PROPUESTA DE MEJORA EN LOS RETRASOS DE DESPACHO DE AERONAVES EN TRES AEROPUERTOS DE MEDIANA OPERACIÓN DE COLOMBIA USANDO MINERIA DE DATOS.

(CASO DE APLICACIÓN DATA AÑO 2017)

CARLOS MARIO IMBACHI VELASCO

UNIVERSIDAD SANTO TOMÁS

DIVISIÓN DE INGENIERIAS

FACULTAD DE INGENIERIA INDUTRIAL

BOGOTA

2022

PROPUESTA DE MEJORA EN LOS RETRASOS DE DESPACHO DE AERONAVES EN TRES AEROPUERTOS DE MEDIANA OPERACIÓN DE COLOMBIA USANDO MINERIA DE DATOS.

(CASO DE APLICACIÓN DATA AÑO 2017)

CARLOS MARIO IMBACHI VELASCO

Proyecto de investigación para optar al título de Ingeniero Industrial

Ing. Luis Manuel Pulido Moreno

UNIVERSIDAD SANTO TOMÁS

DIVISIÓN DE INGENIERIAS

FACULTAD DE INGENIERIA INDUTRIAL

BOGOTA

2022

Agradecimientos

A mi asesor y buen amigo Ingeniero Luis Manuel por su tiempo y ayuda en el desarrollo de la investigación, a el Doctor Jefferson por su enorme apoyo durante la mayoría de la carrera, a mis padres por impulsarme y ayudarme a culminarla, a Bella por su apoyo y motivación incondicional a lo largo de todos los años, y principalmente a Dios y la virgen de Aránzazu por permitirme llegar hasta aquí.

TABLA DE CONTENIDO

LISTA DE TABLAS

LISTA DE ILUSTRACIONES

RESUMEN

Esta investigación se basa en un análisis de las operaciones de tres terminales aéreas colombianas, Aeropuerto Internacional José María Córdova (Medellín), Aeropuerto Internacional Gustavo Rojas Pinilla (San Andrés) y el Aeropuerto Internacional Matecaña (Pereira), cuya operación internacional no ha sido analizada ni trabajada en otros proyectos similares. Mediante la exploración descriptiva se encontró los problemas que causan los retrasos y para las alternativas de solución se plantearon las técnicas de regresión logística, redes neuronales y XGboosting, donde cualquier técnica es aceptable y validada por sus resultados y por la métrica utilizada en la analítica de datos para este tipo de modelamiento.

Palabras Claves:

Factores operacionales.
Aeropuerto Internacional José María Córdova
Aeropuerto Internacional Gustavo Rojas Pinilla
Aeropuerto Internacional Matecaña
Exploración descriptiva
Regresión logística
Redes neuronales
XGboosting

ABSTRACT

This research is based on an analysis of the operations of three Colombian air terminals, José María Córdova International Airport (Medellín), Gustavo Rojas Pinilla International Airport (San Andrés) and Matecaña International Airport (Pereira), whose international operation does not It has not been analyzed or worked on in other similar projects. Through the descriptive exploration, the problems that cause the delays were found and for the solution alternatives, the techniques of logistic regression, neural networks and XGboosting were proposed, where any technique is acceptable and validated by its results and by the metric used in the analysis of data for this type of modelling.

Keywords:

Operational factors.
Jose Maria Cordova International Airport
Gustavo Rojas Pinilla International Airport
Matecana International Airport
Descriptive exploration
Logistic regression
neural networks
xgboosting

INTRODUCCIÓN

Un aeropuerto es una puerta de entrada a un país. Las partes interesadas del aeropuerto que pertenecen a todas las entidades que manejan y prestan servicios a los pasajeros aéreos se enfrentan a una tarea cuesta arriba para proporcionar un servicio cualitativo y eficiente. Siempre están buscando formas de reducir costos y generar ingresos y, al mismo tiempo, esforzarse por brindar un servicio de calidad [1].

La satisfacción del pasajero, la experiencia ideal en el aeropuerto y la distribución de información en tiempo real para gestionar las interrupciones son los tres componentes comerciales clave para que los aeropuertos tengan éxito comercial y operacionalmente. El autoservicio de pasajeros para reducir el tiempo de espera, maximizar las instalaciones de la terminal, rastrear a los pasajeros utilizando balizas Bluetooth y proporcionar información útil para el aeropuerto son algunos ejemplos de soluciones aisladas adoptadas por los aeropuertos para mejorar los servicios de pasajeros, pero estos no son lo suficientemente holísticos [2].

Según lo argumentado por SITA en 2015, una de las figuras más importantes del mundo en comunicaciones de transporte aéreo y tecnología de la información, existen varios motivos por los que los aeropuertos no han podido ofrecer dicha plataforma en los últimos años. Una razón principal es la falta de conciencia situacional común. El sistema aeroportuario debería poder rastrear, administrar y compartir información en tiempo real con las partes interesadas necesarias sobre todos sus activos y capacidad para optimizar los servicios de pasajeros. La superposición ineficiente de datos y recursos porque los sistemas y procesos han evolucionado individualmente o con una comunicación mínima entre ellos es la otra razón. Debido a esto, el sistema es capaz de gestionar un campo específico del aeropuerto, pero no puede abordarlo como una cadena de suministro integrada basada en el tiempo [3].

Los datos son el alma de un sistema aeroportuario inteligente. Las decisiones y los resultados que produce el sistema dependen únicamente de los datos que recibe. Debido al diseño de sistemas heredados, los sistemas aeroportuarios actuales no pueden acceder a la gama completa de datos relevantes para la situación o los datos han caducado. Por lo tanto, los sistemas aeroportuarios de próxima generación utilizarán las arquitecturas y tecnologías actuales con datos holísticos de fuente única con acceso compartido al sistema completo. También mejorará el valor, la precisión y la coherencia de los datos. [4].

1. DEFINICIÓN DEL PROBLEMA

1.1. DESCRIPCIÓN DEL PROBLEMA

La demanda del transporte aéreo ha venido en crecimiento de manera considerable y consigo la oferta de las aerolíneas, pues desde las últimas décadas el número de personas en el mundo que usan este método de transporte aumenta a diario. Según la Asociación Internacional del Transporte [5], (IATA por sus siglas en inglés) el tráfico aéreo internacional habrá crecido un 6% al finalizar el año 2019, con respecto al año anterior, es decir que aproximadamente cada 15 años los vuelos y el volumen de sus pasajeros se duplicara. Por lo tanto, a medida que aumenta el transporte aéreo va aumentando su tráfico y congestión tanto en el aire como en los aeropuertos, haciendo que el control y administración de los vuelos sea un reto constante para para las entidades encargadas de ello.

Ilustración 1 *Pasajeros totales 2004 – 2016* [6].

Según la Organización de Aviación Civil Internacional (OACI) [7] el crecimiento de los vuelos y redes a nivel nacional e internacional será anormal, volviendo así la congestión en el tráfico aéreo un tema más común de lo que es hoy

en día, por tal razón la OACI quiere seguir encargándose de la planificación y el desarrollo internacional del transporte aéreo.

Ilustración 2 *Tráfico aéreo para el año 2030* [7].

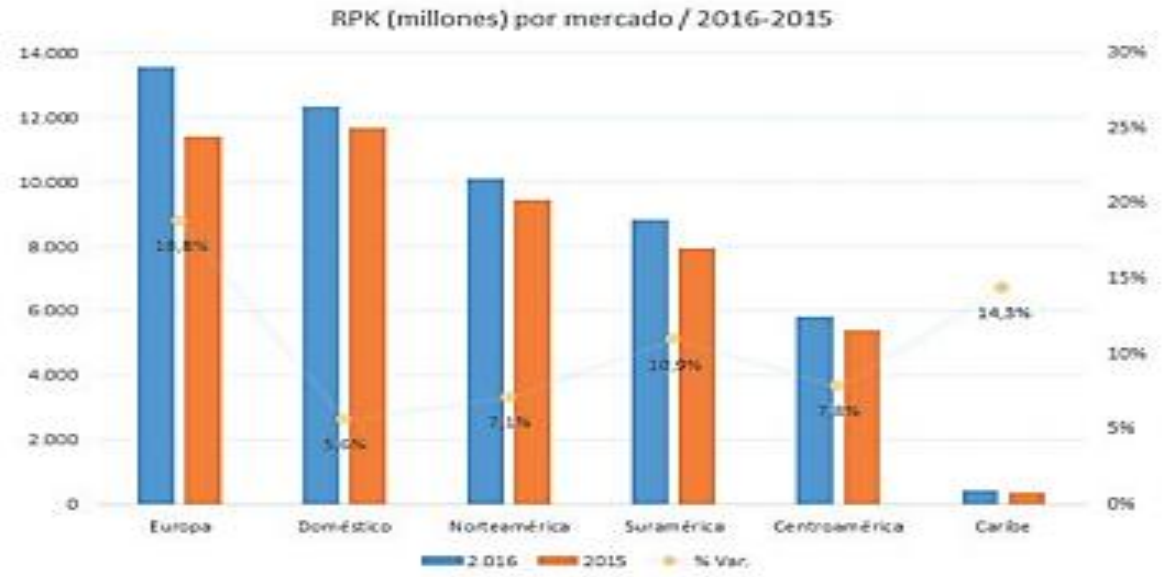

Ilustración 3 *Pasajeros por regiones* [6].

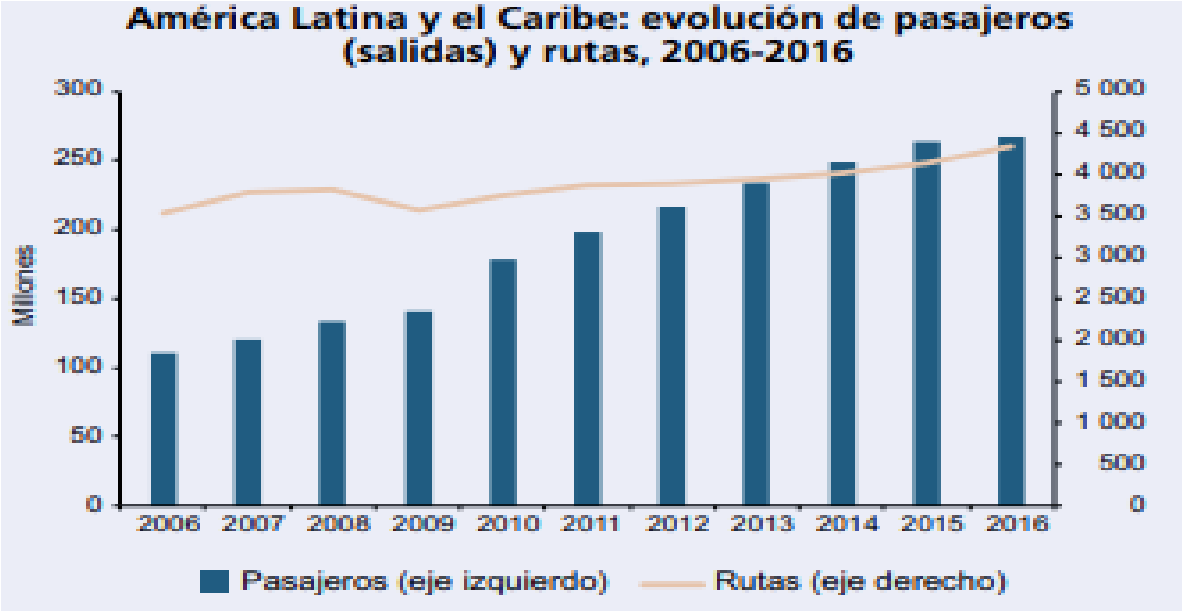

Ilustración 4 *Evolución de pasajeros en Latinoamérica 2006-2016* [8].

IATA [5] señala que en cuanto a la demanda del 2018 a nivel continental en el servicio aéreo, América Latina ocupa el segundo lugar seguido por Asia y superado solamente por Europa, es decir que para los latinoamericanos la demanda aumento un 6,2%, lo que se traduce en que países como Colombia no es la excepción de tener que estar a la vanguardia y afrontar los inconvenientes que se presentan para este medio de transporte donde el problema principal y más visible, no solo en Colombia sino también a nivel mundial, son los retrasos en los vuelos programados.

Según los informes de la CEPAL [8] en los últimos años los viajes en Latinoamérica que en 2006 estaban en 110 millones, en 2016 fueron 266 millones, y según la misma fuente para los próximos 10 o 15 años aquel valor podría ser el doble, potenciando así el tráfico en las aerolíneas y aeropuertos.

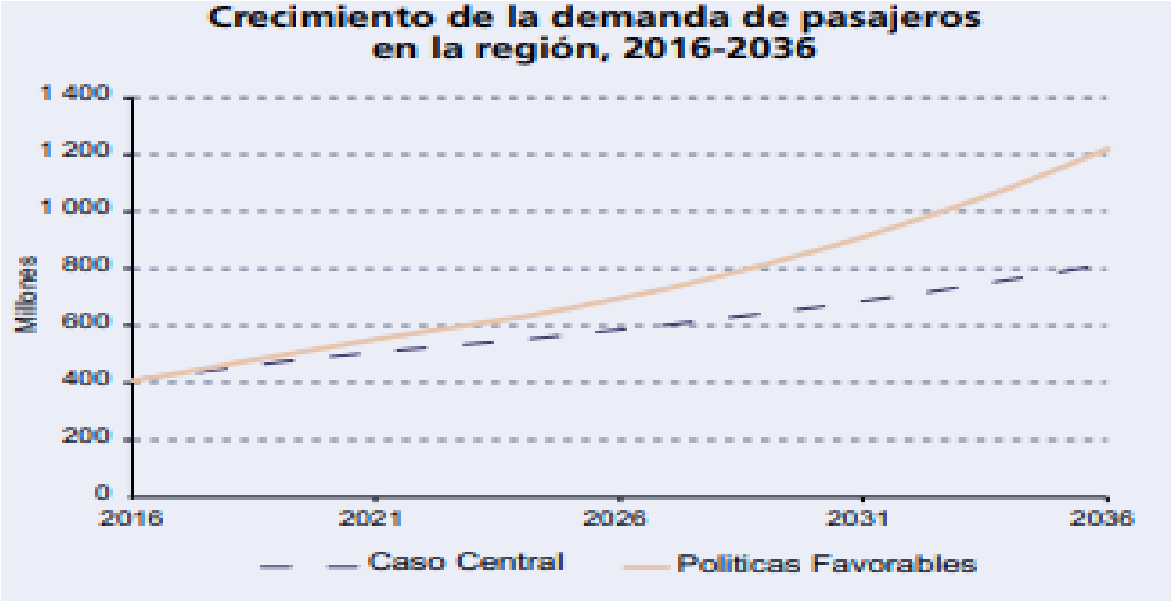

Ilustración 5 *Pronostico de demanda para Latinoamérica* [8].

En Colombia desde siempre se ha concentrado la oferta aérea en la región céntrica, noroccidente y el sur, en donde se encuentran los principales aeropuertos del país, como por ejemplo el aeropuerto internacional el Dorado el cual es el terminal aéreo que más personas y vuelos mueve al año, tan solo en diciembre de 2019 atendieron por hora más de 2.000 viajeros, es decir que al día más de 30.000 [9]. Este alto flujo presente en los aeropuertos hace que en Colombia por 10 vuelos que despegan 4 sean retrasados [10]. La Aerocivil 2017 [6] afirma que se movilizaron un total de 35,77 millones de pasajeros origen-destino en el 2016, lo que representa un crecimiento del 4,81% con relación al año 2015, equivalente a 1,64 millones de personas en donde para aquella oferta el cumplimiento estuvo cerca al 70%. Por lo tanto, esta tendencia nacional de cada vez haber mayor congestión y mayores retrasos en los vuelos impacta muy negativamente la experiencia de las personas que ante este servicio esperan una mayor puntualidad para sus viajes, siendo así necesario buscar y obtener una solución a esta problemática que afecta a un gran número de personas, cuya cifra seguirá creciendo.

2. JUSTIFICACIÓN

Las quejas y reclamos presentadas por los clientes en las aerolíneas y aeropuertos debido a las demoras son cada vez más frecuentes, en donde resalta el descontento, insatisfacción e incluso en algunos casos la desesperación, antes, durante o después de haber utilizado el servicio de transporte aéreo. Esto lleva a que aeropuertos como por ejemplo el dorado dejara de ser visto con tan buenos ojos desde el 2019 [11].

La frecuencia de las demoras también está presente en otros procesos de los aeropuertos, el alto tráfico o mal predicción del clima son tan solo algunos de ellos, obteniendo como consecuencia la creciente inconformidad de los pasajeros quienes viven ya de manera habitual escenas donde se ven obligados hasta a dormir en el suelo mientras esperan su vuelo, en un ambiente donde se respira preocupación, ira, impaciencia e incertidumbre [12].

Llegar a la situación en donde no haya retrasos en los vuelos en cualquier aeropuerto sería lo ideal, por ello es importante solucionar este problema ya que representaría no solo para los pasajeros, sino también para los mismos aeropuertos y aerolíneas una mayor eficiencia en sus operaciones. El impacto sería general, pero en donde más recae seria en los clientes, donde su satisfacción se incrementaría considerablemente pues a manera de ejemplo solo en el Aeropuerto el Dorado estaría impactando al día a 850.000 pasajeros en promedio, incluso en el segundo trimestre del año 2019 hubo 4.297.009 pasajeros saliendo desde este aeropuerto [13].

La mala planificación tiene gran peso en esta problemática pues n diciembre del 2017 se inauguró una ampliación al aeropuerto El Dorado, ampliación que se quedó corta muy rápidamente debido al rápido crecimiento del tráfico aéreo que nadie tuvo en cuenta a la hora de hacer los cálculos, pues tan solo en 2019 este aeropuerto ya estaba movilizando el doble que, en 2010, es decir aproximadamente 45 millones de pasajeros [14].

MES	N° DE PASAJEROS		TOTAL
	Internacional	Domestico	
Abril	483.341	887.709	1.371.050
Mayo	497.760	926.757	1.424.517
Junio	496.224	1.005.218	1.501.442
Total general	1.477.325	2.819.684	4.297.009

Tabla 1 *Pasajeros y operaciones aéreas 2do trimestre de 2019 aeropuerto el Dorado. Elaboración propia según* [13].

Hay que destacar que el mes del año que más presenta retrasos y vuelos cancelados es el mes de diciembre según dijo migración Colombia al periódico Espectador, por ser la época favorita para las vacaciones, y así mismo para las congestiones [15].

En el país la cobertura y servicio es brindada por aeródromos en la mayoría del territorio colombiano, teniendo así presencia en todos los departamentos del país y en el Distrito Capital. De esta manera se logra contar con cobertura al cien por cien a nivel departamental, debido a que la Aerocivil controla los aeropuertos con mayor capacidad, importancia y cantidad de operaciones (ver ilustración 6). Además, según lo menciona la Aerocivil [6] del total de aeropuertos existentes en Colombia, que son 202, hay 11 que califican como aeropuertos internacionales, es decir que la problemática abarcaría no solo los vuelos a nivel nacional, también los que tienen como destino algún país del exterior.

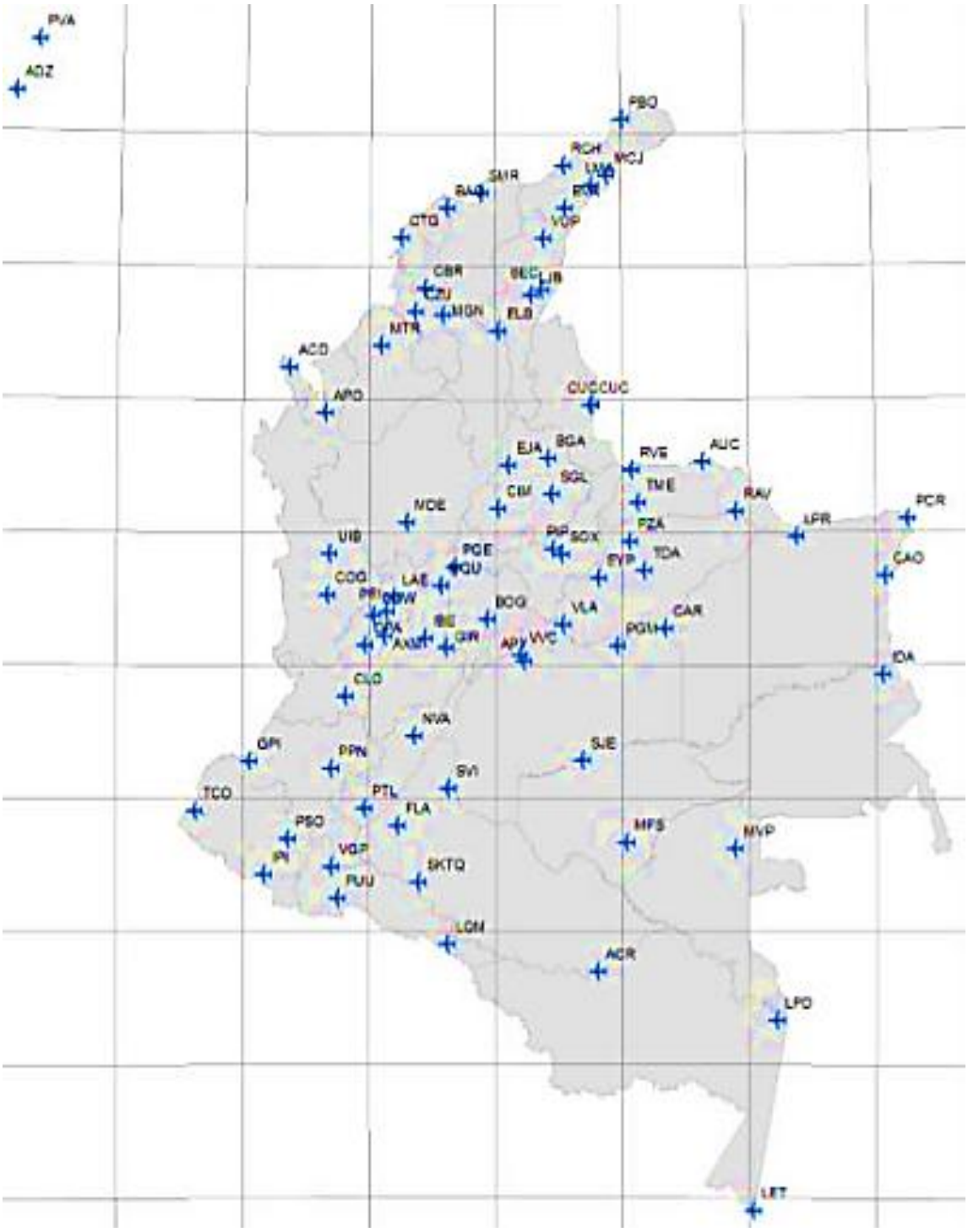

Ilustración 6 *Aeropuertos de Colombia* [6].

2.1. PREGUNTA DE INVESTIGACIÓN

Como se puede apreciar anteriormente el aumento de flujo de pasajeros es un hecho imparable, pero no por ese motivo debería de serlo sus problemas con el retraso de los vuelos, por lo que con mayor razón se hace necesario plantear la siguiente pregunta: *¿Cómo se puede elaborar una propuesta que reduzca los retrasos de vuelos en tres aeropuertos de mediana operación de Colombia utilizando herramientas de analítica de datos?.*

.

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Elaborar una propuesta de mejora que permita reducir el tiempo y número de retrasos en el despacho de aviones en tres aeropuertos de mediana operación de Colombia.

3.2. OBJETIVOS ESPECÍFICOS

- Caracterizar la situación de retrasos en la data del 2017 en tres aeropuertos de mediana operación en Colombia e identificar las variables críticas del proceso, utilizando herramientas de analítica de datos.

- Elaborar alternativas de solución al problema planteado y seleccionar la que más se ajuste a la situación de los tres aeropuertos seleccionados en Colombia.

- Detallar la alternativa de solución escogida y validarla con un modelo de simulación de Montecarlo.

4. MARCO REFERENCIAL

4.1. MARCO HISTORICO Y CONCEPTUAL

4.1.1. Minería de Datos

Los orígenes de la Minería de Datos según Jiawei Han [16] se remontan a los años 50ª, en donde las oficinas de informática hacían resúmenes de información de todo tipo, en especial la información comercial, la cual se hallaba en los archivos de la computadora central, esto con el fin de hacer más fácil el trabajo directo. Es de esa manera como nacen los sistemas de información de dirección, aunque estos eran muy grandes, con poca flexibilidad e imposibles de leer para otras personas. Ya en los años 60ª surgen sistemas que gestionaban bases de datos, pero aun con dificultades para hacer búsquedas, por ende, luego aparecen los motores relacionales que resolvieron los problemas anteriores, pero no quitaron lo demorado y difícil de preparar los informes y depurarlos, eso sumado a que no eran integradas la gran diversidad de base de datos.

Ya en los 70ª empiezan a surgir los conceptos sobre Minería de Datos que al día de hoy están presentes, pues según Olmos Pineda la MD llamada pesca o arqueología de datos, buscaba plantear y encontrar algunas correlaciones sin que hubiese necesidad de proponer antes una hipótesis para cualquier trabajo investigativo, y ante los problemas presentes en la MD la Data Warehouse (DW) llega para solucionarlos a finales de los 80ª, estimulando el avance de los enfoques de la Minería de Datos al empezar a automatizar las tareas y permitir extraer conocimientos inductivos [17].

Y aunque los conceptos varían un poco según los autores, el enfoque de la MD seguía siendo la misma, por ejemplo, desde 1993 [18] definía a la minería de datos como un proceso donde se plantean diferentes búsquedas y obtención de la información, tendencias y patrones que no se conocían antes, todo esto proveniente

de una cantidad de datos muy grande almacenados en una determina base de datos.

Fallad en 1996 definió a la minería de datos como un proceso de alta importancia en la identificación de información novedosa, útil y valida, que permite entender los posibles y distintos patrones dentro de un grupo de datos [19]. Por su parte M. Berry y G. Linoff desde 1997 ven a la minería de datos como una exploración en una gran cantidad de datos que permite analizar lo encontrado mediante procesos y medios semiautomáticos y automatizados para hallar alguna regla o patrón significativo [20].

Peacock en 1998 da tres perspectivas para la Minería de Datos debido a su amplitud. En la primera se refiere a la MD como el hecho de descubrir de manera automática patrones o modelos no obvios ocultos en la base de datos contribuyendo en gran parte a temas de cualquier negocio o empresa (estrategias y objetivos), esto logrado por métodos computacionales que necesitan de muy poca intervención de una persona, y en donde se puede contar con árboles de decisión, algoritmos genéticos, algoritmos de red neuronal artificial y lógica difusa. Como segunda perspectiva, una más amplia, tiene en cuenta lo anterior sumado a la confirmación o pruebas de relaciones dadas en el descubrimiento de los datos, empleando métodos clásicos de estadística y bayesianos y una hipótesis (regresión mínimo cuadrática, análisis discriminante y exploratorio, regresión logística), es decir un proceso semiautomático de minería. Y como tercera perspectiva y la más grande, la MD vendría siendo un proceso donde se descubre conocimiento en base de datos, Knowledge Discovery in Databases, incluyendo el análisis de datos [21].

Por lo tanto, se puede resumir que la minería de datos es la búsqueda de conocimientos en base de datos por medio de métodos estadísticos e inteligencia artificial.

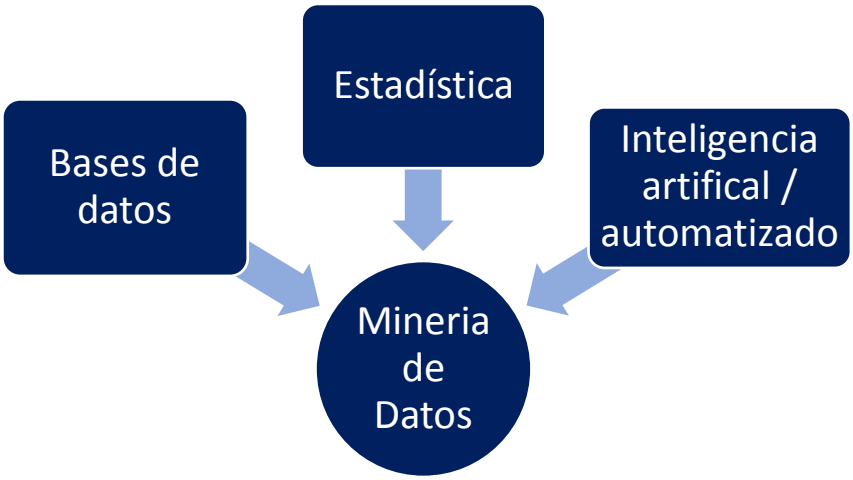

Ilustración 7 *Minería de Datos. Elaboración Propia*

4.1.2. Técnicas de Minería de Datos

Las técnicas de MD se clasifican según Pérez [22] en primera instancia en técnicas de modelado originado por la teoría (variables independientes y dependientes), es decir en base a un conocimiento previo teórico, es por ello que en esta técnica el modelo encontrado en el proceso de minería de determinados datos ha de ser contrastado antes de verlo como válido. Aquí se puede mencionar entre técnicas de este tipo las de análisis discriminante, regresión y asociación, series temporales, análisis de varianza y covarianza. En esta clasificación al aplicar un modelo debería de contar con las fases de:

- Identificación objetiva, donde se identifica el modelo que más se ajuste a los datos, según reglas aplicadas a los mismos.
- Estimación, donde se calcula los parámetros al modelo.
- Diagnosis, donde se contrasta la valides del modelo.
- Predicción, donde se usa el modelo para estimar el valor de las variables dependientes.

Como segunda técnica se halla la originada por los datos, en esta técnica los modelos son creados automáticamente por patrones reconocidos en la base de datos, por ello las variables no son definidas en dependientes o independientes, y no hay tampoco un previo modelo de los datos, pues aquel se logra mediante la combinación del conocimiento que se tiene antes y después de la MD, para luego contrastarse y poder validarlo. Entre las técnicas más usadas se encuentran reducción de la dimensión, la de clasificación como clúster, de escalamiento óptimo y multidimensional, análisis conjunto, árboles de decisión, y redes neuronales; esta última permite ir descubriendo y mejorando modelos al mismo tiempo que se va avanzando en la exploración son las redes neuronales, ya que con ellas se puede encontrar sin factores externos relaciones complejas entre las variables.

Y por último están las técnicas auxiliares, métodos nuevos que se basan en informes y técnicas de estadística descriptiva.

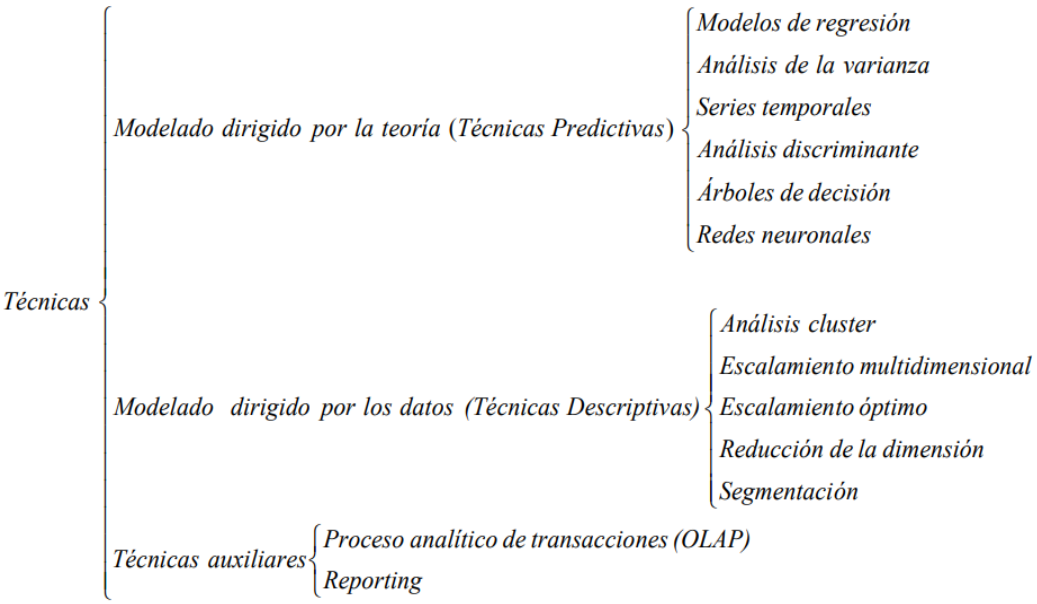

Ilustración 8 *Técnicas de Minería de Datos [22].*

4.1.3. Business Analytics

Para Rouse [23] la business analytics es una exploración metódica e iterativa de los datos de una empresa u organización, teniendo como énfasis el análisis estadístico. El BA o análisis empresarial es usado en empresas que toman decisiones basadas en datos, en donde estos datos son tratados como un activo de la misma empresa, buscando constantemente la manera convertirlo en una ventaja competitiva ante los demás. Para que todo análisis empresarial sea aquella ventaja anhelada necesita tener analistas expertos en tecnología y la empresa, unos altos parámetros de calidad en los datos y la responsabilidad de la organización de usar aquellos datos con el fin de obtener información que influya en las decisiones comerciales.

Los tipos de análisis de negocios son:

- Análisis prescriptivo, donde se usa la información de rendimiento del pasado para producir recomendaciones sobre cómo llevar esas situaciones similares cuando se vuelvan a presentar.
- Análisis descriptivo, (KPI) busca entender el estado actual del negocio mediante los indicadores claves del rendimiento.
- Análisis predictivo, evalúa la probabilidad de sucesos y resultados futuros mediante el análisis de datos de tendencias.

Funcionamiento de la BA:

- Determinar del objetivo comercial del análisis.
- Seleccionar de metodología de análisis
- Extracción de sistemas comerciales, limpieza e integración de un solo repositorio (data mart)
- Adquirir datos que respalden el análisis.

- Analizar en contraste con una pequeña muestra de datos (excel, MD, modelado predictivo, etc.).
- Realizar nuevas preguntas e iterar el proceso de análisis mientras se encuentran patrones y relaciones.
- Cumplir con objetivo comercial.

BA incluye la toma de decisión táctica en consecuencia a eventos no planeados, y por lo general esta decisión esta automatizada para dar respuestas en tiempo real.

Herramientas de análisis de negocios:

- Herramientas de análisis estadístico
- Software de informes de inteligencia empresarial
- Herramientas de visualización de datos
- Grandes plataformas de datos
- Plataformas analíticas de autoservicio

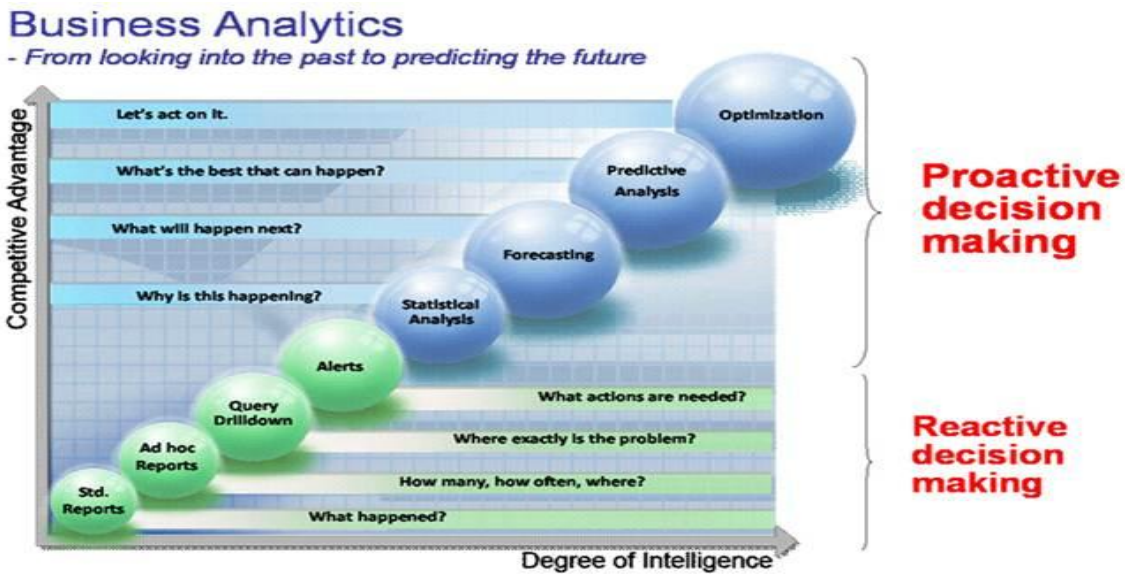

Ilustración 9 *Business Analytics [23].*

Según Ladrero también podemos comprender la BA como un grupo de herramientas que van a ayudar, en cualquier espacio de la organización, a tomar decisiones, a comparar casos, pronosticar amenazas y oportunidad, controlar los recursos, a anticiparse a los resultados entendiéndolos mejor y dándoles sentido al encontrar modelos, anomalías o tendencias.

Las empresas que usan BA obtienen mejores resultados, y las que tienen una política y filosofía basada en la analítica de datos obtienen rendimientos que son el 300% superior al de la competencia. La BA permite a la empresa identificar patrones y tendencias tenues para lograr estar anticipada a cualquier eventualidad controlando hacia buenos resultados.

Business Analytics se basa en un análisis hacia el futuro basándose en la información suministrada por la misma organización y modelos de predicción para mejorar los indicadores de competitividad. Es decir que la BA analiza el AS IS para predecir el futuro de la empresa y de posibles eventos.

En resumen, se puede concluir que el Business Analytics permite tener un panorama claro del futuro en cuanto al funcionamiento de la empresa con el objetivo de que las decisiones a tomar sean asertivas [24].

4.2. MARCO TEORICO

La mayoría de las publicaciones de investigación existentes sobre la optimización de aeropuertos según Laik, Choy y Cheong se clasifican en dos grupos principales. El primer grupo de literatura se centra en los problemas relacionados con el tráfico aéreo en los que el aterrizaje y el despegue de aviones son extremadamente críticos, el segundo grupo de literatura se centra en la capacidad de pasajeros del aeropuerto, así como en la configuración de los diseños del

aeropuerto. Los problemas más comunes mencionados en las revistas son la falta de recursos que afecta la capacidad de las terminales para manejar los vuelos de entrada y salida. La mayor parte de las investigaciones también se centra en el diseño del aeropuerto para garantizar que la llegada y salida de los vuelos se realicen sin mayores problemas. Para trabajos de investigación que se centraron en los pasajeros, la mayoría de ellos manejaron problemas relacionados con las colas y el manejo de equipaje.

La minería de datos abarca muchos campos, incluso se podría afirmar que la MD se puede aplicar a cualquier cosa, por lo tanto, es posible encontrar un gran número de estudios y aplicaciones al mundo del transporte aéreo. Entre los más comunes se encuentran investigaciones realizadas para *pronosticar* retrasos en vuelos, predecir factores que afecten o mejoren el funcionamiento del aeropuerto, *conocer* las tendencias y factores de satisfacción de los pasajeros, entre otros, más sin embargo hallar un estudio o trabajo que busque *disminuir* los retrasos aéreos, es decir encontrar el problema y su solución a partir de la minería de datos evaluando aquel resultado mediante una simulación se torna más escaso, habiendo que recurrir incluso a idiomas diferentes al español [25].

De tal forma entre los diversos aportes que han hecho diferentes autores se puede mencionar a Bogicevic y sus coautores [2] quienes teniendo en cuenta la complejidad de los servicios de la industria aeroportuaria, vieron importancia en identificar qué factores de transporte aéreo son distractores y qué factores mejoran la satisfacción de los pasajeros. Aquel estudio tuvo como objetivo explorar los atributos de calidad del servicio aeroportuario más frecuentemente mencionados y distinguir los factores clave para la satisfacción o insatisfacción de los pasajeros en el contexto del aeropuerto. Para ello Se realizó un análisis de contenido de 1.095 comentarios de viajeros publicados entre 2010 y 2013 en un sitio web de revisión de aeropuertos con el fin de identificar factores de satisfacción o insatisfacción. Se seleccionó al azar los comentarios de los consumidores relacionados con 33 destinos populares. Los resultados de ese estudio indicaron factores de satisfacción

clave en el contexto del aeropuerto, como la limpieza y el ambiente agradable para pasar el tiempo. Por otro lado, el control de seguridad, la señalización confusa y la oferta de restaurantes deficiente se reconocen como factores desfavorables importantes en el entorno del aeropuerto. Los hallazgos del estudio brindan información sobre los factores predominantes de satisfacción, desabastecimiento y desempeño de la calidad del servicio aeroportuario desde la perspectiva de los pasajeros. De esta forma los equipos de administración del aeropuerto pueden usar los resultados del estudio para renovar las instalaciones del aeropuerto y mejorar la calidad del servicio. Según el conocimiento de los autores, este estudio es el primero en utilizar las técnicas de minería de datos visuales para examinar la experiencia de los usuarios del aeropuerto. La visualización produjo resúmenes de comentarios cualitativos en forma de nubes de etiquetas, redes de palabras e imágenes de árboles de palabras que ayudan a descubrir los temas más emergentes de quejas y cumplidos de los viajeros.

Otro tema importante en cuanto al rendimiento de los aeropuertos es el del clima, el cual afecta en gran medida los retrasos en la llegada de los vuelos. Por tal motivo Henriques y Feiteira buscaron predecir estos retrasos en el aeropuerto internacional Hartsfield Jackson de Atlanta, siguiendo una metodología de Knowledge Discovery Database (KDD) y aplicando técnicas de minería a varios datos recopilados como los meteorológicos de 2 meses antes, tráfico aéreo nacional, estadísticas de transporte (rendimiento), zona horaria, tipos de aviones, hora de salida y llegada de los aviones y temporada vacacional, siendo así estos datos las variables independientes, y como variable dependiente utilizaron el retraso de llegada como binario, donde si el tiempo real de llegada es igual o mayor de 15 minutos desde la hora programada, la variable asume el valor 1 de lo contrario, asume el valor de 0. Posteriormente se procesó los datos con el fin de saber si la presencia de alguna mejora o empeora el modelo, es decir se procedió a limpiar la base de datos. Luego se transformaron los datos con el método de min-máx. y el método de filtro. En cuanto a minería de datos para encontrar patrones en la base

y traducirlos en información, de acuerdo al objetivo del trabajo, escogieron tres algoritmos basándose en el criterio de otros autores: bosques aleatorios y multicapas, arboles de decisión y redes neuronales. Finalmente, cada modelo se evaluó y fueron comparados en cuanto a rendimiento, resultando siendo mejor el modelo del Perceptrón multicapa con un 85% de precisión [26].

Nazeri y Zhang usaron también la minería de datos para analizar los impactos climáticos severos en el rendimiento del Sistema Nacional del Espacio Aéreo (NAS). Para este experimento, utilizaron tres fuentes de datos: Pronóstico meteorológico conectivo nacional (NCWF), Rendimiento de la calidad del servicio de la aerolínea (ASQP) (horarios de salida y llegada programados y reales de cada vuelo de diez aerolíneas, el número de cola, los tiempos de encendido/apagado de las ruedas, los tiempos de taxi, la cancelación y la información de desvío) y Sistema mejorado de gestión del tráfico (ETMS) (horarios planificados de salida y llegada, horarios reales de salida y llegada, rutas de vuelo planificadas, rutas de vuelo reales y cancelaciones), suministrado por el Centro nacional de investigación atmosférica. También usaron datos de NCWF de abril a septiembre de 2000 para representar la temporada de clima severo. Se utilizaron 152 días con datos meteorológicos completos. Posteriormente se hicieron un filtro en los datos al enfocarse solo en los vuelos de entre las 6 am y las 11pm, dividiendo este tiempo en cuatro periodos debido a que el clima de un día así se divide, de igual forma se centraron en el espacio aéreo nacional y el noroeste para obtener 36 aeropuertos a estudiar. Con eso pasaron a la fase de extracción de características que consta de cuatro pasos: segmentación de imágenes, extracción de características climáticas, extracción de características de tráfico aéreo y conversión de representación, a los cuales se les aplico un algoritmo de agrupamiento simple y un algoritmo del árbol de decisión para identificar las áreas de clima severo. Y al cuestionasen sobre si los atributos extraídos eran relevantes para el rendimiento del NAS en función de los datos disponibles, realizaron análisis de correlación y regresión (ocho características climáticas y de tránsito aéreo con su medida de rendimiento) para luego aplicar la

agrupación (días con impactos climáticos similares en el rendimiento del NAS) y clasificación (hallaron patrones/reglas climáticas que tienen un impacto significativo en el rendimiento del NAS). Finalmente obtuvieron resultados prometedores sobre las áreas con clima severo, pero sobre todo se demostró que la minería de datos tiene aplicabilidad a los problemas de gestión del tráfico aéreo y además puede representar una primera cercanía a las soluciones significativas de esos problemas [27].

En España en la Universidad de País Vasco se realizó un proyecto, igualmente relacionado, sobre el Pronóstico de Retraso de Vuelo debido al Clima Usando Minería de Datos. Su objetivo fue predecir demoras debido al clima dentro de los 10 días posteriores al vuelo con un 70% de precisión, para ello uso la herramienta de código abierto Weka, que le proporciono aprendizaje automático y la minería de datos. Construyo su modelo inicial usando Weka y Rand, probando diferentes clasificadores. Su primera hipótesis fue que Naive Bayes les proporcionaría los mejores resultados, pero resulta que aquello se logró con el clasificador de análisis discriminante lineal, pues con este logro un 74% de precisión al predecir cuándo se retrasaría un vuelo debido al clima dentro de los 10 días posteriores a la salida de ese vuelo. Mientras que los Bayes solo les dieron una precisión del 69% [28].

De forma más similar a lo que se hará en esta investigación se halló tres trabajos que tuvieron como base de datos la información suministrada por el Departamento de transporte de los Estados Unidos, específicamente la Oficina de estadísticas de transporte. En primer lugar, Bandyopadhyay y Guerrero se centraron en solo dos años de datos, con O ‘Hare en Chicago como el aeropuerto principal y American Airlines como la aerolínea de estudio, teniendo tres enfoques: identificar qué factores causan la mayor cantidad de retrasos, poder predecir si un vuelo individual se retrasará, y si se retrasa cuanto es ese tiempo. Mediante el uso de la regresión lineal, pudieron averiguar cuáles eran los factores de retraso más comunes. Encontraron que usando Naive Bayes pueden hallar la mejor predicción

de cuándo se retrasarán los vuelos. Sin embargo, tuvieron dificultades para predecir en qué magnitud se retrasaría un vuelo. Descubrieron que todos los algoritmos que probaron se recogieron en uno de los dos picos que encontraron, pero nunca en ambos. Creen que, al desarrollar dos modelos separados, uno para cortos y otro para retrasos largos, pueden haber tenido un mayor éxito [29].

Patrick Robert Steele también trabajo partiendo de los datos suministrados por el Departamento de Transporte de Estados Unidos, en donde modeló los problemas de encontrar la ruta de retraso mínimo como un problema de ruta más corta. Algo único en su enfoque es que utilizó el muestreo de la red aérea para modelar el efecto en cascada de los retrasos en los vuelos. También implementó un aspecto de visualización que muestra los resultados de sus algoritmos para que sea más fácil comprender las tendencias significativas de un conjunto de datos tan grande [30].

Y por último Baluch, Bergstra y El-Hajj usando también información proveniente de la misma fuente, estos datos contenían información que incluye la identificación del transportista, el aeropuerto de origen, el aeropuerto de destino, el mes, el día y el año junto con detalles sobre cualquier demora. El objetivo de aquella investigación era analizar cuarenta y ocho conjuntos de datos desde julio de 2012, hasta el conjunto de datos más reciente, junio de 2016, utilizando SQL Business Intelligence Tools de Visual Studio, y así dar respuestas a varias preguntas/problemas planteados desde un inicio como el predecir cuándo es más probable que encuentre retrasos aumentando el conocimiento para que los consumidores conozcan sobre las mejores y más eficientes formas de viajar. Las técnicas de minería de datos usadas fueron la clasificación, la agrupación en clústeres y los algoritmos de árbol de decisión. Finalmente, con los resultados obtenidos llegaron a la conclusión de que el tamaño del aeropuerto no tiene un impacto en los retrasos de los vuelos, también de que algunos operadores tienen retrasos mucho más largos que otros, por ejemplo, la aerolínea American Airlines tiene la demora promedio más alta y Envoy Airlines tiene la más baja. Otro hallazgo

interesante fue que los vuelos que salen tarde generalmente no compensan el tiempo perdido en el aire. Encontraron una fuerte correlación entre el mes, el día del viaje y la demora de salida, así como también el día y el mes en que es mejor viajar. Lo último que encontraron fue que, cuanto más largo es el retraso de la llegada tardía, mayor es el tiempo entre este y el retraso de la salida [31].

5. MARCO METODOLOGICO

5.1. TIPO DE INVESTIGACIÓN

La presente investigación es secuencial y delimitado a la información de la data del 2017 de la Aerocivil. Se elabora dentro del enfoque cuantitativo, basándose en el trabajo de Collado en donde recolecta datos con base en el análisis estadístico y la medición numérica, con el objetivo de describir pautas de comportamiento y probar hipótesis [32].

El diseño del proyecto es no experimental, debido a que no se hace variar de manera intencional las variables independientes para ver qué efectos tienen sobre otras variables. En este trabajo lo que se hace es observar los fenómenos como se dan naturalmente, para así analizarlos en un diseño de investigación transversal descriptivo y correlacional, donde indagamos que incidencia hay en los niveles de las variables de la población limitándonos a fijar relaciones entre variables sin determinar causalidades. Es decir, esta metodología se centra más en el “qué”, en lugar del “por qué” del sujeto de investigación. En otras palabras, su objetivo es describir la naturaleza de un segmento demográfico, sin centrarse en las razones por las que se produce el determinado fenómeno [32].

En cuanto al alcance del proyecto hablamos de exploratorio—descriptivo. Exploratorio debido a que buscamos examinar un problema de investigación poco indagado, en donde suelen haber dudas y temas no tocados antes. Descriptivo por querer medir o recolectar información de forma conjunta o independiente sobre las variables a las que nos referimos [32].

TIPO DE INVESTIGACIÓN	
Alcance	**Exploratorio—descriptivo**
Diseño	**No Experimental**
Enfoque	**Cuantitativo**

Tabla 2 *Tipo de investigación. Elaboración propia.*

5.2. RECOLECCIÓN DE INFORMACIÓN

La información con la que se va a desarrollar la investigación es una base de datos del año 2017, recogida por la Unidad Administrativa Especial de Aeronáutica Civil, en la que se hará todo el proceso de minería de datos. Esta data se recolecta vuelo a vuelo que despega de los terminales aéreos de Colombia, independientemente de su destino.

La base de datos a trabajar fue suministrada por la Aerocivil al ingeniero Luis Manuel Pulido Rico, docente de la Facultad de Ingeniería Industrial de la Universidad Santo Tomas, esto con el fin de desarrollar una investigación que afecte positivamente el algún aspecto relacionado al tráfico aéreo.

5.2.1. Variables

La base de datos posee información de todos los aeropuertos de Colombia en cuanto a:

- Tráfico
- Aerolínea
- Origen
- Destino
- Número del vuelo
- Fecha programada
- Hora programada
- Fecha de remolque
- Hora de remolque
- Demora
- Estado del vuelo
- Motivo de la demora

- Observaciones
- Estatus
- Fecha
- Departamento origen

5.2.2. Población y muestra

Como población se tiene a los Aeropuertos de Colombia, y como muestra a tres de estos aeropuertos, de Medellín, Pereira, y San Andrés, los cuales son de mediana operación comparados con el resto.

5.3. ANÁLISIS DE DATOS

Para el desarrollo las herramientas a usar en el desarrollo de la investigación será programas estadísticos como SPSS de IMB, R Estudio y Excel, los que nos permitieran realizar análisis y tratamientos en grandes volúmenes de datos.

Para el procesamiento de los datos se adaptará la metodología CRISP-DM, la cual es una guía estándar abierta con enfoques que se utiliza comúnmente en el proceso de minería de datos de manera jerárquica en cuatro diferentes etapas:

5.3.1. Comprensión y calidad de la información

En este paso se buscará entender la información y a su vez mejorarla o modificarla para el mismo fin, por lo que la descripción de la data, mejoramiento de su calidad, formatos y las exploraciones se llevaran a cabo con el fin de encontrar significados, reciprocidades y demás datos relevantes para el modelamiento.

5.3.2. Exploración de los datos

Este paso es donde se usa la estadística descriptiva para ir conociendo la naturaleza de la información en la data y a su vez ir preparando la misma para la siguiente etapa. En este se analizará desde lo general a lo especifico teniendo en cuenta las aerolíneas, aeropuertos en general y a estudiar, las demoras en aquellas terminales aéreas, los tiempos de esas demoras y principalmente sus causas

5.3.3. Modelamiento

Esta es la etapa donde se plantean las alternativas que se puedan hallar mediante los diferentes modelos de predicción que se usaran en la investigación, obteniendo donde atacar y buscar solucionar los retrasos en los aeropuertos estudiados.

5.4. HIPOTESIS

Se hallarán las causas de retrasos de los vuelos y con ello su alternativa de donde solucionar y reducir en tiempo y número estos retrasos en los tres determinados aeropuertos de Colombia, en donde la principal causal será el factor climático.

6. RESULTADOS

6.1. COMPRENSIÓN DE LA INFORMACIÓN

6.1.1. Descripción de los datos

La base de datos posee información de los vuelos que se han despachado en 41 aeropuertos de Colombia durante el año 2017, de los cuales 12 son aeropuertos internacionales.

En la data se encuentra 18 tipos de información diferente:

ATRIBUTO	DESCRIPCIÓN
Tráfico	Nacional o internacional
Aerolínea	Nombre de la empresa
Origen	Sigla IATA del aeropuerto donde tiene origen el vuelo
Destino	Sigla IATA del aeropuerto donde tiene destino el vuelo
Número del vuelo	Número del vuelo que se le asigna a la operación
Fecha programada SCORE_UTC	Fecha internacional de salida que tiene programada el vuelo
Hora programada SCORE_UTC	Hora internacional de salida que tiene programada el vuelo
Fecha de remolque_UTC	Fecha internacional en donde se ejecutó el remolque de la aeronave
Hora de remolque_UTC	Hora internacional en donde se ejecutó el remolque de la aeronave
Demora AEROCIVIL	Diferencia entre la Hora de remolque_UTC y la Hora programada SCORE_UTC
Demora (hh:mm)	Diferencia entre la Hora de remolque_UTC y la Hora programada SCORE_UTC
Estado del vuelo	Describe si el vuelo se demoró, canceló, adelantó o se cumplió según lo programado por la Aerocivil
Código	Numero IATA que identifica el motivo de la demora o cancelación de los vuelos
Motivo de la demora	Técnico, operacional, incontrolable
Observaciones	Comentarios adicionales de la operación
Estatus AEROCIVIL	Describe si el vuelo se demoró, cancelo, adelanto o se cumplió según lo programado por la Aerocivil
FECHA	Describe el mes en que se realizó el vuelo
DEPARTA ORIGEN	Departamento donde tiene salida el vuelo

Tabla 3 *Descripción de los datos. Elaboración propia.*

Dentro de la data se encuentra los aeropuertos nacionales e internacionales y sus aerolíneas, en donde encontramos los siguientes:

IATA	NOMBRE	PARAJE	OPERACIONES	DESTINOS
APO	Aeropuerto Antonio Roldan Betancourt	Carepa	3845	7
AUC	Aeropuerto Santiago Pérez	Arauca	1629	3
BSC	Aeropuerto José Celestino Mutis	Bahía Solano	1259	6
BUN	Aeropuerto Gerardo Tobar López	Buenaventura	192	1
CZU	Aeropuerto Las Brujas	Corozal	1705	4
EJA	Aeropuerto Yariguíes	Barrancabermeja	1032	1
EOH	Aeropuerto Olaya Herrera	Medellín	22001	32
EYP	Aeropuerto El Yopal	El Yopal	3496	3
FLA	Aeropuerto Gustavo Artunduaga Paredes	Florencia	1290	5
GPI	Aeropuerto Juan Casiano	Guapi	366	1
IBE	Aeropuerto Perales	Ibagué	2178	2
IPI	Aeropuerto San Luis	Ipiales	290	1
MTR	Aeropuerto Los Garzones	Montería	7796	12
MVP	Aeropuerto Fabio Alberto León Bentley	Mitú	239	1
MZL	Aeropuerto de La Nubia	Manizales	3226	3
NVA	Aeropuerto Benito Salas	Neiva	3446	1
PCR	Aeropuerto German Olano	Puerto Carreño	354	1
PPN	Aeropuerto Guillermo León Valencia	Popayán	1400	1
PSO	Aeropuerto Antonio Nariño	Pasto	2154	2
PUU	Aeropuerto Tres de Mayo	Puerto Asís	797	4
PVA	Aeropuerto El Embrujo	Providencia	1108	1
RCH	Aeropuerto Internacional Almirante Padilla	Riohacha	765	2
RVE	Aeropuerto Los Colonizadores	Saravena	312	3
SJE	Aeropuerto Jorge E. González Torres	San José del Guaviare	154	1
TCO	Aeropuerto Araucanía	Temuco	1311	2
TME	Aeropuerto General Gabriel Vargas Santos	Tame	51	1
UIB	Aeropuerto El Caraño	Quibdó	6788	14
VUP	Aeropuerto Alfonso López Pumarejo	Valledupar	2145	2

VVC	Aeropuerto Vanguardia	Villavicencio	1055	4

Tabla 4 *Aeropuertos nacionales. Elaboración propia.*

IATA	NOMBRE	PARAJE	OPERACIONES	DESTINOS
AXM	Aeropuerto Internacional El Edén	Armenia	3641	4
CUC	Aeropuerto Internacional Camilo Daza	Cúcuta	5853	7
BAQ	Aeropuerto Ernesto Cortissoz	Barranquilla	12721	12
PEI	Aeropuerto Internacional Matecaña	Pereira	9251	11
LET	Aeropuerto Internacional Alfredo Vásquez Cobo	Leticia	1357	7
ADZ	Aeropuerto Internacional Gustavo Rojas Pinilla	San Andrés	10145	9
BGA	Aeropuerto Internacional Palonegro	Bucaramanga	11040	13
BOG	Aeropuerto Internacional El Dorado	Bogotá D.C.	124772	74
CLO	Aeropuerto Internacional Alfonso Bonilla Aragón	Cali	23322	22
CTG	Aeropuerto Internacional Rafael Núñez	Cartagena	18673	13
MDE	Aeropuerto Internacional José María Córdova	Medellín	32201	32
SMR	Aeropuerto Internacional Simón Bolívar	Santa Marta	7138	3

Tabla 5 *Aeropuertos internacionales. Elaboración propia.*

Para llevar un mejor control del desarrollo se procede a nombrar las bases de datos resultantes durante la investigación, siendo la base de datos original la **BASE 0 Inicial**, con 332.506 datos de operaciones aéreas en el 2017 en territorio colombiano, a la cual se empieza a hacer los primeros acercamientos y modificaciones a continuación.

6.2. CALIDAD DE LOS DATOS

Con la data original (BASE 0) se tiene información sobre 332.506 vuelos operados, y debido a que la data contenía espacios vacíos, diferentes formatos, información duplicada o con igual significado, es necesario llevar a cabo el proceso de calidad, transformación y limpieza de datos, y así mismo ir acercándonos a la

información que abarcada por los retrasos, para lo cual se usó la herramienta Excel. Finalizando este paso se obtendría una base de datos nueva con enfoque en las demoras, por lo que se nombró **BASE 1 Demoras,** con un total de 79.431 datos**:**

6.2.1. Complemento y exclusión

- En los datos de *Estatus AEROCIVIL* y *Estado del vuelo*, se procede a eliminar los estados diferentes a “demoras” (253.033 casos), es decir que el 23,9% de las operaciones aéreas del 2017 tuvieron demoras. Cabe destacar que el 61,2% fueron vuelos cumplidos, el 11,7% fueron cancelados, el 3% fueron adelantados, el 0,17% fueron penalizados y el 0,03% son datos no validos

- Para *Departamento de origen* en donde no hay información se procede a rellenar con su información correspondiente basado en el aeropuerto de origen (5 casos).

- En *Código* se hallaron casos vacíos y con código número 0, el cual no está estipulado en los Códigos de demora de causas de incumplimiento de Itinerario de la IATA, por lo cual se procedido a eliminar estos mismos (42 casos).

6.2.2. Integración de formatos

- Para *Fecha programada* y *Fecha de remolque* se encontraron varios formatos como año/mes/día, día/mes/año, diaSEPaño, y un valor numérico que tiene Excel como referencia de fecha, por lo tanto, se procedió a dejar el formato para estos casos de la manera día/mes/año.

- Hora de remolque y hora programada tiene varios formatos, pero debido a que esto no infiere en el valor del tiempo de retrasos (diferencia entre ambas) se decide no hacer cambios por el momento.

- En cuanto a *Demora AEROCIVIL* y *Demora hh:mm* se tiene casos donde la información dice “24h” y en otros el formato es decimal, por lo tanto, se opta por cambiar y dejar esta información en minutos para facilitar el desarrollo del trabajo.

6.2.3. Homogenización

- Se encontró que en *Demora AEROCIVIL* y *Demora hh:mm* hay muchos casos repetidos, casos en donde una tiene información, pero la otra variable no, y casos donde hay errores al momento de calcular el tiempo de demora que tuvo la operación, por lo que se procedió a corregir estos tiempos y a homogenizar ambas variables dejando solo una, nombrada *aTIEMPO DE DEMORA*

- Después de complementar y excluir casos en *Estatus Aerocivil* y *Estado del vuelo*, se encontró con información igual en ambos casos, por lo que se deja en una sola variable esta información, nombrada *bESTADO DEL VUELO.*

- En Motivo de la demora se encontró casos donde la información es escrita en ambos géneros como por ejemplo EXTERNO y EXTERNA, significando lo mismo, dado el caso se hace la homogenización (19.792 casos).

- Como en el caso anterior, para *Trafico* se tiene dos valores con un mismo significado, por lo cual se procede a homogenizar también, dejando así el valor NACIONAL e INTERNACIONAL para esta variable (1.405 casos).

- Se encontró un caso en *Origen* donde el aeropuerto tenía como sigla RNG, entendiéndose que se trata del Aeropuerto Internacional José María Córdova se procede a cambiar la sigla a MDE, para evitar inconvenientes más adelante.

- En *Código* se hallaron casos donde estaba por ejemplo 3 y 03, es decir que tiene varios formatos para el mismo código, por lo que se homogenizo en uno solo (casos

6.3. EXPLORACIÓN DE LOS DATOS

En este paso se realizará la exploración con base en datos y gráficos estadísticos que nos ayudaran a detectar concentraciones, distribuciones y un primer acercamiento a cifras y aproximaciones, de manera resumida para explicar y descubrir patrones con mayor confianza y precisión, primero a nivel general, es decir en la BASE 0 para posteriormente enfocarnos en los retrasos y en tres aeropuertos de mediana operación.

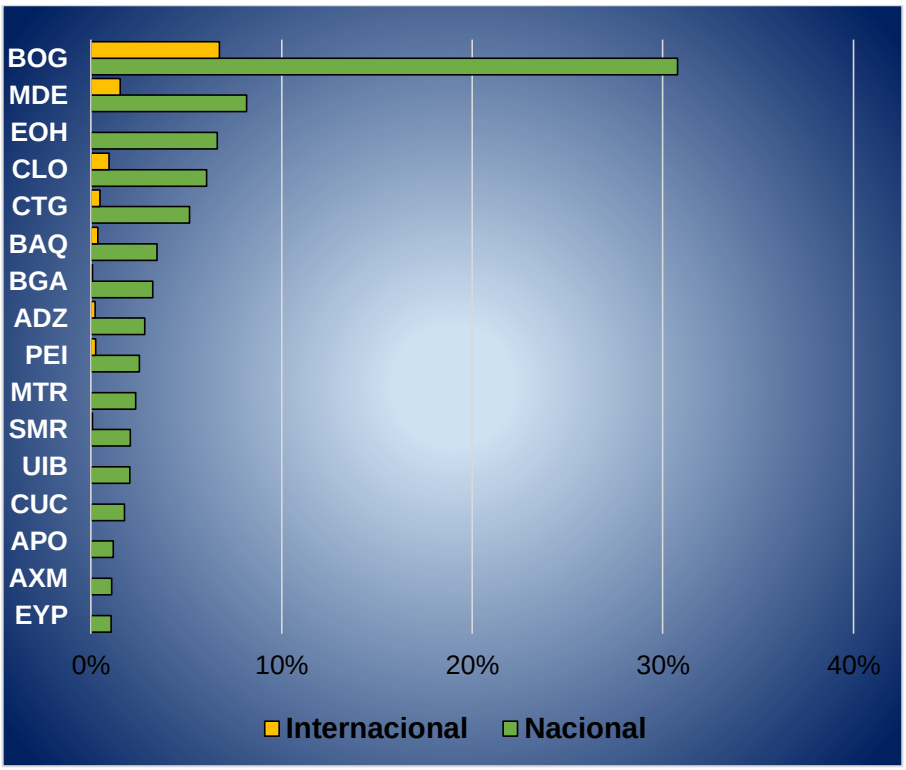

Ilustración 10 *Aeropuertos con mayor tráfico. Elaboración propia.*

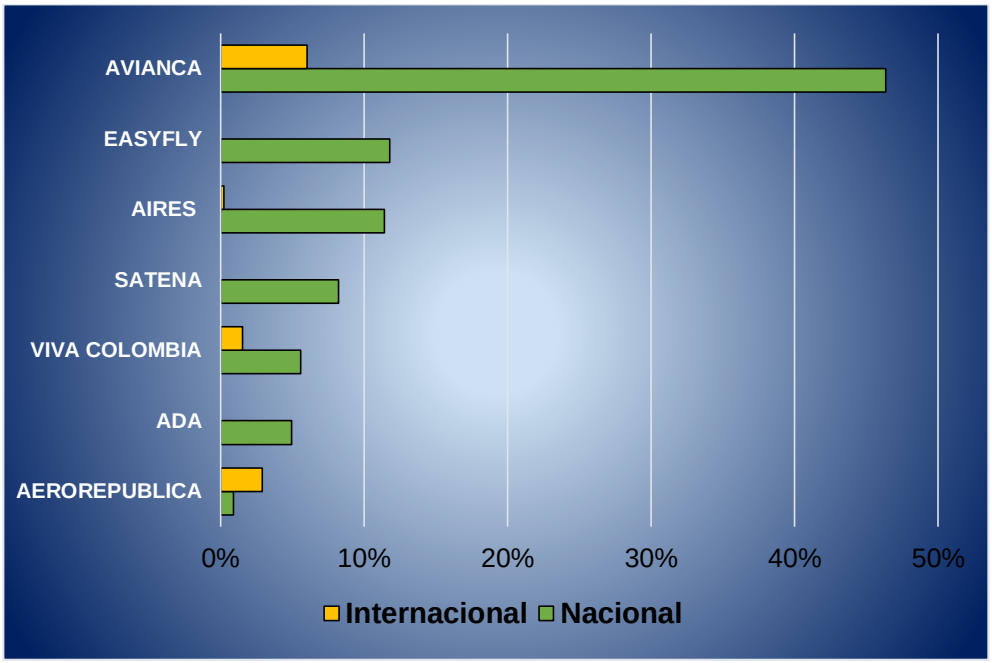

Ilustración 11 *Aerolíneas con mayor tráfico. Elaboración propia.*

Con la ilustración 12 se muestra el porcentaje de tráfico que tuvo cada aerolínea durante el 2017. A nivel nacional la data contiene información sobre Avianca (51,95%), Easyfly (13,23%), Aires (12,80%), Satena (9,21%), Viva Colombia (6,25%), Ada (5,54%) y Aerorepublica (1,03%), donde la primera en mencionar marca una diferencia muy grande respecto a las otras aerolíneas. A nivel internacional solo 4 aerolíneas hacen este tipo de operaciones, donde nuevamente Avianca (56,37%) posee el mayor número de vuelos, seguido por Aerorepublica (27,07%), Viva Colombia (14,20%) y Aires (2,35%).

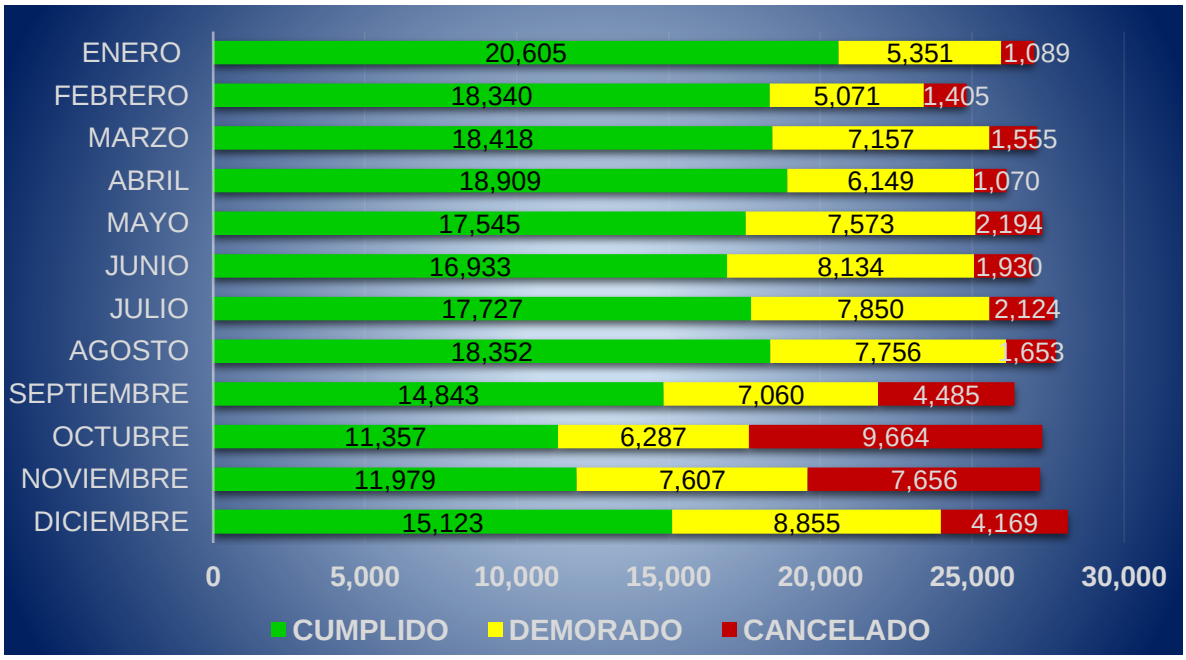

Ilustración 12 *Estados de vuelos por mes. Elaboración propia.*

En la ilustración anterior se muestra el porcentaje del estado de vuelo de las operaciones programadas a lo largo del año 2017, donde la mayoría de operaciones aéreas han sido cumplidas y con demoras, siendo los vuelos cancelados los de menor porcentaje. Hay que resaltar que en cuanto a los vuelos cancelados se ve un incremento grande en los últimos 4 meses del año debido a la huelga de pilotos que hacían parte de la aerolínea Avianca, en donde pedían mejores condiciones laborales. Debido a que el proyecto se basara solo en los vuelos con demoras este evento no perjudica el desarrollo la investigación. Para esta exploración del estado del vuelo no se tuvo en cuenta los datos “penalizado” ni “adelantado” porque no está contemplado en el régimen sancionatorio de las actividades aeronáuticas (RAC 13).

Para los estados de vuelo también se clasifico según los aeropuertos, siendo el aeropuerto de Providencia y Santa Catalina Islas (PVA) el más cumplido de Colombia, mientras que el más demorado viene siendo el de Tame Arauca (TME) y los que más adelantos y cancelaciones tuvieron fueron el de Puerto Carreño y el de Manizales respectivamente.

Para ir entrando en materia de las demoras, se procedió a usar la base de datos delimitada en el ítem 6.1.2, la **BASE 1** (79.431 datos) pues en ella ya se encuentran solamente las demoras de manera más clara y precisa con sus diferentes causas y tiempos a lo largo del año 2017. De tal manera representado en la ilustración 14 y 15 se evidencia que Avianca en tráfico nacional (39.027) e internacional (5.251) tiene mayor cantidad de retrasos que el resto, proporcional a la cantidad de operaciones a nivel general.

Ilustración 13 *Cantidad de demoras en aerolíneas en vuelos nacionales. Elaboración propia.*

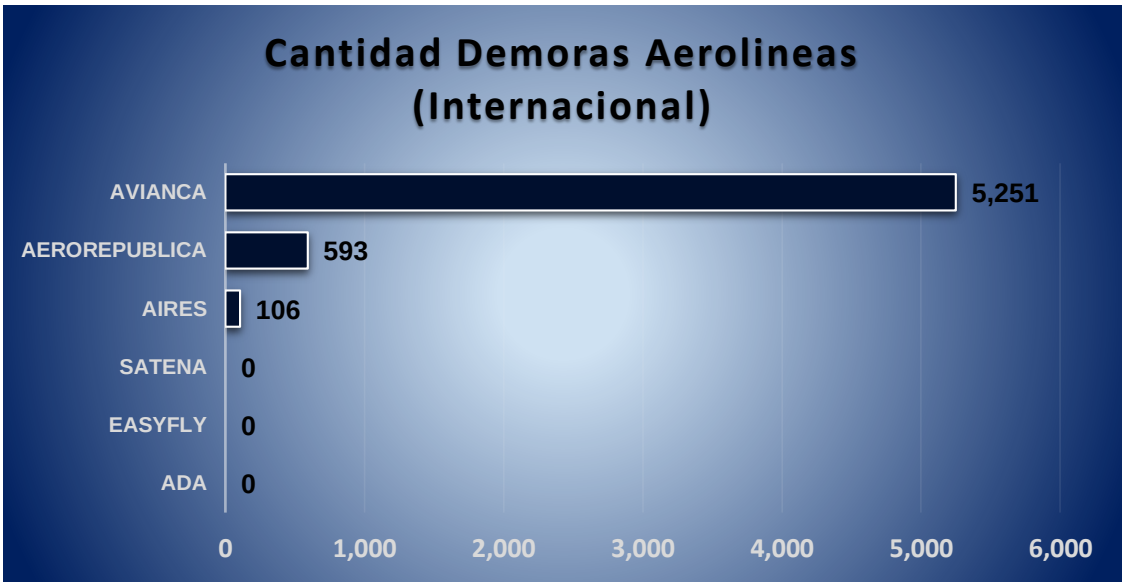

Ilustración 14 *Cantidad de demoras en aerolíneas en vuelos internacionales. Elaboración propia.*

Es así como podemos representar entonces con la ilustración 16 que en las aerolíneas Avianca (53,1%) Easyfly (17,7%) y Satena (11,6%) tienen el 82,3% de las demoras en los vuelos del país a nivel nacional. En cuanto a la operación internacional Avianca posee el 88,3% de las demoras, seguido por Aerorepublica y Aires como se evidencia en la ilustración 17.

Ilustración 15 *Pareto de demoras en aerolíneas en vuelos nacionales. Elaboración propia.*

Ilustración 16 *Pareto de demoras en aerolíneas en vuelos internacionales. Elaboración propia.*

Las cifras de las demoras en los aeropuertos también son representadas gráficamente, entendiendo a simple vista que los aeropuertos internacionales son los que mayores demoras tuvieron, como se evidencia en la ilustración 18.

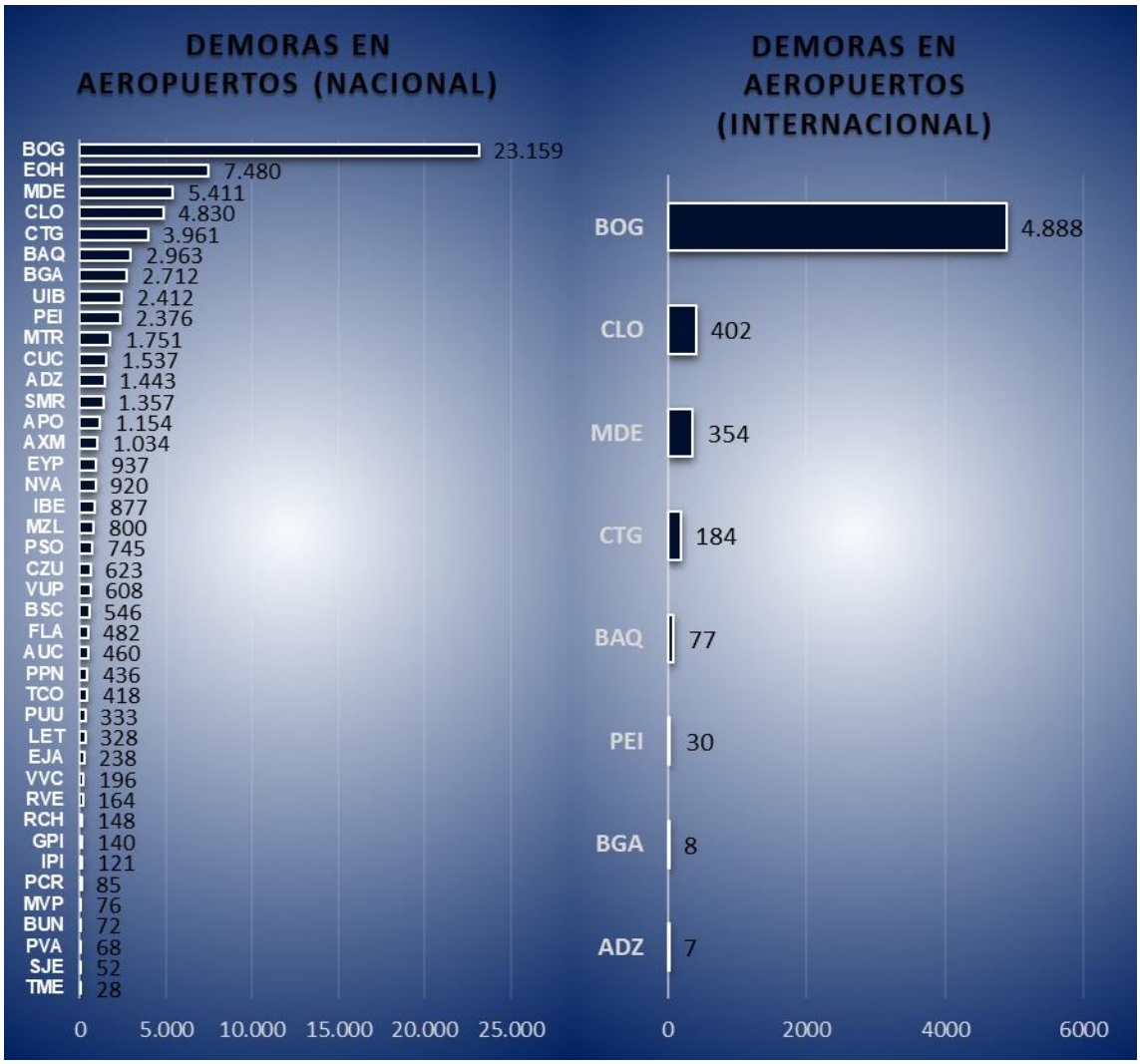

Ilustración 17 *Cantidad de demoras en aeropuertos en vuelos nacionales e internacionales. Elaboración propia.*

Con los aeropuertos los retrasos han estado presente en todos estos terminales aéreos, en donde el 81,6% de todos los retrasos nacionales están en Bogotá (31,5%), Medellín (10,2%), Rionegro (7,4%), Cali (6,6%), Cartagena (5,4%), Barranquilla (4%), Bucaramanga (3,7%), Quibdó (3,3%), Pereira (3%), Montería (2,4%), Cúcuta (2,1%) y San Andrés Isla (2%) como se representa en la ilustración

19. Así mismo el aeropuerto con mayor número de retrasos internacionales (82,2%) es el aeropuerto internacional el Dorado (ilustración 20).

Ilustración 18 *Pareto de demoras en aeropuertos en vuelos nacionales. Elaboración propia.*

Ilustración 19 *Pareto de demoras en aeropuertos en vuelos internacionales. Elaboración propia.*

Con lo anterior se puede observar que el aeropuerto con mayor tráfico y demoras internacional y nacional es notoriamente el de Bogotá, por lo que se decido escoger para este trabajo a tres aeropuertos cuyo tráfico fuese nacional e internacional, y a su vez representara el flujo aéreo restante de lo que abarca Bogotá, es así como en orden descendente de operaciones internacionales, y buscando aeropuertos no trabajados en otros proyectos de investigación similares, se decidió por el Aeropuerto Internacional José María Córdova (Medellín), Aeropuerto Internacional Gustavo Rojas Pinilla (San Andrés) y Aeropuerto Internacional Matecaña (Pereira).

Para la elección de los tres aeropuertos se usó la BASE 1, donde se excluyó a los demás aeropuertos cuyo origen no estuvo en Medellín (MDE), Pereira (PEI) y San Andrés (ADZ), obteniendo 9.621 datos para la nueva base de datos nombrada **BASE 2 Tres aeropuertos.**

Ya con la nueva base de datos nos damos cuenta que la cantidad de demoras es proporcional a la cantidad de operaciones, para MDE y ADZ, excepto para PEI quien tiene un gran número de demoras nacionales, pero es el que menos vuelos nacionales tuvo en aquel año entre estos tres terminales, así se representa en la tabla 6 y gráficamente en la ilustración 20.

AEROPUERTO	TRAFICO	CANTIDAD VUELOS	CANTIDAD RETRASOS	FRECUENCIA RETRASOS
MDE	Internacional	4.125	354	8,58%
	Nacional	28.081	5.411	19,27%
PEI	Internacional	689	30	4,35%
	Nacional	8.562	2.376	27,75%
ADZ	Internacional	373	7	1,88%
	Nacional	9.772	1.443	14,77%

Tabla 6 *Cifras de aeropuertos seleccionados. Elaboración propia.*

Ilustración 20 *Porcentaje de retrasos en cada aeropuerto seleccionado. Elaboración propia.*

Ya con esta nueva base de datos donde solo están los tres aeropuertos que necesitamos y con las demoras solamente, se analizó el comportamiento de los tiempos como se aprecia en la tabla 7, donde se encontró que los tres aeropuertos tienen el mismo tiempo mínimo de retraso de 16 minutos. En cuanto al tiempo máximo de demora si es proporcional a las operaciones. El promedio de demora y la desviación están de la hora (60 minutos) en adelante.

AEROPUERTO	TRAFICO	TIEMPO DE RETRASOS (MIN)					
		TOTAL	MEDIA	MEDIANA	DESVIACION	MIN	MAX
MDE	Internacional	4.125	62	35	105	16	1.298
	Nacional	28.081	59	39	60	16	1.242
PEI	Internacional	689	107	34	166	16	790
	Nacional	8.562	79	50	82	16	914
ADZ	Internacional	373	112	26	146	16	363
	Nacional	9.772	60	34	87	16	1.142

Tabla 7 *Análisis descriptivo del tiempo de retardo en aeropuertos de interés. Elaboración propia.*

Las causas de las demoras se analizaron inicialmente en conjunto para los tres aeropuertos en el tráfico internacional, como se representa en la ilustración 22, encontrándose que el acumulado de la frecuencia absoluta hasta el 80% de los códigos de demora definidos por el IATA presentes son:

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
73	11,00%	11,00%	AEROPUERTO ALTERNO DE DESTINO O EN RUTA	Incontrolables	Externo
81	9,21%	20,20%	ATFM (Air Traffic Flow Management) DEBIDO A LIMITACION DE CAPACIDAD DEL SISTEMA DE ATC EN RUTA O A ALTA DEMANDA	RAC	Externo
41	7,93%	28,13%	DEFECTOS DEL AVION	Técnicas	Interno
83	5,63%	33,76%	ATFM DEBIDO A RESTRICCIÓN EN AEROPUERTO DE DESTINO	Incontrolables	Externo
71	5,37%	39,13%	AEROPUERTO DE SALIDA.	Incontrolables	Externo
87	4,86%	43,99%	INSTALACIONES AEROPORTUARIAS	AGA	Externo
89.1	4,60%	48,59%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	RAC	Externo
3	3,84%	52,43%	INCONTROLABLES CLIENTES	Operacionales	Externo
93	3,58%	56,01%	ROTACIÓN DE AERONAVES	Operacionales	Interno
84	3,58%	59,59%	ATFM DEBIDO A CONDICIONES METEOROLÓGICAS EN EL AEROPUERTO DE DESTINO.	Incontrolables	Externo
51	3,58%	63,17%	DAÑO DURANTE OPERACIONES DE VUELO	Técnicas	Externo
65	3,32%	66,50%	SOLICITUDES ESPECIALES DE LA TRIPULACIÓN TÉCNICA	Operacionales	Interno
72	3,07%	69,57%	AEROPUERTO DE DESTINO	Incontrolables	Externo
52	3,07%	72,63%	DAÑO DURANTE OPERACIONES DE TIERRA	Técnicas	Externo
42	2,30%	74,94%	MANTENIMIENTO PROGRAMADO	Técnicas	Interno
4	2,05%	76,98%	VUELOS BORRADOS	Operacionales	Interno
67	2,05%	79,03%	FALTANTE DE TRIPULACIÓN DE CABINA	Operacionales	Interno
63	2,05%	81,07%	ABORDAJE TARDÍO DE LA TRIPULACIÓN TÉCNICA O PROCEDIMIENTOS DE SALIDA DEMORADOS	Operacionales	Interno

Tabla 8 *Causas de demoras internacionales en los tres aeropuertos. Elaboración propia.*

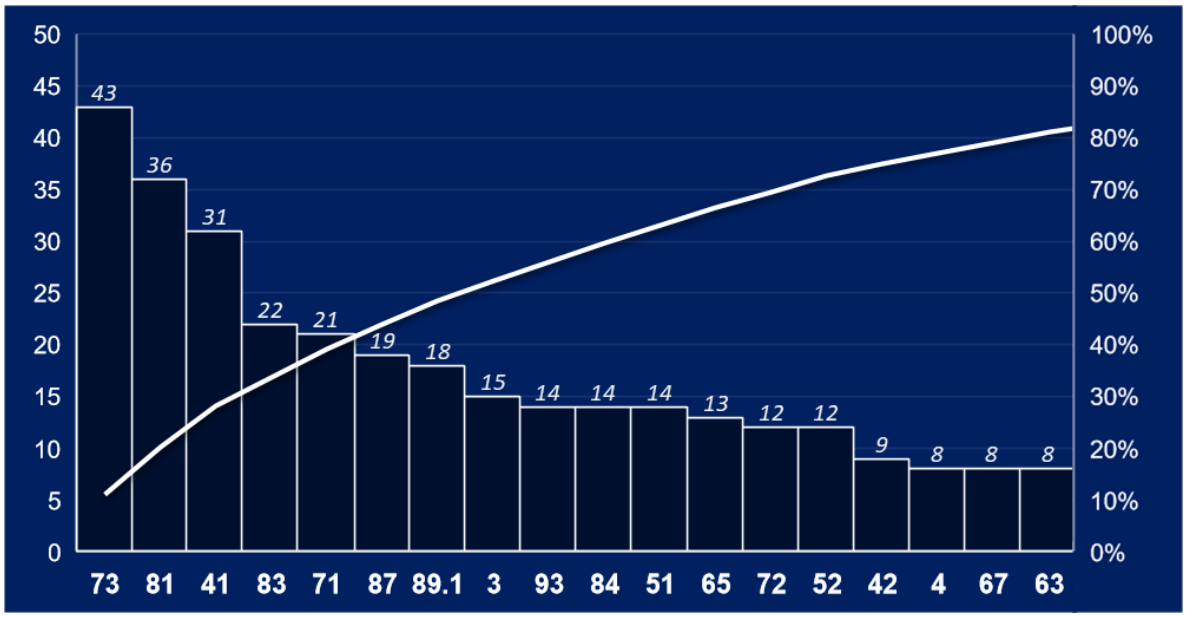

Ilustración 21 *Pareto de causas internacional en los tres aeropuertos. Elaboración propia.*

Para el tráfico nacional también se buscó el acumulado de frecuencia absoluta de los tres aeropuertos en conjunto, para así ver en que códigos hay hasta el 80% de las causas de demoras en estas terminales aéreas representado en la ilustración 23, encontrando que esta cantidad de las causalidades de retrasos son:

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
81	15,42%	15,42%	ATFM (Air Traffic Flow Management) DEBIDO A LIMITACION DE CAPACIDAD DEL SISTEMA DE ATC EN RUTA O A ALTA DEMANDA	RAC	Externo
41	10,50%	25,92%	DEFECTOS DEL AVION	Técnicas	Interno
72	9,33%	35,24%	AEROPUERTO DE DESTINO.	Incontrolables	Externo
89.1	7,25%	42,49%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	RAC	Externo
71	6,83%	49,32%	AEROPUERTO DE SALIDA.	Incontrolables	Externo
83	6,42%	55,74%	ATFM DEBIDO A RESTRICCIÓN EN AEROPUERTO DE DESTINO	Incontrolables	Externo
87	4,54%	60,28%	INSTALACIONES AEROPORTUARIAS	AGA	Externo
52	4,11%	64,39%	DAÑO DURANTE OPERACIONES DE TIERRA	Técnicas	Externo
84	3,04%	67,43%	ATFM DEBIDO A CONDICIONES METEOROLÓGICAS EN EL AEROPUERTO DE DESTINO.	Incontrolables	Externo

65	2,88%	70,31%	SOLICITUDES ESPECIALES DE LA TRIPULACIÓN TÉCNICA	Operacionales	Interno
73	2,74%	73,06%	AEROPUERTO ALTERNO DE DESTINO O EN RUTA	Incontrolables	Externo
3	2,70%	75,75%	INCONTROLABLES CLIENTES	Operacionales	Externo
51	2,43%	78,18%	DAÑO DURANTE OPERACIONES DE VUELO	Técnicas	Externo
15	1,71%	79,89%	ABORDAJE	Operacionales	Interno
88	1,69%	81,58%	RESTRICCIONES EN EL AEROPUERTO DE DESTINO	Incontrolables	Externo

Tabla 9 *Causas de demoras nacionales en los tres aeropuertos. Elaboración propia.*

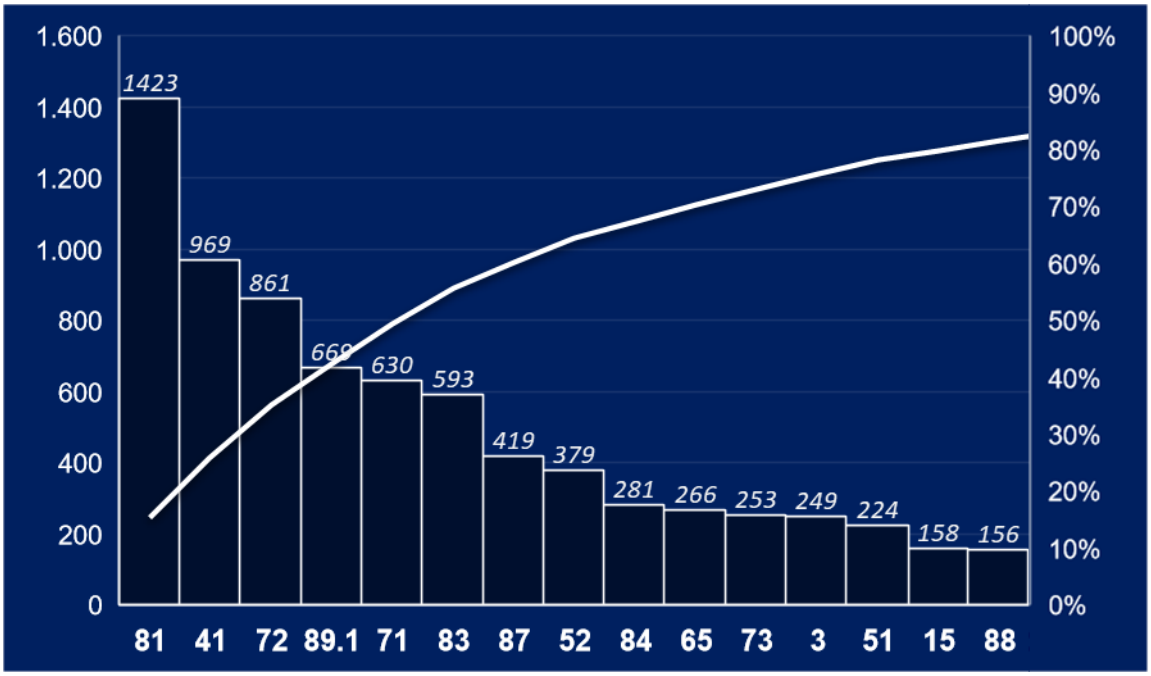

Ilustración 22 *Pareto de causas nacional en los tres aeropuertos. Elaboración propia.*

Para cada aeropuerto se hizo el mismo procedimiento anterior, tanto para el tráfico nacional como el internacional, y encontrándonos que las causas más comunes para los tres aeropuertos en el tráfico internacional son 73 (externa), 81 (externa), 41 (interna) y 71 (externa), así mismo para el tráfico nacional las causas de demoras más comunes en estos aeropuertos son 81 (externa), 41 (interna), 89

(externa) y 71 (externa). Otra particularidad que se encontró en las causas de retrasos es que el análisis de las causas en conjunto para los tres aeropuertos se comporta de manera similar a las causas del aeropuerto de Medellín, esto debido a que es el aeropuerto con mayor cantidad de vuelos.

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
73	10,45%	10,45%	AEROPUERTO ALTERNO DE DESTINO O EN RUTA	Incontrolables	Externo
81	9,32%	19,77%	ATFM (Air Traffic Flow Management) DEBIDO A LIMITACION DE CAPACIDAD DEL SISTEMA DE ATC EN RUTA O A ALTA DEMANDA	RAC	Externo
41	7,34%	27,12%	DEFECTOS DEL AVION	Técnicas	Interno
83	6,21%	33,33%	ATFM DEBIDO A RESTRICCIÓN EN AEROPUERTO DE DESTINO	Incontrolables	Externo
87	5,08%	42,66%	INSTALACIONES AEROPORTUARIAS	AGA	Externo
71	4,24%	37,57%	AEROPUERTO DE SALIDA.	Incontrolables	Externo
3	3,95%	50,56%	INCONTROLABLES CLIENTES	Operacionales	Externo
89.1	3,95%	46,61%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	RAC	Externo
65	3,67%	64,41%	SOLICITUDES ESPECIALES DE LA TRIPULACIÓN TÉCNICA	Operacionales	Interno
51	3,67%	60,73%	DAÑO DURANTE OPERACIONES DE VUELO	Técnicas	Externo
72	3,39%	67,80%	AEROPUERTO DE DESTINO	Incontrolables	Externo
84	3,39%	57,06%	ATFM DEBIDO A CONDICIONES METEOROLÓGICAS EN EL AEROPUERTO DE DESTINO.	Incontrolables	Externo
93	3,11%	53,67%	ROTACIÓN DE AERONAVES	Operacionales	Interno
52	3,11%	70,90%	DAÑO DURANTE OPERACIONES DE TIERRA	Técnicas	Externo
42	2,54%	73,45%	MANTENIMIENTO PROGRAMADO	Técnicas	Interno
4	2,26%	75,71%	VUELOS BORRADOS	Operacionales	Interno
67	2,26%	77,97%	FALTANTE DE TRIPULACIÓN DE CABINA	Operacionales	Interno
63	2,26%	80,23%	ABORDAJE TARDÍO DE LA TRIPULACIÓN TÉCNICA O PROCEDIMIENTOS DE SALIDA DEMORADOS	Operacionales	Interno

Tabla 10 *Causas de demoras internacionales en MDE. Elaboración propia.*

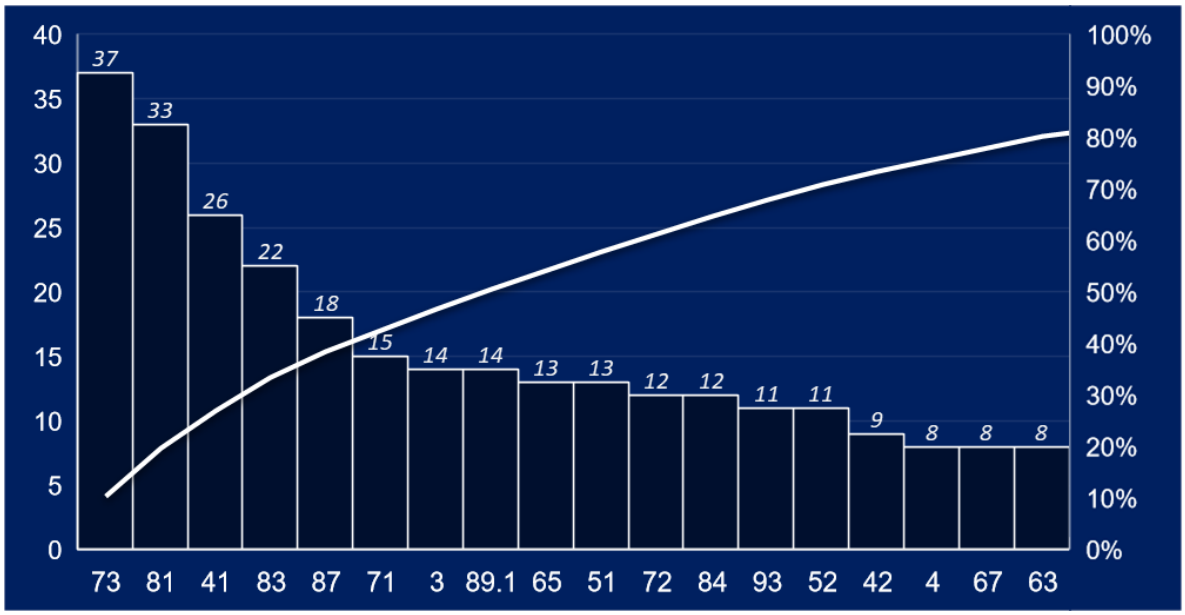

Ilustración 23 *Pareto de causas internacional en MDE. Elaboración propia.*

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
81	15,51%	15,51%	ATFM (Air Traffic Flow Management) DEBIDO A LIMITACION DE CAPACIDAD DEL SISTEMA DE ATC EN RUTA O A ALTA DEMANDA	RAC	Externo
41	10,61%	26,11%	DEFECTOS DEL AVION	Técnicas	Interno
89.1	8,21%	34,32%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	RAC	Externo
72	7,65%	41,97%	AEROPUERTO DE DESTINO.	Incontrolables	Externo
83	6,36%	48,33%	ATFM DEBIDO A RESTRICCIÓN EN AEROPUERTO DE DESTINO	Incontrolables	Externo
71	6,06%	54,39%	AEROPUERTO DE SALIDA.	Incontrolables	Externo
87	4,79%	59,18%	INSTALACIONES AEROPORTUARIAS	AGA	Externo
52	4,68%	63,85%	DAÑO DURANTE OPERACIONES DE TIERRA	Técnicas	Externo
65	4,03%	67,88%	SOLICITUDES ESPECIALES DE LA TRIPULACIÓN TÉCNICA	Operacionales	Interno
84	3,33%	71,21%	ATFM DEBIDO A CONDICIONES METEOROLÓGICAS EN EL AEROPUERTO DE DESTINO.	Incontrolables	Externo
3	2,90%	74,11%	INCONTROLABLES CLIENTES	Operacionales	Externo
51	2,85%	76,95%	DAÑO DURANTE OPERACIONES DE VUELO	Técnicas	Externo

15	2,31%	79,26%	ABORDAJE	Operacionales	Interno
43	1,57%	80,84%	MANTENIMIENTO NO PROGRAMADO	Técnicas	Interno

Tabla 11 *Causas de demoras nacionales en MDE. Elaboración propia.*

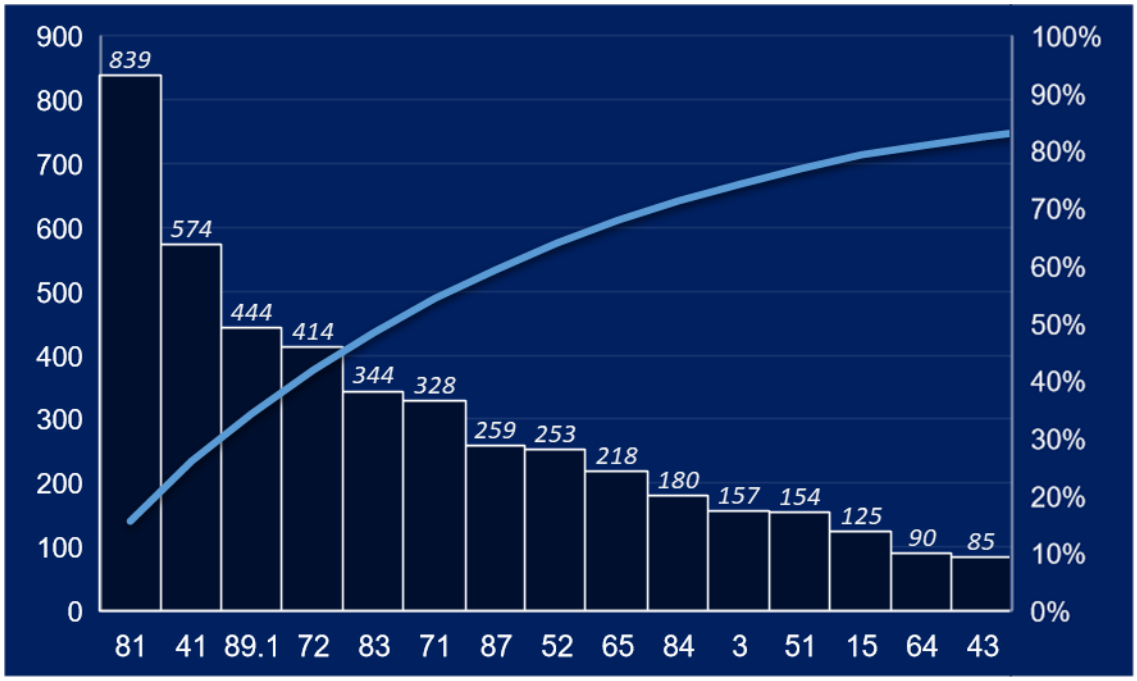

Ilustración 24 *Pareto de causas nacional en MDE. Elaboración propia.*

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
71	20,00%	20,00%	AEROPUERTO DE SALIDA.	Incontrolables	Externo
73	13,33%	33,33%	AEROPUERTO ALTERNO DE DESTINO O EN RUTA	Incontrolables	Externo
89.1	13,33%	46,67%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	RAC	Externo
41	10,00%	56,67%	DEFECTOS DEL AVION	Técnicas	Interno
81	10,00%	66,67%	ATFM (Air Traffic Flow Management) DEBIDO A LIMITACION DE CAPACIDAD DEL SISTEMA DE ATC EN RUTA O A ALTA DEMANDA	RAC	Externo
84	6,67%	73,33%	ATFM DEBIDO A CONDICIONES METEOROLÓGICAS EN EL AEROPUERTO DE DESTINO.	Incontrolables	Externo

3	3,33%	76,67%	INCONTROLABLES CLIENTES	Operacionales	Externo
33	3,33%	80,00%	EQUIPO DE CARGUE	Operacionales	Interno
51	3,33%	83,33%	DAÑO DURANTE OPERACIONES DE VUELO	Técnicas	Externo
93	3,33%	86,67%	ROTACIÓN DE AERONAVES	Operacionales	Interno
52	3,33%	90,00%	DAÑO DURANTE OPERACIONES DE TIERRA	Técnicas	Externo
52.1	3,33%	93,33%	DAÑO DURANTE OPERACIONES DE TIERRA,	Técnicas	Interno
87	3,33%	96,67%	INSTALACIONES AEROPORTUARIAS	AGA	Externo
89.3	3,33%	100,00%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	Incontrolables	Externo

Tabla 12 *Causas de demoras internacionales en PEI. Elaboración propia.*

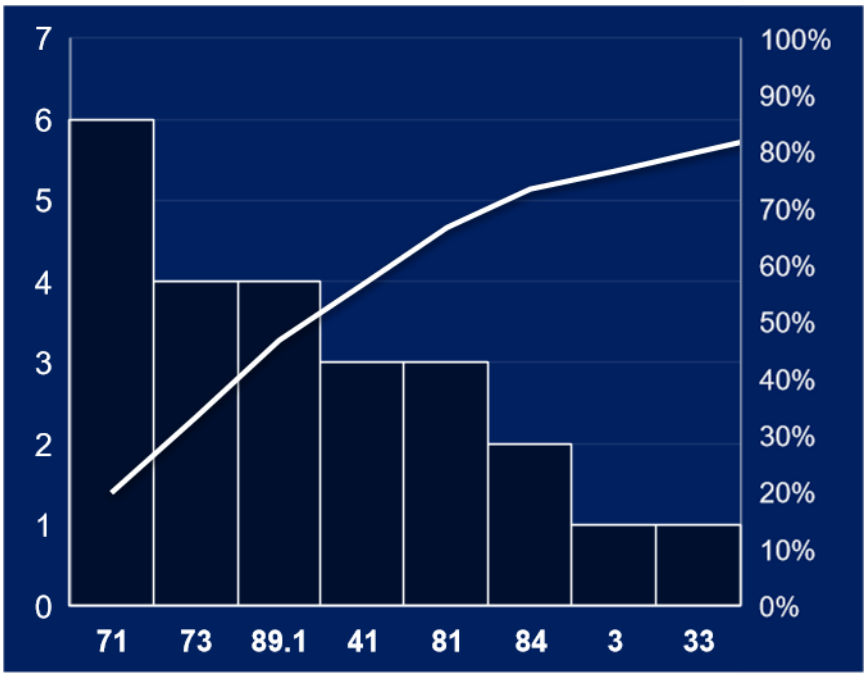

Ilustración 25 *Pareto de causas internacional en PEI. Elaboración propia.*

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
72	16,84%	16,84%	AEROPUERTO DE DESTINO.	Incontrolables	Externo
81	12,25%	29,08%	ATFM (Air Traffic Flow Management) DEBIDO A LIMITACION DE CAPACIDAD DEL SISTEMA DE	RAC	Externo

			ATC EN RUTA O A ALTA DEMANDA		
71	8,92%	38,01%	AEROPUERTO DE SALIDA.	Incontrolables	Externo
83	8,71%	46,72%	ATFM DEBIDO A RESTRICCIÓN EN AEROPUERTO DE DESTINO	Incontrolables	Externo
41	7,91%	54,63%	DEFECTOS DEL AVION	Técnicas	Interno
73	6,65%	61,28%	AEROPUERTO ALTERNO DE DESTINO O EN RUTA	Incontrolables	Externo
89.1	4,42%	65,70%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	RAC	Externo
46	4,34%	70,03%	CAMBIO DE AVION	Técnicas	Interno
87	3,87%	73,91%	INSTALACIONES AEROPORTUARIAS	AGA	Externo
84	2,86%	76,77%	ATFM DEBIDO A CONDICIONES METEOROLÓGICAS EN EL AEROPUERTO DE DESTINO.	Incontrolables	Externo
52	2,65%	79,42%	DAÑO DURANTE OPERACIONES DE TIERRA	Técnicas	Externo
85	2,02%	81,44%	MEDIDAS DE SEGURIDAD MANDATORIAS	Incontrolables	Externo

Tabla 13 *Causas de demoras nacionales en PEI. Elaboración propia.*

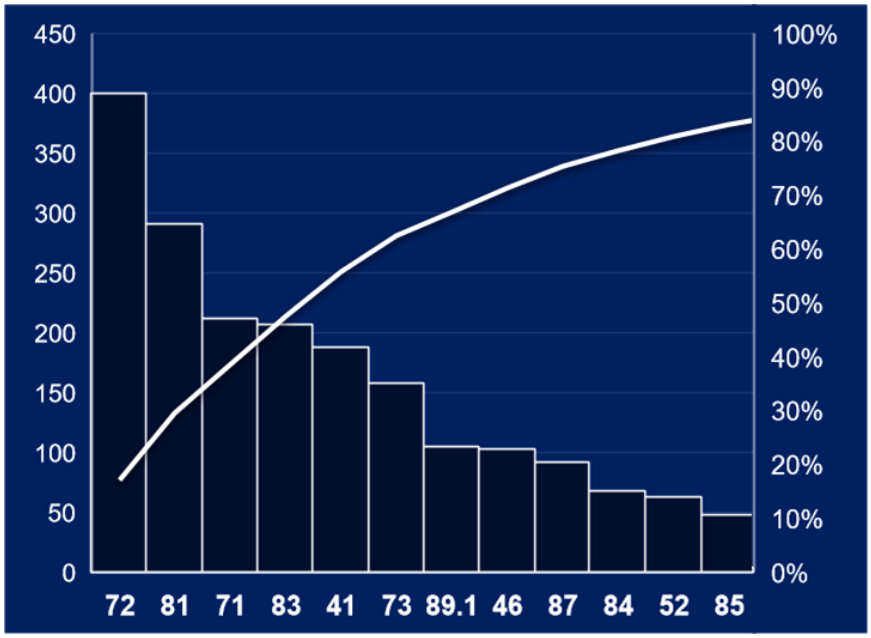

Ilustración 26 *Pareto de causas nacional en PEI. Elaboración propia.*

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
41	28,57%	29%	DEFECTOS DEL AVION	Técnicas	Interno
73	28,57%	57%	AEROPUERTO ALTERNO DE DESTINO O EN RUTA	Incontrolables	Externo
93	28,57%	86%	ROTACIÓN DE AERONAVES	Operacionales	Interno
97	14,29%	100%	MOVIMIENTO SINDICAL DENTRO DE LA COMPAÑÍA	Incontrolables	Externo

Tabla 14 *Causas de demoras internacionales en ADZ. Elaboración propia.*

CODIGO	CANTIDAD %	% ACOMULADO	EVENTO	MOTIVO	CAUSA
81	20,30%	20,30%	ATFM (Air Traffic Flow Management) DEBIDO A LIMITACION DE CAPACIDAD DEL SISTEMA DE ATC EN RUTA O A ALTA DEMANDA	RAC	Externo
89.1	8,32%	28,62%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	RAC	Externo
41	14,35%	42,97%	DEFECTOS DEL AVION	Técnicas	Interno
71	6,24%	49,20%	AEROPUERTO DE SALIDA.	Incontrolables	Externo
87	4,71%	53,92%	INSTALACIONES AEROPORTUARIAS	AGA	Externo
52	4,37%	58,28%	DAÑO DURANTE OPERACIONES DE TIERRA	Técnicas	Externo
88	4,02%	62,30%	RESTRICCIONES EN EL AEROPUERTO DE DESTINO	Incontrolables	Externo
3	3,47%	65,77%	INCONTROLABLES CLIENTES	Operacionales	Externo
72	3,26%	69,02%	AEROPUERTO DE DESTINO.	Incontrolables	Externo
83	2,91%	71,93%	ATFM DEBIDO A RESTRICCIÓN EN AEROPUERTO DE DESTINO	Incontrolables	Externo
84	2,29%	74,22%	ATFM DEBIDO A CONDICIONES METEOROLÓGICAS EN EL AEROPUERTO DE DESTINO.	Incontrolables	Externo
51	1,94%	76,16%	DAÑO DURANTE OPERACIONES DE VUELO	Técnicas	Externo
89	1,59%	77,75%	RESTRICCIONES EN EL AEROPUERTO DE ORIGEN CON O SIN RESTRICCIONES DE AFTM	AGA	Externo
15	1,52%	79,28%	ABORDAJE	Operacionales	Interno
11	1,46%	80,73%	CHEQUEO TARDIO	Operacionales	Interno

Tabla 15 *Causas de demoras nacionales en ADZ. Elaboración propia.*

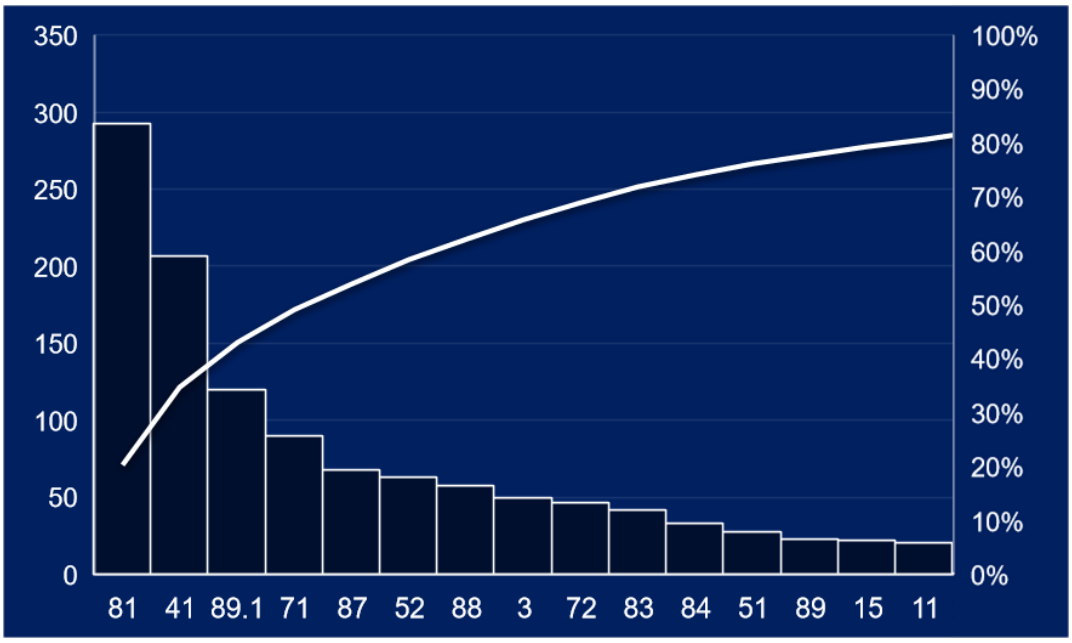

Ilustración 27 *Pareto de causas nacional en ADZ. Elaboración propia.*

Después de conocer la causalidad de cada aeropuerto es necesario identificar las demoras según el tipo de causa, externa o interna, para empezar a entrar al modelamiento. Por lo que usando los análisis anteriores se obtuvo la tabla 16, donde se representa de manera resumida para cada aeropuerto los códigos de demora según el tipo de tráfico nacional o internacional y según el tipo de causa externa o interna y además su representación porcentual, de la cual cabe destacar que la mayoría de causas son por factores que no se pueden controlar, es decir que son externas, mientras que las causas internas son pocas en los tres aeropuertos.

	MDE				PEI				ADZ			
	Internacional		Nacional		Internacional		Nacional		Internacional		Nacional	
	Internas	Externas	Internas	Externas	Internas	Externas	Internas	Externas	Internas	Externas	Internas	Externas
CODIGOS DEMORAS	41	73	41	81	41	71	41	72	41	73	41	81
	65	81	65	89.1	33	73	46	81	93	97	15	89.1
	93	83	15	72	93	89.1		71			11	71
	42	87	43	83	52.1	81		83				87
	4	71		71		84		73				52
	67	3		87		3		89.1				88
	63	89.1		52		51		87				3
		51		84		52		84				72
		72		3		87		52				83
		84		51		89.3		85				84
		52										51
												89
FRECUENCIA	23,45%	56,78%	18,52%	62,32%	20,00%	80,00%	12,25%	69,19%	57,14%	42,86%	17,33%	63,41%
ACOMULADA	80,23%		80,84%		100,00%		81,44%		100,00%		80,74%	

Tabla 16 *Resumen del tipo de causa en los tres aeropuertos. Elaboración propia.*

Teniendo ya definido el tipo de causas en cada aeropuerto y para cada tráfico, pasamos a realizar el análisis descriptivo de los tiempos de estas mismas causas aeropuerto por aeropuerto encontrándonos con los siguientes resultados:

ANALISIS DESCRIPTIVO DE CAUSAS INTERNAS EN MDE							
TRAFICO	CAUSA	FRECUENCIA	TIEMPO TOTAL (min)	MEDIA (min)	DESVIACIÓN (min)	MIN (min)	MAX (min)
INTERNACIONAL	**41**	7,34%	4.000	153,85	255,44	19	1.298
	65	3,67%	1.198	92,15	164,47	16	608
	93	3,11%	413	37,55	27,50	17	115
	42	2,54%	538	59,78	29,11	22	114
	4	2,26%	251	31,38	17,55	19	74
	67	2,26%	1.111	138,88	310,12	16	906
	63	2,26%	247	30,88	10,55	18	46
NACIONAL	**41**	10,61%	38.070	66,32	62,20	16	492
	65	4,03%	8.461	38,81	26,01	16	171
	15	2,31%	4.513	36,10	23,88	16	135
	43	1,57%	5.765	67,82	49,14	17	234

Tabla 17 *Análisis descriptivo causas internas MDE. Elaboración propia.*

ANALISIS DESCRIPTIVO DE CAUSAS INTERNAS EN PEI							
TRAFICO	CAUSA	FRECUENCIA	TIEMPO TOTAL (min)	MEDIA (min)	DESVIACIÓN (min)	MIN (min)	MAX (min)
INTERNACIONAL	41	10,00%	147	49	38,16	25	93
	33	3,33%	391	391	N/A	391	391
	93	3,33%	38	38	N/A	38	38
	52.1	3,33%	20	20	N/A	20	20
NACIONAL	41	7,91%	15477	82,32	68,59	16	340
	46	4,34%	8417	81,72	56,73	17	306

Tabla 18 *Análisis descriptivo causas internas PEI. Elaboración propia.*

ANALISIS DESCRIPTIVO DE CAUSAS INTERNAS EN ADZ							
TRAFICO	CAUSA	FRECUENCIA	TIEMPO TOTAL (min)	MEDIA (min)	DESVIACIÓN (min)	MIN (min)	MAX (min)
INTERNACIONAL	41	28,57%	389	194,50	238,29	26	363
	93	28,57%	37	18,50	3,54	16	21
NACIONAL	41	14,35%	15501	74,88	62,49	16	382
	15	1,52%	736	33,45	16,49	16	83
	11	1,46%	579	27,57	12,44	16	56

Tabla 19 *Análisis descriptivo causas internas ADZ. Elaboración propia.*

ANALISIS DESCRIPTIVO DE CAUSAS EXTERNAS							
AEROPUERTO	TRAFICO	FRECUENCIA	TIEMPO TOTAL (min)	MEDIA (min)	DESVIACIÓN (min)	MIN (min)	MAX (min)
MDE	INTERNACIONAL	56,76%	10.070	50	41	16	259
	NACIONAL	62,34%	197.471	59	59	16	1.142
PEI	INTERNACIONAL	79,98%	2.615	109	174	16	790
	NACIONAL	69,19%	136.950	83	85	16	914
ADZ	INTERNACIONAL	42,86%	356	119	141	26	281
	NACIONAL	63,42%	53.003	58	90	16	1.142

Tabla 20 *Análisis descriptivo causas externas en aeropuertos seleccionados. Elaboración propia.*

6.4. BUSQUEDA DE PREDICCIÓN A LA PROBLEMÁTICA

Para la exploración de la solución se decide utilizar el modelo para identificar los vuelos afectados por retrasos o cancelaciones en el aeropuerto El Dorado de Bogotá, Colombia, realizado por Pulido, Arias, Chavarro y Ramírez [33] que tiene las siguientes características:

Software utilizado

- KNIME: plataforma de minería de datos que permite la creación de ETL y modelos en un entorno visual. Está construido bajo la plataforma Eclipse.
- Microsoft SQL Server: sistema de gestión de base de datos relacional, desarrollado por la empresa Microsoft.
- Power BI: herramienta de visualización interactivas y capacidades de inteligencia empresarial, desarrollado por la empresa Microsoft.
- R Studio: es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, dedicado a la computación estadística y gráficos.

6.4.1. Variables escogidas

Teniendo en cuenta que la mayoría de las variables del set de datos obtenido son categóricas, se crearon una serie de atributos adicionales con el objetivo que aporten al entendimiento del negocio.

Los nombres de las variables de color verde fueron creados a partir de la información de las bases de datos de reporte de cumplimiento de itinerarios y tráfico aéreo mensual.

#	Atributo	Descripción	Tipo
1	**Afectado (variable objetivo)**	Variable Objetivo. Describe si el vuelo fue afectado por un retraso mayor a 15 min o por una cancelación.	Categórica
2	**Tráfico**	Nacional o Internacional	Categórica
3	**Aerolínea**	Nombre comercial de la empresa.	Categórica
4	**Origen**	Corresponde a la sigla IATA del aeropuerto donde se origina el trayecto.	Categórica
5	**Destino**	Corresponde a la sigla IATA del aeropuerto donde termina el trayecto.	Categórica
6	**Año**	Año en el que está programado el vuelo	Categórica
7	**Mes**	Mes en el que está programado el vuelo	Categórica
8	**Día del mes**	Día del mes en el que está programado el vuelo	Categórica
9	**Día de la semana**	Día de la semana en el que está programado el vuelo.	Categórica
10	**Hora (Minutos)**	Hora en el que está programado el vuelo. Valor numérico de 0 (00:00:00) a 1439 (23:59:00)	Número
11	**Distancia**	Distancia en kilómetros entre los aeropuertos del respectivo trayecto.	Número
12	**Pasajeros**	Cantidad de pasajeros promedio para una ruta definida por mes y aerolínea.	Número
13	**Sillas Disponibles**	Cantidad de sillas disponibles promedio para una ruta definida por mes y aerolínea.	Número
14	**Tiempo Vuelo**	Tiempo promedio del vuelo.	Número
15	**Días última afectación**	Cantidad de días transcurridos desde la última afectación del mismo número de vuelo.	Número
16	**Total, afectado semana anterior**	Cantidad total de veces en las que el vuelo fue afectado durante la semana anterior.	Número
17	**Afectado semana anterior**	Indica si el número de vuelo de una aerolínea fue afectado la semana anterior.	Número
18	**Afectado día anterior**	Indica si el número de vuelo de una aerolínea fue afectado el día anterior.	Número
19	**Afectada semana actual**	Cantidad de veces en las que el vuelo fue afectado durante los siete días previos al día programado de salida.	Número
20	**Cantidad vuelos Hora**	Cantidad de vuelos programados para despegar y aterrizar en el Aeropuerto El Dorado de Bogotá.	Número

Tabla 21 *Variables numéricas y categóricas [33].*

6.4.2. Selección de técnicas y supuestos

Para la etapa de modelamiento se proponen las siguientes técnicas:

- **Regresión logística:** método que permite estimar la probabilidad de una variable cualitativa binaria en función de un conjunto de variables predictoras que pueden ser, tanto continuas como categóricas.

- **Redes neuronales:** es un procesador distribuido en paralelo de forma masiva con una propensión natural a almacenar conocimiento experimental y convertirlo en disponible para su uso.

- **XGBoosting:** es una variante o aplicación específica del concepto de “Gradient boosting”, que a su vez es una técnica empleada para problemas de regresión y clasificación, cuyo resultado es un modelo del ensamblaje de modelos (árboles de decisión) más débiles. El aporte específico de XGBoosting (acrónimo derivado de la expresión eXtreme Gradient Boosting) parte de la aplicación de métodos adicionales de regularización para lograr resultados de calidad y de forma muy rápida.

6.4.3. Conjunto de datos de entrenamiento y prueba.

La base de datos se divide en dos conjuntos de datos:

- Entrenamiento: subconjunto para entrenar el modelo (75% de los datos).
- Prueba: subconjunto para probar el modelo entrenado (25% de los datos).

Los subconjuntos de datos son suficientemente grandes para generar resultados significativos desde el punto de vista estadístico.

6.4.4. Modelamiento

Los modelos por implementar se evaluarán con métricas como accuracy, precisión, sensitividad y especificidad. El accuracy indica el porcentaje que el modelo logra predecir correctamente, tanto los vuelos afectados (TP) como los que cumplen con el itinerario programado (TN); la precisión indica el porcentaje de vuelos que se logran predecir con afectación (TP) del total de predicciones realizadas; mientras que la sensitividad es el porcentaje de vuelos que se predicen con afectación (TP), teniendo en cuenta los valores de referencia de la base de datos de prueba que estaban marcados con afectación. Es decir, la precisión tiene en cuenta los falsos positivos (FP) y la sensitividad los falsos negativos (FN). Finalmente, la especificidad expresa el porcentaje de acierto en los vuelos que cumplen con el itinerario (TN), con relación al total de vuelo no afectados en la base de datos de referencia. Para este caso se le dará mayor importancia a los modelos que logren mayor sensitividad, ya que es más importante identificar aquellos vuelos que posiblemente se van a ver afectado que aquellos que cumplirán con el itinerario.

La curva ROC y el área bajo la curva (AUC) también permiten establecer la efectividad de los modelos. La curva ROC es de gran utilidad dado que compara la tasa positiva verdadera (TPR) frente a la tasa positiva falsa (FPR) y el AUC mide toda el área bidimensional por debajo de la curva ROC, proporciona una medición agregada del rendimiento en todos los umbrales de clasificación posibles. Un modelo que clasifica perfectamente las dos clases tiene un 100% de sensibilidad y especificidad, por lo que el área bajo la curva es 1 y un modelo que predice por debajo de lo esperado por azar, tiene AUC menor de 0.5.

Es importante recordar que los vuelos cumplidos son aquellos que tienen demoras o adelantos inferiores a quince (15) minutos con relación a la hora de salida programada por itinerario. Para la creación de los modelos se analizan vuelos con retrasos mayores a 15, 60 y 90 minutos. Por la importancia que tienen en la operación y las multas para las aerolíneas, se profundizará en el análisis y predicción de vuelos con retrasos superiores a 60 minutos.

6.4.4.1. Regresión logística

Una de las principales aplicaciones de la regresión logística es la de clasificación binaria, en donde las observaciones se clasifican en un grupo u otro dependiendo del valor que tomen las variables empleadas como predictoras. Es importante aclarar que, aunque la regresión logística permite clasificar, dicha clasificación se obtiene al modelar el logaritmo de la probabilidad de pertenecer a cada grupo. La asignación final se hace en función de las probabilidades predichas.

Para la interpretación de los Odd Ratio se tendrá en cuenta que estos van de cero (0) a infinito. Cuando es igual a uno (1) no hay asociación entre las variables, los valores menores a uno (1) indican una asociación inversa entre las variables y que los que sean mayores a uno (1) señalan una relación positiva entre las mismas. Cuando los valores sean menores que uno (1), la interpretación se realizará calculando la inversa del Odd Ratio.

Por medio de la regresión logística, se crearon y evaluaron tres modelos: 1. Sin balanceo, 2. Con sobre muestreo, 3. Step sin balanceo, para los vuelos afectados con retrasos de 15, 60 y 90 minutos.

El logaritmo de odds de variables como distancia, hora a la que está programado el vuelo (HoraMin) y cantidad de vuelos por hora (CntVuelosHora) están negativamente relacionados con la afectación de los vuelos. La mayor parte de estas corresponde a variables que tienen que ver con el origen del vuelo.

El resto de las variables del conjunto de datos tienen una relación positiva con los vuelos afectados.

Se presentan a continuación los modelos con mejor balance entre sensitividad y especificidad obtenidos en los siguientes puntos de corte:

AEROPUERTO	PUNTO DE CORTE	VUELOS NO AFECTADOS	VUELOS AFECTADOS
MDE	0.16	18118	852
PEI	0.19	9024	325
ADZ	0.18	5023	185

Tabla 22 *Modelos Regresión logística. Elaboración propia.*

En la siguiente tabla se presenta las métricas obtenidas con el modelo Logit con los datos sin balancear a 60 minutos. El accuracy superior a un 60% indicaría una predicción aceptable para la proporción entre los positivos reales predichos por el algoritmo y todos los casos positivos.

AEROPUERTO	PUNTO DE CORTE	ACCURACY	SENSITIVIDAD	ESPECIFICIDAD
MDE	0.16	*0.6685*	*0.6764*	*0.6669*
PEI	0.19	*0.62897*	*0.64879*	*0.6272*
ADZ	0.18	*0.65123*	*0.66879*	*0.65478*

Tabla 23 *Métricas obtenidas con Logit. Elaboración propia.*

6.4.4.2. Redes neuronales

Una red neuronal es un procesador distribuido en paralelo de forma masiva con una propensión natural a almacenar conocimiento experimental y convertirlo en disponible para su uso. Se asemeja al cerebro en dos aspectos:

- El conocimiento se adquiere por la red mediante un proceso de aprendizaje.
- Las fuerzas de conexión interneuronal, conocidas como ponderaciones sinápticas, se utilizan para almacenar el conocimiento.

Las redes neuronales son un método ideal en muchas aplicaciones de minería de datos predictiva por su potencia, flexibilidad y facilidad de uso. Para el presente análisis se utiliza la arquitectura feedforward, donde las conexiones de la

red fluyen unidimensionalmente desde la capa de entrada hasta la capa de salida sin ciclos de retroalimentación

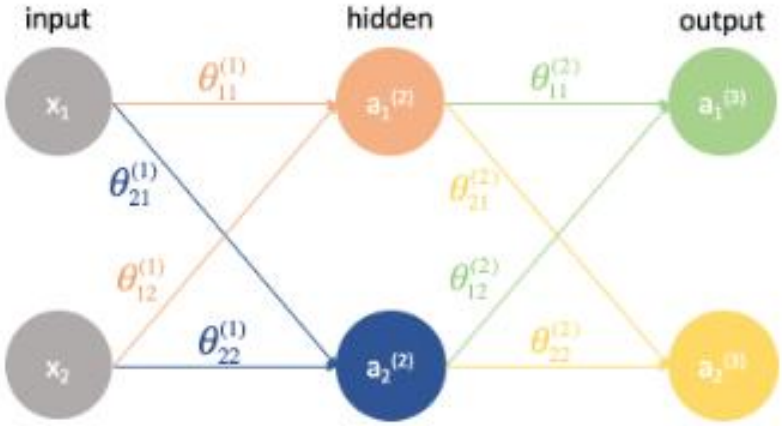

Ilustración 28 *Red Neuronal feedforward [33].*

- La capa de entrada, también denominada sensorial, está compuesta por neuronas que reciben datos o señales precedentes del entorno (predictores).
- La capa oculta contiene nodos no observables. El valor de cada unidad oculta es una función de los predictores.
- La capa de salida se compone de neuronas que proporcionan la respuesta de la red neuronal.

Con la técnica de redes neuronales feedforward se tiene como objetivo clasificar si un vuelo que despega o se dirige a los aeropuertos de interés se afecta (se cancela o se retrasa).

Para el diseño de redes neuronales se recomienda el uso de una sola capa oculta y en algunos casos específicos dos. El aumentar el número de capas ocultas y neuronas hace más lento el entrenamiento, puede aumentar drásticamente el número de mínimos locales y causar sobre entrenamiento.

Para definir el número de neuronas en la capa oculta se utiliza un método iterativo desde un perceptrón (1 neurona en la capa oculta) hasta 8 neuronas. La

selección del mejor modelo se realiza mediante las métricas de accuracy, sensitividad y especificidad.

Con cinco (5) neuronas en la capa oculta se logra el mejor resultado en accuracy, sensitividad y especificidad. A partir de 3 neuronas los resultados son comparables y las diferencias en las métricas evaluadas varían entre 1 y 2 puntos porcentuales.

Para la creación de los modelos finales se tendrá la siguiente arquitectura feedforward:

- Capa de entrada: 110 neuronas.
- Capa oculta: 5 neuronas.
- Capa de salida: 1 neurona.

Los parámetros adicionales para la creación del modelo son: decay = 0.5, rang = 0.1 y maxit = 2.000. El valor de los parámetros mencionados se ajustó con la iteración y creación de modelos con condición de retraso de 60 minutos.

Se presentan a continuación los modelos con mejor balance entre sensitividad y especificidad obtenidos en los siguientes puntos de corte:

AEROPUERTO	PUNTO DE CORTE	VUELOS NO AFECTADOS	VUELOS AFECTADOS
MDE	0.16	19525	1025
PEI	0.17	10924	452
ADZ	0.17	6439	235

Tabla 24 *Modelos Redes neuronales. Elaboración propia.*

En la siguiente tabla se presentan las métricas obtenidas del modelo neuronal con los datos sin balancear a 60 minutos.

AEROPUERTO	PUNTO DE CORTE	ACCURACY	SENSITIVIDAD	ESPECIFICIDAD
MDE	0.16	0.6840	0.6873	0.6846
PEI	0.17	0.6725	0.6725	0.6785
ADZ	0.17	0.6685	0.6764	0.6669

Tabla 25 *Métricas obtenidas con redes neuronales. Elaboración propia.*

6.4.4.3. XGboosting

Leo Brieman en sus trabajos de investigación orientados a mejorar métodos de clasificación, observó la ventaja de redefinir los pesos para las variables consideradas en los procesos de entrenamiento de modelos considerando iteraciones previas. Ese primer intento llevó a lo que él llamó “arcing” (o dar forma de arco) y fue tomado por Freund and Schapire generando la primera versión del algoritmo conocido hoy como Adaboost, el cual de manera iterativa redefine los pesos de una base de datos de entrenamiento considerando la historia de los errores de intentos previos, construye un nuevo clasificador y usa el error de clasificación dar un nuevo peso a las muestras clasificadas erróneamente. Como consecuencia, se obtiene un clasificador robusto a partir de la combinación de varios clasificadores débiles.

Posteriormente, Friedman por propia cuenta y Mason, Baxter, Bartlett and Frean en una iniciativa independiente desarrollaron investigaciones para proponer algoritmos basados en el análisis de los errores residuales (diferencia entre valor de la predicción y valor real) y que, de manera iterativa y similar a un algoritmo de optimización, ejecutaban procesos considerando un gradiente descendente, de tal manera que cada iteración llevaba a la reducción de los residuales a un mínimo. De ahí la denominación de “Gradient Boosting” a estas técnicas de clasificación basadas en la optimización de los residuales.

Más recientemente, en 2014 Tien propone una mejora a la metodología de Gradient Boosting considerando:

- La modificación de un árbol inicial apuntando a reducir el valor del error residual en cada iteración.
- El uso de un algoritmo “Greedy Aproximado” (un algoritmo “greedy” riguroso implicaría un gran consumo de poder computacional).
- La posibilidad de tener el proceso ejecutado por varios equipos al mismo tiempo (parallel learning).
- Weighted Quantile Sketch: generación de “histogramas aproximados” para ubicar las observaciones con bajo nivel de confianza en cuantiles específicos para ello, dejando el resto de las muestras en los cuantiles con la mayor cantidad de observaciones para hacer más eficiente la evaluación de los registros significativos.
- Sparsity-Aware Split Finding: técnicas para procesar adecuadamente “missing values” o datos faltantes en los registros en cuestión.
- Cache-Aware Access: XGBoost restringe los cálculos de los gradientes (primera derivada) y de los hessians (segunda derivada), conceptos clave para los cálculos generales, para la memoria caché del sistema que se esté utilizando.
- Blocks for Out-of-Core Computation: manejo particular de bases de datos de gran tamaño, permitiendo comprimirlas en el caso de tener que hacer uso de los recursos en disco duro. Esto partiendo de la base de que es mejor emplear un tiempo descomprimiendo la data de entrenamiento ubicada en el disco duro que esperar a la lectura de toda la base.

Para la implementación de este tipo de modelo se dispuso de un código que permitió revisar diferentes valores de la rata de aprendizaje y así poder lograr el mejor resultado.

La librería xgboost ofrece varias opciones de ajuste de manera iterativa. También permite observar según el resultado del modelo las variables de mayor importancia y que ofrecen una guía de los verdaderos factores de peso en este análisis.

De igual manera, como se procedió con los modelos de regresión logística y de red neuronal, se ejecutaron ejercicios de modelación para condiciones de afectación de 15, 60 y 90 minutos.

Se presentan a continuación los modelos con mejor balance entre sensitividad y especificidad obtenidos en los siguientes puntos de corte:

AEROPUERTO	PUNTO DE CORTE	VUELOS NO AFECTADOS	VUELOS AFECTADOS
MDE	0.22	19525	1025
PEI	0.21	10924	452
ADZ	0.21	6439	235

Tabla 26 *Modelos XGboosting. Elaboración propia.*

En la siguiente tabla se presentan las métricas obtenidas con el modelo XGBoosting que contiene los datos sin balancear a 60 minutos.

AEROPUERTO	PUNTO DE CORTE	ACCURACY	SENSITIVIDAD	ESPECIFICIDAD
MDE	0.22	0.6746	0.6935	0.6707
PEI	0.21	0.6323	0.6352	0.6658
ADZ	0.21	0.6601	0.6732	0.6714

Tabla 27 *Métricas obtenidas con XGboosting. Elaboración propia.*

6.4.5. Comparativo de los resultados obtenidos con los modelos utilizados.

Medida	*Regresión Logística*	*Redes Neuronales*	*XGBoosting*
Punto de corte	0.16	0.16	0.22
Accuracy	*0.6685*	0.6840	0.6746
Sensitividad	*0.6764*	0.6873	0.6935
Especificidad	*0.6669*	0.6846	0.6707
AUC	0.72	0.74	0.70

Tabla 28 *Resultados de los modelos en MDE. Elaboración propia.*

Medida	*Regresión Logística*	*Redes Neuronales*	*XGBoosting*
Punto de corte	0.19	0.17	0.21
Accuracy	*0.62897*	0.6725	0.6323
Sensitividad	*0.64879*	0.6725	0.6352
Especificidad	*0.6272*	0.6785	0.6658
AUC	0.71	0.75	0.71

Tabla 29 *Resultados de los modelos en PEI. Elaboración propia.*

Medida	*Regresión Logística*	*Redes Neuronales*	*XGBoosting*
Punto de corte	0.18	0.17	0.21
Accuracy	0.65123	0.6685	0.6601
Sensitividad	0.66879	0.6764	0.6732
Especificidad	0.65478	0.6669	0.6714
AUC	0.71	0.76	0.70

Tabla 30 *Resultados de los modelos en ADZ. Elaboración propia.*

Con relación a la validez de los modelos con la técnica de la simulación de Montecarlo, el entendimiento del fenómeno de retrasos que ha sido investigado desde el año 2018 por los grupos de investigación, ha demostrado que, en el fenómeno de los retrasos de los vuelos, sugerir una mejora es un desarrollo complicado, lo se sustenta con el enorme conjunto de variables definidos por la IATA y que en su mayoría son de carácter externo y por lo tanto muy difíciles de controlar.

En este mismo camino, la literatura que acompaña el fenómeno de retrasos, desde las aproximaciones de la analítica de datos, se ha enriquecido con modelos de predicción con una cantidad de técnicas diferentes y que en este trabajo se presenta el desarrollo de las tres (3) técnicas más utilizadas por otros autores y con los resultados obtenidos en la predicción de retrasos, se considera que quedan validados por sí mismo los modelos y permite escoger el que el interesado prefiera.

Desde la misma complejidad de cada uno de los modelos, y los resultados presentados, se considera que quedan validados en sus resultados, sobre la base de prueba y por lo tanto validarlos con la simulación de Montecarlo, va a convertirse en un proceso innecesario, ya que los resultados están validos por la métrica utilizada en la analítica de datos para este tipo de modelamiento.

6.5. SIMULACION DE MONTECARLO A LA PROBLEMÁTICA

Para el planteamiento de la simulación de Montecarlo, se toma la decisión de simular solamente las causas de origen interno que en términos generales sería las que se podrían disminuir con un proceso de revisión y mejora, ya que las externas son consideradas como incontrolables.

Partiendo de lo anterior, se presenta a continuación los porcentajes que se pretende utilizar en la simulación de Montecarlo para cada Aeropuerto, y para cada escenario internacional y nacional:

AEROPUERTO	INTERNACIONAL	NACIONAL
MDE	23,44%	18,52%
PEI	16,66%	12,25%
ADZ	42,86%	17,33%

Tabla 31 *Porcentaje de causas internas utilizadas en la simulación de Montecarlo. Elaboración propia.*

La simulación inicialmente se formuló de la siguiente manera para cada aeropuerto para cada escenario nacional e internacional:

DISTRIBUCION DE LA DEMNADA			DISTRIBUCION PROBABILIDAD ACUMULADA DEMANDA		
FALLA	P(X)		FALLA	LI	LS
41	0,0734		41	0	0,0734
65	0,0367		65	0,0734	0,1101
93	0,0311		93	0,1101	0,1412
42	0,0254		42	0,1412	0,1666
4	0,0226		4	0,1666	0,1892
67	0,0226		67	0,1892	0,2118
63	0,0226		63	0,2118	0,2344

Ilustración 29 *Formulación inicial. Elaboración propia.*

Para cada escenario se programaron las fallas de casa interna que se detectaron en la caracterización y su porcentaje determinado en la base de datos de la Aeronáutica Civil para el año 2017. En el cuadro siguiente se llevó el acumulado porcentual de las fallas.

Posterior a esto, se desarrolló un escenario de simulación con los siguientes componentes:

# FALLA	ALEATORIIO	FALLA	DURACION DEL RETRASO	DURACION ACUMULADA

Ilustración 30 *Componentes del escenario de simulación. Elaboración propia.*

Se decidió simular un total de trescientas sesenta y cinco fallas en cada aeropuerto por cada escenario internacional y nacional que se definieron en la primera columna # FALLA.

En el espacio ALEATORIO se utilizó la función aleatoria de Excel para generar la simulación aleatoria de Montecarlo.

Con el resultado del aleatorio, el porcentaje obtenido se buscaba en la tabla de probabilidad acumulada el valor y de acuerdo con el ordenamiento generaba la falla de origen interno programada.

La duración del retraso se programó con la función de Excel ALEATORIO.ENTRE(), con valores entre 15 y 60 minutos, considerados como los límites de especificación de la operación aeronáutica a nivel mundial.

Finalmente se acumuló la duración simulada de los retrasos para al final de cada aeropuerto en sus dos escenarios para generar el resultado final.

TOTAL DURACION	2.474.561	MINUTOS

Ilustración 31 *Total duración simulada. Elaboración propia.*

El desarrollo de estas simulaciones se puede apreciar en el anexo electrónico que acompaña este documento (anexo 3).

Una vez explorado este recurso se explica por qué el grupo de trabajo decidió no utilizarlo como herramienta válida para justificar los resultados:

- El presente proyecto de grado fue presentado y aprobado por el comité de grado de la Facultad de Ingeniería Industrial de la Universidad Santo Tomas en el año 2019, momento en que el Director del proyecto, Ingeniero Luis Manuel Pulido Moreno se encontraba en su primera aproximación al tema de retrasos en aeropuertos y en el grupo de trabajo, sin una guía adecuada, se decidió utilizar esta herramienta con el propósito de utilizar un concepto propio de la Ingeniería Industrial como lo es el de la simulación.

- Durante los años 2019 y el año 2021, el grupo de investigación de Ingeniería Civil, dirigido por el Ingeniero Oscar Díaz (dueño de los datos) y del cual es miembro el director del proyecto de grado, Ingeniero Luis Manuel Pulido Moreno, logro una compresión diferente del tratamiento de los retrasos en el despegue en la operación aeronáutica, que se ha visto reflejada en publicaciones y en la participación en ponencias internacionales.

- En este lapso de tiempo, entre los productos elaborados con los datos del grupo de investigación de Ingeniería Civil, se desarrollaron dos (2) tesis: Una tesis en pregrado en Ingeniería Industrial, 2020 elaborada por Diaz y Sandoval en la Universidad Santo Tomas y otra, elaborada por Pulido Et al en el 2020 en la Maestría de Analítica de la Universidad Javeriana en que los dos trabajos coinciden en que por la complejidad de la operación aeronáutica,

diversos autores han utilizado algoritmos de predicción, explorando en los dos trabajos algoritmos validados en la literatura como el Random Forest y la Regresión Logística entre otros y no logrando propuestas en cuanto a la mejora de la operación, dado que había que entre otras cosas, realizar procesos de mejora continua con cada aerolínea y con cada aeropuerto.

- Finalmente, luego de la primera e ingenua aproximación del problema propuesto con el proyecto de grado, no hay ninguna evidencia en la literatura encontrada, que el problema de los retardos en aeropuertos, sea tratado con simulación de Montecarlo.

7. CONCLUSIONES

Se puede afirmar que la aerolínea con mayores problemas en las operaciones aéreas es Avianca debido a la cantidad de vuelos tan grandes que tienen en el país y cuanto a los tiempos de retrasos en los vuelos en los tres aeropuertos seleccionados para la investigación (MED, PEI, ADZ), siempre estarán sobre la hora o muy por encima.

Las causas de las demoras que se deben abordar y solucionar primordialmente en las operaciones aéreas de los aeropuertos estudiados para factores externos (factores que no se pueden controlar) son 81: *debido a limitación de capacidad del sistema de atc en ruta o a alta demanda,* y 73: *aeropuerto alterno de destino o en ruta.* Mientras que para factores internos es la causa 41: *defectos del avión,* en donde se puede llegar a mejorar con mayor facilidad debido a que esto se puede controlar.

En el presente trabajo se reconocieron factores que influyen en la afectación (retrasos) de vuelos que despegaron de los aeropuertos de Medellín, Pereira y San Andrés en el año 2017 del vuelo, condiciones operativas, de causalidad según el código internacional de la IATA y de temporalidad. Cada uno de los aeropuertos presenta en el contenido de este documento su detalle.

Es importante comentar, que es posible que se presente un fenómeno de sesgo en los resultados, toda vez que no hay una seguridad de la captura o consignación de la información, que obedece única y exclusivamente a funcionarios de la Aero Civil en cada aeropuerto nacional y cuyos errores se hay detectado en este trabajo, como otros elaborados anteriormente entre otros por el trabajo de grado de Santiago Sandoval en la Universidad Santo Tomas, los trabajos de investigación de Diaz y Pulido desde el año 2018 y la Maestría de Analítica en la Javeriana.

Los modelos implementados mejoran su desempeño a medida que se aumenta el horizonte temporal para clasificar los vuelos retrasados, entre 15, 60 y 90 minutos. La diferencia de los resultados obtenidos entre los cortes de 60 y 90 minutos es marginal y teniendo en cuenta la cantidad de vuelos con retrasos superiores a 60 minutos, se estableció este último como el mejor escenario de estimación.

La red neuronal es la técnica que presentó mejor desempeño para los tres aeropuertos estudiados en este trabajo.

Los resultados de los modelos de predicción utilizados se evalúan por sí mismos en la métrica de la analítica de datos y por lo tanto no se consideró adecuado utilizar la simulación de Montecarlo para validar dichos resultados.

8. REFERENCIAS

[1] James J.H. Liou, L. Y.-H. (16 de Septiembre de 2010). Using decision rules to achieve mass customization of airline services. *Sciencie Direct*, 680-686. doi: https://doi.org/10.1016/j.ejor.2009.11.019

[2] V. Bogicevic, W. Y. (28 de Octubre de 2013). Airport service quality drivers of passenger satisfaction. *Emerald*, 3-18. Recuperado el 15 de Octubre de 2019, de https://www.emerald.com/insight/content/doi/10.1108/TR-09-2013-0047/full/html

[3] SITA. (4 de Junio de 2015). *Sita*. Recuperado el 15 de Otubre de 2019, de sita.aero: https://www.sita.aero/resources/type/white-papers/intelligent-airport-make-it-a-realit

[4] A H sin Lin M, Zhang Y. Marzo de 2017. *Hub-airport congestion pricing and capacity investment.* Recuperado de doi: 10.1016/j.trb.2017.03.009

[5] IATA Asociación Internacional del Transporte Aéreo. (2019). *Informe 2018.* Montreal: IATA. Recuperado el 25 de Agosto de 2019, de https://www.iata.org/pressroom/pr/Documents/2019-07-31-01-sp.pdf

[6] AEROCIVIL La Unidad Administrativa Especial de Aeronáutica Civil. (2017). *La aviacion en cifras.* Bogotá D.C.: Aerocivil. Recuperado el 29 de 08 de 2019, de http://www.aerocivil.gov.co/Potada/revi.pdf

[7] OACI *La Aviacion Unida, Organismo Especializado de las Naciones Unidas (2016).* Informe anual del Consejo de la OACI. Recuperado de: https://www.icao.int/annual-report-2016/Pages/ES/default_ES.aspx

[8] CEPAL. (2017). Transporte aéreo como motor del desarrollo sostenible en América latina y el caribe: retos y propuestas de política. Obtenido de Transporte aéreo como motor del desarrollo sostenible en América latina y el caribe: retos y propuestas de política. Recuperado de: a https://repositorio.cepal.org/bitstream/handle/11362/43411/1/S1800006_es.pdf

[9] Forbes Staff. (20 de diciembre de 2019). Congestión en aeropuerto El Dorado por llegada simultánea de vuelos. Recuperado de: https://forbes.co/2019/12/20/actualidad/congestion-en-aeropuerto-el-dorado-por-llegada-simultanea-de-vuelo

[10] González Muñoz, Luis Rafael. (2020). El transporte aéreo de pasajeros en Colombia: Aspectos para la competitividad. Trabajo de Grado. Universidad del Rosario. Recuperado de:

https://repository.urosario.edu.co/bitstream/handle/10336/21999/GonzalezMunoz-LuisRafael-2020.pdf;jsessionid=B6874DF3A1DE847FE152CEACBFF39370?sequence=4

[11] Dinero. (abril de 2019). Fallas en el Aeropuerto Internacional el Dorado. Revista Dinero. Recuperado de: https://www.dinero.com/edicion-impresa/pais/articulo/fallas-en-elaeropuerto-internacional-el-dorado-de-bogota/269240

[12] Caracol Radio. (17 de Julio de 2019). Diez horas duró cierre de pista en El Dorado que atrasó varios vuelos. Caracol. Recuperado de: https://caracol.com.co/radio/2019/07/17/nacional/1563327569_982918.html

[13] El Dorado. (2019). *ESTADÍSTICAS DE PASAJEROS Y OPERACIONES AÉREAS.* Bogotá D.C.: El Dorado. Recuperado el 10 de Septiembre de 2019, de https://eldorado.aero/wp-content/uploads/2019/10/Consolidado-3-Trimestre-de-2019.pdf

[14] Vergara. (2019). ¿Porque el aeropuerto el dorado está colapsado? Revista Semana. Recuperado de: https://www.semana.com/economia/articulo/por-que-el-aeropuerto-eldorado-esta-colapsado/632691

[15] El Espectador. (diciembre de 2019). Denuncian demoras y cancelaciones en el aeropuerto El Dorado. Periódico El Espectador. Recuperado de: https://www.elespectador.com/noticias/bogota/denunciandemoras-y-cancelaciones-en-el-aeropuerto-el-dorado-articulo-896779

[16] Han, Jiawei, Kamber, Micheline Morgan Kaufmann. Agosto 2000. Data Mining. Concepts and Techniques. Tercera Edicion. Morgan-Kaufmann Recuperado de: http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf

[17] Olmos Pineda, B. G. (2007). Minería de Datos. *Sciencie Direct*, 1, 2. Recuperado el 22 de Septiembre de 2019, de http://es.scribd.com/doc/93421745/Caso-de-Exito-Mineria-de-Datos#scribd

[18] Thuraisingham, B. (1993). *Web Data Mining and Applications in Business Intelligence and Counter-Terrorism.* Washintong D.C.: Taylot & Francis Group. Recuperado el 22 de Septiembre de 2019, de https://books.google.com.co/books?id=9tNgFBe-

TYYC&pg=PA492&lpg=PA492&dq=Thuraisingham+1993&source=bl&ots=o5nEvOGYeh&sig=ACfU3

[19] U, Fallad. M. (1996). *Advances in Knowledge discovery and Data Mining.* Massachusetts: MIT Press.

[20] M. Berry, G. L. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (Tercera ed.). Washington D.C.: Wiley. Recuperado el 22 de Septiembre de 2019, de http://axon.cs.byu.edu/~martinez/classes/478/readings/DataPrep.pdf

[21] R., P. P. (1998). *Data Mining in Marketing part I.* Washington: MIT press.

[22] Marqués, M. P. (2014). *Minería de datos a través de ejemplos.* Madrid: RC Libros. Recuperado el 2 de Octubre de 2019, de http://www.rclibros.es/pdf/capitulo_mineria.pdf

[23] Rouse, M. (4 de Julio de 2019). *WhatIs.com.* Obtenido de TechTarget: https://www.techtarget.com/contributor/Margaret-Rouse

[24] Ladrero, I. (3 de Junio de 2017). Busines Analytic Baoss. *Sciencie Direct*, https://www.baoss.es/que-es-business-analytics/. Obtenido de Baoss.

[25] Ma, N. L., Choy, M., & Cheong, M. (27 de Julio de 2012). Uncovering interesting business insights through the use of data analytics in Airport. *IEEE Xplore*, 805-809. doi:10.1109 / SRII.2012.91

[26] Roberto Henriques, I. F. (2018). Predictive Modelling: Flight Delays and Associated Factors,Hartsfield–Jackson Atlanta International Airport. *Sciencie Direct*, 639-644. Recuperado el 16 de Octubre de 2019, de https://www.sciencedirect.com/science/article/pii/S1877050918317319

[27] Z. Nazeri, J. Z. (2002). Mining Aviation Data to Understand Impacts of Severe Weather. *IEEE Xplore*, 5-6. doi:10.1109/ITCC.2002.1000441

[28] Adrián, A. S. (2015). Flight delay forecast due to weather using Data Mining. *Repositorio Universidad del Pais Vasco*, 27-36. Recuperado el 16 de Octubre de 2019, de http://hdl.handle.net/10810/15889

[29] Raj Bandyopadhyay, R. G. (2012). Predicting airline delays. *Repositorio de Stanford*, 2-5. Recuperado el 17 de Octubre de 2019, de http://cs229.stanford.edu/proj2012/BandyopadhyayGuerrero-PredictingFlightDelays.pdf

[30] Robert, S. P. (Mayo de 2010). Understanding and Minimizing Flight Delay . *Scholar Woeks College of William and Mary*, 10-40. Recuperado el 17 de Octubre de 2019, de

https://scholarworks.wm.edu/cgi/viewcontent.cgi?article=1711&context=honorstheses

[31] Megan Baluch, T. B.-H. (11 de Enero de 2017). Complex Analysis of United States Flight Data. *IEEE Xplore*, 2-5. doi:10.1109 / CCWC.2017.7868414

[32] Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). Metodología de la Investigación. Sexta edición. Ciudad de México. Recuperado de: http://observatorio.epacartagena.gov.co/wp-content/uploads/2017/08/metodologia-de-la-investigacion-sexta-edicion.compressed.pdf

[33] Pulido, Arias, Chavarro y Ramírez. (2019). MODELO PARA IDENTIFICAR LOS VUELOS AFECTADOS POR RETRASOS O CANCELACIONES EN EL AEROPUERTO EL DORADO DE BOGOTÁ, COLOMBIA. Proyecto de grado de Maestría en Analítica de Datos para la Inteligencia de Negocios. Universidad Javeriana. Uri: http://hdl.handle.net/10554/51109.

9. ANEXOS

9.1. Anexo 1. CÓDIGOS DE DEMORA DE IATA.

9.2. Anexo 2. BASES, TABLAS Y GRAFICAS.

9.3. Anexo 3. SIMULACION DE MONTECARLO.

Para acceder a los anexos de esta investigación dirigirse al siguiente link: https://drive.google.com/drive/folders/1wyAn5F6w0vM7dpSdo4WzdwJgoBIEAgTW?usp=sharing