

**Campus:** Monterrey, México.

**Avance 0. Propuesta de proyecto y firma de convenios**

**Planteamiento del proyecto**

1.3 Describir las fases de la metodología a emplear.

1.4 Definir los objetivos comerciales y traducirlos en objetivos del proyecto.

**Estudiantes**

Andrés Julián López Hurtado - A01793899

Nathalia Milena Prada Hernández - A01793999

Víctor Alejandro Regueira Romero - A01794404

**Profesor titular**

Dra. Grettel Barceló Alonso

**Profesora asistente**

Mtra. Verónica Sandra Guzmán de Valle

**Fecha:** 22 de septiembre de 2024

Contenido

[1 Antecedentes 2](#_Toc177926921)

[1.1 Descripción de los servicios impactados 2](#_Toc177926922)

[1.1.1 Factores diferenciales del servicio de Forza 2](#_Toc177926923)

[1.2 Entidades clave de negocio identificadas 3](#_Toc177926924)

[1.2.1 Entidad: Camión 3](#_Toc177926925)

[1.2.2 Entidad: Cliente 4](#_Toc177926926)

[1.2.3 Entidad: Conductor (Primary Driver) 4](#_Toc177926927)

[1.2.4 Entidad: CSR (Customer Service Representative) 4](#_Toc177926928)

[1.2.5 Driver Manager 4](#_Toc177926929)

[1.2.6 Entidad: Viaje 5](#_Toc177926930)

[1.2.7 Entidad: Carga (Load) 5](#_Toc177926931)

[1.2.8 Entidad: Sistemas de Gestión de Viajes 5](#_Toc177926932)

[1.2.9 Entidad: Sistemas de Monitoreo de Trucks 5](#_Toc177926933)

[1.2.10 Entidad: Sistemas de Monitoreo de Cajas 6](#_Toc177926934)

[1.2.11 Entidad: Orden de Trabajo (Work Order) 6](#_Toc177926935)

[1.2.12 Entidad: Factura (Invoice) 6](#_Toc177926936)

[1.3 SIPOC del proceso de procesos de envío de cargas (loads) 6](#_Toc177926937)

[2 Entendimiento del negocio 7](#_Toc177926938)

[2.1 Formulación del problema: ¿Qué es lo que se intenta resolver? 7](#_Toc177926939)

[2.2 Contexto: ¿Por qué es importante resolver este problema? 7](#_Toc177926940)

[2.3 Objetivos: ¿Cuál es la meta prevista? 7](#_Toc177926941)

[2.4 Preguntas clave: ¿Cuáles son las preguntas claves que deben responderse? 8](#_Toc177926942)

[2.5 Involucrados 8](#_Toc177926943)

[3 Entendimiento de los datos 8](#_Toc177926944)

[3.1 Descripción de los datos: 8](#_Toc177926945)

[3.2 Descripción de la arquitectura 9](#_Toc177926946)

[3.2.1 Fases del Procesamiento de Datos 10](#_Toc177926947)

[3.2.2 Salidas o Consumo 10](#_Toc177926948)

[3.2.3 Enfoque General 11](#_Toc177926949)

[3.3 Ejemplo de datos 11](#_Toc177926950)

[3.3.1 Técnica de ML: 11](#_Toc177926951)

[3.3.2 Identificación de Variables: 11](#_Toc177926952)

[4 Bibliografía 12](#_Toc177926953)

[5 Anexo 13](#_Toc177926954)

# Antecedentes

FORZA Transportation, fundada en 2014, se especializa en ofrecer **servicios de Full Truck Load (FTL)** entre los países del T-MEC (Tratado entre México, Estados Unidos y Canadá). Con una flota moderna de camiones modelo 2020 y posteriores, la empresa garantiza un transporte seguro y confiable para una amplia gama de industrias, que incluyen alimentos, electrónica, automotriz y otros sectores generales. Su riguroso mantenimiento permite minimizar el tiempo de inactividad y asegurar operaciones eficientes.

FORZA Transportation tiene sus raíces en 2007, cuando Roberto Pérez, su Fundador y CEO, inició "Roberto Pérez Trucking" como propietario-operador. Gracias a su dedicación y visión, pasó de gestionar un solo camión a liderar una flota de más de 1000 unidades en la actualidad, consolidando a FORZA Transportation Services, Inc. como un referente en el sector del transporte.

La empresa se distingue por brindar una experiencia excepcional al cliente a través de soluciones personalizables, adaptadas a las necesidades únicas de cada cliente. Ofrece servicios de transporte puerta a puerta por carretera, ferrocarril o una combinación de ambos, llevando los envíos de cualquier lugar a cualquier destino en América del Norte. El enfoque de este proyecto se centrará en el transporte puerta a puerta por carretera.

## Descripción de los servicios impactados

Los servicios de **Full Truck Load (FTL)**, o carga completa, son un tipo de transporte de mercancías en el que se alquila todo el espacio de un camión para un solo cliente o cargador. Esto significa que el envío ocupa la totalidad del vehículo, lo que permite transportar grandes volúmenes de mercancía de una manera más directa y rápida.

Algunas características clave de los servicios de FTL son:

* **Carga exclusiva**: Todo el espacio del camión se destina a un solo cliente o cargamento, lo que elimina la necesidad de compartir el transporte con otras cargas.
* **Menor manipulación**: Al ser un servicio directo, las mercancías no se transfieren entre camiones durante el trayecto, lo que reduce el riesgo de daños o pérdidas.
* **Envío más rápido**: Al no tener que hacer múltiples paradas para recoger o entregar mercancías de otros clientes, los tiempos de tránsito suelen ser más cortos.
* **Ideal para grandes volúmenes**: Es la mejor opción cuando el tamaño o el peso de la carga justifica llenar un camión completo, o cuando el cliente requiere un envío rápido y exclusivo.

### Factores diferenciales del servicio de Forza

* Un agente, de principio a fin: un único punto de contacto para cada envío, eliminando la necesidad de tratar con múltiples proveedores o coordinar varias entregas.
* Cultura de seguridad, ante todo: sea cual sea la necesidad del negocio, se puede confiar en que los conductores mantendrán las medidas de seguridad a lo largo de cada envío.
* Amplias certificaciones: FORZA Transportation Services, Inc. cumple completamente con las pautas del DOT y es una orgullosa empresa de propiedad minoritaria certificada por el Estado de Texas. Sus socios en México también están completamente certificados y cumplen con altos estándares de calidad y seguridad.
* Gestionan la carga utilizando camiones y furgonetas de los últimos modelos.
* Ofrecen visibilidad gracias a la conectividad EDI
* Servicio al cliente dedicado.

## Entidades clave de negocio identificadas

A continuación, se describen las principales **entidades de negocio** que sonfundamentales para el proyecto de **Mejora del Desempeño de un servicio de Flotillas Terrestres con Inteligencia de Negocios y datos Telemáticos** en **FORZA Transportation Services, Inc.** La identificación precisa de estas entidades permitirá comprender mejor los procesos internos y facilitará el desarrollo de estrategias basadas en datos:

### Entidad: Camión

La entidad "Camión" representa la infraestructura esencial para el transporte de mercancías. Se caracteriza por modelos recientes, capacidad de carga variada, sistemas telemáticos para seguimiento y gestión, y características de seguridad avanzadas.

La entidad Camión se divide en dos subentidades que lo componen: Truck (vehículo principal) y Caja (remolque):

#### Subentidad: Truck (Vehículo Principal)

* La entidad Truck representa el vehículo motorizado que forma la parte principal de un camión en la flota. El Truck es el núcleo de las operaciones de FORZA Transportation Services, Inc., y su correcta gestión es clave para el éxito del negocio. Monitorear sus especificaciones, estado, seguridad y mantenimiento permite mejorar la eficiencia operativa, reducir costos de operación y garantizar el cumplimiento de las normativas legales y de seguridad. Actualmente, la flota de Forza está compuesta por aproximadamente 1000 trucks, todos con características similares, excepto por el modelo o año de fabricación.

#### Subentidad: Caja (Remolque de Carga)

* Se refiere a la parte del camión que contiene la carga, también conocida como la caja de carga o el remolque. Puede variar en tamaño y tipo, como caja cerrada, plataforma o refrigerada. Dependiendo de la naturaleza de la carga, la caja puede estar equipada con características específicas, como refrigeración para productos perecederos o estructuras reforzadas para cargas pesadas. En Forza actualmente todas las cajas, aproximadamente 5000, tienen las mismas dimensiones. En algunos casos los clientes frecuentes de Forza tienen un subgrupo de cajas asignadas para su constante utilización.

### Entidad: Cliente

Se refiere a las empresas o personas que solicitan los servicios de transporte. Incluye información sobre sus necesidades, preferencias, historial de envíos y satisfacción, lo cual es esencial para personalizar los servicios.

### Entidad: Conductor (Primary Driver)

La entidad "Conductor" es crucial para la operación diaria de la flota. Se caracteriza por la capacitación en conducción segura, conocimiento de normativas de transporte y habilidades en la gestión de cargas. El conductor normalmente está asociado a un truck.

FORZA Transportation Services, Inc. cuenta con tres tipos de conductores, cada uno con características y funciones específicas:

* **Owner-Operator**: Estos conductores son propietarios de sus propios camiones y operan como contratistas independientes. Tienen la flexibilidad de gestionar sus horarios y rutas, lo que les permite adaptarse a las necesidades de los clientes y maximizar su eficiencia operativa.
* **CDL Driver**: Los conductores con licencia CDL (Commercial Driver's License) están capacitados para manejar camiones comerciales de gran tamaño. Cumplen con las regulaciones de seguridad y transporte, garantizando que las cargas se transporten de manera segura y eficiente.
* **B1 Driver**: Este tipo de conductor está especializado en operar vehículos ligeros que requieren una licencia B1. Su función es manejar cargas más pequeñas o específicas, proporcionando una solución ágil y eficiente para envíos que no requieren un camión completo.

### Entidad: CSR (Customer Service Representative)

La entidad CSR se refiere al representante de servicio al cliente que actúa como el punto de contacto principal entre los clientes y FORZA Transportation Services, Inc. El CSR es responsable de gestionar las comunicaciones con los clientes, coordinar los detalles del envío y brindar soporte antes, durante y después del proceso de transporte. En Forza cada cliente suele tener un solo CSR asignado durante toda su relación con la empresa.

### ****Driver Manager****

El **Driver Manager** en FORZA Transportation Services, Inc. es responsable de **coordinar** a los conductores y asegurar que las operaciones de transporte se desarrollen de manera eficiente y conforme a los estándares de seguridad y calidad de la empresa.

### Entidad: Viaje

La entidad "Viaje" representa el trayecto completo que realizan los camiones para transportar mercancías. Este concepto incluye tanto la planificación y ejecución de las rutas como las operaciones logísticas asociadas a cada viaje. Es fundamental para optimizar el desempeño del servicio, reducir costos y garantizar la satisfacción del cliente.

* Al finalizar un viaje, tanto el **truck** como la **caja** reciben mantenimiento de forma independiente.
* Los viajes pueden contar con un **primary driver** o con un **team** (compuesto por el conductor principal y un acompañante). No existe una regla fija para determinar si el viaje se realizará con un equipo o solo con el conductor principal.
* El estatus de un viaje puede ser:
  + **Open (Abierto)**: El viaje está planificado, pero aún no ha comenzado.
  + **En Route (En Ruta)**: El camión está en camino hacia su destino.
  + **Delivered (Entregado)**: La mercancía ha llegado a su destino final, pero el viaje aún no se ha cerrado formalmente.
  + **Completed (Completado)**: El viaje ha finalizado, la carga ha sido entregada y se ha facturado.

### Entidad: Carga (Load)

Representa los productos o mercancías que se transportan. Incluye características como el tipo de carga (perecedera, frágil, etc.), volumen, peso y requisitos específicos de transporte (refrigeración, protección, etc.). En FORZA, el contenido de la carga es desconocido para la empresa, ya que es responsabilidad del cliente cargar y descargar el contenido de las cajas en el lugar de origen y destino del viaje.

### Entidad: Sistemas de Gestión de Viajes

Se refiere a las herramientas y tecnologías utilizadas por FORZA Transportation Services, Inc. para rastrear y gestionar sus flotas en tiempo real.

* **Allways** es una plataforma de soluciones logísticas y de transporte que se centra en la eficiencia y la optimización de la cadena de suministro. Ofrece servicios de gestión de flotas y planificación de rutas y seguimiento de envíos.
* **Forza Core** es un desarrollo interno diseñado específicamente para la gestión de viajes en FORZA Transportation Services, Inc.
* **Conectividad EDI** se refiere a la capacidad de intercambiar documentos comerciales electrónicamente entre sistemas informáticos de diferentes organizaciones de manera estandarizada y segura. En el contexto de FORZA Transportation Services, Inc., la conectividad EDI es clave para optimizar los procesos logísticos y mejorar la comunicación con los clientes y proveedores.

### Entidad: Sistemas de Monitoreo de Trucks

Se refieren a la tecnología utilizada para supervisar en tiempo real el estado, ubicación y rendimiento de los camiones dentro de su flota. Estos sistemas combinan hardware y software para proporcionar visibilidad detallada y datos telemáticos que ayudan a optimizar la operación de los camiones, mejorando la seguridad, eficiencia y mantenimiento preventivo.

* **Geotab** es una plataforma líder en telemática que proporciona soluciones avanzadas de gestión de flotas para empresas de diversos sectores. Su tecnología permite el monitoreo en tiempo real de vehículos, ofreciendo datos detallados sobre el rendimiento del motor, la ubicación GPS, el consumo de combustible y el comportamiento del conductor.

### Entidad: Sistemas de Monitoreo de Cajas

Los **Sistemas de Monitoreo de Cajas** en FORZA Transportation Services, Inc. son herramientas tecnológicas utilizadas para rastrear y supervisar en tiempo real las **cajas** o **containers** durante todo el proceso de transporte. Estos sistemas permiten obtener datos precisos sobre la ubicación de las cajas. Estos sistemas también cuentan con **Geocercas** (Área del Cliente) e informa que se ha ingresado a una geocerca determinada. Esto permite monitorear la ubicación de las cajas y camiones, asegurando que se cumplan las rutas establecidas y mejorando la gestión logística. Las geocercas son una herramienta valiosa para optimizar la seguridad y el control en tiempo real de las operaciones.

* **Spireon** es el sistema de monitoreo de cajas más reciente implementado por FORZA Transportation Services, Inc. Este sistema proporciona soluciones avanzadas para rastrear y supervisar en tiempo real las cajas o containers utilizados en el transporte de mercancías.
* **Skybitz** es el sistema de monitoreo de cajas más anterior utilizado por FORZA Transportation Services, Inc. Aunque la empresa ha decidido implementar **Spireon** recientemente, **Skybitz** aún sigue en operación y no ha sido decomisionado en todas las cajas.

### Entidad: Orden de Trabajo (Work Order)

La Orden de Trabajo (Work Order) en FORZA Transportation Services, Inc. es un documento clave que formaliza la planificación y ejecución de los viajes de transporte. Este documento es fundamental para la gestión operativa y logística, ya que abarca todos los elementos necesarios para asegurar que cada viaje se realice de manera eficiente y dentro de los estándares establecidos.

### Entidad: Factura (Invoice)

La Factura (Invoice) en FORZA Transportation Services, Inc. es un documento formal que detalla la transacción entre la empresa y sus clientes por los servicios de transporte prestados. Este documento es fundamental para la contabilidad y gestión financiera, así como para mantener la transparencia y confianza con los clientes.

## SIPOC del proceso de procesos de envío de cargas (loads)

El siguiente SIPOC resume el proceso de gestión de solicitudes de servicio de carga, desde el registro inicial hasta la finalización del viaje. Este modelo de alto nivel identifica los proveedores, entradas, procesos clave, salidas y clientes involucrados en cada etapa del ciclo de servicio. El objetivo es garantizar una eficiente coordinación entre el cliente, el equipo de CSR, el Driver Manager y el conductor, asegurando que la solicitud sea registrada, planificada, ejecutada y finalizada con éxito, incluyendo la entrega y la facturación.

| Proveedor | Entrada | Proceso | Salida | Cliente |
| --- | --- | --- | --- | --- |
| Cliente | Solicitud de servicio de carga | Registrar y revisar la solicitud | Solicitud registrada | CSR |
| CSR | Orden de trabajo en Allways | Verificar disponibilidad y planificar el transporte | Fechas y disponibilidad confirmadas | Cliente |
| CSR / Driver Manager | Fechas y detalles del transporte | Asignar y coordinar transporte con conductor | Transporte asignado y programado | Conductor |
| CSR | Estado del viaje y entrega | Supervisar ejecución y confirmar entrega | Entrega confirmada | Cliente |
| CSR | Detalles del viaje completado | Generar factura y cerrar el viaje | Factura generada, viaje completado | Cliente |

# Entendimiento del negocio

## Formulación del problema: ¿Qué es lo que se intenta resolver?

Utilizando los datos generados por el sensor GeoTab instalado en cada camión de la flota, buscamos identificar los factores clave que influyen en el rendimiento del combustible. Este análisis nos permitirá desarrollar estrategias específicas que optimicen el consumo de combustible, lo cual repercutirá positivamente en la reducción de costos operativos y, en consecuencia, en la mejora de la rentabilidad del negocio.

## Contexto: ¿Por qué es importante resolver este problema?

Minimizar los costos operativos es fundamental para asegurar la competitividad de Forza Transportation, ya que una mejora en dichos costos impacta directamente en la reducción de las tarifas cobradas a los clientes, lo que a su vez incrementa la rentabilidad del negocio. Dentro de estos costos, el control y la optimización del rendimiento del combustible, medido en millas por galón o en kilómetros por litro, juega un papel crítico. Por ello, es esencial identificar los factores clave que influyen en el rendimiento del combustible para implementar acciones concretas que permitan maximizarlo. De esta manera, se pueden lograr costos más competitivos frente a los clientes y, al mismo tiempo, mejorar la rentabilidad de la empresa.

## Objetivos: ¿Cuál es la meta prevista?

* Analizar los datos disponibles para comprender la dispersión en el rendimiento del combustible y las causas que la originan.
* Desarrollar un modelo de clasificación con técnicas de aprendizaje no supervisado para identificar los factores más significativos que influyen en el rendimiento del combustible.
* Proponer alternativas de acciones concretas y aplicables al negocio para mejorar la eficiencia en el consumo de combustible. Asimismo, emplear los datos y modelos de inteligencia artificial para establecer metas alcanzables de consumo.
* Crear un dashboard de visualización basado en los datos disponibles, que proporcione insights clave sobre el comportamiento de la flota. Esto facilitará la toma de decisiones basadas en datos en tiempo real.
* Documentar la caracterización de los procesos asociados al problema, para reflejar la situación actual de la compañía en este ámbito. Esta documentación mejorará la comunicación interna y contribuirá a la conservación del conocimiento sobre los procesos de la organización.

## Preguntas clave: ¿Cuáles son las preguntas claves que deben responderse?

* ¿Cuáles son los factores relevantes que, a partir de los datos del sensor GeoTab, influyen en el rendimiento del combustible, y cómo impactan en la eficiencia?
* ¿Qué patrones o causas subyacentes se encuentran al analizar la dispersión en los datos de rendimiento de combustible, y cómo influyen en la variabilidad observada?
* ¿Cómo contribuye el modelo de clasificación basado en aprendizaje no supervisado a destacar los factores más significativos en el consumo de combustible, y qué descubrimientos se obtienen de este modelo?
* ¿Qué acciones concretas se proponen a partir del análisis de los datos y los modelos de inteligencia artificial, y de qué manera estas acciones mejoran la eficiencia del consumo de combustible en el negocio?
* ¿Cómo se determinan las metas de consumo de combustible basadas en los datos y modelos de IA, y en qué medida estas metas son alcanzables y relevantes para mejorar la rentabilidad del negocio?
* ¿De qué forma resulta útil el dashboard de visualización creado para ofrecer insights sobre el comportamiento de la flota, y cómo facilita la toma de decisiones basadas en los datos generados?
* ¿Cómo contribuye la documentación de los procesos relacionados con el problema a mejorar la comunicación interna y a conservar el conocimiento dentro de la organización, y qué aspectos específicos de la situación actual de la empresa se clarifican a través de esta caracterización?

## Involucrados

El departamento de Ciencia de Datos está involucrado en este proyecto. Abraham Rincón está a cargo del equipo y Sebastián Cardona actúa como asesor. Actualmente, se encuentran en una etapa de exploración de los datos y de propuestas de mejora para los procesos. Dado que no se tienen mapeados los procesos de envío de cargas (loads), se deberá considerar el mapeo de estos procesos e identificar los puntos críticos dentro de la operación para optimizarla y garantizar un flujo adecuado de información y recursos.

# Entendimiento de los datos

## Descripción de los datos:

El contexto de los datos se basa en la información obtenida de los dispositivos GPS del sistema Geotab. Estos datos están estructurados en tres tablas principales:

* **silver\_consumption:** Contiene información relacionada con el consumo de combustible u otros recursos del vehículo, posiblemente con datos como cantidad de combustible utilizado, eficiencia, y patrones de consumo. Estos datos ya han pasado por una limpieza y enriquecimiento para garantizar que estén listos para análisis detallados.
* **silver\_idling:** Esta tabla incluye datos sobre el tiempo de inactividad de los vehículos, como las instancias en las que el motor permanece encendido sin movimiento. Estos datos también han sido procesados para eliminar inconsistencias y prepararlos para un análisis más profundo.
* **bronze\_safety:** A diferencia de las tablas anteriores, esta contiene información en su estado crudo, sin procesamiento. Probablemente incluye eventos relacionados con la seguridad, como frenadas bruscas, aceleraciones rápidas o cualquier otra métrica relacionada con el comportamiento del conductor y la seguridad del vehículo.

Aunque las tablas se encuentren en la capa silver, es necesario que pasen por un control adicional de calidad para garantizar que los datos estén libres de anomalías y duplicados. Este proceso asegura que los datos sean precisos, consistentes y adecuados para el análisis o para su uso en modelos de machine learning (ML) y herramientas de Business Intelligence (BI). Esto es esencial para evitar problemas en los resultados finales derivados de posibles errores en los datos originales.

Estos datos corresponden a los meses de julio y agosto, y aún deben ser procesados y transformados para generar resultados que puedan ser utilizados en análisis de negocio (BI) o en modelos de machine learning (ML).

## Descripción de la arquitectura

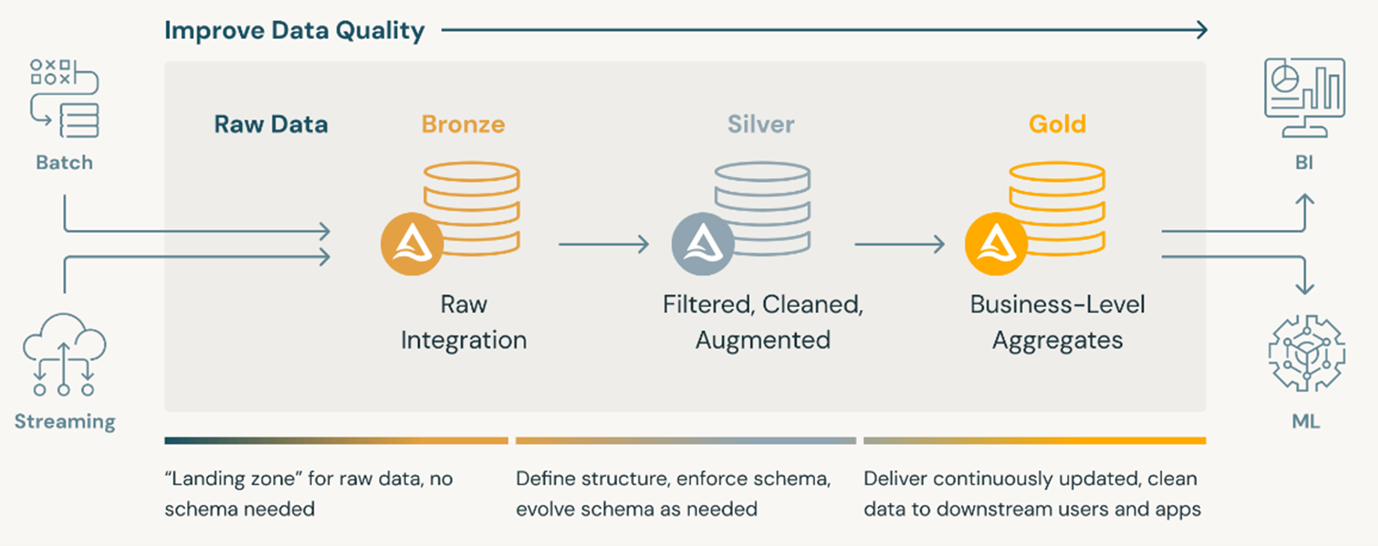


Figura Arquitectura Medallón

La arquitectura representada en la Figura 1, sigue un enfoque típico de procesamiento de datos basado en la mejora de la calidad de estos a través de diferentes capas o fases, desde la ingesta de datos hasta la entrega de resultados finales para su consumo en análisis de negocio o modelos de machine learning (ML). A continuación, se detallan las partes clave de la arquitectura:

### Fases del Procesamiento de Datos

1. **Ingesta de Datos Crudos (Raw Data)**:
   * **Origen de los Datos**: Los datos provienen de dos fuentes principales: procesamiento por lotes (batch) y procesamiento en tiempo real o continuo (streaming).
   * **Integración de Datos en Bronce**: En esta fase, los datos son integrados en bruto, sin que se apliquen transformaciones ni estructuras predefinidas. Se almacena en un "landing zone" o zona de aterrizaje donde no se necesita un esquema predeterminado. Es la fase inicial donde los datos se capturan directamente desde su origen.
2. **Capa de Bronce (Bronze Layer)**:
   * Los datos en crudo son almacenados en su forma más básica, sin estructura ni procesamiento adicional.
   * **Propósito**: Capturar la data tal como llega, para poder ser procesada posteriormente. Es ideal para realizar auditorías o revisiones a nivel crudo.
3. **Capa de Plata (Silver Layer)**:
   * **Transformación**: En esta etapa, los datos son filtrados, limpiados y enriquecidos (augmented). Aquí se aplican estructuras o esquemas para darles forma y se depuran errores, duplicados o inconsistencias en los datos.
   * **Propósito**: Los datos son procesados para asegurar que cumplan con ciertos estándares de calidad, siendo aptos para su consumo analítico. También se puede aplicar integración con otras fuentes de datos en esta fase.
4. **Capa de Oro (Gold Layer)**:
   * **Datos Agregados a Nivel de Negocio**: En esta fase, los datos están listos para ser consumidos a nivel de negocio, ya que han sido transformados en agregados significativos y relevantes para los usuarios finales.
   * **Propósito**: Facilitar la entrega de información clave a los usuarios finales (BI, ML, etc.), con un enfoque en obtener métricas y reportes para la toma de decisiones. Los datos están completamente procesados y son fácilmente accesibles para el consumo de las aplicaciones downstream.

### Salidas o Consumo

* **Business Intelligence (BI)**: Los datos se entregan a las herramientas de BI para su análisis, generación de reportes y visualizaciones.
* **Machine Learning (ML)**: Los datos también se pueden integrar en modelos de machine learning, alimentando algoritmos para la creación de predicciones o insights automáticos.

### Enfoque General

* **Mejora Continua de la Calidad de Datos**: A lo largo del proceso, la arquitectura busca mejorar la calidad de los datos mediante un ciclo de procesamiento que va desde la captura en bruto hasta la refinación y entrega de resultados listos para el análisis.
* **Evolución del Esquema**: Conforme los datos pasan por las diferentes capas, los esquemas se definen y evolucionan para asegurar que los datos filtrados y limpiados tengan la estructura adecuada para su procesamiento posterior.

Esta arquitectura sigue el patrón de **"Data Lakehouse",** donde se combinan los beneficios de un data lake para almacenar datos en bruto y de un data warehouse para proporcionar datos estructurados y listos para su uso en análisis avanzados.

## Ejemplo de datos

### Técnica de ML:

* **Supervisado (Regresión, Clasificación)**:
  + **Regresión**: Predecir valores continuos como el costo de ralentí, el consumo de combustible o la distancia recorrida.
  + **Clasificación**: Clasificar comportamientos de manejo o niveles de riesgo basados en las excepciones de gestión de riesgos (e.g., "excesive speeding", "hard braking").
* **No supervisado**: Puede aplicarse para análisis de clusters que identifiquen patrones similares de comportamiento de conducción sin etiquetas predefinidas.

### Identificación de Variables:

**Entradas:**

* **unique\_id**: Identificador único del vehículo o evento.
* **date**: Fecha en la que se registraron los datos.
* **name**: Nombre o número del vehículo.
* **distance**: Distancia recorrida.
* **idlingDuration**: Duración del tiempo en ralentí.
* **idlingPercent**: Porcentaje del tiempo total que estuvo en ralentí.
* **CurrentFuelPrice**: Precio actual del combustible.
* **RiskManagementExceptionRule2,3,4**: Excepciones en la gestión de riesgos (e.g., exceso de velocidad, frenado brusco, aceleración brusca).
* **RiskManagementExceptionRule2Count,3Count,4Count**: Número de veces que ocurrieron esas excepciones.

**Salida (posibles objetivos de predicción):**

* **IdlingCost**: Costo del ralentí (variable de salida en modelos de regresión).
* **performance**: Puede ser una variable de salida en modelos que busquen optimizar el rendimiento del vehículo.
* **RiskManagement**: Se puede utilizar para predicción de comportamientos riesgosos (e.g., modelos de clasificación para predecir si un conductor caerá en exceso de velocidad o frenado brusco).

Este análisis permite una evaluación del comportamiento de los vehículos y conductores, optimización de consumo de combustible y gestión de riesgos en la operación.

# Bibliografía

* Visengeriyeva, L., Kammer, A., Bär, I., Kniesz, A., y Plöd, M. (2023). CRISP-ML(Q). The ML Lifecycle Process. MLOps. INNOQ. https://ml-ops.org/content/crisp-mlLinks to an external site.
* Mssaperla. (2024, 1 marzo). *What is the medallion lakehouse architecture? - Azure Databricks*. Microsoft Learn. <https://learn.microsoft.com/en-us/azure/databricks/lakehouse/medallion>

# Anexo

A blue and green cover with white text

Description automatically generated

A document with text on it

Description automatically generated

A document with signature on it

Description automatically generated