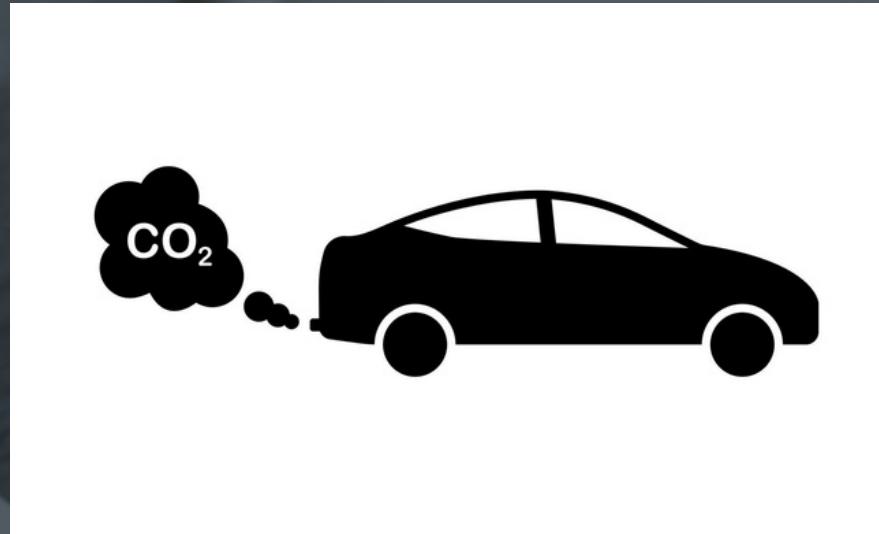


PROUESTA

**PLATAFORMA DE ANÁLISIS INTERACTIVO Y
MODELO DE PREDICCIÓN DE EMISIONES DE
CARBONO EN TAXIS DE NYC: INFORMACIÓN
ACCESIBLE PARA DECISIONES SOSTENIBLES Y
BASADAS EN DATOS**

AIMPEC

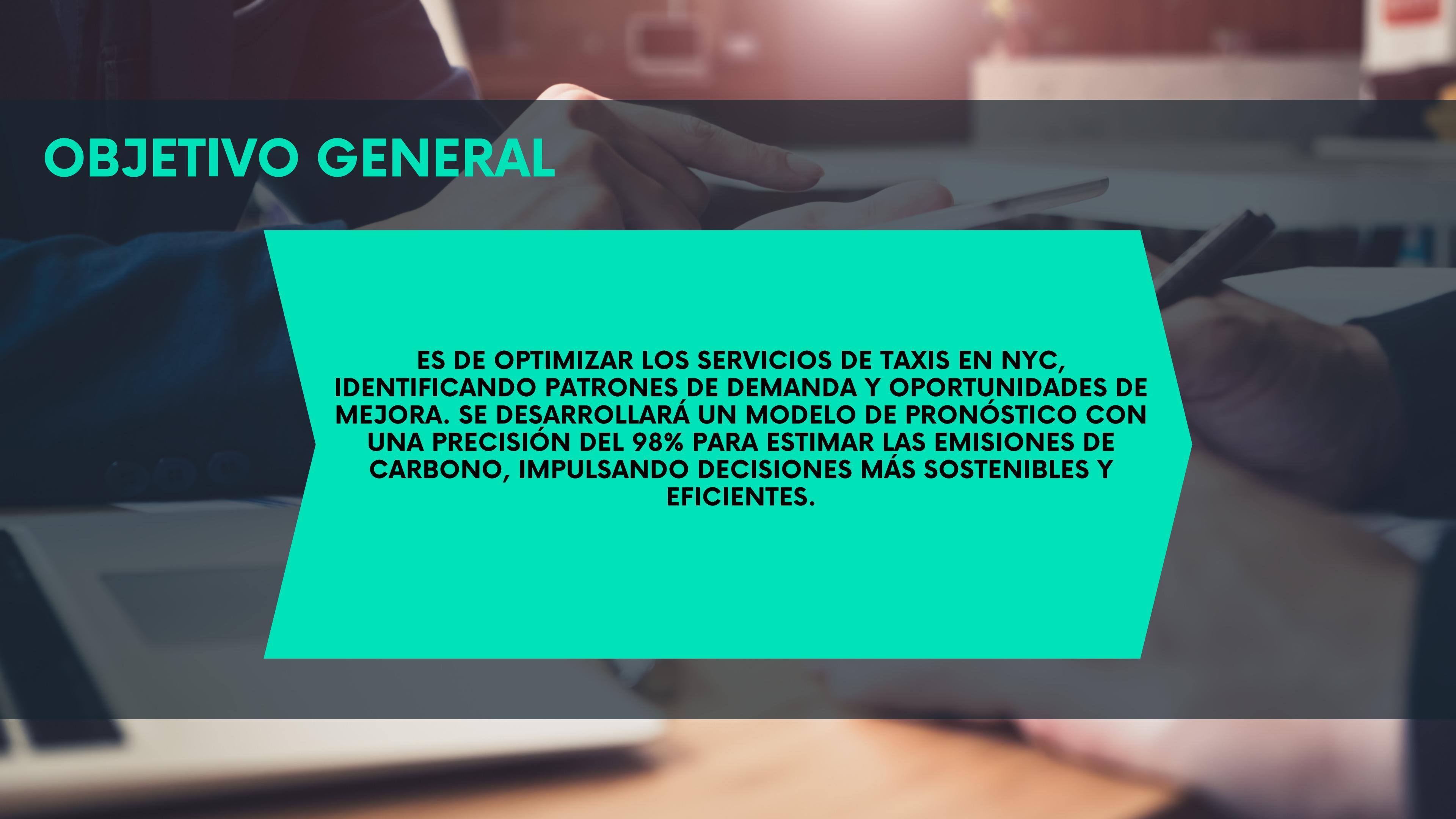
"NYC Taxis & Carbon Emission"



Lienzo Lean Canvas



OBJETIVO GENERAL



ES DE OPTIMIZAR LOS SERVICIOS DE TAXIS EN NYC, IDENTIFICANDO PATRONES DE DEMANDA Y OPORTUNIDADES DE MEJORA. SE DESARROLLARÁ UN MODELO DE PRONÓSTICO CON UNA PRECISIÓN DEL 98% PARA ESTIMAR LAS EMISIONES DE CARBONO, IMPULSANDO DECISIONES MÁS SOSTENIBLES Y EFICIENTES.

OBJETIVO ESPECÍFICO

SOE-01

SOE-02

SOE-03

INDICADORES DE EXITO

KPI 1

REDUCCION DE CO₂

Proyeccion a reducir -15% en 3 meses.

KPI 2

EFICIENCIA DE LOS CONDUCTORES

Incrementos en +5% en 3 meses.

KPI 3

AUMENTO DE INGRESOS

Incrementos de capital en +10% en 3 meses.

KPI 4

AUMENTO DE VIAJES COMPARTIDOS

Incrementos en +10% en 3 meses.

	Indicador de éxito	Objetivo
I-01	<p>Reducción de la emisión de contaminantes.</p> <p>Objetivo: Reducir la emisión de contaminantes en un 15% en 3 meses.</p> <p>Formula</p> $\%PR = (Pi - Pf) / (Pi * 100)$ <p>Donde:</p> <p><i>PR: Reducción de la emisión de contaminantes.</i></p> <p><i>Pi: Polución al inicio (polluting_agent al inicio del período).</i></p> <p><i>Pf: Polución al final (polluting_agent al final del período).</i></p>	OE-01

Mejora de la eficiencia de los conductores

Objetivo: Aumentar las horas promedio de trabajo por conductor en un 5% en 3 meses.

Formula

$$\%ED = ((Df - Di) / Di) * 100$$

I-03

Donde:

OE-03

ED: Eficiencia del conductor.

Di: Horas promedio de trabajo por conductor al inicio (avg_hours_per_day_per_driver al inicio del periodo).

Df: Horas promedio de trabajo por conductor al final (avg_hours_per_day_per_driver al final del periodo).

Aumento de los ingresos

Objetivo: Aumentar el farebox_per_day en un 10% en 3 meses.

Formula.

$$\%RI = ((If - Ii) / Ii) * 100$$

OE-03

Donde:

RI: Aumento de los ingresos.

Ii: Ingresos al inicio (farebox_per_day al inicio del periodo).

If: Ingresos al final (farebox_per_day al final del periodo).

I-02

Aumento de viajes compartidos

Objetivo: Aumentar los viajes compartidos en un 10% en 3 meses.

Formula

$$\%AT = ((Tf - Ti) / Ti) * 100$$

OE-02

Donde:

AT: Aumento de viajes compartidos.

Ti: Porcentaje de viajes compartidos al inicio (trips_per_day_shared al inicio del periodo).

Tf: Porcentaje de viajes compartidos al final (trips_per_day_shared al final del periodo).

INSPECCIÓN ÁGIL

THE PRODUCT VISION BOARD



Visión

La visión del producto es brindar una solución integral que optimice los servicios de taxis en NYC, utilizando análisis de datos y técnicas de aprendizaje automático, para impulsar decisiones sostenibles y mejorar la eficiencia del negocio.



Usuarios

.La empresa de servicios de transporte de pasajeros.
.Consultoras y empresas del sector de transporte y movilidad interesadas en acceder a datos y análisis relacionados con el funcionamiento de los servicios de taxis en NYC.
.Organismos públicos y reguladores que requieren información detallada sobre las emisiones de carbono generadas por los servicios de taxis.
.Estudiantes e investigadores en el campo de ciencias de datos y transporte interesados en estudiar el funcionamiento de los servicios de taxis en NYC.



Necesidades

.Necesidad de optimizar los servicios de taxis.
.Necesidad de tomar decisiones basadas en datos.
.Necesidad de cumplir con regulaciones y políticas medioambientales.
.Necesidad de acceder a datos confiables y procesados



Producto

.Plataforma integral que incluye un dashboard interactivo y un modelo de predicción.
.Esta plataforma permitirá a los usuarios acceder y visualizar de manera intuitiva los indicadores clave relacionados con los servicios de taxis en NYC, así como obtener pronósticos confiables sobre las emisiones de carbono generadas por la flota de vehículos.



Beneficios el negocio

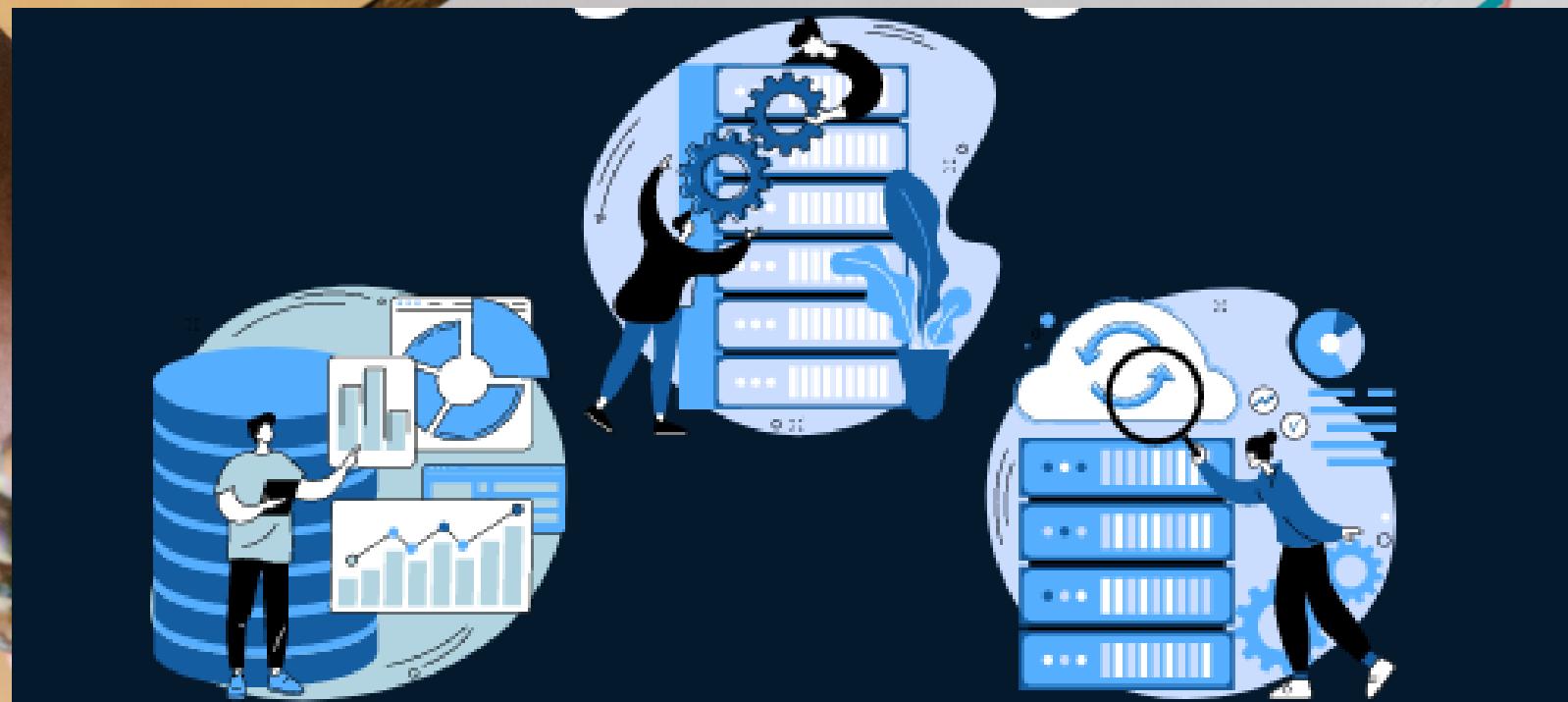
.Optimización de los servicios de taxis.
.Toma de decisiones basadas en datos: Al tener acceso a datos precisos y análisis detallados, el negocio podrá tomar decisiones estratégicas más informadas.
.Cumplimiento de regulaciones medioambientales: Al contar con un modelo de predicción preciso para estimar las emisiones de carbono generadas por la flota de taxis, el negocio podrá cumplir con las regulaciones medioambientales vigentes y reducir las sanciones económicas asociadas.

DEFINICION DE ROLES

DATA ANALYST



DATA ENGINEER



DATA SCIENCE



PROJECT MANAGER



HISTORIAS DE USUARIO

HISTORIAS DE USUARIO				
Identificador de la Historia	Nombre de la Historia	Como (rol)	Necesito (característica)	Para (valor de negocio)
HU001	Limpieza de datos	Data Engineer	Deseo limpiar y preprocesar los datos para garantizar la calidad y fiabilidad de los mismos.	Mejora la calidad de los datos y la confiabilidad de los resultados para la ingestión en mi Base de datos.
HU002	Integración de datos	Data Engineer	Deseo integrar datos de diferentes fuentes para tener una visión completa del problema.	Proporciona una visión integral y mejora la precisión de los análisis y mi relación final en mi Base de Datos.
HU003	Análisis exploratorio de datos	Data Analyst	Deseo analizar los datos para identificar tendencias y patrones.	Revela insights valiosos y guía la toma de decisiones.
HU004	Modelo de predicción	Data Scientist	Deseo desarrollar un modelo de predicción para estimar las emisiones de carbono.	Ayuda a la empresa a planificar estrategias sostenibles y cumplir con las regulaciones medioambientales.
HU005	Desarrollo de Dashboard	Data Analyst	Deseo desarrollar un dashboard interactivo para visualizar los indicadores.	Facilita la interpretación de los datos y la comunicación de los resultados.
HU006	Validación de Dashboard	Data Analyst	Deseo validar y ajustar el dashboard para asegurarme de que cumple con las necesidades del usuario.	Asegura que la solución es útil y relevante para el cliente.
HU007	Despliegue en la nube	Arquitecto Cloud	Necesito desplegar la solución en la nube para que esté accesible.	Permite el acceso remoto y garantiza la escalabilidad de la solución.

HISTORIAS DE USUARIO POR ÉPICA

HISTORIAS DE USUARIO										
Nº	Épica	Identificador de la Historia	Nombre de la Historia	Descripción de la Historia de Usuario	Estado	Dimensión /Esfuerzo	Iteración (Sprint)	Prioridad	Criterio de Aceptación	Comentarios
Como (rol)										
1	ETL	HU001	Limpieza de datos	Data Engineer	Planificada	88	Sprint 1	89	CA1	Leandro J.
2		HU002	Integración de datos	Data Engineer	Planificada	70	Sprint 2	61.5	CA2	Pedro O.
4	EDA	HU003	Análisis exploratorio de datos	Data Analyst	Planificada	42	Sprint 2	44.5	CA3	Juliana C.
5		HU004	Modelo de predicción	Data Scientist	Planificada	67	Sprint 3	40	CA4	Ilbert A.
6	Dashboard	HU005	Desarrollo de Dashboard	Data Analyst	Planificada	51	Sprint 3	38	CA5	Juliana C.
7		HU006	Validación de Dashboard	Data Analyst	Planificada	50	Sprint 3	34	CA6	Juliana C.
9	Cloud	HU007	Despliegue en la nube	Arquitecto Cloud	Planificada	65	Sprint 3	33	CA7	David H.

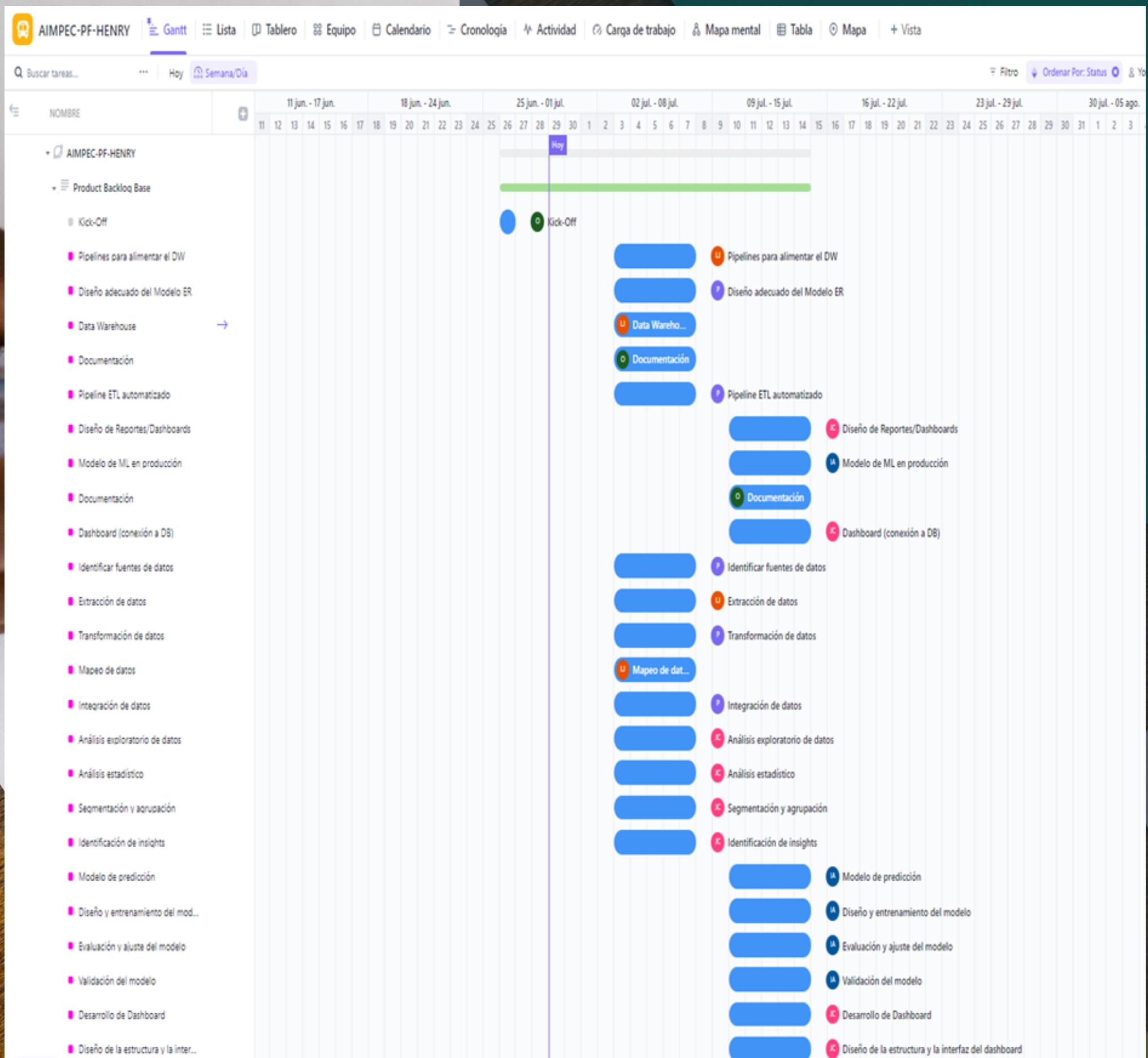
TABLERO SCRUM BASE

AIMPEC-PF-HENRY Gantt Lista Tablero Equipo Calendario Cronología Actividad Carga de trabajo

Buscar tareas... Filtro Agrupar por: N

#	NOMBRE DE LA TAREA	PERSONA ASIGNADA	ESTADO	FECHA LÍMITE	PRIORIDAD
1	Crear Lienzo Lean Canvas	OscarHospital	DONE	Ayer	■
2	Completar los Issues y Uniq...	Leandro Gastón Jofré	OPEN		■
3	Completar Special Advanta...	Juliana Caceres	OPEN		■
4	Completar KPIs	Ilbert Alarcon	OPEN		■
5	Completar Solution,Cost St...	OscarHospital	OPEN		■
6	Completar Lienzo Sources ...	pedrooria23@gmail.com	OPEN		■
7	Kick-Off	OscarHospital	OPEN	Hace 3 días	■
8	Organización y División de tar...	OscarHospital	DONE	Hace 2 días	■
9	Entendimiento de la situación ...	Juliana Caceres	DOING	Mañana	■
10	Objetivos	Ilbert Alarcon	DOING	Mañana	■
11	Alcance	Ilbert Alarcon	DOING	Hoy	■
12	Objetivos y KPIs asociados (pla...	Ilbert Alarcon	DONE	Hoy	■
13	Creación del Repositorio Github	Ilbert Alarcon	DONE	Ayer	■
14	Solución propuesta	OscarHospital	DONE	Ayer	■
15	Incluir stack tecnológico	Leandro Gastón Jofré	DONE	Hoy	■
16	Creación del Equipo de trabaj...	Juliana Caceres	DONE	Ayer	■
17	Creación del Cronograma gen...	OscarHospital	DOING	Mañana	■
18	Análisis preliminar de calidad ...	pedrooria23@gmail.com	DOING	Mañana	■
19	Realizar un análisis inicial de lo...	Leandro Gastón Jofré	DOING	Mañana	■
20	Eliminación de datos irrelevant...	Leandro Gastón Jofré	DOING	Mañana	■
21	Tratamiento de valores faltantes	Leandro Gastón Jofré	DOING	Mañana	■
22	Limpieza de errores y valores a...	pedrooria23@gmail.com	DOING	Mañana	■
23	Verificación de integridad y co...	pedrooria23@gmail.com	DOING	Mañana	■
24	Pipelines para alimentar el DW	Leandro Gastón Jofré	TO DO	7/7/23	■
25	Diseño adecuado del Modelo ...	pedrooria23@gmail.com	TO DO	7/7/23	■
26	Data Warehouse	Leandro Gastón Jofré	TO DO	7/7/23	■

DIAGRAMA DE GANNT



CRONOLOGÍA DEL PROYECTO

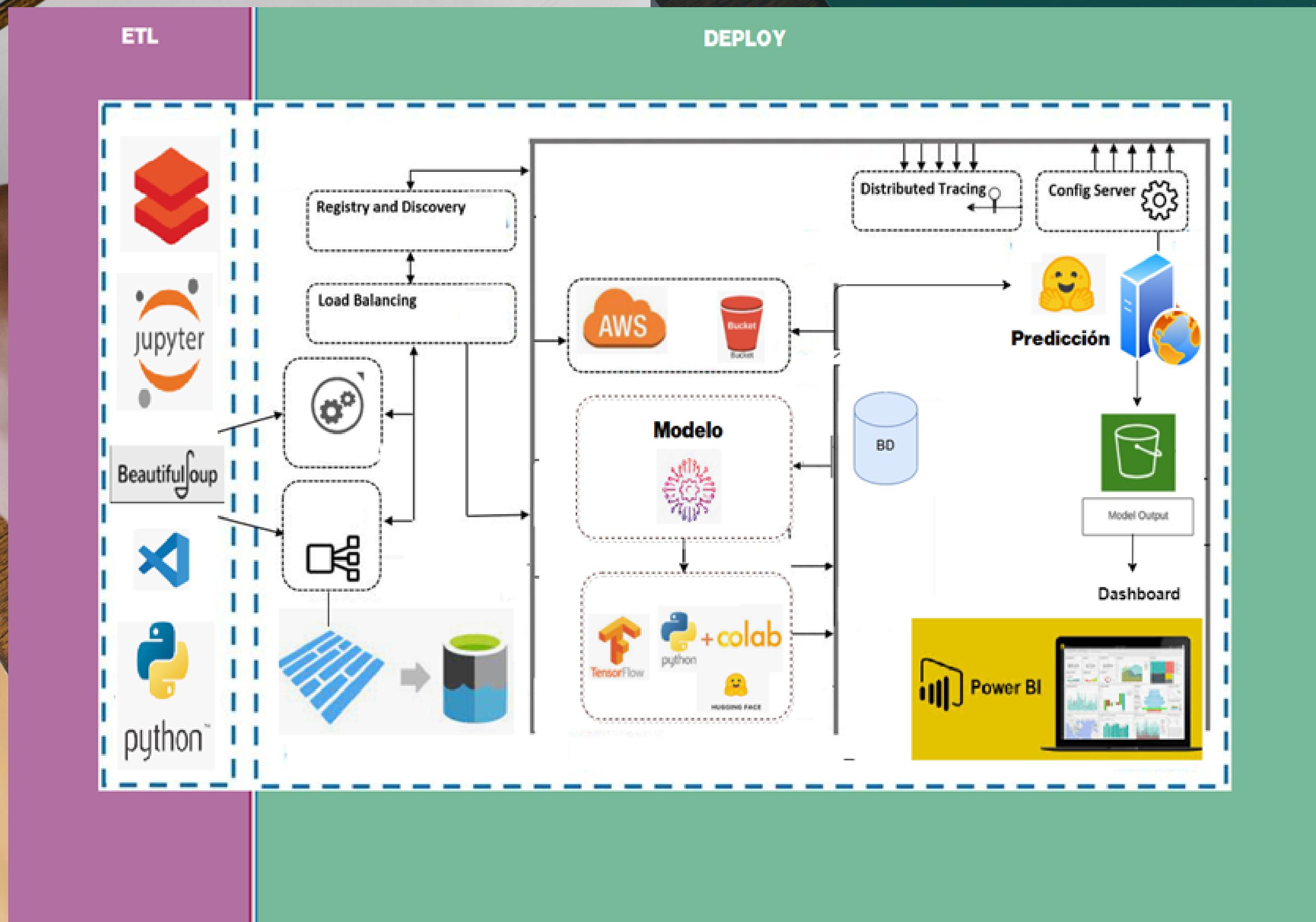
AIMPEC-PF-HENRY | Gantt | Lista | Tablero | Equipo | Calendario | Cronología | Actividad | Carga de trabajo | Mapa mental | Tabla | Mapa | + Vista

Buscar tareas... Hoy - + Día | Filtro | Agrupar por Nivel

JUN. 2023 JUL. 2023

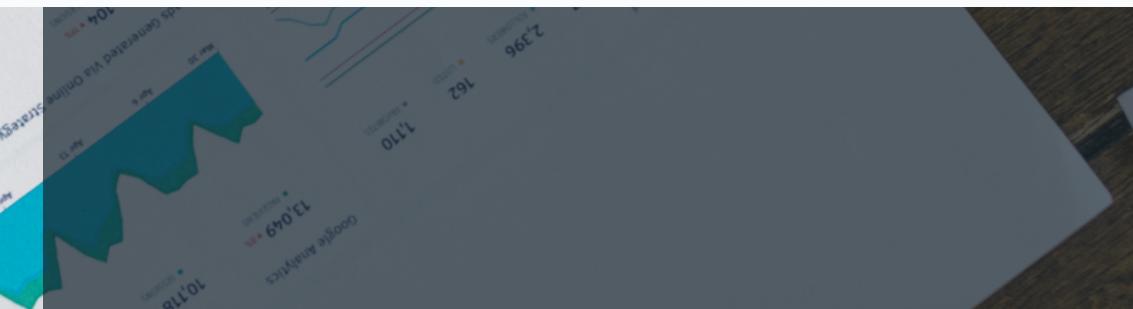
Actividad	Fecha
Crear Lienzo Lean Canvas	2023-06-26
Kick-Off	2023-06-26
Análisis preliminar de calidad de datos	2023-06-26
Creación del Cronograma general - Gantt	2023-06-26
Eliminación de datos irrelevantes	2023-06-26
Entendimiento de la situación propuesta	2023-06-26
Limpieza de errores y valores atípicos	2023-06-26
Objetivos	2023-06-26
Realizar un análisis inicial de los datos para...	2023-06-26
Tratamiento de valores faltantes	2023-06-26
Verificación de integridad y coherencia	2023-06-26
Alcance	2023-06-26
Incluir stack tecnológico	2023-06-26
Objetivos v KPIs asociados (ol...	2023-06-26
Creación del Equipo	2023-06-26
Creación del Reocr...	2023-06-26
Solución propuesta	2023-06-26
Organizaci...	2023-06-26
Análisis estadístico	2023-07-03
Analisis exploratorio de datos	2023-07-04
Data Warehouse	2023-07-05
Diseño adecuado del Modelo ER	2023-07-06
Documentación	2023-07-07
Extracción de datos	2023-07-08
Identificación de insights	2023-07-09
Identificar fuentes de datos	2023-07-10
Integración de datos	2023-07-11
Mapeo de datos	2023-07-12
Pipeline ETL automatizado	2023-07-13
Pipelines para alimentar el DW	2023-07-14
Segmentación v acrioación	2023-07-15
Transformación de datos	2023-07-16
Aprobación final	2023-07-17
Configuración de interacciones y filtros	2023-07-18
Configuración de la infraestructura en la nube	2023-07-19
Configuración de seguridad	2023-07-20
Dashboard (conexión a DB)	2023-07-21
Desarrollo de Dashboard	2023-07-22
Despliegue en la nube	2023-07-23
Diseño de la estructura y la interfaz del dashboard	2023-07-24
Diseño de Reportes/Dashboards	2023-07-25
Diseño y entrenamiento del modelo	2023-07-26
Documentación	2023-07-27
Evaluación y ajuste del modelo	2023-07-28
Modelo de ML en producción	2023-07-29
Modelo de predicción	2023-07-30
Monitoreo v gestión	2023-07-31
Prueba de funcionalidad	2023-08-01
Retroalimentación del usuario	2023-08-02
Validación de Dashboard	2023-08-03
Validación del modelo	2023-08-04
Verificación de datos	2023-08-05

ARQUITECTURA TÉCNICA BASE TO-BE

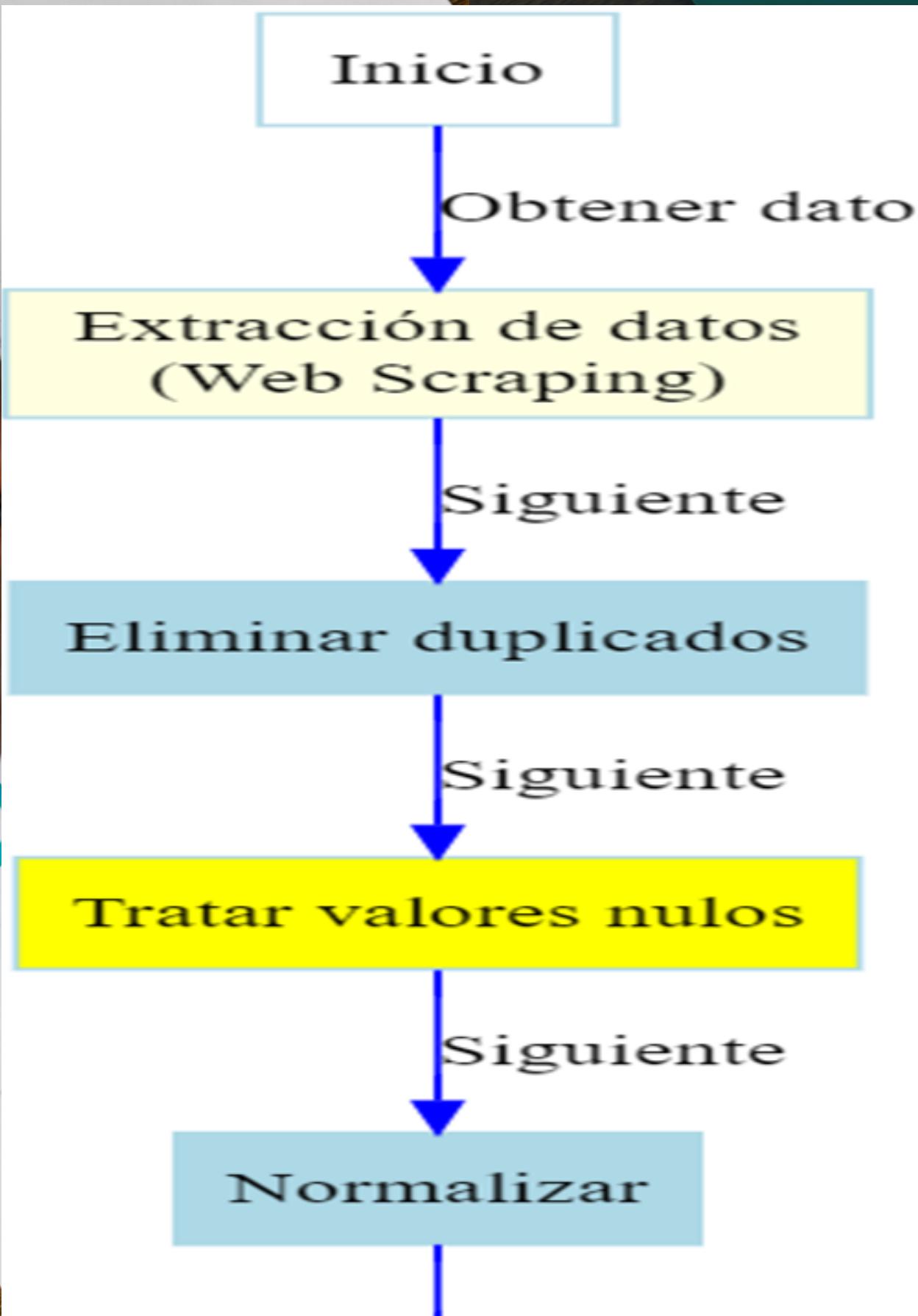


HISTORIAS DE USUARIO PARA EL SPRINT 1

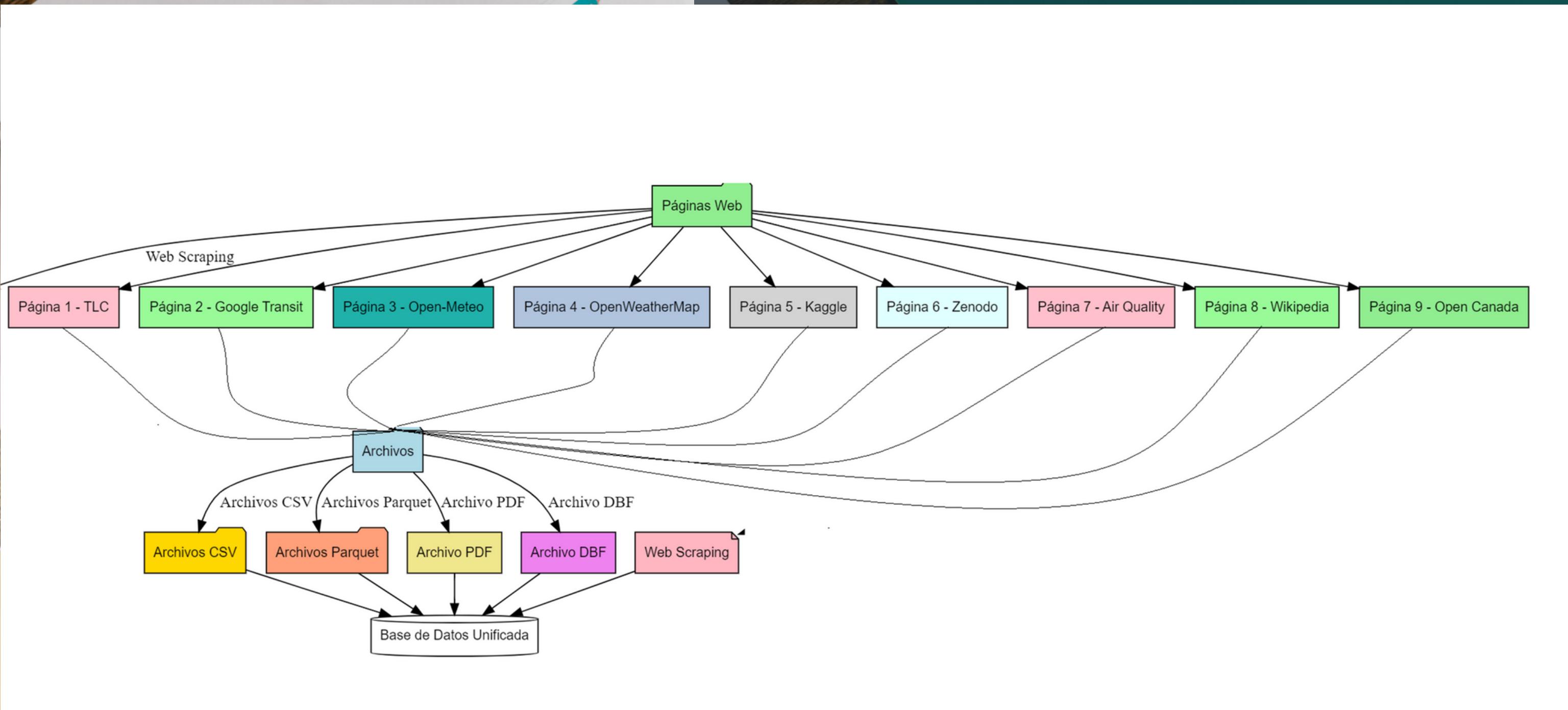
DOING	10 TAREAS	PERSONA ASIGNADA	FECHA LÍMITE	PRIORIDAD
	Análisis preliminar de calidad de datos entregable 01	P	Mañana	!
	Realizar un análisis inicial de los datos para comprender su estructura, calidad y características epica01 hu001 sprint01	U	Mañana	!
	Eliminación de datos irrelevantes epica01 hu001 sprint01	U	Mañana	!
	Tratamiento de valores faltantes epica01 hu001 sprint01	U	Mañana	!
	Limpieza de errores y valores atípicos epica01 hu001 sprint01	P	Mañana	!
	Verificación de integridad y coherencia epica01 hu001 sprint01	P	Mañana	!
	+ Nueva tarea			



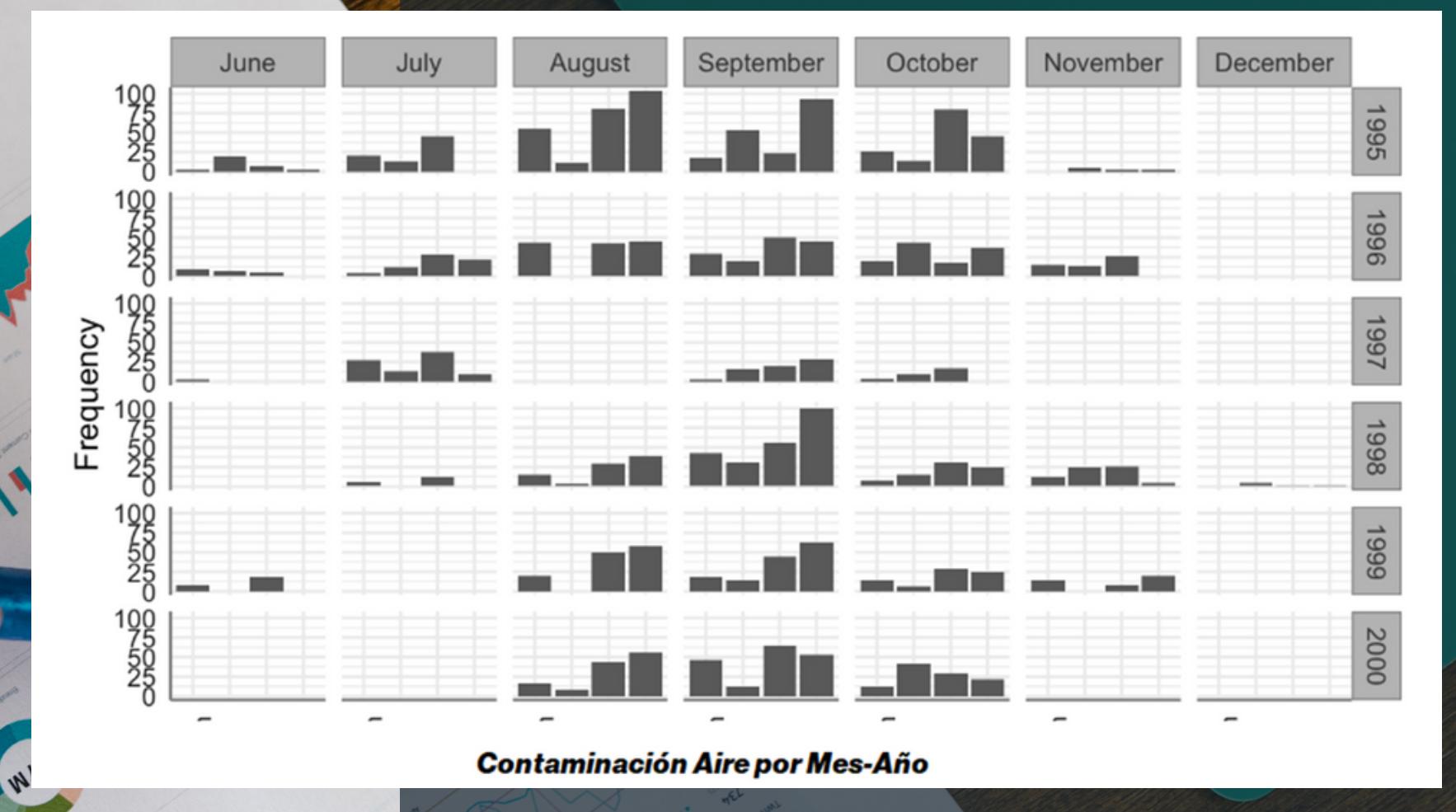
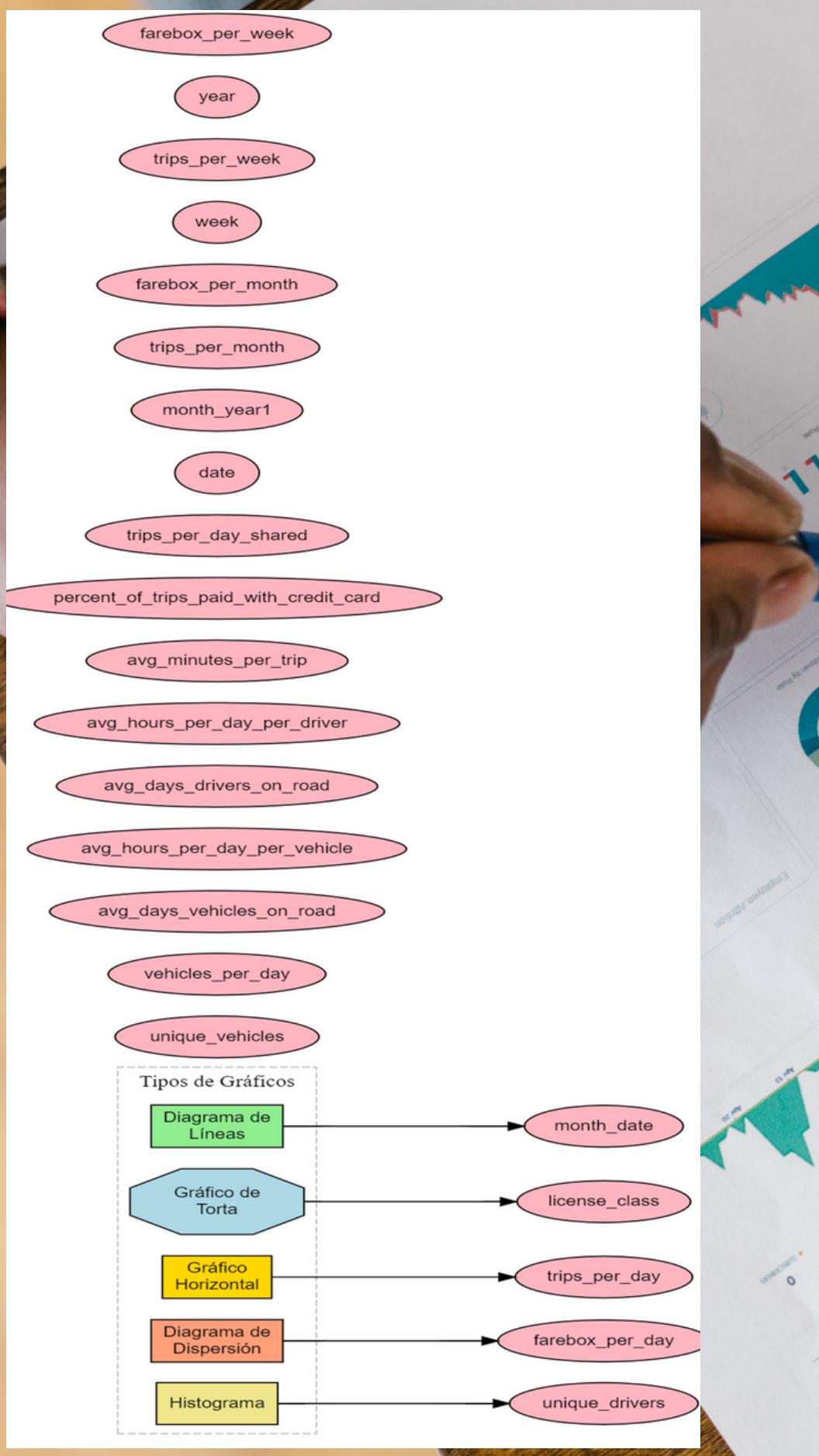
CONSTRUCCIÓN DEL SPRINT 1-MOCKUP HU001



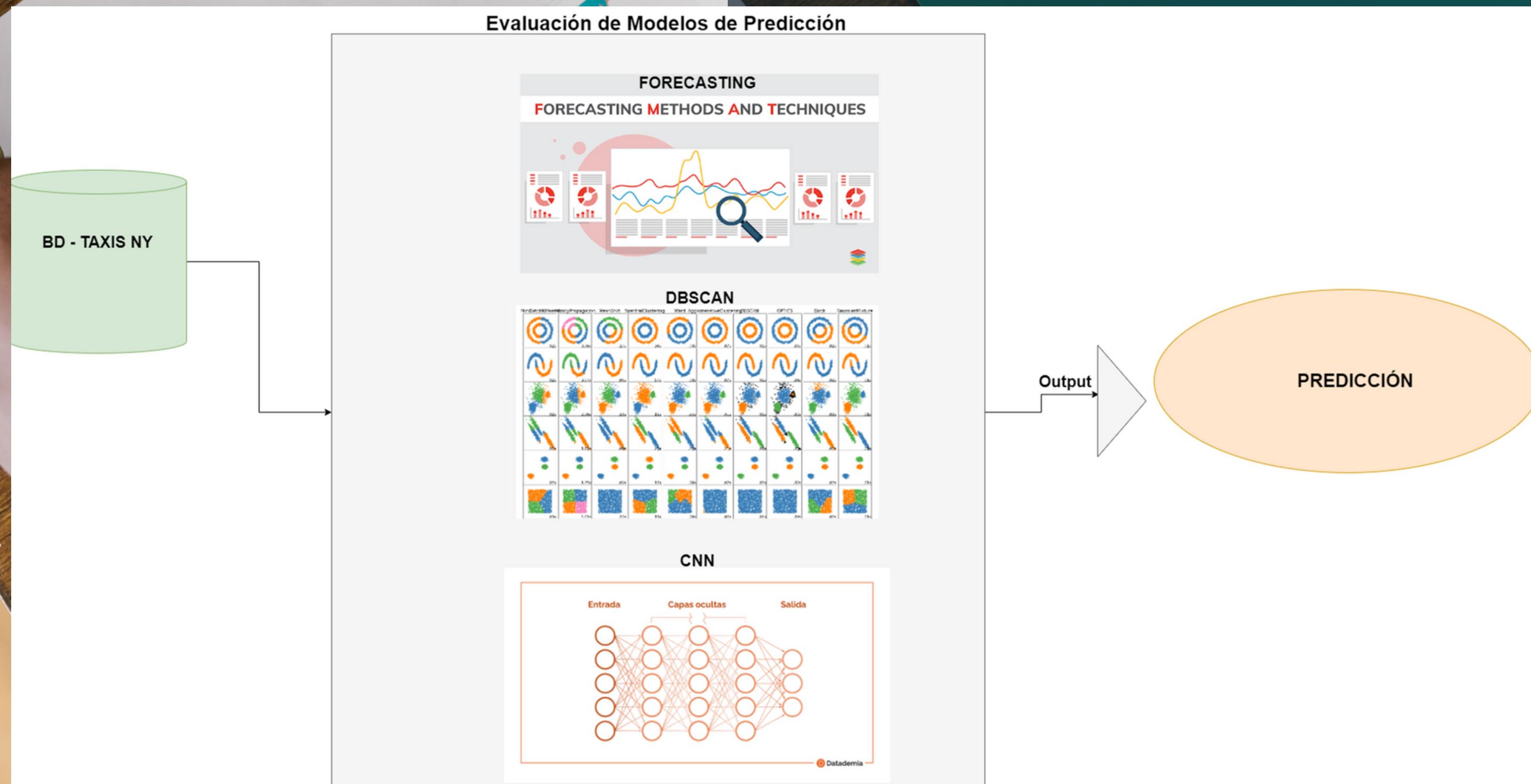
CONSTRUCCIÓN DEL SPRINT 2-MOCKUP HU002



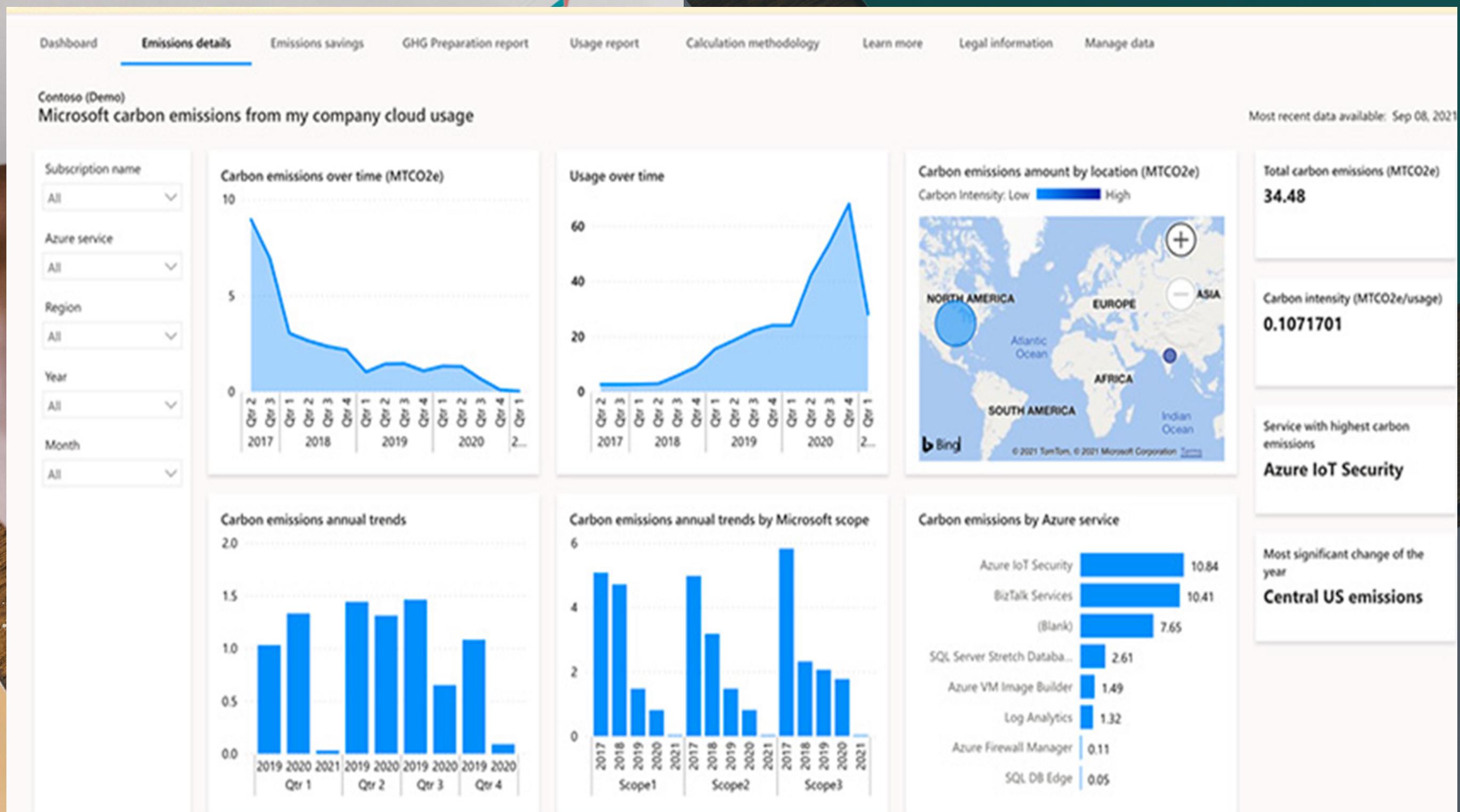
CONSTRUCCIÓN DEL SPRINT 2-MOCKUP HU003



CONSTRUCCIÓN DEL SPRINT 3-MOCKUP HU004



CONSTRUCCIÓN DEL SPRINT 3-MOCKUP HU005



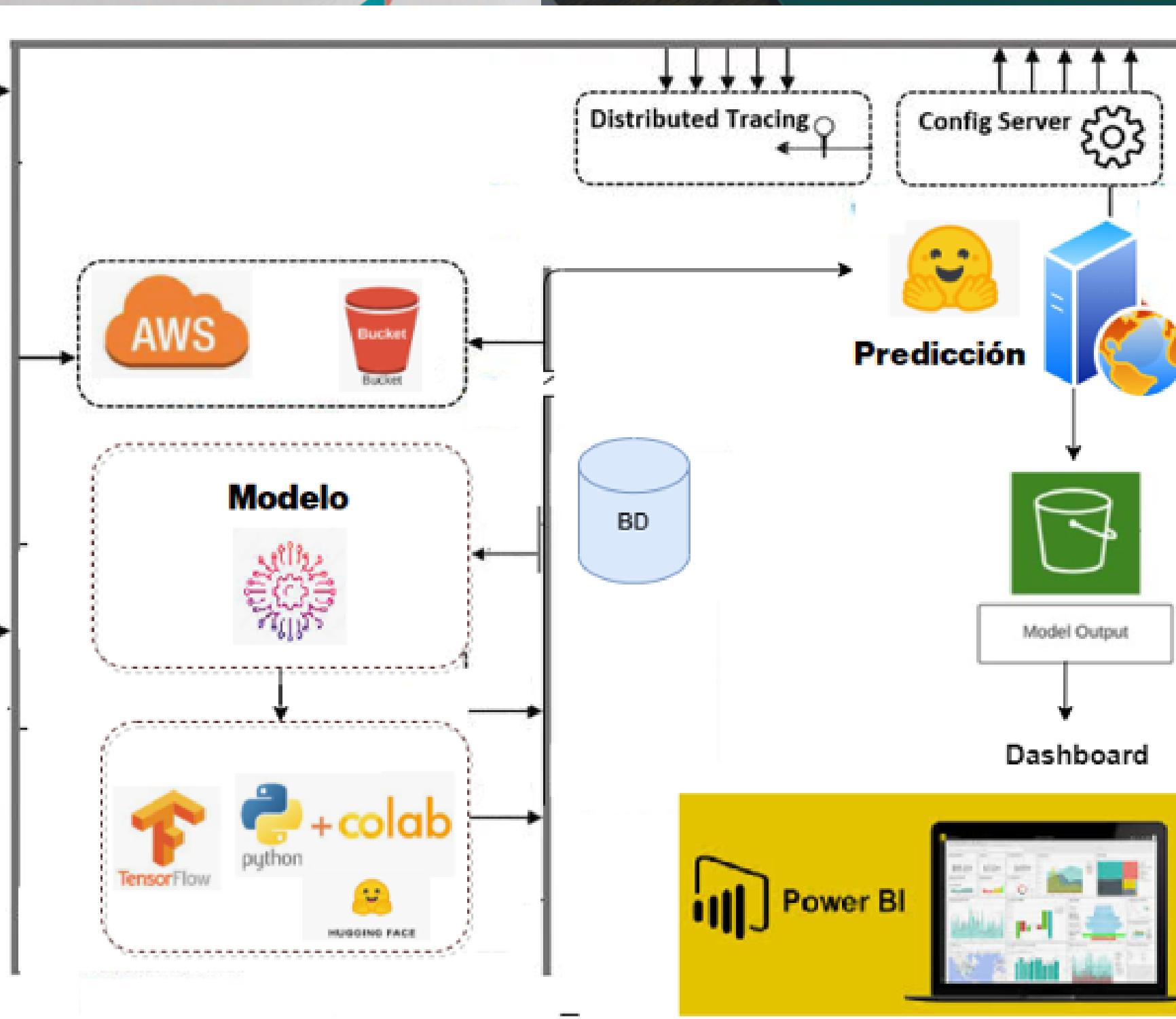
CONSTRUCCIÓN DEL SPRINT 3-MOCKUP HU006



CONSTRUCCIÓN DEL SPRINT 3-MOCKUP HU006



CONSTRUCCIÓN DEL SPRINT 3-MOCKUP HU007



PRODUCT BACKLOG REFINADO

AIMPEC-PF-HENRY | Gantt | Lista | Tablero | Equipo | Calendario | Cronología | Actividad | Carga de trabajo | +

Buscar tareas... | ... | Filtro | Ordenar por | Agrupar por: Estado | Subtareas

OPEN 1 | TO DO 34 | DOING 10 | DONE 7

Product Backlog Base
Kick-Off
Hace 3 días - Hace 3 días 4h ⭐ 2
+ AÑADIR SUBTAREA

Product Backlog Base
Pipelines para alimentar el DW
lun. - jul. 7 4h ⭐ 3
entregable02
+ AÑADIR SUBTAREA

Product Backlog Base
Diseño adecuado del Modelo ER
lun. - jul. 7 4h ⭐ 3
entregable02
+ AÑADIR SUBTAREA

Product Backlog Base
Data Warehouse
lun. - jul. 7 4h ⭐ 3
entregable02
+ AÑADIR SUBTAREA

Product Backlog Base
Documentación
lun. - jul. 7 6h ⭐ 3
entregable02
+ AÑADIR SUBTAREA

Product Backlog Base
Realizar un análisis inicial de los datos para comprender su estructura, calidad y características
Hace 2 días - Mañana 15h ⭐ 2
epica01 hu001 sprint01
+ AÑADIR SUBTAREA

Product Backlog Base
Eliminación de datos irrelevantes
Hace 2 días - Mañana 4h ⭐ 3
epica01 hu001 sprint01
+ AÑADIR SUBTAREA

Product Backlog Base
Tratamiento de valores faltantes
Hace 2 días - Mañana 4h ⭐ 2
epica01 hu001 sprint01
+ AÑADIR SUBTAREA

Product Backlog Base
Limpieza de errores y valores atípicos
Hace 2 días - Mañana 4h ⭐ 2

Roles y responsabilidades
Hace 2 días - Ayer 1h ⭐ 5
entregable 01
+ AÑADIR SUBTAREA

Product Backlog Base
Incluir stack tecnológico
Hace 2 días - Hoy 1h ⭐ 5
entregable 01
+ AÑADIR SUBTAREA

Product Backlog Base
Organización y División de tareas
Hace 2 días - Hace 2 días 2h ⭐ 3
entregable 01
+ AÑADIR SUBTAREA

Product Backlog Base
Creación del Repositorio Github
Hace 2 días - Ayer 2h ⭐ 1
entregable 01
+ AÑADIR SUBTAREA

4/5

DATA ENGINEERING APLICADA

01

Objetivos.

Generar la vision de la data en base a lo que solicito el cliente.

02

Extraccion.

Utilizar herramientas de extracción de datos como web scraping para recopilar información de sitios web.

03

Limpieza.

Estandarizar formatos de datos para garantizar la coherencia y facilitar la comparación.

04

Inspeccion Final.

Optimizar campos y detalles finales.

05

Muestreo.

Seleccionar conjuntos de informacion para su evaluacion con las predicciones pertinentes.

06

Observaciones.

Detalles encontrados en los datasets.

1 OBJETIVOS

DISEÑAR Y MANTENER UNA
ARQUITECTURA ESCALABLE Y SEGURA.

2 EXTRACCION

INFORMACION DE CALIDAD PARA PODER PRODUCIR
MEJORES METRICAS.

3 LIMPIEZA

NORMALIZAR Y DEPURAR REDUNDANCIAS.

4 INSPECCION FINAL

ESTABLECER UNA PRIMER VERSION FUNCIONAL.

5 MUESTREO

PEQUEÑAS MUESTRAS PARA GRANDES
RESPUESTAS.

6 OBSERVACIONES

DETALLES, POR MENORES Y DATOS
FALTANTES.



DE NOSOTROS
GRACIAS