

DATA OPS MIND

Fueling innovation through data intelligence



EQUIPO:



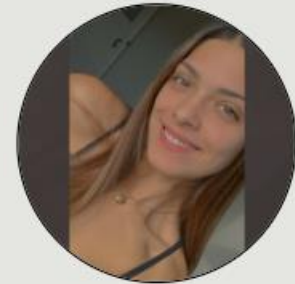
Martin Ushiñahua

Data Scientist



Rossella Manzi

Functional Analyst



Clara Santillan

Data Analyst



José Abrahan Bonilla

Data Engineer



Jorge Luis Garcia

Data Scientist

NECESIDADES EMPRESARIALES

- Fortaleciendo la toma de decisiones basadas en datos.



OPORTUNIDADES DE INVERSIÓN

- Datos de Yelp y Google Maps
- Análisis de sentimientos
- Datos demográficos

MEJORAR EL DESEMPEÑO DE NEGOCIOS LOCALES

- Encontrar áreas de mejora
- Entender preferencias del cliente
- Estrategias de fidelización.



CAMPAÑAS DE MARKETING

- Sistema de recomendación
- Maximizar el retorno de la inversión
- Segmentación del público
- Decisiones basadas en datos

DATA OPS MIND

- DIAGRAMA DE GANTT



SPRINT 1:



Sprint 1, Puesta en marcha el proyecto y trabajando con los ...			28h	31/jul	10/Aug	0%	
1	Entendimiento de las necesidades	Martin	3h	31/jul	02/Aug	0%	Ma
2	Objetivos: específicos del trabajo y del grupo	Jorge	5h	02/Aug	03/Aug	0%	Jo
3	Alcance y fuera de alcance	Jorge	2h	03/Aug	04/Aug	0%	Jo
4	KPI's	Ro, Cl	10h	04/Aug	06/Aug	0%	Rr Cl
5	Crear repositorio en Github	José A	1h	02/Aug	02/Aug	0%	JA
6	Stack tecnológico	José A	2h	04/Aug	04/Aug	0%	JA
	Entregables:	Todos	-	03/Aug	07/Aug	0%	To
8	Más de 4 KPI's (ver doc sobre KPI en slack)	Todos	-	03/Aug	03/Aug	0%	To
9	Tecnologías a usar	Todos	-	07/Aug	07/Aug	0%	To
10	Documentación de alcance del proyecto	Todos	-	07/Aug	07/Aug	0%	To
11	Eda preliminar, calidad de dato	Todos	-	07/Aug	07/Aug	0%	To
12	Repo de github.	Todos	-	07/Aug	07/Aug	0%	To
13	Planificación y estimación de esfuerzos. Gantt	Todos	-	07/Aug	07/Aug	0%	To
14	Presentación para la demo	José A	5h	08/Aug	08/Aug	0%	JA
15	DEMO	Todos	-	10/Aug	10/Aug	0%	To

Sprint 1, Puesta en marcha el proyecto y trabajando con los datos:	
Entendimiento de las necesidades	
Objetivos: específicos del trabajo y del grupo	
Alcance y fuera de alcance	
KPI's	
Crear repositorio en Github	
Stack tecnológico	
Entregables:	
Más de 4 KPI's (ver doc sobre KPI en slack)	
Tecnologías a usar	
Documentación de alcance del proyecto	
Eda preliminar, calidad de dato	
Repo de github.	
Planificación y estimación de esfuerzos. Gantt	
Presentación para la demo	
DEMO	

SPRINT 2:



Sprint 2, Data engineering:		45h	14/Aug	24/Aug	0%		
18	Arquitectura propuesta y diagrama	José A.	10h	14/Aug	14/Aug	0%	Ja
19	Diccionario de datos	Martin	10h	16/Aug	16/Aug	0%	Ma
20	Creación de DW	JA, Jo	20h	18/Aug	18/Aug	0%	Ja, Jo
21	Automatizar DW, carga incremental.	Todos	5h	19/Aug	19/Aug	0%	To
Entregables		Todos	-	21/Aug	21/Aug	0%	To
23	DW automatizado con carga inicial	Todos	-	21/Aug	21/Aug	0%	To
24	Al menos dos tablas de hecho y 5 dimensionales (aplic...	Todos	-	21/Aug	21/Aug	0%	To
25	Carga incremental del dato (video)	Todos	-	21/Aug	21/Aug	0%	To
26	Uso de herramientas de BigData o servicios Cloud.	Todos	-	21/Aug	21/Aug	0%	To
27	DEMO	Todos	-	24/Aug	24/Aug	0%	To

Sprint 2, Data engineering:	
Arquitectura propuesta y diagrama	
Diccionario de datos	
Creación de DW	
Automatizar DW, carga incremental.	
Entregables	
DW automatizado con carga inicial	
Al menos dos tablas de hecho y 5 dimensionales (aplica?)	
Carga incremental del dato (video)	
Uso de herramientas de BigData o servicios Cloud.	
DEMO	

In Sync

SPRINT 3:



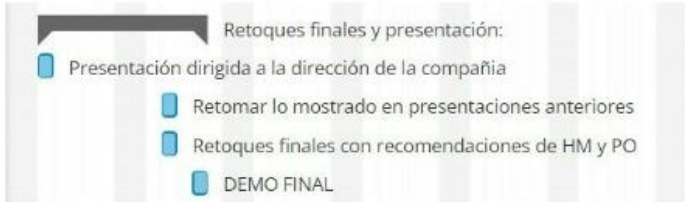
Sprint 3, Etapa de Analytics y ML:			79h	28/Aug	12/Sep	0%	
30	✓ Dashboard	Ro, CI	20h	28/Aug	04/Sep	0%	Rx CI
31	✓ Reportes	CI, Ro	20h	28/Aug	04/Sep	0%	CI Ro
32	✓ Story Telling	CI, Ro, JA	15h	01/Sep	04/Sep	0%	CI Rr JA
33	✓ Modelo de ML y sus ajustes	Ma, Jo	24h	28/Aug	04/Sep	0%	Ma Jo
34	✓ Preparación Demo general del proyecto	Todos	-	11/Sep	12/Sep	0%	To
35	✓ Entregables	Todos	-	06/Sep	06/Sep	0%	To
36	✓ Story Telling	Todos	-	06/Sep	06/Sep	0%	To
37	✓ Dashboard y reportes	Todos	-	06/Sep	06/Sep	0%	To
38	✓ Insights encontrados	Todos	-	06/Sep	06/Sep	0%	To
39	✓ Relacionar los KPI con datos hallados relevantes	Todos	-	06/Sep	06/Sep	0%	To
40	✓ Implementar modelo de ML	Todos	-	06/Sep	06/Sep	0%	To
41	✓ Implementar un reporte con visualización geodráfica...	Todos	-	06/Sep	06/Sep	0%	To
42	✓ Plus del proyecto	Todos	-	06/Sep	06/Sep	0%	To
43	✓ Documentación	José A	-	28/Aug	04/Sep	0%	JA
44	✓ DEMO	Todos	-	07/Sep	07/Sep	0%	To



SPRINT 4:

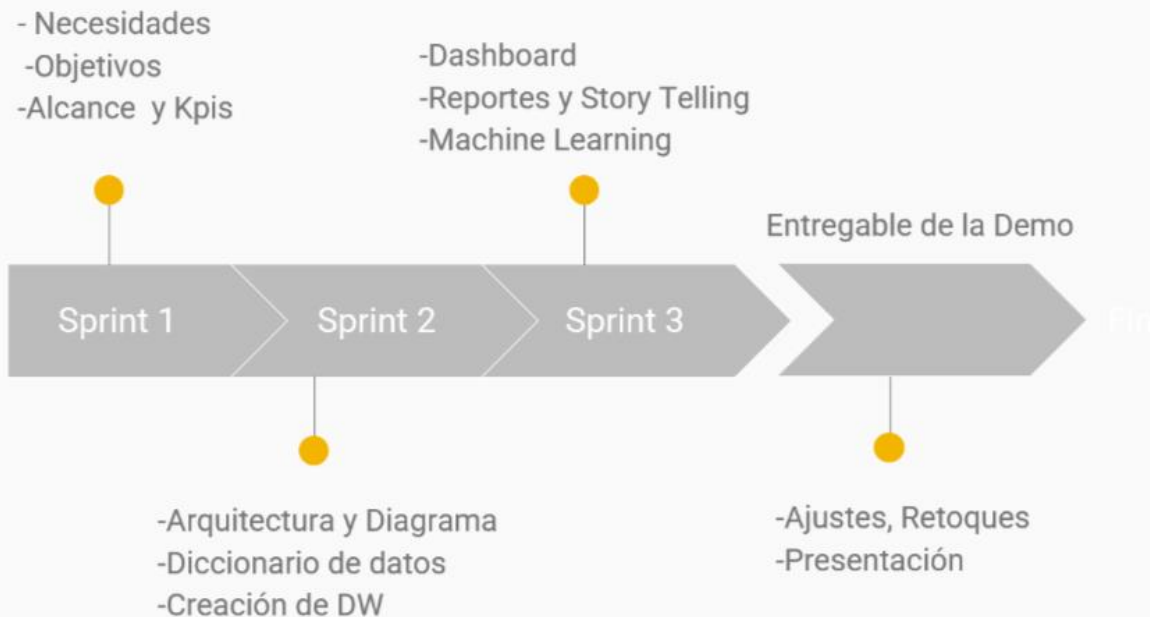


Retoques finales y presentación:			-	11/Sep	21/Sep	0%	
47	✓ Presentación dirigida a la dirección de la compañía	Todos	-	11/Sep	11/Sep	0%	To
48	✓ Retomar lo mostrado en presentaciones anteriores	Todos	-	19/Sep	19/Sep	0%	To
49	✓ Retoques finales con recomendaciones de HM y PO	Todos	-	19/Sep	19/Sep	0%	To
50	✓ DEMO FINAL	Todos	-	21/Sep	21/Sep	0%	To



OBJETIVOS GENERALES

- Desarrollar un sistema de recomendación para usuarios en todo el territorio de los Estados Unidos.
- Analizar los datos y desarrollar un Dashboard interactivo para respaldar decisiones empresariales.

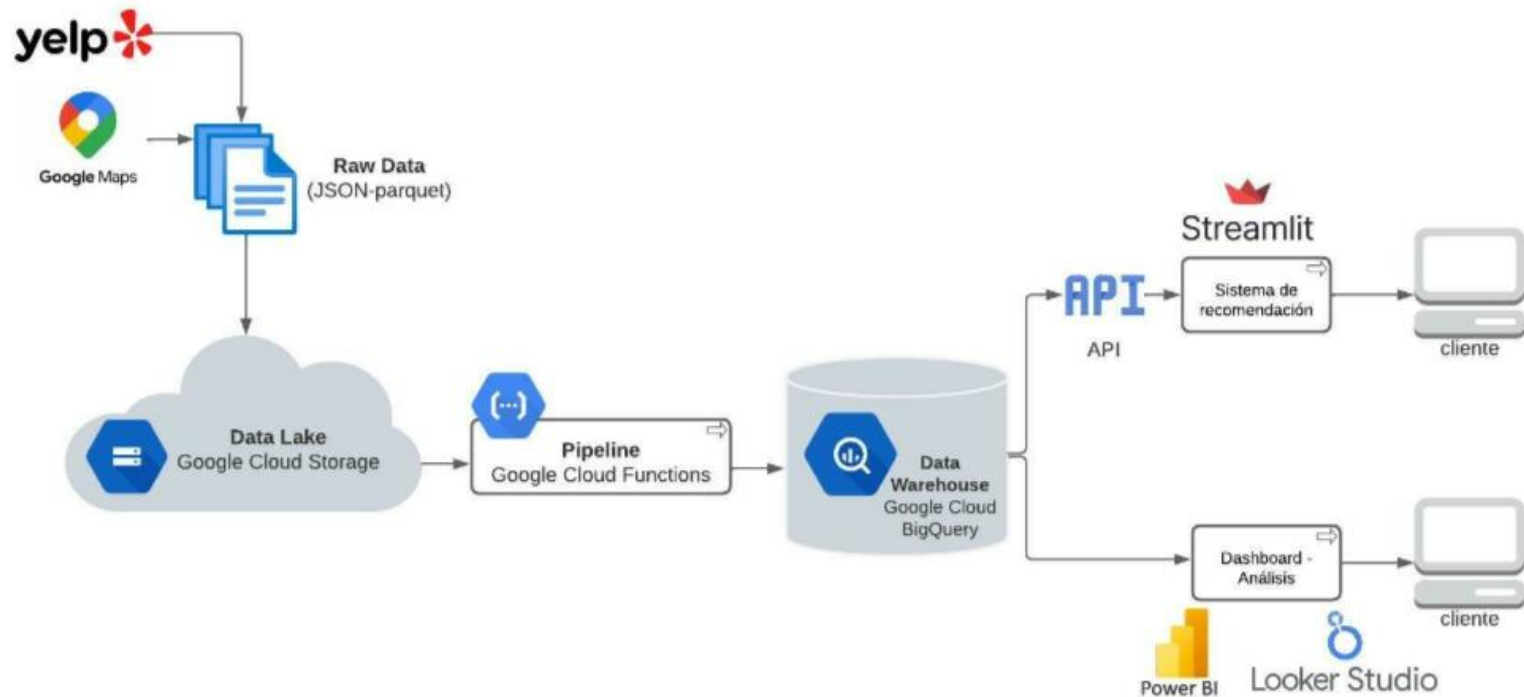


- TECNOLOGÍAS USADAS ➡

DATA WAREHOUSE

Disponibilización de los datos

DATA OPS MIND 



PIPELINE

Procesamiento de los datos

DATA OPS MIND 



Google Cloud Functions -
Automatización del ETL



Pyspark - procesamiento de
grandes volúmenes de datos

PIPELINE

Procesamiento de los datos

DATA OPS MIND 



Geopy - Trabajar con coordenadas



Pandas y Numpy - En equipo con
Spark para procesar los datos

CATEGORIZACIÓN

Establecer rubros (lines)

DATA OPS MIND 

Category

['Senior citizen center']

['Massage therapist']

['Grocery store' 'ATM']

['Gas station']

['Chinese restaurant' 'Restaurant']

['Golf course']



Line

Health and Hospitals

Health and Beauty

Shopping

Automotive

Restaurants

Sports

CATEGORIZACIÓN

Establecer rubros (lines)

DATA OPS MIND 



k-means clustering
Elbow Method



18 Categorías

Category

['Senior citizen center']

['Massage therapist']

['Grocery store' 'ATM']

['Gas station']



OpenAI API
Integration

18 Lines



Lagunas
de palabras



NLP

Análisis de sentimientos

DATA OPS MIND 

Love this stadium for
soccer events. Great
place to see the
tortillas fly!

I went in for a simple
mani-pedi. I walked away
paying \$120 because he
did extra things that I did
not agree to I would not
mean to leave until I paid
for it. Very disappointing



Transformers

POSITIVE -0.99985

NEGATIVE - 0.9997307

Porcentaje de reseñas positivas (PRP)



PRP < 40%	'Very Dissatisfied 😞'
40% <= PRP < 60%	'Dissatisfied 😞'
60% <= PRP < 90%	'Satisfied 😊'
PRP >= 90%	'Very Satisfied 😄'

DATA OPS MIND

- Sistema de recomendación y Dashboards

