

# EXAMEN CIENCIA DE DATOS L3/L4

Agosto 2021

HENRY SANCHEZ ALVARADO

# Contenido

**1 – Sección A**

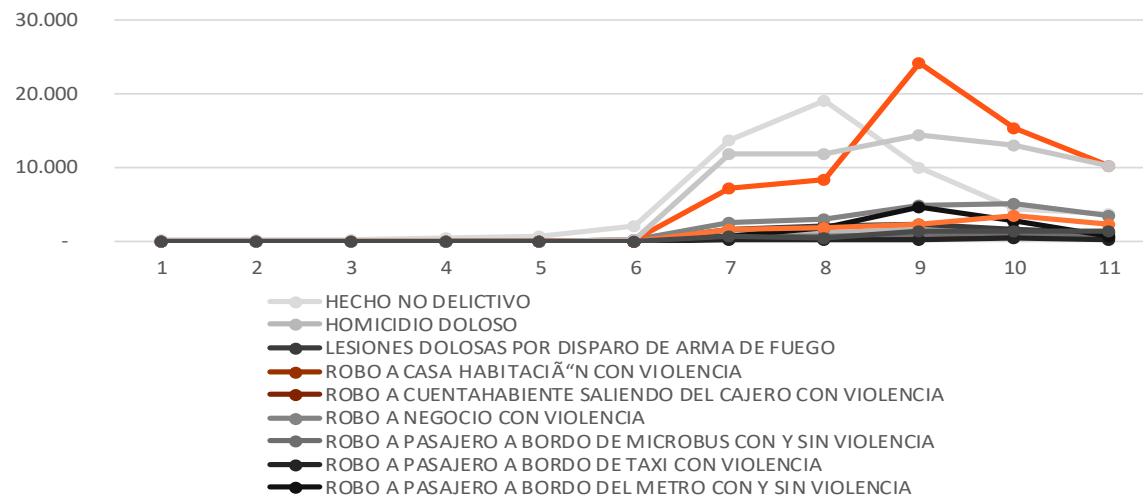
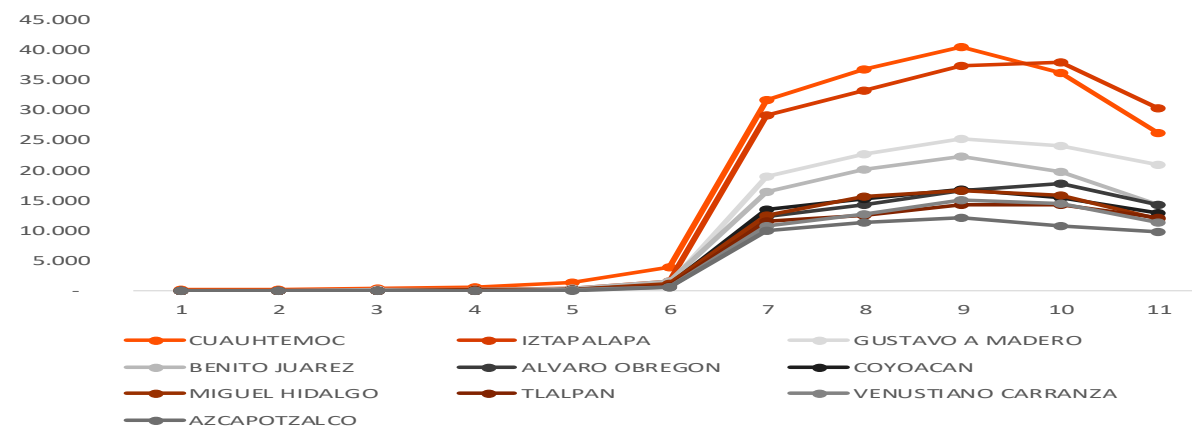
**2 – Sección B**

## ¿Qué pruebas identificarías para asegurar la calidad de estos datos?

1. Pruebas de Complitud, para asegurar que cada registro tenga los datos minimos necesarios para el análisis, es decir pordemos tener la Alcaldía y la Categoría del Delito, pero si no temenos el delito como tal, este registro tendría que ser eliminado para un análisis minucioso.
2. Pruebas de validez, para segurar que los valores de cada campo sean consecuentes. Es decir, edades positivas, si son datos históricos deberían ser máximo del año 2021, las ubicaciones geográficas que correspondan a la alcaldía.
3. Pruebas de coherencia, para asegurar la calidad de los datos se require que los campos contenga información acorde a ese campo. Es decir, que Alcaldía tenga solo alcaldías y no otra información.
4. Pruebas de precision, para asegurar que los datos representan el valor del objeto en análisis.

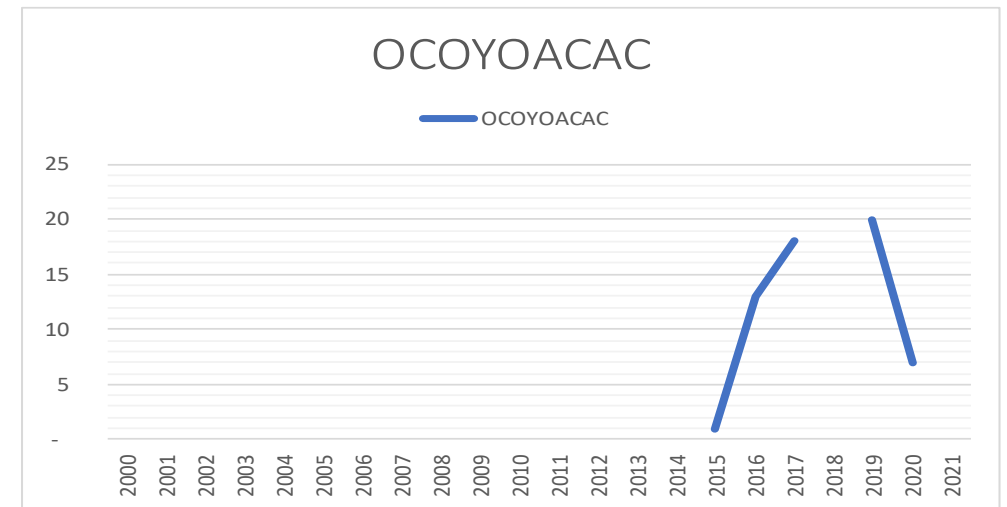
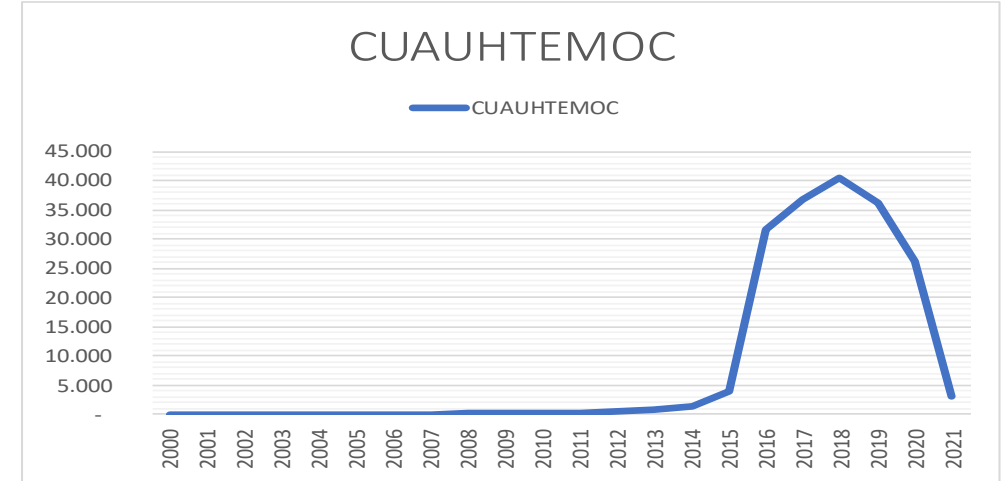
# Identifica los delitos que van a la alza y a la baja en la CDMX

1. Para este análisis consider los datos desde el 2000 hasta el 2020, excluyo el 2021 porque estoy analizando los datos de manera anual y además consider las 10 principales alcaldías debido a que concentran la mayor cantidad de incidencias.
2. Como se puede observar en el gráfico, para el año 2020 casi todas las alcaldías presnetaron una baja en sus indicencias totales, este fenomeno se puede explicar con las cuarentena y otras politicas frente a la pandemia del COVID 19.
3. Para revisar los datos por categoría de delito, se han excluido aquellos que presentan una baja incidencia, y Tambien se excluyó DELITO DE BAJO IMPACTO, debido a que concentra la mayor cantidad de incidencias. Y en este caso se considero desde el 2010 hasta el 2020. fecha en la que tambien se observa una tendencia a la baja.



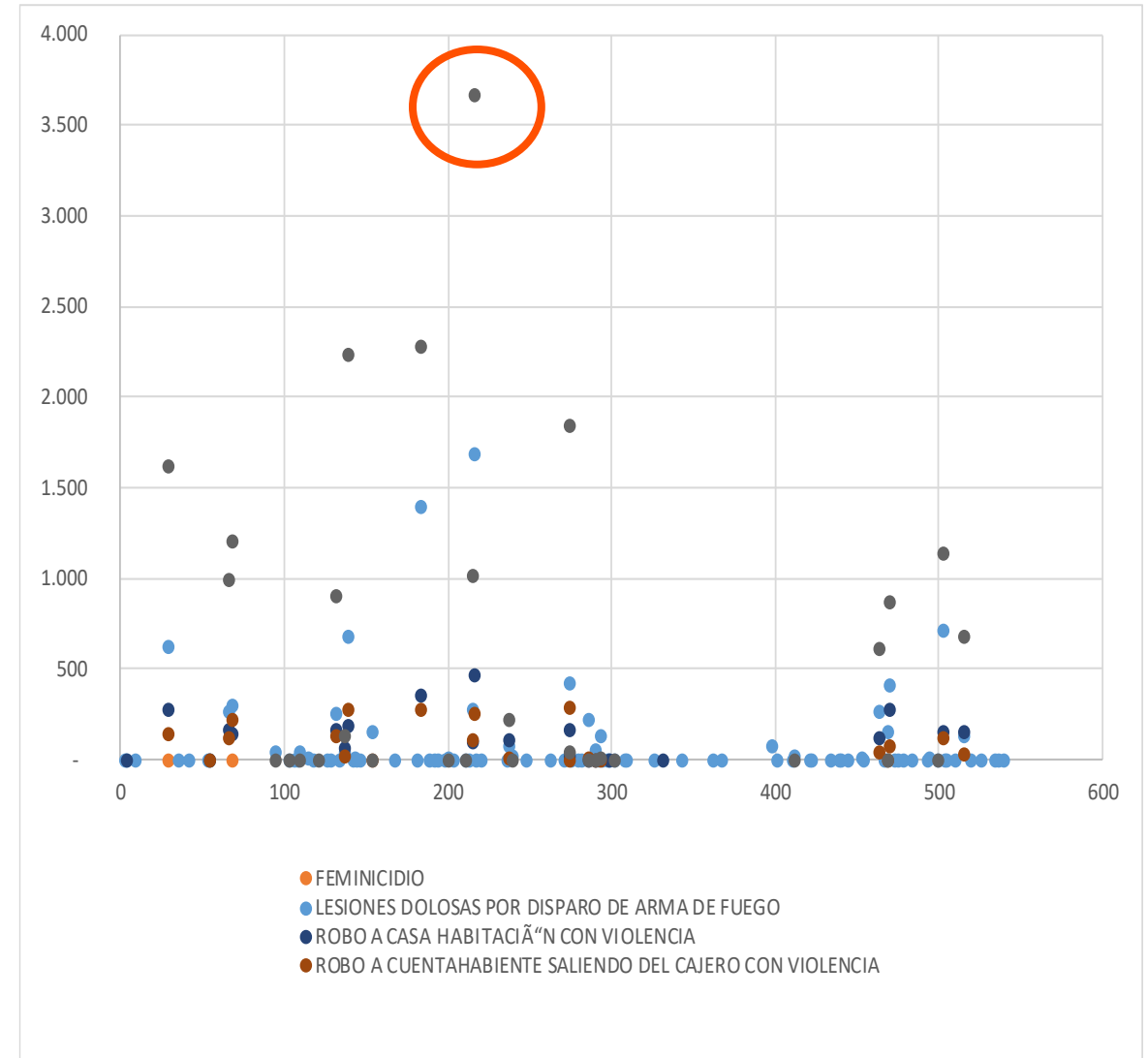
¿Cuál es la alcaldía que más delitos tiene y cuál es la que menos? ¿Por qué crees que sea esto?

- Según la información analizada, Cuauhtemoc es la alcaldía que presenta una mayor frecuencia de delitos (considerando los últimos 11 años).
- Las razones pueden ser diversas, en Cuauhtemoc se encuentra desde Tepito hasta los edificios más altos de la CDMX sobre la Avenida reforma. Según los informes periodísticos, en esta alcaldía se presenta desde delincuencia común, pasando por narcomenudeo hasta crimen organizado. Al concentrar muchas de las actividades principales de la economía Mexicana, es normal que también se concentre los eventos delictivos.
- Por otro lado, en la base de datos pude observar que hay alcaldías que no pertenecen a la CDMX y que presentan muy pocos datos. He considerado como 59 el número mínimo de incidencias. Y corresponde a la alcaldía de Ocoyoacac. La razón puede ser la ubicación geográfica de esta, pues se encuentra en las proximidades de Toluca (alcaldía que también presenta incidencias bajas).



¿Cuál es la alcaldía que más delitos tiene y cuál es la que menos? ¿Por qué crees que sea esto?

- Para este análisis, he optado por retirar los delitos que presentan mucha frecuencia en todas las alcaldías (DELITO DE BAJO IMPACTO, ROBO DE VEHÍCULO CON Y SIN VIOLENCIA, ROBO A TRANSEUNTE EN VÍA PÚBLICA CON Y SIN VIOLENCIA, etc).
- Este filtrado lo hago con el fin de encontrar realmente aquellos delitos que SI resaltan en una alcaldía.
- El hallazgo principal es que en Iztapalapa presenta una mayor incidencia en ROBO A NEGOCIO CON VIOLENCIA.
- Es decir, en esta alcaldía este delito es el más representativo frente a las otras alcaldías.



## Diseña un indicador que mida el nivel de “inseguridad”

- Un indicador interesante sería medir las incidencias por cada 1000 ó 100,000 habitantes, pero al no contar con la población por alcaldía este indicador queda descartado.
- Dado que tenemos las fechas de la incidencia, el indicador que propongo es:

Número de delitos en los últimos 3 meses

- Este indicador es dinámico, dado que se cada día se puede actualizar considerando los últimos 90 días.

Alcaldía	Número Delitos- 3 meses
IZTAPALAPA	7.612
CUAUHTEMOC	6.378
GUSTAVO A MADERO	5.181
ALVARO OBREGON	3.757
BENITO JUAREZ	3.537
COYOACAN	3.314
MIGUEL HIDALGO	2.979
TLALPAN	2.916
VENUSTIANO CARRANZA	2.758
AZCAPOTZALCO	2.644
IZTACALCO	1.985
XOCHIMILCO	1.668
TLAHUAC	1.507
LA MAGDALENA CONTRERAS	1.106
CUAJIMALPA DE MORELOS	861

# Contenido

**1 – Sección A**

**2 – Sección B**



# Ruta para desarrollar una solución al problema de La Michoacana

1




Establecemos objetivos

2



Proponemos una metodología

3




Identificamos insights

4



Feedback para involucrar a las personas

5



Lanzamiento de la solución

## a. ¿de qué tipo de problema se trata?

- Se trata de un problema de investigación de operaciones, en la que buscamos minimizar los costos de reposición de paletas en las máquinas de La Michoacana sujeto a algunas restricciones y objetivos.
- En el caso se menciona que existe información histórica diaria desde hace 5 años, entonces surge la posibilidad de utilizar modelos de Series Temporales (modelos GARCH debido a la varibilidad de los datos).
- Debido a que contamos con datos diarios y se requieren indicadores para tomar decisiones en el día a día, debemos considerar la posibilidad de aplicar modelos de segmentación o clasificación para los CLIENTES, UBICACIÓN DE MÁQUINAS, ETC.
- Se debe buscar un balance para máquina entre costo de transporte y costo de energía.

### Diagnóstico y objetivos

- **4000** máquinas instaladas.
  - **1 MXN** es el costo de mantener cada paleta.
  - **100 MXN** costo para reponer cada máquina.
  - **OBJETIVOS:**
    - **Mínimizar costo de transporte y energía**
    - Mantener días/máquina sin disponibilidad menor al 2%.
- 5 años de** información histórica.

### Hechos

- Llenar la maquina a tope tiene costos altos de energia y costos bajos de transporte
- Llenar las maquinas según la necesidad del día siguiente tiene un costo de energía bajo y un coste de transporte alto

Sea:

$x_i$ : Número de paletas disponibles en la máquina  $i$

$r_i$ : Número de paletas para reponer en la máquina  $i$

$T_i$ : Capacidad total de la máquina  $i$

$d_i$ : variable dummy (0 cuándo no se repone y 1 cuando se repone la máquina)

Entonces se cumple que  $T_i = x_i + r_i$

*Costo de la máquina  $i$*

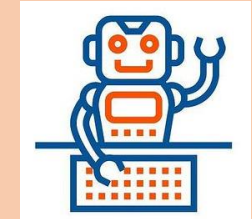
$$\text{Costo}_i = \text{CostoFijo}_i + \text{CostoVariable}_i$$

$$\text{Costo}_i = 1\text{MXN} * x_i + 100\text{MXM} * d_i$$

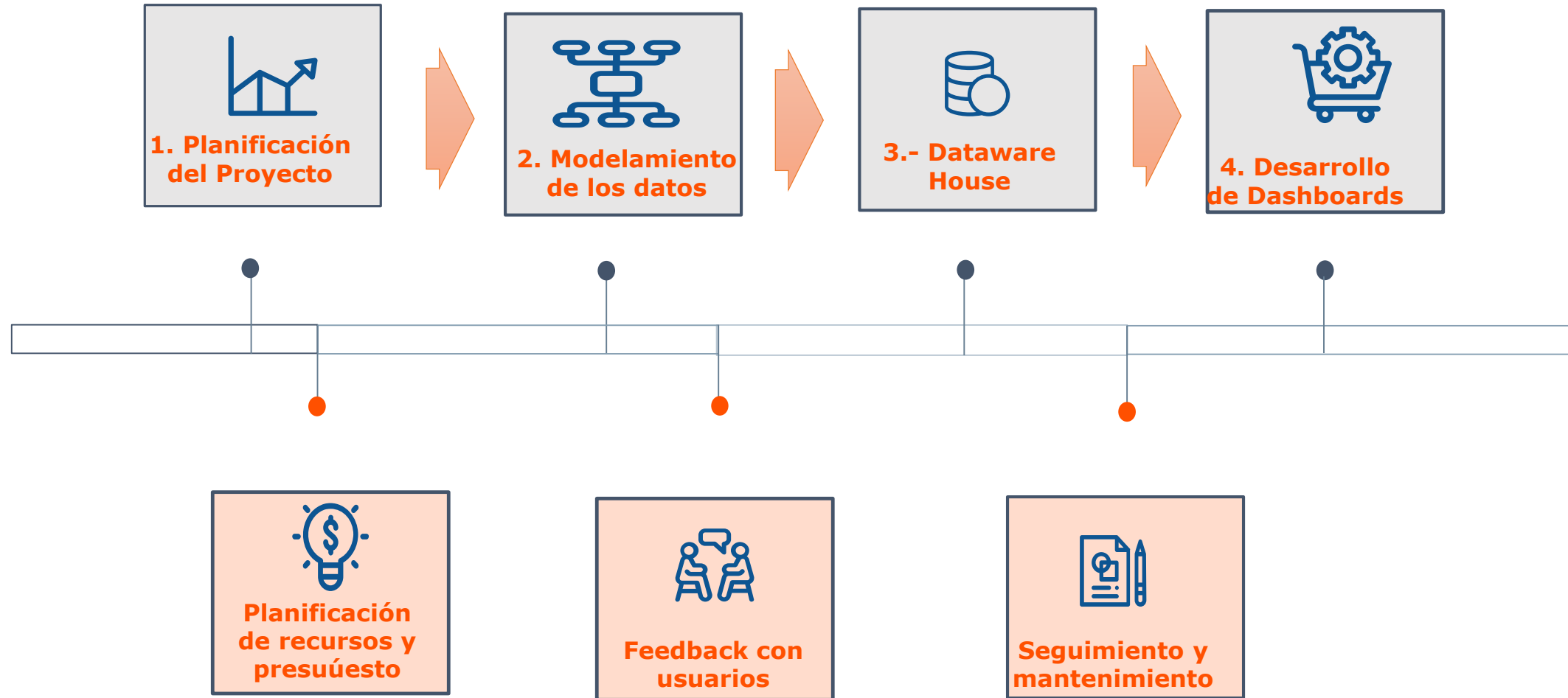
$$\text{CostoTotal} = \sum_{i=1}^n x_i + \sum_{i=1}^n 100 * d_i$$

b. ¿Cómo te imaginas una solución funcional al problema? ¿De qué partes está conformada? ¿Cómo interactúan esas partes? ¿Qué supuestos y riesgos ves en tu planteamiento?

- La Michoacana dispone de una Base de Datos que se actualiza diariamente, debe tratarse de algún ERP cómo SAP o similar.
- Habría que verificar si estos datos están bajo un modelo eficiente y si es necesario se debe considerar la implementación de un Dataware House
- Luis necesita indicadores y métricas para tomar decisiones en el día a día, por tal razón se puede proponer la implementación de reportes Dashboard que se actualicen de forma automática todos los días.
- Estos reportes Dashboard deben contener los algoritmos y modelos que se plantearon en la optimización de este caso, por tanto se debe considerar un procedimiento ETL para convertir los datos del ERP en información que se mostrarán en los reports Dashboard.
- Estamos suponiendo que los datos están alojados en un ERP como SAP, tal vez esta no sea la realidad de La Michoacana y se deba implementar algún Sistema para tener la actualización de los datos y poder hacer la conexión con los Dashboard.
- El riesgo operative está latente en cada etapa de está solución (error humano, falla en los procesos)



# Diagrama de la solución propuesta



### c. ¿Qué métodos o algoritmos utilizarías durante el Desarrollo de esa solución?

- Una de las principales variables identificadas es el número de paletas disponibles en cada máquina y se trata de una serie temporal.
- Se puede modelar y encontrar un modelo matemático para estimar el consumo de paletas para cada máquina según sus características.
- Al tratarse de un modelo de optimización, el coste y la energía se podrían minimizar utilizando métodos de optimización dinámica o investigación de operaciones.

Se debe hacer seguimiento a las siguientes variables:

- Los indicadores de costos de transporte y energía por máquina.
- El número de Paletas disponibles y la capacidad total de la máquina.
- El consume promedio por máquina.
- La calidad de los datos.
- Las variables meterológicas de las ubicaciones de las máquinas.
- El patron de consume de los suscriptores.
- Alertas tempranas para medir el desabastecimiento de las máquinas.



## Costos



- Revisión diaria del avance real de los costos.
- Seguimiento de indicadores.

## Comité Compra Cierta



- Digitalización del registro de solicitudes para el CCC.
- Seguimiento y revision de las solicitudes para el CCC.

## Forecast y Budget



- Digitalización de formularios.
- Revisión y seguimiento.
- Alertas tempranas.

## Contratos (en curso)



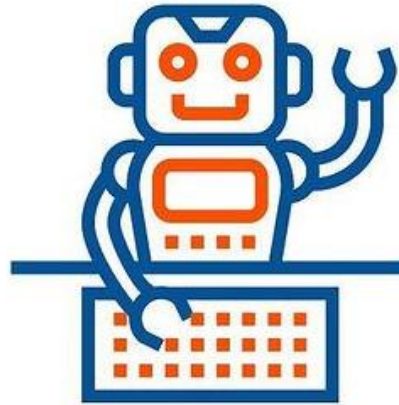
- Desarrollar módulos para el seguimiento de contratos.
- Desarrollo de modulo para licitaciones.

1. Empezamos con identificando la situación actual de La Michoacana.
2. Identificamos los objetivos y las actividades.
3. Definimos los recursos, roles y los presupuestos necesarios.
4. Elaboramos el cronograma de reuniones y entregables con el cliente.
5. El feedback del cliente en todo momento es importante para no caer en retrabajo.
6. Por la naturaleza del problema los recursos que serán necesarios serán:
  - Gerente de Proyecto
  - Analista de Datos
  - Ingeniero de Datos
  - Desarrolladores

**Toma de decisiones  
(La Michoacana)**



**Automatización de  
reportes  
(OPI Analitics)**



**Digitalización  
(OPI Analitics y La  
Michoacana)**



- Una vez que se construyeron los modelos y los indicadores es importante hacer seguimiento a estos.
- Se puede establecer metas para cada indicador.
- El problema plantea una meta principal, el desabastecimiento día/máquina menor al 2%.
- En el día a día se deben hacer los ajustes y mejoras necesarias para mantener los objetivos.

# ¿Cómo podemos medir el impacto?



## **Impacto**

¿Podemos definir claramente en esta iniciativa el valor para el negocio?



## **Factibilidad**

¿Podemos hacerlo? ¿Se justifica llevarlo a cabo?



## **Soporte**

¿Contamos con los recursos adecuados para hacer de esta iniciativa un éxito?

