

**Índice**

[**Introducción**](#_heading=h.klegz8queykc) **2**

[**Escalabilidad**](#_heading=h.7lnnlh7720bl) **2**

[**Flujo de trabajo on Cloud**](#_heading=h.mhckarqvg31d) **2**

[**Dashboard**](#_heading=h.e1c4rugol8rh) **3**

[**Diseño**](#_heading=h.1tex9goikmuy) **3**

[**Análisis Descriptivo**](#_heading=h.ulc2qonsv7gn) **4**

[**Machine Learning**](#_heading=h.cfiv8eqwfoh3) **5**

[**Análisis de la serie temporal**](#_heading=h.orbtz8yk9sam) **6**

[**Componente Estacional**](#_heading=h.rgktbu9i3p4) **7**

[**Componente de Tendencia**](#_heading=h.xptpmnf8kgqf) **8**

[**Forecasting**](#_heading=h.nh2d9fjokpex) **9**

[**Conclusiones**](#_heading=h.6nn72b89hcq1) **10**

[**Diagrama de Gantt**](#_heading=h.oihkprm1c23) **10**

[**Producto Final**](#_heading=h.swh9o8wov3dw) **11**

# Introducción

La presente documentación tiene como objetivo detallar las tareas y actividades realizadas durante la tercera semana del proyecto de análisis para el ecommerce Olist.

Esta semana, el enfoque se centró en:

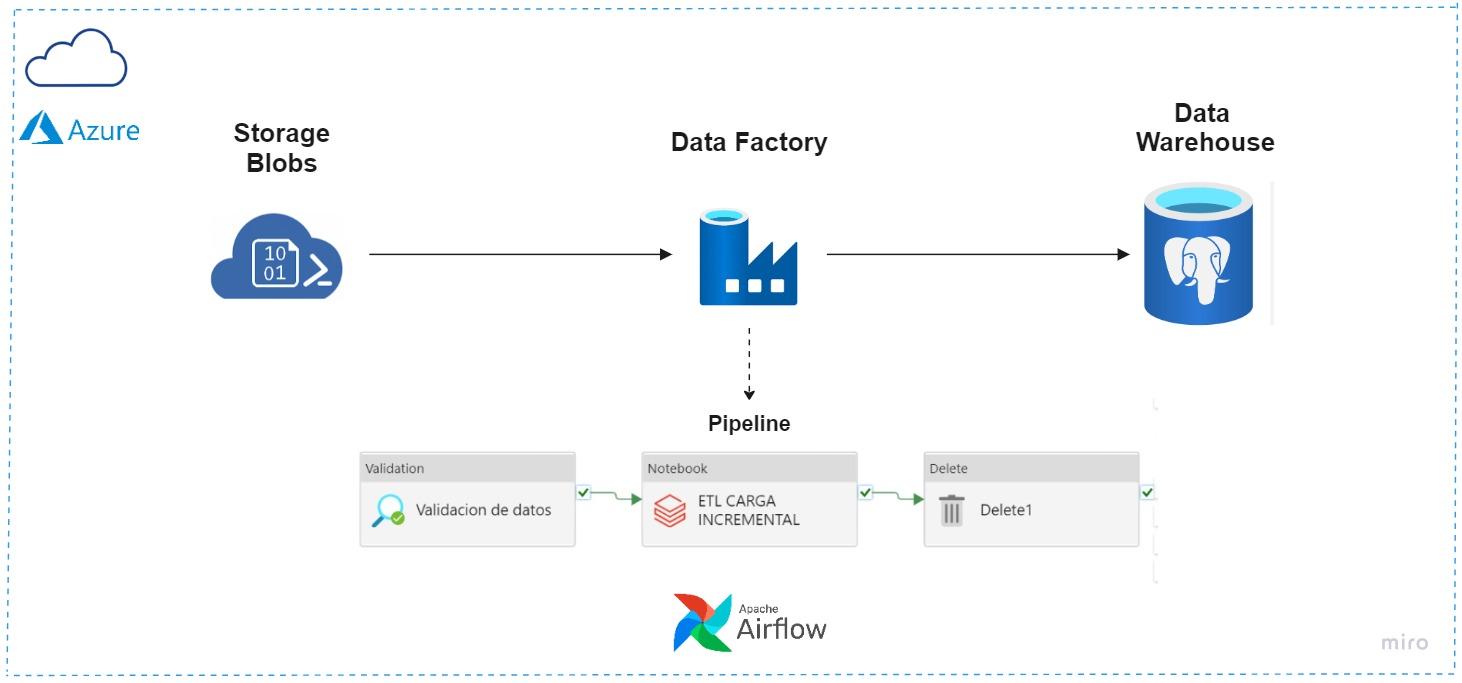
* Probar la escalabilidad del proyecto on cloud luego de la exitosa carga inicial
* Data Analytics, realizando un análisis minucioso de los datos, definiendo las plantillas del Dashboard y ultimando los detalles de la narrativa con la que se presentará el proyecto final.
* Machine Learning, con la implementación de modelos de aprendizaje para mejorar la precisión y el rendimiento del análisis.

# Escalabilidad

## Flujo de trabajo on Cloud

En la semana 2 del proyecto, se realizó la carga inicial de datos a través de un pipeline de ETL en un entorno local (on premise) para probar y validar el proceso.

Sin embargo, al considerar la escalabilidad del proyecto, se decidió migrar la carga de datos al cloud (Azure) para aprovechar las ventajas de almacenamiento, procesamiento y disponibilidad que ofrece esta opción. Esto permitirá una mayor flexibilidad y escalabilidad en el futuro, para manejar una mayor cantidad de datos.



# Dashboard

## Diseño

Como punto de partida para la realización del Dashboard, utilizamos plantillas que realizamos con Adobe Illustrator, utilizando el logo de Olist y los colores institucionales.

Partimos de dos modelos, uno para la presentación del Dashboard y otro que utilizamos en las distintas páginas, donde analizamos los datos en función de los KPIs planteados.

Dichos modelos son los siguientes:

Esta plantilla la utilizamos para la presentación del Proyecto.



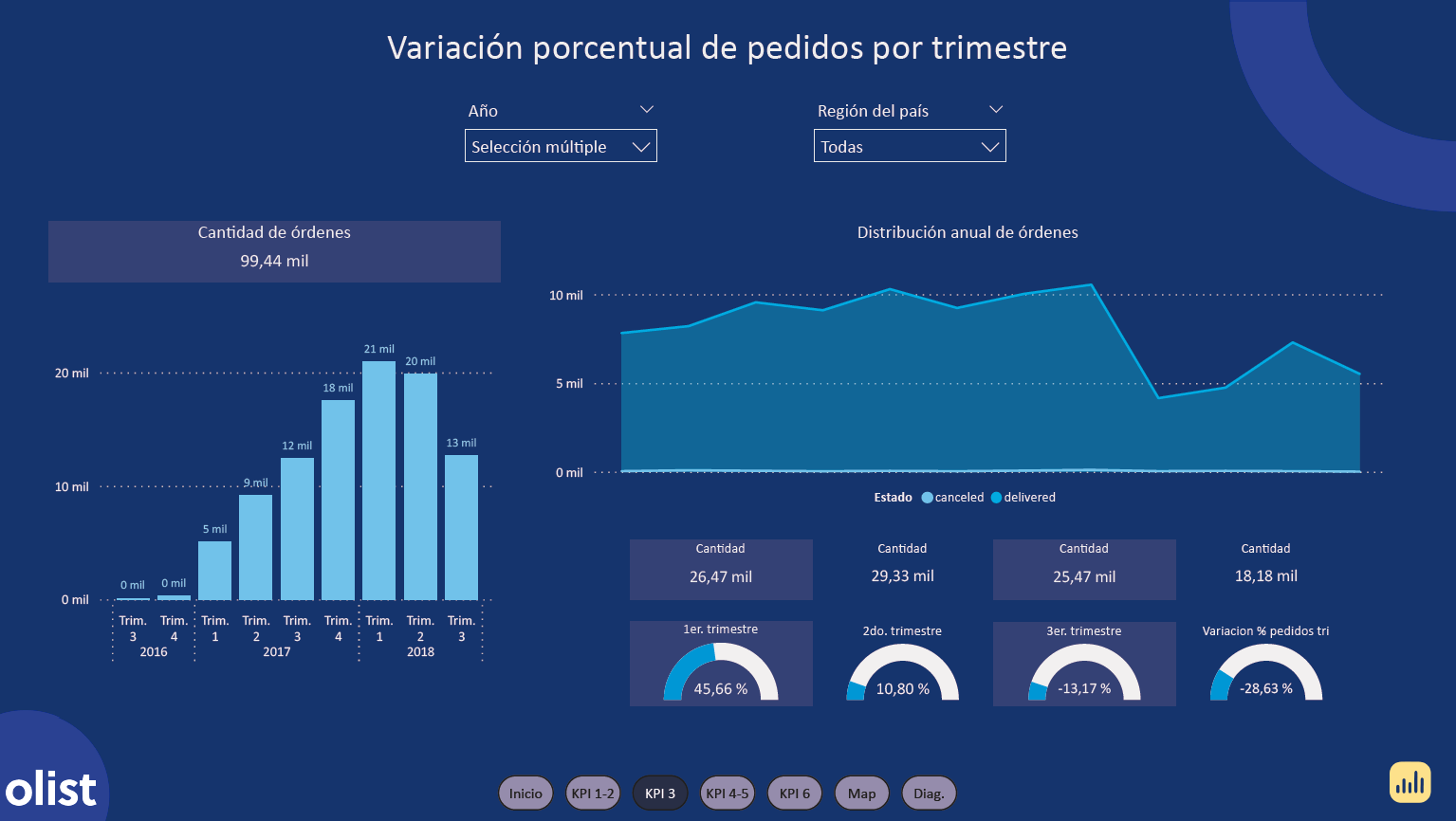
Esta es la plantilla utilizada para las páginas con el análisis de datos.



Por último, mostramos un diseño preliminar (con la utilización de las herramientas gráficas de Power Bi) de cómo quedaría una página para analizar los datos y estudiar un KPI.

La aplicación de cada herramienta gráfica queda sujeta al tipo de datos y análisis que vamos a estudiar, siendo de barras, líneas, métricas y tablas los modelos a los que más recurrimos.

El mockup de una página sería el siguiente:



## Análisis Descriptivo

Nuestros KPIs, ordenados cronológicamente en el proceso de compra de un usuario, están enfocados en 3 estadíos de la operación:

* Los ingresos generados por las ventas.
* El tiempo de demora que tiene la entrega del producto.
* El review (reseña) del usuario.

Cada página se enfoca en el análisis descriptivo de los datos para conocer el comportamiento de dichos KPIs. Así hemos podido obtener información trascendente del desarrollo de Olist en el período de tiempo que comprenden los datasets.

Es importante destacar que cada KPI arroja información relevante, pero ha sido necesario establecer un entorno de datos que complementen su lectura. Así fue posible conocer desde una perspectiva macro la situación, obteniendo un acceso más profundo de los datos, permitiendo que encontremos relaciones de causalidad que no habíamos visto antes.

**Obtención de Insights**

Podemos mencionar como insights los siguientes puntos:

* Las ventas se ven afectadas por fechas especiales (efemérides) o acontecimientos especiales (Black Friday): este último ha sido claramente observado en los datos relacionados al 24 de noviembre de 2017. Ese día las operaciones en la plataforma resultan ser un outlier que afecta las métricas y análisis de los datos.
* La distribución de proveedores y clientes en la geografía de Brasil es un factor clave en la determinación del tiempo de envío de los productos. Las distintas regiones del país tienen diferentes comportamientos.
* El nivel de satisfacción de los clientes está fuertemente relacionado a la región del país donde vive, la disponibilidad de proveedores en un radio cercano a su localización y la demora en el envío del producto.

Cuanto más profundizamos en el análisis de los datos, más relaciones de causalidad encontramos, lo que nos ha permitido obtener diagnósticos y recomendaciones que creemos serán de utilidad para Olist.

# Machine Learning

Uno de los problemas que deberían afrontar todas las empresas es el estudio de las series de tiempo, debido a que su manejo es de vital importancia en la planeación estratégica de cualquier negocio.

Se propone para este problema, analizar la variable aleatoria de ventas de Olist, en función del tiempo para los periodos disponibles (fines de 2016, 2017 y parte de 2018).

En este sentido se pretende la construcción de un modelo de Forecasting para predecir ventas futuras. Considerando que va a ser necesaria la revisión periódica del mismo para el reentrenamiento.

## Análisis de la serie temporal

¿Qué consideramos?

2017 - 201

### 

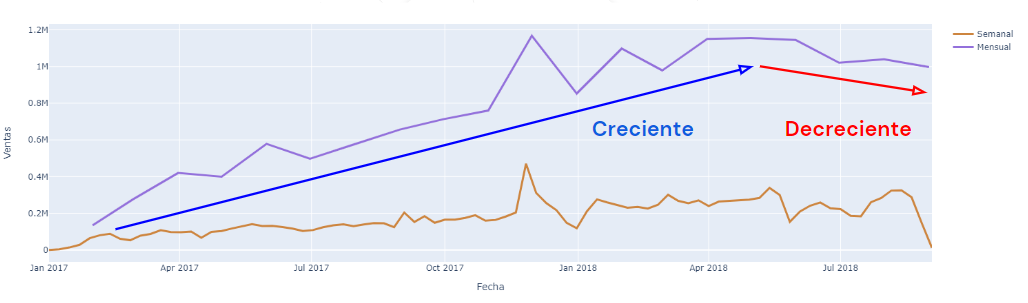
### Componente Estacional

### 

### 

### 

### Componente de Tendencia

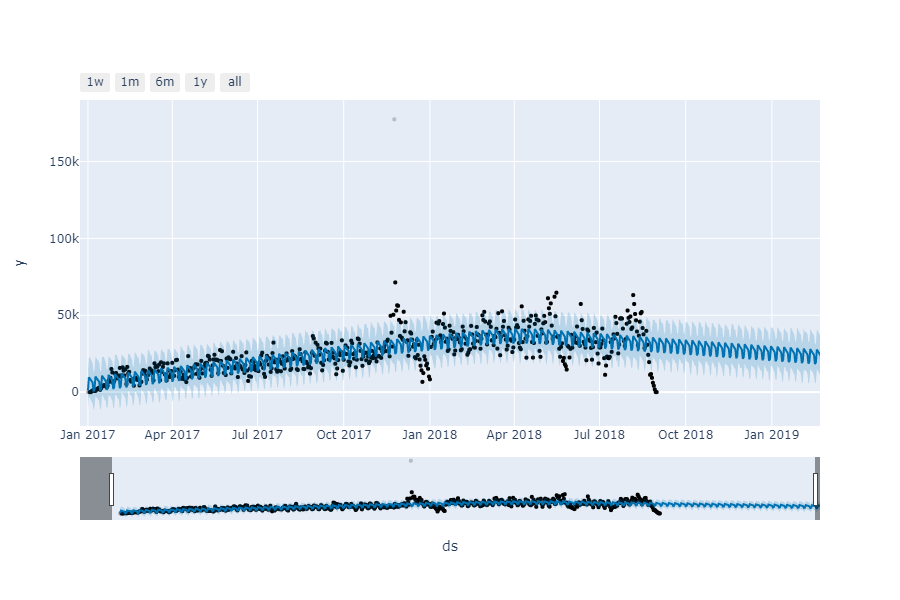


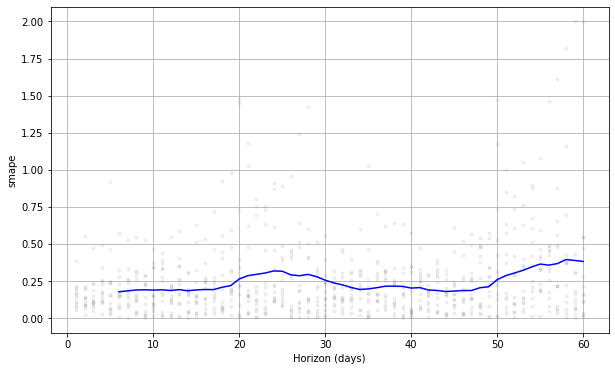
## 

## 

## Forecasting

Propuesta - Análisis de sentimiento - que se va a hacer



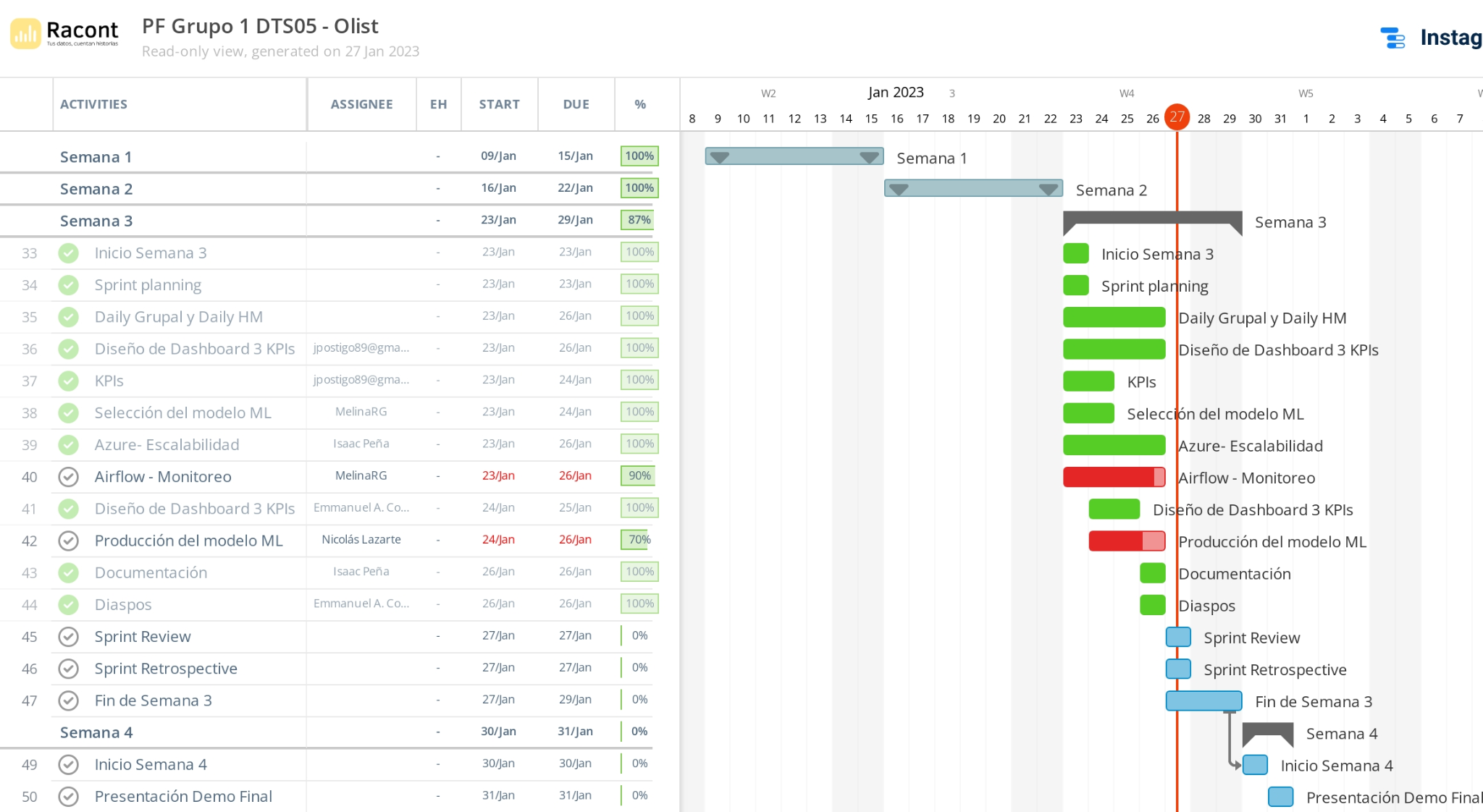


# Conclusiones

Estamos llegando a la recta final del desarrollo del proyecto. Esta semana sin dudas fue fundamental para ver el cambio de cómo los **datos** se están transformando en **información**. A su vez la integración, interpretación y aplicación de dicha información va a resultar en **conocimiento** que nos va a permitir:

* Obtener una visión general del rendimiento del negocio, para poder identificar problemas y oportunidades de mejora.
* Monitorear el progreso y detectar cambios en los patrones de datos.
* Facilitar la comunicación de los resultados y hallazgos a los miembros del equipo y a los líderes de la empresa.
* Poder justificar las acciones que se van a recomendar y proponer.
* Obtener capacidad de procesamiento para grandes cantidades de datos.
* Hacer predicciones y generar recomendaciones basadas en los patrones y tendencias detectadas.
* *Identificar oportunidades de negocio y mejorar la eficiencia operativa de la empresa.*

## Diagrama de Gantt



## 

## Producto Final

