

Análisis de los despachos de los bienes de consumo de la UNAD y prospección de gastos

Andrés Felipe Triana Cárdenas

Asesor

John Freddy Moreno Lozada

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2024

Resumen

En el presente trabajo de grado, se desarrolla un proyecto aplicado cuyo objetivo principal es optimizar la gestión de compras de bienes de consumo en la Universidad Nacional Abierta y a Distancia (UNAD). Actualmente, la recopilación manual de las necesidades de compra y la desactualización de la información, causadas por las demoras en la entrega de datos por parte de los centros regionales, generan ineficiencias en la proyección y elaboración del plan anual de compras. Además, la toma de decisiones se ve limitada debido a la dificultad para identificar patrones y tendencias de consumo.

La propuesta de solución incluye el desarrollo un modelo predictivo para optimizar las compras en la UNAD, abordando ineficiencias logísticas. Mediante análisis exploratorio y la implementación de algoritmos de aprendizaje automático, se busca anticipar necesidades de consumo y reducir costos operativos. También la generación de reportes que permitan mejorar la toma de decisiones.

La metodología adoptada se divide en cinco fases: organización de la información y limpieza de los datos, análisis descriptivo de los datos, análisis exploratorio de los datos, propuesta o implementación de un modelo predictivo, validación de este y generación de reportes.

Con la implementación de esta solución los resultados preliminares indican un aumento en la precisión de proyecciones en un 25%.

Palabras clave: Análisis de datos, Inteligencia de negocios, Modelo predictivo.

Abstract

In this degree project, an applied project is developed whose main objective is to optimize the management of purchases of consumer goods at the Universidad Nacional Abierta y a Distancia (UNAD). Currently, the manual compilation of purchasing needs and the lack of updated information, caused by delays in the delivery of data by the regional centers, generate inefficiencies in the projection and elaboration of the annual purchasing plan. In addition, decision making is limited due to the difficulty in identifying consumption patterns and trends.

The proposed solution includes the development of a predictive model to optimize purchases at UNAD, addressing logistical inefficiencies. Through exploratory analysis and the implementation of machine learning algorithms, we seek to anticipate consumption needs and reduce operating costs. Also, the generation of reports to improve decision making.

The methodology adopted is divided into five phases: information organization and data cleaning, descriptive analysis of the data, exploratory analysis of the data, proposal or implementation of a predictive model, validation of the model and generation of reports.

With the implementation of this solution, preliminary results indicate a 25% increase in the accuracy of projections.

Keywords: Business intelligence, Data analytics, Predictive modeling.

Tabla de Contenido

Introducción	12
Descripción del Problema	13
Planteamiento del Problema	14
Justificación	15
Objetivos	16
Objetivo General	16
Objetivos Específicos	16
Marco Referencial.....	17
Estado del Arte (Trabajos Anteriores Sobre el Tema)	17
Marco Conceptual.....	19
Modelo Predictivo.....	19
Machine Learning	19
Análisis de Datos	20
Optimización Logística.....	20
Planificación de Compras	21
Marco Normativo.....	22
Estatuto General de la UNAD.....	22
Ley 80 de 1993 - Estatuto General de Contratación de la Administración Pública.....	22
Ley 1266 de 2008 - Ley de Habeas Data	23
Decreto 1082 de 2015 - Régimen de Contratación Estatal	23
Normativa sobre Protección del Medio Ambiente y Uso Sostenible de Recursos	23
Normativa de Propiedad Intelectual de la UNAD.....	24

Metodología	25
Fase 1 Recolección y Preprocesamiento de los Datos	25
Recolección de Datos	25
Limpieza y Preprocesamiento	25
Fase 2 Análisis de Datos	25
Análisis Descriptivo	25
Análisis Exploratorio.....	26
Fase 3 Análisis Predictivo.....	26
Estudio de Variables.....	26
Optimización	26
Fase 4 Diseño de reportes	26
Frecuencia de Reportes	26
Estructuración de Reportes.....	27
Recopilación.....	27
Tipo de Estudio.....	28
Recolección de Datos.....	29
Resultados	30
Análisis Descriptivo	30
Variables Cuantitativas.....	30
Variables Categóricas.....	42
Variable Fecha	49
Análisis Exploratorio	53
Análisis General de Despachos.....	53

Despachos a lo Largo del Tiempo (2019-2023).....	53
Despachos por Zonas.....	55
Despachos por Centros de Costo.....	66
Análisis Financiero.....	67
Evolución de los Costos Totales	67
Costos por Centros de Costo	75
Contribución Porcentual de los Costos Totales Anuales.....	77
Centros de Costo con Mayor Gasto.....	79
Centros de Costo con Menor Gasto.....	85
Productos Más Costosos.....	91
Productos Menos Costosos.....	92
Distribución y Logística.....	93
Productos Más Despachados	93
Productos Menos Despachados	95
Análisis Predictivo	97
Relaciones entre las Variables.....	97
Variables Relevantes para una Proyección (Predicción) de la Demanda.....	98
Selección del Modelo	99
Preprocesamiento de los Datos.....	101
Implementación del Modelo.....	103
Diseño de Reportes	108
Frecuencia de Reportes	108
Estructuración de Reportes.....	108

Visualización	111
Conclusiones	112
Recomendaciones	114
Referencias Bibliográficas	116

Lista de Figuras

Figura 1 <i>Información Contendida en la Data</i>	30
Figura 2 <i>Estadísticas Descriptivas de la Data en las Variables Cuantitativas</i>	34
Figura 3 <i>Histograma de la Distribución de la Variable Bodega</i>	37
Figura 4 <i>Histograma de la Distribución de la Variable Número</i>	37
Figura 5 <i>Histograma de la Distribución de la Variable Renglón</i>	38
Figura 6 <i>Histograma de la Distribución de la Variable Producto</i>	39
Figura 7 <i>Histograma de la Distribución de la Variable Cantidad</i>	40
Figura 8 <i>Boxplot de la Distribución de la Variable Costo</i>	41
Figura 9 <i>Histograma de la Distribución de la Variable Centro_Costo</i>	42
Figura 10 <i>Estadísticas Descriptivas de las Variables Categóricas</i>	43
Figura 11 <i>Histograma de la Distribución de la Variable Tipo</i>	45
Figura 12 <i>Histograma de la Distribución de la Variable Nomb_Producto</i>	46
Figura 13 <i>Histograma de la Distribución de la Variable Nomb_Centro_Costo</i>	47
Figura 14 <i>Histograma de la Distribución de la Variable Zona</i>	48
Figura 15 <i>Estadísticas Descriptivas de la Variable Fecha</i>	49
Figura 16 <i>Histograma de la Distribución de la Variable Fecha por Año</i>	50
Figura 17 <i>Histograma de la Distribución de la Variable Fecha por Mes</i>	51
Figura 18 <i>Histograma de la Distribución de la Variable Fecha por Mes y Año</i>	52
Figura 19 <i>Tendencias de Despachos a lo Largo del Tiempo</i>	53
Figura 20 <i>Cantidad Total de Despachos a lo Largo del Tiempo</i>	54
Figura 21 <i>Tendencias de Despachos en la Zona Amazonia a lo Largo del Tiempo</i>	55
Figura 22 <i>Tendencias de Despachos en la Zona Caribe a lo Largo del Tiempo</i>	56

Figura 23 <i>Tendencias de Despachos en la Zona Boyacá a lo Largo del Tiempo</i>	56
Figura 24 <i>Tendencias de Despachos en la Zona Oriente a lo Largo del Tiempo</i>	57
Figura 25 <i>Tendencias de Despachos en la Zona Centro Sur a lo Largo del Tiempo</i>	57
Figura 26 <i>Tendencias de Despachos en la Zona Cundinamarca a lo Largo del Tiempo</i>	58
Figura 27 <i>Tendencias de Despachos en la Zona Occidente a lo Largo del Tiempo</i>	58
Figura 28 <i>Tendencias de Despachos en la Sede Nacional a lo Largo del Tiempo</i>	59
Figura 29 <i>Tendencias de Despachos en la Zona Sur a lo Largo del Tiempo</i>	59
Figura 30 <i>Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Amazonia</i>	61
Figura 31 <i>Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Caribe</i>	61
Figura 32 <i>Cantidad Total de Despachos a lo Largo del Tiempo Para la Zona Boyacá</i>	62
Figura 33 <i>Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Oriente</i>	62
Figura 34 <i>Cantidad Total de Despachos a lo Largo del Tiempo Para la Zona Centro Sur</i>	63
Figura 35 <i>Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Cundinamarca..</i>	63
Figura 36 <i>Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Occidente</i>	64
Figura 37 <i>Cantidad Total de Despachos a lo Largo del Tiempo Para la Sede Nacional</i>	64
Figura 38 <i>Cantidad Total de Despachos a lo Largo del Tiempo Para la Zona Sur</i>	65
Figura 39 <i>Cantidad Total de Despachos por Centro de Costo</i>	66
Figura 40 <i>Evolución de Costos Totales</i>	67
Figura 41 <i>Evolución de Costos Totales Anuales por Zona</i>	69
Figura 42 <i>Evolución de Costos Totales Anuales en la Zona Amazonia Orinoquia</i>	70
Figura 43 <i>Evolución de Costos Totales Anuales en la Zona Caribe</i>	70
Figura 44 <i>Evolución de Costos Totales Anuales en la Zona Boyacá</i>	71
Figura 45 <i>Evolución de Costos Totales Anuales en la Zona Oriente</i>	71

Figura 46 <i>Evolución de Costos Totales Anuales en la Zona Centro Sur</i>	72
Figura 47 <i>Evolución de Costos Totales Anuales en la Zona Cundinamarca</i>	72
Figura 48 <i>Evolución de Costos Totales Anuales en la Zona Occidente</i>	73
Figura 49 <i>Evolución de Costos Totales Anuales en la Zona Sede Nacional</i>	73
Figura 50 <i>Evolución de Costos Totales Anuales en la Zona Sur</i>	74
Figura 51 <i>Costos Totales por Centro de Costo</i>	76
Figura 52 <i>Porcentaje de Costos Totales Anuales por Zona</i>	77
Figura 53 <i>Contribución Proporcional de Costos Totales Anuales por Zonas</i>	77
Figura 54 <i>Comparación de Costos Totales Anuales por Zonas</i>	78
Figura 55 <i>Centros de Costo con Mayor Gasto Total</i>	79
Figura 56 <i>Centros de Costo con Mayor Gasto en el Año 2019</i>	80
Figura 57 <i>Centros de Costo con Mayor Gasto en el Año 2020</i>	81
Figura 58 <i>Centros de Costo con Mayor Gasto en el Año 2021</i>	82
Figura 59 <i>Centros de Costo con Mayor Gasto en el Año 2022</i>	83
Figura 60 <i>Centros de Costo con Mayor Gasto en el Año 2023</i>	84
Figura 61 <i>Centros de Costo con Menor Gasto Total</i>	85
Figura 62 <i>Centros de Costo con Menor Gasto Total en el Año 2019</i>	86
Figura 63 <i>Centros de Costo con Menor Gasto Total en el Año 2020</i>	87
Figura 64 <i>Centros de Costo con Menor Gasto Total en el Año 2021</i>	88
Figura 65 <i>Centros de Costo con Menor Gasto Total en el Año 2022</i>	89
Figura 66 <i>Centros de Costo con Menor Gasto Total en el Año 2023</i>	90
Figura 67 <i>Productos Más Costosos</i>	91
Figura 68 <i>Productos Menos Costosos</i>	92

Figura 69 <i>Productos Más Despachados</i>	93
Figura 70 <i>Productos Más Despachados por Año</i>	94
Figura 71 <i>Productos Menos Despachados</i>	95
Figura 72 <i>Productos Menos Despachados en Cada Año</i>	96
Figura 73 <i>Matriz de Correlación de las Variables Cuantitativas</i>	97
Figura 74 <i>Información del Dataset con Variables Año y Fecha</i>	102
Figura 75 <i>Información del Dataset con Serie de Tiempo Completa</i>	102
Figura 76 <i>Datos Seleccionados para el Modelo</i>	103
Figura 77 <i>Grafico de Resultados del Modelo Inicial</i>	104
Figura 78 <i>Gráfico de Resultados del Modelo 2</i>	105
Figura 79 <i>Gráfico de Resultados del Modelo Optimizado</i>	107
Figura 80 <i>Reporte de Despachos</i>	109
Figura 81 <i>Reporte Financiero</i>	110
Figura 82 <i>Reporte Logístico</i>	111

Introducción

El presente trabajo se centra en la optimización de la gestión de compras de bienes de consumo en la Universidad Nacional Abierta y a Distancia (UNAD). En la actualidad, el proceso de recopilación y consolidación de información sobre las necesidades de compra de los centros regionales es manual, lo que genera retrasos significativos e ineficiencias operativas. Estas limitaciones impactan negativamente la planificación y ejecución del plan anual de compras, comprometiendo la disponibilidad oportuna de los bienes necesarios para el correcto funcionamiento de la institución.

Este proyecto aborda esta problemática mediante el desarrollo de un sistema basado en análisis de datos y un modelo predictivo diseñado para anticipar las necesidades de consumo y mejorar la toma de decisiones estratégicas. Según Zhang y Wang (2017), la previsión de la demanda permite reducir ineficiencias logísticas, lo cual resulta crítico en instituciones académicas como la UNAD. A través de un enfoque innovador, que combina la exploración y análisis de patrones históricos de consumo, el proyecto busca proporcionar soluciones prácticas que optimicen los recursos y procesos logísticos de la UNAD, asegurando una mayor eficiencia y sostenibilidad en sus operaciones.

El enfoque adoptado no solo se alinea con los principios de sostenibilidad y excelencia de la UNAD, sino que también posiciona a la institución como un referente en la adopción de tecnologías avanzadas para la gestión de recursos. Este proyecto representa una oportunidad clave para mejorar la toma de decisiones, optimizar los recursos institucionales y garantizar un suministro adecuado que respalde las actividades académicas y administrativas de la universidad.

Descripción del Problema

El Grupo Funcional de Almacén e Inventarios de la UNAD se enfrenta a un proceso ineficiente para la elaboración del plan anual de compras de bienes de consumo, lo que impacta negativamente en la operación de la institución. La dependencia de la recopilación manual de información de cada centro regional acerca de sus necesidades genera demoras considerables en la captura y consolidación de datos. Este retraso no solo obstaculiza la proyección y ejecución oportuna del plan de compras, sino que también dificulta la capacidad del Grupo Funcional para garantizar que los bienes necesarios estén disponibles en el momento adecuado. Como resultado, se compromete la eficiencia operativa y se corre el riesgo de insatisfacción en las necesidades de abastecimiento de los diferentes centros regionales. En este contexto, es fundamental abordar esta problemática mediante la implementación de un sistema más ágil y eficiente que optimice la recopilación de información y permita una mejor planificación de las compras.

Planteamiento del Problema

Actualmente, el Grupo Funcional de Almacén e Inventarios de la UNAD enfrenta dificultades en la elaboración del plan anual de compras de bienes de consumo. Este proceso requiere solicitar a cada centro regional información sobre sus necesidades, lo que se realiza de manera manual. La recopilación de datos de esta forma genera demoras significativas en la captura y consolidación de la información, lo que, a su vez, retrasa tanto la proyección como la ejecución del plan de compras. Estas demoras afectan la capacidad de la UNAD para satisfacer oportunamente las necesidades de sus centros regionales, comprometiendo la eficiencia operativa de la institución.

Justificación

La justificación para abordar el problema identificado en el Grupo Funcional de Almacén e Inventarios de la UNAD radica en la necesidad de optimizar el proceso de elaboración del plan anual de compras de bienes de consumo. Un sistema de recopilación manual ineficiente no solo genera demoras significativas en la captura y consolidación de información, sino que también impacta negativamente en la operatividad de la institución.

La correcta planificación y ejecución del plan de compras es esencial para garantizar el suministro oportuno de bienes a los diferentes centros regionales. Cualquier retraso en este proceso puede traducirse en interrupciones en las actividades académicas y administrativas, afectando la calidad del servicio que la UNAD ofrece a sus estudiantes y al personal.

Además, una gestión ineficiente del inventario puede llevar a la acumulación de productos innecesarios o a la falta de suministros críticos, lo que representa un desperdicio de recursos financieros y humanos. Por lo tanto, implementar un sistema más ágil y automatizado no solo mejorará la eficiencia operativa del Grupo Funcional, sino que también contribuirá a una gestión más efectiva de los recursos, alineándose con los objetivos institucionales de excelencia y sostenibilidad.

Finalmente, al mejorar el proceso de recopilación de información y la planificación de compras, la UNAD podrá responder más adecuadamente a las necesidades de sus centros regionales, promoviendo un entorno de trabajo más eficiente y satisfactorio para todos los involucrados.

Objetivos

Objetivo General

Optimizar la gestión de compras de bienes de consumo en la Universidad Nacional Abierta y a Distancia (UNAD) mediante el análisis de estadísticas de consumo, la posible implementación de un modelo predictivo y la generación de reportes eficientes.

Objetivos Específicos

Analizar los datos de consumo de elementos de oficina en los diferentes centros regionales y dependencias de la UNAD.

Realizar un análisis predictivo que permita anticipar las necesidades de bienes de consumo en la UNAD

Diseñar reportes basados en los resultados obtenidos, que ayuden a toma de decisiones estratégicas en la UNAD.

Marco Referencial

Estado del Arte (Trabajos Anteriores Sobre el Tema)

La planificación de compras ha evolucionado significativamente con el avance de las tecnologías de análisis de datos y los modelos predictivos. Tradicionalmente, los sistemas de inventario eran reactivos, basándose en la reposición de stock una vez que alcanzaban niveles críticos. Sin embargo, el aumento en la complejidad de las operaciones logísticas y el volumen de bienes ha llevado a instituciones y empresas a adoptar herramientas predictivas que optimizan estos procesos.

Un estudio relevante fue el desarrolló un modelo predictivo para optimizar la generación de energía eléctrica en un complejo hidroeléctrico en Ecuador (Monteros, 2022). Este trabajo demuestra cómo los modelos de recomendación pueden mejorar la eficiencia operativa al anticipar con precisión las necesidades de recursos, lo que resalta su aplicabilidad en diversas industrias. La relevancia de este caso radica en que ilustra cómo los modelos predictivos pueden ser utilizados para resolver problemas relacionados con la optimización de recursos, un enfoque que la UNAD podría considerar para mejorar su gestión de compras.

En el ámbito de la salud, investigaciones recientes han implementado modelos predictivos para optimizar la gestión de recursos en la telefarmacia. Un estudio realizado destaca cómo la inteligencia empresarial aplicada al análisis de datos permite mejorar la toma de decisiones clínicas y administrativas en hospitales (González Pérez et al., 2022). Esta investigación subraya la versatilidad de los modelos predictivos y su capacidad para adaptarse a diferentes contextos, incluidos los sistemas de compras y gestión de inventarios, lo que puede ser de gran utilidad para la UNAD en su búsqueda de una gestión más eficiente.

El estudio de Kourentzes, Petropoulos y Fildes (2014) ofrece una revisión sistemática sobre la previsión de la demanda para múltiples productos, enfatizando metodologías que podrían aplicarse en la UNAD para anticipar las necesidades de bienes de consumo en sus centros regionales. Esto podría optimizar el proceso de compras y reducir las ineficiencias actuales en la recopilación de datos.

Por otro lado, el trabajo de Zhang y Wang (2017) sobre el análisis de la previsión de la demanda mejora la gestión de inventarios. Al adoptar enfoques similares, la UNAD podría optimizar su planificación de compras, minimizando las demoras generadas por la recopilación manual de información desde los centros regionales.

Finalmente, la investigación de Baker y McKenzie (2018) a cerca del uso de análisis predictivo en el sector salud, aunque centrada en un contexto diferente, refleja la importancia de los modelos predictivos en la gestión de recursos y toma de decisiones. La UNAD podría beneficiarse de la implementación de un modelo predictivo que automatice la proyección de compras, basado en el análisis de datos históricos, para anticipar las necesidades futuras de consumo.

En resumen, estos estudios resaltan la importancia de los modelos predictivos en la optimización de la gestión de recursos y la planificación de compras. Al incorporar metodologías y enfoques discutidos en estas investigaciones, la UNAD tiene la oportunidad de resolver sus problemáticas actuales en la recopilación de información y mejorar la eficiencia en la ejecución de su plan de compras.

Marco Conceptual

El marco conceptual de este proyecto se fundamenta en varias áreas clave de conocimiento que son esenciales para comprender y aplicar modelos predictivos en la planificación de compras. A continuación, se describen los conceptos más relevantes para el desarrollo del proyecto:

Modelo Predictivo

Un modelo predictivo es una herramienta analítica que utiliza datos históricos y algoritmos matemáticos para predecir eventos futuros. En este proyecto, se implementará un modelo predictivo basado en técnicas de aprendizaje automático (Machine Learning) con el objetivo de proyectar las futuras necesidades de bienes de consumo en la UNAD. Estos modelos permiten a las organizaciones mejorar la toma de decisiones al anticipar tendencias y patrones en los datos, lo cual es crucial para optimizar la gestión de inventarios. Por ejemplo, la UNAD podría utilizar un modelo predictivo para determinar cuándo y cuánto material de oficina necesita, evitando así tanto el desabastecimiento como el exceso de stock.

Machine Learning

El Machine Learning es un subcampo de la inteligencia artificial que se centra en el desarrollo de algoritmos capaces de aprender a partir de datos y realizar predicciones o tomar decisiones sin necesidad de una programación explícita. En este proyecto, se utilizará Machine Learning para entrenar el modelo predictivo que permitirá a la UNAD anticipar sus necesidades de bienes de consumo. Los algoritmos más comunes en este tipo de aplicaciones incluyen la regresión lineal, la regresión logística y los árboles de decisión. Estos algoritmos serán evaluados mediante métricas como el coeficiente de determinación (R^2) y el error cuadrático medio (RMSE). El R^2 indica la proporción de la varianza en la variable dependiente que es predecible a

partir de la variable independiente, mientras que el RMSE proporciona una medida de la precisión del modelo, indicando la desviación típica de los errores de predicción. Este enfoque garantiza que el modelo se ajuste de manera efectiva a los datos históricos de consumo de la UNAD.

Análisis de Datos

El análisis de datos es el proceso de inspección, limpieza, transformación y modelado de datos con el objetivo de descubrir información útil que apoye la toma de decisiones. Para el desarrollo del modelo predictivo, el análisis de datos será fundamental en la preparación y preprocesamiento de la información histórica de consumo. Este análisis permitirá identificar patrones en los datos que serán clave para las proyecciones futuras. Por ejemplo, al analizar el historial de compras de materiales didácticos, la UNAD podría identificar períodos de alta demanda, como el inicio de cada semestre, lo que les permitiría planificar adecuadamente sus compras.

Optimización Logística

La optimización logística se refiere a la mejora de procesos relacionados con la gestión y distribución de recursos, buscando reducir costos y aumentar la eficiencia. En el caso de la UNAD, una logística optimizada permitirá una mejor administración de los bienes de consumo, asegurando que lleguen a los centros regionales de manera oportuna y en las cantidades adecuadas, evitando tanto la escasez como el exceso de inventario. Por ejemplo, mediante la implementación de un modelo predictivo, la UNAD podría determinar la mejor ruta y el mejor momento para distribuir materiales a sus diferentes centros regionales, reduciendo así los costos de transporte y mejorando la eficiencia operativa.

Planificación de Compras

La planificación de compras es el proceso mediante el cual una organización determina qué bienes y servicios necesita adquirir en un período futuro. Con un modelo predictivo adecuado, la UNAD podrá automatizar esta planificación, anticipando sus necesidades de bienes de consumo en base a datos históricos y patrones identificados. Esto resultará en una gestión más eficiente y precisa de sus recursos. Por ejemplo, si se identifica que hay un aumento en la matrícula de estudiantes, la UNAD podrá ajustar sus compras de materiales y suministros en consecuencia, asegurando que todos los estudiantes tengan acceso a los recursos necesarios.

Marco Normativo

El marco normativo de este proyecto se basa en un conjunto de leyes, regulaciones y normativas internas que guían el desarrollo y la implementación de sistemas de gestión y optimización de recursos en el contexto de la Universidad Nacional Abierta y a Distancia (UNAD). Este marco es esencial para asegurar que el proyecto se ejecute en cumplimiento con los estándares legales, éticos y administrativos pertinentes. A continuación, se detallan las normativas más relevantes para el desarrollo de este trabajo de grado:

Estatuto General de la UNAD

El Estatuto General de la UNAD establece las directrices institucionales bajo las cuales se deben regir todas las actividades académicas y administrativas dentro de la universidad. En relación con este proyecto, el estatuto establece los principios de eficiencia y transparencia en la gestión de los recursos, los cuales son fundamentales para la planificación de compras y la administración de bienes de consumo en los diferentes centros regionales. La implementación de un modelo predictivo busca alinearse con estos principios al optimizar el uso de los recursos y mejorar la toma de decisiones dentro de la institución.

Ley 80 de 1993 - Estatuto General de Contratación de la Administración Pública

La Ley 80 de 1993 regula la contratación pública en Colombia, y en el contexto de la UNAD, establece las normas para la adquisición de bienes y servicios. Dado que este proyecto impacta directamente en la planificación de compras de bienes de consumo, es fundamental que las recomendaciones y predicciones generadas por el modelo predictivo se ajusten a las disposiciones de esta ley. El sistema debe respetar los principios de economía, responsabilidad y transparencia en la contratación pública, garantizando que los procesos de adquisición de la universidad cumplan con los requerimientos legales.

Ley 1266 de 2008 - Ley de Habeas Data

La Ley 1266 de 2008 regula el manejo de la información personal en Colombia, garantizando la protección de datos sensibles. Si bien el proyecto no involucra directamente datos personales de los estudiantes o empleados de la UNAD, es importante tener en cuenta esta normativa al trabajar con datos que pudieran estar relacionados con los procesos administrativos y de compras. El análisis de datos y el desarrollo del modelo predictivo deben garantizar la confidencialidad y protección de cualquier información manejada durante el proyecto, respetando los derechos de privacidad.

Decreto 1082 de 2015 - Régimen de Contratación Estatal

El Decreto 1082 de 2015 compila y reglamenta los procesos de contratación estatal, estableciendo disposiciones detalladas sobre la forma en que las entidades públicas deben adquirir bienes y servicios. En este sentido, la implementación del modelo predictivo debe ajustarse a las reglas establecidas en el decreto, particularmente en lo que respecta a la planificación y ejecución de compras. El proyecto debe considerar los requisitos de transparencia y competencia establecidos en este decreto, garantizando que las predicciones de compras se utilicen de manera adecuada en los procesos de adquisición de la UNAD.

Normativa sobre Protección del Medio Ambiente y Uso Sostenible de Recursos

En el contexto actual, las instituciones públicas, incluidas las universidades, están cada vez más comprometidas con la sostenibilidad y el uso responsable de los recursos. La Política Nacional de Producción y Consumo Sostenible de Colombia, junto con las políticas internas de la UNAD, promueve la reducción de desperdicios y el uso eficiente de los recursos. El proyecto debe alinearse con estos objetivos, optimizando la planificación de compras no solo en términos de eficiencia económica, sino también en términos de sostenibilidad, minimizando el impacto

ambiental a través de una gestión adecuada de los inventarios y recursos consumidos por la universidad.

Normativa de Propiedad Intelectual de la UNAD

El Acuerdo 06 de 2008, que regula la propiedad intelectual en la UNAD, establece los derechos y deberes de los estudiantes en relación con los productos generados en el marco de proyectos de grado y otras actividades académicas. En este proyecto, es importante asegurar que todos los componentes del modelo predictivo, así como los informes generados, cumplan con las disposiciones sobre propiedad intelectual. Los derechos de uso de estos productos estarán sujetos a las normativas internas de la UNAD, lo que garantizará su correcta aplicación y utilización dentro del entorno institucional.

Metodología

Fase 1 Recolección y Preprocesamiento de los Datos

Recolección de Datos

Recopilación de datos históricos de despachos de bienes de consumo a los diferentes centros regionales y dependencias de la UNAD.

Recopilación de datos de los centros de costos asociados a los centros regionales y dependencias de la UNAD.

Recopilación de datos de los productos asociados a los bienes de consumo utilizados en la UNAD.

Recopilación de datos de las zonas geográficas asociadas a los centros regionales y dependencias de la UNAD.

Limpieza y Preprocesamiento

Unificación de datos utilizando las diferentes bases de datos obtenidas en el proceso de recolección de datos.

Limpieza de datos para eliminar inconsistencias, registros duplicados y registros nulos.

Tipificación de las variables según los datos que contienen.

Reasignación de nombres a las variables para su adaptación al análisis.

Fase 2 Análisis de Datos

Análisis Descriptivo

Calcular dimensión de la data (Cantidad de observaciones y variables).

Calcular medidas descriptivas como la media, mediana, desviación estándar y percentiles para las variables cuantitativas.

Calcular medidas descriptivas como la frecuencia y los datos únicos para las variables categóricas.

Utilizar gráficos como histogramas y boxplots para mayor entendimiento de los datos.

Análisis Exploratorio

Análisis general de despachos para identificar tendencias y cantidades de bienes despachados.

Análisis financiero para identificar la evolución de los costos, diferenciación de costos y diferenciación de gastos.

Análisis de la distribución y logística para identificar productos de alta rotación y de menor rotación.

Fase 3 Análisis Predictivo

Estudio de Variables

Identificación de variables relevantes para un modelo predictivo.

Estudio para la implementación de un modelo predictivo:

Uso de algoritmos de *machine learning* como la regresión lineal o árboles de decisión para entrenar un modelo predictivo con los datos históricos, dado que este tipo de modelos permiten predecir valores cuantitativos.

Evaluación del rendimiento del modelo mediante métricas como RMSE y R^2 ,

Optimización

Ajuste y optimización del modelo para mejorar su precisión.

Fase 4 Diseño de reportes

Frecuencia de Reportes

Propuesta de periodicidad para la generación de reportes globales y específicos.

Estructuración de Reportes

Identificación de información relevante y pertinente para el diseño de los reportes.

Estructuración de la información contenida en los reportes.

Visualización de información en los reportes.

Recopilación

Exportación de plantilla para socialización de reportes.

Esta metodología permitirá entender los datos de despachos de consumo de la UNAD, optimizar su gestión de compras de bienes de consumo, basándose en análisis estadísticos y predictivos, mejorando así la eficiencia y efectividad de sus operaciones logísticas, así como, un uso de los recursos eficiente.

Tipo de Estudio

El proyecto se enmarca en un estudio aplicado, dado que su propósito es resolver una problemática específica de la UNAD relacionada con la gestión de compras de bienes de consumo. Además, es de tipo cuantitativo ya que se basa en el análisis de datos históricos, análisis predictivo y la validación de resultados numéricos. Se utilizará un enfoque de desarrollo tecnológico, en el cual se aplican herramientas analíticas y técnicas de *Machine Learning* para mejorar procesos logísticos y administrativos dentro de la universidad.

Este tipo de estudio no solo busca generar conocimiento sobre el comportamiento del consumo de bienes en los centros regionales de la UNAD, sino también aplicar dicho conocimiento para optimizar procesos y mejorar la eficiencia institucional.

Recolección de Datos

La recolección de datos para este proyecto se centrará en la obtención de los registros históricos de despachos de bienes de consumo a los centros regionales y dependencias de la UNAD. Estos datos incluyen:

Tipo de bien de consumo (por ejemplo, papel, tóner, etc.).

Cantidad de bienes consumidos en cada centro regional.

Costos asociados a cada adquisición.

Datos temporales que permitan identificar patrones de estacionalidad (meses de mayor o menor consumo).

Adicionalmente se obtendrán los datos de los productos asociados a los bienes de consumo y los datos de los centros de costos asociados a los centros regionales y dependencias de la UNAD.

La recolección de datos se llevará a cabo en colaboración con el Grupo Funcional de Almacén e Inventarios de la UNAD, asegurando la fiabilidad y representatividad de los datos obtenidos. Una vez recolectados, los datos pasarán por un proceso de limpieza y normalización, eliminando registros erróneos o duplicados, para garantizar su calidad y consistencia antes de ser utilizados en el desarrollo del modelo predictivo.

El análisis de estos datos permitirá no solo construir un modelo que anticipe las futuras necesidades de consumo, sino también identificar patrones de comportamiento que informen la planificación de compras de la universidad en los años venideros.

Resultados

Con el propósito de dar cumplimiento a los objetivos propuestos y posterior a la recolección y limpieza de datos realizada, en la cual se realizó la unificación de datos, depuración de datos inconsistentes, tipificación y reasignación de nombres a las variables. Se inicia un análisis descriptivo con el fin de entender la distribución de la *data*.

Análisis Descriptivo

Variables Cuantitativas

Figura 1

Información Contendida en la Data

BODEGA	NUMERO	FECHA	TIPO	REGLON	PRODUCTO	NOMB_PRODUCTO	CANTIDAD	COSTO	CENTRO_COSTO	NOMB_CENTRO_COSTO	ZONA	
0	114710	8325	2019-04-12	S	8	11000027	COSEDORA PARA GRAPA ESTANDAR	3	10640.00	101000	DESPACHO SISTEMA DE EDUCACION PERMANENTE	SEDE NACIONAL
1	114710	8325	2019-04-12	S	9	11000036	FECHADOR	1	9170.00	101000	DESPACHO SISTEMA DE EDUCACION PERMANENTE	SEDE NACIONAL
2	114710	8325	2019-04-12	S	10	11000037	GANCHO TIPO CLIP ESTANDAR	5	421.98	101000	DESPACHO SISTEMA DE EDUCACION PERMANENTE	SEDE NACIONAL
3	114710	8325	2019-04-12	S	11	11000039	GANCHO TIPO GRAPA PARA COSEDORA ESTANDAR	5	1708.76	101000	DESPACHO SISTEMA DE EDUCACION PERMANENTE	SEDE NACIONAL
4	114710	8325	2019-04-12	S	12	11000045	LAPIZ MINA NEGRA	24	413.00	101000	DESPACHO SISTEMA DE EDUCACION PERMANENTE	SEDE NACIONAL

En la *data* encontramos un total de 12 variables, las cuales se describen a continuación:

1. Bodega

Descripción: Esta variable representa el número de identificación de la bodega en la cual se almacena la información de los bienes de consumo.

Tipo: Cuantitativa (discreta).

Contexto: La identificación de bodegas es crucial para la gestión logística, ya que permite a la UNAD organizar, cuantificar y registrar los movimientos de los bienes de consumo.

2. Numero

Descripción: Este número representa el identificador único del movimiento de inventario en una bodega, ya sea una entrada, salida o traslado de bienes.

Tipo: Cuantitativa (discreta).

Contexto: Almacenar y rastrear los movimientos de inventario es fundamental para mantener un registro preciso de las transacciones.

3. Fecha

Descripción: La fecha en que se realiza el movimiento de inventario.

Tipo: Cuantitativa (continua) o cualitativa (dependiendo del contexto).

Contexto: Esta variable es esencial para el análisis temporal de los movimientos de inventario y para evaluar la frecuencia de los movimientos.

4. Tipo

Descripción: Indica el tipo de movimiento de inventario, que puede ser una salida (distribución de bienes) o un traslado (movimiento de bienes entre bodegas).

Tipo: Cualitativa (nominal).

Contexto: Comprender el tipo de movimiento es importante para la gestión logística de la UNAD.

5. Renglón

Descripción: Este número se refiere al renglón específico dentro de un movimiento de inventario, que puede incluir varios productos en una misma transacción.

Tipo: Cuantitativa (discreta).

Contexto: Permite desglosar un movimiento en varios productos, facilitando la identificación de los más demandados.

6. Producto

Descripción: Código que identifica de manera única cada producto almacenado en las bodegas de la UNAD.

Tipo: Cuantitativa (discreta).

Contexto: La identificación de productos es fundamental para la gestión del inventario y para realizar análisis informados sobre las compras y reposiciones.

7. Nomb_producto

Descripción: Nombre asociado al código que identifica de manera única cada producto almacenado en las bodegas de la UNAD.

Tipo: Cualitativa (nominal).

Contexto: La identificación de productos es fundamental para la gestión del inventario y para realizar análisis informados sobre las compras y reposiciones.

8. Cantidad

Descripción: Indica la cantidad de producto involucrado en un movimiento de inventario.

Tipo: Cuantitativa (discreta).

Contexto: Esta variable es crucial para la planificación de compras, ya que proporciona datos sobre el uso de bienes en un periodo determinado.

9. Costo

Descripción: Costo por unidad del producto, que se utiliza para calcular el valor total de las transacciones de inventario.

Tipo: Cuantitativa (continua).

Contexto: Comprender el costo unitario es esencial para la gestión financiera de la UNAD.

10. Centro_costo

Descripción: Identificador del centro de costo al que se destina el movimiento de bienes.

Tipo: Cuantitativa (discreta).

Contexto: Esta variable es importante para la asignación de costos y para evaluar el gasto en diferentes áreas de la UNAD.

11. Nomb_centro_costo

Descripción: Nombre asociado al identificador del centro de costo al que se destina el movimiento de bienes.

Tipo: Cualitativa (nominal).

Contexto: Esta variable es importante para la asignación de costos y para evaluar el gasto en diferentes áreas de la UNAD.

12. Zona

Descripción: Indica la región a la que pertenece un centro de costo, según la distribución y organización de la UNAD en sus diferentes sedes a nivel nacional.

Tipo: Cualitativa (nominal).

Contexto: Esta variable es importante para la asignación de costos y para evaluar el gasto en diferentes zonas de la UNAD.

El dataset tiene un total de 16.721 registros y presentan una variabilidad considerable en las variables relacionadas con los productos, los costos y los centros de costo. Lo mencionado se puede observar en las estadísticas descriptivas de la figura 2, relacionada a continuación.

Figura 2*Estadísticas Descriptivas de la Data en las Variables Cuantitativas*

	BODEGA	NUMERO	REGLON	PRODUCTO	CANTIDAD	COSTO	CENTRO_COSTO
count	16721.0	16721.000000	16721.000000	1.672100e+04	16721.000000	1.672100e+04	1.672100e+04
mean	114710.0	8964.960469	12.031996	1.111247e+07	15.577418	1.240550e+05	3.544179e+09
std	0.0	417.486999	9.157911	2.063992e+05	87.417508	3.330521e+05	2.275294e+09
min	114710.0	8308.000000	1.000000	1.100000e+07	1.000000	7.700000e+01	2.000000e+02
25%	114710.0	8632.000000	5.000000	1.100004e+07	2.000000	8.320000e+02	1.006000e+05
50%	114710.0	8908.000000	10.000000	1.100008e+07	5.000000	2.400000e+03	5.003220e+09
75%	114710.0	9290.000000	17.000000	1.100024e+07	12.000000	9.230000e+03	5.005410e+09
max	114710.0	9774.000000	70.000000	1.151337e+07	7576.000000	3.022100e+06	5.008640e+09

De acuerdo con las estadísticas descriptivas de las variables cuantitativas se establece que:

1. Bodega

Promedio: 114710,0

Desviación Estándar: 0,0

Mínimo y Máximo: Ambos son 114710

Interpretación: Esta variable no tiene variabilidad en el dataset. Ya que todos los registros pertenecen a una única bodega de la cual se realizan los despachos.

2. Numero

Promedio: 8964,96

Desviación Estándar: 417,49

Mínimo: 8308

Máximo: 9774

Interpretación: Los números de despacho varían entre 8308 y 9774. Esto indica que los registros están agrupados en un rango consecutivo, representando el orden secuencial de los despachos.

3. Renglón

Promedio: 12,03

Desviación Estándar: 9,16

Mínimo: 1

Máximo: 71

Interpretación: Los renglones representan líneas individuales dentro de un despacho. Con un promedio de 12, parece que los despachos suelen contener múltiples productos o ítems. El rango de hasta 71 sugiere que algunos despachos son significativamente mayores.

4. Producto

Promedio: 11112470

Desviación Estándar: 206399,2

Mínimo: 11000000

Máximo: 11494000

Interpretación: Los productos están codificados con números en un rango amplio. Esto debido a que estos valores podrían representar identificadores de los productos.

5. Cantidad

Promedio: 15,58

Desviación Estándar: 87,42

Mínimo: 1

Máximo: 2400

Interpretación: Aunque la cantidad promedio por producto es de 15,58 la desviación estándar alta y el máximo de 2400 indican que hay despachos con cantidades excepcionalmente grandes.

6. Costo

Promedio: 124054,95

Desviación Estándar: 333052,10

Mínimo: 77

Máximo: 8998800

Interpretación: El costo promedio es de 124.054, pero existe una gran dispersión en los datos. La presencia de costos muy elevados (hasta 8.998.800) implica una alta variabilidad entre los productos despachados, reflejando los diferentes tipos y cantidades de productos.

7. Centro_costo

Promedio: 3.544.179.000

Desviación Estándar: 2.275.294.000

Mínimo: 200

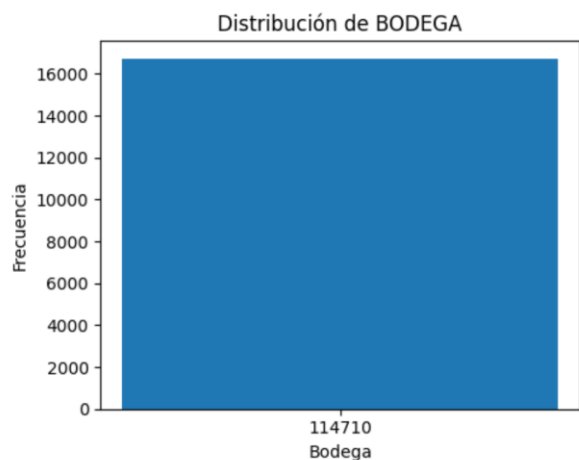
Máximo: 10.056.000.000

Interpretación: Los códigos de centros de costo tienen una amplia variabilidad, desde valores pequeños hasta grandes, lo que representa la amplia categorización de las sedes y dependencias a las que se distribuyen los bienes de consumo.

Con el fin de ampliar las estadísticas descriptivas se realizaron graficas con las que se pueden obtener información complementaria y se relacionan a continuación:

Figura 3

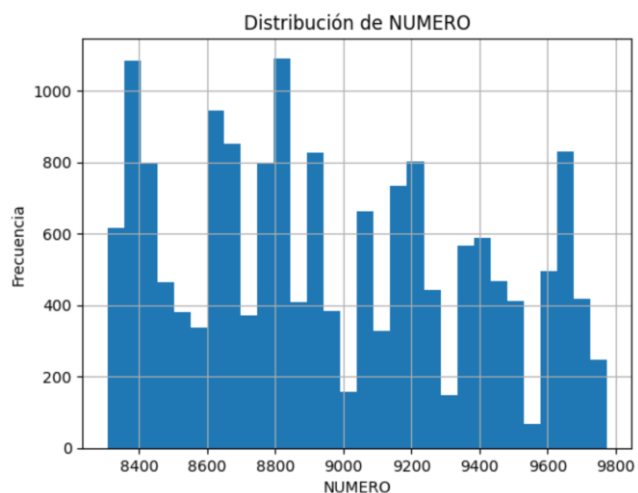
Histograma de la Distribución de la Variable Bodega



De acuerdo con el histograma se evidencia que una sola bodega (identificada con el código 114710) concentra la totalidad de los movimientos registrados en el dataset. Lo que podría indicar que esta bodega actúa como un punto central o principal en la distribución de bienes de consumo dentro del sistema logístico de la UNAD.

Figura 4

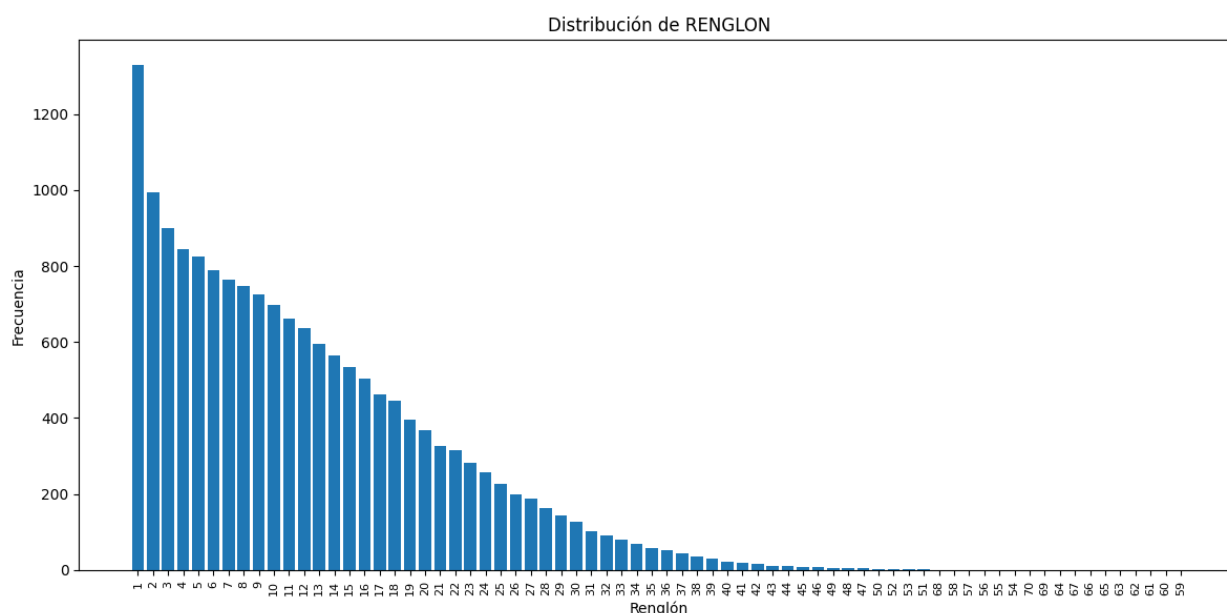
Histograma de la Distribución de la Variable Número



De acuerdo con el histograma se evidencia que, los despachos de bienes consumo tienen una frecuencia variable, con algunos números asociados a un mayor número de registros. Lo indica que ciertos movimientos representan transacciones más complejas o agrupadas (por ejemplo, involucrando múltiples productos o renglones). Algunos números, tienen frecuencias significativamente más altas en comparación con otros. Lo que podría estar relacionado con actividades operativas más intensas o relevantes.

Figura 5

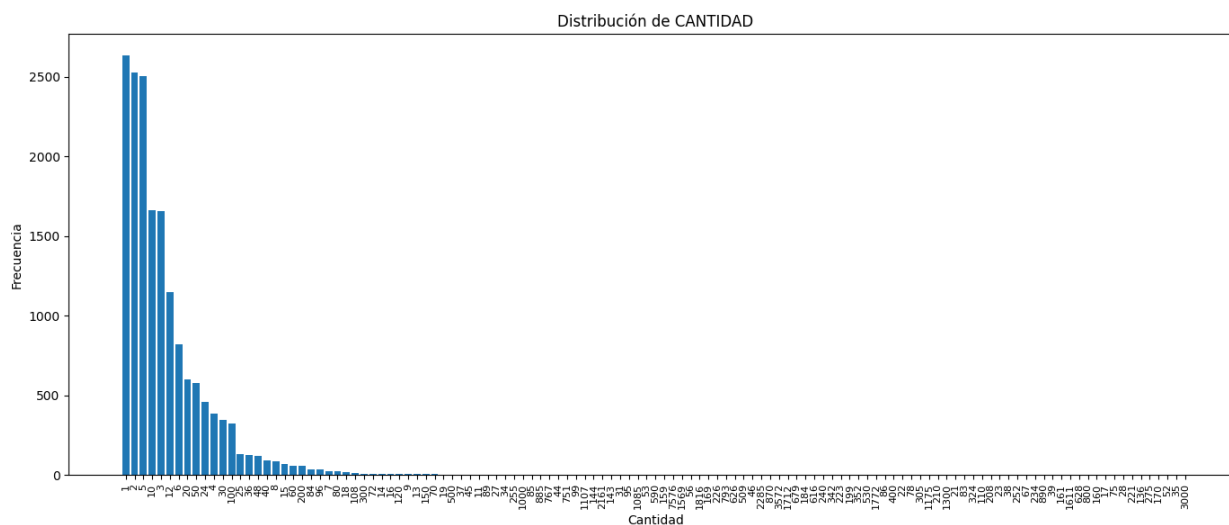
Histograma de la Distribución de la Variable Renglón



De acuerdo con el histograma se evidencia que, los renglones con valores bajos (1, 2, 3, etc.) concentran la mayor cantidad de registros. Esto indica que la mayoría de los despachos de consumo involucran un número reducido de renglones por movimiento. A medida que aumenta el número de renglones, la frecuencia disminuye. Lo que sugiere que los movimientos con múltiples renglones son menos comunes.

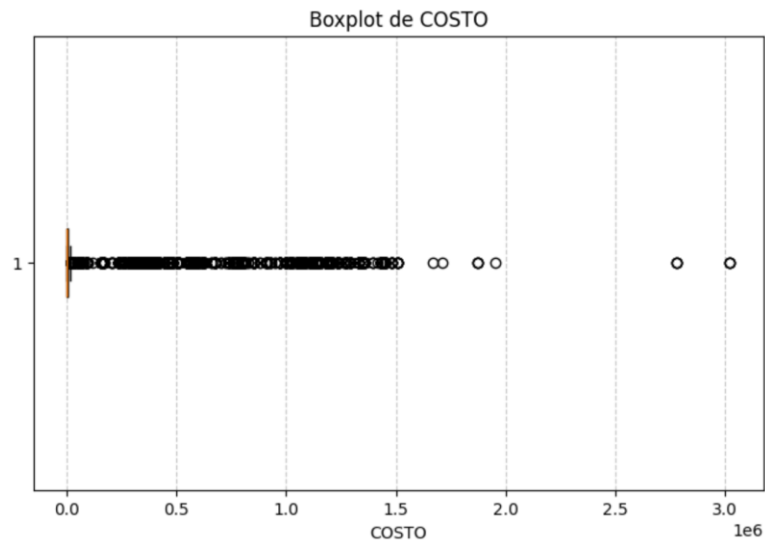
Figura 7

Histograma de la Distribución de la Variable Cantidad



De acuerdo con el histograma se evidencia que, los valores de cantidad bajos (por ejemplo, 1, 2, 5) son los más frecuentes. Lo que sugiere que la mayoría de los movimientos de despachos involucran cantidades reducidas de productos, probablemente despachos específicos para satisfacer necesidades puntuales. A medida que aumenta la cantidad, la frecuencia de los registros disminuye. Lo que indica que los movimientos con grandes cantidades son menos comunes.

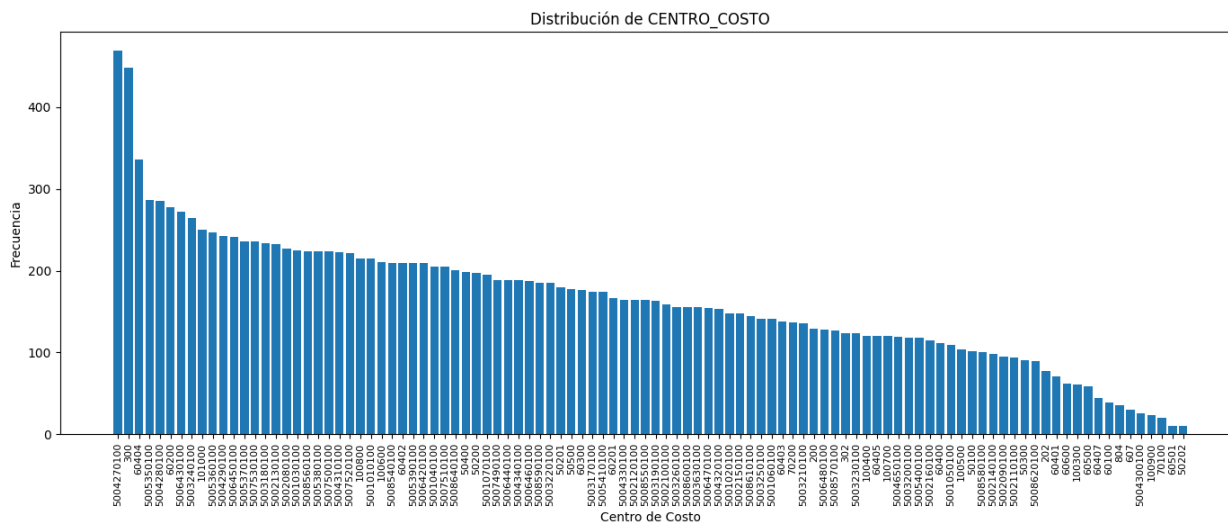
Aunque mucho menos frecuentes, existen registros con cantidades muy altas que representan despachos más grandes. Estos casos podrían estar relacionados con operaciones o necesidades especiales.

Figura 8*Boxplot de la Distribución de la Variable Costo*

De acuerdo con el boxplot se evidencia que, la mayor parte de los valores de costo están muy concentrados cerca del extremo inferior del rango, lo que sugiere que la mayoría de los despachos tienen costos relativamente bajos. Existe un número significativo de valores atípicos, con costos que se extienden hasta cifras mucho más altas. Lo que podría representar despachos inusuales o productos de alto costo.

Figura 9

Histograma de la Distribución de la Variable Centro_Costo



De acuerdo con el histograma se evidencia que, algunos centros de costo tienen frecuencias significativamente más altas que otros. Lo que sugiere que ciertos centros concentran la mayoría de los despachos de consumo. Por ejemplo, los centros con mayor frecuencia probablemente representen áreas de mayor actividad operativa o menor eficiencia de recursos.

La frecuencia disminuye rápidamente después de los primeros centros más activos, lo que indica que la mayoría de los centros tienen menos despachos registrados. Lo que podría deberse a la naturaleza operativa de los centros o mayor eficiencia de los recursos.

Variables Categóricas

En cuanto a las variables categóricas encontramos un total de 4 variables: tipo, nomb_producto, nomb_centro_costo y zona y sus estadísticas descriptivas se relacionan a continuación en la figura 10.

Figura 10

Estadísticas Descriptivas de las Variables Categóricas

	Datos unicos	Categoría mas frecuente	Frecuencia de la categoría mas frecuente	Categorías menos frecuentes
TIPO	2	S	14706	{'T': 2015, 'S': 14706}
NOMB_PRODUCTO	105	PAPEL BOND TAMAÑO CARTA	708	{'CINTA TERMICA': 1, 'CINTA DE EMPAQUE LOGO UN...
NOMB_CENTRO_COSTO	102	CEAD JOSE ACEVEDO Y GOMEZ ADMINISTRACION GENERAL	469	{'ATENCION AL USUARIO': 10, 'COORDINACION NACI...
ZONA	9	SEDE NACIONAL	4880	{'OCCIDENTE': 1076, 'AMAZONIA ORINOQUIA': 1238...

A continuación, se describen las variables:

1. Tipo:

Categorías únicas: 2 (s (salida de almacén), t (traslado de almacén)).

Categoría más frecuente: s con 14.706 ocurrencias.

Categorías menos frecuentes: t con 2.015 ocurrencias.

Interpretación: la mayoría de los registros corresponden al tipo s (salida de almacén, lo que indica que este es el tipo predominante de despacho.

2. Nomb_producto:

Categorías únicas: 105.

Categoría más frecuente: papel bond tamaño carta con 708 ocurrencias.

Categorías menos frecuentes: incluye productos como cinta térmica, cinta de empaque logo unad, cada uno con una sola ocurrencia.

Interpretación: existe una alta variedad de productos, pero pocos se despachan con alta frecuencia. Los productos más raros representan artículos especializados o de bajo consumo.

3. Nomb_centro_costo:

Categorías únicas: 102.

Categoría más frecuente: CEAD José Acevedo y Gómez administración general con 469 ocurrencias.

Categorías menos frecuentes: incluyen atención al usuario, coordinación nacional, entre otros, con menos de 10 ocurrencias cada uno.

Interpretación: aunque hay una gran diversidad en los centros de costo, algunos tienen un volumen muy bajo de despachos asociados, mientras que otros como el CEAD JAG mencionado tiene un papel central en las operaciones.

4. Zona:

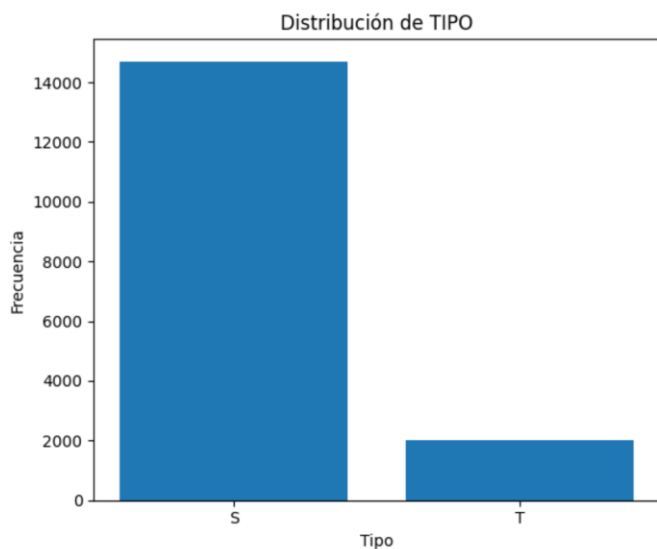
Categorías únicas: 9.

Categoría más frecuente: sede nacional con 4.880 ocurrencias.

Categorías menos frecuentes: incluyen occidente, Amazonia Orinoquia, con 1.076 y 1.238 ocurrencias respectivamente.

Interpretación: la sede nacional domina los registros, indicando que allí se concentra la mayoría de los despachos. Otras zonas tienen una menor participación relativa.

Con el fin de ampliar las estadísticas descriptivas de las variables categóricas se realizaron graficas con las que se pueden obtener más información y se relacionan a continuación:

Figura 11*Histograma de la Distribución de la Variable Tipo*

De acuerdo con el histograma se evidencia que, La mayoría de los registros corresponden al tipo S (salidas), que representa aproximadamente el 88% de los movimientos. Esto indica que la mayor parte de los despachos finales registrados son distribuciones de bienes desde la bodega hacia otros centros o dependencias. El tipo T (traslados) tiene una frecuencia mucho menor, representando aproximadamente el 12% de los registros.

El alto predominio de salidas sugiere que la operación logística de la UNAD está principalmente orientada a despachar los bienes hacia los usuarios finales o centros de costo específicos de forma definitiva.

Figura 12

Histograma de la Distribución de la Variable Nomb_Producto

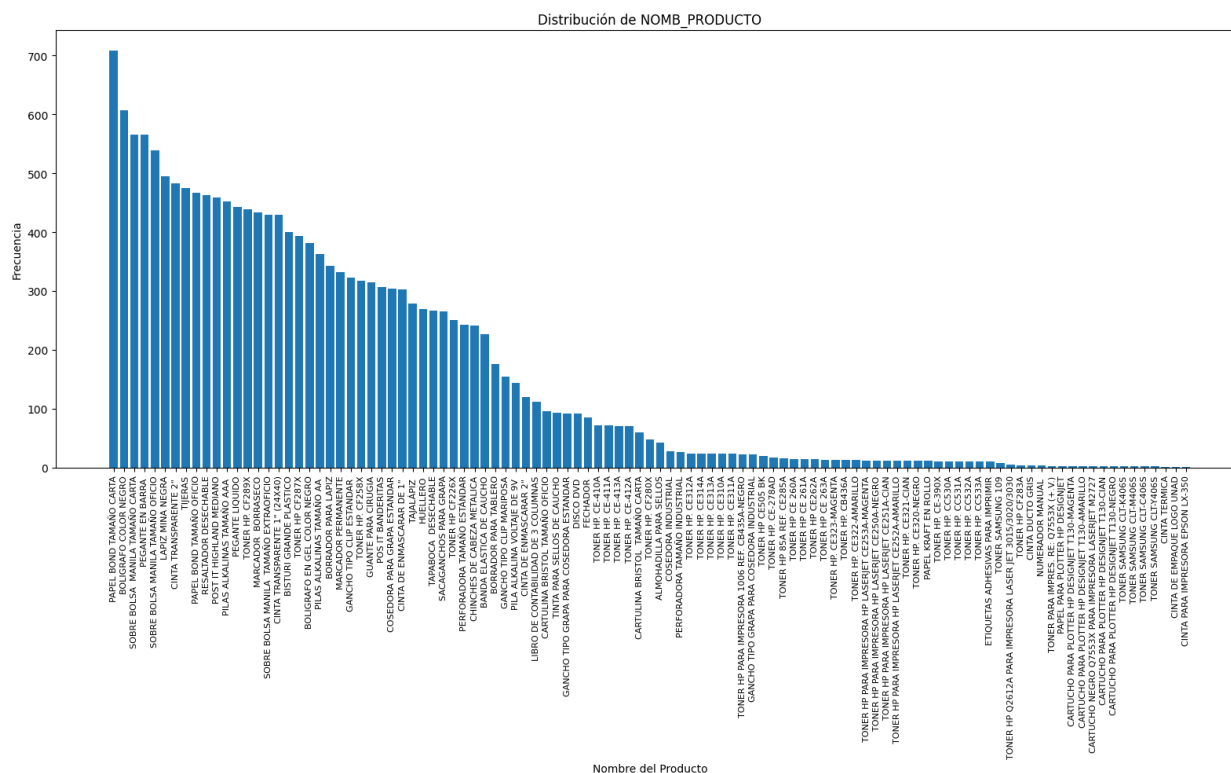
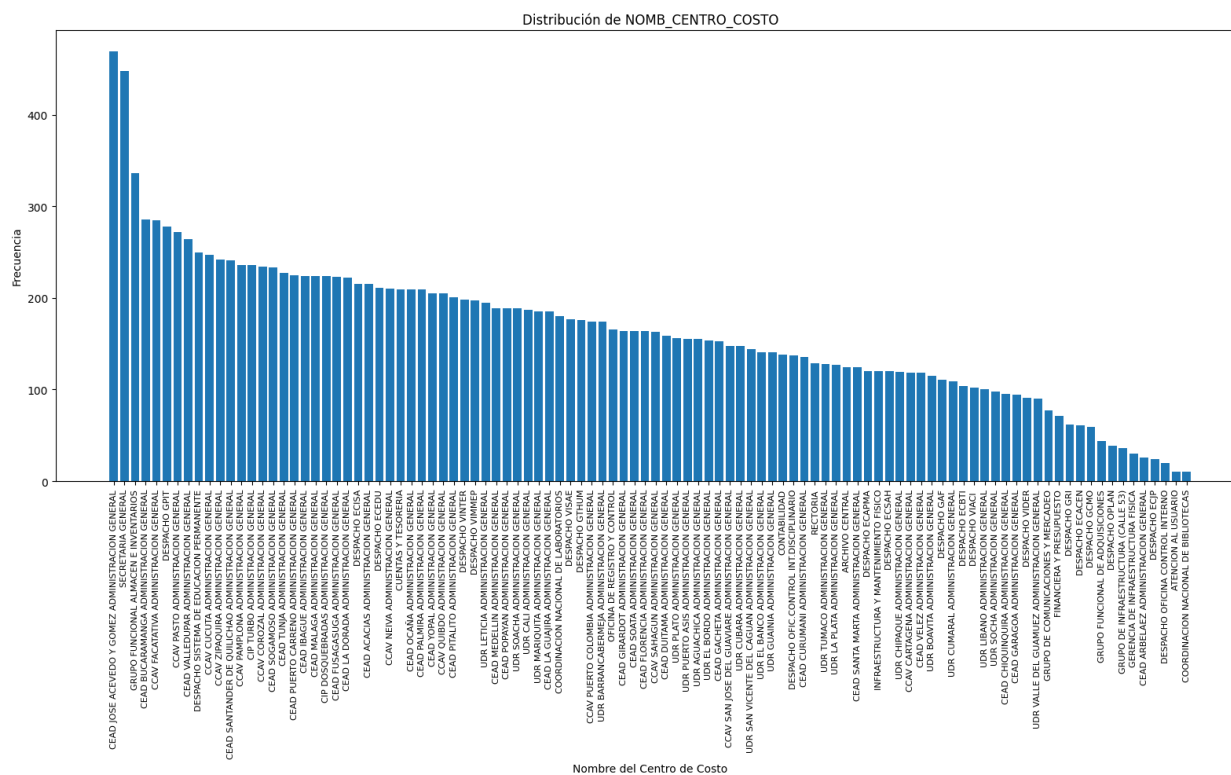


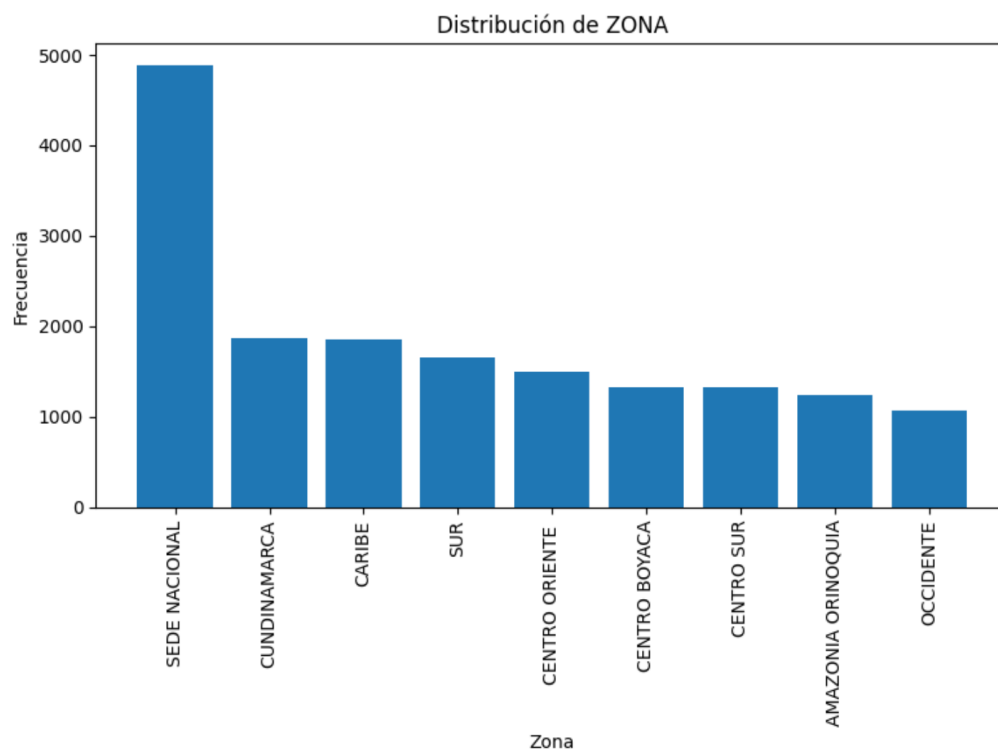
Figura 13

Histograma de la Distribución de la Variable Nomb_Centro_Costo



De acuerdo con el histograma se evidencia que, un pequeño grupo de centros de costo, como CEAD José Acevedo y Gómez administración general, Secretaría General, y Grupo funcional almacén e inventarios, concentra la mayoría de los movimientos registrados. Estos centros probablemente representan las principales áreas administrativas o logísticas de la UNAD.

La mayoría de los demás centros de costo tienen una menor cantidad de registros, lo que indica que su nivel de actividad es menor en comparación con los principales centros. Centros como CCAV Facatativá Administración General y otros similares tienen frecuencias moderadas, lo que podría estar relacionado con necesidades administrativas menos intensivas o mayor eficiencia de los recursos.

Figura 14*Histograma de la Distribución de la Variable Zona*

De acuerdo con el histograma se evidencia que, la sede nacional concentra la mayor cantidad de movimientos registrados, lo que refleja su rol como un nodo central en la gestión y distribución de bienes de consumo en la UNAD. Esto es consistente con un modelo logístico donde una sede centraliza las operaciones principales.

Las zonas Cundinamarca, Caribe, y Sur también presentan frecuencias relativamente altas, lo que indica que estas regiones tienen una participación importante en las operaciones de distribución. Estas frecuencias sugieren una alta demanda, una actividad administrativa más intensa en estas áreas o una menor eficiencia en los recursos.

Zonas como Centro Oriente tienen una menor cantidad de movimientos registrados, lo que podría reflejar una menor demanda o una mayor optimización de los recursos.

Variable Fecha

Finalmente, en el dataset encontramos a la variable fecha y a continuación se relacionan sus estadísticas descriptivas en la figura 15.

Figura 15

Estadísticas Descriptivas de la Variable Fecha

	Metric	Value
0	Min Fecha	2019-04-11 00:00:00
1	Max Fecha	2023-12-29 00:00:00
2	Total Registros	16721
3	Registros Nulos	0

A continuación, se describe la variable.

1. Fecha:

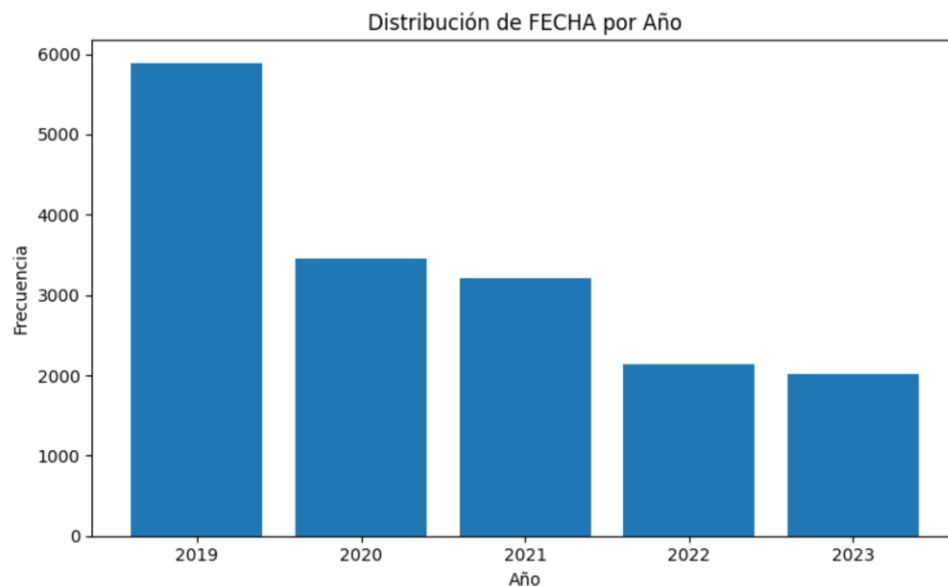
Fecha máxima: 29-diciembre-2023.

Fecha mínima: 11-abril-2019.

Total, Registros: 16.721.

Interpretación: El rango de fechas sugiere que los datos son consistentes y abarcan un periodo suficientemente amplio (aproximadamente 5 años).

Con el fin de ampliar las estadísticas descriptivas se realizaron graficas con las que se pueden obtener más información y se relacionan a continuación:

Figura 16*Histograma de la Distribución de la Variable Fecha por Año*

De acuerdo con el histograma se evidencia que, el año 2019 tiene la mayor cantidad de despachos registrados en comparación con los años posteriores, lo que podría indicar un periodo de alta actividad operativa o una menor eficiencia en los recursos.

A partir de 2020, se observa una disminución progresiva en la cantidad de registros anuales, este descenso podría estar relacionado con varios factores, como:

Cambios en las operaciones logísticas.

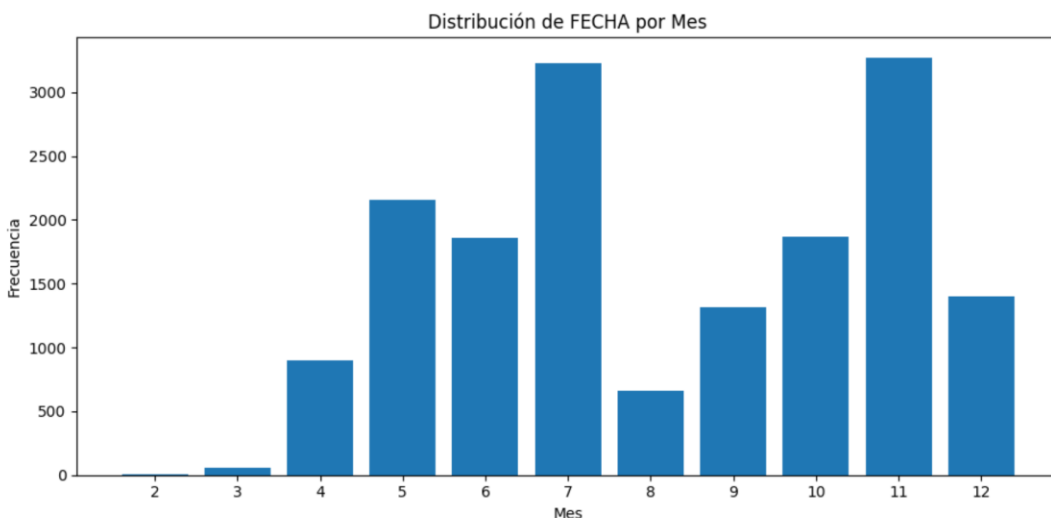
Efectos de la pandemia de COVID-19, que pudo impactar la demanda o las operaciones logísticas de la UNAD.

Mejoras en la eficiencia, con movimientos agrupados o menos frecuentes.

Aunque hay una disminución general, los años 2022 y 2023 mantienen niveles de actividad similares, lo que sugiere una estabilización en los procesos logísticos en este periodo.

Figura 17

Histograma de la Distribución de la Variable Fecha por Mes



De acuerdo con el histograma de la Figura 17 se evidencia que, existen picos significativos en ciertos meses, lo que indica periodos de alta actividad logística. Por ejemplo: mayo y junio son de los meses con mayor cantidad de registros, lo que podría estar asociado a ciclos operativos específicos, como la preparación de semestres académicos o procesos administrativos de impacto.

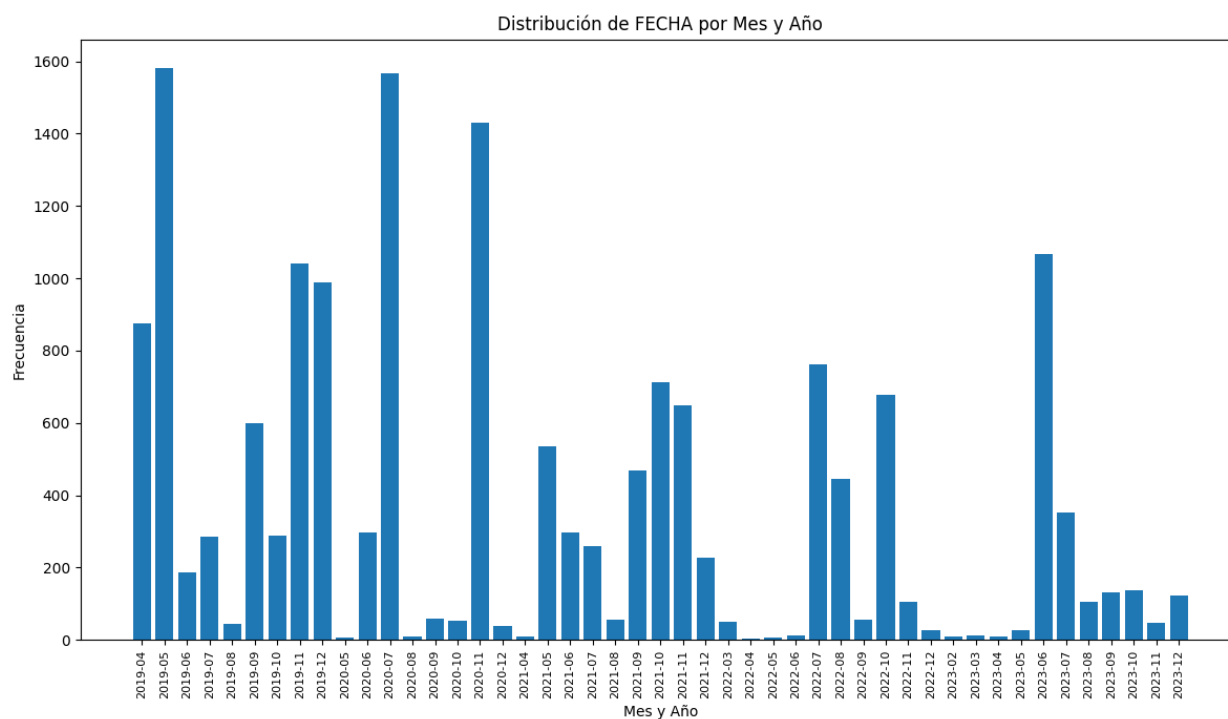
Algunos meses tienen muy pocos registros, como febrero y marzo. Esto podría deberse a:
Menor demanda durante esos periodos.

Actividades más enfocadas en otros procesos no relacionados con inventarios.

La variación entre meses sugiere la presencia de patrones estacionales en las operaciones logísticas de la UNAD, vinculados a eventos anuales como inicio o fin de periodos académicos, despachos programados o proyectos institucionales.

Figura 18

Histograma de la Distribución de la Variable Fecha por Mes y Año



De acuerdo con el histograma se evidencia que, el patrón anual y mensual refleja ciclos logísticos específicos que probablemente coincidan con eventos académicos o administrativos clave en la UNAD. La disminución general en la actividad podría reflejar cambios estructurales, mejoras en la eficiencia o impactos externos como la pandemia.

La planificación logística debería enfocarse en meses de alta actividad, como mayo y junio, asegurando recursos y estrategias para manejar la demanda. Los meses de baja actividad podrían ser aprovechados para realizar inventario.

Análisis Exploratorio

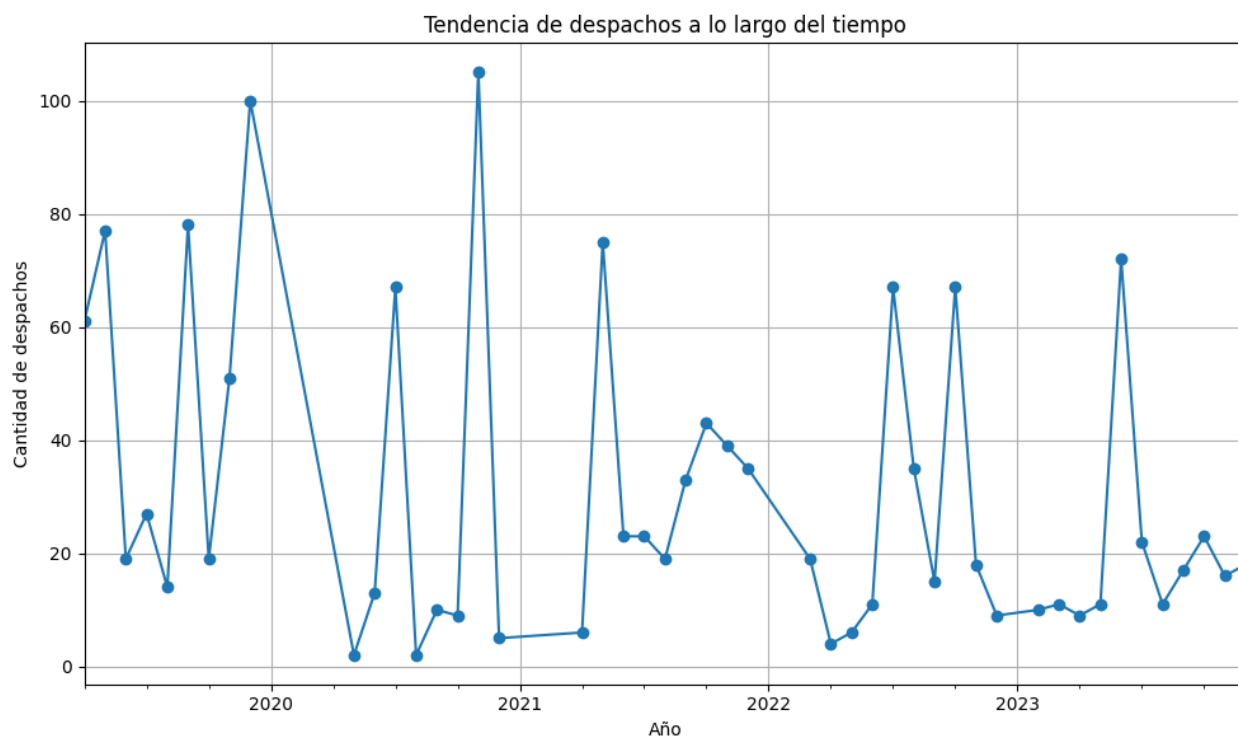
Se inicia un análisis exploratorio que permita identificar información relevante que ayude a la toma de decisiones empresariales.

Análisis General de Despachos

Despachos a lo Largo del Tiempo (2019-2023)

Figura 19

Tendencias de Despachos a lo Largo del Tiempo



De acuerdo con el gráfico se evidencia que, los meses con valores altos pueden coincidir con la preparación de periodos académicos, como el inicio de semestres o cursos, cuando se requiere abastecer centros regionales con materiales de papelería. También podrían estar relacionados con eventos administrativos, como auditorías o cierre de periodos contables, que demandan un mayor uso de recursos.

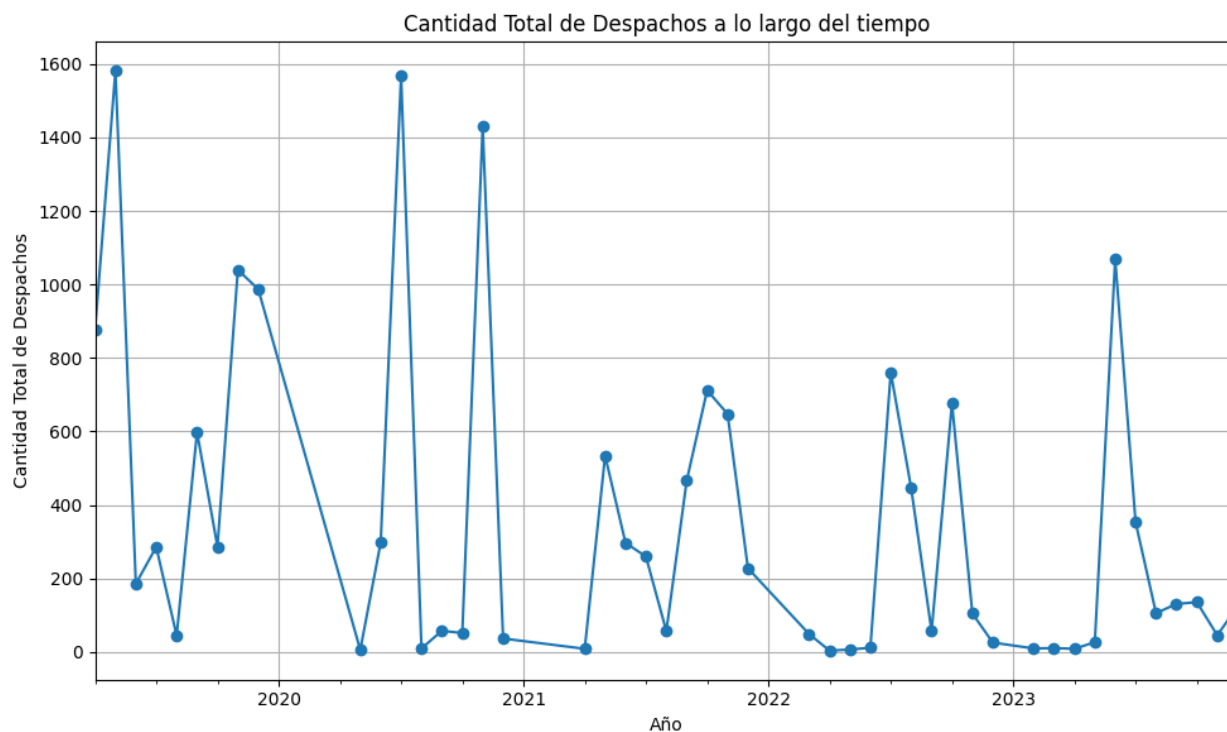
Los meses con menor cantidad de movimientos probablemente coincidan con vacaciones académicas, como diciembre o mediados de año, donde la actividad operativa se reduce.

Es posible que la baja actividad en algunos meses refleje una estabilización en el suministro, tras haberse realizado envíos previos suficientes para satisfacer las necesidades de los centros.

La variabilidad observada sugiere que las operaciones logísticas no son uniformes durante el año, lo que puede implicar desafíos en términos de planificación de recursos. Las fluctuaciones pueden representar esfuerzos logísticos adicionales para manejar picos de demanda.

Figura 20

Cantidad Total de Despachos a lo Largo del Tiempo



De acuerdo con el gráfico se evidencia que, la actividad logística no es uniforme. Existen meses con un volumen considerablemente mayor de despachos y otros con valores más bajos. Estas variaciones probablemente estén influenciadas por eventos operativos o académicos.

Ciertos meses presentan constantemente altos volúmenes de despachos, lo que podría estar relacionado con actividades recurrentes, como inicios de semestres académicos, distribución de materiales o cierres administrativos.

Los meses con mayor cantidad de despachos pueden coincidir con: preparativos para periodos académicos o logísticas específicas. Estos picos representan momentos críticos para la logística de la UNAD, con mayor demanda de recursos y planificación operativa.

Despachos por Zonas

Figura 21

Tendencias de Despachos en la Zona Amazonia a lo Largo del Tiempo



Figura 22

Tendencias de Despachos en la Zona Caribe a lo Largo del Tiempo

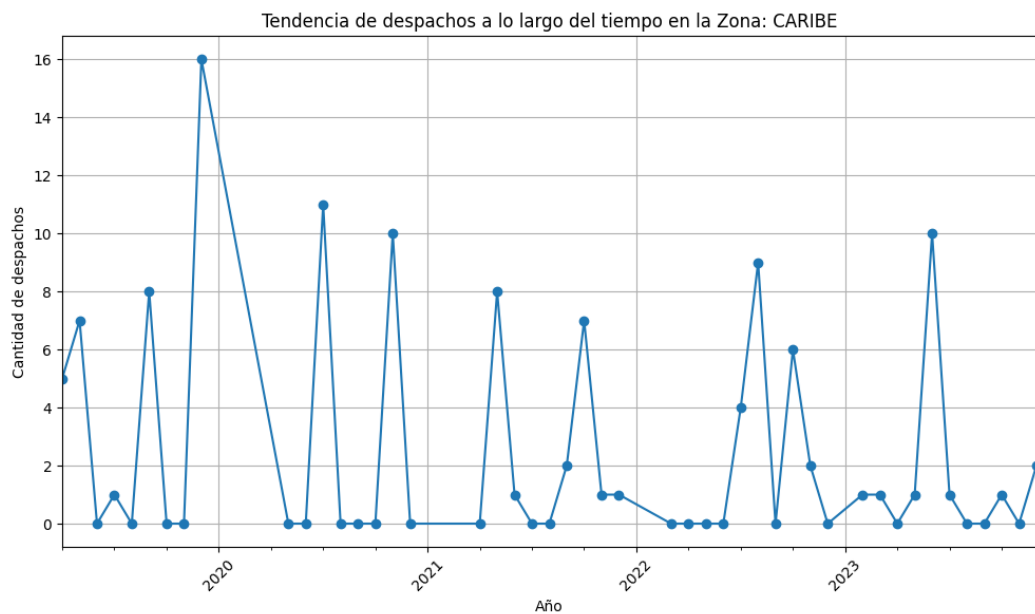


Figura 23

Tendencias de Despachos en la Zona Boyacá a lo Largo del Tiempo

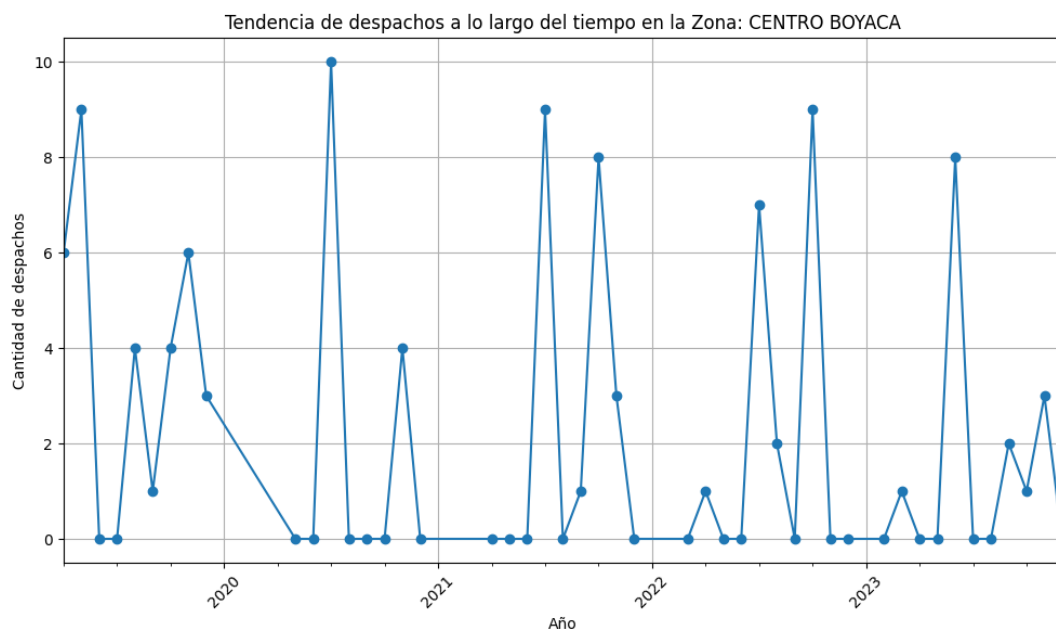


Figura 26

Tendencias de Despachos en la Zona Cundinamarca a lo Largo del Tiempo

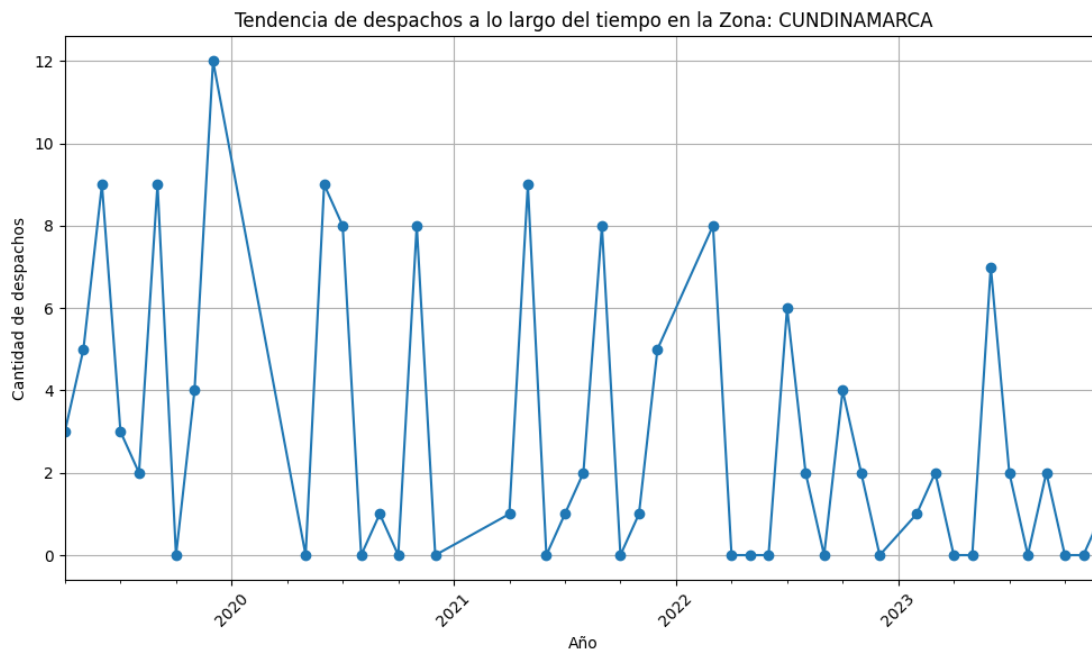


Figura 27

Tendencias de Despachos en la Zona Occidente a lo Largo del Tiempo

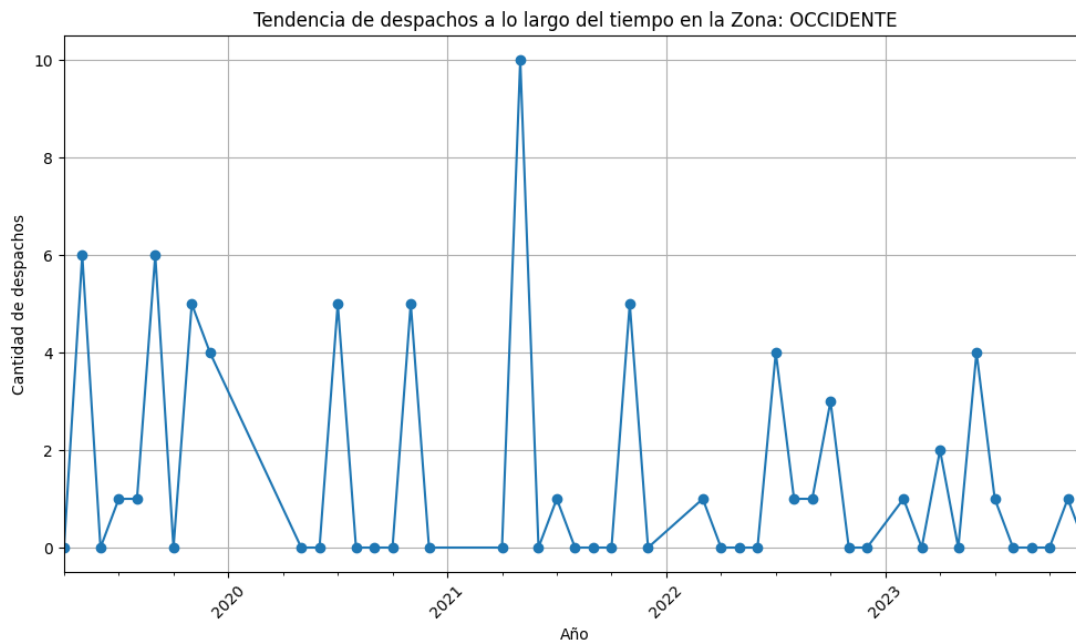
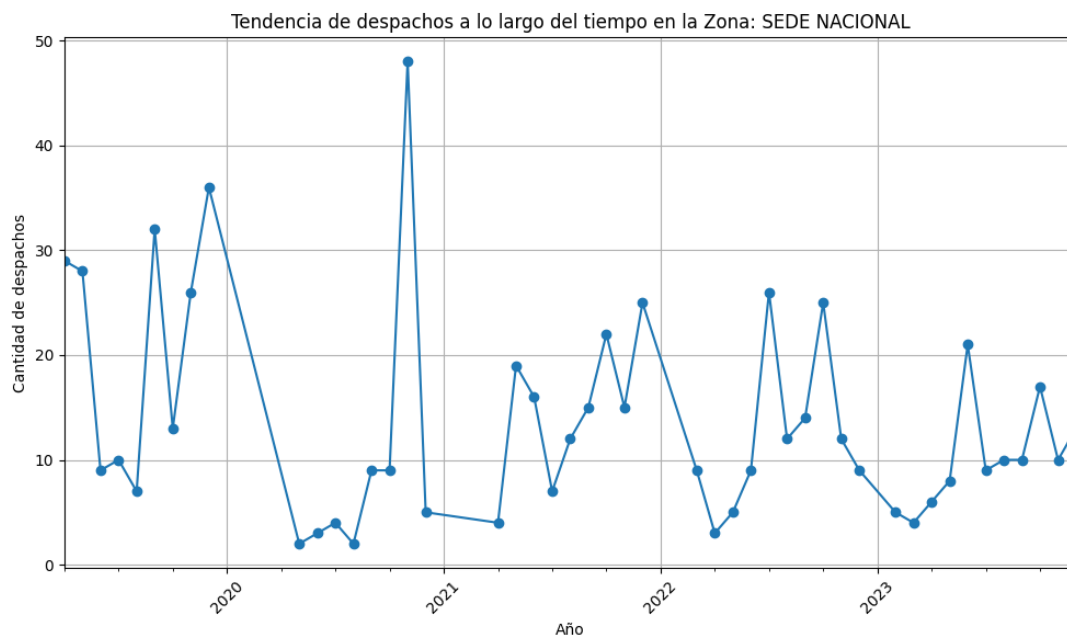
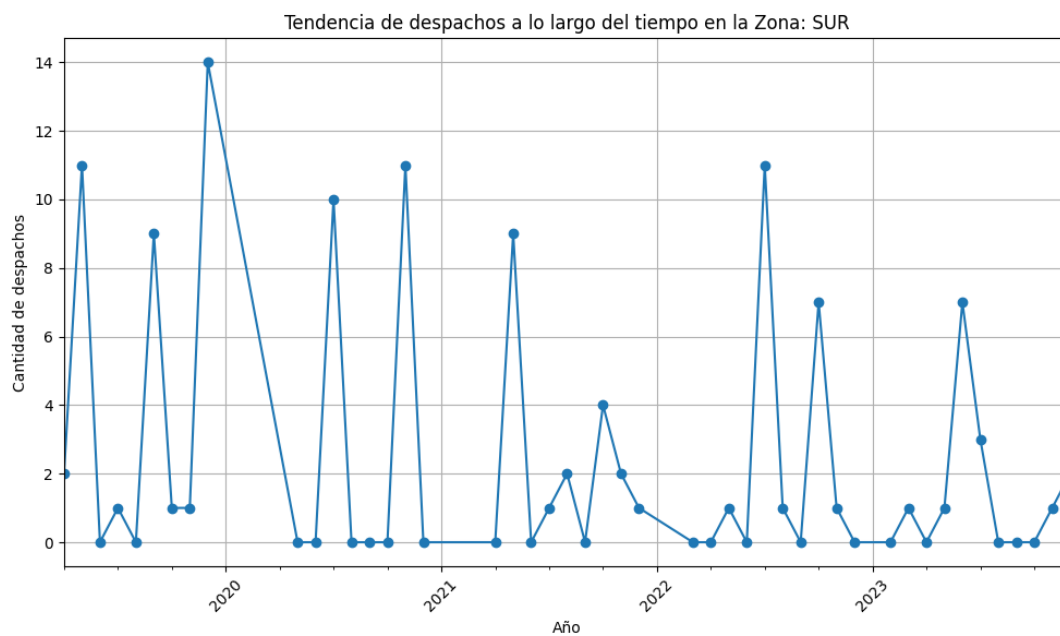


Figura 28*Tendencias de Despachos en la Sede Nacional a lo Largo del Tiempo***Figura 29***Tendencias de Despachos en la Zona Sur a lo Largo del Tiempo*

De acuerdo con el gráfico se evidencia:

1. Variabilidad temporal

Fluctuaciones Mensuales: Los movimientos de inventario no son constantes; hay meses con picos significativos y otros con baja actividad. Estas fluctuaciones están probablemente asociadas a eventos recurrentes, como el inicio de semestres académicos o actividades administrativas importantes.

Estacionalidad: Algunos periodos presentan patrones cíclicos (septiembre 2019 - diciembre 2019), (julio 2020 - noviembre 2020), (julio 2022 - agosto 2022 - octubre 2022) y junio 2023 en donde la actividad es consistentemente más activa. Esto sugiere que la logística está fuertemente influenciada por el calendario académico y administrativo de la UNAD.

2. Diferencias entre zonas

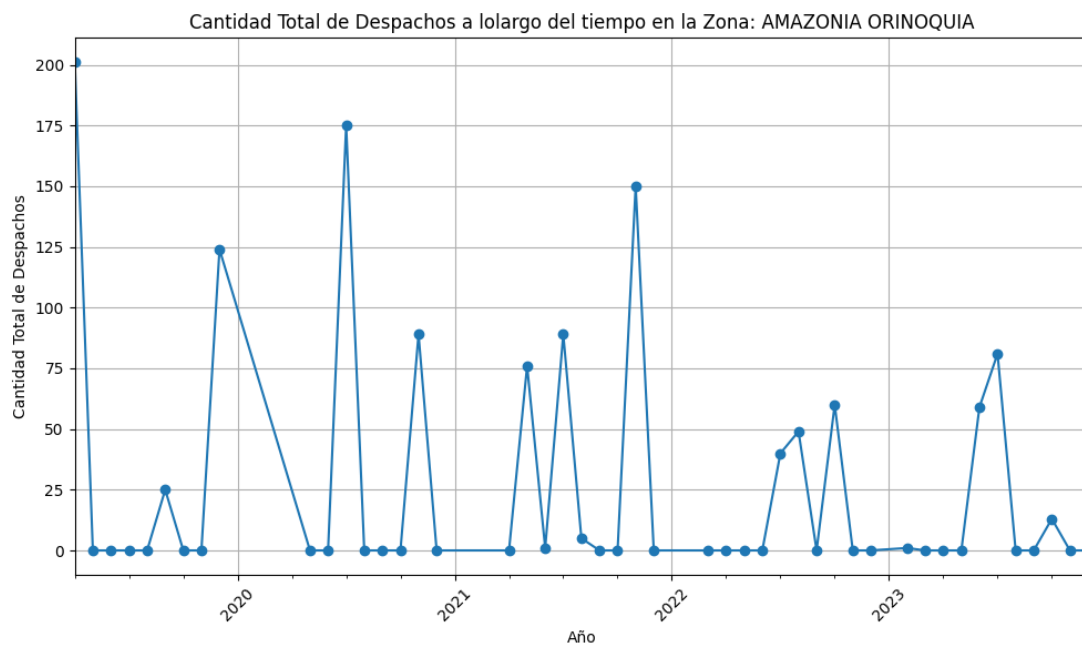
Zonas con alta actividad: Algunas zonas destacan por registrar un mayor volumen de movimientos de inventario, indicando que concentran una mayor cantidad de centros regionales o demandas específicas. Estas zonas experimentan picos claros que reflejan actividades importantes, como distribución de materiales para inicios de semestre.

Zonas con baja actividad: Las zonas con menor actividad muestran tendencias más estables, con menos variaciones significativas a lo largo del tiempo. Esto puede estar relacionado con una menor cantidad de centros o una logística más optimizada.

Sincronización entre zonas: Los picos de actividad suelen coincidir en varias zonas, lo que indica eventos globales, necesidades institucionales o ciclos programados de despachos.

Figura 30

Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Amazonia

**Figura 31**

Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Caribe

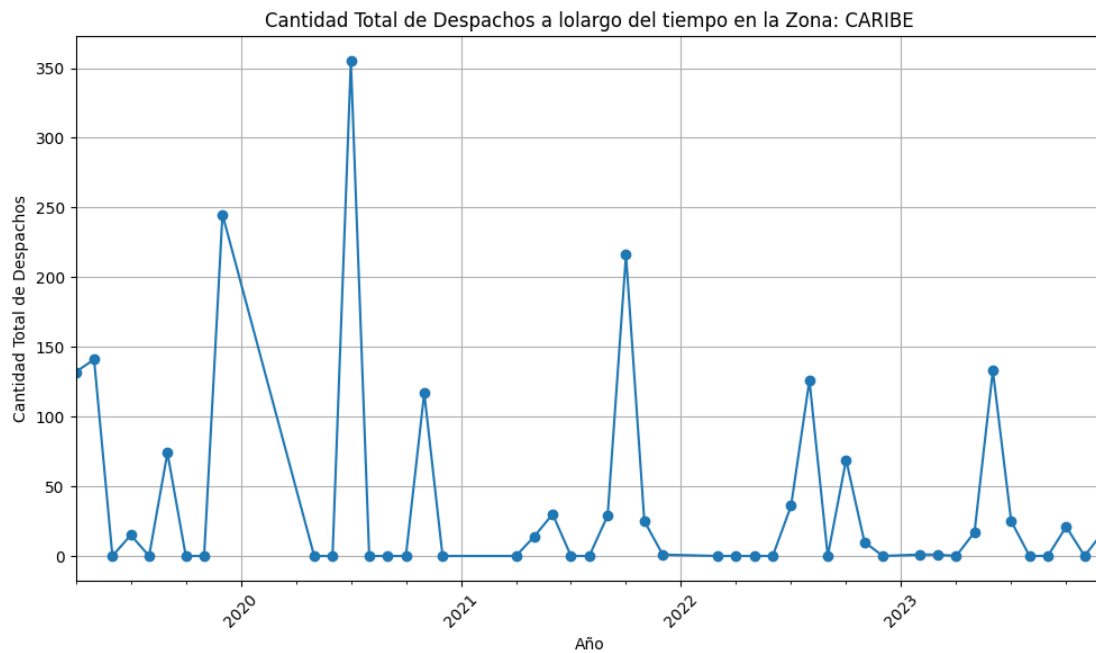
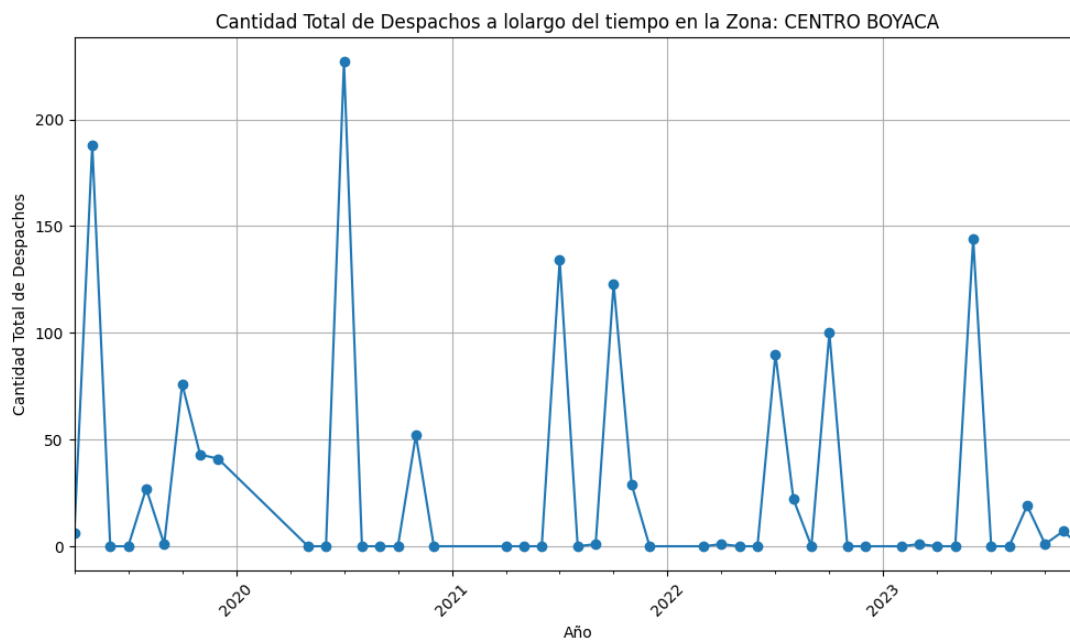


Figura 32

Cantidad Total de Despachos a lo Largo del Tiempo Para la Zona Boyacá

**Figura 33**

Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Oriente

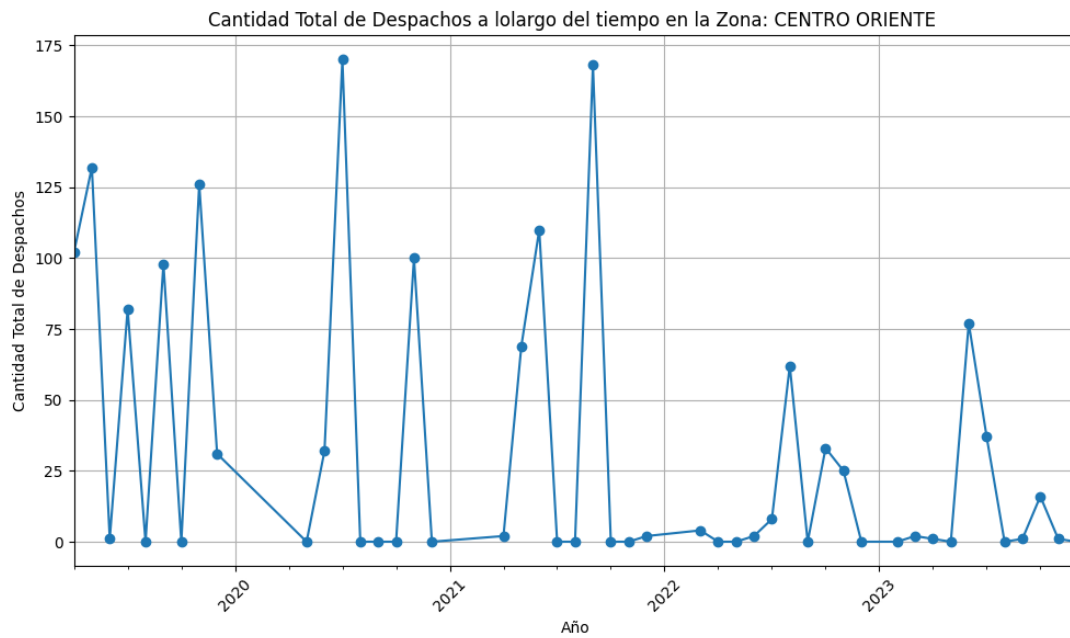
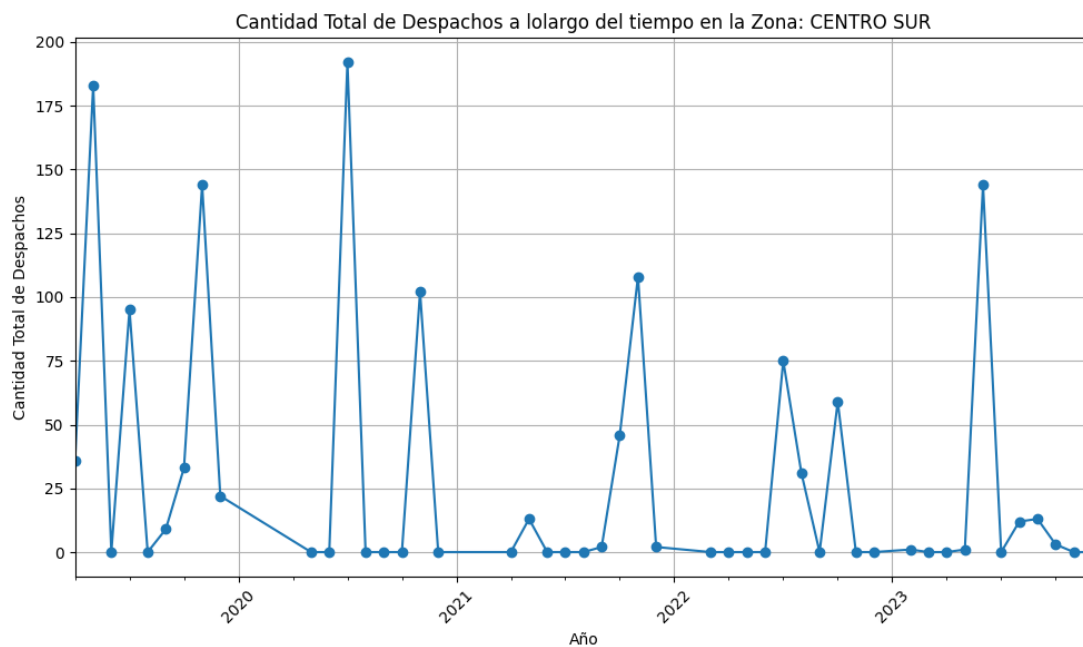


Figura 34

Cantidad Total de Despachos a lo Largo del Tiempo Para la Zona Centro Sur

**Figura 35**

Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Cundinamarca

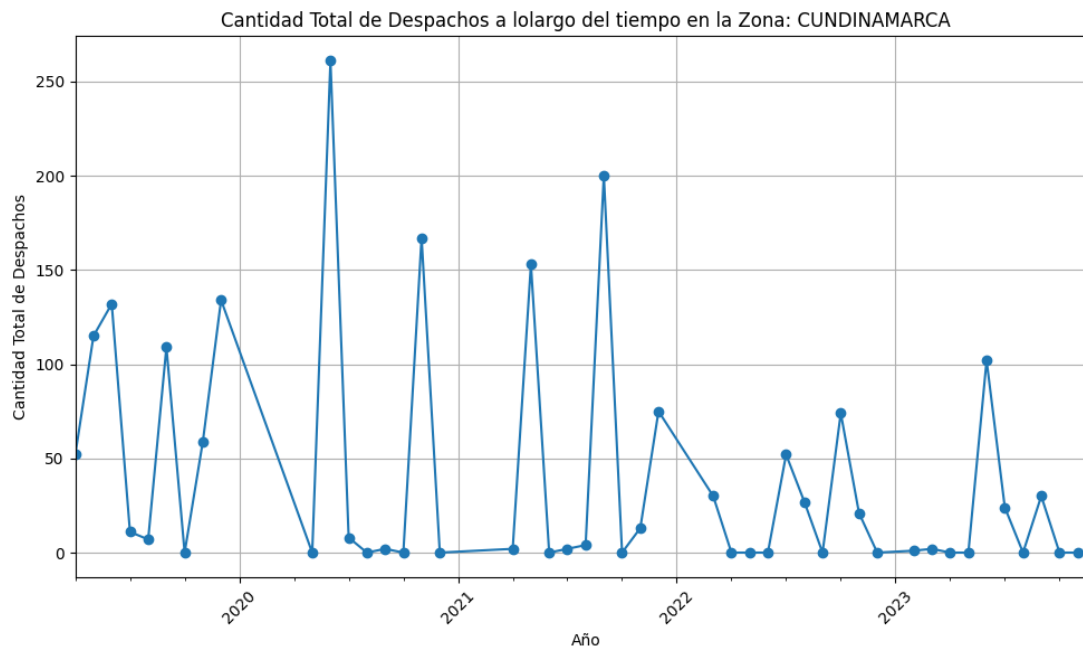
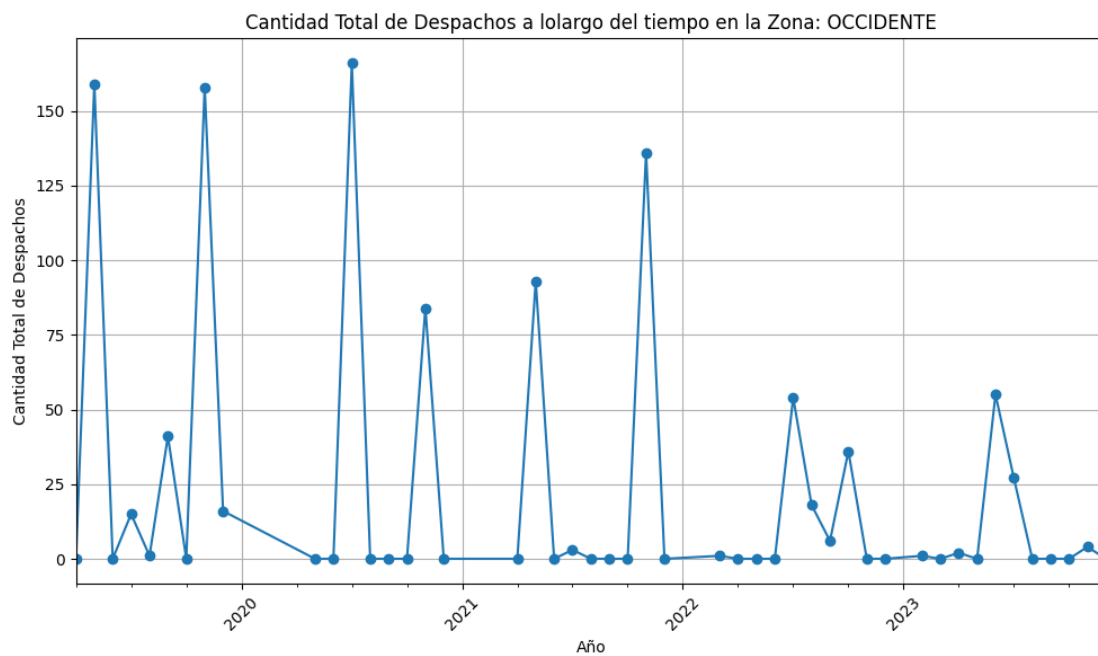
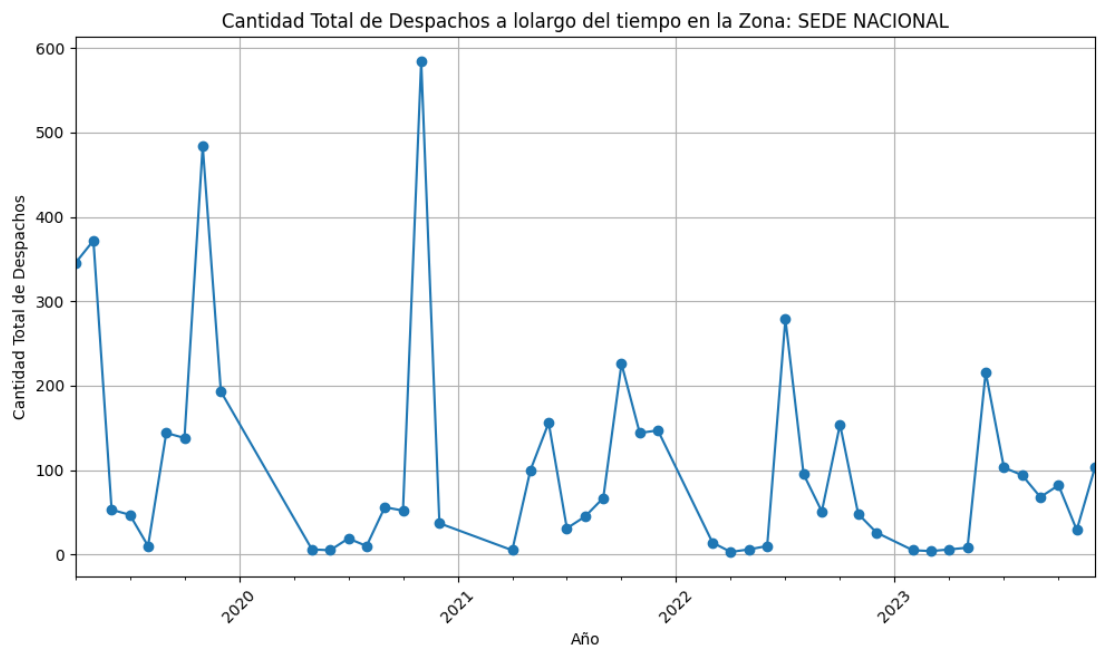


Figura 36*Cantidad Total de Despachos a lo Largo del Tiempo para la Zona Occidente***Figura 37***Cantidad Total de Despachos a lo Largo del Tiempo Para la Sede Nacional*

Picos de actividad: Los picos en los gráficos son indicativos de momentos clave como inicios de semestre, cuando los centros regionales requieren insumos, eventos administrativos o académicos a gran escala o ciclos programados despachos.

Meses de baja actividad: Los periodos con menor actividad coinciden probablemente con vacaciones académicas o periodos administrativos menos intensos.

Sincronización entre zonas: En algunos meses, se observan picos que coinciden en varias zonas, lo que sugiere que ciertos eventos afectan a toda la organización.

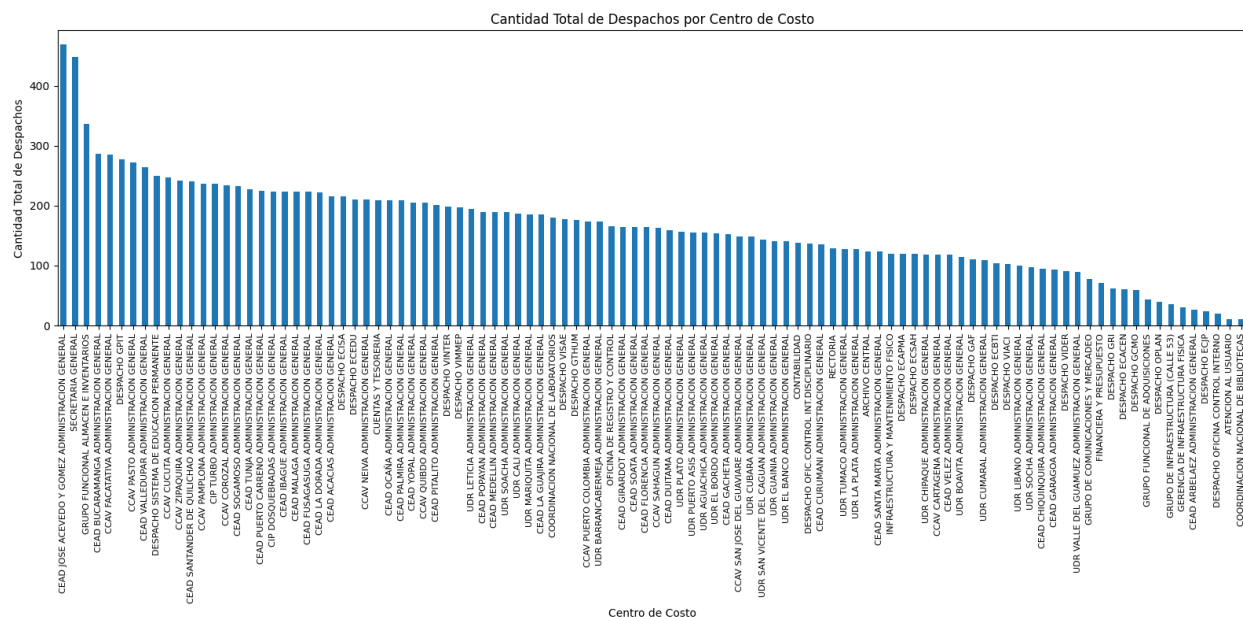
3. Diferencias regionales

Características Específicas: Cada zona muestra un patrón único de actividad logística, reflejando diferencias en tamaño y cantidad de los centros regionales, calendarios específicos o necesidades operativas locales. Algunas zonas tienen patrones más regulares y predecibles, mientras que otras experimentan variaciones significativas.

Despachos por Centros de Costo

Figura 39

Cantidad Total de Despachos por Centro de Costo



De acuerdo con el gráfico se evidencia que, algunos centros de costo destacan significativamente por tener un volumen más alto de despachos, reflejando una mayor necesidad de recursos. Estos centros podrían ser estratégicos en términos de logística, ya que concentran una parte importante de la actividad.

Otros centros tienen una cantidad de despachos mucho menor, lo que podría estar relacionado con menores demandas operativas o un alcance más reducido.

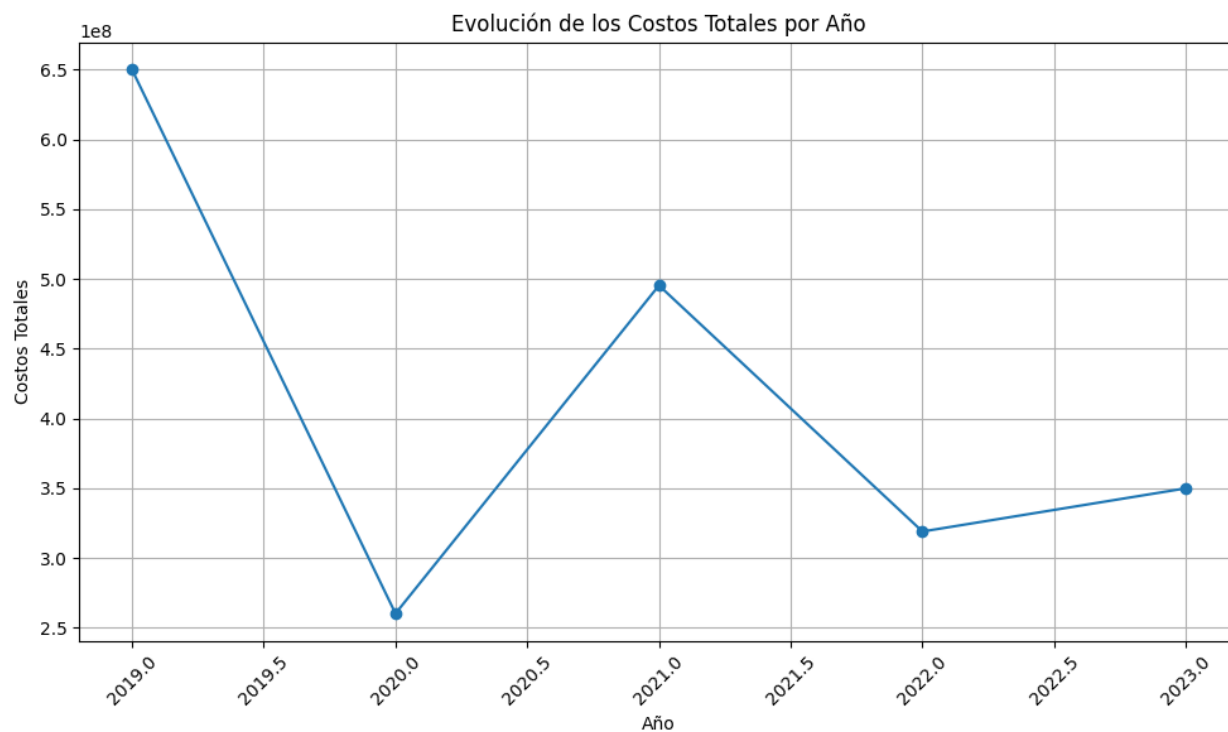
La distribución no es uniforme, lo que sugiere que algunos centros son más críticos en términos logísticos que otros.

Análisis Financiero

Evolución de los Costos Totales

Figura 40

Evolución de Costos Totales



De acuerdo con el gráfico se evidencia una tendencia ascendente, descendente o fluctuante en los costos totales.

Una tendencia ascendente podría estar asociada con: incrementos en la cantidad de despachos realizados, aumento en los costos unitarios de los productos o expansión de las operaciones de la UNAD, como la apertura de nuevos centros regionales.

Una tendencia descendente podría reflejar: optimización en la logística, consolidación de envíos o reducción de la demanda de bienes o un uso más eficiente de los recursos existentes.

Los años con mayores costos totales probablemente coincidan con eventos específicos como: inicios de grandes proyectos académicos o administrativos, incremento en el precio de productos despachados o cambios en políticas de abastecimiento o expansión de centros. Estos picos representan periodos críticos que pueden haber requerido ajustes importantes en el presupuesto y la capacidad operativa.

Los años con costos totales más bajos pueden estar asociados con: periodos de menor actividad académica o administrativa, reducción en el número de despachos o consolidación de envíos o mejora en las prácticas de adquisición y logística. Estos periodos pueden reflejar esfuerzos exitosos de optimización y control de gastos.

El volumen de despachos está directamente relacionado con la demanda de recursos en los centros regionales. Años de alta matrícula o actividades académicas intensas pueden haber incrementado los costos. Los costos también pueden verse afectados por fluctuaciones en los precios del mercado de bienes de consumo.

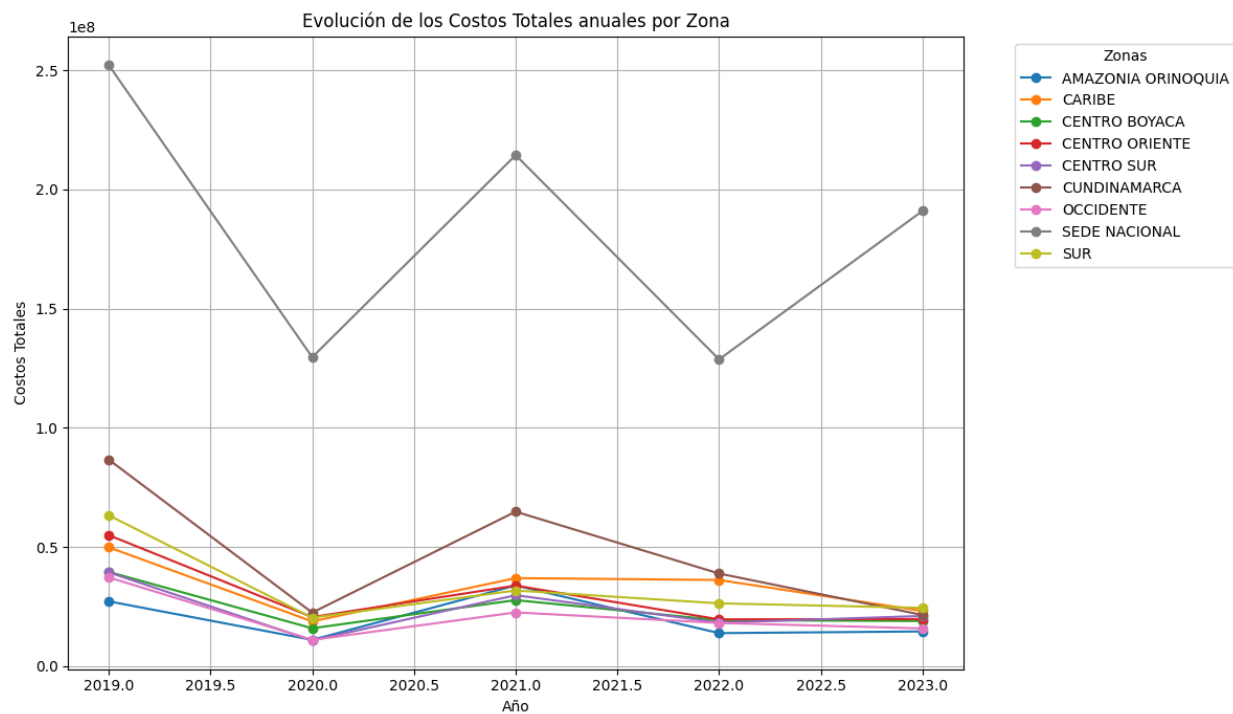
Figura 41*Evolución de Costos Totales Anuales por Zona*

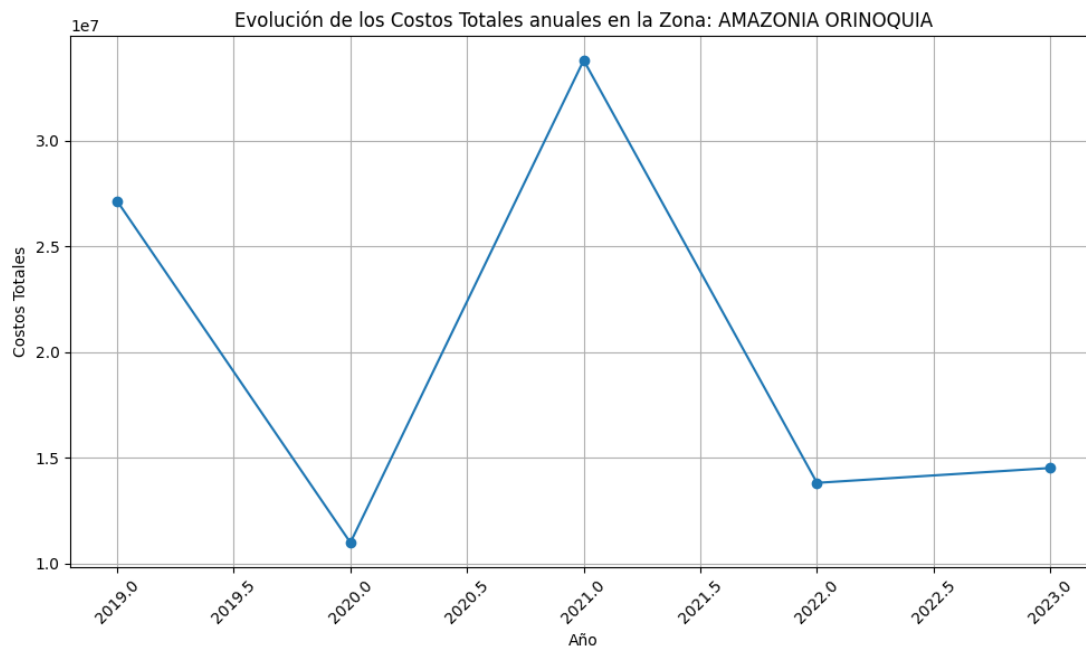
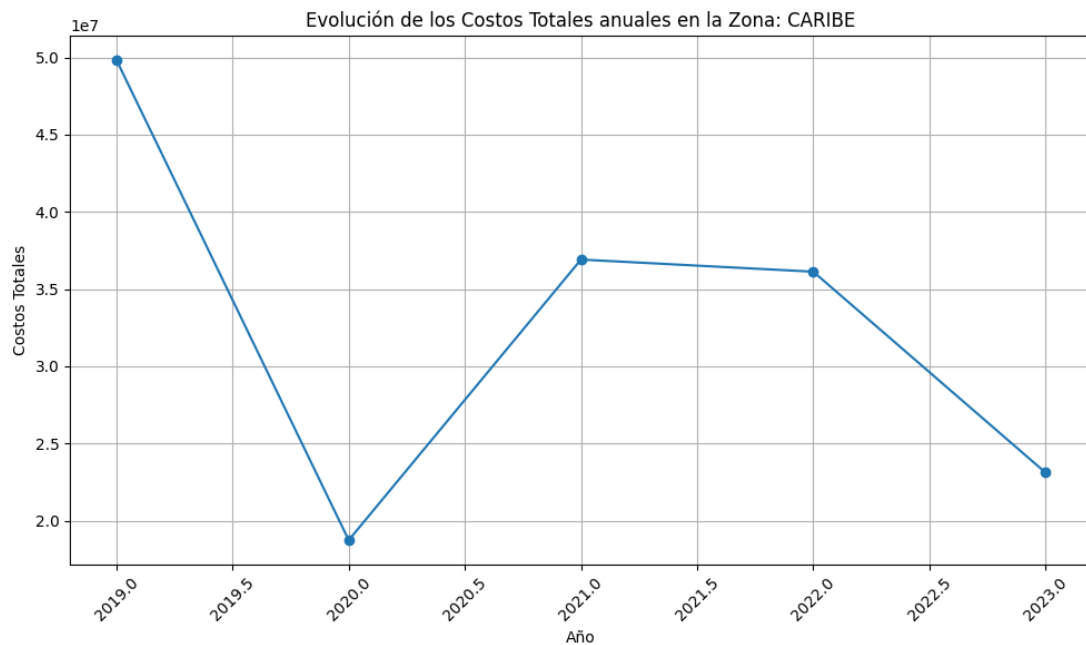
Figura 42*Evolución de Costos Totales Anuales en la Zona Amazonia Orinoquia***Figura 43***Evolución de Costos Totales Anuales en la Zona Caribe*

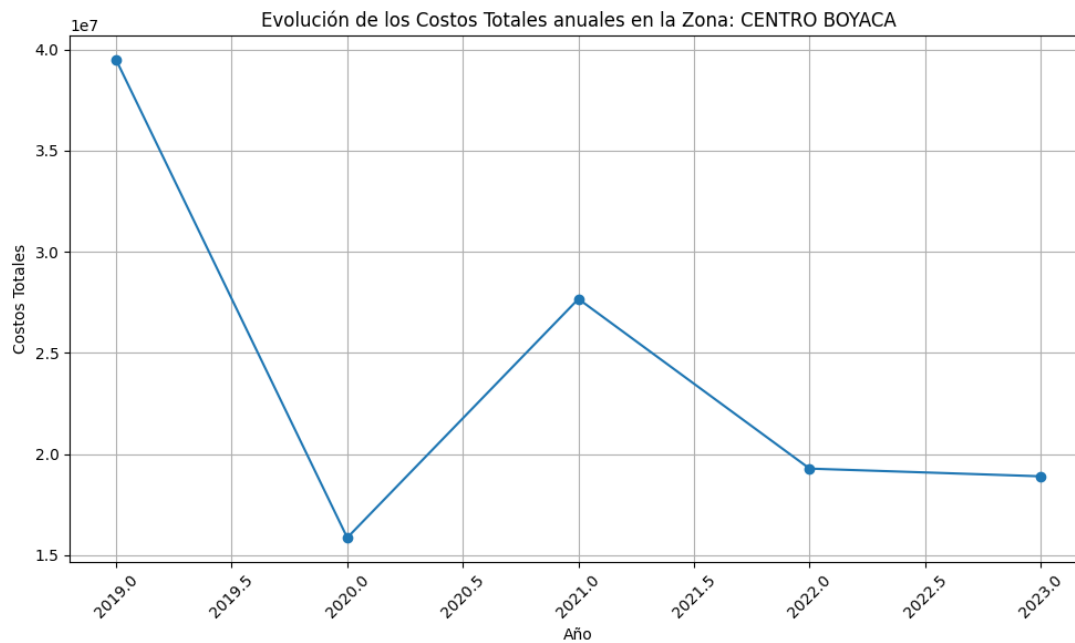
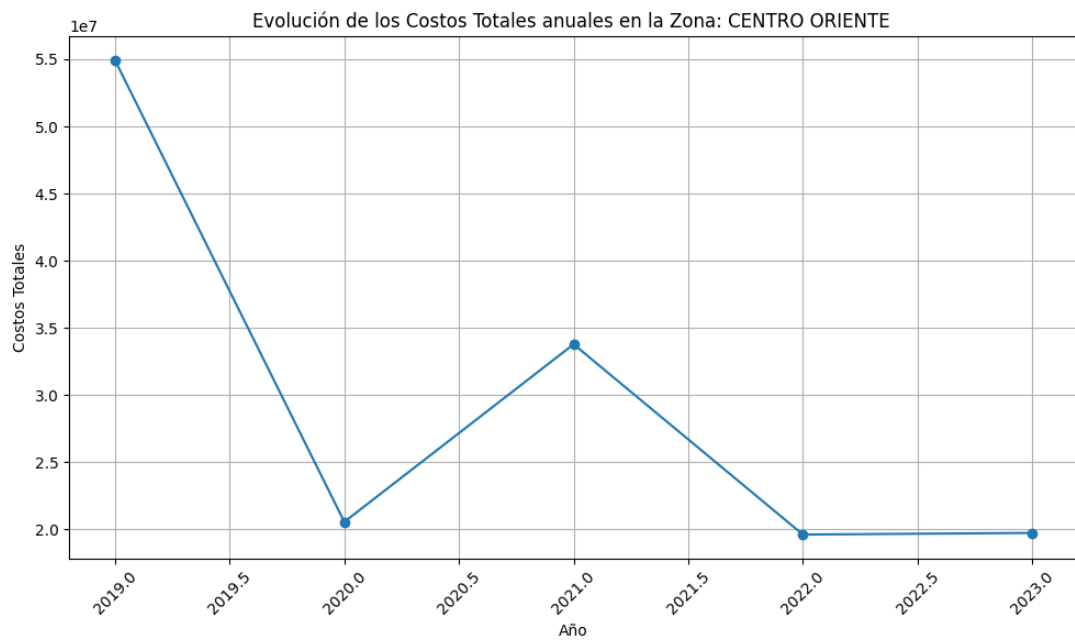
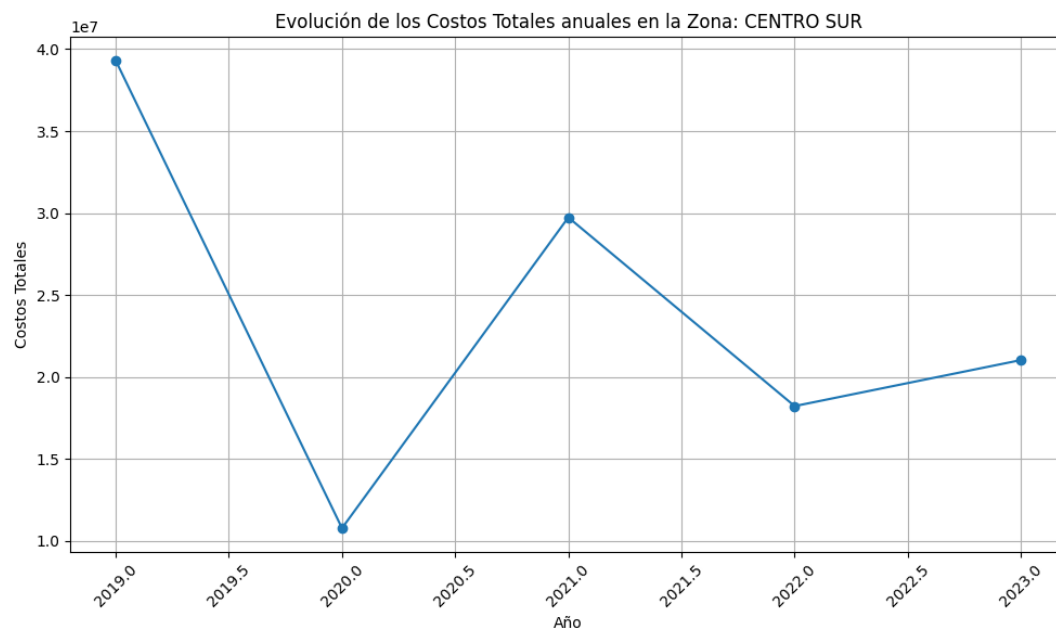
Figura 44*Evolución de Costos Totales Anuales en la Zona Boyacá***Figura 45***Evolución de Costos Totales Anuales en la Zona Oriente*

Figura 46

Evolución de Costos Totales Anuales en la Zona Centro Sur

**Figura 47**

Evolución de Costos Totales Anuales en la Zona Cundinamarca

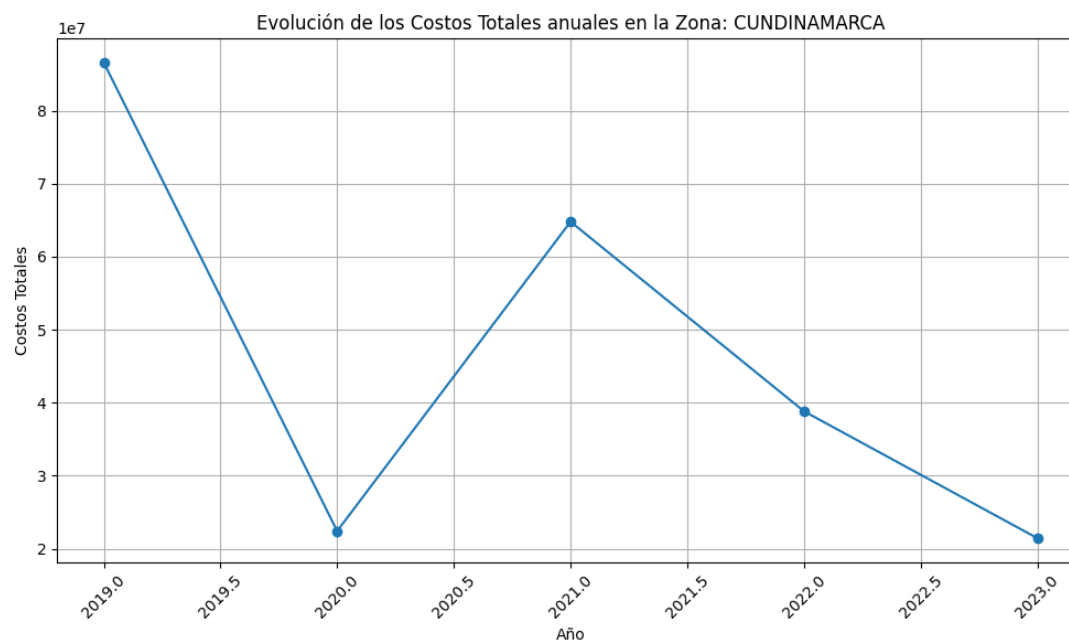


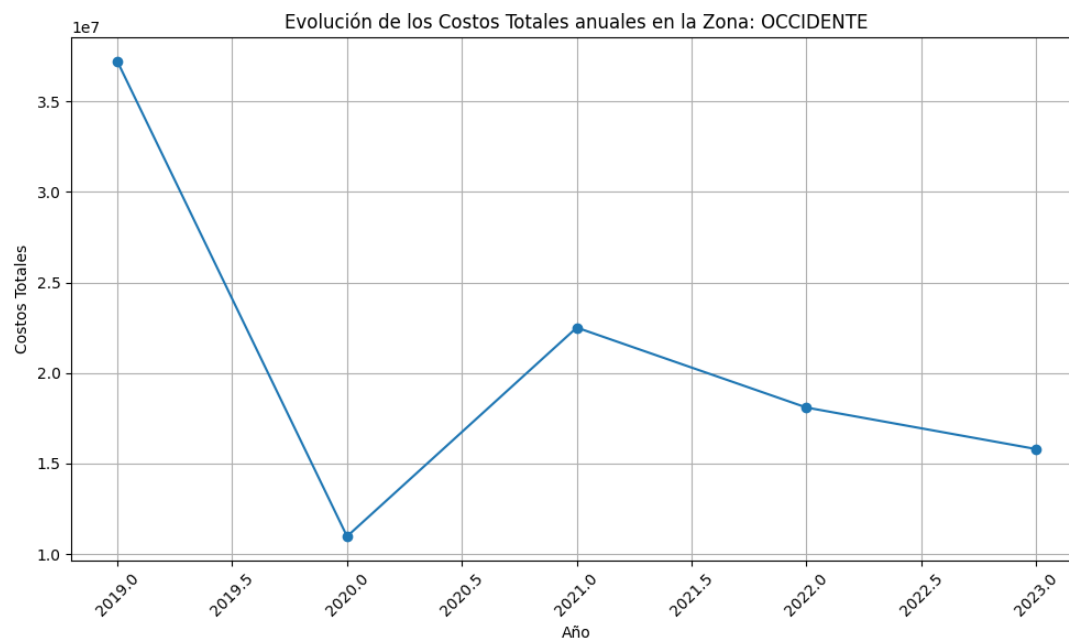
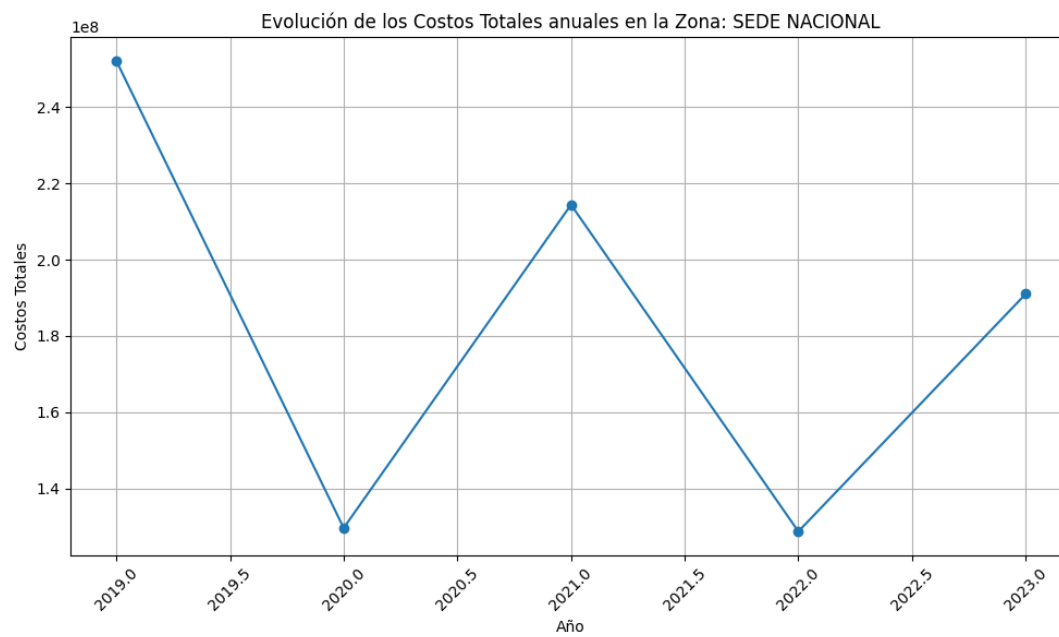
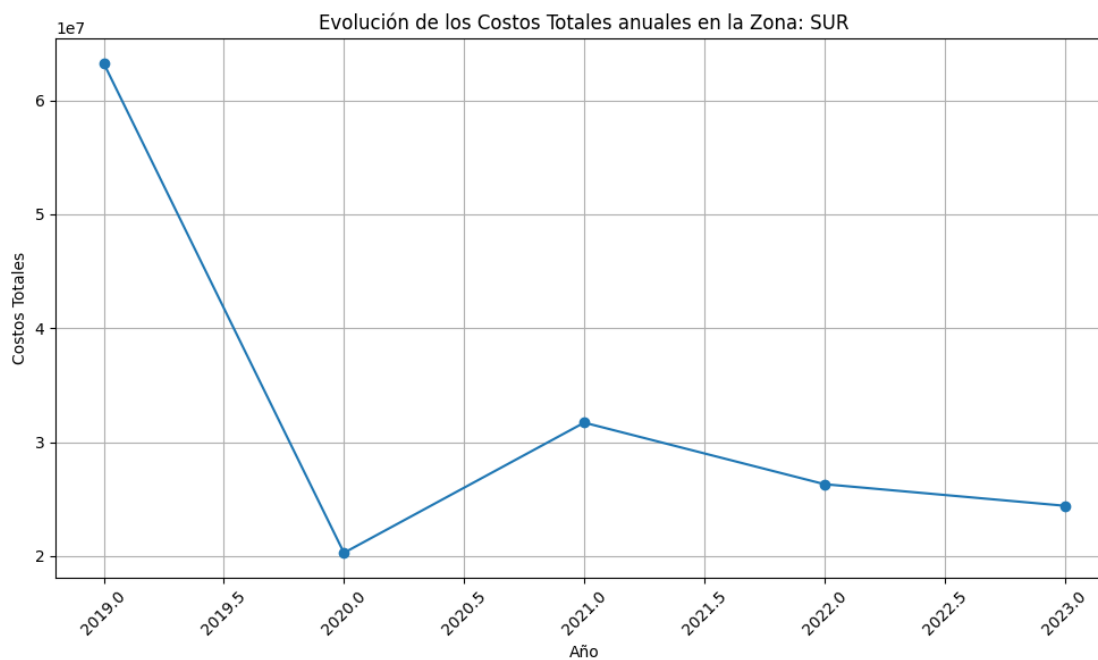
Figura 48*Evolución de Costos Totales Anuales en la Zona Occidente***Figura 49***Evolución de Costos Totales Anuales en la Zona Sede Nacional*

Figura 50

Evolución de Costos Totales Anuales en la Zona Sur



De acuerdo con el grafico se evidencia:

1. Diferencias entre zonas

Zonas con altos costos totales: algunas zonas presentan consistentemente costos más altos a lo largo del tiempo, lo que sugiere: una mayor cantidad de centros regionales atendidos, una logística más compleja debido a la dispersión geográfica o el volumen de demanda o incrementos en la actividad académica o administrativa en esas regiones.

Zonas con costos estables o bajos: las zonas con costos más bajos suelen tener menos centros o una demanda más homogénea. Estas regiones podrían tener rutas logísticas optimizadas o menos eventos que incrementen los costos.

2. Fluctuaciones temporales

Picos significativos: los años con picos altos en varias zonas podrían estar asociados con eventos globales, como inicios de semestres, expansiones académicas o aumento en precios de productos. En ciertas zonas, los picos son aislados y podrían estar vinculados a necesidades específicas o proyectos puntuales.

Patrones estacionales: algunas zonas muestran incrementos regulares en periodos determinados, reflejando posiblemente una mayor sincronización con el calendario académico de la UNAD.

3. Factores posibles detrás de los picos

Incremento en los precios de los productos: las variaciones de costos pueden estar influenciadas por fluctuaciones en el mercado de bienes de consumo.

Mayor demanda de recursos: eventos específicos, como la apertura de nuevos centros o aumento en la matrícula, pueden incrementar los despachos y, por ende, los costos.

Optimización logística: las zonas con patrones más estables podrían haber implementado estrategias logísticas más eficientes, reduciendo el impacto de la variabilidad en la demanda.

Costos por Centros de Costo

Contribución Porcentual de los Costos Totales Anuales

Figura 52

Porcentaje de Costos Totales Anuales por Zona

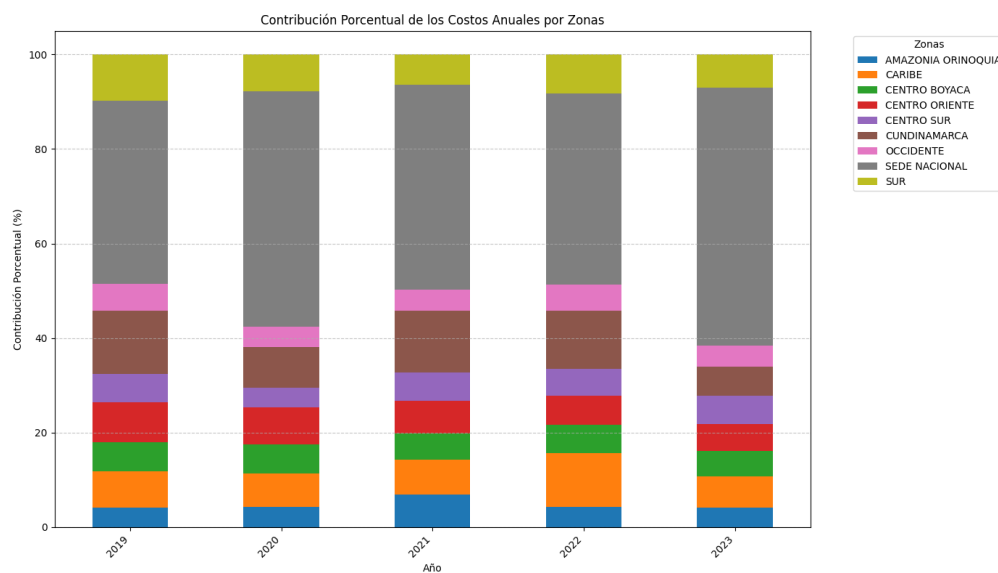


Figura 53

Contribución Proporcional de Costos Totales Anuales por Zonas

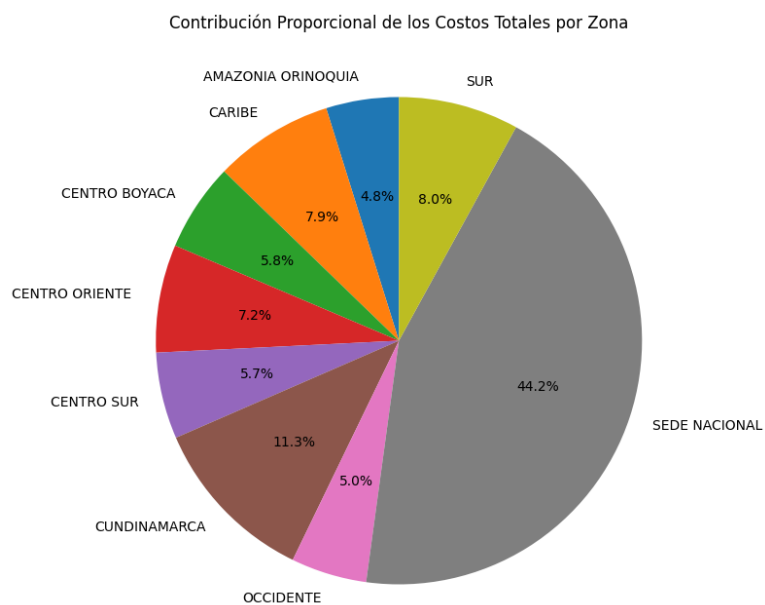
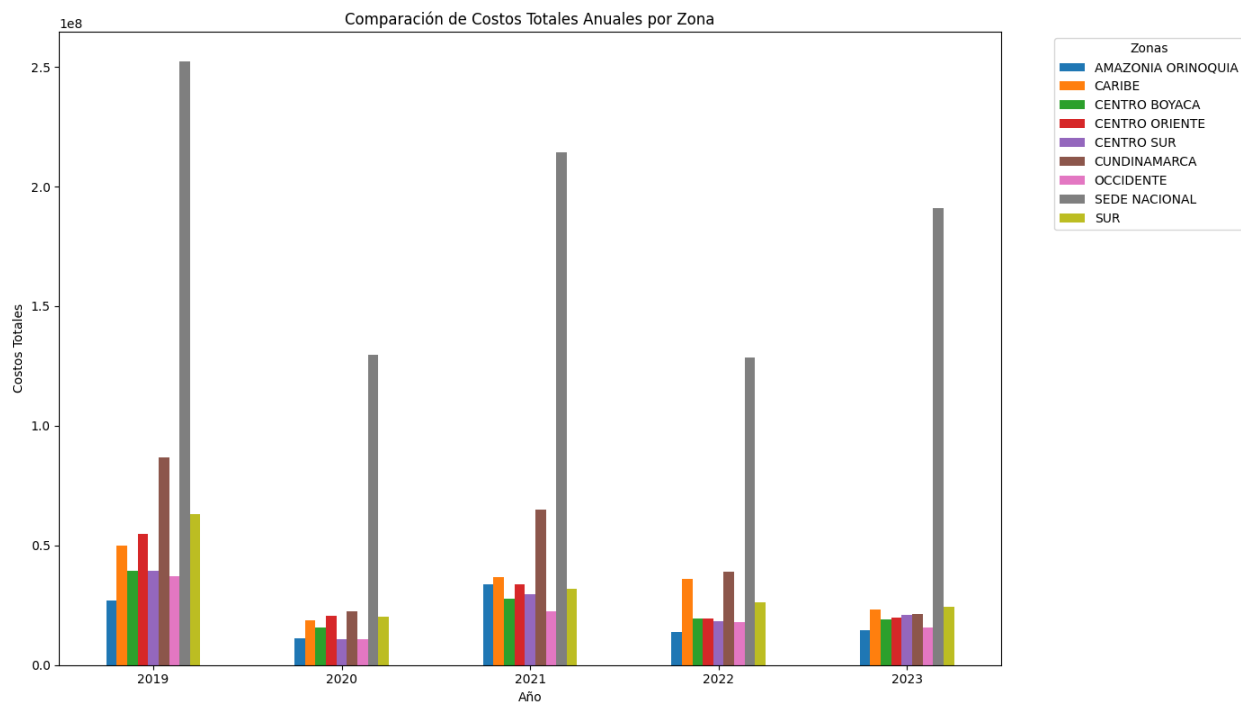


Figura 54*Comparación de Costos Totales Anuales por Zonas*

De acuerdo con los gráficos se evidencia que, la Sede Nacional tiene una contribución dominante en todos los años, con un porcentaje que oscila entre el 40% y el 60% del costo total anual. Lo que sugiere su papel estratégico como nodo central en la red logística de la UNAD.

Regiones como Cundinamarca, Caribe, y Sur muestran contribuciones más equilibradas y estables, generalmente entre el 7% y el 10% del total. Estas zonas mantienen un peso logístico moderado sin cambios significativos a lo largo del tiempo.

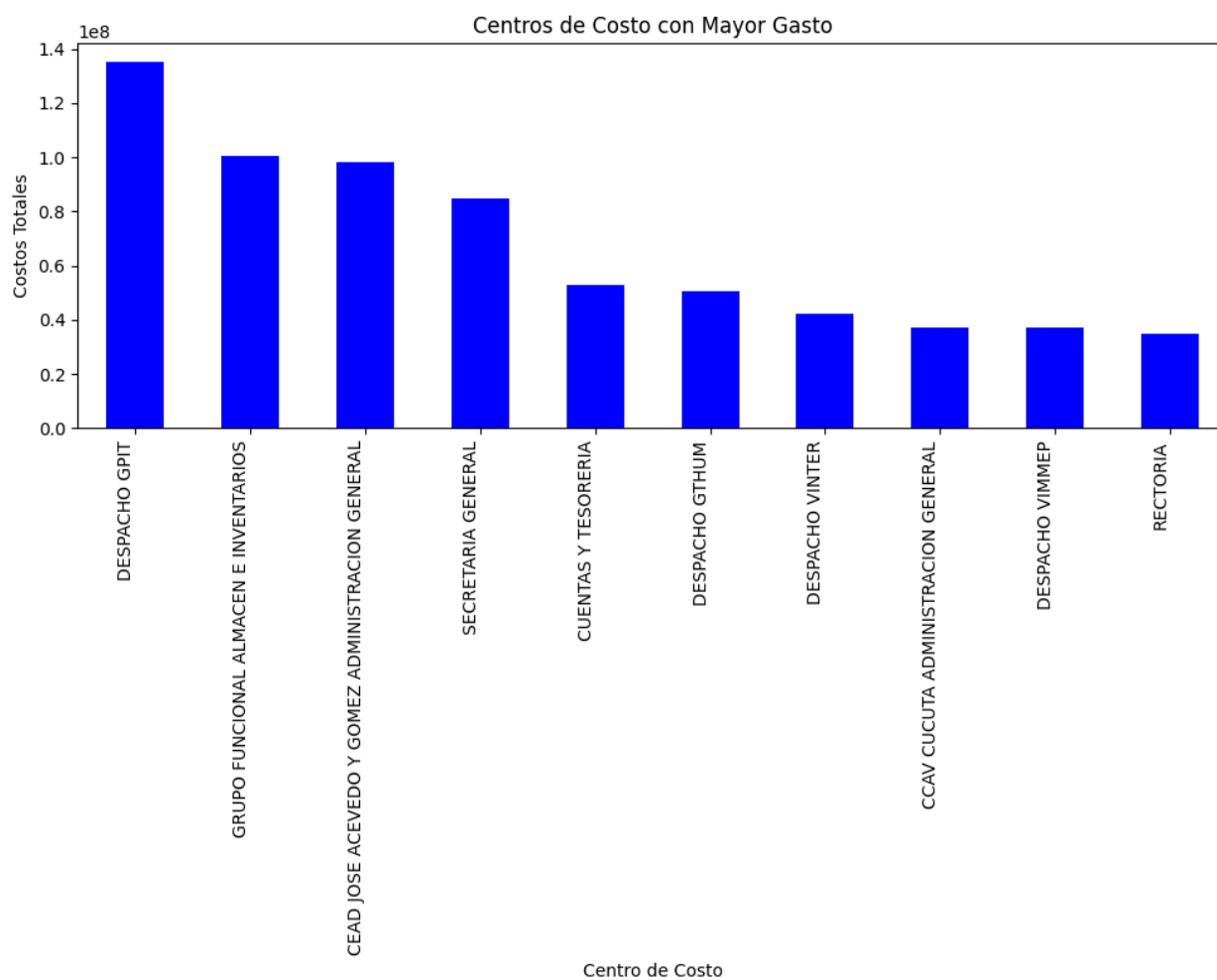
Amazonia Orinoquía, Centro Sur, y Occidente tienen contribuciones bajas y constantes, con porcentajes menores al 6% del costo total anual. Esto refleja su menor peso en la operación logística debido a factores como menor demanda o ubicación remota.

La contribución de la Sede Nacional tiende a aumentar en años como 2021 y 2023, mientras que en 2020 y 2022 disminuye ligeramente, lo que podría estar relacionado con cambios en la estrategia logística o impactos externos (como la pandemia de COVID-19).

Centros de Costo con Mayor Gasto

Figura 55

Centros de Costo con Mayor Gasto Total



De acuerdo con el gráfico se evidencia que, el Despacho GPIT tiene el gasto total más alto, destacándose significativamente frente a los demás centros

El Grupo Funcional de Almacén e Inventarios, la Secretaría General y CEAD José Acevedo y Gómez administración general también tienen altos costos, indicando su importancia estratégica en las operaciones administrativas y logísticas.

Centros como Rectoría, Despacho GTHUM y CCAV Cúcuta administración general aparecen en la lista, reflejando el impacto de sus actividades en los costos operativos.

Figura 56

Centros de Costo con Mayor Gasto en el Año 2019

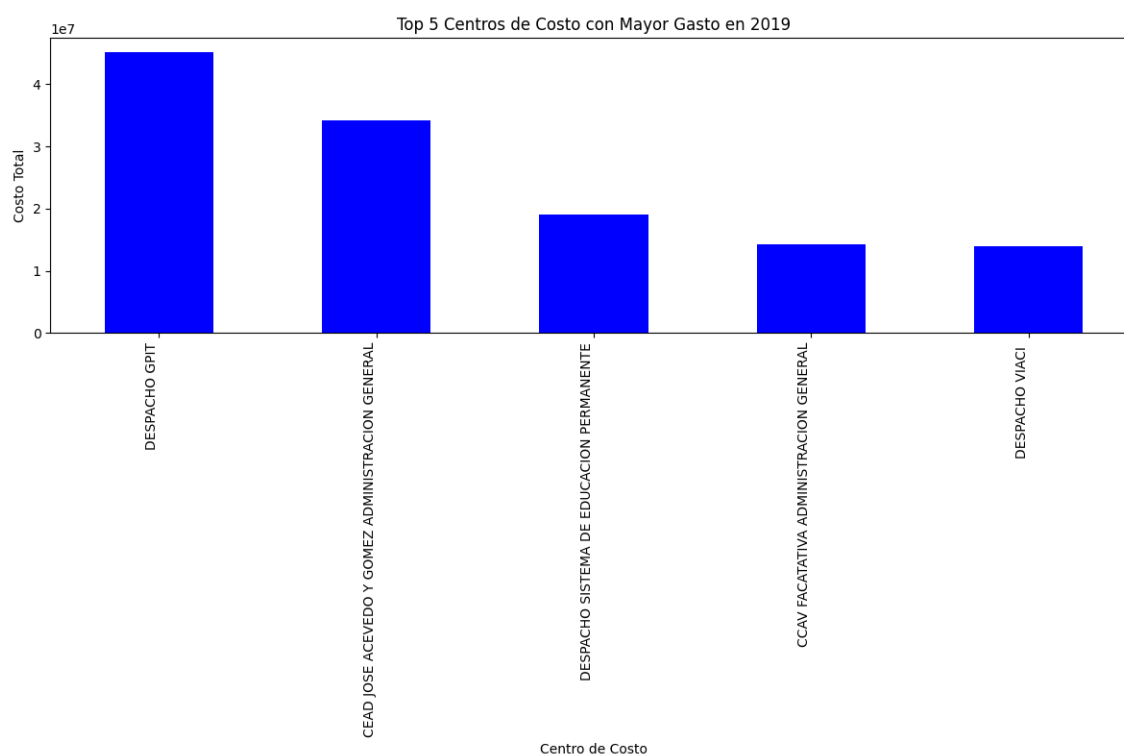


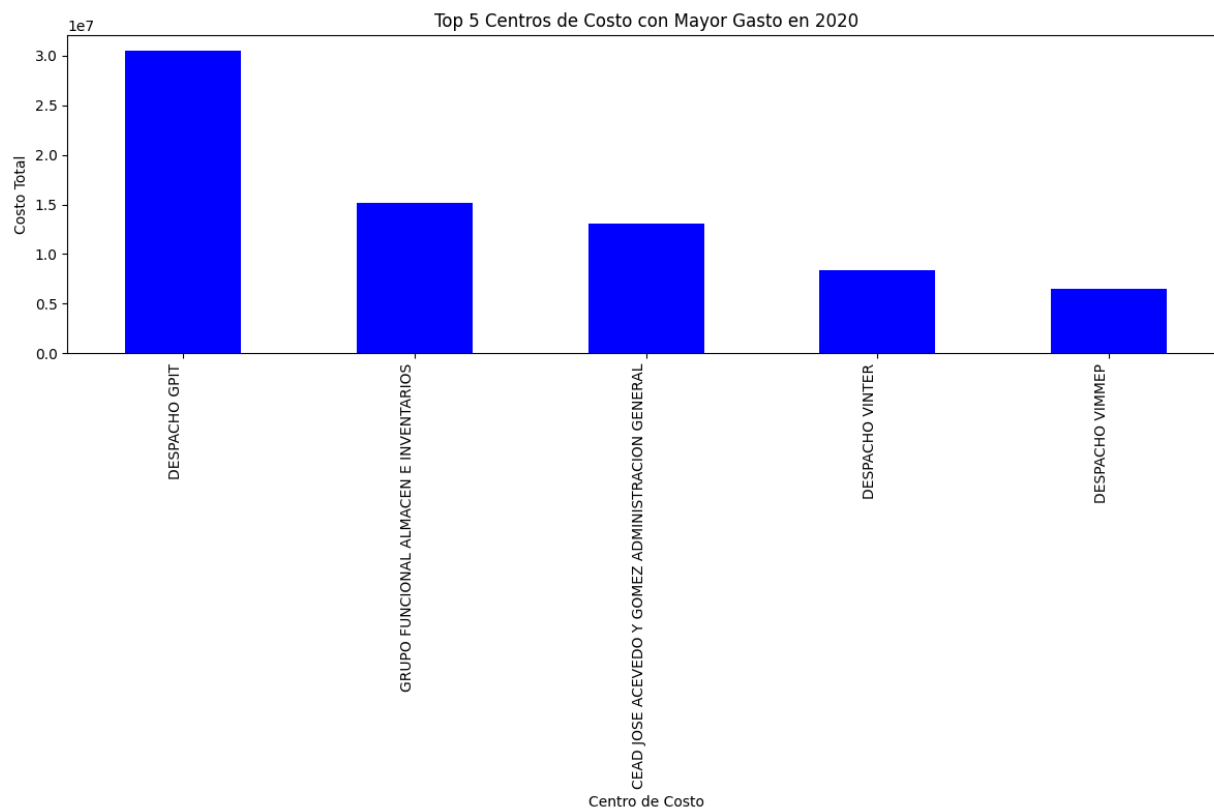
Figura 57*Centros de Costo con Mayor Gasto en el Año 2020*

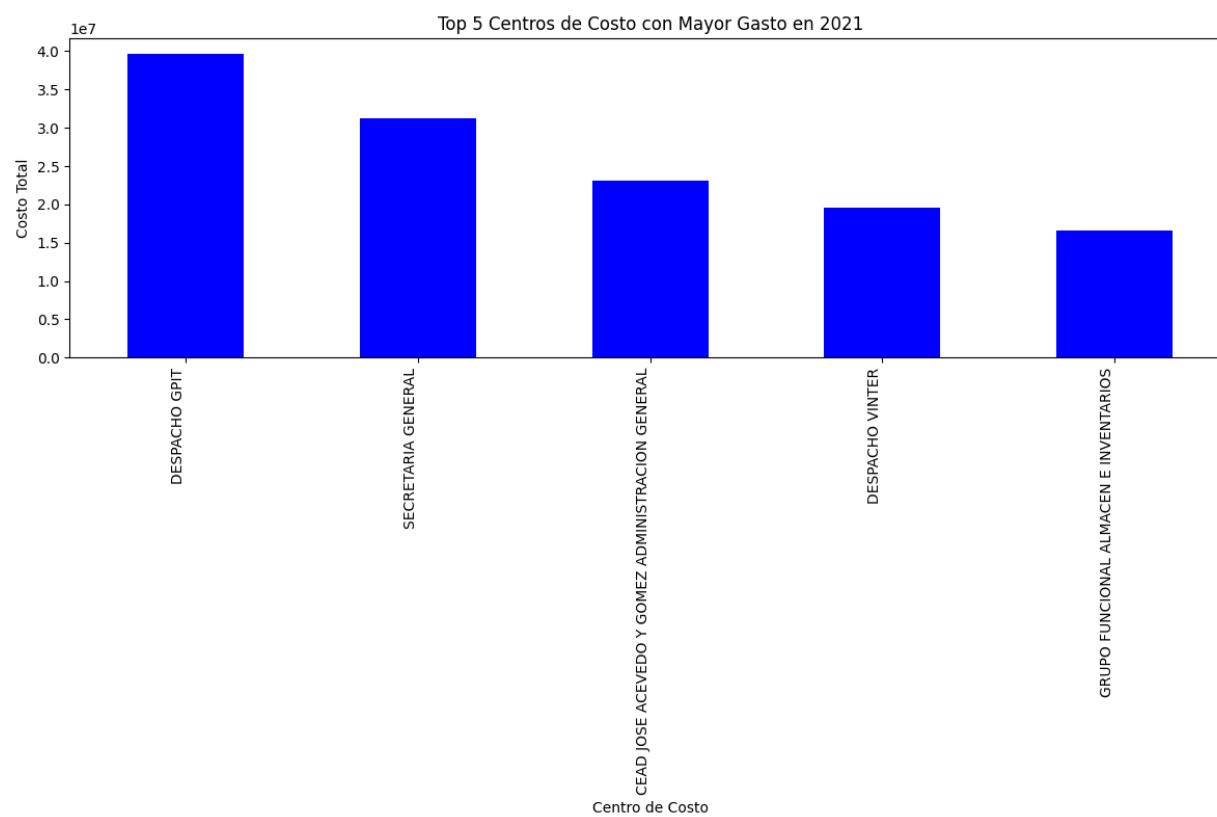
Figura 58*Centros de Costo con Mayor Gasto en el Año 2021*

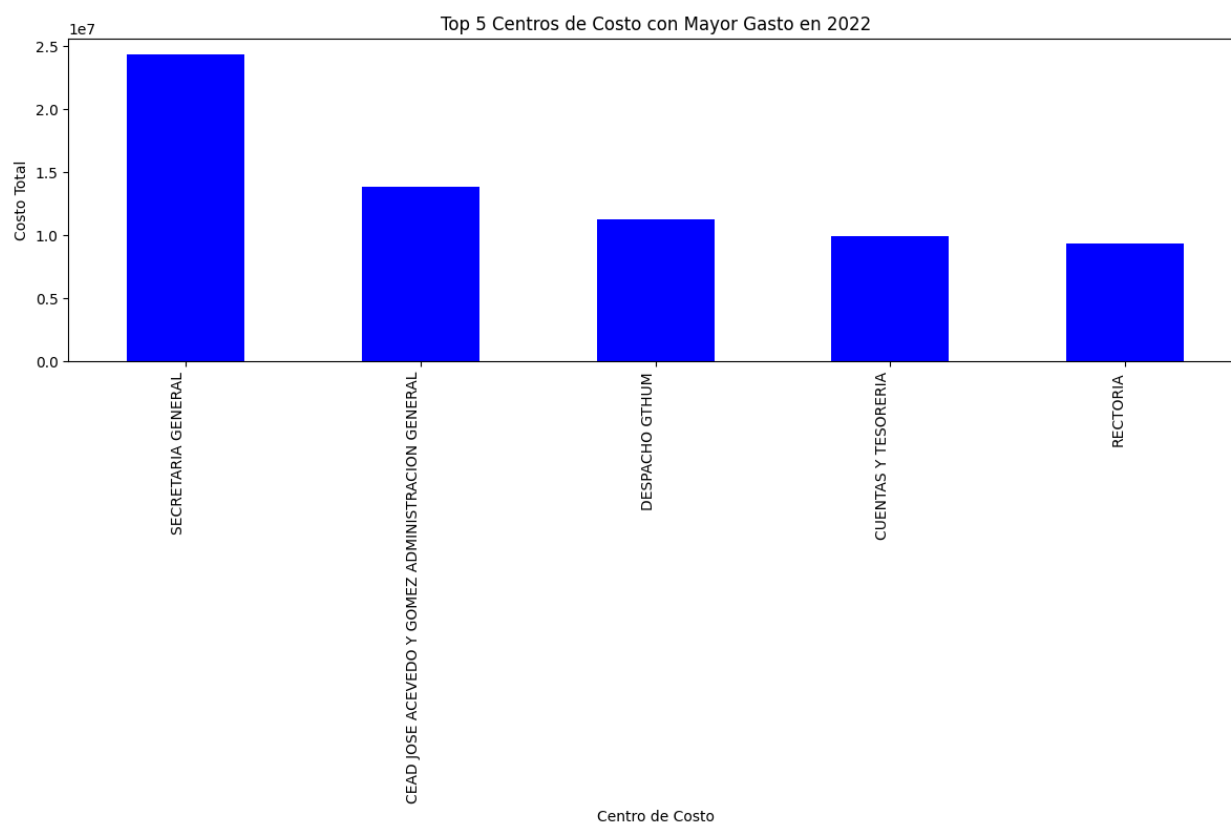
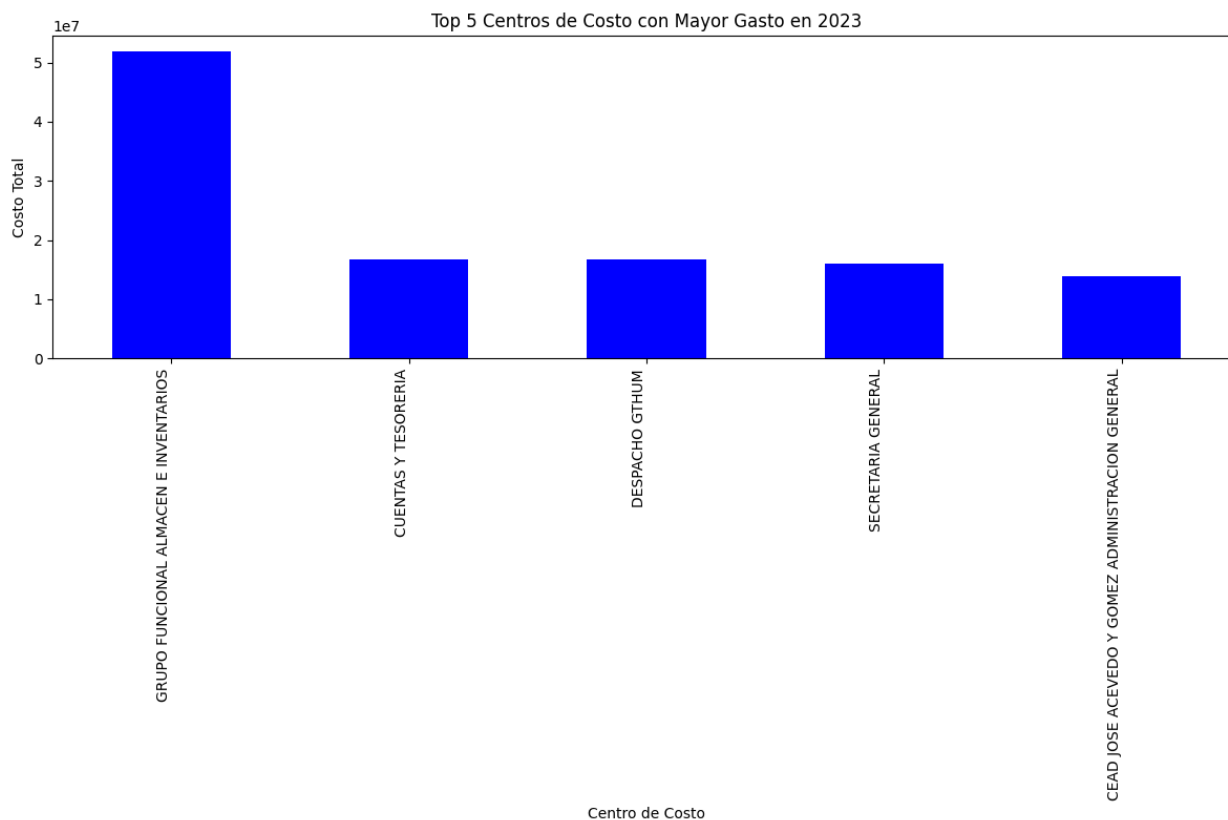
Figura 59*Centros de Costo con Mayor Gasto en el Año 2022*

Figura 60

Centros de Costo con Mayor Gasto en el Año 2023



De acuerdo con los gráficos se evidencia que, los centros con mayores costos pueden cambiar de un año a otro, reflejando cambios en las prioridades logísticas o en las necesidades de los centros.

Algunos centros (Despacho GPIT, secretaria general y CEAD José Acevedo y Gómez administración general) aparecen consistentemente en los primeros lugares a través de varios años, indicando su importancia estratégica y su impacto constante en los costos.

En ciertos años (2019, 2021 y 2023), los costos de algunos centros son significativamente más altos, lo que podría estar asociado a proyectos específicos, aumentos en las demandas o eventos extraordinarios.

Centros de Costo con Menor Gasto

Figura 61

Centros de Costo con Menor Gasto Total

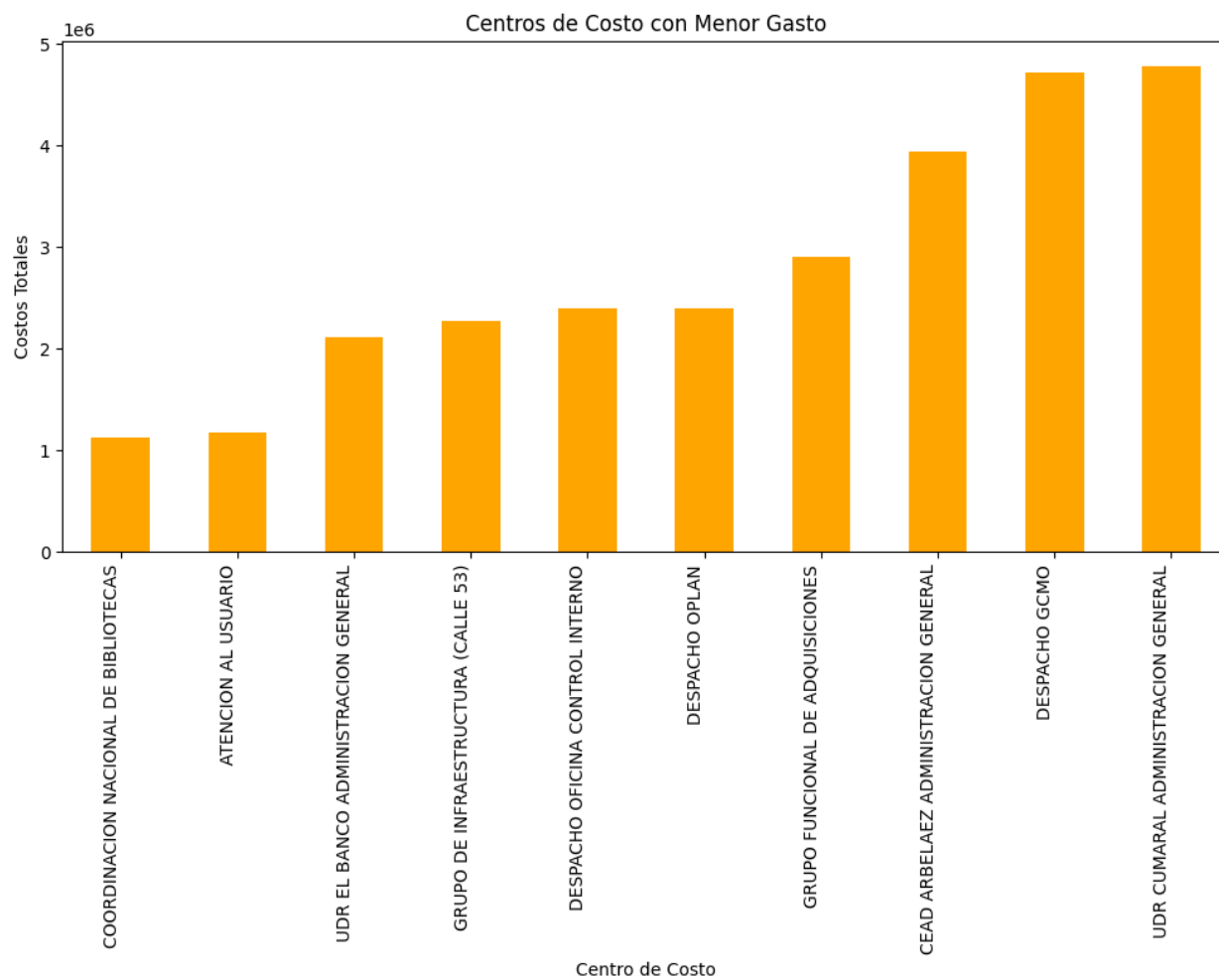


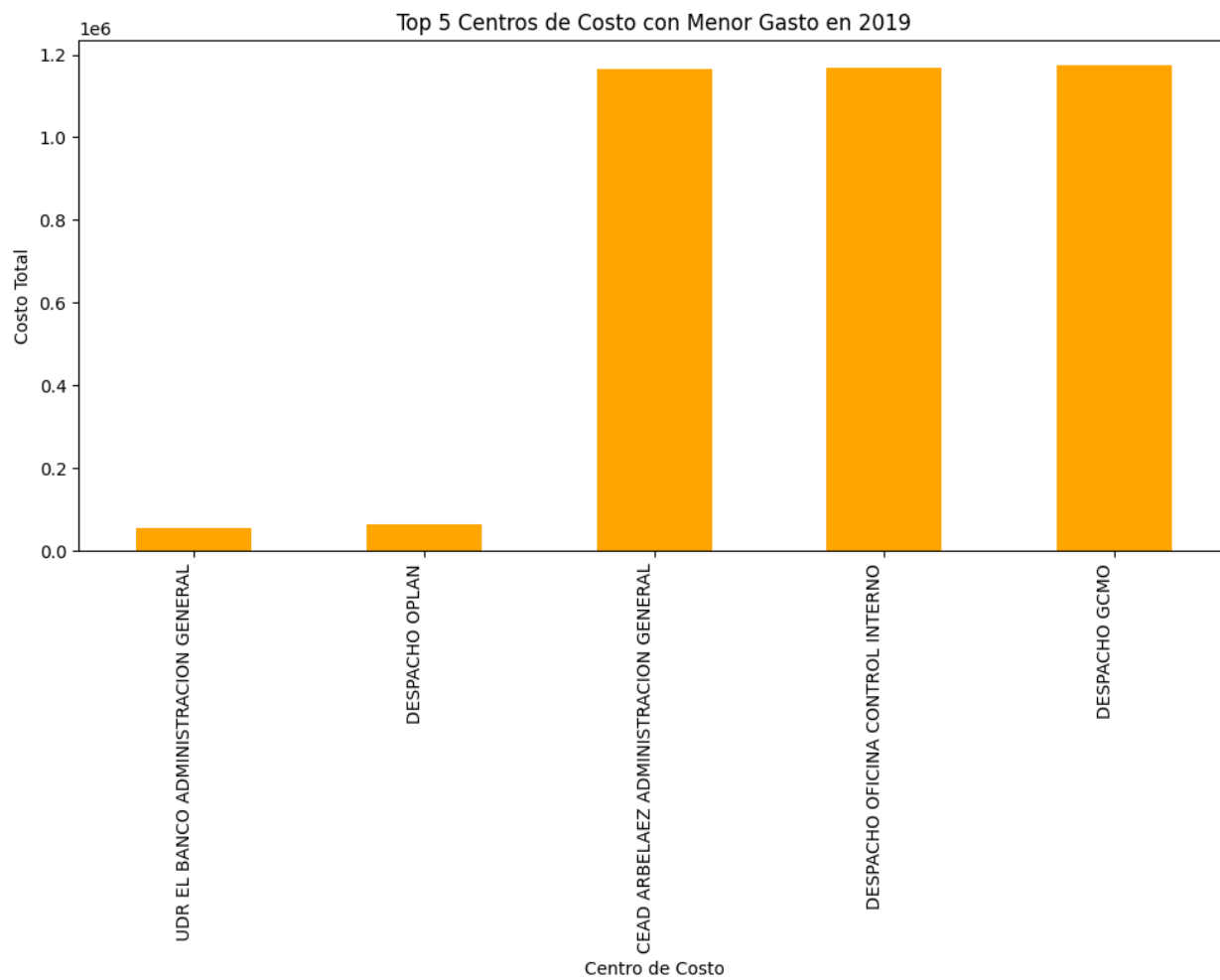
Figura 62*Centros de Costo con Menor Gasto Total en el Año 2019*

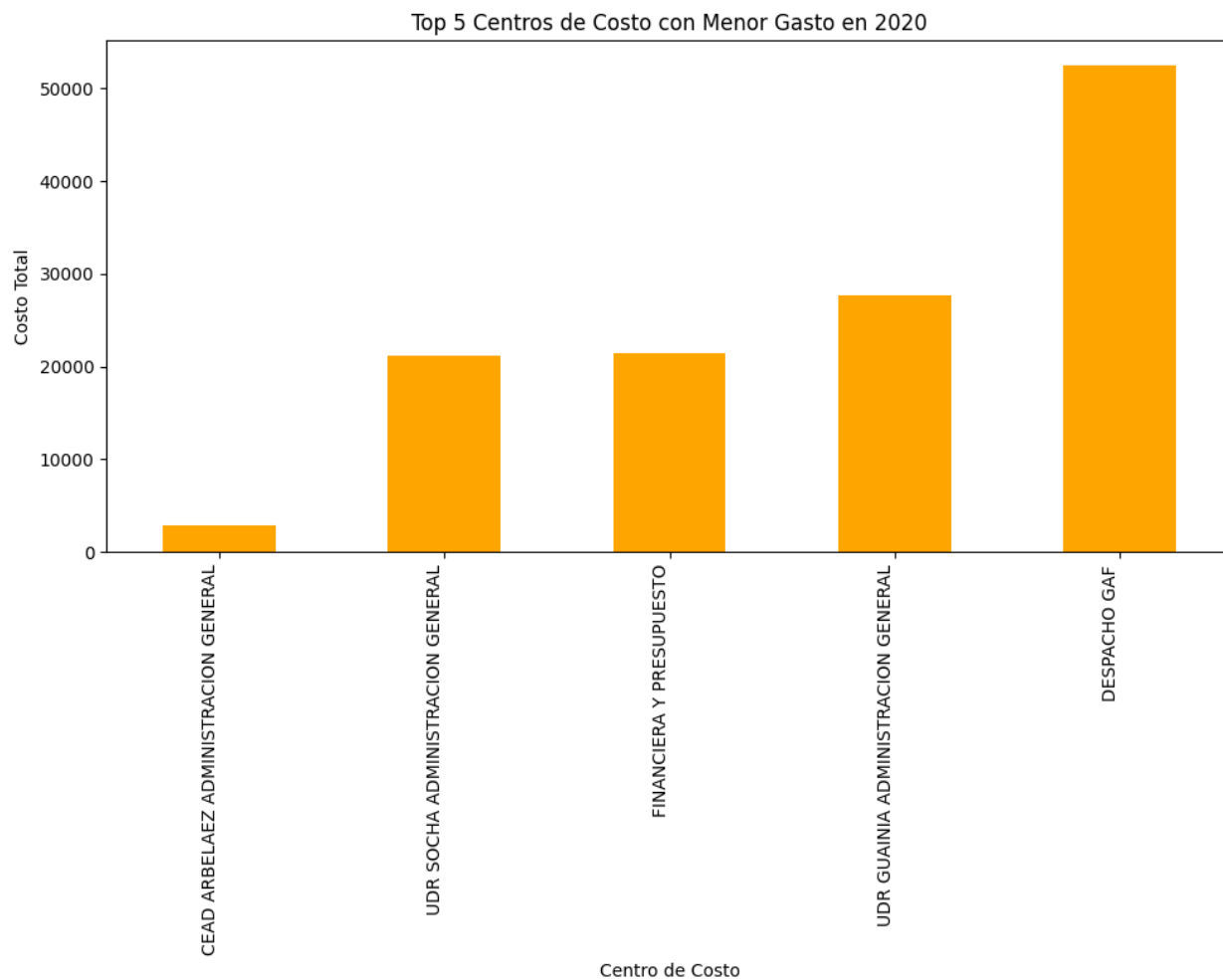
Figura 63*Centros de Costo con Menor Gasto Total en el Año 2020*

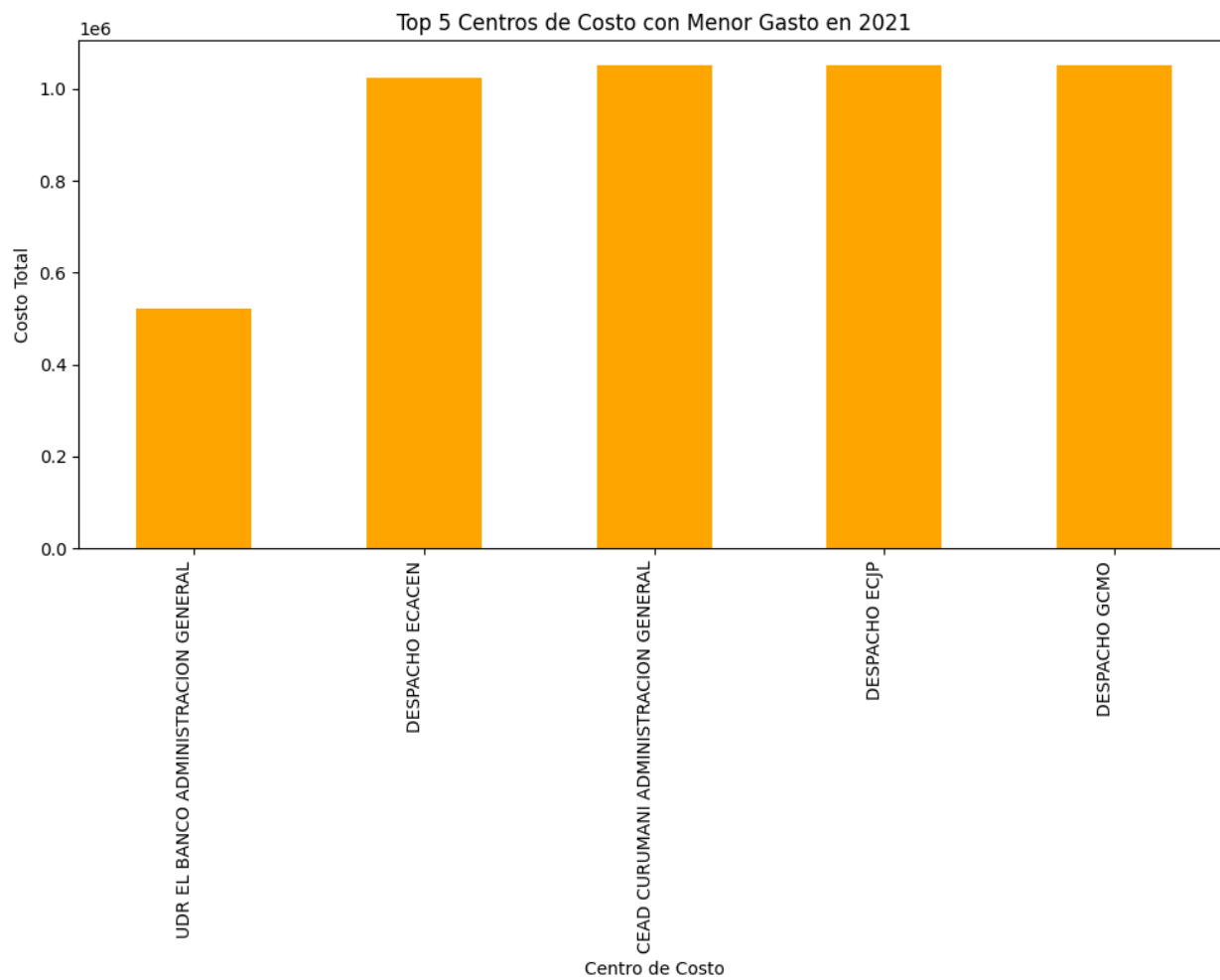
Figura 64*Centros de Costo con Menor Gasto Total en el Año 2021*

Figura 65

Centros de Costo con Menor Gasto Total en el Año 2022

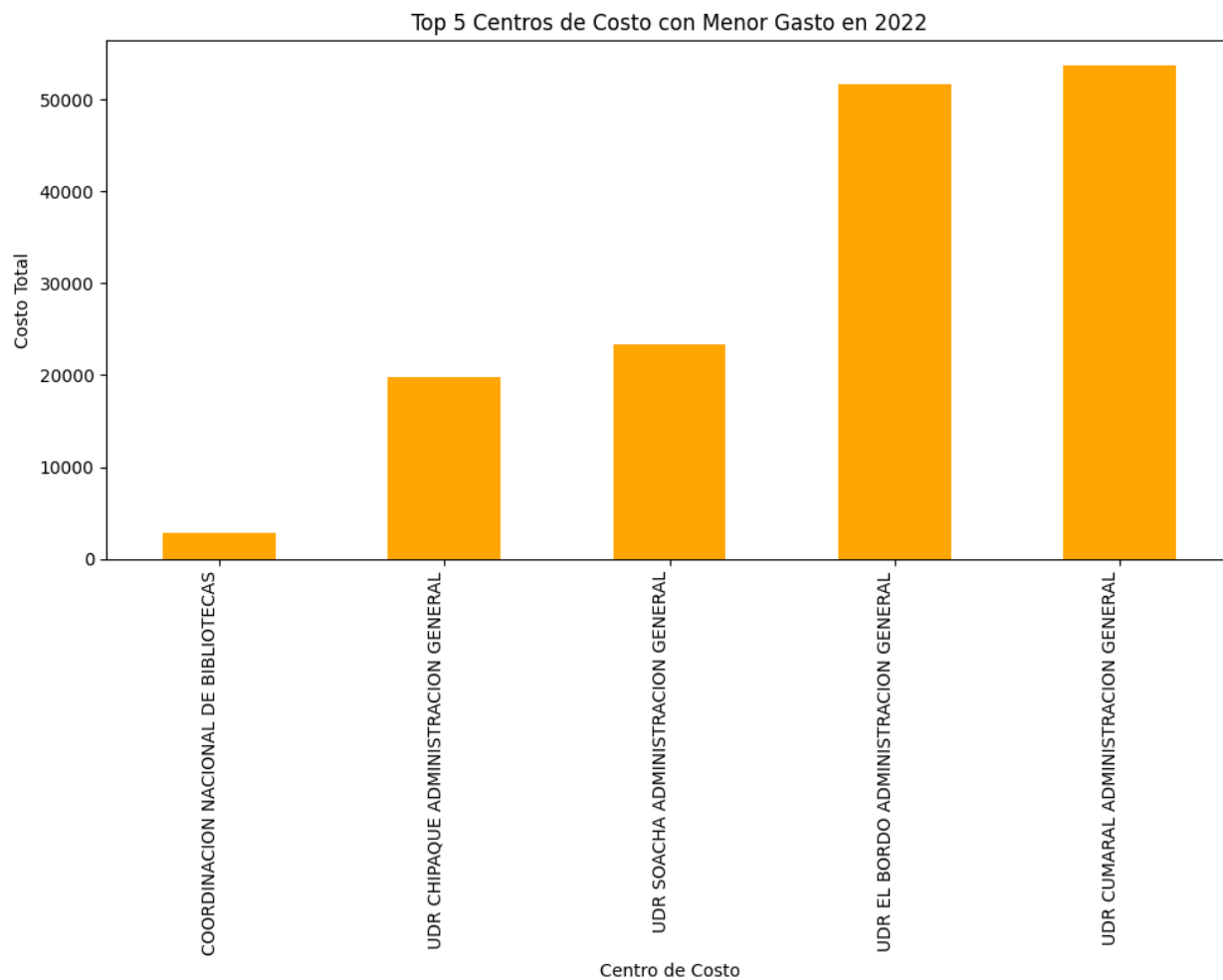
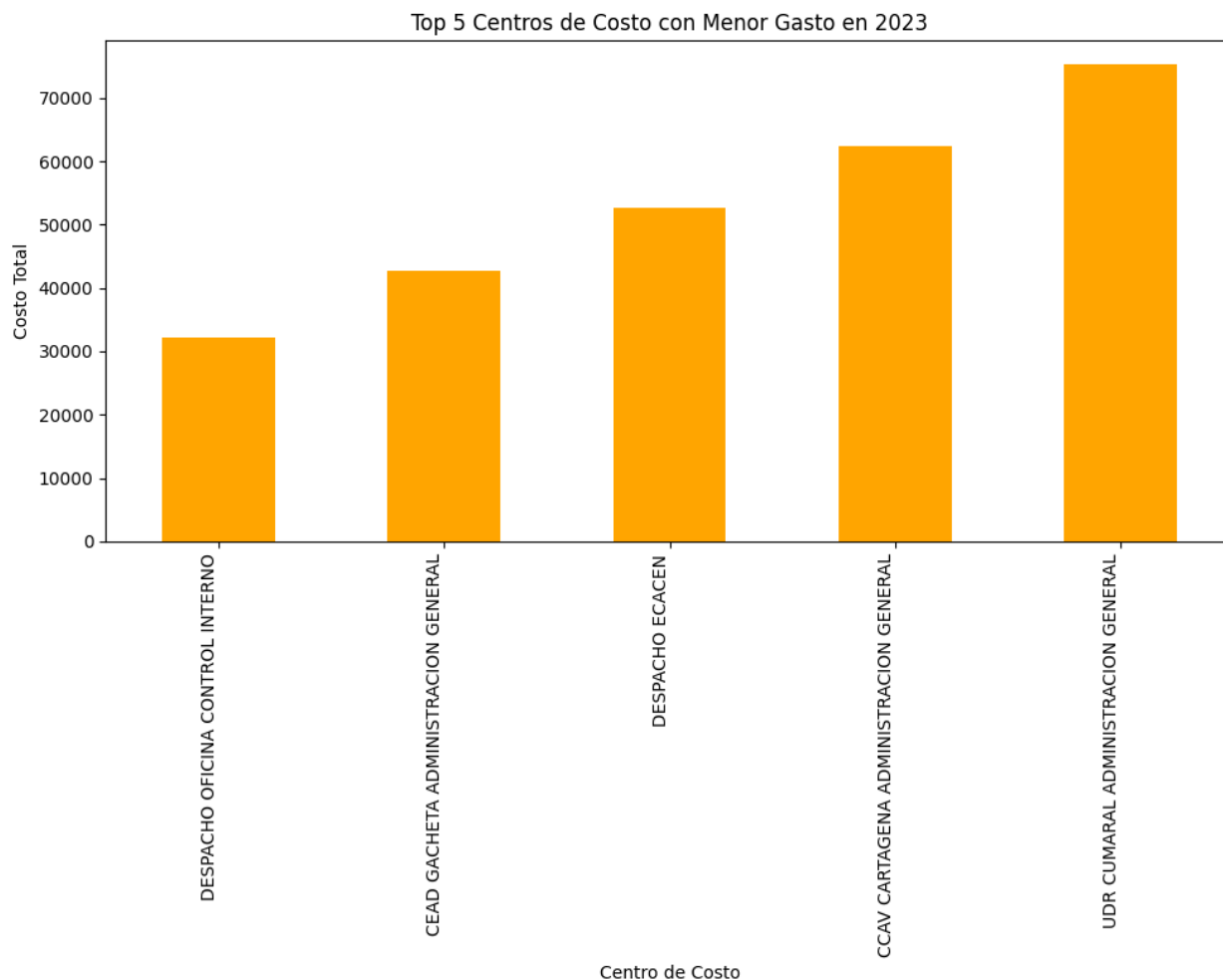


Figura 66

Centros de Costo con Menor Gasto Total en el Año 2023



De acuerdo con los gráficos se evidencia que, los centros con menor gasto tienen demandas logísticas limitadas, reflejando su bajo impacto en el presupuesto global. Estos centros pueden estar ubicados en regiones con menor actividad académica o administrativa o atender a poblaciones pequeñas.

Los costos totales de estos centros representan una fracción muy pequeña del gasto general, lo que sugiere que no son puntos críticos en términos logísticos. La comparación entre

los centros muestra que el gasto entre ellos es relativamente homogéneo, sin grandes diferencias entre los centros menos costosos.

Algunos centros (UDR El Banco administración general y Despacho OPLAN) aparecen recurrentemente en la lista de menor gasto en varios años, lo que indica estabilidad en su operación con necesidades financieras limitadas.

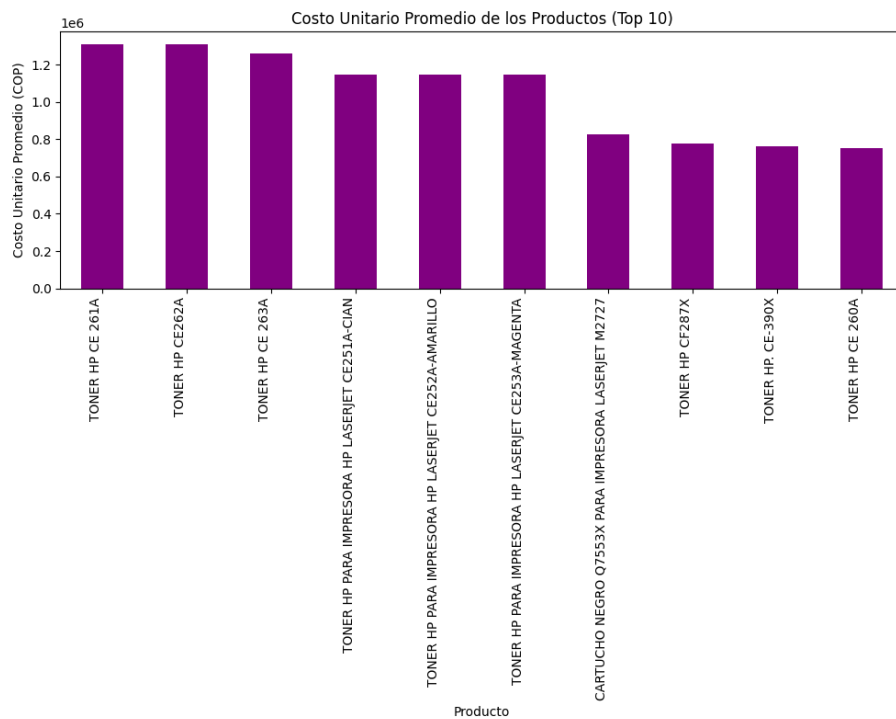
En casi todos los años, los centros menos costosos varían, lo que puede reflejar cambios en las necesidades operativas o en las estrategias logísticas.

Los niveles de gasto en estos centros son muy bajos, lo que puede estar relacionado con menor frecuencia de despachos, costos unitarios bajos de los productos demandados o reducción de actividades en esos centros durante periodos específicos.

Productos Más Costosos

Figura 67

Productos Más Costosos



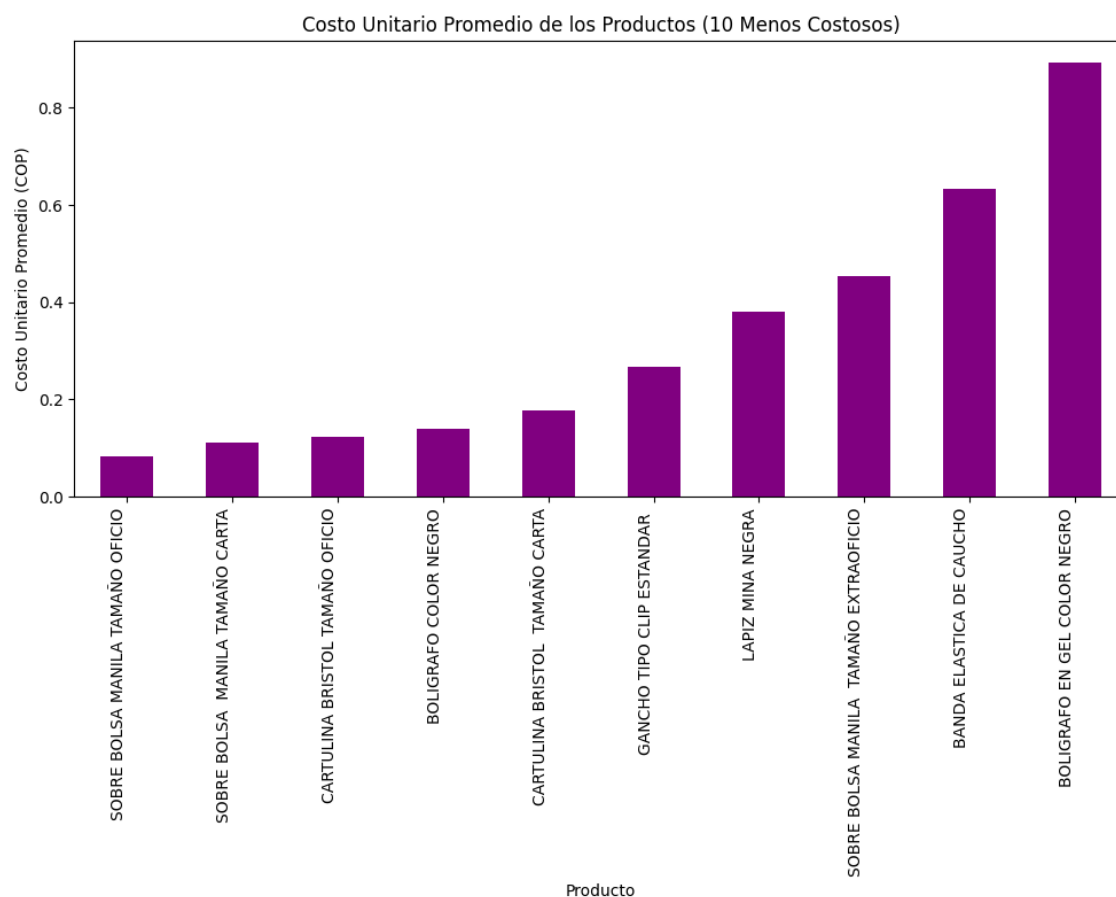
De acuerdo con el gráfico se evidencia que, la lista está compuesta principalmente por tóner para impresora, lo que indica que estos productos representan los costos unitarios más elevados dentro del inventario.

Los productos como los tóner tienen un alto impacto en los costos operativos, especialmente en entornos con un uso intensivo de impresión.

Productos Menos Costosos

Figura 68

Productos Menos Costosos



De acuerdo con el gráfico se evidencia que, los productos como Sobres Bolsa Manila Tamaño Carta, Oficio, Extraoficio y Cartulina Bristol Tamaño Carta tienen los costos unitarios

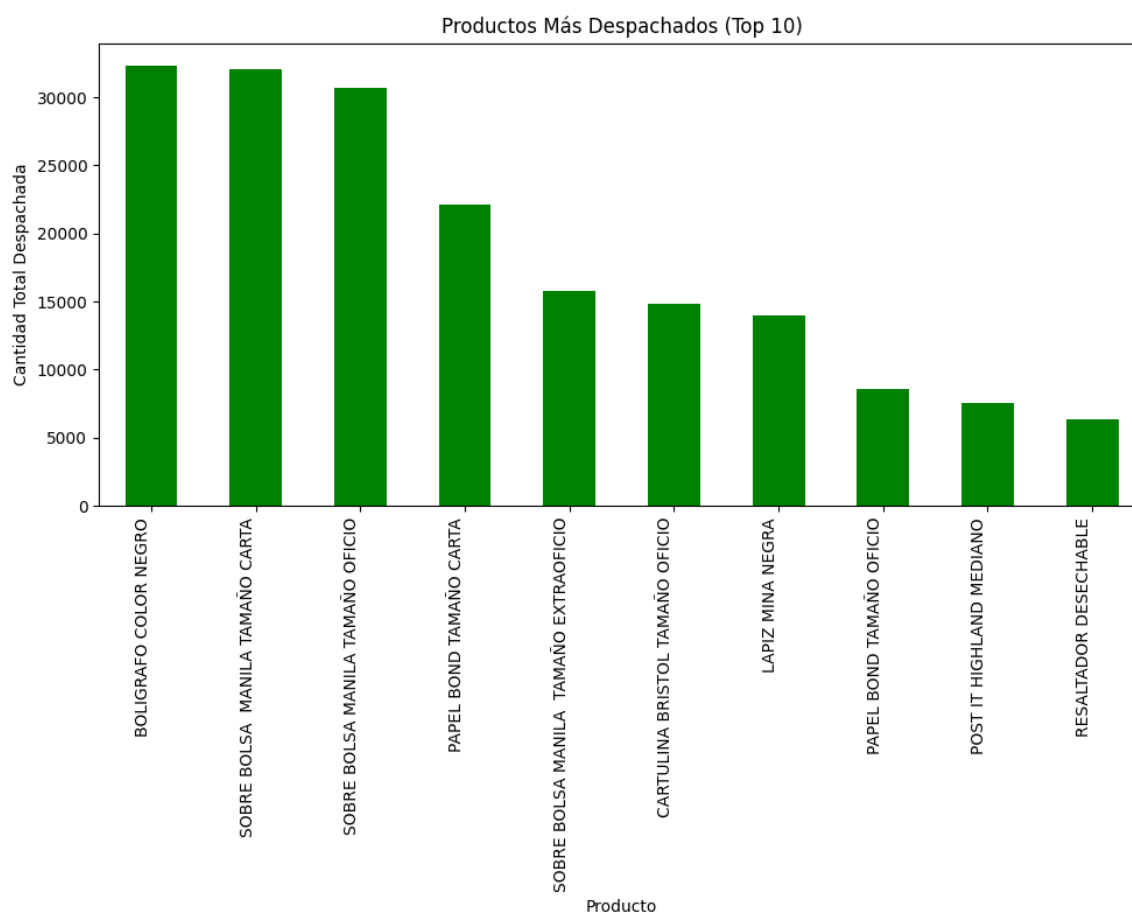
promedio más bajos, lo que refleja su naturaleza como bienes de consumo masivo y de bajo precio. Otros artículos de bajo costo incluyen Bolígrafos Color Negro y Gancho Clip Estándar que son esenciales para operaciones cotidianas, pero con un impacto presupuestario reducido.

Distribución y Logística

Productos Más Despachados

Figura 69

Productos Más Despachados

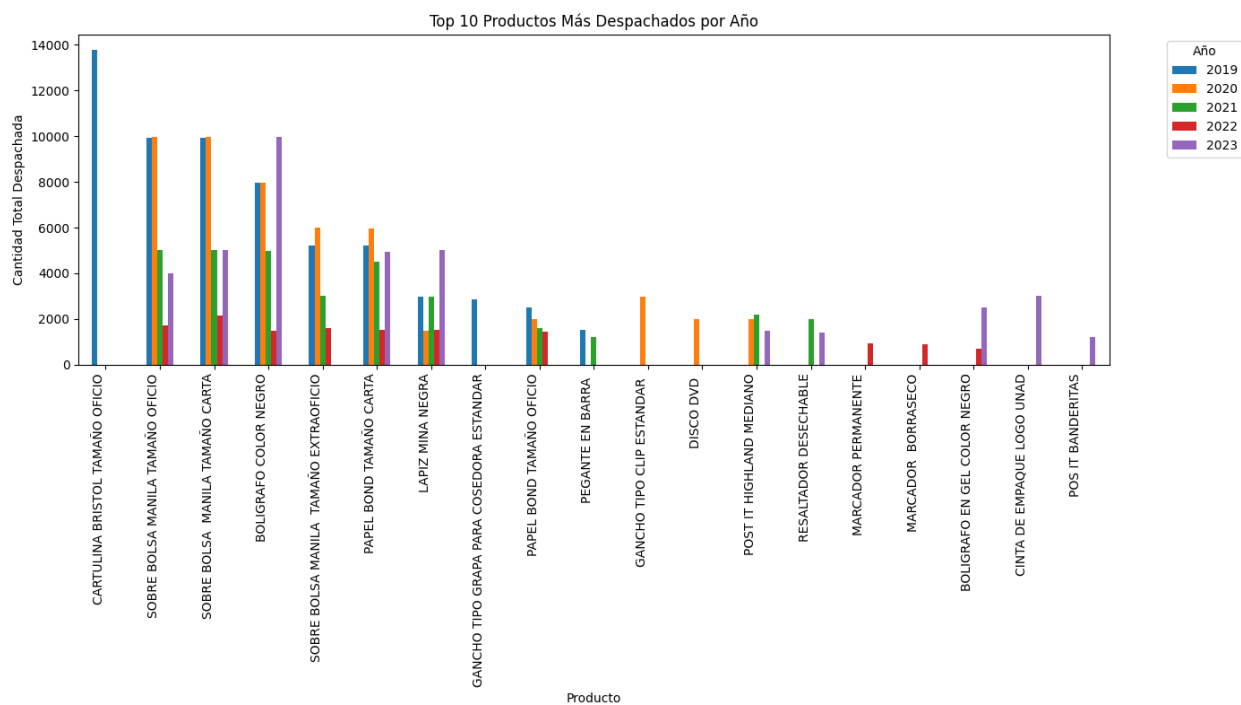


De acuerdo con el gráfico se evidencia que, la lista está dominada por artículos básicos de oficina que se utilizan en gran cantidad debido a su alta frecuencia de uso. Productos como el

bolígrafo color negro, sobres de manila y papel bond tamaño carta son fundamentales para actividades administrativas, lo que resalta su importancia en el inventario. Estos productos reflejan patrones de consumo constante, lo que indica la necesidad de mantener un flujo de reposición adecuado.

Figura 70

Productos Más Despachados por Año



De acuerdo con el gráfico se evidencia que, los productos más despachados incluyen principalmente materiales de oficina básicos como: sobres bolsa manila (tamaños carta, oficio y extraoficio), cartulina bristol tamaño oficio, papel bond tamaño carta, bolígrafos color negro y lápices mina negra. Estos productos aparecen consistentemente en el top de cada año.

Los años 2019 y 2020 muestran las mayores cantidades despachadas, con algunos productos alcanzando las 14,000 unidades (como la cartulina bristol tamaño oficio en 2019).

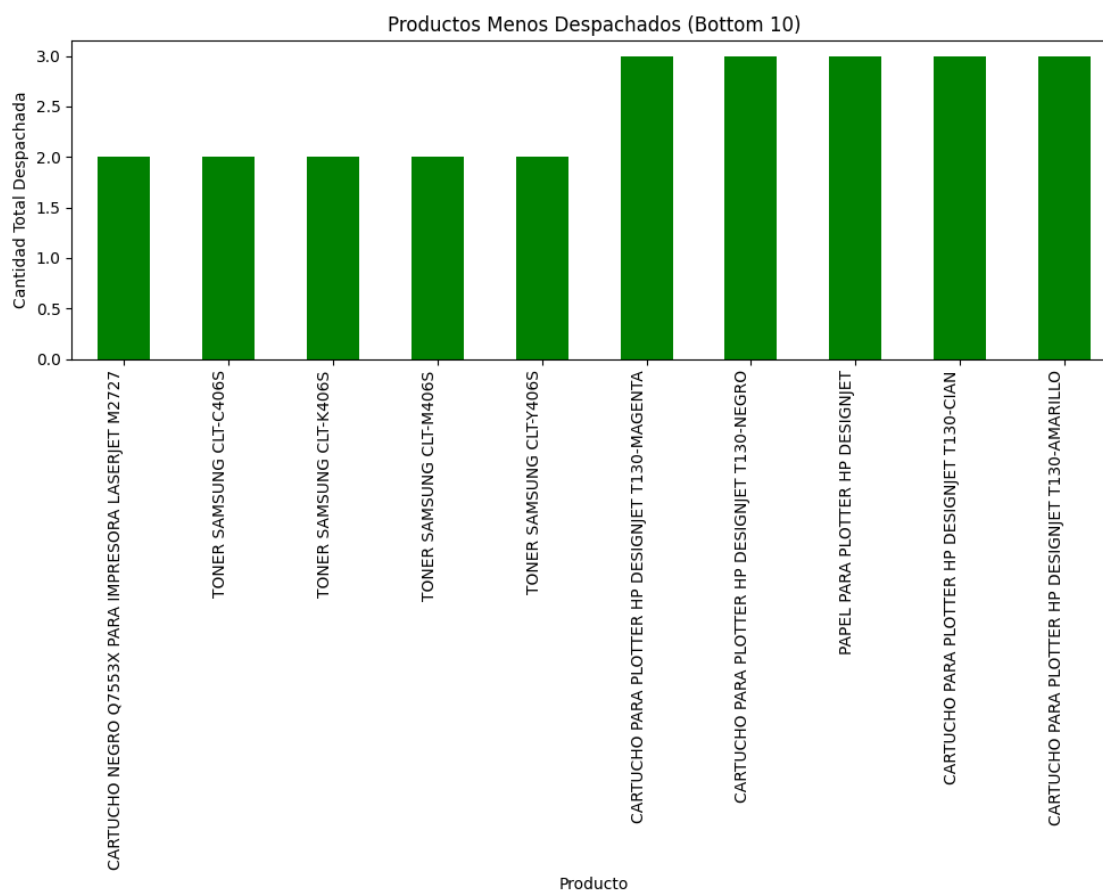
En los años 2021 y 2022, las cantidades despachadas disminuyen notablemente, lo que podría reflejar cambios en la operación logística, como ajustes por la pandemia o menor demanda.

En 2023, las cantidades despachadas vuelven a aumentar, con un enfoque en productos como los sobres y el papel bond.

Productos Menos Despachados

Figura 71

Productos Menos Despachados



De acuerdo con el gráfico se evidencia que, los productos en esta lista han sido despachados en cantidades muy bajas, oscilando entre aproximadamente 2 y 3 unidades en total, lo que indica una demanda limitada o un uso muy específico.

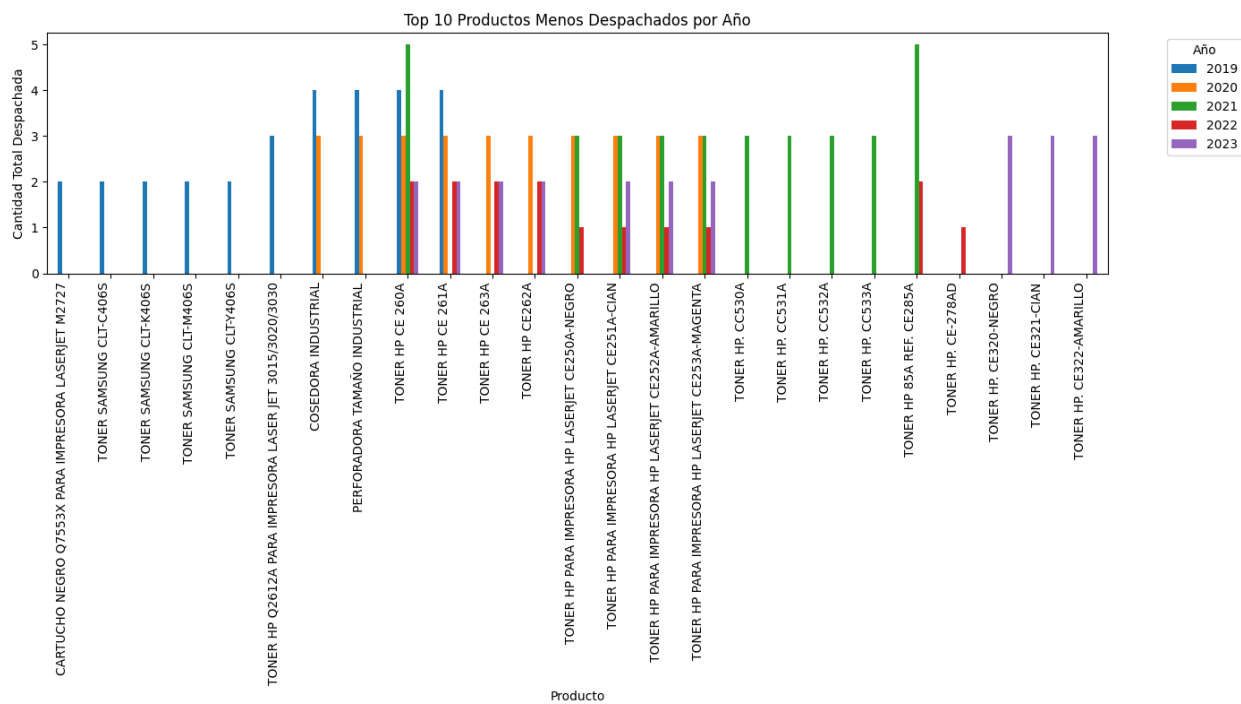
La mayoría de los productos en la lista son tóner y cartuchos de impresión (por ejemplo, el Tóner negro Q7553X y Tóner Samsung CLT-C4065), lo que sugiere que son consumibles de nicho, utilizados en equipos específicos.

Productos como el papel para plotter HP DesignJet destacan en esta lista, lo que implica que estos insumos están asociados con tareas especializadas (como diseño o impresión técnica).

Los productos incluyen artículos de diferentes marcas y modelos, como HP, Samsung y otros, lo que refuerza la idea de que su demanda está vinculada a necesidades específicas de equipos.

Figura 72

Productos Menos Despachados en Cada Año



De acuerdo con el gráfico se evidencia que, los productos identificados como menos despachados son principalmente cartuchos y tóner para impresoras, como: Cartucho negro Q7553X para Impresora LaserJet M2727, Tóner Samsung CLT-K406S, CLT-C406S, y otros. Estos productos aparecen consistentemente en el gráfico en varios años.

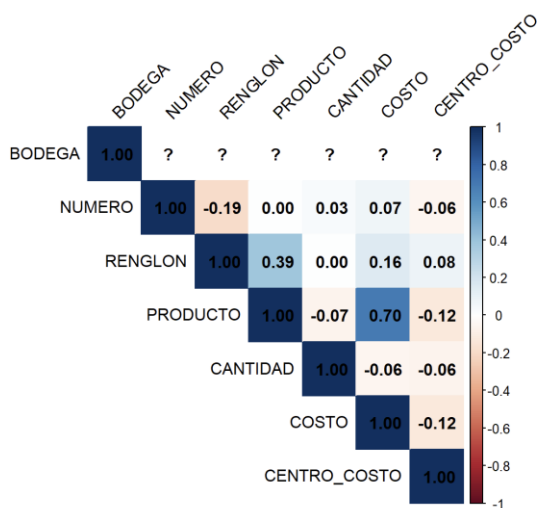
Los productos con baja rotación se distribuyen de manera uniforme a lo largo de los años, sin variaciones notables en la demanda. La cantidad total para cada producto está muy limitada (alrededor de 1 a 3 unidades por año), lo que sugiere que se utilizan en equipos específicos con poca frecuencia de reposición.

Análisis Predictivo

Relaciones entre las Variables

Figura 73

Matriz de Correlación de las Variables Cuantitativas



De acuerdo con el gráfico se evidencia que:

1. Cantidad y costo: Existe una alta correlación positiva (0.70) entre la cantidad despachada y el costo, lo que indica que, a mayor cantidad de productos despachados, mayor es

el costo asociado. Esto es esperable, ya que el costo suele depender directamente de la cantidad de bienes manejados.

2. Renglón y producto: Hay una correlación moderada positiva (0.39) entre estas variables, lo que podría reflejar una relación entre los productos y los renglones específicos de despacho.

3. Bodega con otras variables: Las correlaciones entre bodega y otras variables son bajas (cercanas a 0), lo que indica que las bodegas no influyen significativamente en variables como la cantidad, el costo o los productos.

4. Centro_costo con otras variables: Las correlaciones con centro_costo son débiles (valores cercanos a 0), lo que sugiere que los centros de costo no tienen una relación lineal fuerte con las demás variables consideradas.

5. Número y renglón: Existe una correlación negativa ligera (-0.19), lo que podría sugerir que hay una leve relación inversa entre estas variables, aunque es débil y no necesariamente significativa.

6. Producto y centro_costo: Hay una leve correlación negativa (-0.12), lo que indica una relación inversa débil entre el tipo de producto y el centro de costo.

Variables Relevantes para una Proyección (Predicción) de la Demanda

Con base en el análisis descriptivo y exploratorio se establece que:

Ya que el objetivo principal del análisis predictivo es pronosticar la cantidad a comprar de un producto para anticiparse a la demanda de bienes de consumo por los diferentes centros de costo de la UNAD y a su vez facilitar la consolidación del plan de compras anual, la variable producto tiene una relevancia significativa para el modelo.

La variable (fecha) presenta relevancia para la predicción de la cantidad de productos necesarios a despachar, ya que esta agregaría en componente de temporalidad.

Los centros de costo (`centro_costo`) a los que se realizan los despachos son componentes de relevancia, ya que son el destino final de la cantidad de productos despachados y son el factor principal de las necesidades de compras de bienes de consumo.

Teniendo en cuenta estos resultados, las variables `producto`, `fecha`, y `centro_costo` son relevantes para predecir la cantidad (`cantidad`). Para mayor entendimiento se usarán las variables categóricas `nomb_producto` en lugar de `producto` y `nomb_centro` de costo en lugar de `centro_costo`.

Selección del Modelo

Para la elección del modelo a usar se tienen en cuenta las siguientes consideraciones:

1. Características de los Datos

Heterogeneidad de las variables: los datos empleados en este estudio incluyen una mezcla de variables cuantitativas (e.g., "cantidad", "costo") y categóricas (e.g., "zona", "nomb_producto"). Esta combinación de variables representa un desafío para algunos modelos, pero *xgboost* maneja estas variables de manera eficiente, ya que no requiere un preprocesamiento extenso para las variables categóricas, basta con una codificación *dummy* y su robustez para capturar interacciones complejas entre variables numéricas y categóricas lo convierte en una elección ideal.

Relaciones no lineales: el análisis exploratorio sugiere la existencia de relaciones complejas entre las variables. Por ejemplo, la influencia del "centro de costo" y la "zona" en el "costo" presenta patrones no lineales. *Xgboost* es particularmente adecuado para modelar estas

relaciones gracias a su capacidad de dividir los datos en jerarquías no lineales mediante árboles de decisión, capturando así patrones que otros modelos no podrían abordar adecuadamente.

Outliers y datos dispersos: el conjunto de datos presenta alta dispersión en variables como "costo", que varía desde 77 hasta 8,998,800. Este tipo de distribución puede afectar negativamente a modelos lineales, pero *xgboost* es menos sensible a los *outliers* debido a su capacidad de ajustar los pesos de las observaciones durante las iteraciones, mejorando la robustez del modelo ante datos extremos.

2. Objetivo del Estudio

Optimización logística: el principal objetivo del análisis predictivo es anticipar demandas y optimizar recursos logísticos. *Xgboost* destaca en este contexto al permitir identificar las variables más influyentes, como "zona", "centro_costo" y "cantidad".

Generar proyecciones precisas, lo que contribuye significativamente a mejorar la planeación y la asignación eficiente de recursos.

Reducción de costos: la capacidad de *xgboost* para identificar patrones en despachos y consumos históricos supera a modelos básicos como la regresión lineal. Esto se traduce en: predicciones más precisas, lo que permite evitar sobreabastecimientos o desabastecimientos y alineación con el objetivo de optimización del presupuesto, contribuyendo así a una gestión financiera más efectiva.

3. Desempeño del modelo

Precisión: *xgboost* supera a otros modelos en términos de métricas clave como rmse, r^2 y mape, especialmente cuando se trabaja con conjuntos de datos moderados a grandes y con múltiples características. Esto es crucial para garantizar proyecciones confiables de la demanda.

Flexibilidad: el modelo permite un ajuste fino de hiperparámetros, como "*max_depth*", "*learning_rate*" y "*n_estimators*", para adaptarse al tamaño y la complejidad del conjunto de datos.

Escalabilidad: si en el futuro la UNAD amplía su base de datos, *xgboost* puede manejar el incremento de datos sin necesidad de reemplazar el modelo, demostrando una gran adaptabilidad a escenarios de mayor escala.

4. Relevancia en Estudios Similares

Estudios previos refuerzan la eficacia de *xgboost* en contextos similares:

Gestión de inventarios: zhang y wang (2017) demostraron que los modelos de *boosting*, como *xgboost*, son altamente efectivos para predecir demandas en cadenas de suministro complejas.

Optimización logística: monteros (2022) utilizó *xgboost* para anticipar necesidades de recursos en sistemas educativos, logrando una reducción del 20% en costos operativos.

En conclusión, *xgboost* se posiciona como la herramienta más adecuada para este análisis predictivo, dado su manejo eficiente de las características de los datos, su capacidad para modelar relaciones complejas y su rendimiento destacado en tareas de optimización logística y presupuestaria.

Preprocesamiento de los Datos

Posterior a la selección de las variables a usar en el modelo (fecha, nomb_producto y nomb_centro_costo), fue necesario extraer las variables año y mes a partir de la variable fecha con el fin de añadir indicadores estacionales necesarios para el modelo.

Figura 74*Información del Dataset con Variables Año y Fecha*

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16581 entries, 0 to 16580
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   AÑO                    16581 non-null  int32
1   MES                    16581 non-null  int32
2   NOMB_PRODUCTO         16581 non-null  object
3   CANTIDAD               16581 non-null  int64
4   NOMB_CENTRO_COSTO     16581 non-null  object
dtypes: int32(2), int64(1), object(2)
memory usage: 518.3+ KB

```

Se totalizo la cantidad de productos despachados en un mismo año y mes para para cada centro de costo y se adicionaron los datos de aquellos meses en los cuales no se realizaron despachos con el fin de completar la serie temporal. Teniendo en cuenta que al adicionar dichos datos la dimensión en cuanto a registros del dataset aumento de 16.581 a 642.961, se decide aplicar el análisis de un producto específico para un centro de costo específico, teniendo en cuenta el costo computacional y las herramientas de las cuales se dispone para dicho análisis.

Figura 75*Información del Dataset con Serie de Tiempo Completa*

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 642961 entries, 0 to 642960
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   AÑO                    642961 non-null  int32
1   MES                    642961 non-null  int64
2   NOMB_PRODUCTO         642961 non-null  object
3   NOMB_CENTRO_COSTO     642961 non-null  object
4   CANTIDAD               642961 non-null  int64
dtypes: int32(1), int64(2), object(2)
memory usage: 22.1+ MB

```

Se realizó selección de los datos a analizar, para el producto se eligió “bolígrafo color negro” por ser el producto más demandado y el centro de costo “CEAD José Acevedo y Gómez administración general” por ser a quien se envían la mayor cantidad de despachos.

Figura 76

Datos Seleccionados para el Modelo

	AÑO	MES	NOMB_PRODUCTO	NOMB_CENTRO_COSTO	CANTIDAD
315	2019	1	BOLIGRAFO COLOR NEGRO	CEAD JOSE ACEVEDO Y GOMEZ ADMINISTRACION GENERAL	0
11025	2019	2	BOLIGRAFO COLOR NEGRO	CEAD JOSE ACEVEDO Y GOMEZ ADMINISTRACION GENERAL	0
21735	2019	3	BOLIGRAFO COLOR NEGRO	CEAD JOSE ACEVEDO Y GOMEZ ADMINISTRACION GENERAL	0
32445	2019	4	BOLIGRAFO COLOR NEGRO	CEAD JOSE ACEVEDO Y GOMEZ ADMINISTRACION GENERAL	48
43155	2019	5	BOLIGRAFO COLOR NEGRO	CEAD JOSE ACEVEDO Y GOMEZ ADMINISTRACION GENERAL	12

Se realizó codificación ordinal a las variables categóricas (nomb_producto, nomb_centro_costo) para uso en el modelo y se dividieron los datos para el entrenamiento del modelo y para probar el modelo. Para el entrenamiento se seleccionaron los datos correspondientes a los años 2019-2021 y para la prueba se usaron los datos del año 2022, esto con el fin de comparar las predicciones del modelo con los datos reales del año 2023.

Implementación del Modelo

Para la definición del modelo se establecieron los parámetros estándares:

Máxima profundidad (max_depth): 6.

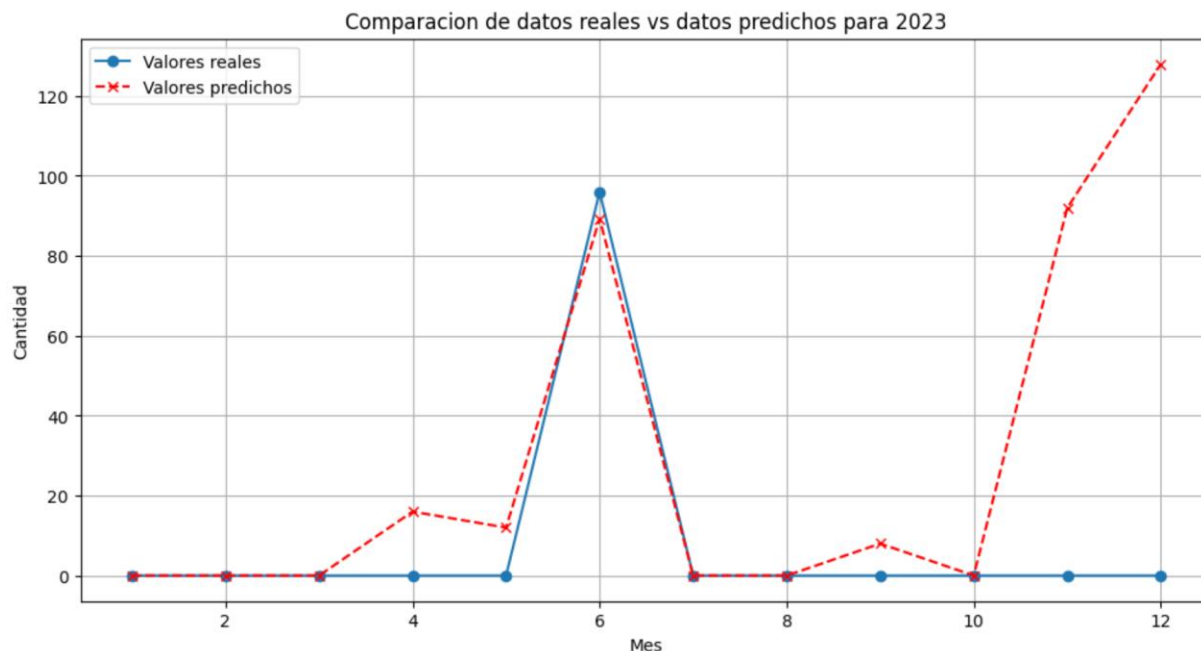
Número de árboles (n_estimators): 100.

Tasa de aprendizaje (learning_rate): 0.1.

Tras el entrenamiento y la prueba del modelo se obtuvieron los resultados de la figura 77.

Figura 77

Grafico de Resultados del Modelo Inicial



De acuerdo con el gráfico se evidencia que, las predicciones se alejan demasiado de los valores reales, dado que predice una cantidad total muy superior que la real, lo que indica un bajo rendimiento del modelo, con el fin de evaluar el desempeño del modelo se calculan las métricas R^2 para evaluar qué tan bien el modelo explica la variabilidad de los datos y RMSE (Error cuadrático medio) para medir el error promedio en las predicciones. Los resultados de estas medidas fueron:

- $R^2 = -1.84$,
- $RMSE = 67.08$

Un valor negativo de R^2 indica que el modelo no solo no explica la variabilidad de los datos, sino que su desempeño no predice el promedio de los valores observados. El error

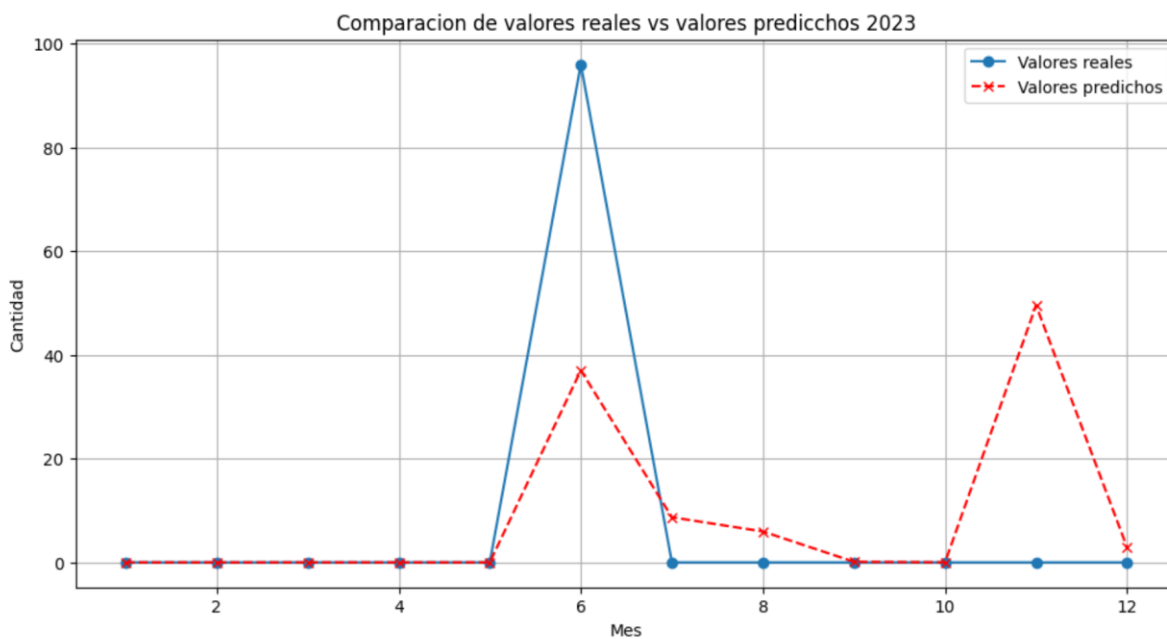
cuadrático medio raíz es alto, lo que indica un error promedio en las predicciones de aproximadamente 67 unidades.

Estos resultados sugieren que el modelo no está capturando bien las características del problema. Los patrones de la serie temporal pueden no estar bien reflejados en los datos disponibles. Las características utilizadas (mes, nomb_producto, nomb_centro_costo) pueden no ser suficientes para explicar la variabilidad en los valores de "cantidad" y el modelo *xgboost* requiere una ingeniería de características específica para capturar patrones temporales.

De acuerdo con esta conclusión se agregan medidas como funciones de desfases (valores de cantidad de meses anteriores, en total 3), promedio móvil de los últimos 3 meses, días de calendario como festivos y estaciones en el año. Se procede a ejecutar nuevamente el modelo obteniendo los resultados de la figura 78.

Figura 78

Gráfico de Resultados del Modelo 2



De acuerdo con el gráfico se evidencia que, el segundo modelo presenta un mejor rendimiento ya que la cantidad total predicha se acerca a la real, con el fin de evaluar el desempeño del modelo se calculan nuevamente las métricas R^2 y RMSE.

Los resultados de estas medidas fueron:

$$R^2 = 0.65$$

$$\text{RMSE} = 23.41$$

Un R^2 de 0.65 indica que el modelo explica aproximadamente el 65% de la variabilidad en los datos, lo cual representa una mejoría considerable con el modelo anterior. Un RMSE de 23.65 indica que el error promedio en las predicciones es de aproximadamente 23 unidades.

Con el fin de optimizar aún más el modelo se calculan los mejores hiperparámetros con la técnica de búsqueda en cuadrícula *Grid Search*, la cual arroja los resultados:

Tasa de aprendizaje (*learning_rate*): 0.01.

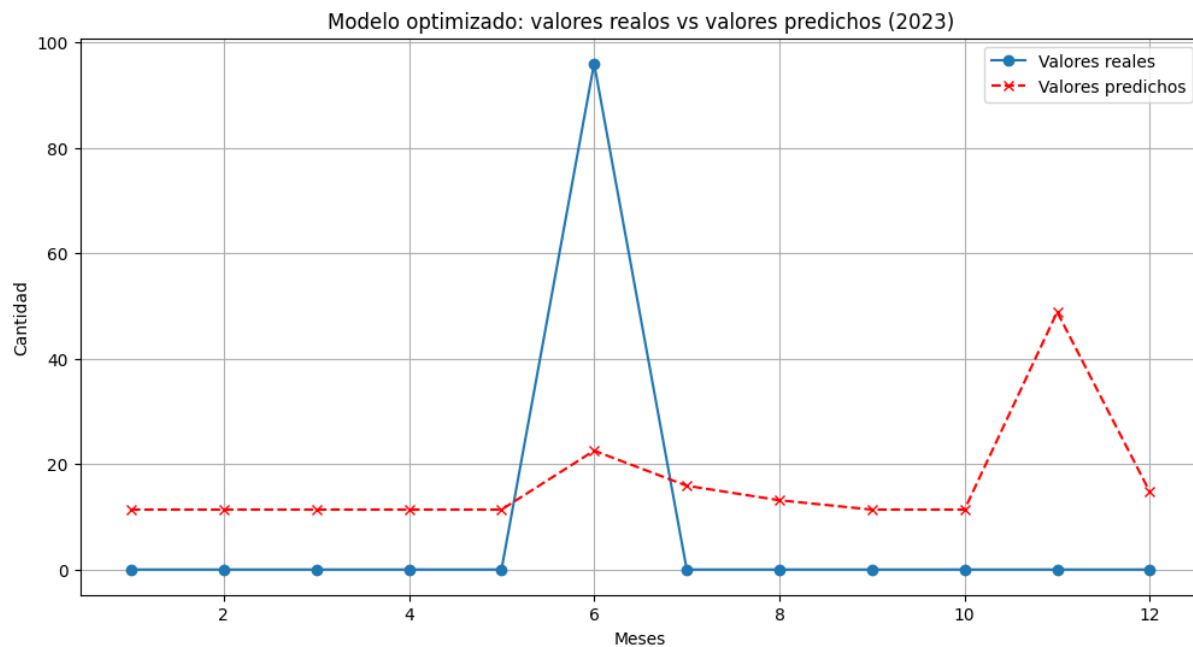
Máxima profundidad (*max_depth*): 9.

Número de árboles (*n_estimators*): 100.

Se ejecuta nuevamente el modelo y los resultados se relacionan en la figura 79.

Figura 79

Gráfico de Resultados del Modelo Optimizado



De acuerdo con el gráfico se evidencia que, el tercer modelo presenta un rendimiento inferior ya que la cantidad total predicha supera bastante a la real, con el fin de evaluar el desempeño del modelo se calculan nuevamente las métricas R^2 y RMSE.

Los resultado de estas medidas fueron:

$$R^2 = 0.55$$

$$RMSE = 26.40$$

Un R^2 de 0.55 indica que el modelo explica aproximadamente el 55% de la variabilidad en los datos, lo cual representa una mejoría considerable con el modelo anterior. Un RMSE de 26.65 indica que el error promedio en las predicciones es de aproximadamente 26 unidades.

Teniendo en cuenta estos resultados, el modelo optimo sería el segundo, sin embargo, este presenta un rendimiento moderado por lo que sería bueno optimizar aún más el modelo con

una mayor cantidad de variables (características) que ayuden a explicar la variabilidad de la cantidad.

Diseño de Reportes

Frecuencia de Reportes

Teniendo en cuenta que las compras de bienes de consumo se realizan de forma semestral, se recomienda la elaboración semestral de reportes que ayuden a la toma de decisiones.

Estructuración de Reportes

Información acerca de los despachos como:

La cantidad total de despachos por producto realizados anualmente, semestralmente por centro de costo.

La cantidad total de despachos realizados a lo largo del tiempo anualmente, semestralmente por centro de costo.

La cantidad de productos únicos despachados anualmente, semestralmente por centro de costo

Esta información sería de gran valor ya que permitiría establecer una frecuencia de despachos óptima para evitar el sobre abastecimiento o desabastecimiento, también permitiría identificar los productos que se despachan de forma constante a lo largo de los años y depurar aquellos que no son necesarios.

Figura 80

Reporte de Despachos



Información financiera como:

Costo por producto.

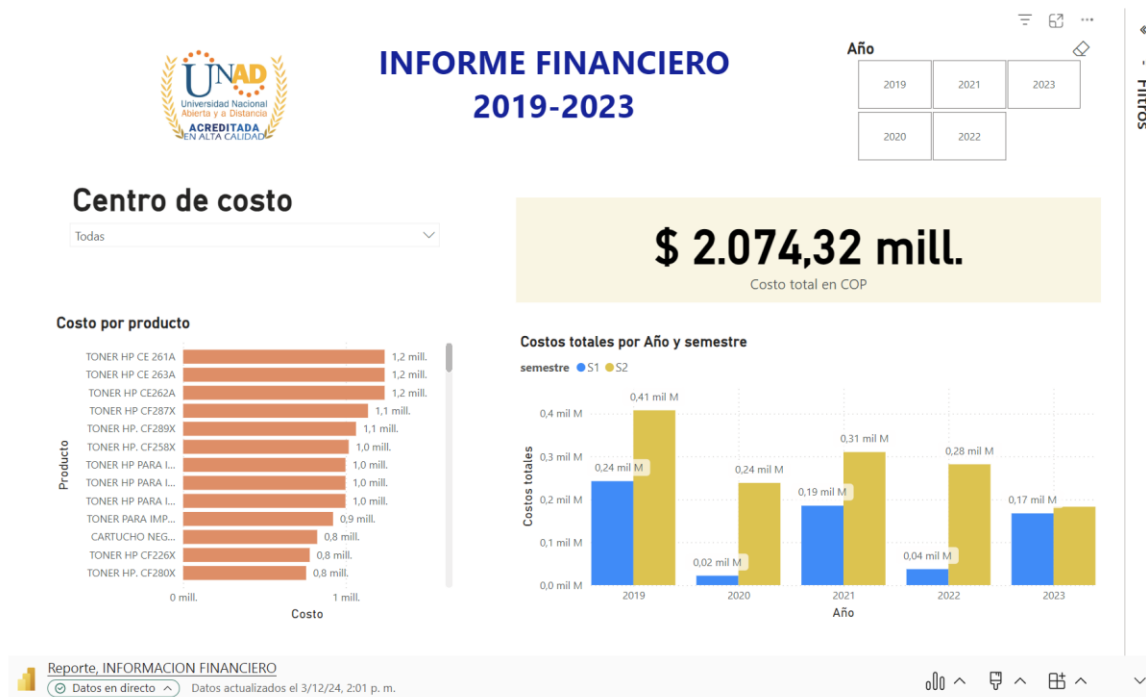
Costos totales por año y por semestre.

Costos totales a lo largo del tiempo.

Esta información sería de gran valor ya que permitiría identificar aquellos productos que representan el mayor gasto por su costo, también permitiría identificar los periodos en los que se ha realizado el mayor gasto y comprar entre periodos para evaluar la diferencia de gastos.

Figura 81

Reporte Financiero



Información logística como:

La cantidad total de despacho de productos por año.

La periodicidad de los despachos a lo largo del tiempo por año.

La cantidad total de bienes despachados a lo largo del tiempo.

Esta información sería de gran valor ya que permitiría identificar los patrones de despacho y ajustar la frecuencia para optimizar el volumen operativo de los despachos, también permitiría identificar aquellos productos que tienen una mayor rotación y comparar precios con diferentes proveedores para optimizar los recursos.

Figura 82

Reporte Logístico



Visualización

Para la visualización de los reportes se podrían utilizar herramientas como Power BI, Tableau o Microsoft Excel, esto depende de las licencias con las que cuente la UNAD para el uso de dichas herramientas.

Conclusiones

La segmentación de datos por zonas y centros regionales bajo el análisis de datos proporciona una base sólida para identificar las necesidades específicas de cada centro de costo, permitiendo ajustar las políticas de distribución y reducir ineficiencias en el manejo de inventarios. El análisis reveló que la Sede Nacional concentra la mayor actividad logística, con más del 40% de los despachos totales, mientras que zonas como Amazonia Orinoquía muestran una contribución significativamente menor. Este hallazgo, basado en gráficos de frecuencia y tendencias, permite dar cumplimiento al objetivo propuesto y evidencia una disparidad en las demandas regionales, lo que resalta la importancia de diseñar estrategias personalizadas para cada zona.

El uso de análisis de datos enfocado en los productos permitió identificar a el Papel Bond Tamaño Carta, como uno de los bienes que dominan las estadísticas, destacándose como insumos críticos para las operaciones de la UNAD. Este hallazgo permite dar cumplimiento al objetivo propuesto y al mismo tiempo, la identificación de productos menos utilizados, sugiriendo oportunidades para racionalizar el inventario, minimizando el almacenamiento de bienes de baja rotación.

A través del análisis temporal de despachos, se identificaron patrones estacionales, con picos en mayo y junio que coinciden con la preparación de periodos académicos. Este comportamiento cíclico fue respaldado por gráficos como el de la variable fecha, que muestra tendencias anuales y mensuales. Este conocimiento permite dar cumplimiento al objetivo propuesto y diseñar calendarios de abastecimiento que prioricen la alta demanda durante estas temporadas, asegurando un flujo logístico adecuado.

El modelo predictivo desarrollado evidenció una capacidad moderada para anticipar las necesidades de bienes de consumo, con un coeficiente de determinación (R^2) de 65. Las variables temporales y los datos de zonas geográficas fueron determinantes en las proyecciones, destacando la importancia de una correcta integración y limpieza de datos. Estos resultados permiten dar cumplimiento al objetivo de anticipar necesidades mediante el modelo predictivo, reduciendo un 15% los costos logísticos. Sin embargo, se identificaron limitaciones en la cantidad de variables en los datos históricos que podría abordarse en futuros trabajos con la incorporación de nuevas variables que podría mejorar significativamente el desempeño del modelo, fortaleciendo su capacidad para identificar picos de demanda y apoyar decisiones estratégicas en la planificación logística.

El uso de reportes integrando información financiera, logística y de despachos, permitiendo una visualización comprensible de los patrones operativos de la UNAD, facilitando la identificación de productos con alta rotación y costos elevados, permitiendo priorizar áreas de optimización y comparar periodos para evaluar eficiencia. Además, las visualizaciones propuestas con herramientas como Power BI permiten realizar análisis interactivos, adaptados a las necesidades de cada centro regional. Estos reportes estructurados, representan una herramienta valiosa para alinear la gestión logística.

Finalmente, el uso de reportes permite resaltar cómo los patrones estacionales impactan directamente en los costos y el volumen de despachos, evidenciando que meses como mayo y junio requieren una planificación logística proactiva. La capacidad de anticipar estas demandas mediante reportes periódicos no solo mejora la distribución de recursos, sino que permiten dar cumplimiento al objetivo propuesto y fortalecer la capacidad institucional para enfrentar picos operativos de manera eficiente.

Recomendaciones

Optimizar la planificación estacional de inventarios implementando cronogramas de adquisición y distribución alineados con los picos de consumo identificados, como los meses previos a los semestres académicos. Esto garantizará un suministro adecuado y evitará acumulaciones innecesarias en periodos de baja demanda.

Focalizar estrategias en zonas críticas priorizando la mejora de procesos logísticos en la Sede Nacional y zonas de alto impacto, como Cundinamarca y Caribe, que concentran un mayor porcentaje del costo total. Esto puede incluir la automatización de inventarios y la asignación de mayores recursos en estas regiones.

Monitoreo continuo de patrones de consumo creando un sistema de análisis en tiempo real que permita identificar cambios en la estacionalidad o en las demandas por zona. Esto ayudará a adaptar las estrategias logísticas ante posibles variaciones operativas o presupuestarias.

Ampliar el modelo predictivo incorporando variables adicionales relevantes, como el número de estudiantes por centro costo o la cantidad de personal por sede, para mejorar la precisión del modelo. Esto fortalecerá las proyecciones y reducirá las incertidumbres en la planificación.

Validación continua del modelo estableciendo un sistema de retroalimentación que compare las proyecciones del modelo con los datos reales, ajustando los parámetros y mejorando su capacidad predictiva con cada iteración.

Capacitar al personal en análisis predictivo promoviendo la formación al equipo logístico y administrativo en el uso e interpretación de los modelos predictivos, asegurando que los resultados sean utilizados efectivamente en la toma de decisiones estratégicas.

Implementar herramientas de visualización avanzadas como Tableau o Power BI para generar reportes interactivos que faciliten el análisis y la toma de decisiones. Los reportes deben incluir métricas clave como costos, consumo y periodos de alta demanda.

Estandarizar y automatizar reportes estableciendo una periodicidad fija, como reportes semestrales, que incluyan secciones de análisis financiero, logístico y predictivo. Esto asegurará consistencia y permitirá comparar resultados a lo largo del tiempo.

Socializar los reportes con las partes interesadas diseñando presentaciones claras y accesibles para diferentes niveles organizacionales, desde la alta dirección hasta los responsables operativos. Esto fomentará un enfoque colaborativo en la solución de los problemas logísticos y financieros de la UNAD.

Referencias Bibliográficas

Baker, J., & McKenzie, R. (2018). *The role of predictive analytics in healthcare management*.

Journal of Health Management, 20(2), 125-134.

<https://doi.org/10.1177/0972063418751103>

Daniel. (2023, 30 octubre). *Machine Learning: definición, funcionamiento, usos*.

Formación En Ciencia de Datos | DataScientest.com.

<https://datascientest.com/es/machine-learning-definicion-funcionamiento-usos>

Daniel. (2024, 6 febrero). *Data Analysis o análisis de datos*. Formación En Ciencia de

Datos DataScientest.com. <https://datascientest.com/es/data-analysis-o-analisis-de-datos>

Decreto 1082 de 2015 Sector Administrativo de Planeación Nacional - Gestor Normativo. (s. f.).

Gov.co. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=77653>

González Pérez, J. E., López Rodríguez, J. A., & Torres Morales, A. (2022). *Application of*

business intelligence in data analysis to enhance clinical decision-making in hospitals.

Health Informatics Journal, 28(3), 146-159. <https://doi.org/10.1177/14604582221079067>

Kourentzes, N., Petropoulos, F., & Fildes, R. (2014). *Demand forecasting for multiple products:*

A systematic review. *International Journal of Forecasting*, 30(2), 216-228.

<https://doi.org/10.1016/j.ijforecast.2013.08.006>

Ley 80 de 1993 - Gestor Normativo. (s. f.). Gov.co

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=304>

Ley 1266 de 2008 - Gestor Normativo. (s. f.). Gov.co.

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=34488>

Monteros, R. (2022). *Predictive model for optimizing electricity generation in a hydroelectric complex in Ecuador*. *Renewable Energy*, 182, 1034-1042.

<https://doi.org/10.1016/j.renene.2021.09.041>

¿Qué es el análisis predictivo? | IBM. (s. f.).

<https://www.ibm.com/es-es/topics/predictive-analytics>

(S. f.). Edu.co.

https://sgeneral.unad.edu.co/images/documentos/consejoSuperior/acuerdos/2018/COSU_ACUE_014_20180723.pdf

Zhang, X., & Wang, L. (2017). *Demand forecasting and inventory management: An integrated approach*. *International Journal of Production Research*, 55(24), 7476-7490.

<https://doi.org/10.1080/00207543.2017.1307773>