

4. Entregables

En esta sección el entregables consiste una **presentación ejecutiva** dirigida a un CEO, en formato **pdf** para comunicar en una charla de no más de 10 minutos (7 diapositivas) los resultados del análisis realizado. Dicha presentación deberá considerar los siguientes elementos:

Estructura del reporte por secciones:

- Objetivos (Problemas a resolver de acuerdo a lo que `Oilst` solicitó al equipo de `Brasil IA Consulting`),
- Métodos de análisis y fuentes de datos empleadas.
- Principales resultados del análisis, que deberán cumplir:
 - Ser comunicados de manera corta, clara y concreta para una audiencia no experta en estadística,
 - Deben ilustrarse y fundamentarse a través de tablas, visualizaciones estáticas o dinámicas, construidas con Python.
- Conclusiones
- Recomendaciones de acciones que Brasil IA Consulting recomienda a Oilst para afrontar el problema de retrasos en su ordenes
 - Dichas recomendaciones deben ser acciones concretas a implementar,
 - Se deben ligar a los resultados del análisis.
- Limitaciones: deben mencionarse obstáculos en el análisis, que puedan afectar la interpretación de los resultados del negocio.

El reporte deberá denominarse `5_reporte_brasil_ia_consulting.pdf`

Objetivos:

- Identificar las causas de los retrasos en las entregas de las ordenes de Olist.
- Identificar el impacto monetario causado por los retrasos en las entregas
- Identificar las zonas (ciudades, regiones) con más retrasos en las órdenes.

Métodos de análisis y fuentes de datos empleadas

Para este análisis, se utilizaron análisis cuantitativos, así como análisis estadísticos, primero se reunió toda la información disponible relacionada con el proceso de entrega de compras por parte de los clientes, información de los productos, origen de las compras, es decir, de que ciudad/región se realizan las compras, se identifica el tipo de pago utilizado, y reunimos información geográfica de Brasil.

Una vez reunida toda la información, se une en una sola fuente de la verdad usando script de Python, calculando algunos datos para comprender mejor el análisis; se clasifican los retrasos de acuerdo a los días que llevan, en on time, short delay y long delay de las ordenes entregadas.

Además, se busca identificar el costo del impacto de las ordenes con retraso prolongado, su ubicación, es decir, que ciudad/región es la mas afectada por estos retrasos.

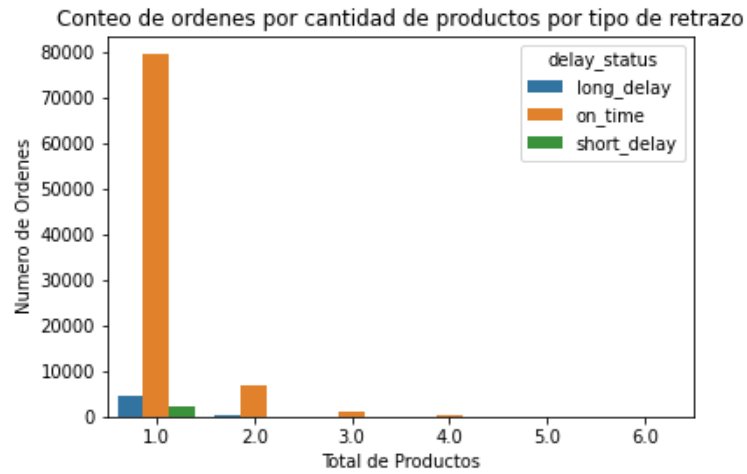
Principales resultados del análisis.

Una vez reunida la información, la procesamos para obtener una sola fuente de la verdad que analizaremos.

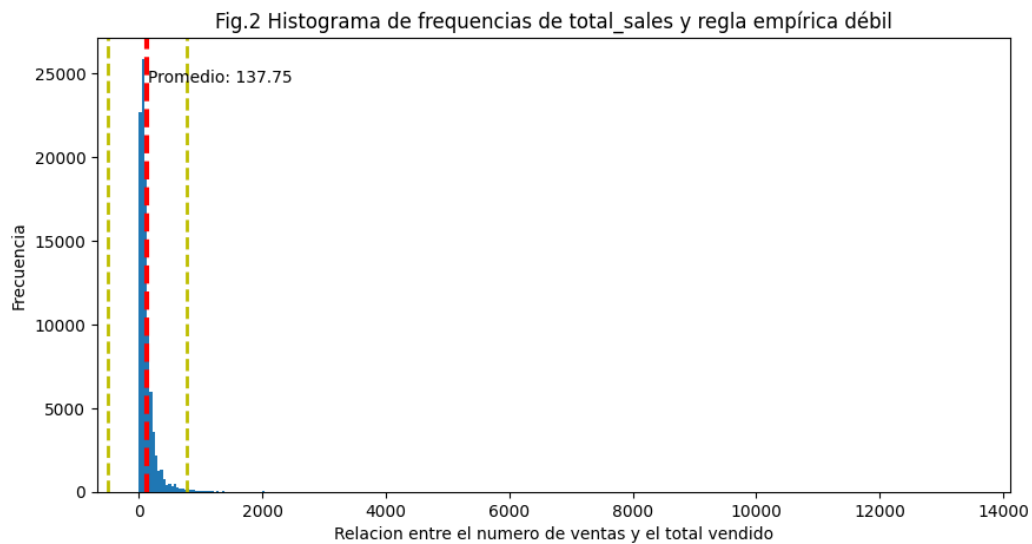
- Analizamos las ordenes por retraso por trimestre con el fin de identificar cuando comenzaron las incidencias de ordenes con retraso.



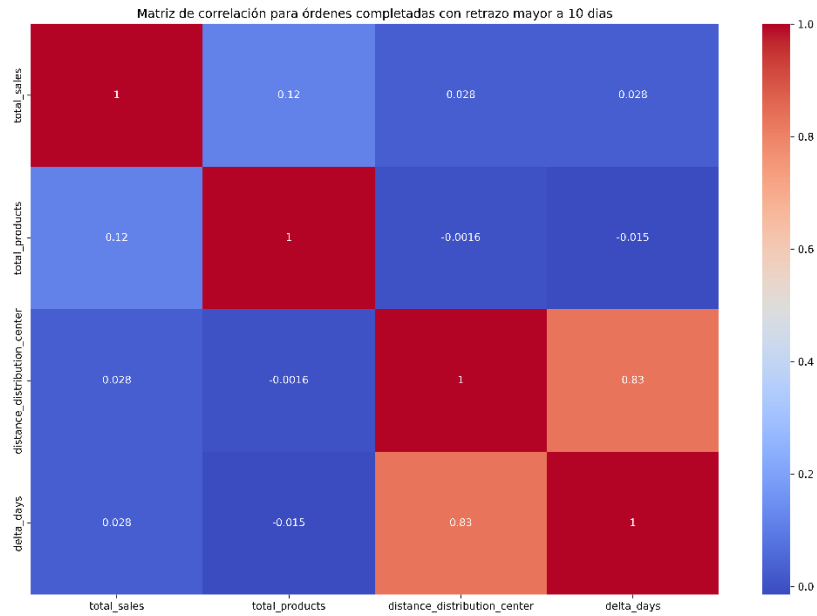
- Después deseamos conocer el volumen del impacto, decir cuantos productos contienen las ordenes entregadas por tipo de retraso, aquí deseamos conocer el impacto de la afectación.



- Para medir el impacto del costo de las ordenes con retraso, primero identificamos el costo de las ordenes y su frecuencia usando un histograma, de este modo observamos el valor de venta mas frecuente.



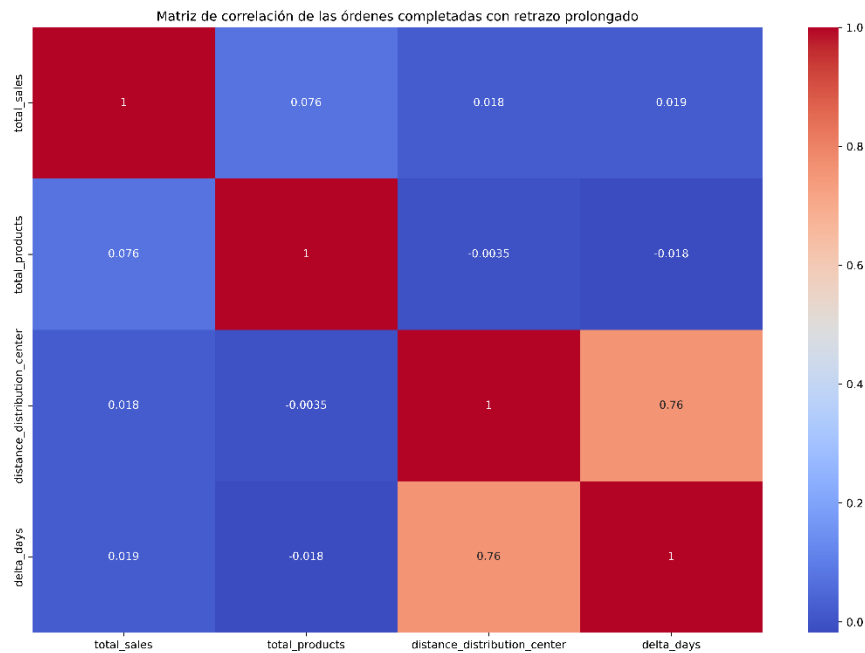
- Ahora deseamos conocer la relación entre las variables numéricas, con el fin de determinar que variable afecta más cual, esto lo hacemos mediante una matriz de correlación, esto nos dice que las ordenes con retraso prolongado están directamente relacionadas con la distancia al centro de distribución, esto es, que una orden se retrasa por que su destino se encuentra muy alejado del centro de distribución.



- En el histograma anterior, solo observamos el total de compra mas frecuente, ahora analizamos la misma variable pero por las ordenes con retraso corto y prolongado, con el fin de medir el impacto en la venta de las ordenes con algún tipo de retraso.

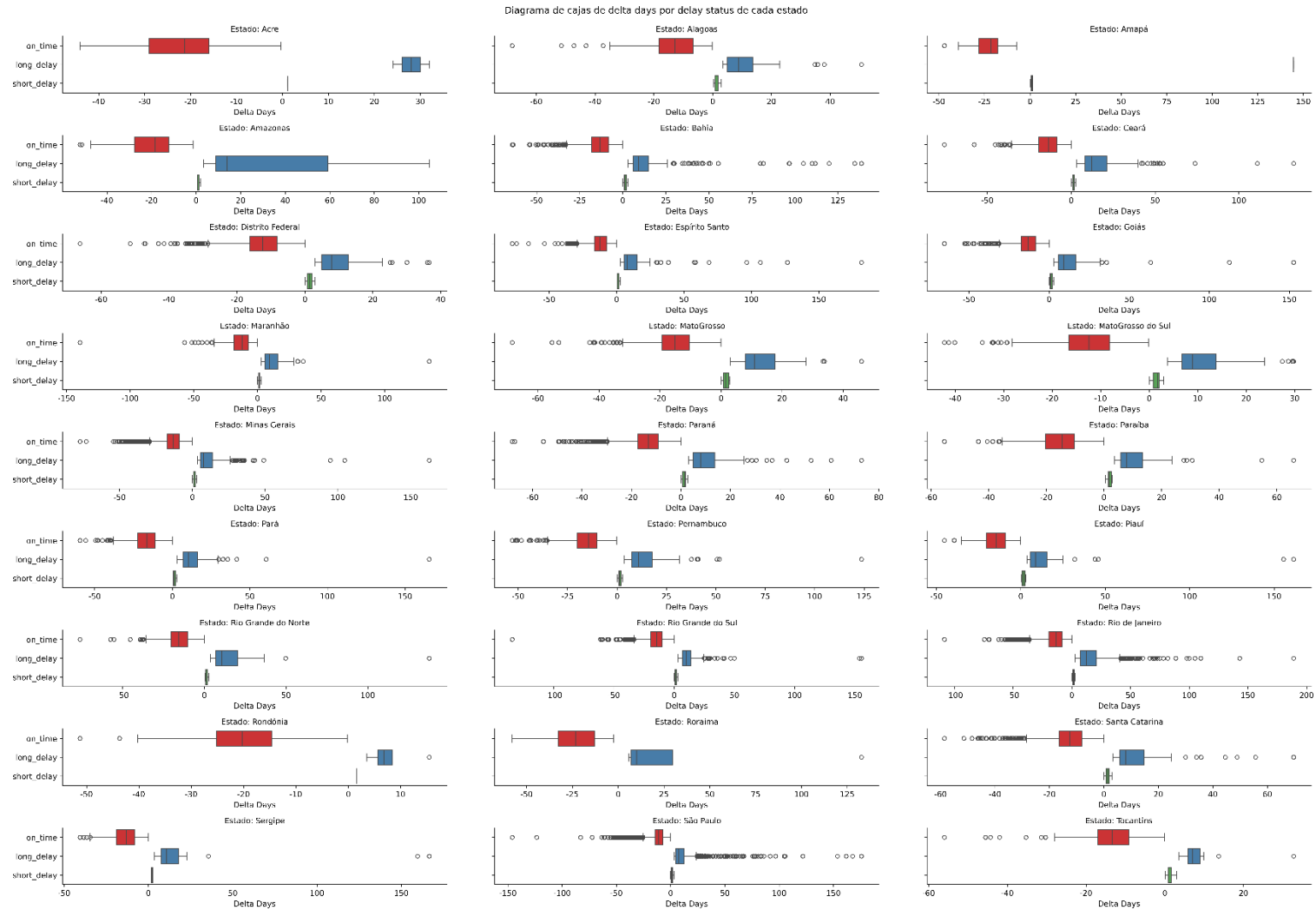


- Asimismo, para comprender mejor la relación de las ordenes con retraso prolongado contra las demás variables, usamos una matriz de correlación:

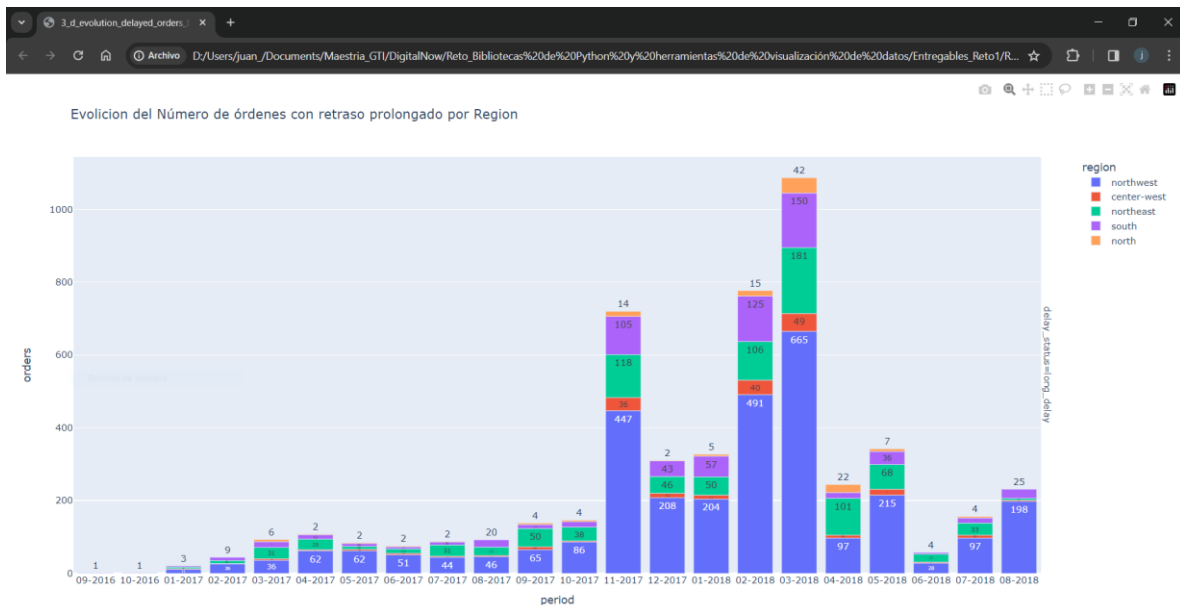


Como observamos, en la matriz anterior identificamos una fuerte relación entre la distancia al centro de distribución con las ordenes con algún tipo de retraso, en esta matriz vemos que existe una fuerte relación entre los las ordenes con retraso prolongado y la distancia al centro de distribución.

- Ya que hemos identificado, el impacto de las ordenes con retraso, tanto en valor de venta y sabemos que la distancia al centro de distribución en un factor, ahora necesitamos identificar cual es el estado mas afectado y que tan afectado es, para esto utilizamos un diagrama de cajas, este nos permite visualizar varios datos por estado, y asi identificar cual es el mas afectado y que tanto.

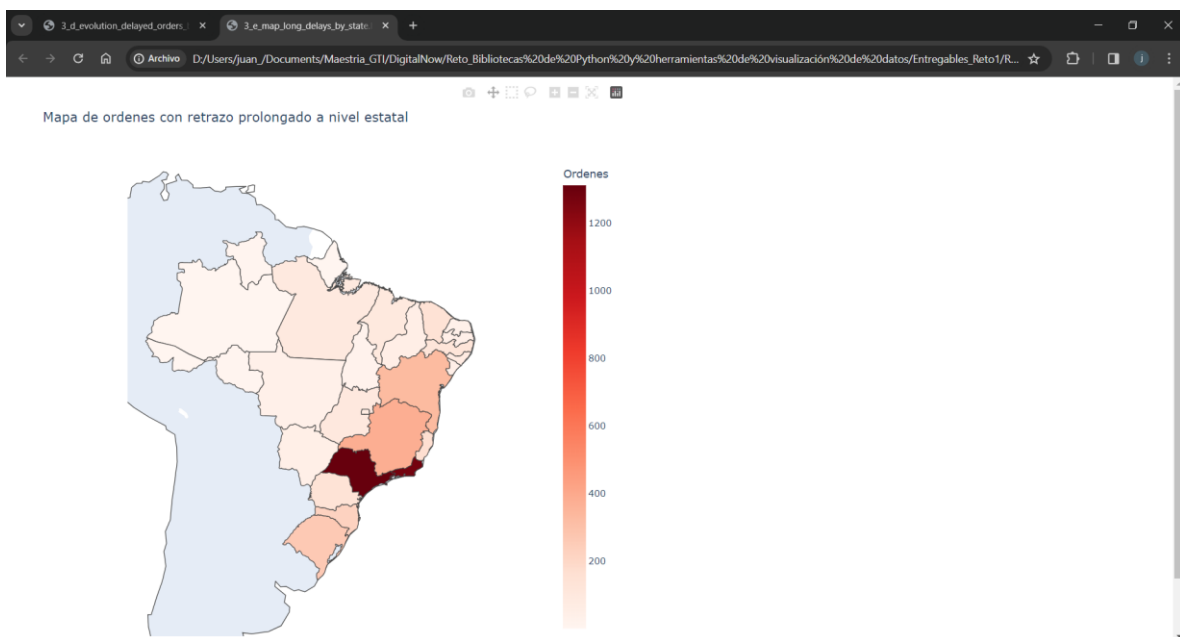


- Por otro lado, es necesario, conocer la evolución del numero de ordenes por región, con el fin de identificar en qué momento comenzó el incremento de las ordenes y cuales son las regiones con mayor demanda.



Aquí podemos ver como la región northwest, experimento un alza en el numero de ordenes en el mes de noviembre de 2017.

- Sin embargo, para identificar donde s ubica el mayor impacto de las ordenes con retraso prolongado, creamos este mapa, ese nos dice que el estado de Sao Paulo se concentran más de 1200 órdenes con retraso prolongado.

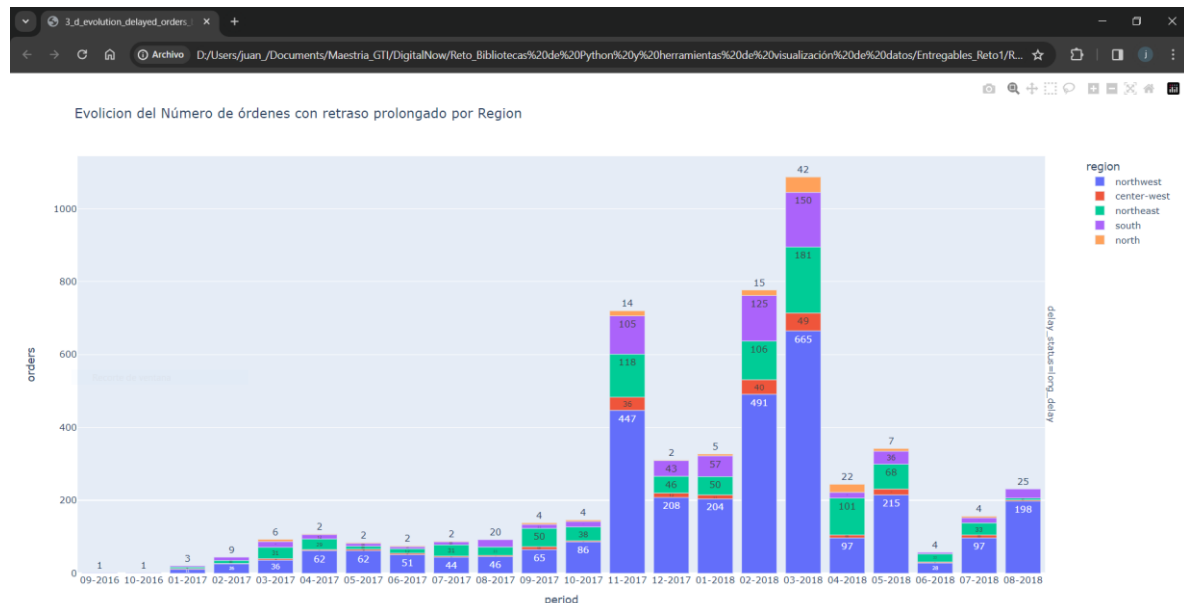


Conclusiones

Después de analizar la información, aplicar ciertas metodologías, operaciones estadísticas, cuantitativas, crear gráficas, concluimos que el estado mas afectado por las ordenes con retraso prolongado es Sao Paulo, pero cual es la causa; el retraso en las ordenes se debe que el centro de destrucción que debe atender esas ordenes se encuentra un poco lejos de su destino, aunque esto se debe a un incremento en las ordenes, es decir, los centros de distribución se vieron superados en el numero de ordenes que recibieron y esto ocasiono que lagunas ordenes se desviaran a otros centros de distribución cuya mayor distancia ocasiono un retraso prolongado.

Recomendaciones

Dicho lo anterior, y después de analizar los resultados, recomendamos aplicar modelos de machine learning a los datos, para predecir el incremento en las compras, como podemos observar, este incremento de da en los meses de noviembre, diciembre, enero.



Aplicando estos modelos, podemos identificar mas o menos en que porcentaje crecerán las compras, y asi realinear las estrategias para responder a este crecimiento, es decir, la contratación de empleados temporales, acuerdos con empresas de mensajería, para acelerar el proceso de entregas, asimismo, si identificamos cuales son los productos con mayor demanda, podríamos realizar compras anticipadas, y llenar los almacenes con estos productos.