ser flexible y adaptarse a la disponibilidad de los datos. A veces, los datos existentes pueden no cumplir con los requisitos necesarios, por lo que es importante buscar formas de complementar con otras fuentes y utilizar la información disponible de manera efectiva.

Este documento representa una valiosa herramienta que ofrece una guía detallada y práctica para aquellas personas interesadas en maximizar el valor de los datos en el ámbito de la movilidad. Al seguir las recomendaciones y casos de uso presentados, se puede lograr una mejora significativa en las fases de planeación, monitoreo y análisis de proyectos de transporte. Se espera que esta guía sea una herramienta útil para la comunidad interesada en el tema, brindando el conocimiento necesario para impulsar el desarrollo de sistemas de transporte más inteligentes y adaptados a las necesidades de las personas usuarias.

Índice

	Introducción	7
	Objetivos	7
	Estructura del documento	8
	Uso de datos de las ERT para proyectos de movilidad	10
	Planeación	10
	Monitoreo	11
	Evaluación	13
	Planeación y actividades previas al acceso a la información	15
	Estrategia de evaluación y definición de objetivos e indicadores	15
	Convenios de transferencia de información	18
	Necesidades para el uso de información	20
	Necesidades de la información	20
	Interoperabilidad	20
	Calidad de los datos	20
	Diccionario de datos	21
	Seguridad	21
	Necesidades de tecnología	22
	Herramientas de programación	22
	Necesidades de hardware	24
	Almacenamiento de datos	25
	Transferencia, procesamiento y análisis de la información	28
	Transferencia de información	28
	Exploración	32
	Procesamiento de datos	34
	Transformación de los datos	34
	Almacenamiento	34
	Análisis e interpretación	35
	Recomendaciones clave	38
	Conclusiones	39
	Bibliografía	40



Introducción

Las Empresas de Redes de Transporte (ERT) son compañías que proveen servicios de movilidad a través de aplicaciones digitales, que conectan a personas usuarias con los diversos servicios en tiempo real. El ecosistema de ERT considerado incluye empresas de taxi por aplicación, conocidas como ride hailing, empresas de micromovilidad como de bicicletas compartidas, de auto compartido y empresas que ofrecen tecnología para el transporte público, entre otras. Este documento presenta una visión general del panorama sobre la generación y el uso de datos en las Empresas de Redes de Transporte, analiza las principales tendencias y ofrece una visión detallada de las necesidades, oportunidades y actividades para trabajar con esta información.

Para entender mejor el contexto en el que se desarrolla la generación y el uso de datos en las ERT, es importante considerar algunos factores clave. En primer lugar, la demanda de transporte ha crecido significativamente en las últimas décadas, como consecuencia del aumento de la población, el crecimiento económico, y el aumento de la movilidad (Kreuzer & Wilmsmeier, 2014). Esto ha generado una mayor presión sobre la red de transporte, y ha exigido una mayor eficiencia en su gestión. En segundo lugar, la tecnología ha avanzado rápidamente en los últimos años, y ha proporcionado nuevas herramientas y soluciones para el análisis de datos en las ERT. Por ejemplo, la tecnología de Internet de las cosas (IoT) ha permitido generar, recopilar y procesar grandes cantidades de datos en tiempo real.

Asimismo, la regulación y la política pública también han influido en la generación y uso de datos de las ERT. Algunas autoridades locales o regionales han implementado políticas que promueven el uso de estos datos, por ejemplo, con el desarrollo de plataformas de datos abiertos, o la implementación de sistemas de tarificación dinámica.

En la actualidad, la generación y el uso de datos se ha convertido en un eje fundamental para el funcionamiento y éxito de las ERT. Estos datos proporcionan información valiosa sobre el rendimiento, la eficiencia, la calidad, la seguridad y otros aspectos clave del transporte, lo que permite a las ERT tomar decisiones informadas, mejorar sus operaciones y ofrecer un servicio de alta calidad a sus clientes. Además, esta información puede ser de gran utilidad para autoridades y organizaciones externas que busquen: **(i) monitorear** los servicios para la toma de decisiones sobre la operación, **(ii) evaluar** los impactos de estos servicios en distintos temas como las emisiones de gases de efecto invernadero o sus potenciales beneficios sociales y **(iii) utilizar** estos datos como complemento a estudios para la planeación de distintos servicios de movilidad.

El desarrollo de los servicios de ERT se alinea con uno de los ejes de la Visión 2025 del BID, la digitalización de servicios. Esta iniciativa promueve y cataliza inversiones que mejoran la conectividad y fomentan el despliegue de herramientas digitales que aumenten el uso de la tecnología en América Latina y el Caribe (LAC). Algunos factores que han acelerado la digitalización en los distintos sectores de la movilidad son el crecimiento de las cadenas de suministro y la pandemia de 2020. Ello ha provocado un crecimiento del mercado electrónico en LAC, y muchas de estas inversiones se materializan en fintechs y servicios digitales como lo son las ERT.

Objetivos

El objetivo general del documento es facilitar la gestión de datos generados por las ERT en las fases de planeación, monitoreo y evaluación de proyectos de movilidad. Esto, partiendo desde el desarrollo de una estrategia de evaluación, la gestión legal del uso de la información y, finalmente, la presentación de herramientas y tecnologías necesarias para su análisis.

Los objetivos específicos son los siguientes:

- › Definir los criterios para la evaluación de proyectos de movilidad;
- › Identificar las tecnologías y herramientas necesarias para el análisis de datos de las ERT;
- › Mostrar cómo estos datos pueden ser utilizados para mejorar la toma de decisiones en la planificación y gestión de proyectos de movilidad en la ciudad.

Estructura del documento

El documento se estructura en cinco secciones categorizadas de la siguiente forma:

Uso de datos de las ERT para proyectos de movilidad

Describe la importancia del análisis de datos en las ERT, con enfoque en tres ejes (planeación, monitoreo y evaluación).

Necesidades previas al acceso de la información

Contiene recomendaciones para generar una estrategia de evaluación ante un proyecto y las necesidades legales para el acceso a la información por parte de una ERT.

Necesidades para el uso de la información

Identifica las necesidades técnicas, de software y hardware para el desarrollo de un proyecto con datos de movilidad.

Transferencia, procesamiento y análisis

Presenta una descripción cronológica de los pasos necesarios recomendados para llevar a cabo el procesamiento de los datos generados por las ERT.

Lecciones aprendidas y conclusiones

Recapitulación de las áreas de oportunidad así como los aprendizajes, útiles para futuros proyectos.

Para este trabajo se han retomado experiencias de tres pilotos desarrollados por el ITDP a través del programa Ideamos, el cual fue impulsado por ITDP en conjunto con el Banco Interamericano de Desarrollo (BID) y BID Lab, en donde se colaboró con ERT para promover la sostenibilidad e inclusión social de sus servicios en México. Esta información se acompaña a lo largo del documento a través de cuadros azules para cada uno de los casos de estudio, una breve descripción de los pilotos se presenta a continuación.

Movin Reforma

Consistió en desarrollar un Plan de Movilidad Institucional (PMI) con 8 empresas ubicadas en la Avenida Reforma de la Ciudad de México. Dentro del piloto se implementaron 23 estrategias de promoción a modos sustentables y se colaboró con 3 ERT para la implementación de servicios de vanpool, carpool y bicicletas compartidas. Su objetivo fue mejorar la calidad de vida de las personas empleadas y reducir externalidades negativas como congestión vehicular, contaminación del aire, ruido, entre otras.

Digitalización del Transporte Concesionado en Mérida

El proyecto fue desarrollado en la ciudad de Mérida en colaboración con el Instituto de Movilidad y Desarrollo Urbano Territorial (IMDUT) y la empresa de tecnología Embsoft. Los objetivos del piloto fueron mejorar los hábitos de conducción y disminuir los eventos de conducción brusca y el exceso de velocidad mediante un sistema de alertas; así como evaluar el impacto en el nivel de accesibilidad urbana a actividades nocturnas con un sistema de rutas nocturnas recién implementadas.

Moveup

Proyecto realizado en colaboración con la empresa Vemo y tres instituciones de educación superior ubicadas en Puebla y Cholula. El proyecto puso en marcha una ruta entre las tres universidades mediante un minibus eléctrico con accesibilidad universal y perspectiva de género a través de una aplicación digital (app). Además, como parte del proyecto también se digitalizaron siete rutas de transporte desde y hacia las universidades participantes.

El análisis de datos en las ERT no está exento de desafíos y obstáculos. La planeación, monitoreo y evaluación de proyectos de movilidad llevados a cabo por gobiernos u organizaciones suelen llevar dificultades para acceder a los datos que necesitan, para integrar diferentes fuentes de datos y para analizarlos de forma eficiente y escalable. Sin embargo, al existir tantos actores relacionados con diferentes escalas de análisis, los beneficios de la generación de datos de transporte se convierten en una red de interacción que puede impactar más allá de un nivel local o regional.

Con este documento se espera contribuir a un entorno donde la toma de decisiones se apoye cada vez más en la información y en el análisis de datos. Es por esto que resulta crucial que empresas, instituciones y organizaciones de la movilidad cuenten con prácticas eficientes de recolección, almacenamiento y análisis que permitan una adecuada gestión de la movilidad urbana.



Uso de datos de las ERT para proyectos de movilidad

Actualmente, el uso de datos se ha convertido en un factor esencial para el desarrollo de proyectos de movilidad. En el contexto de la movilidad, existen dos tipos de proyectos que pueden ser considerados: los proyectos públicos de movilidad y los proyectos de Mobility as a Service (MaaS). Los primeros se refieren a los proyectos llevados a cabo por entidades gubernamentales o de interés público, que buscan mejorar el transporte y la movilidad en las ciudades. Por otro lado, los proyectos de MaaS se enfocan en la integración de diferentes modos de transporte y servicios en una sola plataforma digital, con el objetivo de mejorar la accesibilidad y la eficiencia en el uso del transporte. Es importante diferenciar ambos tipos de proyectos ya que sus objetivos y estrategias son diferentes, aunque en muchos casos se pueden complementar. En el caso específico de las ERT, estas tecnologías pueden ser una herramienta valiosa tanto para proyectos públicos de movilidad como para proyectos de MaaS, ya que pueden proporcionar información en tiempo real sobre el uso de transporte y la movilidad de las personas usuarias.

Las empresas de transporte están recolectando información a través de diversas fuentes, como dispositivos IoT y aplicaciones móviles (*apps*) desde donde se ofrecen servicios de movilidad como taxis, bicicletas compartidas, vans compartidas (*vanpool/microtransit*), entre otros. Esta información puede ser utilizada en la generación de indicadores para la medición de objetivos de un proyecto, y también puede utilizarse para comprender las necesidades de oferta y demanda en el caso de las ERT.

Existen grandes retos a los que se enfrentan las políticas de movilidad y transporte, uno de ellos está relacionado con la importancia de disponer de datos actualizados y confiables de distintos tipos para medir el éxito de las políticas y estrategias y, en su caso, reajustarlas. Para ello, tanto las ERT, las organizaciones y las ciudades necesitan cifras completas, relativas no sólo a indicadores cuantitativos, como el número de usuarios de algún modo de transporte, la congestión de automóviles en horas pico o las bicicletas disponibles en una estación. Las cifras deben incluir también factores como, por ejemplo, la percepción de seguridad, el comportamiento intermodal o las opciones de movilidad de viajes de recreación o turísticos en el caso de algunas localidades. Algunos de estos datos son realmente difíciles de obtener y requieren de muchos recursos. Para obtenerlos, es necesario implementar un enfoque integral que incluya las etapas de planeación, monitoreo y evaluación. De esta forma será más sencillo localizar los indicadores de interés que el proyecto necesite, ya que éstos pueden variar de acuerdo con los objetivos del mismo.

En esta sección, nos centraremos en los beneficios de utilizar los datos en estas tres etapas clave. Veremos cómo las ciudades, empresas de transporte y otros actores pueden aprovechar esta información para mejorar la eficiencia y la efectividad de sus servicios, optimizar recursos y al mismo tiempo reducir las emisiones de gases de efecto invernadero.

Planeación

La planeación de un proyecto se interpreta como un proceso para desarrollar una estrategia para alcanzar un futuro deseable, para resolver problemas y para facilitar la acción (Mitchell, 2018). Hasta la década pasada la planeación de la movilidad se veía como un problema de ingeniería y administrativo (Hodson & Marvin, 2010) y la responsabilidad principal de la planeación recaía en los gobiernos.

Sin embargo, la introducción de nuevas tecnologías ha abierto nuevas posibilidades para que las personas usuarias, empresas de transporte y gobiernos interactúen de nuevas maneras, más inteligentes y sustentables, a través de la transición de la movilidad coordinada por el gobierno a un sistema multifacético donde se ejerce la coordinación con la ayuda de otros actores, como las empresas responsables del transporte, quienes ofrecen servicios de movilidad como servicio (e.g.: Ecobici, Uber, Dezza) o empresas de tecnología (e.g.: Google, Waze, Where is my Transport) (Melis *et al.*, 2016; Paiva *et al.*, 2021). Un ejemplo del uso de estos datos es la identificación de las rutas transitadas por personas que utilizan la bicicleta u otros servicios de micromovilidad para planear la creación de nuevos carriles confinados para ciclistas y personas usuarias de vehículos de micromovilidad (Abduljabbar *et al.*, 2021).

Una de las maneras para comenzar con la transición a nuevas formas de planeación es a través de la recopilación de datos mediante apps celulares y usando datos generados por instrumentos como GPS (*Global Positioning System*), teléfonos celulares, sensores, etc., para obtener información de orígenes y destinos de las personas, trazas precisas de las rutas de transporte, capacidad utilizada de las unidades de transporte, disponibilidad de bicicletas compartidas, tránsito en tiempo real, entre otras. Esta información (véase **Tabla 1**) es procesada por los gobiernos y empresas responsables del transporte para ayudar a su planeación y es compartida con las personas usuarias para planificar los viajes, comparar tarifas, etc. (Melis *et al.*, 2016).

Tabla 1. Fuentes de información big data útiles para la planeación de servicios de transporte

Fuente de información	Herramientas	Información generada
Tarjeta de transporte público	<ul style="list-style-type: none"> Tarjeta de transporte público 	<ul style="list-style-type: none"> Flujos de origen-destino Tiempo de viaje
GPS	<ul style="list-style-type: none"> Dispositivos GPS 	<ul style="list-style-type: none"> Posición del dispositivo Densidad de dispositivos Velocidad de dispositivos
Datos de seguimiento de apps	<ul style="list-style-type: none"> Datos de redes sociales Datos de ubicación de apps en segundo plano 	<ul style="list-style-type: none"> Tiempo de viaje Flujo de origen-destino Consumo de energía Aceleración del dispositivo
Detector en calles	<ul style="list-style-type: none"> Detector por inducción Tubos neumáticos de carretera Radar de microondas Radar infrarrojo Casetas de peaje 	<ul style="list-style-type: none"> Posición del vehículo Velocidad del vehículo Densidad de vehículos Clasificación de vehículos
Sensores remotos de área extendida	<ul style="list-style-type: none"> GPS Sensores aéreos Antenas de servicio celulares 	<ul style="list-style-type: none"> Tiempo de viaje Flujo de origen-destino
Vehículos conectados autónomos	<ul style="list-style-type: none"> Diversos sensores 	<ul style="list-style-type: none"> Velocidad de vehículos Ubicación de vehículos
Puntos de acceso WI-FI	<ul style="list-style-type: none"> Teléfonos celulares 	<ul style="list-style-type: none"> Ubicación de dispositivos

Fuente: Adaptado de Zhu et al. (2019).

La planeación inicial es fundamental en la gestión de uso de datos porque permite definir los objetivos y las metas que se quieren alcanzar a largo plazo, así como los pasos que se deben seguir para lograrlos. También permite identificar las limitaciones que puedan existir en el proyecto a implementar. En este sentido, es importante destacar que la planeación no sólo se enfoca en la definición de objetivos y estrategias, sino que también permite detectar posibles obstáculos y dificultades que puedan surgir.

Algunos ejemplos de limitaciones que se pueden presentar en el uso de datos de ERT son la falta de infraestructura adecuada, la insuficiencia de recursos, la incompatibilidad tecnológica entre diferentes sistemas, las barreras idiomáticas y culturales, entre otros. Por lo tanto, la identificación temprana de estas limitaciones durante la etapa de planeación puede ayudar a diseñar soluciones más efectivas y a evitar retrasos y costos innecesarios en la implementación del proyecto.

Monitoreo

El monitoreo de los datos de servicios de movilidad es esencial para asegurar la eficiencia del transporte, así como para mantener un estándar alto de la calidad del servicio para las personas usuarias. A lo largo del tiempo, la implementación del monitoreo ha evolucionado significativamente, desde la recopilación manual de datos hasta la integración de sistemas automatizados en tiempo real. Hoy en día, se están utilizando tecnologías avanzadas, como la inteligencia artificial y el aprendizaje automático para mejorar aún más la calidad del monitoreo y, con ello, la toma de decisiones. Por ejemplo, los sistemas de recomendación de rutas que usan las preferencias y necesidades de cada persona usuaria o la detección y prevención de siniestros viales, que se apoyan de instrumentos de visión como cámaras para detectar patrones de conducción peligrosos (Yeo & Shafahi, 2017).

El objetivo del monitoreo en las ERT es asegurar la seguridad y eficiencia de los servicios prestados, mediante la supervisión en tiempo real de la operación de los vehículos y la infraestructura. Algunas ventajas de realizar un monitoreo constante incluyen la detección temprana de fallas y la optimización de la planificación de la operación.

Los indicadores, así como el método de visualización de la información, dependerán del tipo de proyecto. Para proyectos de transporte público, las agencias gubernamentales pueden contar con un sistema de control y monitoreo, como el de la **figura 1**, que les permite dar seguimiento en tiempo real al servicio a través de la información generada por sensores de ascenso y descenso, dispositivos AVL¹, entre otros. Dentro de este centro de control se realiza un monitoreo a distintos niveles de operación, desde una visión general con indicadores como el índice de pasajeros por kilómetro (IPK), la cantidad de pasajeros, etc., hasta un monitoreo detallado, con indicadores desagregados con datos sobre el tráfico en tiempo real, la ubicación de los vehículos, las tasas de retención y otras métricas relevantes.

Los sistemas de control y monitoreo cuentan con sistemas y software que presentan información gráfica en tiempo real e histórica de la operación del transporte. Adicionalmente, suelen contar con sistemas de análisis geoespacial, bases de datos, software de análisis estadístico y sistemas de comunicación. En la siguiente sección se explican las posibles necesidades de software y hardware para el desarrollo y análisis de proyectos de movilidad. Todos estos componentes deben trabajar en conjunto y en tiempo real para brindar una visibilidad completa de la información de transporte, lo que permite tomar decisiones informadas y mejorar la eficiencia y sostenibilidad de las redes de transporte.

Figura 1. Centro de Control y Monitoreo del Sistema Metropolitano de Movilidad Amable y Sostenible del Gobierno de Yucatán (2022)



Otro ejemplo de monitoreo es para el rebalanceo de bicicletas en un servicio de bicicleta pública sin anclaje. Generalmente estos servicios presentan una alta demanda en algunas zonas y horarios, mientras que en otras zonas sus bicicletas están subutilizadas. Debido a esto, es necesario optimizar la disponibilidad de las bicicletas según la demanda en tiempo real de las mismas. Esta tarea es compleja debido a distintos factores que influyen en los patrones de uso, como la hora del día, el día de la semana, la ubicación y el clima. Las empresas de bicicletas compartidas utilizan algoritmos y datos históricos para informar y guiar sus esfuerzos de reequilibrio. Esto demuestra la importancia del uso de datos en tiempo real para llevar a cabo tareas complejas como el reequilibrio de bicicletas (Stephen Gossett, 2020; Xiaomei Tan, 2018).

1 Un dispositivo AVL (Automatic Vehicle Location) es un sistema que utiliza GPS para transmitir su ubicación mediante el uso de internet.

Evaluación

El concepto de evaluación se puede entender como una investigación sistemática del impacto, valor o importancia de un objeto u objetivo. La práctica de la evaluación ha cambiado radicalmente en las tres últimas décadas: se han desarrollado nuevos métodos y enfoques, y ahora se utiliza para proyectos y públicos cada vez más diversos (Jones, 2021).

La fase de evaluación de la gestión o el uso de datos de ERT consiste en examinar los mismos para comprender la información recopilada a través del monitoreo. Esta etapa es crucial para entender y mejorar la eficiencia y la efectividad de los servicios de transporte y para tomar decisiones informadas. La evaluación puede ser realizada a través de técnicas de análisis de datos, métricas de rendimiento y otros indicadores de desempeño. Esta planificación se abordará más adelante en el subcapítulo “Estrategia de evaluación y definición de objetivos e indicadores”.

Previo a desarrollar una estrategia de evaluación, se debe tener una idea clara de cómo se pueden responder las preguntas mostradas en la **tabla 2**. En ella se encuentran algunas de las respuestas para el proyecto “Rodando Ayuda”, un proyecto dentro del proyecto Ideamos llevado a cabo en 2020 durante la pandemia por COVID-19.

Tabla 2. Preguntas a realizar para el desarrollo del proceso de evaluación con un ejemplo de las respuestas del proyecto Rodando Ayuda

Pregunta	Respuestas
¿Qué se evaluará?	Rodando Ayuda, un proyecto de envío de despensas a población en situación de vulnerabilidad económica por COVID-19, en el que se priorizó el uso de modos sostenibles como la bicicleta mecánica, de carga, eléctrica y la moto eléctrica.
¿Qué indicadores se utilizarán para juzgar el rendimiento del proyecto?	Sociales: Despensas entregadas, personas beneficiadas, personas repartidoras participantes, contagios. Ambientales: Kilómetros recorridos, emisiones de CO2, ahorro de emisiones. Económicos: Donaciones recibidas, costo por despensas entregadas, ahorro en costos.
¿Qué niveles deben alcanzar los indicadores para que se considere que el proyecto ha tenido éxito?	Sociales: Entregar un mínimo de 500 despensas sin contagios por parte de las personas participantes. Ambientales: Reducir al menos en un 60% las emisiones de CO2 y otros contaminantes de los envíos. Económicos: Alcanzar al menos un monto de 100 mil pesos MXN en donaciones.
¿Qué pruebas indicarán que el rendimiento de los indicadores tienen relación con los objetivos?	Sociales: Evidencias fotográficas, reportes de repartidoras, tablas de excel. Ambientales: Registro de kilómetros, análisis de datos (programación interna). Económicos: Recibos de pagos, caracterización de donaciones en especie.

Fuente: Elaboración propia.

Como se mencionó, durante la evaluación es importante la definición de indicadores clave y éstos pueden variar de acuerdo al actor involucrado. En el caso de una ciudad o un gobierno local, éstos ayudan a tomar decisiones informadas sobre el mejoramiento de la red de transporte y están más relacionados con la planificación urbana y la promoción de modos de transporte más sustentables. En el caso de las ERT, estos indicadores suelen estar más orientados a mejorar la eficiencia en el servicio, mejorar la experiencia de las personas usuarias y reducir los costos de operación, sin que esto excluya la necesidad de evaluar indicadores ambientales, sociales, etc. A continuación, se muestra una tabla con algunos indicadores clave en la etapa de evaluación.

Tabla 3. Ejemplos de indicadores para los actores involucrados en la evaluación de datos de transporte.

Actor involucrado	Categoría	Indicador
Gobierno o Autoridades Locales	Tráfico	<ul style="list-style-type: none"> a. Velocidad de vehículos b. Incidentes de tránsito c. Frecuencia de retorno
	Transporte público	<ul style="list-style-type: none"> a. Frecuencia del servicio b. Capacidad de las unidades c. Popularidad de los modos de transporte
	Viajes	<ul style="list-style-type: none"> a. Patrones de viaje b. Distancias promedio recorridas c. Tiempos de viaje
	Ambientales	<ul style="list-style-type: none"> a. Calidad del aire b. Emisiones de gases contaminantes c. Niveles de ruido
ERT	Eficiencia operativa	<ul style="list-style-type: none"> a. Tasa de puntualidad b. Costos operativos por kilómetros c. Tasa de ocupación de los vehículos
	Satisfacción del cliente	<ul style="list-style-type: none"> a. Niveles de reclamaciones b. Tasa de retención de clientes c. Porcentaje de usuarios frecuentes
	Impacto Ambiental	<ul style="list-style-type: none"> a. Huella de carbono de la flota b. Consumo de combustible c. Eficiencia energética

Fuente: Elaboración propia.

Para llevar a cabo esta etapa, generalmente se comparan los datos obtenidos con los objetivos establecidos y se identifican tendencias y desviaciones. Finalmente, se utiliza la información recopilada para identificar áreas de mejora, así como desarrollar soluciones para aumentar la eficiencia y la efectividad de los servicios de transporte. Con esta información, una ciudad puede identificar áreas en las que se pueden promover modos de transporte más sostenibles, como la bicicleta y la caminata. En la siguiente sección se muestran las actividades a realizar y definir para contar con una estrategia de medición.

Planeación y actividades previas al acceso a la información

Esta sección se centra en las actividades que deben llevarse a cabo antes de acceder a la información proporcionada por las empresas de movilidad. Se abordarán todas las actividades necesarias para elaborar una propuesta de evaluación que mida el impacto del proyecto o intervención, según los objetivos planteados.

Antes que nada, es importante destacar la importancia de comprender el problema que se busca resolver y los posibles impactos directos e indirectos de su solución. En este sentido, se entiende que tanto los objetivos del proyecto, como el plan de trabajo general ya han sido definidos. Sin embargo, también se asume que aún no se cuenta con una planificación detallada del plan de trabajo, por lo que es probable que la estrategia de evaluación a realizar y las actividades posteriores deban ser ajustadas una vez que se obtenga más información acerca del proyecto.

Estrategia de evaluación y definición de objetivos e indicadores

Como primer paso, es esencial desarrollar un plan detallado para medir el cumplimiento de los objetivos del proyecto. Este plan debe incluir los indicadores o KPI a medir, los métodos de recolección de datos, las fechas de recolección y la metodología para analizar la información. Así se garantiza una adecuada medición de los objetivos del proyecto y es posible tomar decisiones informadas para su mejora continua. Además, permitirá tener una planeación adecuada del tiempo y recursos necesarios para su medición. En la figura 4 se encuentran los elementos clave necesarios para tener una estrategia de evaluación del proyecto integral. Los contenidos de esta sección están basados en una revisión de distintas fuentes, como la *Guía para la Elaboración de la Matriz de Indicadores para Resultados* (CONAPO, 2013), donde se presenta una metodología para la definición de indicadores a través de 10 pasos, el Portal de Evaluación de impacto del BID (BID, 2020), donde se incluyen recursos para diseñar, implementar, recolectar datos, analizar y difundir los resultados de evaluaciones de impacto y la guía *Cómo Desarrollar un Plan de Monitoreo y Evaluación* (The Compass for SBC, s/f) que incluye recomendaciones y ejemplos para elaborar un plan de evaluación de una intervención.

Figura 4. Contenidos de la estrategia de evaluación.

Contenido de la estrategia de evaluación



Fuente: Elaboración propia.

Antes del desarrollo de la estrategia de evaluación, se recomienda al lector explorar las secciones cinco y seis para contar con un entendimiento general de las necesidades para el uso de la información, así como los pasos necesarios para esta tarea. A continuación, se da una descripción de cada una de las etapas mostradas en la tabla anterior.

Los **objetivos de evaluación** deben ser claros y estar bien definidos; adicionalmente, se deberá asegurar que cualquier objetivo de medición esté relacionado con los objetivos generales del proyecto. Por ejemplo, un objetivo puede ser “reducir el tiempo de viaje de las personas usuarias del transporte corporativo en un 20%”. Este objetivo específico ayuda a enfocar la evaluación en medir un resultado específico del proyecto y a establecer una meta cuantificable para determinar si el proyecto ha sido efectivo o no en la reducción de los tiempos de traslado.

El **marco de medición** es un componente clave de la estrategia de evaluación, ya que servirá para realizar las peticiones de información a las contrapartes. En esta sección, se definirán los indicadores y las fuentes de datos que se utilizarán para medir el progreso y el impacto del proyecto. Los elementos que deben incluirse en el marco de medición son:

- 1. Indicadores:** son las variables que se utilizarán para medir el progreso y el impacto del proyecto. Deben ser claros, relevantes y medibles.
- 2. Fuentes de datos:** son los lugares o los medios de donde se obtendrán los datos necesarios para medir los indicadores. Puede ser información generada por las ERT desde sus apps o vehículos, encuestas, registros administrativos, entrevistas, entre otros.
- 3. Periodicidad de la medición:** se refiere a la frecuencia con la que se medirán los indicadores. La periodicidad dependerá de la frecuencia en que el indicador se actualiza, de la capacidad de análisis así como de las necesidades de medición.
- 4. Responsables de la medición:** son las personas o las entidades encargadas de recolectar y analizar los datos necesarios para medir los indicadores.
- 5. Niveles de desagregación:** se refiere al nivel de detalle en el que se medirán los indicadores. Pueden ser a nivel general, regional, de género, entre otros.
- 6. Información general:** incluye otros aspectos del indicador, por ejemplo su relación con los objetivos, el resultado esperado, el tipo de indicador, entre otra información relevante de acuerdo a las necesidades del proyecto.

El desarrollo de los puntos anteriores debe ser diseñado cuidadosamente para asegurar la validez y la fiabilidad de los datos obtenidos. Además, debe ser actualizado periódicamente a lo largo del proyecto para asegurar que sigue siendo relevante y útil para medir el progreso y el impacto del proyecto. La definición final de los indicadores deberá estar orientada a cómo se planearán, evaluarán o monitorearán las intervenciones y no en agotar la lista de indicadores sin que estén relacionados en alguna medida con los objetivos. Para que un indicador sea exitoso debe ser construido con un público objetivo claro, debe tener una base científica, debe basarse en información disponible, debe de ser sensible al cambio del tema en cuestión, fácilmente entendible y, por último, debe ser usado para la toma de decisiones. Además de los puntos anteriores, es importante que los indicadores sean interpretados y presentados ante la audiencia objetivo y transmitir de una manera más eficiente el mensaje, los impulsores del cambio, los escenarios probables y las acciones que pueden tomarse para incentivar o desincentivar los cambios (Biodiversity Indicator Partnership, 2010).

También se deberá considerar la necesidad de contar con una línea base de información previa a la intervención o implementación de las acciones del proyecto, con el fin de tener una línea de referencia para la evaluación del impacto. Será posible reconsiderar y adaptar los elementos del Marco de Medición según las necesidades y capacidades del equipo de trabajo. En el documento *Impact Evaluations and Development* (2009) de la OCDE se incluye una descripción exhaustiva de las necesidades para la evaluación del impacto de un proyecto. Específicamente, el capítulo “Outline of Principles of Impact Evaluation” incluye recomendaciones útiles para la definición del marco de medición. Por otro lado, el *Manual para la construcción de indicadores* (CONEVAL, 2013) presenta una serie de pasos metodológicos para la construcción de indicadores, desde la definición de los objetivos y resultados esperados, hasta la selección de las fuentes de información y la construcción de los propios indicadores con una metodología SMART².

2 La metodología SMART (Specific, Measurable, Achievable, Relevant, Time-bound) es un marco claro y conciso para la creación de objetivos e indicadores. SMART. Para obtener más información sobre los Objetivos SMART, visite: <https://asana.com/resources/smart-goals>.

Por otra parte, los datos públicamente disponibles son una fuente de información complementaria que puede ser útil para la parte de planeación de proyectos, éstos generalmente son recolectados por agencias gubernamentales o por empresas de tecnología. Ejemplos de este tipo de fuentes son los censos de población y vivienda o el Directorio Nacional de Unidades Económicas (DENUE) del Instituto Nacional de Estadística y Geografía (INEGI). Estos datos pueden ser reutilizados por ésta u otras agencias para transformarlos o generar nueva información, como en el caso del Inventario Nacional de Viviendas, donde el INEGI espacializa la información recolectada durante el censo e incluye nueva información como disponibilidad de banquetas o la accesibilidad a las calles; Otro ejemplo es el caso del Índice de Marginación hecho por el Consejo Nacional de Población (CONEVAL), donde recuperan los datos del Censo de Población y Vivienda y, a través de análisis multivariados, generan información nueva sobre las carencias que sufre la población a escalas estatal, municipal, localidad y a nivel AGEB³ urbano. Esta información permite complementar los indicadores y los análisis hechos para introducir variables sociales, económicas y demográficas y, así, permitir un entendimiento más profundo de los sistemas analizados, su impacto, así como nuevos niveles de desagregación.

Otra fuente de información adicional es aquella recabada por empresas de tecnología. Empresas como Google, Waze, o Uber recopilan datos en tiempo real y los almacenan para su análisis. Esta información después se comparte a través de una API (*Application Programming Interface*), datos tabulados o reportes. Ejemplos de estas aplicaciones son los reportes de movilidad de Uber (<https://movement.uber.com/>), donde a través de los datos de miles de millones de viajes recolectados son usados para obtener estimaciones de tiempos de viaje entre zonas de las ciudades, la velocidad de los viajes dependiendo de la fecha y hora, o mapas de las zonas principales de origen y destino de los viajes. Otro uso de los datos recopilados fueron los reportes de movilidad de Google durante el inicio de la pandemia por COVID-19 (<https://google.com/covid19/mobility/>), donde los datos de ubicación de las personas usuarias de teléfonos inteligentes se procesaron para conocer los patrones de movimiento a diferentes tipos de destino (hogares, lugares de trabajo, de abastecimiento, parques, etc.). Dentro del ITDP México se desarrolló un visualizador de accesibilidad urbana (<https://accesibilidad.ideamos.mx/>) en el que se utilizan los datos de la API de Google (<https://developers.google.com/maps>) para obtener información sobre tiempo de viaje para modelar la accesibilidad en diferentes condiciones de congestión vehicular.

El **Plan de Recolección de Datos** se refiere al conjunto de acciones y pasos necesarios para obtener la información y para acceder a ella. En esta parte del plan se deberá definir la forma en que se tendrá acceso a la información para cada indicador planteado. Se pueden utilizar bases de datos, carpetas de almacenamiento en la nube, servicios de almacenamiento de datos, o bien, el uso o desarrollo de una interfaz de programación de aplicaciones (API, por sus siglas en inglés) para extraer datos de servicios externos. Es importante que se defina claramente el acceso a los datos, incluyendo las limitaciones de privacidad y seguridad, así como la calidad y confiabilidad de los datos obtenidos. En el siguiente capítulo se presenta una descripción de las posibles necesidades de hardware y software.

Por último, en la estrategia de evaluación se deberá tener un **Plan de análisis de datos**, en el que se deberán incluir los métodos y técnicas para transformar los datos obtenidos en los indicadores de medición de impacto. Esto incluye la selección de herramientas estadísticas y de visualización de datos, así como la definición de los criterios y medidas que se utilizarán para interpretar los resultados. También se deben definir los plazos y los responsables de la realización del análisis de datos. Es importante considerar que las organizaciones poseedoras de la información cuentan con agendas y compromisos que suelen no estar sincronizados con los de los proyectos, por lo que es necesario mantener una comunicación clara y constante sobre las necesidades de medición para que la información pueda ser entregada en los plazos requeridos. Para la definición de la metodología para la evaluación de impacto, se recomienda revisar los capítulos tres y cuatro de la guía *Impact Evaluations and Development* (2009) de la OCDE, donde se describen aspectos clave para articular la relación del impacto con la intervención realizada y se detallan diferentes metodologías de análisis de evaluación.

También se deberá tener claro en el Plan de análisis de datos la forma en que se compartirán los resultados a los diferentes *stakeholders*, así como la periodicidad del análisis. Es posible que algunos análisis requieran ser desarrollados en la etapa de planeación pero no en la etapa de monitoreo. Por lo que es necesario definir su periodicidad y su forma de ser compartidos, por ejemplo a través de tableros interactivos o reportes recurrentes.

Dependiendo del tipo y cantidad de datos que se recolectan, puede ser necesario contar con servidores o con computadoras de alta capacidad destinadas al almacenamiento de datos y el cómputo. Es importante destacar que el uso de tecnología y automatización no sólo simplifica y automatiza el proceso de evaluación, sino que también ayuda a garantizar la precisión, fiabilidad y replicabilidad de

3 AGEB: Área Geoestadística Básica: Extensión territorial que corresponde a la subdivisión de las áreas geoestadísticas municipales. Constituye la unidad básica del Marco Geoestadístico Nacional y, dependiendo de sus características, se clasifican en dos tipos: rural o urbana.

los procesos realizados. Además, al contar con las herramientas adecuadas, es posible ahorrar tiempo y recursos valiosos durante el proceso de evaluación. En la siguiente sección se presenta un resumen de las posibles necesidades de hardware y software.

Caso 1: Digitalización del transporte Concesionado

La importancia de la flexibilidad en la estrategia de evaluación

Durante la fase de planeación del proyecto, se desarrolló la estrategia de evaluación estableciendo objetivos, recursos y tiempos necesarios a partir de diferentes hipótesis y posibles líneas de trabajo. Por ejemplo, en un posible piloto a realizar fue incluir una función en la app del sistema para visualizar el nivel de ocupación de las unidades de transporte. Se trabajó en esta funcionalidad durante varios meses: El equipo de Vinden identificó que la complejidad de desarrollo de esta funcionalidad era alta, por lo que decidió no desarrollarse a pesar del trabajo realizado. Durante este proyecto fue necesario adaptar el plan de trabajo y la estrategia de evaluación según las necesidades cambiantes del proyecto y de las contrapartes, según su validación, presupuesto, capacidades del equipo y tiempo de implementación. La flexibilidad para adaptar la estrategia de evaluación fue clave para atender las nuevas necesidades y limitaciones.

Las peticiones de información fueron adaptadas a cada contraparte involucrada en el proyecto, a las necesidades específicas de cada etapa del proyecto y de las capacidades técnicas de los equipos. Una metodología útil para solicitar información a instancias gubernamentales fue contar con una tabla como la RR, donde se solicita la información de manera centralizada y con campos de información adicionales que permiten el seguimiento por las contrapartes. Es importante destacar que, en este proyecto, la solicitud de información fue recurrente y no sólo se realizó al inicio del proyecto, por lo que contar con medios para gestionar estas solicitudes y su cumplimiento fue clave para obtener la información necesaria.

Para la evaluación del proyecto, se definió una estrategia que incluyó la revisión de la información recopilada y la elaboración de informes periódicos que permitieron realizar ajustes en la estrategia según los resultados obtenidos. En el caso de alertas, se utilizaron principalmente datos cuantitativos, como los datos de sensores y GPS de los autobuses de transporte concesionado, mientras que en rutas nocturnas se utilizó información cualitativa, como encuestas y muestras.

Convenios de transferencia de información

La colaboración en proyectos de movilidad entre gobiernos, organismos como ITDP y ERT es fundamental para el desarrollo de soluciones innovadoras y efectivas en el ámbito del transporte. Sin embargo, esta colaboración implica la necesidad de gestionar adecuadamente la información y los datos compartidos entre las partes involucradas. En esta sección se abordan algunos elementos clave para establecer medidas legales para garantizar el acceso a la información y para delimitar el uso que se le puede dar a ésta.

Para el desarrollo de convenios o contratos de colaboración entre empresas y ERT, es necesario contar con la asesoría legal de abogados especializados en la materia. La negociación de los términos de estos acuerdos puede requerir una considerable cantidad de tiempo y esfuerzo para alcanzar un consenso entre las partes involucradas. La privacidad de los datos es uno de los aspectos más importantes a considerar en estos convenios, ya que la información recopilada puede contener datos sensibles y privados de las personas usuarias. Es necesario establecer claramente las políticas y protocolos de privacidad de datos para garantizar la seguridad de la información, así como la protección de los derechos de las personas usuarias. Los contratos también deben incluir cláusulas específicas sobre el uso y divulgación de los datos, y cómo se protegerán y compartirán los derechos de propiedad intelectual (OMPI, s/f).

Acuerdo de confidencialidad o (NDA)

Un acuerdo de confidencialidad (por sus siglas en inglés Non Disclosure Agreement), también conocido como acuerdo de no divulgación, es un contrato legal entre al menos dos entidades que se utiliza para compartir información confidencial o conocimientos con un propósito específico y restringir su uso público. Establece una relación comercial entre las partes involucradas para proteger cualquier secreto comercial o información privada de una empresa. En el contexto de la gestión de datos con ERT, los acuerdos de confidencialidad se utilizan para proteger los datos personales de los participantes de cada empresa al impedir su divulgación y para evitar que las ERT compartan públicamente los datos obtenidos durante la implementación de las estrategias de movilidad.

Contrato prestación de servicios de consultoría

El contrato de prestación de servicios de consultoría tiene como objetivo principal definir las actividades que se realizarán entre las partes involucradas, estableciendo el alcance de los servicios y las obligaciones correspondientes de cada una de ellas. Este contrato debe incluir los gastos y cuotas que se comprometen a realizar, así como los términos y condiciones generales, los derechos de propiedad intelectual y las medidas de confidencialidad correspondientes. Además, es importante que este tipo de contrato también incluya una garantía para asegurar el cumplimiento de las obligaciones acordadas.

Caso 2: MOVIN Reforma

Eficientar el proceso legal y administrativo

Para llevar a cabo este piloto fue necesaria la firma de un contrato de prestación de servicios para cada una de las nueve empresas con las que se desarrolló el Plan de Movilidad Institucional, así como un acuerdo de confidencialidad con las tres ERT participantes. Este acuerdo, a su vez, incluía los alcances de la fase de transferencia de datos con el objetivo de poder monitorear el desarrollo del proyecto y evaluar su impacto una vez terminado. Además, solicitar los datos de manera anonimizada fue útil para evitar retrasos importantes relacionados con el uso de datos personales. Algunos de los puntos más relevantes que incluyeron estos contratos fueron, para el caso de las empresas:

- › Obligaciones de las partes
- › Uso de nombre y logotipos (sólo en caso de ser necesarios para el proyecto)
- › Uso de datos personales
- › Confidencialidad, titularidad de materiales y propiedad intelectual
- › Vigencia, terminación anticipada o incumplimiento

En el caso de las ERT, algunas cláusulas incluían lo siguiente:

- › Periodicidad de la transferencia de datos
- › Formatos de transferencia (JSON, XLSX, CSV)
- › Manejo de conflictos
- › Alcances del soporte técnico
- › Definiciones de los indicadores a compartir

En algunos casos, la transferencia por una API estaba condicionada a un tiempo mínimo de desarrollo, pruebas e implementación, por lo que para este piloto se firmó una carta compromiso en donde se establecieron formas de transferencia manual hasta que se concluyera el desarrollo de la API como sustitución del medio de traspaso.

Finalmente, fue necesario que, además del equipo legal y administrativo involucrado, se hiciera una revisión general de los alcances e indicadores de los contratos por parte del equipo técnico, esto para garantizar que toda la información necesaria para el análisis de datos estuviera incluida y fuera suficiente, precisa, consistente y oportuna.

Necesidades para el uso de la información

Para hacer un uso eficiente y efectivo de los datos de ERT, es fundamental que los diferentes actores involucrados (operadores, autoridades, personas expertas y personas usuarias) puedan compartir información de manera interoperable, es decir, que los datos puedan ser transferidos y utilizados por diferentes sistemas y aplicaciones sin dificultad. Además, es crucial contar con datos de alta calidad, que sean precisos, actualizados y confiables, para que los resultados del análisis y las decisiones que se tomen a partir de ellos sean también de alta calidad.

En esta sección se abordarán también las necesidades de tecnología, incluyendo el software y hardware necesarios para el almacenamiento, procesamiento y análisis de los datos de ERT. Asimismo, se explorarán diferentes métodos de transferencia y procesamiento de información, desde la transferencia manual de datos hasta el uso de API y procesamiento automatizado de datos. Finalmente, se abordarán aspectos importantes del análisis y la interpretación de los datos, así como la necesidad de una comunicación efectiva para compartir los resultados obtenidos y tomar decisiones informadas en proyectos de movilidad.

Necesidades de la información

Los datos a usar deben de cumplir con ciertos requerimientos que garanticen su calidad, accesibilidad y compatibilidad. Estos requerimientos son las necesidades de información que se deben considerar para eficientar y facilitar los procesos a llevar a cabo para la planeación, monitoreo y evaluación. Los requerimientos mostrados a continuación podrán ser discutidos con las contrapartes para ser considerados según las necesidades específicas de cada proyecto.

Interoperabilidad

La interoperabilidad se define como “la capacidad de usar datos de diferentes fuentes en una forma estandarizada y contextualizada” (Steele & Orrell, 2017). Existen dos tipos niveles de interoperabilidad: **i)** la primera es la interoperabilidad semántica que se refiere a la capacidad de que la información obtenida sea útil, con significado y comparable entre sí, aun si viene de distintas fuentes y la **2)** interoperabilidad técnica, que se refiere a la capacidad de que diferentes componentes de software trabajen en cooperación entre ellos, incluyendo ocasiones en que se usen distintas herramientas de análisis.

Una de las principales tareas para alcanzar la interoperabilidad es contar con una estructura organizada de manera que los datos sean leíbles tanto para humanos como para computadoras. Algunas de las convenciones más importantes para el uso de datos son: **(I)** usar archivos de formato abierto, evitando así el uso de archivos con formato propietario, por ejemplo de Excel. El uso de formatos abiertos, por ejemplo CSV, JSON, XML, GeoJSON, GeoPackage, ARROW, etc.⁴, cuenta con documentación comprensiva y pueden ser leídos por todos los software de análisis de datos. El archivo tipo CSV es probablemente el formato más fácil de utilizar para personas expertas y no expertas, es un formato de texto plano que puede ser leído por cualquier proyecto. **(II)** Usar nombres de archivos y de campos en minúsculas, evitando el uso de acentos y caracteres especiales (ñ o ç). **(III)** Usar formatos ISO para fechas, es decir, formato año-mes-día. Entre otros, como asegurarse de ser consistente con los nombres de campos o columnas y usar un formato común para datos faltantes, por ejemplo NaN.

Un aspecto particularmente relevante para las ERT son los formatos tipo GTFS (General Transit Feed Specification), GBFS (General Bike Feed Specification) y MDS (Micromobility Data Specification). Su utilidad yace en estandarización de datos, lo que permite que cualquier persona o software pueda leerlos, analizarlos, generar visualizaciones y modificarlos para sus necesidades. Éstos son estándares de datos de formato abierto cuyo objetivo principal es que los operadores de transporte público y de micromovilidad (MDS) puedan compartir sus itinerarios, rutas, estaciones, tarifas, y otras características del transporte público (GTFS) y de los sistemas de bicicletas compartidas (GBFS) con gobiernos, personas usuarias y plataformas de datos como Google Maps u otros planeadores de viajes.

Calidad de los datos

El objetivo principal de esta fase es garantizar que los datos sean precisos y confiables, lo que a su vez puede mejorar la calidad de los resultados y reducir el riesgo de tomar decisiones basadas en información incorrecta. Para garantizar la calidad de los datos, es necesario examinar sus características, algunas de ellas se muestran en la **tabla 4** (Morales & Orrell, 2018).

⁴ Los formatos Geojson y Geopackage son formatos de datos geoespaciales que describen objetos, eventos u otras características con una ubicación sobre o cerca de la superficie de la tierra. Este tipo de datos combina información de ubicación (coordenadas), sus atributos y/o información temporal (What Is Geospatial Data?, s/f)

Tabla 4: Criterios de calidad y características a examinar en la fase de QA

Nombre	Descripción
Integridad	La medida en que el modelo captura todas las entidades y relaciones pertinentes.
Simplicidad	La capacidad de proporcionar una representación concisa de todos los aspectos clave de los datos con pocas entidades, relaciones y atributos y con una pequeña cantidad de instancias de los mismos.
Flexibilidad	La habilidad para agregar nuevas entidades y atributos sin comprometer la integridad de los datos y aplicaciones existentes.
Legibilidad	La adherencia general a convenciones de nombres, así como la calidad general de las definiciones.
Precisión	Cercanía de agrupación entre un grupo de resultados.
Exactitud	Cercanía de una medición al verdadero valor.
Consistencia	Sin conflictos entre los valores de las variables entre diferentes fuentes de datos.
Únicos	Sin registros repetidos.
Oportunos	Los datos han sido medidos y actualizados para mantener su disponibilidad para usarse cuando sea necesario.

Fuente: (Morales & Orrell, 2018)

Esta validación se conoce como Quality Assurance (QA), es crucial en la exploración de datos ya que asegura que los datos utilizados sean de alta precisión. En esta fase, se lleva a cabo una serie de pruebas para identificar cualquier problema que pueda haber surgido en la recopilación, limpieza o análisis de los datos. Las pruebas pueden incluir verificación de la coherencia de los datos, la precisión de las variables, la ausencia de valores nulos, la presencia de valores extremos y la comparación de los resultados con otras fuentes confiables.

Diccionario de Datos

Una buena práctica al trabajar con datos es solicitar un diccionario de datos de la información adquirida. Se trata de un documento que contiene los metadatos con la información general de la información que se comparte, incluyendo el nombre de los campos, su descripción, periodicidad, tipo de dato y, en algunos casos, las reglas de negocio para la generación de la información de cada campo.

Contar con un diccionario de datos es importante para lograr un entendimiento rápido de la información con la que se trabaja, comprendiendo así las relaciones que existen entre la información, las reglas con las que se genera y el significado de cada campo. Esto facilita la interoperabilidad de los datos entre diferentes sistemas y organizaciones.

Seguridad

Proteger la integridad y el acceso a la información es esencial para la gestión y uso de los datos. Es fundamental salvaguardar la privacidad de los datos y de las personas, evitar ataques cibernéticos y proteger la infraestructura tecnológica. A continuación, se presentan prácticas recomendadas para garantizar la seguridad de la información, tanto en computadoras personales como en servidores en la nube. Es importante tener en cuenta que la seguridad es responsabilidad de todos los actores involucrados. Al aplicar estas medidas y compartirlas con las contrapartes, es posible mitigar algunos riesgos. De acuerdo a cada proyecto puede ser necesario garantizar la seguridad de los sistemas informáticos de manera efectiva y confiable, así como contar con la orientación y asesoría de un profesional en el campo de la seguridad informática.

1. Mantener el software actualizado. Actualizar el software es una manera de prevenir ataques cibernéticos que aprovechan vulnerabilidades en el software.
2. Contar con medidas de seguridad de red. Implementar medidas de seguridad como firewalls y

sistemas de detección o prevención de intrusiones para proteger la red contra ataques externos.

3. Implementar medidas de seguridad en la transferencia de datos. En caso de que se necesite compartir los datos utilizar medios de comunicación encriptados, y sólo permitir el acceso a las personas usuarias autorizadas para el uso de los datos con contraseñas seguras o limitando el acceso a través de las herramientas utilizadas.
4. Guardar cautela al abrir vínculos en internet o en archivos adjuntos. Evitar dar clic en páginas de internet o en archivos adjuntos sospechosos. A menudo estos sitios y archivos pueden contener virus o llevar a páginas capaces de robar información local. Es recomendable no abrir los sitios ni abrir archivos, a menos de que exista confianza en el remitente.
5. Bloquear el acceso a la información cuando no se encuentre en uso: esto aplica tanto al acceso físico como al acceso virtual. De preferencia mantener computadoras, tabletas, papeles y memorias USB en lugares seguros y bloquear la computadora con contraseña en caso de alejarse de ella.
6. Realizar respaldos de la información. Mantener una copia separada de la información en un lugar seguro, ya sea de manera física en disco duro o en la nube, permite recuperar el acceso a la información en caso de pérdidas accidentales.

Necesidades de tecnología

Una vez que se cuenta con un plan de monitoreo y evaluación y con los datos necesarios para medir, se requiere contar con las herramientas adecuadas para su análisis. En esta sección, se mencionan las herramientas más comunes para el análisis de datos y sus usos. También se describen de manera rápida las opciones de software que pueden ser usadas para la automatización de los análisis y, por último, se presenta una descripción de las herramientas de software y hardware que pueden ser requeridas.

Herramientas de programación

Los lenguajes de programación son la herramienta más poderosa para el análisis de datos debido a la flexibilidad que ofrecen. Su flexibilidad yace en la capacidad de la persona usuaria para reusar los análisis debido a que al programarlos en un proceso automatizado, éste puede reusarse para los mismos análisis o adaptarse y ampliarse para nuevos. El código también permite que otras personas puedan utilizar o auditar el código y los análisis, lo cual realza la transparencia del proceso. En la **tabla 5** se encuentra un resumen de las principales características, herramientas y funciones para el uso de lenguaje de programación en proyectos de análisis.

Tabla 5: Tabla resumen con las principales características de los lenguajes de programación

Lenguaje	Descripción	Librerías principales	Entorno de desarrollo (IDE)
Python	Lenguaje ampliamente usado que permite la automatización del procesamiento y análisis de datos	Pandas: manipulación de tablas. Numpy, Scikit: análisis estadístico. Geopandas: manipulación y análisis de información geoespacial	Jupyter, Hex
R	Lenguaje diseñado originalmente para el análisis estadístico.	Tidyverse: Colección de librerías diseñadas para la ciencia de datos.	RStudio
Sql	Permite comunicarse con bases de datos y hacer consultas, así como generar estadísticas básicas a partir de los datos almacenados de una manera computacionalmente eficiente.	PostGIS: manipulación y análisis de información geoespacial.	Dbeaver Adminer PgAdmin

Fuente: Elaboración propia.

Estos lenguajes a menudo se complementan con **entornos de desarrollo integrado** (IDE, por sus siglas en inglés) como Rstudio o Jupyter. En la figura 5 se encuentra una captura del entorno Rstudio. Los IDE permiten interactuar con el código de una manera más fácil y visual, contando con una interfaz que permite visualizar el código y el resultado, por ejemplo gráficas o tablas. Además, facilitan la depuración

del código gracias a que cuentan con espacios separados para el código fuente y los resultados y/o gráficos desarrollados. También cuentan con herramientas para estilizar y acomodar el código que permite mejorar su legibilidad.

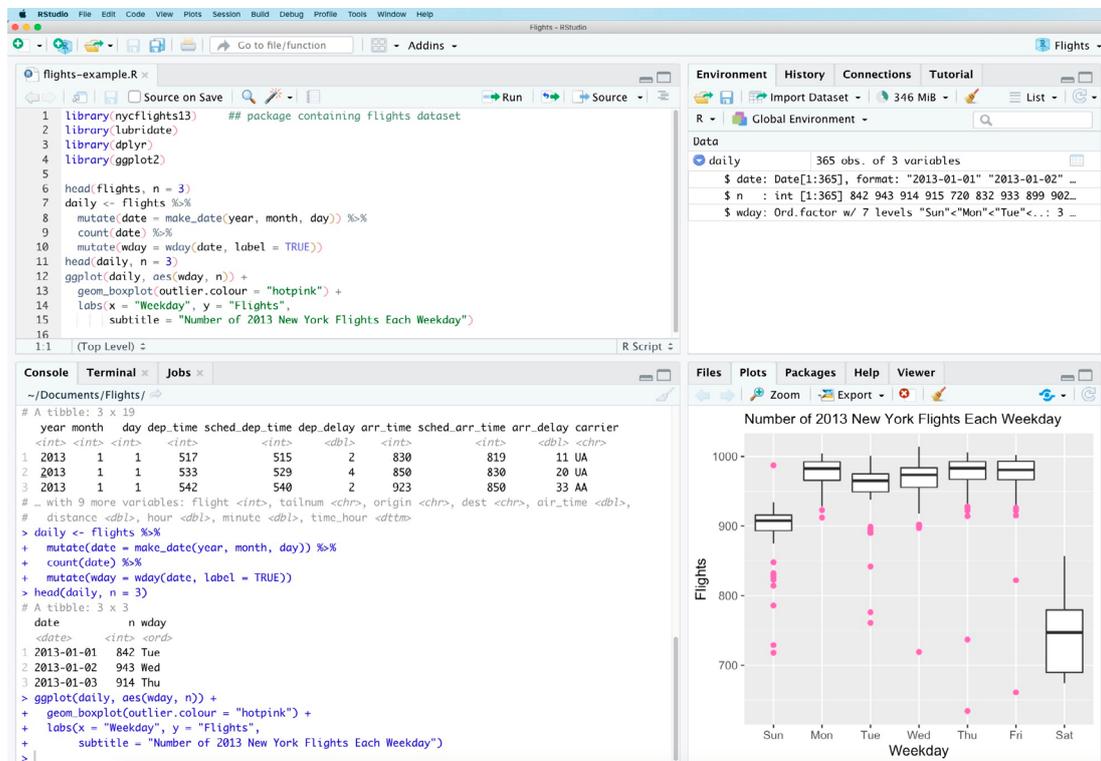
Una de las herramientas de software libre más populares es Anaconda, una plataforma de código abierto para el lenguaje de programación Python y R, diseñada para la ciencia de datos y la computación científica en general. Proporciona un gestor de paquetes (conda) que facilita la instalación y gestión de librerías que contienen software estadístico, de visualización, entre otros. Además, incluye una interfaz gráfica de usuario (GUI) que contiene aplicaciones instaladas de forma predeterminada, algunas de ellas: Jupyter Notebook, RStudio, PyCharm, Spyder, Orange 3 e IBM Watson.

La selección del lenguaje para el proyecto dependerá de las capacidades y conocimientos del equipo en cada lenguaje. En general, tanto R como Python pueden utilizarse para llevar a cabo las tareas de automatización necesarias para el proyecto. Sin embargo, es importante destacar que R tiene una fortaleza en el campo estadístico de alta especialidad, mientras que Python se destaca en la automatización y la integración con otros servicios.

Software de análisis geoespacial

El análisis geoespacial es una herramienta clave en proyectos de movilidad, ya que permite entender y visualizar patrones y tendencias en el comportamiento de las personas usuarias o de los servicios en el espacio y en el tiempo. Existen diversas herramientas para llevar a cabo este tipo de análisis, siendo algunas de las más destacadas ArcGIS y Qgis, ambas herramientas son similares y permiten realizar análisis geoespaciales y visualizaciones tipo mapa a través de una interfaz gráfica amigable, con la diferencia de que Qgis es un software de código abierto y gratuito y funciona en cualquier tipo de sistema operativo. Existen también librerías en Python (Geopandas), R (sf, terra) y SQL (PostGIS) que permiten realizar los mismos tipos de análisis a través de código, la ventaja principal de usar estas librerías es que los análisis y procesos son replicables y automatizables.

Figura 5. Captura del entorno de programación IDE Rstudio



Necesidades de Hardware

La evaluación, monitoreo y diagnóstico de proyectos de movilidad requieren del uso intensivo de datos, por lo que es fundamental contar con la infraestructura adecuada para su procesamiento y almacenamiento. En este apartado se establecerán las necesidades de cómputo y almacenamiento requeridas para llevar a cabo estas tareas de manera efectiva y eficiente.

Servidores para cómputo de datos

Los servidores de cómputo proveen la capacidad de procesar grandes volúmenes de datos mediante algoritmos y software especializados. Los servidores deben tener una alta potencia y velocidad para realizar operaciones complejas y rápidas sobre los datos. Además, deben tener una alta disponibilidad y recuperación ante posibles fallas o interrupciones. Los servidores de datos podrán realizar los procesos de cómputo de manipulación de datos mostrados en la sección 6 de este documento y, además, podrán almacenar archivos y bases de datos.

Uno de los primeros elementos a considerar para validar la necesidad de contar con servidores, y no realizar análisis en computadoras personales, es el volumen y la complejidad de los datos que se manejan. Los proyectos de movilidad suelen involucrar grandes cantidades de datos provenientes de diversas fuentes y formatos, que requieren operaciones complejas y rápidas para su integración y análisis y, adicionalmente, suele ser necesario realizar análisis de manera recurrente, por lo que contar con la capacidad de realizar tareas de cómputo de manera ininterrumpida es crucial. Estas operaciones pueden sobrepasar la capacidad y la velocidad de una computadora personal, generando demoras o errores en los resultados. Existen tres opciones principales para el desarrollo de procesos de alta intensidad de cómputo:

Servidor local: es posible realizar los procesos con un servidor físico. Un servidor local tiene la ventaja de que no depende de una conexión a internet para acceder a los datos, y que ofrece un mayor control y seguridad sobre los mismos. Sin embargo, un servidor local también tiene inconvenientes, como el costo inicial de adquirir el hardware, las capacidades técnicas para su configuración y el espacio necesario para alojar el dispositivo, y el mantenimiento y soporte técnico requeridos.

Servidor en la nube (Infraestructura como Servicio, IAAS): Una opción para optimizar el uso de los servidores es el cómputo en la nube, que consiste en acceder a recursos informáticos alojados en internet mediante proveedores especializados. El cómputo en la nube ofrece ventajas como la escalabilidad, la flexibilidad y el ahorro de costos. Sin embargo, también implica desafíos como la seguridad, la privacidad y la latencia de los datos (el tiempo necesario para que los datos se actualicen). Una opción para optimizar el uso de los servidores es el cómputo en la nube, que consiste en acceder a recursos informáticos alojados en internet mediante proveedores especializados (“Cloud-Based or In-House Servers?”, 2019). Algunos proveedores de estos servicios son AWS (Amazon Web Services) con su servicio EC2, Google Compute Engine, Azure Virtual Machines de Microsoft, Digital Ocean con su servicio de servidores virtuales, entre otros.

Cómputo en la nube sin servidor (Plataforma como Servicio, PAAS): Se trata de un modelo de computación en la nube para ejecutar código, sin necesidad de contar con servidores. El cómputo sin servidor ofrece ventajas como la escalabilidad, la flexibilidad y el ahorro de costos. Además, elimina las tareas de gestión y mantenimiento de la infraestructura, como las actualizaciones del sistema operativo y los parches, la gestión de la seguridad, la planificación de la capacidad, el monitoreo del sistema, entre otros. El cómputo sin servidor no significa que no haya servidores involucrados, el nombre se refiere a que los servidores son invisibles para el desarrollador, que no los ve, no los administra ni interactúa con ellos de ninguna manera. El proveedor de servicios en la nube se encarga de asignar y liberar los recursos informáticos según la demanda cuando se ejecuta el código.

La selección de la opción de cómputo a usar deberá ser guiada por un profesional de datos, considerando el presupuesto y duración del proyecto, así como los criterios clave mostrados en la **tabla 6**. Una de las mayores ventajas del uso de cómputo sin servidor para este tipo de proyectos, es que usualmente los procesos de datos no se realizan continuamente, sino con una frecuencia programada, por ejemplo cada hora o cada día, por lo que contar con un servidor para realizar estas tareas, que normalmente sólo se ejecutan por varios minutos al día, no es conveniente debido a los costos. En este sentido, usar servicios de cómputo sin servidor, por ejemplo AWS Lambda o Google Cloud Functions, puede significar una reducción de costos, ya que el costo por el servicio sólo se realiza por el tiempo en que se ejecuta el proceso medido en milisegundos de ejecución y en los recursos de cómputo requeridos. Por otro lado, si el tiempo de ejecución es alto, o es demasiado frecuente la ejecución de código, es posible que contar con un servidor físico o en la nube sea una mejor opción.

Tabla 6: Comparación de las opciones de cómputo según seis criterios clave

Criterio	Servidor local	Servidor en la nube	Cómputo sin servidor
Escalabilidad	Limitada por hardware	Escalable en función de la capacidad contratada	Escalable automáticamente en función de la demanda
Mantenimiento	Responsabilidad del usuario	Responsabilidad compartida con el proveedor	Responsabilidad del proveedor
Costos	Altos costos iniciales de hardware y mantenimiento	Costos variables en función del uso y capacidad contratada	
Configuración	Requiere configuración y administración por parte del usuario	Configuración y administración por parte del usuario o proveedor	Configuración y administración mínima por parte del usuario
Confiabilidad	Dependiente del hardware y el entorno del usuario	Confiabilidad alta gracias a la redundancia y copias de seguridad	Confiabilidad alta gracias a la tolerancia a fallos

Fuente: Elaboración propia.

Almacenamiento de datos

El almacenamiento es uno de los componentes clave de esta infraestructura, ya que se encarga de guardar los datos en un lugar accesible y protegido. El almacenamiento debe ser capaz de adaptarse a las necesidades cambiantes de los proyectos de movilidad, que pueden generar o requerir diferentes tipos y cantidades de datos según sus objetivos y características. Además, el almacenamiento debe ser rentable y escalable, es decir, que permita ajustar la capacidad según la demanda sin incurrir en costos innecesarios o desperdiciar recursos.

Los datos de los proyectos de movilidad pueden tener diferentes formatos y estructuras según su origen y propósito. Una forma de almacenar estos datos es mediante archivos, como los descritos en la sección “Interoperabilidad”, guardados en un disco duro local o en la nube. Los archivos permiten guardar y transferir datos fácilmente, pero tienen algunas limitaciones para consultar y manipular los datos, como el volumen de datos que pueden almacenar. El almacenamiento de datos puede realizarse de tres maneras: **(I)** almacenamiento de archivos en computadoras locales, **(II)** almacenamiento de archivos en servidores y, por último, **(III)** almacenamiento a través de servicios de almacenamiento donde se cuenta con una interfaz para cargar y descargar archivos almacenados en un centro de datos a través de un portal web. Estos servicios mantienen toda la infraestructura y únicamente se paga por la cantidad de información almacenada, el tipo de almacenamiento y por la frecuencia de acceso. Algunos de los proveedores de almacenamiento en la nube más conocidos son Amazon S3, Microsoft Blob, Google Storage, entre otros. A continuación, se muestra una tabla comparativa de los servicios. En última instancia la selección del proveedor dependerá del costo de uso y de la experiencia del equipo de trabajo con cada servicio.

Tabla 7. Proveedores principales de almacenamiento en la nube y sus características principales

Proveedor de almacenamiento en la nube	Almacenamiento gratuito	Precio por almacenamiento (USD)
Amazon Web Services	5 GB	\$0.023/GB al mes
Microsoft Azure	5 GB	\$0.0184/GB al mes
Google Cloud Storage	15 GB	\$0.020/GB al mes
IBM Cloud	25 GB	\$0.0227/GB al mes

Nota: el costo mostrado se refiere al almacenamiento tipo estándar y está actualizado a marzo del 2023.

El almacenamiento de datos en servidores o computadoras personales tiene limitaciones en cuanto a la capacidad de almacenamiento y la escalabilidad, lo que puede ser un problema cuando se requiere almacenar grandes cantidades de datos. Además, la seguridad y la gestión de copias de seguridad pueden ser desafiantes para las personas usuarias sin experiencia. En contraste, los servicios de almacenamiento en la nube como los descritos en la **tabla 7** ofrecen una solución escalable y segura para almacenar grandes cantidades de datos, sin preocuparse por la capacidad y la gestión de la infraestructura subyacente. Además, los servicios en la nube ofrecen flexibilidad y facilidad de acceso a los datos desde cualquier lugar y en cualquier momento.

Otra opción para guardar la información es a través del uso de bases de datos, que son conjuntos de datos organizados en forma digital para una o más personas usuarias. Las bases de datos pueden tener diferentes modelos según la forma en que se relacionan los datos, como relacional, jerárquico o no relacional. Están diseñadas para facilitar el acceso, la búsqueda y el análisis de los datos mediante un lenguaje especializado llamado SQL. Ofrecen más ventajas que los archivos en términos de eficiencia, seguridad e integridad, sin embargo, para su uso es necesario contar con conocimiento técnico para su configuración y con servidores destinados para esta tarea; además, su tiempo de implementación suele ser mucho mayor. Las bases de datos usadas comúnmente son MySQL y Postgres, ambas herramientas son gratuitas y permiten ser instaladas en servidores en la nube o locales como los descritos en la sección anterior.

También es posible usar servicios de bases de datos en la nube, por ejemplo AWS RDS (Relational Database Service), Azure SQL, Google SQL o Firebase o IBM DB2. Todos estos servicios simplifican su configuración, mantenimiento y escalabilidad, ofreciendo una base de datos como un servicio donde la personal usuaria sólo se encarga de administrar el software y no su infraestructura o configuración. Estos servicios suelen ser más costosos que usar servidores para el mismo propósito, y sólo se recomiendan cuando se requiere una alta flexibilidad de escalamiento en la infraestructura y es necesario contar con medidas de seguridad y recuperación de archivos, aunque estas dos últimas también pueden integrarse en bases de datos gestionadas manualmente, pero con un esfuerzo significativo.

La selección sobre si es necesario usar una base de datos en el proyecto dependerá principalmente de las capacidades del equipo, del volumen de datos a gestionar, del nivel de complejidad de los datos, es decir, qué tan relacionados están entre sí, de la frecuencia de actualización y de la seguridad requerida. También será necesario preguntarse por cuánto tiempo se requiere almacenar la información; si se trata de un proyecto corto, es posible que la mejor opción sea usar servicios en la nube. La decisión sobre su uso deberá ser guiada por un profesional de datos, considerando los criterios antes mencionados.

Caso 3: Moveup

Selección del hardware y software adecuadas para el piloto

La ERT AllRide desarrolló una API con el objetivo principal de monitorear el uso de las rutas universitarias diariamente durante los tres meses de implementación. También, para tomar medidas que aumentaran su uso o también para ajustar el servicio de acuerdo a las necesidades de oferta y demanda de la comunidad universitaria.

Dado que era necesario generar información sobre los viajes de manera automatizada para que los líderes del proyecto pudieran tomar decisiones más informadas, se decidió programar un proceso de limpieza, procesamiento y análisis de la información a través del servicio AWS Lambda. La información requería actualizarse sólo una vez al día, por lo que no fue necesario contar con un servidor dedicado para esta tarea. Además, se eligió almacenar la información generada diariamente en formato JSON en el servicio AWS S3; esto permitió una implementación rápida del proceso sin la necesidad de una base de datos.

El uso de un servicio de cómputo sin servidor resultó beneficioso al evitar la configuración de un servidor local o en la nube, lo que significó un ahorro de costos considerable. El costo por el procesamiento junto con el de almacenamiento de archivos no superó los 10 USD durante los tres meses de monitoreo. Asimismo, el uso de servicios en los que no se requiere gestionar la infraestructura permitió ahorrar tiempo en esta tarea y, por tanto, centrarnos en la optimización y análisis de los resultados obtenidos.

Para la visualización de los datos, se seleccionó la plataforma HEX, similar a Jupyter, que permite compartir webapps interactivas con visualizaciones y análisis para público no técnico. Con esta plataforma, fue posible ahorrar tiempo en la creación de un dashboard de datos y compartir resultados actualizados con los líderes del proyecto. Es importante destacar que aunque HEX es una herramienta de pago, ITDP cuenta con una licencia gratuita debido a su condición de organización sin fines de lucro.

Es necesario considerar que para la configuración y uso de servidores, así como el desarrollo de procesos automatizados, se requiere un equipo humano capacitado y multidisciplinario que pueda realizar las siguientes tareas:

- › Diseñar y planificar el proceso de datos, definiendo las fuentes de datos, la infraestructura y las herramientas necesarias para el procesamiento y análisis de los datos.
- › Implementar y gestionar los procesos de datos, realizando las tareas de ingesta, transformación y almacenamiento de los datos. Esto implica tener conocimientos en programación, configuración de bases de datos y sistemas operativos.
- › Analizar e interpretar los resultados, aplicando metodologías apropiadas para medir el impacto y el desempeño de las soluciones implementadas. Esto implica tener conocimientos en estadística, aprendizaje automático y visualización de datos.
- › Comunicar y reportar los resultados del pipeline de datos a los diferentes actores involucrados en el proyecto. Esto implica tener habilidades en redacción, presentación y gestión de la información generada.

Transferencia, procesamiento y análisis de la información

El uso de datos de empresas de movilidad implica un proceso de lectura, transformación y análisis de información que puede ser complejo y demandante. En esta sección se describirán los pasos principales para transferir o leer los datos de las ERT, explorar la información, procesarla y por último analizarla e interpretarla. El uso de herramientas de programación no es estrictamente necesario para realizar estas tareas, pero puede mejorar la eficiencia y la calidad del proceso, al conjunto de scripts que realizan estas tareas se le denomina Data Pipeline o Canalización de Datos. Un Data Pipeline es un proceso que transfiere automáticamente datos del origen al destino. Es decir, es una serie de pasos que se realizan de manera automatizada para mover datos de un lugar a otro, aplicando ciertas transformaciones en el camino (Snowflake, s/f). En caso de que no se elija o sea posible automatizar el pipeline de datos, los siguientes pasos tendrán que hacerse de manera manual a través de hojas de cálculo y otras herramientas de manera reiterativa.

Figura 6: Diagrama del proceso sugerido de análisis de datos.



Fuente: Elaboración propia.

Transferencia de información

La transferencia de datos es el proceso de trasladar datos de un lugar a otro, estos datos transferidos pueden transformarse en el proceso de migración o llegar a su destino tal cual. Cuando esta fase de transferencia da lugar a dos copias de los datos, se habla de una “replicación de datos”. En cambio, si la fuente de datos original va a quedar obsoleta, se denomina “migración de datos”.

Cualquier transferencia de datos conlleva un cierto grado de riesgo. Existe el riesgo de pérdida de datos, el riesgo de corrupción de datos y la exposición potencial a terceros. Al planificar una transferencia de datos, se deben tener en cuenta los aspectos mostrados en la **tabla 8**.

Tabla 8. Consideraciones durante el proceso de transferencia manual o automática

Característica	Descripción
Seguridad	Los datos son más vulnerables cuando están en tránsito, podrían ser interceptados por un tercero que podría extraer información sensible de la exportación.
Disponibilidad	Los datos deben estar disponibles para todas las personas usuarias y procesos. Esto significa que el origen y el destino deben actualizarse según las necesidades del proyecto. Las transferencias asíncronas suelen impactar en menor medida el rendimiento.
Fiabilidad	Cualquier tipo de transferencia periódica de datos debe seguir el calendario de forma fiable. Una interrupción de la programación puede provocar un fallo del sistema o la pérdida de datos del repositorio destino.
Eficacia	La transferencia de datos conlleva un coste en términos de recursos, cuando se utiliza un servicio en la nube (como AWS), también hay un coste financiero. Un uso ineficiente de los recursos puede limitar la viabilidad y sostenibilidad del proyecto.
Latencia	Si la transferencia de datos es ineficiente, la latencia (el tiempo que tarda un dato en ser transmitido) se incrementará, lo que puede resultar en retrasos en la toma de decisiones y en la respuesta a eventos críticos, como en el caso de datos de transporte en tiempo real.
Conformidad	Los requisitos de conformidad aseguran que la transferencia cumpla con las regulaciones y estándares aplicables. Esto es fundamental para garantizar la seguridad y privacidad de los datos, así como para prevenir sanciones y multas.

Una transferencia de datos implica al menos dos pasos: En primer lugar, se obtienen los datos de la fuente original, lo que se denomina extracción. Después, los datos se escriben en el destino, proceso conocido como carga. Estos pasos pueden realizarse de forma manual o automática, de acuerdo a las necesidades del proyecto y también al tipo de origen de datos. Una transferencia manual debe usarse cuando: **(I)** no es posible automatizar el proceso, por ejemplo cuando la información está almacenada como archivos tipo CSV o similares en computadoras o servidores, o **(II)** se trata de una transferencia puntual donde el beneficio de automatizar la tarea no resulta tan claro, debido al tiempo necesario para la automatización del proceso. La exportación manual podría ser útil en proyectos donde no se cuenta con grandes volúmenes de datos, pues permitirá ahorrar costos de transferencia a través de servicios de terceros.

Por otro lado, también es posible realizar una transferencia automática de la información; ésta será necesaria cuando: **(I)** la transferencia de la información se necesita de manera continua, por ejemplo cada hora, día o semana; **(II)** cuando la información está almacenada en bases de datos, y **(III)** cuando la contraparte, por ejemplo la ERT, está en la capacidad de generar una API para la lectura de la información. También será necesario evaluar la complejidad de la transferencia de información, los requerimientos de seguridad y el volumen de la información. A continuación, se describen algunos procesos de automatización:

- › **Codificación:** en algunos casos, puede ser necesario escribir una pequeña aplicación o script para extraer datos de una fuente de datos. Este enfoque es necesario cuando la fuente de datos no tiene una API disponible o cuando la API existente no proporciona la funcionalidad requerida. Esta aplicación suele escribirse en Python o R. Algunas técnicas de codificación incluyen *web scraping* o *web crawling*, que consisten en la captura de información directamente desde la pantalla de una aplicación, desde un sitio web o desde un archivo HTML (figura 7). Es importante tener en cuenta que el uso de estas técnicas debe hacerse siempre de manera ética y legal, respetando los términos de servicio y las leyes de protección de datos que correspondan.

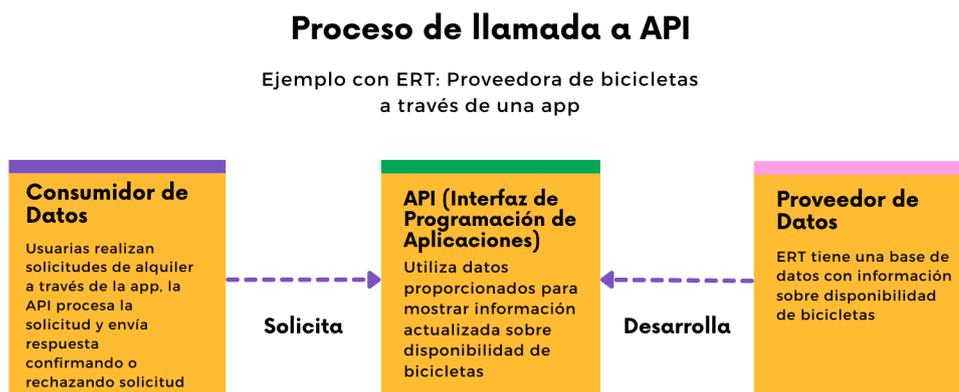
Figura 7. Diagrama de transferencia por Scraping.



Fuente: Elaboración propia.

► **Transferencia desde una API:** Una “Interfaz de Programación de Aplicaciones” (API, por sus siglas en inglés) permite que varias aplicaciones se comuniquen entre sí. Esencialmente, una API comprende un conjunto de funciones que permite a las aplicaciones acceder a datos e interactuar con software externo, sistemas operativos y otros servicios. En palabras más sencillas, una API funciona como intermediaria entre dos máquinas o aplicaciones diferentes que intentan compartir información (figura 8). Las API pueden ser diseñadas para distintos propósitos, como el monitoreo del tráfico, la obtención de indicadores de transporte o la planificación de rutas. Los datos suelen exportarse en forma de archivo, como JSON, XML o CSV. Para leer y consumir datos de una API, es necesario en primer lugar que la empresa dueña de la información la haya desarrollado, tener acceso a su documentación para comprender su estructura y saber cómo enviar y recibir solicitudes de información. Es importante tener conocimientos básicos de programación y familiaridad con algún lenguaje para desarrollar un script que consuma su información de manera automatizada. Además, se necesitan credenciales de acceso válidas, como una API Key, para autenticarse en la API y acceder a sus datos

Figura 8. Diagrama de llamada a una API para el caso de una ERT.



Fuente: Elaboración propia.

› **Transferencia desde bases de datos:** Si la información a usar se encuentra dentro de una base de datos y no se cuenta con una API, es posible desarrollar un proceso automatizado para leer la información de la misma. La transferencia de datos suele automatizarse sobre todo cuando se trabaja con big data. En la **tabla 5** se encuentra una descripción de las herramientas que pueden ser usadas para la lectura de bases de datos. En ella se pueden crear sentencias en SQL que se ejecuten de manera automática y se almacenen los resultados en una tabla o archivo de salida.

Caso: MOVIN Reforma

Estrategias de transferencia de información

Para medir los viajes en este piloto, el ITDP estableció como requisito para las ERT el desarrollo de una API que permitiera monitorear las estrategias de movilidad a través de la información generada por los diferentes servicios. Para el aprovechamiento de la información requerida fue necesario crear un pipeline de datos con un proceso de limpieza, procesamiento y carga de información.

Para las estrategias de Dezza (sistema de bicicletas sin anclaje) y AllRide (plataforma de carpool) sí fue posible contar con una API, en ambos casos la transferencia se pudo hacer de manera diaria a través de AWS Lambda, que ejecutaban funciones desarrolladas en Python una vez al día para hacer una solicitud de los viajes realizados y extraer indicadores más importantes, finalmente estos indicadores eran almacenados en la nube (AWS S3). En contraste, para medir los viajes de otras estrategias, como Ecobici (sistema de bicicletas compartidas en la CDMX), no se contó con una API debido al corto tiempo de implementación y el presupuesto del proyecto, por lo que se desarrolló un script que realizaba un scrapping de un archivo HTML. Este archivo era enviado por la persona usuaria y el programa extraía información específica, en este caso una tabla con el historial de viajes de la persona usuaria.

Finalmente, la información generada de las tres ERT se cargaba a un sistema de tableros para que las empresas dieran seguimiento al desempeño de las estrategias de manera visual e interactiva. El uso de un sistema de tableros permitió compartir insights del desempeño a un público no técnico de manera rápida y efectiva.

Algunos de los retos en este piloto fueron la falta de documentación de las respuestas arrojadas por las API. Casos de formatos de fechas inconsistentes fueron recurrentes, así como viajes duplicados. Dado que no había un indicador que permitiera identificar los viajes de cada empresa participante, se tuvo que presentar los resultados de manera agregada.

Uno de los aprendizajes de este proyecto es la importancia de la correcta delimitación de la estructura general de una API previo a su desarrollo e implementación. Esto incluye una documentación que ilustre la estructura general de la API, para facilitar la extracción de la información necesaria y evitar una transferencia con indicadores que limiten el tiempo y el costo de la misma.

Exploración

La exploración de datos, en inglés llamado Exploratory Data Analysis o EDA, es un proceso crítico donde se realizan investigaciones iniciales sobre los datos para descubrir patrones, detectar anomalías, probar hipótesis y comprobar supuestos con la ayuda de estadísticas resumidas y representaciones gráficas. Esta fase proporciona un análisis a primera vista de las fuentes de datos disponibles y trata de dar sentido a la información antes de profundizar en ellos. En lugar de intentar ofrecer una visión precisa, como la que se obtiene con el análisis de datos, la exploración de datos se centra en identificar tendencias clave y variables significativas.

La exploración de datos se utiliza a menudo como primer paso antes de realizar análisis que requieren más recursos y contar con el entendimiento previo de la información. Este enfoque es útil cuando se trabaja con conjuntos de datos grandes y desconocidos, o cuando el equipo de análisis está tratando de averiguar dónde buscar información útil y evaluar la calidad de la información obtenida. Esta exploración ayuda a responder tres preguntas importantes en los proyectos que se plantean:

- *¿Cuán confiables son los datos?*
- *¿Cuáles son las variables más importantes?*
- *¿Cuáles son las relaciones entre esas variables?*

En la mayoría de los casos, se deberán de utilizar técnicas manuales para realizar análisis exploratorios de datos en conjuntos. Esto se hace principalmente en Python o R, o simplemente visualizando los datos en una aplicación como Excel. Ya sea de forma manual o automatizada, se deben de realizar al menos las siguientes actividades en la fase de exploración (IBM, 2021b):

I) Lectura de datos: Consiste en importar los datos a una herramienta de análisis de datos como Python, R, Excel, etc. Es más eficiente cargar un archivo CSV en una tabla porque permite una fácil visualización y manipulación de los datos. Es posible filtrar, ordenar y buscar en los datos de manera interactiva, lo que es especialmente útil para conjuntos de datos grandes. Además, la biblioteca pandas de Python es altamente optimizada para trabajar con datos tabulares, lo que permite una carga y manipulación rápida de archivos CSV en Jupyter Notebook.

II) Limpieza de datos: Es el proceso de corrección o eliminación de datos incorrectos, corruptos, con formato incorrecto, duplicados o incompletos dentro de un conjunto. Cuando se combinan varias fuentes de datos, hay muchas posibilidades de que los datos estén duplicados o mal etiquetados. Si los datos son incorrectos, los resultados y los algoritmos no son confiables. No hay una forma absoluta de prescribir los pasos exactos en el proceso de limpieza de datos, pero una serie de pasos útiles son los siguientes (IBM, 2021a):

- › **Eliminar duplicados u observaciones irrelevantes:** La duplicación se produce frecuentemente en la fase de recopilación de datos, cuando se combinan conjuntos de datos de fuentes diferentes. O bien se pueden eliminar columnas vacías o datos que desde un inicio y gracias a la estrategia de evaluación sabemos que no serán de utilidad.
- › **Corregir errores estructurales:** Éstos se producen al observar nomenclaturas incorrectas o tipografías inconsistentes, lo que da lugar a categorizaciones incorrectas. Un ejemplo común es la clase de datos "N/A" que puede aparecer como "NaN", "NA", "Null", entre otros, debiéndose analizar en la misma categoría.
- › **Filtrar outliers o valores atípicos:** A menudo estos datos pueden observarse a simple vista ya que no parecen encajar con los valores típicos. Sin embargo, es importante tener una razón legítima para eliminarlos, como la certeza de una incorrecta introducción de la información. A veces estos outliers pueden confirmar o refutar la teoría en la que se está trabajando.

› **Tratar valores faltantes:** Es importante no ignorar los datos faltantes o vacíos, ya que muchos algoritmos no reconocen este tipo de valores. Existen muchas opciones por las que puede optar, algunas de ellas se muestran en la **tabla 9**.

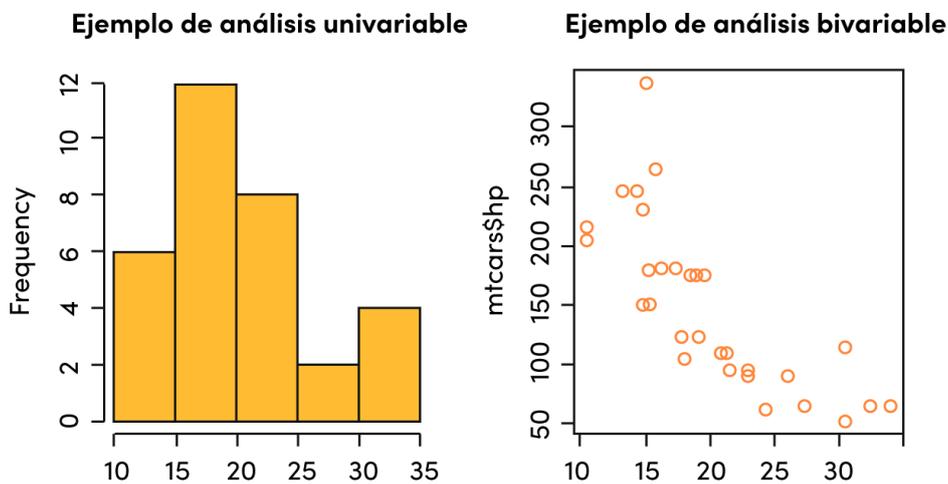
Tabla 9. Técnicas para tratar valores nulos

Método	Consideraciones
Eliminar datos omitidos	Pérdida de información significativa
Sustituir datos omitidos basándose en observaciones u operaciones (como promedios o regresiones)	Pérdida de integridad de los datos debido a las suposiciones
Modificar el proceso o algoritmo para navegar entre valores nulos	Compromete la eficiencia del proceso y la calidad de los patrones o relaciones

› **Validación de los datos:** Al finalizar el proceso de limpieza, debería poder responder preguntas como ¿tienen sentido los datos?, ¿prueban o refutan alguna teoría de trabajo?, ¿es posible identificar patrones? Si no es así, puede existir un problema en la calidad de los datos.

III) Análisis exploratorios: Consiste en la creación de las primeras visualizaciones para comprender mejor los patrones encontrados. Las gráficas más útiles para los análisis univariados suelen ser los análisis univariados como histogramas, diagramas de caja o los gráficos de tallo y hoja. En contraste, los gráficos de dispersión, las barras apiladas y los mapas de calor suelen ser una opción más adecuada para visualizar los análisis bivariados (**figura 9**).

Figura 9. Figura ilustrativa de análisis univariado y bivariado



Fuente: Elaboración propia.

V) Identificación y selección de variables: En esta fase se deben seleccionar los campos de información más importantes para el análisis subsecuente y eliminar aquellos que no son relevantes dentro de la estrategia de evaluación.

Procesamiento de datos

En esta etapa de la gestión de datos, los datos en bruto ya se han ingerido desde una serie de fuentes, como puede ser una API, bases de datos, archivos de datos, entre otros y además se han explorado por lo que ya se cuenta con un entendimiento general de la información. Usualmente la información usada en la fase de procesamiento proviene en formatos que no son directamente aptos para su análisis, conteniendo campos adicionales o errores. Por lo que es necesario realizar tareas que atiendan a estos problemas típicos de la información. Estos procesos suelen denominarse procesos ETL: Extracción de la información, Transformación de la misma y de Carga (*load* en inglés) a sistemas de almacenamiento de datos. A continuación, se describen las tareas principales de la transformación de datos para su explotación en la fase de análisis.

Transformación de los datos

Durante este paso, se ejecutan una serie de tareas o *scripts* para procesar los datos en el formato requerido para facilitar su manipulación y análisis. Estas tareas consisten en flujos de trabajo automatizados, como la elaboración de informes o cálculo de indicadores. Por ejemplo, un flujo de datos puede partir de una API de una ERT, y la etapa de transformación de datos tendrá como objetivo a extraer los campos clave para el análisis que se desee desarrollar en un formato de archivo útil para su manipulación como JSON (Science Direct, s/f). A continuación, se enumeran tipos de transformaciones que el equipo puede realizar durante el procesamiento de datos:

- › **Agregación de datos:** este proceso busca agrupar datos en diferentes niveles de agregación, por ejemplo, por semana o día, o también por otros campos de la información, por ejemplo género, zona de uso, etc.
- › **Filtrado de datos:** el objetivo del filtrado de datos es reducir una fuente de datos a sólo lo que el usuario necesita, eliminando los datos repetidos, irrelevantes o demasiado sensibles.
- › **Unión de datos:** Es el proceso de tomar diferentes tipos de datos, como diferentes bases de datos y fusionarlos en la misma estructura o esquema. Es una de las funciones más importantes de la transformación de datos. Una "unión" es una operación del lenguaje de bases de datos SQL que permite conectar dos o más tablas de bases de datos mediante sus columnas coincidentes o claves.
- › **División de datos:** La división de datos puede ser útil con una gran cantidad de datos recopilados durante un periodo de tiempo significativo.
- › **Revisión de formato:** Las revisiones de formato solucionan los problemas derivados de los diferentes tipos de datos de los campos, por ejemplo que un campo numérico esté guardado como de texto o viceversa.
- › **Reestructuración de claves:** Cuando las tablas de datos tienen claves con significados ya incorporados, pueden surgir problemas graves si no se conoce el significado de los parámetros, normalmente esto sucede cuando se quieren categorizar variables binarias, como valores falso y verdadero. En caso de que la documentación no incluya el valor de las claves, una reestructuración será necesaria para poder categorizar correctamente los datos.

Almacenamiento

Como último paso previo al análisis de la información está el almacenamiento, que consiste en cargar la información extraída y transformada a la base de datos o servidor de almacenamiento. La carga es un proceso continuo que puede realizarse de forma completa (cuando se cargan todos los datos por única vez en el almacén) o de forma incremental (a medida que se actualiza el almacén con nueva información). El almacenamiento incremental es particularmente útil en entornos donde los datos cambian constantemente, ya que permite actualizar el almacén de forma oportuna y eficiente. Este proceso es importante porque permite la disponibilidad de los datos para su posterior uso en análisis y visualizaciones.

El guardado de datos puede realizarse no sólo como paso previo al análisis e interpretación, sino en cada etapa mostrada en la figura 6, incluso de ser necesario podrá realizarse varias veces dentro de una etapa para evitar la pérdida de datos. En caso de que este proceso se realice a través de un proceso automatizado, este almacenamiento puede ser incorporado dentro del código del pipeline de datos para actualizar bases de datos o usando los formatos de datos descritos en la sección anterior. El respaldo de datos también puede realizarse manualmente a través del uso de hojas de cálculo, como Excel o Google Sheets, donde se pueden crear diferentes tablas para almacenar la información.

Caso: Moveup

La importancia de la validación de la información

Para analizar el servicio y evaluar el éxito del piloto, se desarrolló un pipeline de datos que incluyó la lectura de la información a través de una API, procesamiento, limpieza y visualización de la información. Esto permitió obtener insights valiosos para la toma de decisiones durante el piloto y una vez finalizado.

El uso de la información proveniente de la API presentó desafíos; en primer lugar, la documentación de la API era insuficiente, lo que dificultó su comprensión y uso efectivo. Para abordar esta situación, se llevaron a cabo varias reuniones con la ERT AllRide, para obtener una explicación detallada de la información. Por ejemplo, se requirió una clarificación de términos, incluso básicos como "kilómetros recorridos", a fin de asegurar una interpretación precisa de los datos para la toma de decisiones, por ejemplo, determinar si se trataba de kilómetros recorridos en el día o en la jornada del conductor, entre otros posibles significados no descritos en la documentación.

Durante la fase de exploración, se detectó información inconsistente que necesitaba ser validada con la ERT. Por ejemplo, el campo "tasa de ocupación" mostraba valores por encima del 100%, lo cual no era coherente con el máximo nivel esperado. Se consultó con la contraparte y se corrigió el problema en sus bases de datos. Asimismo, se identificaron viajes realizados en días en los que no se ofrecía servicio, los cuales resultaron ser viajes de prueba y no debían ser considerados en los análisis. Estos hallazgos destacan la importancia de tener un conocimiento detallado de la información que se está generando y no limitarse a utilizar los datos sin un análisis crítico previo.

Contar con información clara sobre la documentación de las API fue un aprendizaje de este piloto y también se destaca la importancia de invertir tiempo en la fase de exploración para tener un entendimiento pleno de la información, identificar posibles errores de ésta, evitando así llegar a conclusiones erróneas por no comprender completamente la información recibida.

Análisis e Interpretación

El objetivo del análisis de datos es obtener información útil y aprovechable. Este análisis, independientemente de que los datos sean cualitativos o cuantitativos, pretende describir y resumir la información, identificar relaciones, comparar variables y prever resultados. Marshall y Rossman (1999) describen el análisis de datos como el proceso de dar orden, estructura y significado a la masa de datos recopilados. Lo describen como un proceso desordenado, ambiguo y lento, pero también creativo y fascinante. En términos generales, aunque no se desarrolla de forma lineal, es la actividad de dar sentido a los datos, interpretarlos, lo que implica una búsqueda de afirmaciones generales entre varias categorías.

Al recopilar datos, nos enfrentamos a la inevitable pregunta de qué es mejor, una perspectiva cualitativa o cuantitativa, que aunque estimulante para algunos, podría causar estragos en la planificación y ejecución de proyectos. Los datos que se representan en formato verbal o narrativo son datos cualitativos. Este tipo de datos se recogen en grupos de discusión, entrevistas, cuestionarios abiertos y otras situaciones menos estructuradas. Una forma sencilla de ver los datos cualitativos es pensar en ellos en forma de palabras. Los datos cuantitativos son datos que se expresan en términos numéricos, estos valores numéricos pueden corresponder a una categoría o etiqueta específica.

Cuando se utiliza una metodología cuantitativa, normalmente se pone a prueba una teoría mediante la comprobación de una hipótesis. En la investigación cualitativa, se explora la aplicación de una teoría o modelo en un contexto diferente o se espera que surja una teoría o modelo a partir de los datos.

En cuanto al análisis cualitativo y cuantitativo de los datos, Kreuger y Neuman (2006) ofrecen un útil resumen de las diferencias y similitudes entre los métodos cualitativos y cuantitativos de análisis de datos. Según estos autores, los análisis cualitativos y cuantitativos son similares en cuatro aspectos, ya que ambas formas de análisis de datos implican:

1. Inferencia: el uso del razonamiento para llegar a una conclusión basada en pruebas.
2. Un método público: revelar su diseño de estudio de alguna manera.
3. Comparación como proceso central: identificación de patrones o aspectos que son similares o diferentes.
4. Esforzarse por evitar errores, conclusiones falsas e inferencias engañosas.

El análisis de datos cualitativos está menos estandarizado, ya que la gran variedad de enfoques de la investigación cualitativa se corresponde con los numerosos enfoques del análisis de datos, mientras que los investigadores cuantitativos eligen entre un conjunto especializado y estándar de técnicas de análisis de datos, normalmente éstos corresponden a técnicas estadísticas.

Posterior al análisis es importante darle una interpretación de valor a los mismos, para ello, utilizar herramientas de visualización es lo más común. Visualizar datos es, literalmente, crear y luego considerar una presentación visual de los datos. Técnicamente, no es un análisis, ni un sustituto del análisis, pero a menudo puede proporcionar una base eficaz para orientar los análisis posteriores.

Hay muchas herramientas de visualización de datos disponibles, adaptadas a distintos niveles de experiencia. Entre las herramientas más populares que requieren pocos o ningún conocimiento de programación se encuentran Google Charts, Tableau, Datawrapper, entre otros. En caso de estar familiarizado con Python y R, también existen muchas bibliotecas y paquetes de visualización de datos disponibles. Por ejemplo Plotly, Seaborn, Matplotlib, ggplot2, Charts.js.

Divulgación los datos

En pocas palabras, divulgar los datos es el acto de poner a disposición de otras personas u organizaciones los hallazgos realizados en el análisis de datos. La compartición de datos apoyados de la tecnología hace posible que los datos se distribuyan fácilmente por toda una organización.

Las instituciones y empresas necesitan cada vez más datos para resolver problemas, perfeccionar operaciones y tomar las decisiones mejor informadas. Para que esto sea una realidad, los datos deben ser accesibles y comprensibles para todos o de lo contrario es posible que nunca se lleguen a hacer descubrimientos y reflexiones importantes.

Es común encontrar retos en términos de análisis de datos cuando se hacen pilotos con múltiples actores. El primero de ellos es la cultura que rodea a la puesta en común de datos. Tradicionalmente, las organizaciones tanto públicas como privadas han adoptado una cultura de "propiedad" de los datos frente a una cultura de "puesta en común", lo que dificulta la transmisión de información entre organizaciones. Esto nos lleva al segundo gran reto: estandarizar los datos para que todos los actores puedan acceder a ellos y utilizarlos. Comúnmente, los datos tienen un formato diferente en los distintos niveles de una organización, y para ser entendidos, personal técnico como ingenieros de datos tienen que transformarlos constantemente para que sean comprensibles y utilizables, lo que a menudo puede llevar días o incluso semanas.

El último gran reto es la capacidad de búsqueda. Incluso si las organizaciones tienen todos sus datos accesibles en un solo lugar, puede ser un proceso meticuloso tamizar a través de todos esos datos para encontrar los puntos que son relevantes para su proyecto o piloto.

Sin embargo, a medida que mejora la tecnología de los datos y las empresas se dan cuenta de la importancia de compartirlos, estas barreras y retos tradicionales se están derrumbando. La tecnología y el software más recientes disponibles hoy en día permiten a las organizaciones almacenar datos en lagos de datos flexibles, automatizar la gestión de datos, transformar y transferir datos sin problemas y compartirlos como productos utilizables.

Caso: Digitalización del Transporte Concesionado

Datos desafiantes: cómo superar obstáculos en análisis

Durante la ejecución del piloto de la Digitalización del Transporte Concesionado y sus subpilotos de alertas y rutas nocturnas, fue necesario realizar análisis de datos para establecer las líneas bases y evaluar los resultados de las intervenciones realizadas. Una vez establecidos los objetivos, recursos y tiempos, así como los planes de trabajo y las líneas bases y estrategias de evaluación fue necesario recolectar datos de distintas fuentes y con diferentes formatos y grados de calidad para realizar los análisis de línea base y de evaluación.

Previo a los análisis fue necesario explorar los diferentes conjuntos de datos para asegurar que contaban con la calidad y el contenido necesario para hacer los análisis propuestos, así como que se contaba con claridad en el significado de los distintos campos. Esto representó un reto debido a que la calidad y la manera en la que se recopilaban los datos no permitía una interoperabilidad completa (e.g.: las preguntas realizadas entre encuestas de calidad podían variar; dependiendo de la encuesta, las respuestas podían ser abiertas o tener respuestas predefinidas), entre las distintas fuentes de datos o la manera de transferir los datos era a través de archivos individuales. De la misma manera, no se contó con un diccionario de datos desde el principio, lo que dificultó la generación de los análisis planteados en las primeras etapas del proyecto. Estos retos encontrados dentro de los datos hicieron que la estrategia de evaluación requiriera de ajustes constantes para adaptarse tanto a las necesidades del proyecto como a la disponibilidad de información.

Para lograr la interoperabilidad entre los datos se siguieron distintos pasos dependiendo de la dificultad para hacer comparables las distintas fuentes. En ocasiones, la información obtenida de encuestas con preguntas ligeramente diferentes pero compatibles pudo ser utilizada sin modificarse, otras veces fue necesario replantear los indicadores para quitar niveles de desagregación o modificar la granularidad en la temporalidad. En otros casos, los indicadores tuvieron que ser modificados con mayor profundidad debido a que las fuentes diferían en el tipo de dato, unas fuentes incluían variables numéricas mientras que otras contenían variables discretas, esto nos obligó a crear nuevas variables para poder hacer compatibles las diferentes fuentes y cambiar el análisis de cuantitativo a cualitativo. En las ocasiones en que no existían datos concretos para generar los indicadores fue necesario plantear nuevos indicadores indirectos (proxy) para poder entender de una manera aproximada los fenómenos analizados. Por último, existieron casos en los que no fue posible obtener los datos necesarios, lo que nos llevó a desestimar los datos por completo.

Esta experiencia demostró la necesidad de generar procesos iterativos en la planeación de los proyectos y sus estrategias de evaluación, ya que permitió adaptar los proyectos y los indicadores a las necesidades de medición y las limitaciones de información. Demostró la utilidad de la socialización del proceso entre personas miembros del proyecto y de terceros para compartir los objetivos de los proyectos y las necesidades de información y de desarrollo tecnológico. De igual importancia fue la búsqueda documental de proyectos similares para conocer las metodologías e indicadores utilizados. Por último y de vital importancia fue la capacidad de manejo de los datos del equipo y el entendimiento de las herramientas estadísticas para lograr comunicar el impacto de los proyectos de manera efectiva con personas no expertas.

Recomendaciones clave

La implementación del uso de datos en colaboración con gobiernos y ERT ha abierto un amplio abanico de oportunidades en el desarrollo de proyectos y evaluaciones, así como en la toma de decisiones informadas. No obstante, también existen diversos retos que han dejado importantes lecciones aprendidas obtenidas de las tres experiencias con ERT desarrolladas. En esta sección, se detallan algunas recomendaciones recopiladas a partir de estas lecciones, con el fin de compartir conocimientos y experiencias que puedan ser útiles en futuros proyectos de movilidad basados en datos.

- › **La importancia de un conocimiento detallado de los datos.** Conocer con detalle la metodología de generación de datos y su significado, es crucial para el aprovechamiento de la información. El desarrollo de análisis exploratorios y estadísticos a realizar es fundamental para contar con un conocimiento de la manera en que se está generando la información para el desarrollo y monitoreo de los proyectos, de otra forma, es posible que se generen conclusiones erróneas o análisis de datos incorrectos a raíz de una falta de entendimiento de la información. Además, el uso de herramientas estadísticas adecuadas permite obtener resultados más precisos, confiables y relevantes para la toma de decisiones. El uso de estas herramientas facilita la identificación de patrones, tendencias y relaciones entre los datos, así como la evaluación de hipótesis y la medición de resultados.
- › **Necesidades técnicas de un equipo de datos.** La gestión y procesamiento de datos en proyectos de movilidad requiere de un equipo técnico capacitado y con habilidades específicas en el manejo de herramientas y tecnologías para la gestión de datos. Un equipo de personas expertas en datos y estadística puede realizar procesos más ágiles y efectivos, lo que se traduce en un mejor aprovechamiento de los recursos y la obtención de un mayor valor de la información. Además, como se presentó en las experiencias de los pilotos, un equipo técnico bien estructurado puede ayudar a minimizar los errores y disminuir los costos asociados con el análisis de datos.
- › **Adaptación al proyecto según la disponibilidad de la información.** La disponibilidad de datos es fundamental para las acciones de planeación, monitoreo y evaluación. Sin embargo, es importante tener en cuenta que los datos disponibles no siempre cumplen con los requisitos necesarios para el análisis o la toma de decisiones o que sencillamente los datos requeridos no existen o son demasiado caros o difíciles de generar. Por esta razón, es esencial contar con flexibilidad para adaptarse a la información existente y buscar formas de complementar con otras fuentes.
- › **Transferencia de información automática.** El uso de una API para obtener la información permite monitorear y evaluar los proyectos de una manera automatizada, lo cual a su vez permite generar visualizaciones interactivas e infografías para comunicar resultados y recomendaciones tanto para personas expertas como para el público en general. Esto requiere una negociación con la ERT que genera los datos, ya que este desarrollo implica destinar recursos humanos y tecnológicos. Este proceso suele tomar al menos un par de meses y requiere de revisiones detalladas de quien desarrolla la API y quien consume los datos. También, es posible encontrar errores en la información, campos faltantes e información incongruente, por lo que se debe contar con mecanismos de verificación, por ejemplo visitas en campo, junto con requisitos claros de la información y un calendario de trabajo para asegurar el aprovechamiento de la información.
- › **Adopción de estándares de movilidad.** Es importante que las ciudades y las ERT impulsen la adopción de estándares abiertos como GTFS, MDS o GBFS, que permiten la interoperabilidad de los datos de diferentes fuentes y sectores. Su uso tiene distintos beneficios: para las personas usuarias les permite contar con acceso a la información de los servicios de transporte a través de planeadores de viaje como Google Maps, Moovit, Where is My Transport, etc. A las ciudades les permite gestionar la oferta de transporte y tener una visión completa de la dinámica de oferta y demanda de los servicios. A las empresas de movilidad como las ERT les permite compartir su información con las ciudades para cumplir con los requerimientos de operación y además les permite analizar su información con herramientas de análisis de software especializadas que tienen como insumo la información en estos estándares.
- › **Aprovechamiento de la tecnología.** La selección adecuada de tecnologías y herramientas es clave para el éxito del proyecto, ya que afecta significativamente la eficiencia y la efectividad del procesamiento y análisis de datos. Es recomendable explorar opciones y evaluar sus fortalezas y debilidades antes de tomar una decisión final. El uso de herramientas de código abierto reduce los costos y proporciona una documentación extensa. Las herramientas actuales, como Jupyter o RStudio, reducen el tiempo de desarrollo y permiten compartir resultados con públicos no expertos.



Conclusiones

Este documento, resultado de la gestión de datos de las ERT durante el programa Ideamos, explora las prácticas para el aprovechamiento de esta información en las fases de planeación, monitoreo y evaluación de los proyectos de movilidad. En las cuatro secciones de contenido se han identificado las necesidades y herramientas necesarias para llevar a cabo un proyecto exitoso de datos, desde los elementos necesarios para la definición correcta del marco de medición, necesidades legales, técnicas, de software y de hardware, y por último, se describe un proceso con los pasos necesarios para la obtención, procesamiento y uso de la información.

Adicionalmente, se presentan a lo largo del documento tres pilotos que ofrecen valiosas lecciones aprendidas y recomendaciones clave para su consideración en otros proyectos de movilidad. Estas experiencias muestran cómo el uso de datos puede mejorar la calidad de vida de las personas, reducir externalidades negativas y fomentar la movilidad sustentable no sólo en México, sino que puede extenderse a otros países similares y a la región de LAC en general.

Al adoptar enfoques y mejores prácticas desarrolladas a partir de las experiencias de este estudio, otras ciudades pueden mejorar su eficiencia en la toma de decisiones basadas en datos y, a su vez, mejorar la movilidad urbana en su región. Es importante mencionar que, la colaboración entre entidades gubernamentales, empresas de transporte y actores clave del sector es esencial para crear un ecosistema de datos sólido que promueva la movilidad sustentable y el desarrollo urbano sostenible en toda la región. Al establecer colaboraciones y alianzas estratégicas en el ámbito regional, se puede fomentar la adopción de estándares y buenas prácticas, permitiendo que más ciudades y países se beneficien del potencial de los datos para la transformación positiva de la movilidad.

Es importante resaltar que, aunque el aprovechamiento de los datos se ha vuelto más accesible y económico, la disponibilidad de datos no es suficiente para lograr un impacto significativo en la mejora de la movilidad de las personas sin la voluntad por parte de las empresas, gobiernos y organismos para utilizarlos de manera efectiva. Por lo tanto, es necesario que las empresas y entidades responsables del transporte reconozcan la importancia de los datos y se comprometan a compartirlos y utilizarlos para la mejora continua del transporte. Además, es crucial que las ciudades establezcan marcos normativos claros y consistentes para garantizar que estos datos puedan ser usados para la planificación de la movilidad en las ciudades, sin vulnerar la privacidad de las personas usuarias.

La documentación de prácticas, herramientas y casos de uso para el aprovechamiento de datos a lo largo de los tres años de duración del programa, ha permitido utilizar una gran cantidad de información útil sin destinar grandes recursos. Este documento es una guía útil para aquellos que buscan maximizar el valor de sus datos y, en última instancia, contribuir a ciudades con servicios de transporte más inclusivos y sostenibles. Con esto, se cumple el objetivo de este documento y se espera que sea de gran utilidad para la comunidad interesada en el tema.

Bibliografía

Abduljabbar, R. L., Liyanage, S., & Dia, H. (2021). The role of micro-mobility in shaping sustainable cities: A systematic literature review. *Transportation Research Part D: Transport and Environment*, 92, 102734. <https://doi.org/10.1016/j.trd.2021.102734>

Hodson, M., & Marvin, S. (2010). Can cities shape socio-technical transitions and how would we know if they were? *Research Policy*, 39(4), 477–485. <https://doi.org/10.1016/j.respol.2010.01.020>

IBM. (2021a, marzo 8). *Tratamiento de los valores atípicos y extremos*. <https://www.ibm.com/docs/es/spss-modeler/18.2.0?topic=tab-handling-outliers-extreme-values>

IBM. (2021b, agosto 4). *¿Qué es el análisis exploratorio de datos?* IBM.com. <https://www.ibm.com/mx-es/cloud/learn/exploratory-data-analysis>

Jones, H. (2021). Sustainable Infrastructure Project Evaluation. En W. Leal Filho, A. M. Azul, L. Brandli, A. Lange Salvia, & T. Wall (eds.), *Industry, Innovation and Infrastructure* (pp. 1–13). Springer International Publishing. https://doi.org/10.1007/978-3-319-71059-4_60-1

Kreuger L. & Neuman W. L. (2006). *Social work research methods : qualitative and quantitative approaches : with research navigator*. Pearson/Allyn and Bacon.

Kreuzer, F. M., & Wilmsmeier, G. (2014). *Eficiencia energética y movilidad en América Latina y el Caribe*.

Marshall, C. and Rossman, G.B. (1999) *Designing Qualitative Research*. 3rd Edition, International Educational and Professional Publisher, California

Marshall, C., & Rossman, G. B. (1999). *Designing Qualitative Research*. 3rd Edition, International Educational and Professional Publisher, California.

Melis, A., Prandini, M., Sartori, L., & Callegati, F. (2016). Public Transportation, IoT, Trust and Urban Habits. En F. Bagnoli, A. Satsiou, I. Stavrakakis, P. Nesi, G. Pacini, Y. Welp, T. Tiropanis, & D. DiFranzo (eds.), *Internet Science* (Vol. 9934, pp. 318–325). Springer International Publishing. https://doi.org/10.1007/978-3-319-45982-0_27

Mitchell, B. (2018). *Resource and environmental management*. Oxford University Press.

OMPI. (s/f). *Acuerdos de transferencia de tecnología*. Recuperado el 21 de marzo de 2023, de <https://www.wipo.int/technology-transfer/es/agreements.html>

Paiva, S., Ahad, M. A., Tripathi, G., Feroz, N., & Casalino, G. (2021). Enabling Technologies for Urban Smart Mobility: Recent Trends, Opportunities and Challenges. *Sensors*, 21(6), Article 6. <https://doi.org/10.3390/s21062143>

Science Direct. (s/f). *Data Processing Stage—An overview*. Recuperado el 14 de marzo de 2023, de <https://www.sciencedirect.com/topics/computer-science/data-processing-stage>

Snowflake. (s/f). *What is a Data Pipeline?* Recuperado el 14 de marzo de 2023, de <https://www.snowflake.com/guides/data-pipeline>

Steele, L., & Orrell, T. (2017). The frontiers of data interoperability for sustainable development. *Development Initiatives* (2017). https://www.publishwhatyoufund.org/wpcontent/uploads/2017/11/JUDS_Report_Web_061117.pdf.

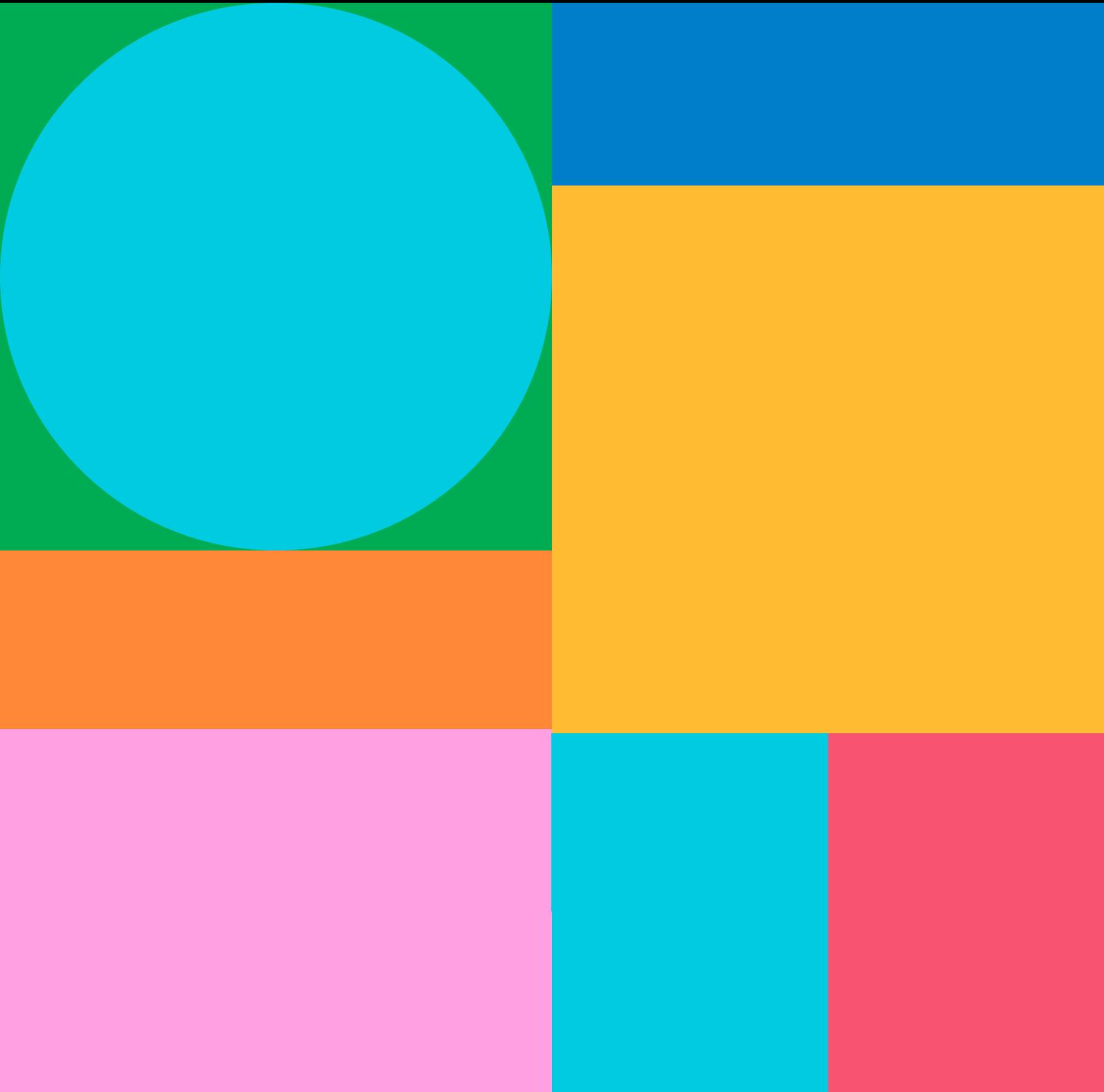
Stephen Gossett. (2020, agosto). *Bike-Sharing Presents a Huge Data Challenge*. Built In. <https://builtin.com/data-science/bike-share-rebalancing>

The Compass for SBC. (s/f). *How to Develop a Monitoring and Evaluation Plan*. The Compass for SBC. Recuperado el 23 de marzo de 2023, de <https://thecompassforsbc.org/how-to-guide/how-develop-monitoring-and-evaluation-plan>

Xiaomei Tan. (2018, enero 17). *Bike-sharing data and cities: Lessons from China's experience*. Global Environment Facility. <https://www.thegef.org/blog/bike-sharing-data-and-cities-lessons-chinas-experience>

Yeo, H., & Shafahi, Y. (2017). Automated Monitoring and Control of Public Transit Systems In Smart Transportation Systems (pp. 57-78). Springer, Cham.

Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 383-398. <https://doi.org/10.1109/TITS.2018.2815678>



ideam***s**

 **ITDP**

 **BID**

 **BID | LAB**