



# ENTREGA 3

PROYECTO FINAL

Portafolio de Soluciones de Analítica

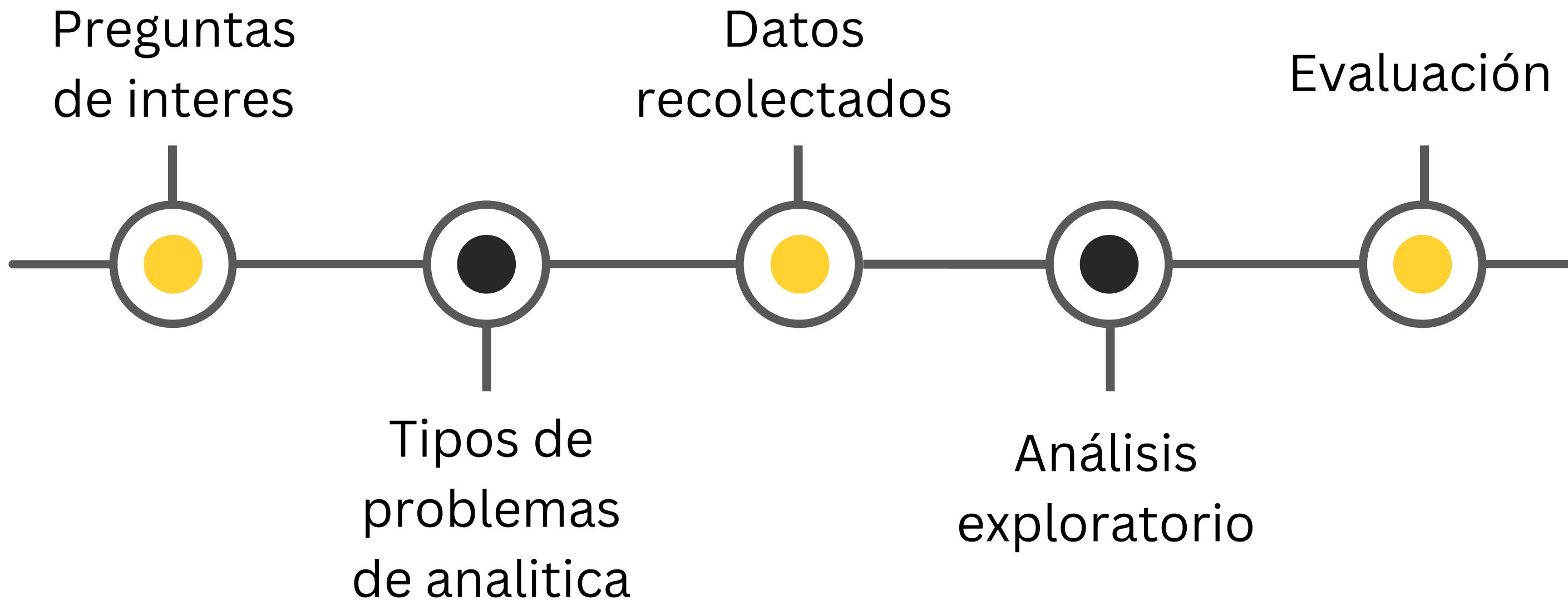
# Introducción

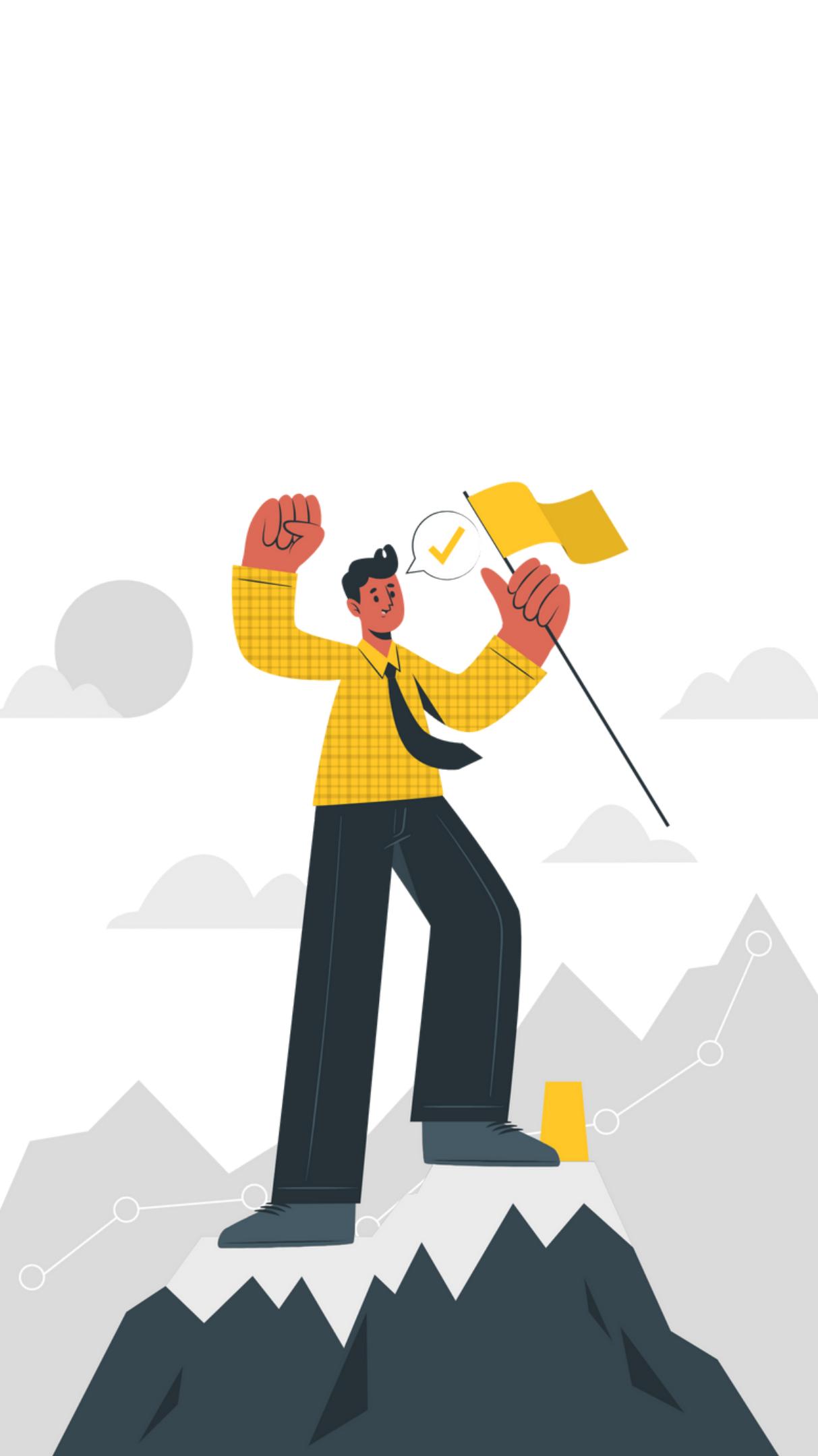
*LIMPIK: e-commerce que vende productos de la canasta familiar, vinos y bisutería.*

Realizan ventas en línea, por catálogo y en su tienda física. Tienen interés en aplicar analítica e inteligencia artificial en su negocio.



# METODOLOGIA





# Objetivos del negocio

*Tienen como principal objetivo:*

Ampliar el entendimiento sobre sus clientes actuales y nuevos que ingresan a la compañía.

Han realizado estudios de mercado para llegar a perfilar clientes y definir otros aspectos.

# Preguntas de interés



1

¿Cuál de mis medios de venta es más frecuentado?

2

¿Qué perfiles de cliente tenemos?

3

Identificar clientes que aceptarían una campaña de mercado a futuro

4

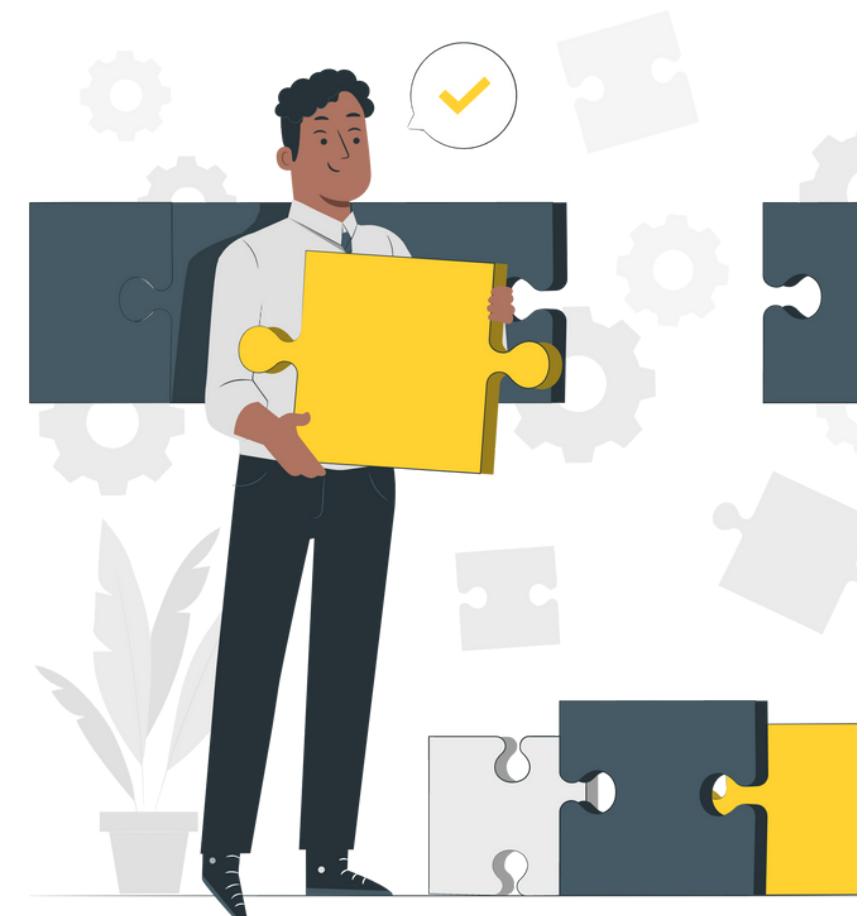
¿Los niños afectan en el dinero gastado?

## Tipos de problema de analítica

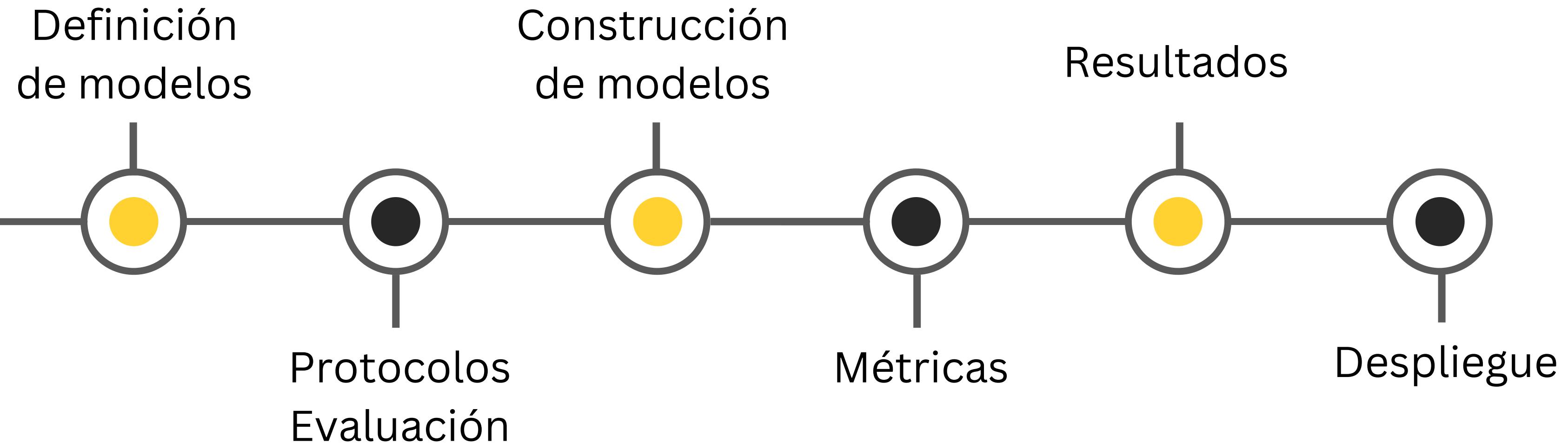
En base a lo anterior, podemos determinar los objetivos de analítica que se definirán para el proyecto:

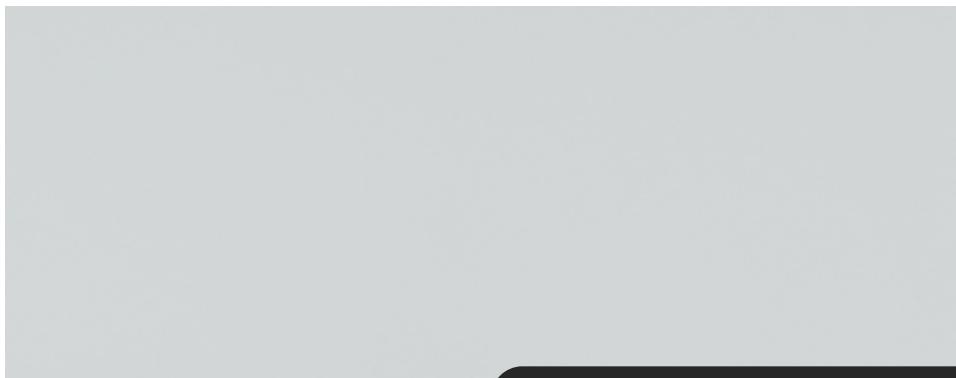
- ✓ Elaborar un modelo de Modelos Random Forest para predecir cuántas compras realizará un cliente por medio del catálogo, página web y tienda física.
- ✓ Elaborar un modelo de clasificación para identificar los usuarios que podrían aceptar una campaña de marketing/oferta a futuro, en base al número de campañas aceptadas y diferentes características del comprador.

- ✓ Construir un modelo de clasificación de clientes acorde a los artículos que adquiere para generar perfiles de compradores
- ✓ Elaborar un modelo de clasificación por perfiles para predecir el dinero gastado por un cliente del comercio, en base al número de niños y jóvenes que viven en su hogar.



# FLUJO DE ANALÍTICA





# Modelos

*Técnicas de modelamiento*

Elaboración de diferentes modelos de predicción que nos faciliten llegar a una respuesta correcta a los objetivos de negocio

Regresión Lineal

XGBoost

Decision Tree

SVM



Random Forest



K-Means

[1-0]

Regresión Logística

# Protocolos

*Protocolos de evaluación para los  
modelos*



## Random Forest

$R^2$ , MSE y RM SE.



## K-Means

Estimadores: método de codo, silueta y el índice Calinski-Harabasz



## Regresión logística

Exactitud, precisión, TPR, FPR, puntaje f1 y la matriz de confusión.



## Veamos Jupyter!

En nuestros notebooks se encuentra todo el código fuente y proceso de modelado para cumplir los objetivos de analítica

# Resultados

Con base en los modelos y evaluaciones hechas, llegamos a los siguientes resultados

-  Se obtuvo un perfil de comprador el cual es casi seguro que no aceptaría una campaña
-  Se obtuvo un perfil de comprador que posiblemente acepte campañas
-  Se obtuvo que la cantidad de jóvenes y niños por hogar, no es un factor determinante en el consumo
-  Tenemos un error cuadrático demasiado alto para nuestra predicción de consumo de acuerdo a la cantidad de jóvenes y niños por hogar.
-  Al intentar realizar predicciones de ventas con respecto a la tienda fisica no seran tan exactos como al realizar predicciones por catalogo
-  Por el bajo ajuste de los modelos al intentar predecir la cantidad de comprar que un cliente realizara por cualquier medio, no se realizara una predicción efectiva.
-  Definimos 3 perfiles de compradores diferentes, cada uno con sus respectivas características

# Despliegue

*Plan de Despliegue*

Usar una página web simple para consumir los diferentes modelo como recursos, en donde se podrá interactuar ingresando los datos de un comprador.

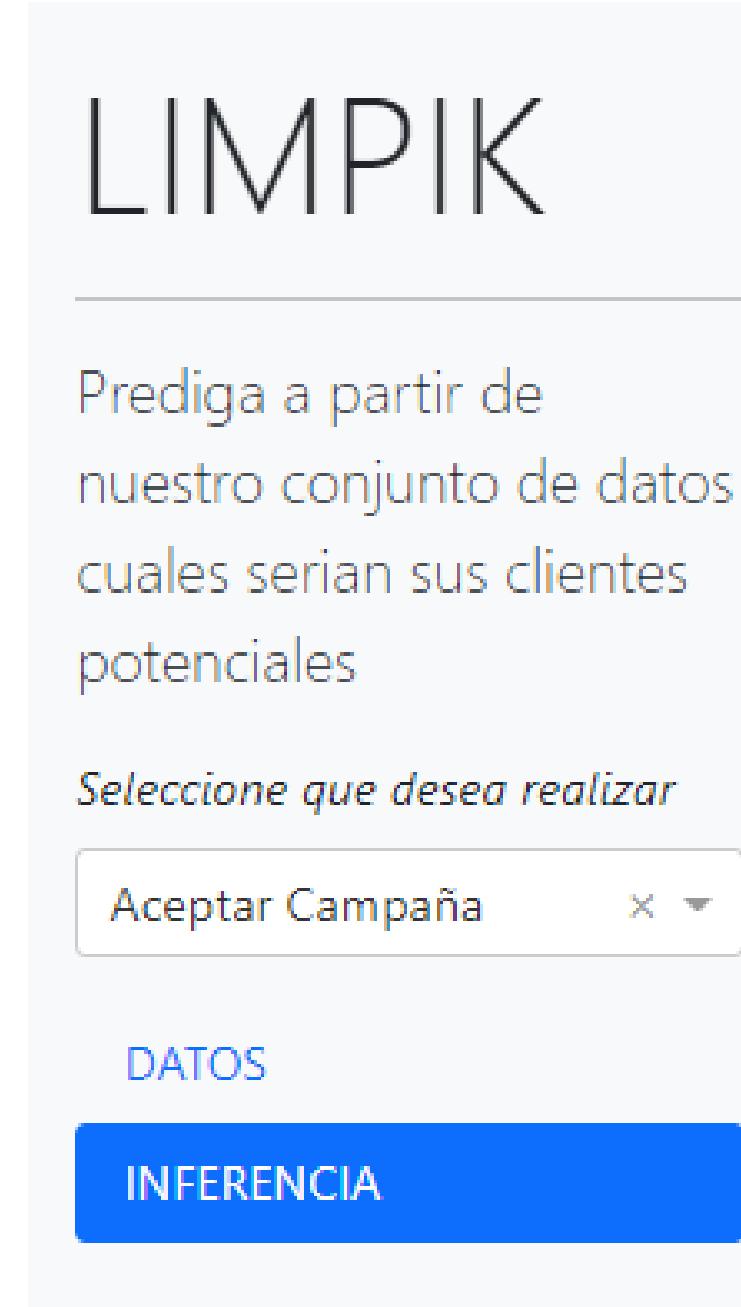


 Número de compras

 Perfiles compadores

 Aceptar Campañas

 Dinero Gastado



## Servicios

*Servicios desplegados en página web*

Se realizó un despliegue usando Flask y Dash. La página web es sencilla, permite seleccionar el tipo de predicción a realizar (INFERENCIA) o si desea ver el set de datos usado (DATOS)

**La importación de los modelos a la web se realizó con joblib**



# Análisis Final

*Resultados obtenidos vs Objetivos de Negocio*

Finalmente se generaron y desplegaron 6 modelos, de los cuales los modelos de regresión para el objetivo 1 no se acoplan en gran medida a los datos (**máx. 74%**)

# Análisis Final



## Segmentación Adecuada

Por medio de los modelos K-Means de los objetivos de analítica 2 y 4



## Clasificación en campañas

Se logró elaborar un modelo eficiente para identificar clientes que no aceptarían, pero no para quienes sí aceptan campañas



## Nivel de Confiabilidad

Con los modelos generados si bien se cumple, no se otorgan niveles de confianza alto para tomar decisiones

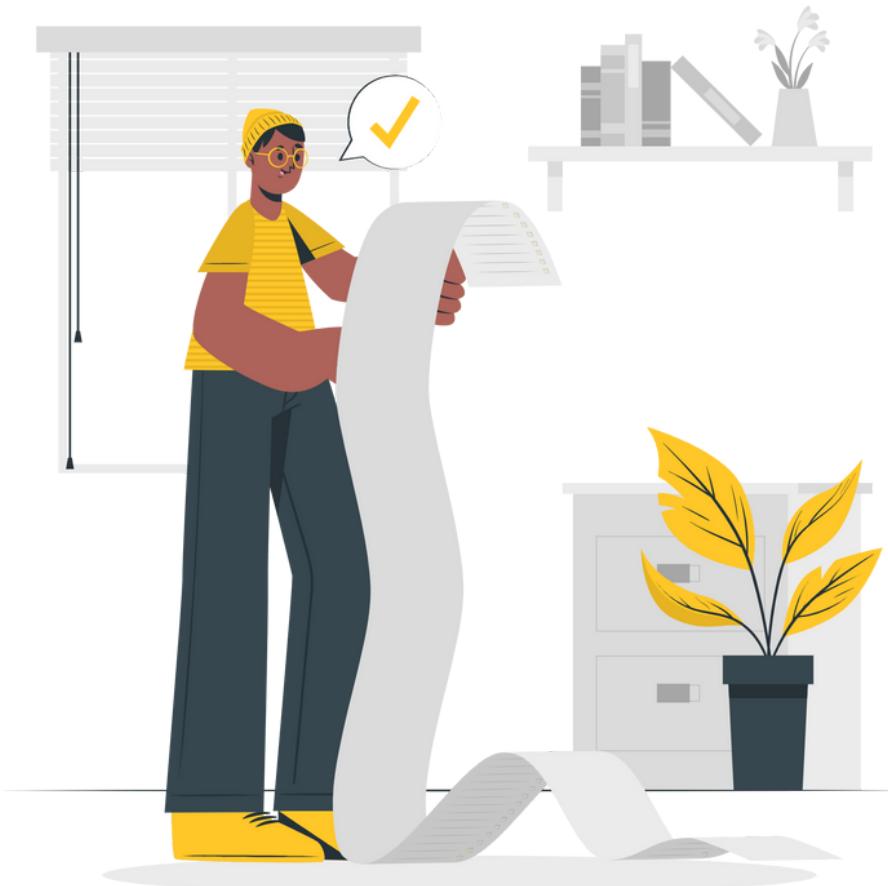


## Cimientos

Se han logrado unos cimientos para trabajos futuros en base a un mayor entendimiento de los clientes

**¡Se cumplieron los  
objetivos de analítica  
y negocio!**

Más o menos...



Pero



### Ingeniería de atributos

Realizar este proceso para obtener más información del dataset y poder realizar mejores predicciones



### Métodos más avanzados

Emplear modelos diferentes a los tradicionales para poder construir soluciones más precisas y efectivas

## Oportunidades de mejora

# GRACIAS!

PROYECTO FINAL

