

A woman with dark hair is looking down at a tablet computer she is holding. The background is a blurred city street at night, with lights from buildings and streetlights visible through a window. The entire image has a blue color overlay.

LA SAMBA DEL DATA ANÁLISIS.

BIBLIOTECAS DE PYTHON Y HERRAMIENTAS
DE VISUALIZACIÓN DE DATOS

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 más afectada por estos retrasos.



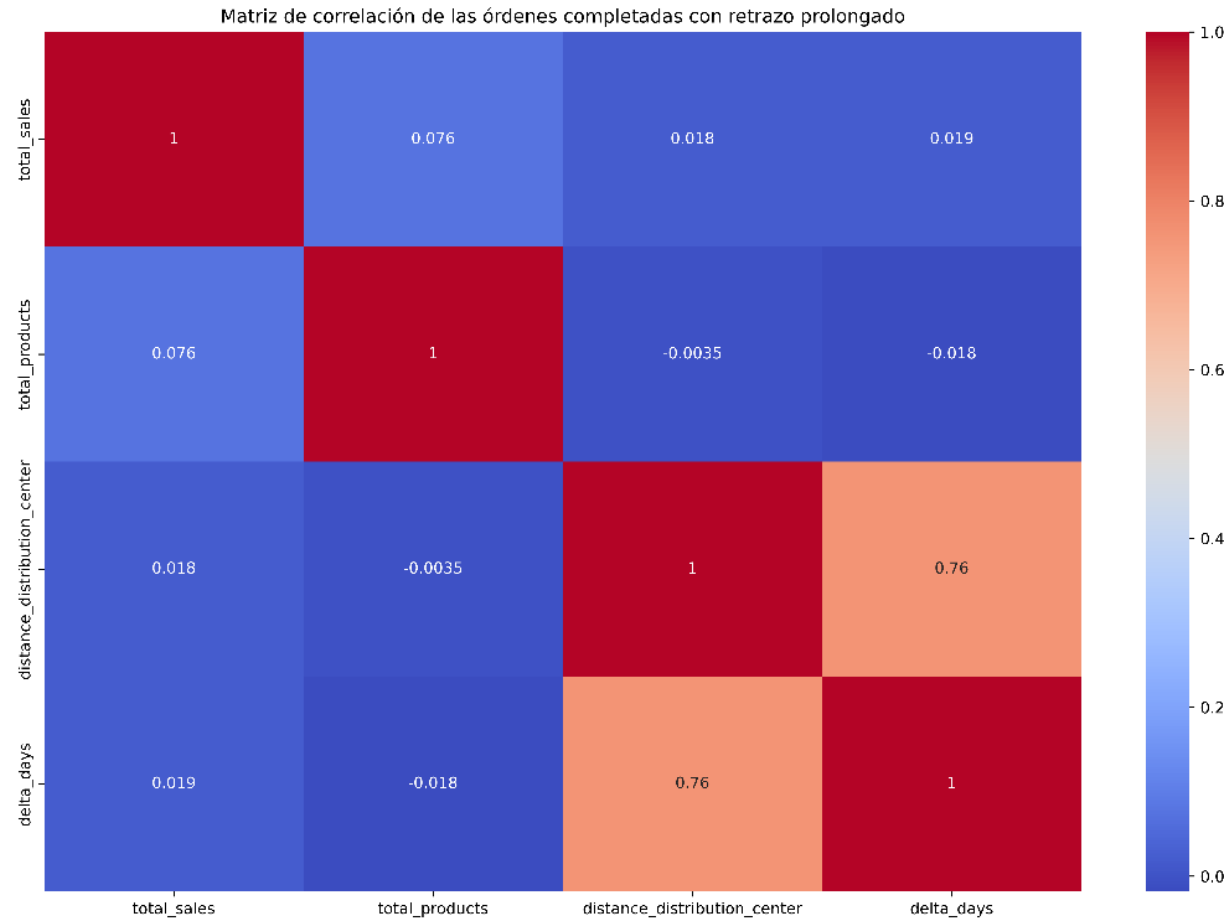
PRINCIPALES RESULTADOS DEL ANÁLISIS

ANALIZAMOS LAS ORDENES POR RETRASO POR TRIMESTRE CON EL FIN DE IDENTIFICAR CUANDO COMENZARON LAS INCIDENCIAS DE ORDENES CON RETRASO.

Proporción de ventas por `delay_status` por quarter



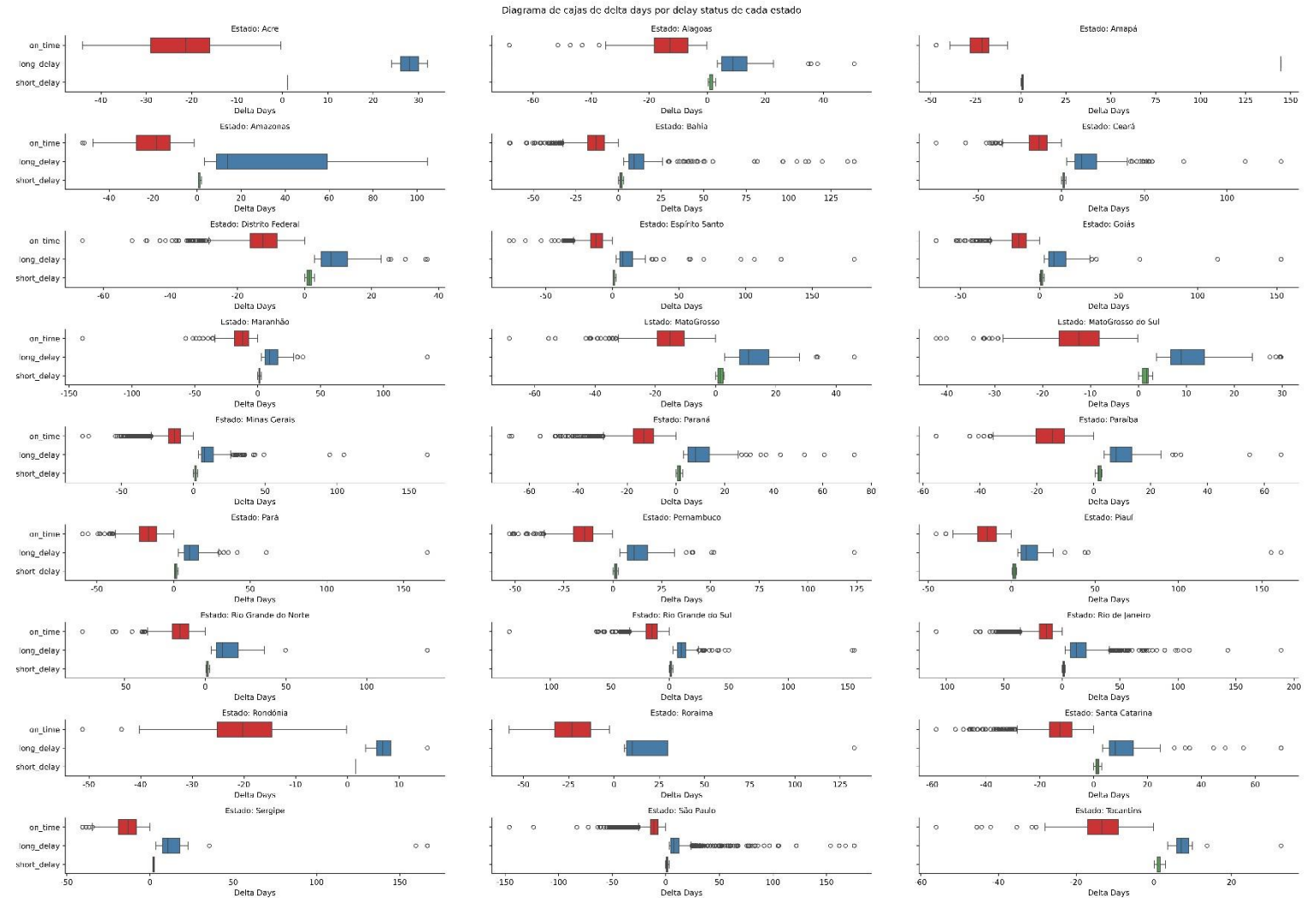
PARA COMPRENDER MEJOR LA RELACIÓN DE LAS ORDENES CON RETRASO PROLONGADO CONTRA LAS DEMÁS VARIABLES, USAMOS UNA MATRIZ DE CORRELACIÓN



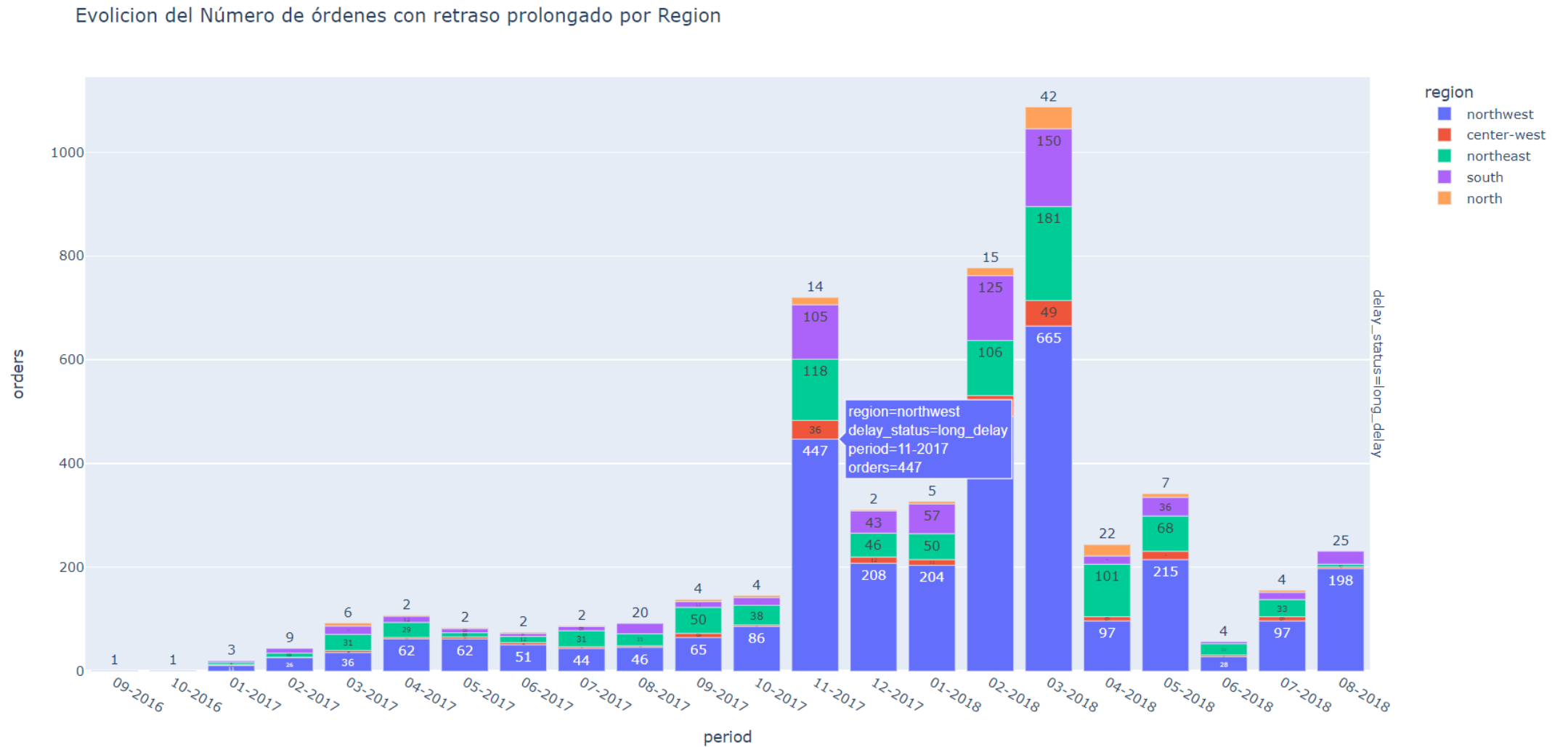
Como observamos, en la matriz anterior identificamos una fuerte relación entre la distancia al centro de distribución con las ordenes con algún tipo de retraso, en esta matriz vemos que existe una fuerte relación entre los las ordenes con retraso prolongado y la distancia al centro de distribución.

DIAGRAMA DE CAJAS DE DELTA DAYS POR DELAY STATUS DE CADA ESTADO

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.



EVOLUCIÓN DEL NÚMERO DE ÓRDENES CON RETRASO PROLONGADO POR REGION

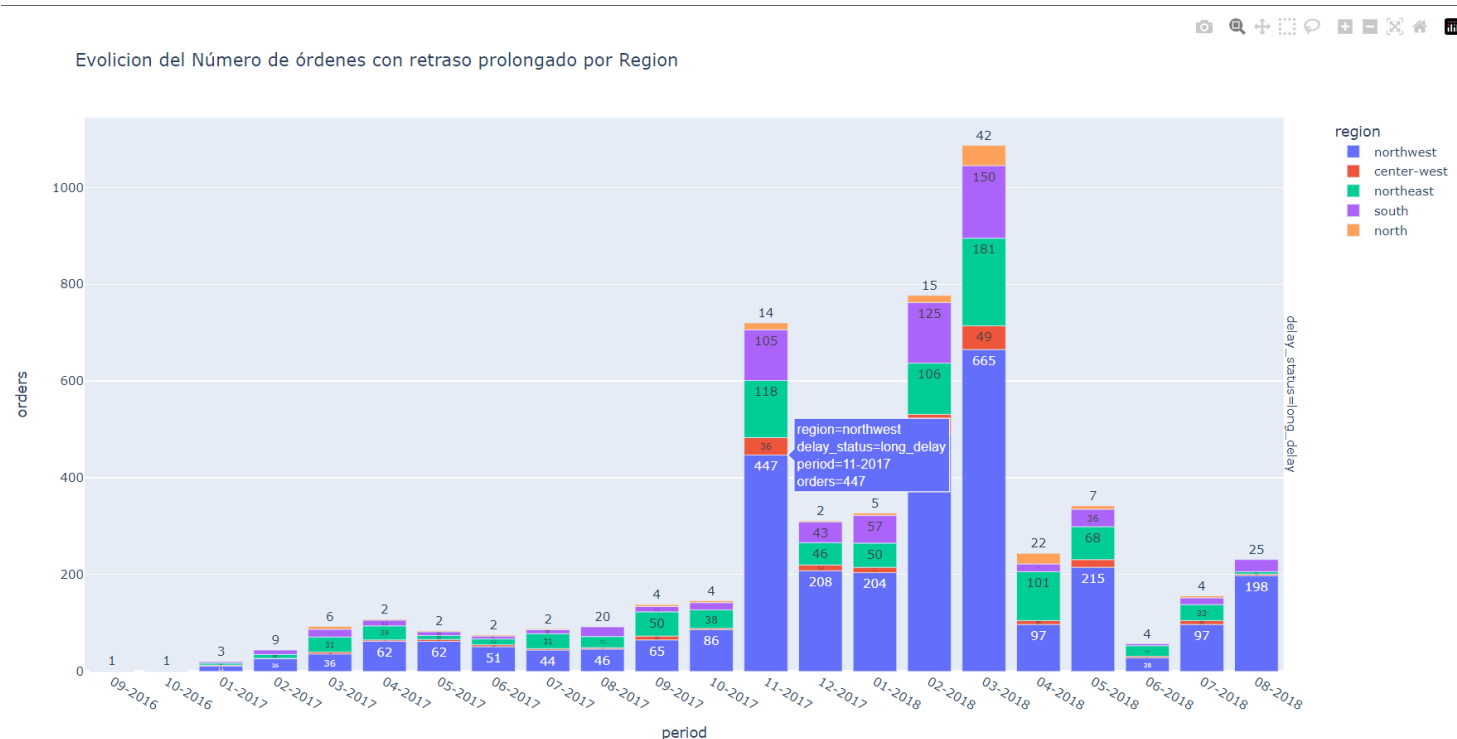


CONCLUSIONES

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.