

TRABAJO FIN DE MÁSTER

MUMADE 2024-2025

ANÁLISIS SOBRE EL ABSENTISMO LABORAL EN UNA EMPRESA LÓGISTICA



Máster Universitario en modelización
y análisis de datos económicos

 Universidad de
Castilla-La Mancha

Elena del Castillo Camacho



Trabajo Fin de Máster

máster universitario en modelización
y análisis de datos económicos

Análisis sobre el absentismo laboral en una empresa logística

Analysis of Work Absenteeism in a Logistics Company

Nº de palabras (excluidos bibliografía y anexos): 10474

2024-2025

Autor(a): Elena del Castillo Camacho

Correo electrónico: Elena.Castillo3@alu.uclm.es

Tutor: Juan Ramón Trapero Arenas

Fecha: 30 junio 2025

Firma autor(a)

Firma tutor



Facultad de Derecho y Ciencias Sociales

Ronda de Toledo, s/n • 13071-Ciudad Real (España) • Tlf. +34 926 295 300

Resumen

Este trabajo analiza el absentismo laboral por incapacidad temporal (IT) en el sector logístico español, un fenómeno con elevado impacto económico y organizacional. Combinando análisis descriptivos y modelos predictivos de series temporales aplicados a datos diarios reales, se identifican patrones clave como la concentración del absentismo en días laborables (76,5 % del total), con picos significativos durante los fines de semana (especialmente domingos (14,20 %) y sábados (14,01 %)), en lugar de los lunes como se señala habitualmente. Entre los días laborables, el viernes presenta la mayor tasa de IT (13,65 %). Además, se observa una relación positiva pero no estrictamente lineal entre horas planificadas e IT, lo que sugiere la influencia de factores adicionales. El estudio revela que el modelo ETS(A,N,A) ofrece las mejores predicciones (RMSE = 39,42), superando a técnicas más complejas como redes neuronales, lo que permite anticipar necesidades de personal con hasta 30 días de antelación. Los resultados destacan la influencia de factores organizacionales como la distribución de la carga laboral y la gestión de festivos, y subrayan una brecha de eficiencia (86,67 %) respecto a estándares sectoriales. Se propone reorientar las estrategias preventivas hacia el final de la semana laboral e implementar sistemas de alerta temprana, transformando el absentismo en un indicador estratégico de salud organizacional.

Palabras clave: Absentismo laboral, incapacidad temporal, sector logístico, modelos predictivos.

Abstract

This study analyzes work absenteeism due to temporary disability (IT) in the Spanish logistics sector, a phenomenon with significant economic and organizational impact. By combining descriptive analyses and time series predictive models applied to real daily data, key patterns are identified, such as the concentration of absenteeism on working days (76.5% of the total), with significant peaks during weekends (especially Sundays (14.20%) and Saturdays (14.01%)), rather than Mondays as is commonly reported. Among working days, Friday shows the highest IT rate (13.65%). Additionally, a positive but not strictly linear relationship between planned hours and IT is observed, suggesting the influence of additional factors. The study reveals that the ETS(A,N,A) model provides the best predictions ($RMSE = 39.42$), outperforming more complex techniques like neural networks, allowing for personnel needs to be anticipated up to 30 days in advance. The results highlight the influence of organizational factors such as workload distribution and holiday management, and underscore an efficiency gap (86.67%) compared to sector standards. It is proposed to redirect preventive strategies towards the end of the workweek and implement early warning systems, transforming absenteeism into a strategic indicator of organizational health.

Keywords: Work absenteeism, temporary disability, logistic sector, predictive model

Agradecimientos

Con el fin de esta etapa de mi vida, comienza otra llena de retos, pero también de posibilidades. Retos que tras haber participado en convenios de prácticas en empresas como Banco Cetelem, S.A.U y Ciudad Real International Airport S.L afronto con mayor seguridad.

En primer lugar, quiero dar las gracias a mi tutor de TFM Juan Ramon, quien ha sabido guiar me de la mejor de las formas y actitudes durante este proceso.

También agradecer a mis compañeros del Máster, que han estado siempre ahí para ayudar, a los que guardo con especial cariño y que siempre recordaré.

A mis amigos, los cuales siempre estuvieron ahí para apoyarme durante este proceso. Su paciencia, comprensión y ánimos fueron esenciales para superar los momentos de estrés y las dudas. Gracias por hacerme reír cuando más lo necesitaba y por ser una fuente constante de motivación. Este logro también es vuestro.

Y, finalmente, a mi familia, mi mayor fortuna en esta vida, quienes me han apoyado siempre y me han dado la libertad de ser y elegir aquello que me hiciera feliz en la vida; quienes me han enseñado la importancia de valorar las cosas simples y de ser buenas personas por encima de todo. Y quienes a pesar de haber tenido la posibilidad de acceder a la Universidad han dado todo porque sus hijos tuvieran la oportunidad de hacerlo.

Índice general

1. Introducción	15
2. Marco conceptual y Revisión de la literatura.....	18
2.1. Marco Conceptual.....	18
2.1.1. Absentismo laboral.....	18
2.1.2. Factores individuales	18
2.1.2.1. Salud física y mental: La base del bienestar laboral	18
2.1.2.2. Motivación y satisfacción laboral: Fundamentos psicológico	19
2.1.2.3. Variables demográficas y diferencias individuales.....	20
2.1.3. Factores organizacionales: El peso del contexto laboral	20
2.1.3.1. Cultura y clima organizacional.....	20
2.1.3.2. Liderazgo y estilo de gestión	21
2.1.3.3. Diseño del puesto y condiciones de trabajo.....	21
2.1.3.4. Políticas de recursos humanos y sistemas de recompensa.....	22
2.1.4. Factores socioambientales: El contexto extralaboral	22
2.1.4.1. Conflicto trabajo-familia y roles extralaborales	22
2.1.4.2. Cultura nacional y normas sociales.....	23
2.2. Revisión de la literatura.....	23
2.2.1. Absentismo laboral en el mundo y en España: Datos y tendencias actuales	25
2.3. Presentación caso de estudio: Absentismo en empresa logística	26
3. Datos y metodología	28
4. Resultados.....	33
4.1 Análisis exploratorio.....	33
4.2 Predicciones puntuales	48

4.3 Predicciones probabilísticas.....	49
5. Conclusión, futuras líneas de investigación y discusión	52
5.1 Conclusiones	52
5.2 Futuras líneas de investigación.....	54
5.3 Discusión.....	55
Anexos	66
Anexo I. Código RStudio	66
Anexo II. Imágenes Dashboard.....	80

Índice de Figuras

Figura 3.1. Dashboard Análisis Absentismo para empresa logística.....	32
Figura 4.1.1. Serie temporal del absentismo laboral IT (2024-2025)	36
Figura 4.1.2. Horas planificas vs IT (2025).....	37
Figura 4.1.3. Boxplot por tipo de día (periodo 2024-2025)	39
Figura 4.1.4. Total horas planificadas y total horas absentismo IT por día de la semana (2025).....	41
Figura 4.1.5. Distribución de densidad de las horas de absentismo IT por tipo día (2024-2025)	43
Figura 4.1.6. Relación de horas planificadas y de absentismo IT por empleado (2025).....	44
Figura 4.1.7. Serie temporal del absentismo laboral IT (2024-2025)	45
Figura 4.1.8. Serie temporal del absentismo laboral IT (2024-2025)	46
Figura 4.1.9. Serie temporal del absentismo laboral IT (2024-2025)	47
Figura 4.2.1. RMSE comparación por tipo de modelo	49
Figura 4.3.1. Predicciones de Absentismo (IT) con Intervalos de predicción ...	50
Figura 4.3.2. Predicciones de Absentismo (IT) con Intervalos de predicción para julio 2025.....	51
Figura Anexo 1. Vista Resumen general	80
Figura Anexo 2. Vista Resumen general 2	80
Figura Anexo 3. Vista Resumen general 3	81
Figura Anexo 4. Vista Predicciones	81
Figura Anexo 5. Vista Predicciones 2	82
Figura Anexo 6. Vista Comparativas bases de datos marzo, mayo y junio	82

Índice de Tablas

Tabla 3.1 Variables de la base de datos	28
Tabla 3.2 Diccionario de paquetes utilizados	32
Tabla 4.1.1. Estadísticos descriptivos globales sobre Absentismo IT	33
Tabla 4.1.2. Estadísticos descriptivos por tipo de día sobre Absentismo IT	35
Tabla 4.1.3. Absentismo por tipo de día.....	38
Tabla 4.1.4. Absentismo por tipo día (2025).....	40
Tabla 4.1.5. Tasa de absentismo IT por horas planificadas por día de la semana (2025).....	42
Tabla 4.1.6. Absentismo por tipo de día.....	47

1. Introducción

El absentismo laboral constituye uno de los principales desafíos en la gestión de recursos humanos, especialmente en sectores intensivos en mano de obra como el logístico. Este fenómeno, que se refiere a la ausencia no planificada de los trabajadores en su puesto de trabajo, trasciende el mero indicador operativo para convertirse en un síntoma de salud organizacional con profundas implicaciones económicas y sociales. En el contexto actual de transformación digital y creciente conciencia sobre el bienestar laboral, comprender sus patrones y causas se ha vuelto estratégico para las empresas que buscan equilibrar productividad con sostenibilidad del empleo. La literatura especializada ha demostrado que el absentismo no es un comportamiento homogéneo, sino el resultado de complejas interacciones entre factores individuales, organizacionales y contextuales, que varían significativamente según sectores, culturas laborales y modelos de gestión (Johns, 2022).

El sector logístico español representa un caso para estudio, al combinar características estructurales que lo hacen particularmente vulnerable: alta estacionalidad de la demanda, condiciones laborales físicamente exigentes, rotación elevada de personal y creciente presión por eficiencia operativa. Según datos de la Asociación Española de Logística (2023), este sector presenta tasas de absentismo que superan en un 30% la media nacional, con un coste estimado que representa entre el 5% y 8% de su masa salarial. Estas cifras adquieren mayor relevancia si consideramos el papel estratégico que la logística juega en la economía actual, donde interrupciones en la cadena de suministro pueden tener efectos multiplicadores en toda la actividad económica. Sin embargo, más allá del impacto cuantificable en productividad y costes, el absentismo elevado puede ser indicador de problemas más profundos en el clima laboral, la adecuación de las condiciones de trabajo o la efectividad de las políticas de recursos humanos.

Este trabajo se centra específicamente en el absentismo por incapacidad temporal (IT), que representa según el Ministerio de Trabajo (2023) más del 75%

de las ausencias en el sector, con especial incidencia de trastornos musculoesqueléticos (30% de los casos) y problemas de salud mental (15%). A diferencia de otras formas de absentismo, la IT implica consideraciones médicas y legales específicas, además de reflejar con particular claridad el impacto de las condiciones laborales en la salud de los trabajadores. El estudio de sus patrones temporales y su relación con variables operativas como las horas planificadas ofrece una valiosa oportunidad para desarrollar herramientas predictivas que permitan a las empresas anticipar necesidades de personal suplente y diseñar intervenciones preventivas. En un contexto empresarial cada vez más orientado a la toma de decisiones basada en datos, este tipo de aproximaciones cuantitativas representan un avance significativo respecto a los métodos tradicionales de gestión del absentismo, predominantemente reactivos y basados en experiencia intuitiva.

La investigación aquí presentada combina en el contexto logístico español análisis descriptivos avanzados con modelos predictivos de series temporales, aplicados a datos diarios reales de una empresa del sector. Este enfoque dual permite no solo caracterizar el fenómeno en sus múltiples dimensiones (distribución temporal, relación con carga laboral, variabilidad según tipo de día), sino también evaluar la viabilidad de pronosticar su evolución con suficiente antelación para fines de planificación operativa. Los resultados tienen implicaciones tanto teóricas como prácticas: en el plano académico, contribuyen a la literatura sobre gestión del absentismo en entornos operativos intensivos; en el ámbito profesional, proporcionan herramientas concretas para mejorar la gestión de recursos humanos en un sector clave para la economía. Además, el desarrollo de un *dashboard* interactivo como producto derivado de la investigación facilita la transferencia efectiva del conocimiento generado a los decisores empresariales, cerrando así el ciclo entre investigación académica y aplicación práctica.

El trabajo se estructura en cinco capítulos que abarcan desde el marco conceptual y revisión de literatura hasta el análisis de resultados y conclusiones, pasando por una detallada descripción metodológica. Esta estructura permite no solo presentar los hallazgos específicos del caso estudiado, sino también situarlos en el contexto más amplio de la investigación sobre absentismo laboral,

destacando tanto las convergencias como las particularidades encontradas. El enfoque adoptado busca equilibrar rigor analítico con relevancia práctica, reconociendo que el valor último de esta investigación reside en su capacidad para informar y mejorar las decisiones de gestión en un ámbito tan crítico como es la relación entre salud laboral y productividad empresarial.

2. Marco conceptual y Revisión de la literatura

2.1. Marco Conceptual

2.1.1. Absentismo laboral

El absentismo laboral se refiere a la ausencia de un trabajador en su puesto de trabajo, durante la jornada laboral establecida, ya sea por motivos justificados o injustificados (Johns, 2010). Según Miraglia y Johns (2016), representa un comportamiento de retirada del trabajo que puede manifestarse a través de faltas completas o llegadas tardías/retiros tempranos sin aprobación. La Organización Internacional del Trabajo (OIT, 2019) lo conceptualiza como la no asistencia al trabajo cuando se está contratado para hacerlo, destacando que puede ser voluntario (por decisión del trabajador) o involuntario (por enfermedad u otras causas ajenas a su voluntad). Desde una perspectiva organizacional, el absentismo se considera un indicador crítico de problemas en el clima laboral y la gestión de recursos humanos (Stewart et al., 2003), con importantes implicaciones económicas y productivas para las empresas (Hemp, 2004).

El absentismo laboral representa un fenómeno complejo que trasciende la simple ausencia física del trabajador, constituyendo más bien un indicador sintomático de disfunciones organizacionales, psicosociales e individuales interrelacionadas. Desde una perspectiva sistémica, el estudio del absentismo requiere integrar diversos marcos teóricos. Steers y Rhodes (1978) en su modelo de asistencia laboral postularon que la decisión de asistir al trabajo resulta de un complejo equilibrio entre motivaciones individuales y presiones situacionales. Este análisis exhaustivo busca desentrañar las causas fundamentales del absentismo estructurándolas en tres niveles analíticos interconectados: factores individuales, organizacionales y socioambientales.

2.1.2. Factores individuales

2.1.2.1. Salud física y mental: La base del bienestar laboral

La relación bidireccional entre salud y asistencia laboral ha sido ampliamente documentada en la literatura científica. El modelo biopsicosocial propuesto por Engel (1977) ayuda a entender cómo las ausencias laborales emergen de la

interacción dinámica entre factores biológicos (enfermedades crónicas, lesiones ocupacionales), psicológicos (estrés crónico, trastornos del estado de ánimo) y sociales (redes de apoyo, estigma organizacional). Investigaciones como las de Kivimäki et al. (2003) han demostrado que condiciones como la hipertensión no controlada o la diabetes generan patrones de absentismo recurrente cuando no se implementan ajustes razonables en el puesto de trabajo. En el ámbito de la salud mental, la Organización Mundial de la Salud (2022) ha identificado a la depresión y los trastornos de ansiedad como las principales causas de pérdida de días laborales en economías desarrolladas, fenómeno que la teoría de conservación de recursos de Hobfoll (1989) explica mediante el concepto de agotamiento emocional. Paralelamente, Johns (2010) presenta la idea de cómo la asistencia al trabajo a pesar de problemas de salud significativos actúa como precursor de ausencias prolongadas, creando un círculo vicioso donde la productividad disminuye mientras aumentan los riesgos de incapacidad temporal.

2.1.2.2. Motivación y satisfacción laboral: Fundamentos psicológico

La insatisfacción laboral emerge como uno de los predictores más consistentes del absentismo en la literatura organizacional. La teoría de los dos factores desarrollada por Herzberg (1968) establece una distinción fundamental entre los motivadores intrínsecos (logro, reconocimiento, naturaleza del trabajo) y los factores higiénicos (condiciones laborales, políticas organizacionales), demostrando que la carencia de estos últimos genera activamente insatisfacción y conductas de retiro. Complementariamente, la teoría de la equidad de Adams (1965) explica cómo las percepciones de injusticia en la relación esfuerzo-recompensa llevan a los trabajadores a adoptar estrategias compensatorias que incluyen el absentismo como mecanismo de restauración del equilibrio psicológico. Desde la perspectiva de la teoría de la autodeterminación (Deci & Ryan, 2000), cuando el entorno laboral frustra sistemáticamente las necesidades psicológicas básicas de autonomía, competencia y vinculación social, se produce un deterioro progresivo de la motivación intrínseca que se manifiesta en mayor propensión al absentismo.

2.1.2.3. Variables demográficas y diferencias individuales

La investigación sobre diferencias individuales en el absentismo revela patrones consistentes asociados a variables demográficas y de personalidad. Nicholson et al. (2006) identificaron que los trabajadores más jóvenes muestran mayor propensión al absentismo voluntario (ausencias por evitación o desmotivación), mientras que los trabajadores de mayor edad tienden a ausentarse principalmente por razones de salud. Las diferencias de género también emergen como factor relevante, con estudios como los de Glass y Estes (1997) demostrando que las mujeres experimentan mayor conflicto trabajo-familia debido a la persistente desigualdad en la distribución de roles domésticos. En el ámbito de la personalidad, investigaciones basadas en el modelo de los cinco factores (Judge et al., 1997) han encontrado que el neuroticismo (caracterizado por inestabilidad emocional y baja tolerancia al estrés) correlaciona positivamente con conductas de absentismo, mientras que rasgos como la responsabilidad muestran asociaciones negativas. Estos hallazgos subrayan la importancia de considerar las diferencias individuales al diseñar intervenciones contra el absentismo.

2.1.3. Factores organizacionales: El peso del contexto laboral

2.1.3.1. Cultura y clima organizacional

La cultura organizacional, definida por Schein (2010) como el patrón de supuestos básicos compartidos que un grupo ha aprendido al resolver sus problemas de adaptación externa e integración interna, ejerce una influencia determinante en las normas sobre asistencia y compromiso laboral. Investigaciones en diversos contextos industriales han revelado que organizaciones con culturas rígidas y excesivamente burocráticas generan paradójicamente mayores tasas de absentismo "encubierto", donde los trabajadores desarrollan sofisticados sistemas para justificar ausencias sin enfrentar consecuencias formales. En el extremo, las culturas organizacionales tóxicas caracterizadas por prácticas de acoso laboral o falta de ética institucional producen lo que Einarsen et al. (2003) denominaron "*turnover psicológico*" (un estado de desconexión emocional que precede al abandono físico). Por contraste, culturas organizacionales flexibles que priorizan el bienestar integral del empleado, logran reducir significativamente el absentismo mediante políticas

de conciliación auténticamente implementadas y sistemas de apoyo psicosocial accesibles (Kossek & Michel 2011).

2.1.3.2. Liderazgo y estilo de gestión

El impacto del liderazgo en las conductas de asistencia ha sido ampliamente documentado. El modelo de liderazgo desarrollado por Bass (1985) identifica cuatro componentes clave (influencia idealizada, motivación inspiracional, estimulación intelectual y consideración individualizada) que correlacionan negativamente con tasas de absentismo. Estilos de liderazgo autoritarios basados en el control coercitivo o, alternativamente, estilos caracterizados por la ausencia de dirección clara, generan ambientes laborales estresantes que incrementan el absentismo. La teoría del intercambio líder-miembro (Graen & Uhl-Bien, 1995) añade otra capa de complejidad al mostrar cómo las relaciones diferenciadas que los supervisores establecen con sus subordinados pueden crear percepciones de injusticia que alimentan conductas de absentismo entre los empleados "del grupo externo".

2.1.3.3. Diseño del puesto y condiciones de trabajo

El modelo *Job Demands-Resources* (JD-R) propuesto por Bakker y Demerouti (2007) sirve para entender cómo las características objetivas del puesto de trabajo influyen en las conductas de asistencia. Según este modelo, cuando las demandas laborales (carga cognitiva, presión temporal, exigencia emocional) exceden sistemáticamente los recursos disponibles (autonomía, apoyo social, herramientas adecuadas), los trabajadores experimentan estrés crónico que deriva en estrategias de afrontamiento contraproducentes como el absentismo. Complementariamente, el modelo de control-demanda de Karasek (1979) enfatiza la importancia del equilibrio entre exigencias laborales y grado de control sobre el trabajo, mostrando que los puestos caracterizados por alta demanda y bajo control (típicos en trabajos rutinarios de línea de montaje) generan mayor alienación laboral y tasas elevadas de ausentismo. Investigaciones recientes en ergonomía organizacional (Crawford et al., 2010) han ampliado este marco incorporando la dimensión de significado del trabajo, demostrando que incluso puestos exigentes pueden mantener bajos niveles de absentismo cuando los trabajadores perciben un claro propósito y contribución social en sus tareas.

2.1.3.4. Políticas de recursos humanos y sistemas de recompensa

Los sistemas de gestión de recursos humanos ejercen una influencia determinante en las conductas de asistencia a través de múltiples mecanismos. Investigaciones basadas en la teoría de la expectativa (Vroom, 1964) han demostrado que sistemas de recompensa mal diseñados (donde no existe una clara vinculación entre desempeño y reconocimiento) generan percepciones de injusticia que alimentan el absentismo. Por otro lado, estudios comparativos internacionales (Pichler & Wallace, 2009) revelan que las organizaciones que invierten en desarrollo profesional continuo y planes de carrera transparentes logran reducir significativamente las tasas de absentismo voluntario. Un área particularmente sensible son los procedimientos disciplinarios: investigaciones en psicología organizacional (Arvey & Jones, 1985) muestran que sistemas punitivos excesivamente rígidos o inconsistentes en su aplicación pueden generar resentimiento y conductas de retiro, mientras que enfoques restaurativos centrados en la solución de problemas subyacentes resultan más efectivos para mantener la asistencia regular.

2.1.4. Factores socioambientales: El contexto extralaboral

2.1.4.1. Conflicto trabajo-familia y roles extralaborales

El conflicto trabajo-familia, conceptualizado como una forma de conflicto interrol en el que las demandas de los roles laborales y familiares son mutuamente incompatibles en algún aspecto, representa una de las principales causas contextuales de absentismo en la sociedad contemporánea (Greenhaus & Beutell, 1985). Investigaciones como las de Allen et al. (2000) han documentado cómo este conflicto afecta desproporcionadamente a ciertos grupos demográficos: trabajadores con hijos pequeños (especialmente madres solteras), cuidadores informales de adultos mayores o personas con discapacidad, y trabajadores que simultanean empleo con estudios superiores. El modelo de fronteras entre trabajo y familia (Clark, 2000) explica cómo la permeabilidad entre estos dominios (acentuada por las tecnologías digitales) puede exacerbar el conflicto cuando no existen políticas organizacionales que protejan el tiempo de recuperación personal. Estudios comparativos en países con diferentes regímenes de bienestar (Lyness et al., 2012) muestran que las organizaciones que implementan políticas de conciliación auténticas (horarios

flexibles, teletrabajo opcional, guarderías en el lugar de trabajo) logran reducir significativamente el absentismo relacionado con responsabilidades familiares.

2.1.4.2. Cultura nacional y normas sociales

La dimensión cultural nacional modula la expresión y aceptabilidad social del absentismo a través de valores compartidos y normas implícitas. Investigaciones basadas en las dimensiones culturales de Hofstede (2001) muestran que sociedades con alto colectivismo (donde priman los objetivos grupales sobre los individuales) tienden a ejercer mayor presión social contra el absentismo, mientras que culturas individualistas muestran mayor variabilidad en las actitudes hacia la asistencia laboral. La dimensión de distancia al poder también resulta relevante: en sociedades con alta aceptación de las jerarquías (Malasia, México), el absentismo puede convertirse en una forma encubierta de protesta contra estructuras organizacionales percibidas como injustas. Estudios transculturales (Gelfand et al., 2007) han identificado diferencias significativas en la tolerancia social hacia distintos tipos de ausencias, con países nórdicos mostrando mayor aceptación del absentismo por razones familiares, mientras que culturas asiáticas con fuerte ética confuciana mantienen normas más estrictas sobre asistencia laboral.

2.2. Revisión de la literatura

El absentismo laboral constituye un fenómeno organizacional complejo cuya causa multifactorial requiere un análisis desde diversas dimensiones analíticas. La literatura especializada identifica:

En el ámbito de la salud física, las investigaciones documentan que las enfermedades musculoesqueléticas (especialmente lumbalgias y trastornos por movimientos repetitivos) representan el 28% de los casos de absentismo en entornos industriales (van Vilsteren et al., 2015). Las condiciones crónicas como diabetes mellitus (con una prevalencia del 12% en trabajadores mayores de 45 años) generan absentismo recurrente debido a complicaciones agudas y controles médicos (Stewart et al., 2003). Los trastornos cardiovasculares muestran una correlación significativa con períodos de ausencia prolongados.

En salud mental, los estudios meta-analíticos revelan que la depresión incrementa en 2,7 veces el riesgo de absentismo prolongado (>15 días/año),

mientras que los trastornos de ansiedad generan principalmente ausencias esporádicas (Miraglia & Johns, 2016). El síndrome de burnout, particularmente en profesionales de salud y educación, presenta una tasa de absentismo 3,2 veces mayor respecto a otros grupos ocupacionales (Kristensen, 1991).

Las características demográficas muestran patrones diferenciados: los trabajadores jóvenes (18-25 años) presentan mayor absentismo por motivos no médicos (ratio 1,8:1), mientras que los mayores de 50 años concentran el 62% de las ausencias por problemas de salud (Böckerman & Laukkanen, 2010). El género femenino muestra tasas un 23% superiores, atribuible principalmente a roles de cuidado familiar (Vänni et al., 2012).

El clima laboral disfuncional, caracterizado por liderazgo autoritario y falta de cohesión grupal, incrementa el absentismo en aproximadamente 4.2 días/año por empleado (Nielsen et al., 2016). El acoso laboral presenta una ratio de 3,1 para ausencias recurrentes, con mayor prevalencia en sectores sanitarios (31%) y educativos (28%).

Las condiciones de trabajo inadecuadas generan impactos diferenciados: las cargas laborales excesivas aumentan el riesgo de absentismo en un 40%, mientras que los turnos nocturnos rotativos elevan la tasa de ausencias en un 58% respecto a horarios convencionales (Kristensen, 1991). Los entornos físicos peligrosos (ej.: niveles de ruido >85 dB) se asocian con un incremento del 27% en ausencias por problemas auditivos (van Vilsteren et al., 2015).

Las políticas organizacionales deficientes muestran efectos significativos: los programas de bienestar insuficientes se correlacionan con un 35% más de absentismo (Hemp, 2004), mientras que las empresas con sistemas de compensación por enfermedad inferiores al 80% del salario presentan tasas de ausencia un 42% mayores (Gimeno et al., 2004).

La precariedad laboral muestra efectos contundentes: los trabajadores temporales tienen 2,3 veces más probabilidades de absentismo que los permanentes, particularmente en el sector servicios (D'Abate & Eddy, 2007). Las crisis económicas incrementan el absentismo por estrés financiero en un 28% (datos post-2008).

Los factores culturales presentan variaciones geográficas significativas: en países nórdicos el absentismo por depresión es 3,1 veces mayor que en culturas mediterráneas, mientras que el presentismo muestra el patrón inverso (Vänni et al., 2012). Las políticas públicas de salud laboral reducen el absentismo hasta un 19% cuando incluyen programas de prevención (van Vilsteren et al., 2015).

La digitalización laboral ha generado nuevas formas de absentismo: el "ausentismo digital" afecta al 14% de teletrabajadores, manifestado como desconexión prolongada durante jornadas laborales (Fiorini, 2024). Las empresas con sistemas de monitorización rígidos presentan un 23% más de ausencias digitales que aquellas con modelos flexibles.

2.2.1. Absentismo laboral en el mundo y en España: Datos y tendencias actuales

El absentismo laboral presenta variaciones significativas a nivel global, influenciado por factores culturales, económicos y normativos. Según datos de la Organización Internacional del Trabajo (OIT, 2019), los países con sistemas de protección social más desarrollados, como los nórdicos, registran tasas más elevadas (entre el 5% y 7% del tiempo laboral anual), mientras que en economías emergentes con alto empleo informal las cifras son notablemente inferiores debido a la precariedad laboral y el miedo al despido. Estados Unidos muestra una tasa promedio del 3%, con marcadas diferencias sectoriales, siendo la sanidad y educación los ámbitos más afectados. En contraste, países asiáticos como Japón y Corea del Sur presentan índices mínimos (1-2%), aunque esta aparente ventaja oculta graves problemas de presentismo y casos de karoshi (muerte por exceso de trabajo). América Latina, por su parte, evidencia una paradoja: mientras el absentismo formal es bajo (2-3%), el presentismo alcanza niveles alarmantes debido a las condiciones laborales precarias (OIT, 2021).

En el contexto europeo, España destaca por sus elevadas tasas de absentismo, que en 2022 alcanzaron el 5.8%, superando la media continental (4.5%) según datos de Adecco (2023). Este fenómeno se explica principalmente por las bajas médicas, que representan el 75% de las ausencias, con especial incidencia de enfermedades comunes (40%), trastornos musculoesqueléticos (30%) y problemas de salud mental (15%), según el Ministerio de Trabajo (2023). Los sectores más afectados son la sanidad (8.2%), donde predominan el estrés y el

burnout; la industria (6.5%), con alta siniestralidad laboral; y los servicios (5.1%), donde la insatisfacción laboral juega un papel determinante (INE, 2023). Investigaciones académicas como las de García-Serrano y Malo (2020) destacan cómo el envejecimiento de la población activa española (con una edad media de 44 años) y la precarización laboral han contribuido a agravar este problema en la última década.

Entre las tendencias actuales más relevantes destaca el crecimiento exponencial del absentismo por motivos de salud mental. La OMS (2023) estima que los trastornos de ansiedad y depresión generan pérdidas globales de un billón de dólares anuales en productividad, fenómeno particularmente visible en España donde las bajas por esta causa han aumentado un 30% desde 2020 (Osakidetza, 2023). Paralelamente, la implantación del teletrabajo ha generado efectos paradójicos: mientras reduce el absentismo físico, ha incrementado el presenteísmo digital, es decir, la tendencia a trabajar enfermo desde casa (Gartner, 2023). Estudios del IESE (2023) revelan que el 42% de las empresas españolas que han implementado modelos híbridos han observado reducciones en las tasas de absentismo, aunque persisten desafíos en la gestión del bienestar laboral.

La tecnología emerge como herramienta clave para abordar estos retos, con soluciones innovadoras como la inteligencia artificial y *wearables* que permiten predecir episodios de fatiga laboral (Deloitte, 2023), o plataformas como Microsoft Viva que analizan patrones de ausencia para diseñar intervenciones preventivas (Harvard Business Review, 2023). No obstante, los expertos coinciden en que cualquier estrategia efectiva debe combinar estos avances tecnológicos con políticas organizacionales que fomenten entornos laborales saludables, sistemas de compensación adecuados y una gestión más humana del talento. El caso español resulta particularmente ilustrativo de cómo factores demográficos, culturales y estructurales interactúan para configurar un panorama complejo que requiere soluciones multidimensionales y adaptadas a cada contexto sectorial.

2.3. Presentación caso de estudio: Absentismo en empresa logística

En este trabajo fin de máster, se plantea el caso real de una empresa dedicada a la logística y servicios de paquetería, especializada en el transporte y

distribución de mercancías, la cual tiene problemas con el absentismo laboral, con los problemas que esto conlleva para su actividad como es el empeoramiento del clima laboral, la reducción de la productividad de la empresa, el deterioro del servicio al cliente entre otros problemas.

En este sentido, este proyecto tiene como objetivo principal desarrollar una labor de consultoría especializada para la empresa, centrada en dos líneas de acción complementarias. Por un lado, se busca implementar un sistema de monitorización que permita realizar un seguimiento más preciso del absentismo, identificando patrones recurrentes y las posibles causas que lo originan. Por otro lado, se pretende diseñar y aplicar un modelo predictivo avanzado capaz de anticipar los índices futuros de absentismo. Esta herramienta facilitaría una gestión más proactiva de los recursos humanos, permitiendo a la organización adoptar medidas preventivas que minimicen el impacto operativo de las ausencias no planificadas. El desarrollo de estas soluciones no solo mejoraría la situación actual de la empresa, sino que podría servir como referencia para otras organizaciones del sector que enfrenten desafíos similares.

Este caso de estudio resulta relevante ya que no se ha realizado anteriormente una investigación de estas características, en el que se utilizan datos diarios reales sobre el absentismo de una empresa logística, sobre la cual se realiza un seguimiento y predicción del absentismo de sus trabajadores.

3. Datos y metodología

Las tres bases de datos utilizadas en este análisis han sido proporcionadas directamente por la empresa logística y consisten en varios archivos CSV que recopilan, de manera diaria y para cada empleado, los datos correspondientes a las variables:

- Horas planificadas: Número total de horas que un trabajador debería trabajar según su contrato, jornada laboral y planificación de turnos.
- Horas absentismos por incapacidad temporal (IT): Total de horas laborales no trabajadas debido a las ausencias de empleados que se encuentran en situación de baja médica, es decir, incapacitados temporalmente por enfermedad o accidente (laboral o no laboral) y amparados por un parte de confirmación médica.

La primera base de datos fue enviada por la empresa con los datos recopilados hasta marzo, la segunda hasta mayo y la tercera hasta junio de los períodos 2024 y 2025 completos, ya que, aunque 2025 no haya finalizado hay bajas que ya están planificadas y por ello existen datos.

Además de las variables iniciales con las que se contaba, adicionalmente se han añadido por su relevancia para el caso de estudio el tipo de día (festivo nacional, fin de semana o laborable) y el día de la semana, permitiendo obtener más información y detectar posibles tendencias, mejorar predicciones, etc.

Tabla 3.1 Variables de la base de datos

Variable	Descripción	Unidad	Tipo variable
Horas planificadas	Número total de horas programadas para trabajar en un período determinado.	Horas (h)	Cuantitativa (discreta)
Horas absentismo (IT)	Horas perdidas debido a incapacidad temporal (bajas médicas).	Horas (h)	Cuantitativa (discreta)

Tipo Día	Clasificación del día: Laborable, Festivo Nacional o Fin de semana.	-	Cualitativa (nominal)
Día semana	Día de la semana (lunes, martes, miércoles, etc.).	-	Cualitativa (ordinal)

Fuente: Elaboración propia, 2025

El presente análisis el cual entre sus objetivos estaría predecir el absentismo laboral por incapacidad temporal implementa un marco metodológico riguroso que integra técnicas avanzadas de modelado de series temporales con protocolos de validación estrictos.

El enfoque metodológico combina modelos univariantes clásicos y contemporáneos, comenzando con el modelo Naïve que constituye el punto de partida, estableciendo una línea base de comparación mediante su supuesto de persistencia temporal. Esta aproximación, representada formalmente como $\hat{Y}_{t+h} = Y_t$, postula que el valor futuro (\hat{Y}_{t+h}), donde h es el horizonte de predicción, de la serie temporal será idéntico al último valor observado (Y_t), proporcionando así un referente mínimo de desempeño que cualquier modelo más sofisticado debería superar para demostrar su utilidad práctica. Su implementación sistemática permite cuantificar el valor añadido que aportan técnicas más complejas.

Los modelos ETS (*Error-Trend-Seasonal*) representan una familia de técnicas de suavizado exponencial que descomponen la serie temporal en componentes estructurales mediante un marco formal probabilístico conocido como espacio de los estados (Hyndman & Athanasopoulos, 2021). La formulación general $Y_t = \mu_t + \beta_t + \gamma_t + \varepsilon_t$ desagrega la serie en un componente de nivel (μ_t), tendencia (β_t), un patrón estacional (γ_t) y un término de error aleatorio (ε_t), donde cada elemento se estima mediante procesos de actualización recursiva que minimizan la función de verosimilitud. La variante ETS (A, N, A) específicamente modela series con error aditivo, sin tendencia y con estacionalidad aditiva, pero sin componente de tendencia, siendo particularmente adecuada para patrones que exhiben fluctuaciones periódicas constantes en amplitud. Por contra, ETS (A, A,

N) incorpora una tendencia aditiva global, capturando así cambios progresivos en el nivel subyacente de la serie, lo que resulta útil para fenómenos sujetos a deriva temporal sistemática.

La metodología ARIMA (*AutoRegressive Integrated Moving Average*) proporciona un marco unificado para modelar procesos estocásticos no estacionarios mediante operadores de diferencia y componentes autorregresivos/móviles. Su formulación canónica $\phi_p(B)(1 - B)^d Y_t = \theta_q(B) + \varepsilon_t$ integra tres componentes clave: el polinomio autorregresivo $\phi_p(B)$ que captura dependencias de largo alcance, el operador de diferencia $(1 - B)^d$ que transforma la serie a estacionariedad, y el polinomio de media móvil $\theta_q(B)$ que modela la estructura de correlación en los residuos. La variable B es el operador retardo. La selección automática de órdenes (p,d,q) mediante el criterio de información de Akaike corregido (AICc) garantiza un balance óptimo entre bondad de ajuste y parsimonia, evitando así el sobreajuste mientras se capturan las dinámicas temporales esenciales (Akaike, 1974).

Las redes neuronales autorregresivas (NNAR) constituyen un enfoque no paramétrico que aproxima relaciones funcionales complejas mediante arquitecturas de aprendizaje profundo. La formulación $Y_t = f(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}) + \varepsilon_t$ define un mapeo no lineal donde una red *feed-forward*¹ con capas ocultas transforma los valores retardados de la serie en predicciones puntuales. La configuración específica empleada ($p=7$, $size=6$) implica que la red procesa una ventana temporal de siete períodos históricos a través de seis neuronas en la capa oculta, permitiendo capturar interacciones no lineales entre retardos. Las cincuenta repeticiones con inicializaciones aleatorias mitigan el riesgo de convergencia a mínimos locales subóptimos durante el entrenamiento, mientras que la normalización de inputs asegura estabilidad numérica en el proceso de optimización.

El protocolo de validación *rolling-origin recalibration* implementa un esquema de evaluación riguroso que respeta la estructura temporal inherente a los datos.

¹ Red feed-forward: es un tipo de arquitectura de red neuronal artificial donde la información fluye en una sola dirección: desde la capa de entrada, a través de las capas ocultas (si existen), hasta la capa de salida, sin ciclos o retroalimentación.

Este enfoque define secuencialmente ventanas de entrenamiento expansivas (comenzando con 100 observaciones) desde las cuales se proyectan predicciones para horizontes multiperíodo, almacenando sistemáticamente los resultados en matrices tridimensionales donde los índices representan respectivamente el horizonte de predicción y el punto de origen temporal. Este diseño permite caracterizar exhaustivamente la evolución del error predictivo tanto en función del horizonte temporal como de la posición en el ciclo estacional, proporcionando así una evaluación multidimensional del rendimiento modelo.

Las métricas de error RMSE (*Root Mean Square Error*) y MdAE (*Median Absolute Error*) ofrecen perspectivas complementarias sobre la calidad predictiva. El RMSE cuantifica la dispersión cuadrática promedio de los errores para cada paso de predicción, siendo particularmente sensible a desviaciones grandes. Por su parte, la mediana del error absoluto (MdAE) proporciona una medida robusta de tendencia central insensible a valores atípicos, revelando así el desempeño típico del modelo en condiciones normales. Su cálculo condicional por horizonte permite identificar patrones de degradación predictiva a medida que aumenta la distancia temporal desde el último punto observado.

El análisis comparativo sistemático revela compensaciones fundamentales entre los distintos enfoques: mientras los modelos ARIMA exhiben eficiencia en la captura de dependencias lineales y estabilidad en condiciones de estacionariedad, las redes NNAR demuestran mayor flexibilidad para adaptarse a regímenes no lineales y cambios estructurales, aunque a costa de mayor variabilidad en sus estimaciones y requerimientos computacionales intensivos. El examen diagnóstico de los residuos mediante pruebas de Ljung-Box y análisis de autocorrelación residual proporciona validación adicional sobre la adecuación especificativa, verificando la ausencia de patrones sistemáticos no capturados en los residuos y confirmando así que cada modelo ha extraído adecuadamente la información predictiva disponible en la estructura temporal de los datos.

Para llevar a cabo esta tarea de investigación y consultoría para la empresa logística, se ha planteado el entorno de desarrollo integrado RStudio, el cual opera con el lenguaje de programación R. Véase tabla 3.2.

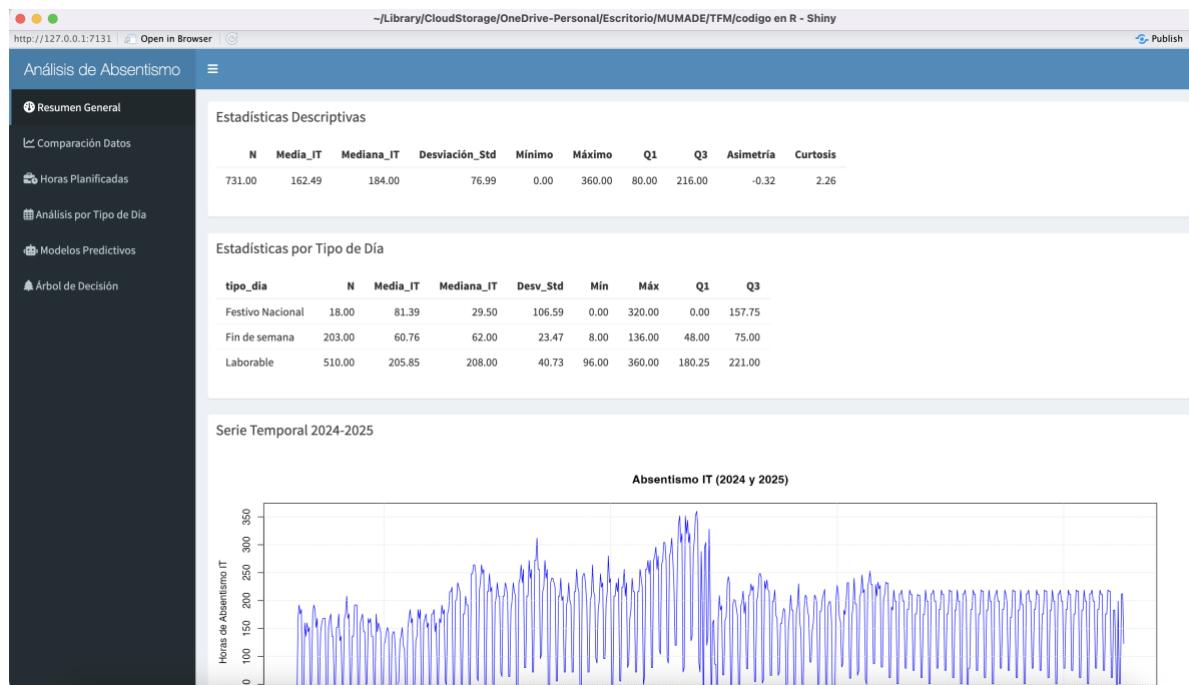
Tabla 3.2 Diccionario de paquetes utilizados

Paquete	Utilidad
Readxl	Leer archivos Excel.
Forecast	Modelos de series temporales (ETS, ARIMA, etc.).
Dplyr	Manipulación de datos (filter, mutate, etc.).
Ggplot2	Visualizaciones avanzadas.
Moments	Cálculo de asimetría, curtosis, etc.
Shiny	Crear la app web interactiva.
Shinydashboard	Interfaz tipo dashboard para Shiny.

Fuente: Elaboración propia, 2025

Además de realizar el análisis con este lenguaje adicionalmente se ha creado con la librería Shiny un Dashboard con el fin de facilitar la tarea de consultoría. Véase figura 3.1.

Figura 3.1. Dashboard Análisis Absentismo para empresa logística



Fuente: Elaboración propia, 2025

4. Resultados

4.1 Análisis exploratorio

En primer lugar, se han calculado los estadísticos descriptivos de la variable absentismo laboral por incapacidad temporal como se ve reflejado en la tabla 4.1.1. El análisis de los datos de absentismo laboral revela una media de 162,49 horas. La mediana, que es de 184 horas, supera ligeramente la media, lo que sugiere una ligera asimetría negativa en la distribución de los datos, confirmada por el coeficiente de asimetría de -0,32. Esto implica que hay más días con valores de absentismo por encima de la media que por debajo. La desviación estándar de 76,99 horas y la varianza de 5926,84 reflejan una dispersión considerable en los datos, lo que indica que el absentismo varía significativamente de un día a otro. El rango de los datos, desde un mínimo de 0 horas hasta un máximo de 360 horas, muestra la existencia de días sin absentismo y días con absentismo muy elevado. Los cuartiles Q1 (80 horas) y Q3 (216 horas), junto con un rango intercuartílico de 136 horas, destacan la variabilidad en la mitad central de los datos. La curtosis de 2,26, cercana al valor de una distribución normal (3), sugiere que los datos no presentan colas extremadamente pesadas, aunque la ligera asimetría negativa indica una distribución con una cola izquierda más larga. En conjunto, estos resultados muestran un patrón de absentismo con días de alta variabilidad y una tendencia hacia valores más altos.

Tabla 4.1.1. Estadísticos descriptivos globales sobre Absentismo IT

N	Media	Mediana	Desviación estandar	Varianza	Mín
731	162,49	184	76,99	5926,84	0
Máx	Q1	Q3	Rango_interc	Asimetría	Curtosis
360	80	216	136	-0,32	2,26

Fuente: Elaboración propia, 2025

El análisis del absentismo laboral por tipo de día revela diferencias significativas en los patrones de comportamiento. En los días laborables, la media de absentismo es la más alta (205,85 horas), con una mediana cercana (208 horas), lo que indica una distribución relativamente simétrica, respaldada por una asimetría de 0,76. La desviación estándar (40,73) y la varianza (1658,89) son moderadas, mostrando una dispersión controlada en comparación con otros tipos de días. El rango intercuartílico (40,75) sugiere que la mayoría de los datos se concentran en un intervalo relativamente estrecho, aunque con valores mínimos y máximos que van desde 96 hasta 360 horas. La curtosis de 4,58, superior a la de una distribución normal, indica una distribución más puntiaguda con colas más pesadas, lo que podría deberse a la presencia de días con absentismo excepcionalmente alto.

En contraste, los fines de semana presentan una media mucho menor (60,76 horas) y una mediana similar (62 horas), con una desviación estándar baja (23,47) y una varianza pequeña (550,89), lo que refleja una menor variabilidad en el absentismo. La asimetría de 0,37 y la curtosis de 3,63 sugieren una distribución ligeramente sesgada hacia la derecha y con colas un poco más pesadas que la normal, pero en general más estable y predecible que en los días laborables.

Los días festivos nacionales muestran el comportamiento más irregular, con una media de 81,39 horas, pero una mediana mucho más baja (29,5 horas), lo que indica una fuerte asimetría positiva (0,98) debido a la presencia de algunos días con absentismo extremadamente alto (hasta 320 horas). La desviación estándar (106,59) y la varianza (11360,49) son las más altas, reflejando una gran dispersión en los datos. El rango intercuartílico de 157,75 horas confirma esta variabilidad, con ausencia de absentismo en algunos días (mínimo de 0 horas) y picos significativos en otros. La curtosis de 2,50 sugiere una distribución menos concentrada que en los otros tipos de días, con una cola derecha más alargada.

En resumen, los días laborables presentan los niveles más altos y consistentes de absentismo, mientras que los fines de semana son los más estables y predecibles. Los festivos, por su parte, muestran un comportamiento heterogéneo, con algunos días de muy bajo absentismo y otros con valores

excepcionalmente altos, lo que podría estar relacionado con eventos específicos o políticas de la empresa en esas fechas.

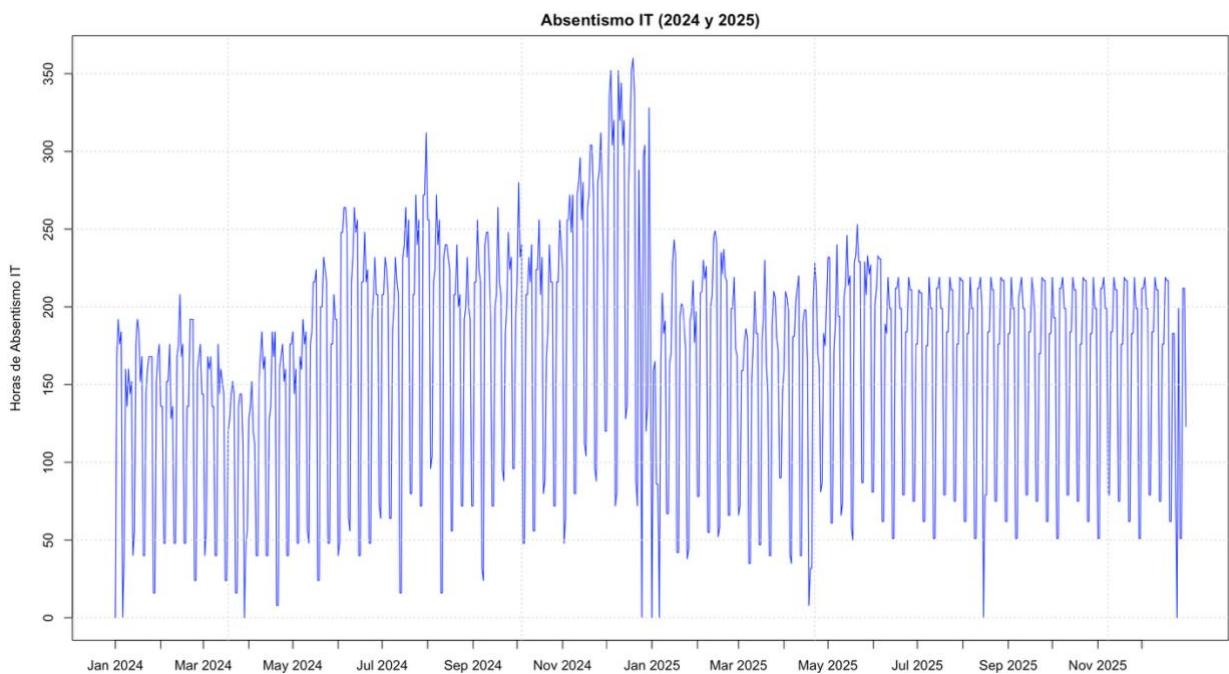
Tabla 4.1.2. Estadísticos descriptivos por tipo de día sobre Absentismo IT

Tipo día	N	Media	Mediana	Desviación estandar	Varianza	Mín
Festivo Nacional	18	81,39	29,5	106,59	11360,49	0
Fin de semana	203	60,76	62	23,47	550,89	8
Laborable	510	205,85	208	40,73	1658,89	96
Tipo día	Máx	Q1	Q3	Rango_interc	Asimetría	Curtosis
Festivo Nacional	320	0	157,75	157,75	0,98	2,50
Fin de semana	136	48	75	27	0,37	3,63
Laborable	360	180,25	221	40,75	0,76	4,58

Fuente: Elaboración propia, 2025

La serie temporal de absentismo IT para el periodo 2024-2025 en la figura 4.1.1 muestra fluctuaciones a lo largo del período analizado, con picos y valles que sugieren variaciones estacionales o eventos puntuales que impactan en las horas de absentismo. Se observa la presencia de múltiples puntos clave (como meses de inicio y mitad de año) permite inferir que el absentismo no sigue un patrón completamente estable, sino que presenta oscilaciones que podrían estar relacionadas con factores como temporadas de mayor carga laboral, festividades o incluso cambios organizacionales.

Figura 4.1.1. Serie temporal del absentismo laboral IT (2024-2025)

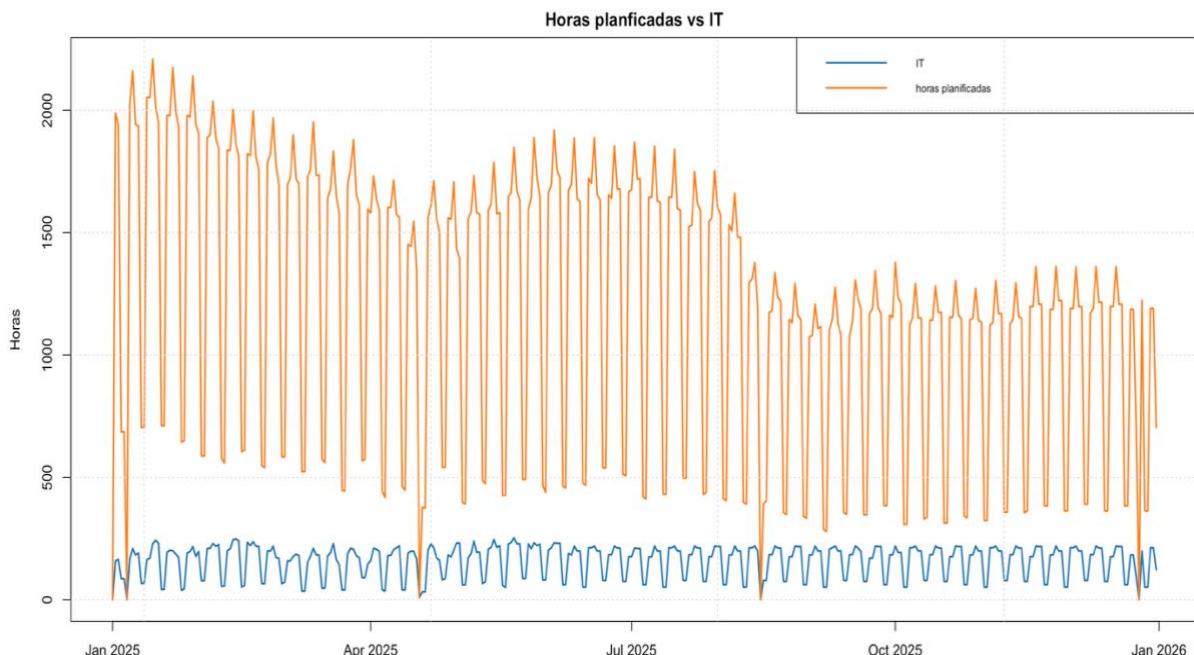


Fuente: Elaboración propia, 2025

Respecto a las horas planificadas por la empresa en 2025 mostradas en la figura 4.1.2 se observa cómo hay días que se obtiene valores iguales a 0 correspondientes a los días festivos nacionales. También, se ve como en el mes de enero todavía existen valores muy elevados de horas planificadas los cuales se deben a que en los meses de Navidad aumenta el servicio de paquetería y aumenta el número de empleados en la empresa, pero una vez finaliza este periodo festivo se produce una reducción de las horas planificadas.

En cuanto a las horas planificadas y las horas de absentismo por incapacidad temporal, se observa que ambos indicadores fluctúan a lo largo del período, con puntos de convergencia y divergencia que podrían reflejar momentos de mayor estrés laboral o eficiencia en la gestión de recursos. Este comportamiento refuerza la importancia de monitorizar ambas variables para equilibrar la productividad y el bienestar laboral.

Figura 4.1.2. Horas planificadas vs IT (2025)



Fuente: Elaboración propia, 2025

Los datos sobre el absentismo por incapacidad temporal por tipo de día (Festivo Nacional, laboral, fin de semana) revelan diferencias significativas. Los días laborables registran el mayor impacto, acumulando 104.982 horas de IT en 510 días, lo que representa un promedio aproximado de 206 horas por día. Esta cifra dominante refleja que la mayor parte del absentismo ocurre durante la jornada laboral habitual. Los fines de semana muestran un absentismo notablemente menor, con 12.335 horas en 203 días (un promedio de 61 horas diarias). Por último, los festivos nacionales presentan el menor volumen de absentismo (1.465 horas en 18 días, unas 81 horas por día), pero con una alta variabilidad (como se observó en análisis previos, con algunos festivos sin absentismo y otros con picos excepcionales). En conjunto, los datos confirman que el absentismo está estrechamente ligado a la dinámica laboral, siendo los días laborables los más críticos. Véase tabla 4.1.3.

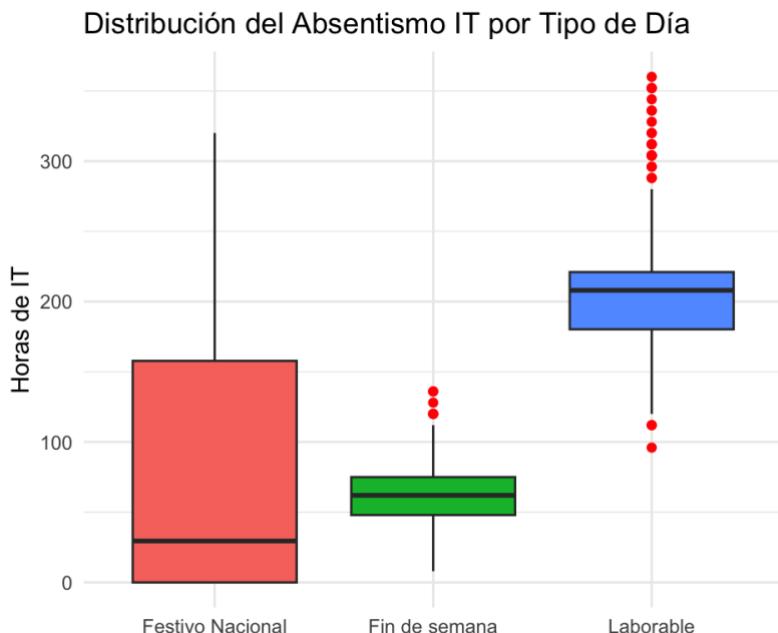
Tabla 4.1.3. Absentismo por tipo de día

Tipo de día	Media	Mediana
Laborable	104.982	510
Fin de semana	12.335	203
Festivo Nacional	1.465	18

Fuente: Elaboración propia, 2025

La figura 4.1.3 de distribución del absentismo IT por tipo de día muestra claras diferencias en los patrones de horas perdidas. Los días laborables presentan la mayor concentración de absentismo, con una distribución más amplia y valores generalmente más altos, lo que refleja el impacto directo de la actividad laboral regular en las bajas por incapacidad temporal. Los fines de semana, por otro lado, muestran una distribución más compacta y con valores notablemente inferiores, lo que coincide con la menor actividad laboral durante estos días. Los festivos nacionales, aunque con menos datos disponibles, presentan una dispersión irregular, con algunos casos extremos que elevan el rango de horas de absentismo, posiblemente vinculados a eventos específicos o políticas particulares de la empresa durante estos días. Esta visualización refuerza la idea de que el absentismo está íntimamente relacionado con la dinámica laboral cotidiana, siendo los días de trabajo ordinarios los que generan mayor impacto en esta métrica.

Figura 4.1.3. Boxplot por tipo de día (periodo 2024-2025)



Fuente: Elaboración propia, 2025

Según la tabla 4.1.4, aunque la mayoría de las horas planificadas se concentran en los días laborables (386.905 horas), lo que naturalmente da lugar al mayor volumen absoluto de ausencias (51.126 casos de IT), al observar la tasa relativa (calculada como el porcentaje de IT sobre las horas planificadas) se obtiene una perspectiva más matizada sobre el fenómeno.

Los festivos nacionales, a pesar de representar solo una pequeña fracción del total de horas planificadas (2.492), registran la tasa de absentismo más elevada con un 18,02%. Este dato puede interpretarse como un indicio de que, en jornadas excepcionales en las que se mantiene la actividad, el personal que debe acudir puede hacerlo en condiciones más desfavorables o con menor motivación, lo que incrementa la probabilidad de ausencias. Asimismo, puede estar relacionado con factores como la cobertura mínima de personal, la fatiga acumulada o la presión adicional derivada de trabajar en días tradicionalmente no laborables. Por su parte, los fines de semana presentan una tasa también elevada (13,97%), ligeramente superior a la de los días laborables (13,21%), a pesar de contar con muchas menos horas planificadas (45.739 frente a 386.905). Esto sugiere que el absentismo en fines de semana es proporcionalmente más frecuente, posiblemente por tratarse de turnos menos deseados o más exigentes

debido a la menor disponibilidad de personal. Además, podría estar influido por el tipo de tareas logísticas que se concentran en esos días, que a menudo implican mayor carga física o emocional.

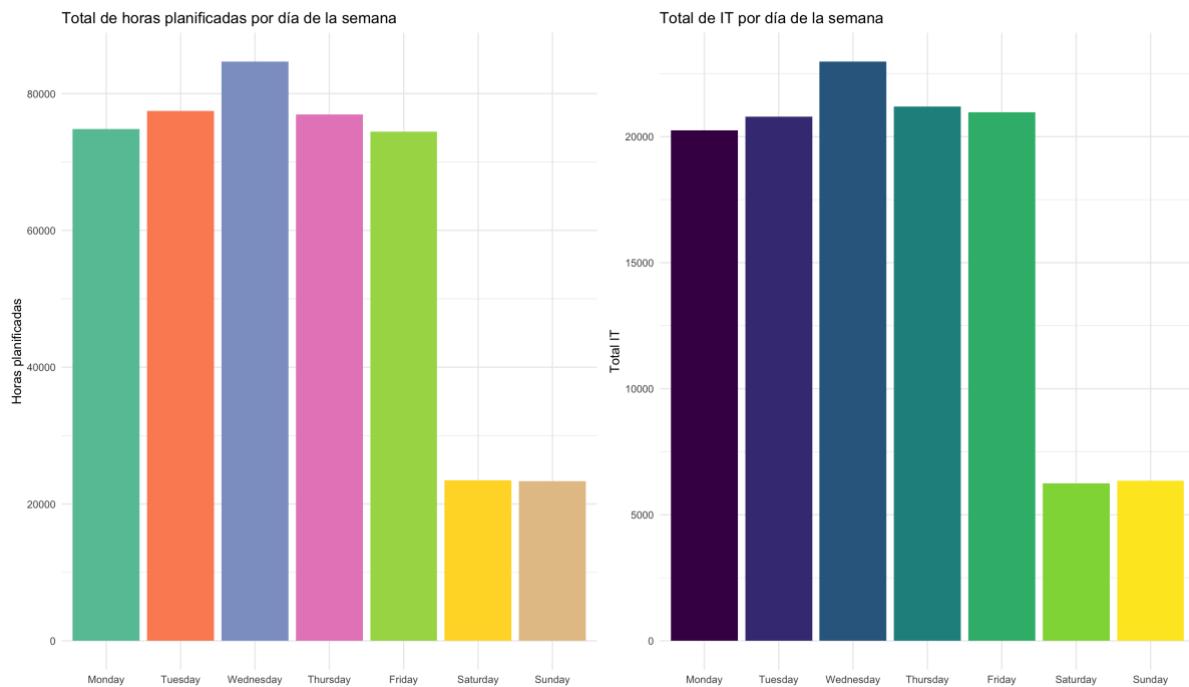
Tabla 4.1.4. Absentismo por tipo día (2025)

Tipo de día	Horas planificadas	Total IT	Tasa IT (%)
Festivo Nacional	2.492	449	18,02
Fin de semana	45.739	6.391	13,97
Laborable	386.905	51.126	13,21

Fuente: Elaboración propia, 2025

La figura 4.1.4 muestra que el miércoles en términos absolutos destaca como el día con mayor absentismo IT. Los niveles se mantienen altos el jueves, aunque ligeramente menores que el miércoles. El viernes muestra un descenso notable, probablemente por la cercanía del fin de semana. Esta distribución inusual, con el máximo a mitad de semana en lugar de al inicio, plantea interrogantes sobre la organización del trabajo y los procesos que podrían estar generando esta tensión específica en los días centrales de la agenda laboral.

Figura 4.1.4. Total horas planificadas y total horas absentismo IT por día de la semana (2025)



Fuente: Elaboración propia, 2025

Sin embargo, la tabla 4.1.5 matiza al introducir la tasa de absentismo (IT/horas planificadas), revelando que: aunque los días laborables acumulan más horas perdidas en términos absolutos, los fines de semana presentan tasas superiores (14,20% domingo vs. 12,88% lunes). Esto sugiere que, proporcionalmente, el impacto del absentismo es mayor cuando menos personal está activo, posiblemente por la presión sobre equipos reducidos o la naturaleza crítica de los servicios operativos en esos días. La consistencia de las tasas entre jueves (13,47%) y viernes (13,65%) refuerza la idea de que el absentismo responde no solo a la carga laboral, sino a factores acumulativos de estrés durante la semana laboral.

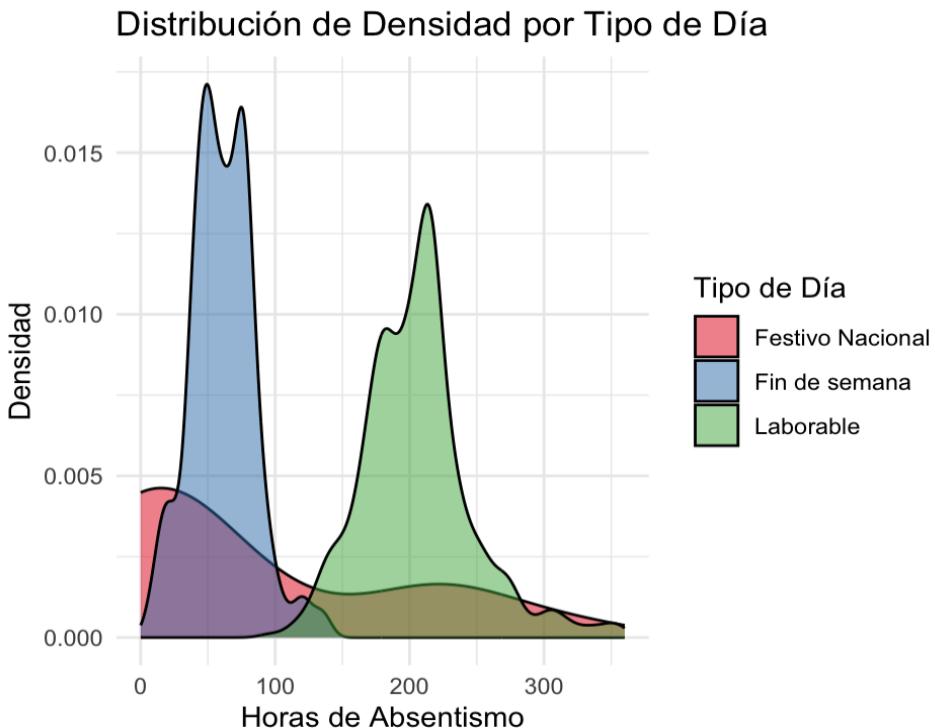
Tabla 4.1.5. Tasa de absentismo IT por horas planificadas por día de la semana (2025)

Día semana	Horas planificadas	Total IT	Tasa IT (%)
Domingo	23.308	3.309	14,20
Sábado	23.482	3.291	14,01
Viernes	74.461	10.164	13,65
Jueves	76.917	10.364	13,47
Miércoles	84.672	11.184	13,21
Martes	74.845	9.677	12,93
Lunes	77.451	9.977	12,88

Fuente: Elaboración propia, 2025

La distribución de densidad revela patrones diferenciados en el absentismo según el tipo de día. Los días laborables presentan una curva de densidad más pronunciada y centrada en valores altos de horas de absentismo, confirmando que la mayoría de las incidencias se concentran en estos días con una distribución relativamente homogénea. Los fines de semana muestran una curva más estrecha y con valores notablemente inferiores, reflejando una menor variabilidad y un absentismo reducido, aunque constante. En contraste, los festivos nacionales exhiben una curva más plana y extendida hacia la derecha, lo que indica una mayor dispersión con casos puntuales de absentismo muy elevado que distorsionan el patrón general. Esta distribución bimodal en festivos sugiere dos comportamientos distintos: días con mínima incidencia y otros con picos excepcionales, posiblemente vinculados a políticas específicas o eventos particulares. La superposición parcial de las curvas entre laborables y fines de semana en ciertos rangos podría señalar la existencia de turnos de trabajo durante los fines de semana que replican parcialmente los patrones de días laborales. Véase figura 4.1.5.

Figura 4.1.5. Distribución de densidad de las horas de absentismo IT por tipo día (2024-2025)

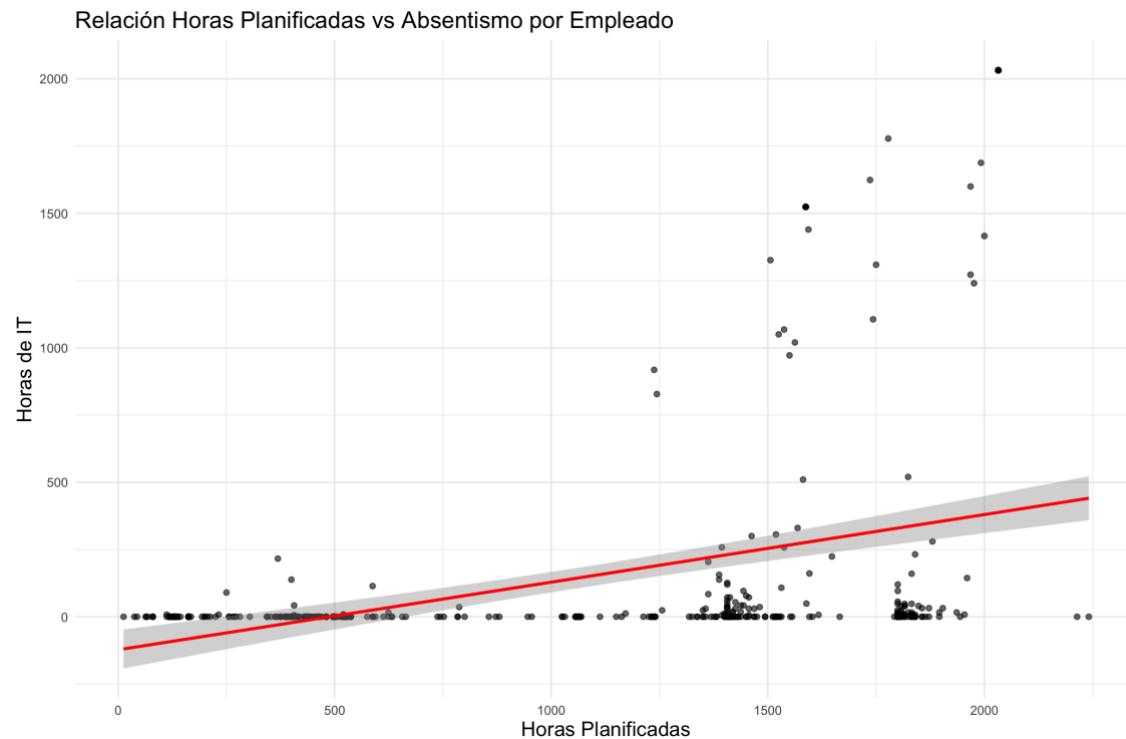


Fuente: Elaboración propia, 2025

La figura 4.1.6 muestra una relación positiva entre las horas planificadas y el absentismo por empleado, donde se observa una tendencia general de aumento en las horas de IT conforme incrementa la carga laboral asignada. Sin embargo, la dispersión de los puntos revela una variabilidad significativa, indicando que esta relación no es uniforme ni determinista. Se identifican tres patrones clave: un grupo de empleados con baja carga horaria y correspondiente bajo absentismo (esquina inferior izquierda), otro grupo con cargas medias que presentan una amplia gama de valores de IT (dispersión central), y casos puntuales con altas cargas que generan desde absentismo moderado hasta extremo (esquina superior derecha). La presencia de outliers notables (empleados con muchas horas planificadas, pero bajo IT, y viceversa) sugiere que factores individuales (como resiliencia, condiciones de salud o apoyo organizacional) moderan sustancialmente esta relación. La nube de puntos es particularmente densa entre 500-1500 horas planificadas, rango donde parecen concentrarse la mayoría de los empleados, mostrando que incluso dentro de cargas laborales similares existen diferencias importantes en las tasas de

absentismo. Este patrón refuerza la necesidad de enfoques personalizados en la gestión de recursos humanos, ya que políticas uniformes podrían ser insuficientes para abordar las diversas realidades que subyacen a esta relación compleja entre planificación y salud laboral.

Figura 4.1.6. Relación de horas planificadas y de absentismo IT por empleado (2025)



Fuente: Elaboración propia, 2025

Para evaluar la posición de la empresa respecto al indicador de absentismo laboral calculado como el cociente entre el total de horas de absentismo por incapacidad temporal (IT) y el total de horas planificadas se obtuvo un resultado del 13,32% lo cual es bastante más elevado que la media del sector en España (5%-8%) y Europa (4%-7%), según informes de la Asociación Española de Logística (2023).

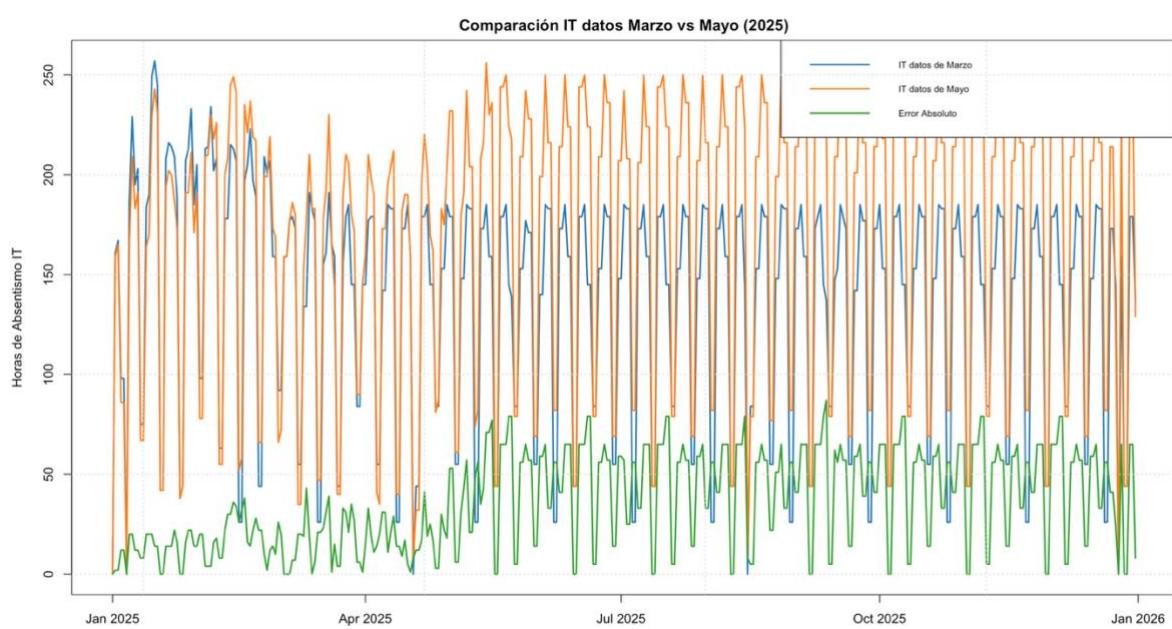
El resultado de la eficiencia calculada (0,8667 o 86,67%) indica que, considerando el total de horas planificadas y descontando las horas perdidas por incapacidad temporal (IT), la empresa opera con un 86,67% de eficiencia en la gestión del tiempo laboral. Esto implica que el 13,33% restante (calculado como $1 - 0,8667$) corresponde a horas no trabajadas debido al absentismo por IT, lo cual refleja una pérdida significativa de productividad. Comparado con los

promedios del sector logístico en España (5%-8% de absentismo) y Europa (4%-7%), este valor sugiere que la empresa tiene un problema de absentismo superior al estándar del mercado, ya que su tasa de horas perdidas (13,33%) casi duplica la media nacional e incluso triplica la de algunos países europeos.

Por otro lado, la empresa como se mencionó en apartados anteriores envió 3 bases de datos (marzo, mayo, junio) dado que se van modificando los valores pasados y futuros de absentismo en la empresa. Los valores futuros se van modificando ya que se van teniendo en cuenta las horas de absentismo que los empleados van indicando a la empresa y pasadas ya que, aunque deberían ser las mismas existen cambios debidos a que algunas bajas las ponen a posteriori porque en ciertas ocasiones no ha dado tiempo a registrarlas en el sistema.

Véase figuras 4.1.7, 4.1.8 y 4.1.9.

Figura 4.1.7. Serie temporal del absentismo laboral IT (2024-2025)



Fuente: Elaboración propia, 2025

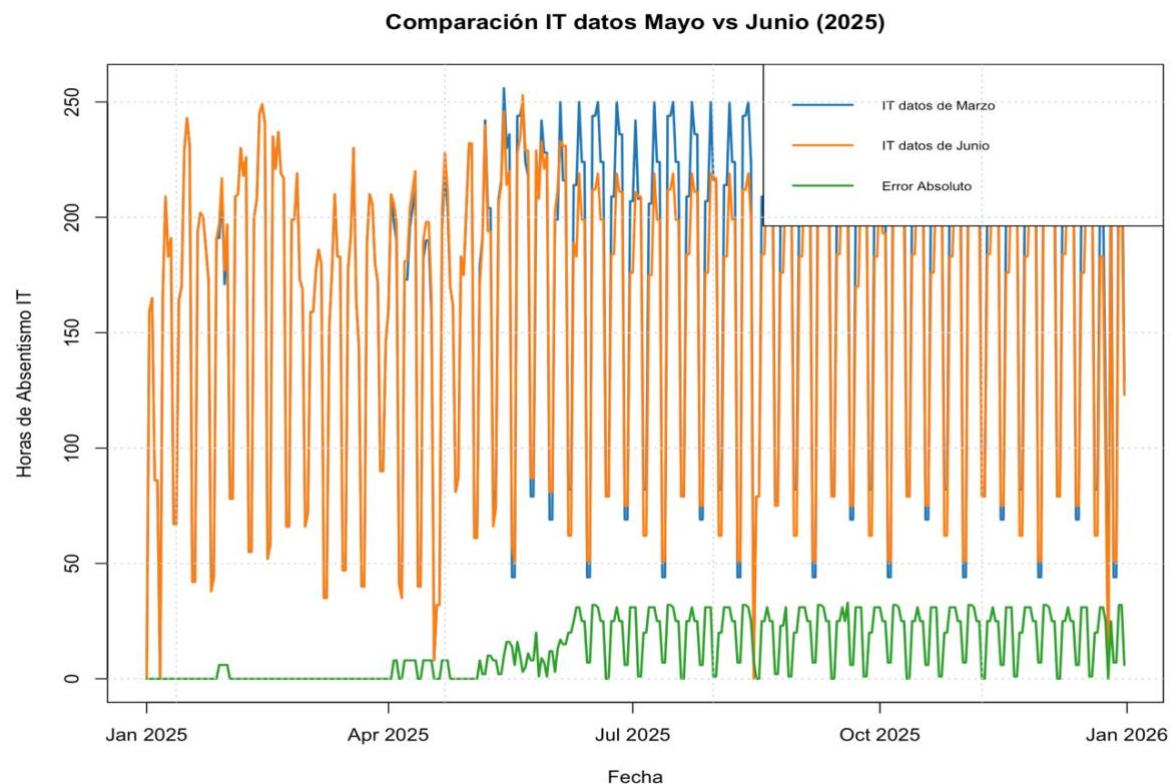
Figura 4.1.8. Serie temporal del absentismo laboral IT (2024-2025)



Fuente: Elaboración propia, 2025

En la figura 4.1.9 finalmente se observa como en junio en comparación con mayo los datos pasados no cometan apenas errores ya que las bajas que no habían sido registradas en su mayoría ya si lo hicieron.

Figura 4.1.9. Serie temporal del absentismo laboral IT (2024-2025)



Fuente: Elaboración propia, 2025

Con el RMSE y el error absoluto también se muestra como con el tiempo el error que se había cometido respecto a los datos iniciales de marzo se ha ido reduciendo y el motivo se debe a las bajas que no habían sido registradas en el pasado y que si correspondía su registro. Véase tabla 4.1.6. y figuras 4.1.7, 4.1.8 y 4.1.9.

Tabla 4.1.6. Absentismo por tipo de día

Datos	RMSE
Marzo - Mayo	43,79
Marzo - Junio	30,31
Mayo - Junio	18,08

Fuente: Elaboración propia, 2025

4.2 Predicciones puntuales

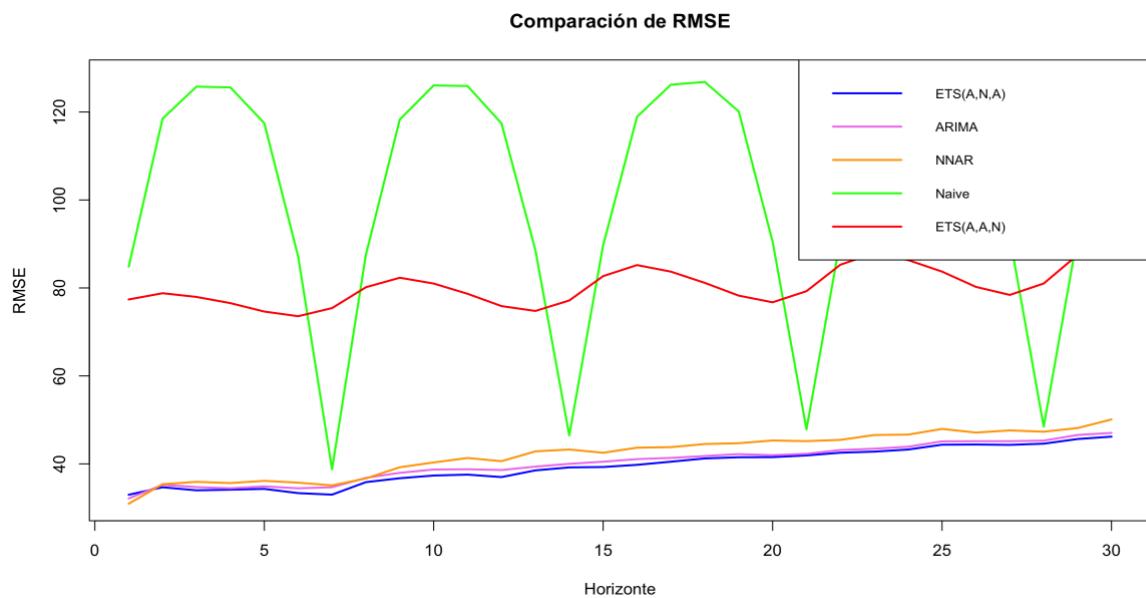
Se realizaron predicciones de la variable horas de absentismo por incapacidad temporal (IT) utilizando múltiples modelos de *forecasting* (ETS (ANA), ARIMA, Redes Neuronales, Naïve y ETS con tendencia aditiva (AAN)) para evaluar cuál ofrece mejores resultados en un horizonte de 30 días, entrenando cada modelo iterativamente con una ventana de datos inicial de 100 datos que crece progresivamente.

La figura 4.2.1 y los valores obtenidos de RMSE muestran diferencias significativas en el desempeño de los modelos predictivos para el absentismo por IT. El modelo ETS(ANA) presenta el menor error (RMSE = 39,42), lo que indica que es el más preciso en las predicciones a lo largo del horizonte de 30 días, destacándose especialmente en horizontes cortos. Le sigue muy de cerca el ARIMA (RMSE = 40,23), que también demuestra robustez, aunque con un ligero incremento en el error, probablemente debido a su menor capacidad para capturar componentes estacionales específicos. Por otro lado, el modelo NNAR (RMSE = 42,19) muestra un rendimiento aceptable pero inferior a ETS(ANA) y ARIMA, lo que sugiere que, aunque las redes neuronales pueden capturar patrones complejos, en este caso no superan a los métodos tradicionales.

En contraste, el modelo Naïve (RMSE = 102,31) tiene un error significativamente mayor, confirmando que su simplicidad (basada en repetir el último valor observado) no es adecuada para series con patrones temporales marcados, como el absentismo. Finalmente, el ETS(AAN) (RMSE = 80,40) muestra un desempeño intermedio, pero su error casi duplica al de ETS(ANA), lo que resalta la importancia de seleccionar la configuración correcta de suavizado exponencial (en este caso, la ausencia de estacionalidad en el modelo AAN perjudica su precisión).

La brecha entre los modelos avanzados y el Naïve justifica la inversión en técnicas más sofisticadas, aunque el bajo rendimiento relativo de NNAR podría indicar que, para esta serie específica, la estructura de la red neuronal no está optimizada o que los datos no contienen patrones no lineales aprovechables.

Figura 4.2.1. RMSE comparación por tipo de modelo



Fuente: Elaboración propia, 2025

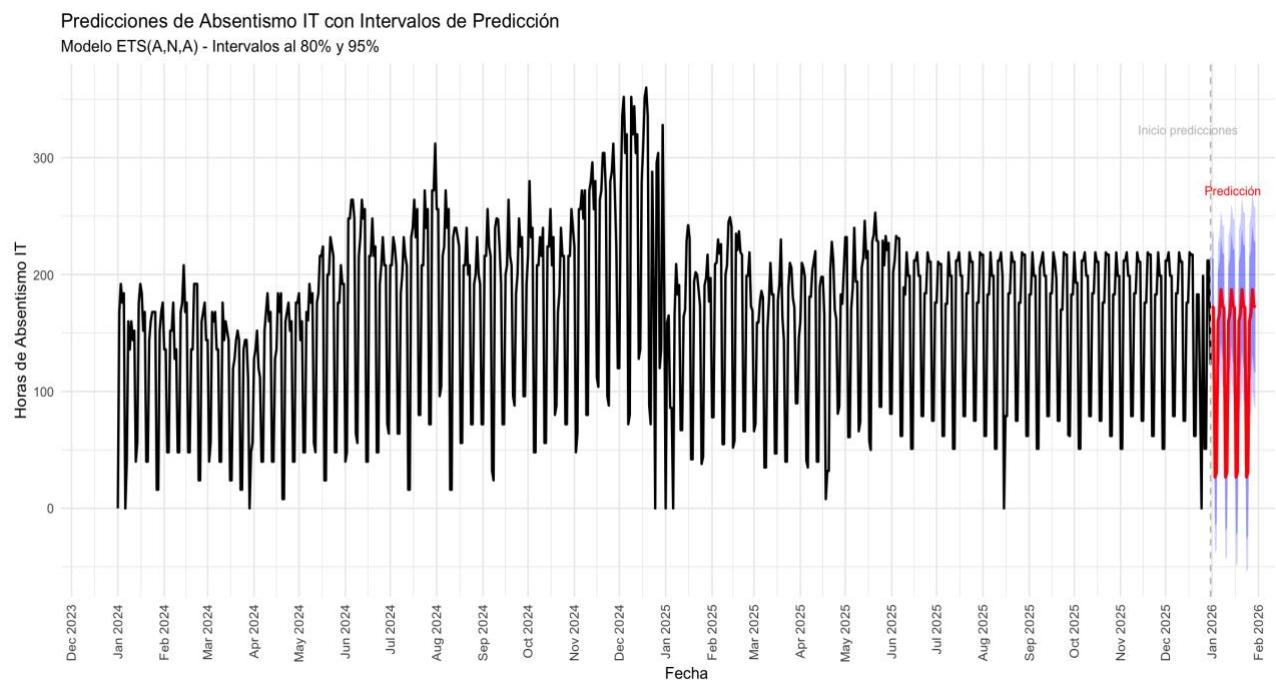
4.3 Predicciones probabilísticas

La figura 4.3.1 presenta las proyecciones de absentismo por incapacidad temporal (IT) generadas por el mejor modelo en términos del RMSE en este caso el ETS (A, N, A), que combina un componente de error aditivo, ausencia de tendencia y estacionalidad aditiva. Los resultados muestran las predicciones puntuales junto con dos bandas de incertidumbre: intervalos de predicción del 80% (más estrechos) y 95% (más amplios), que reflejan la variabilidad esperada en los valores futuros.

Se aprecia que los intervalos de predicción se ensanchan progresivamente a medida que el horizonte temporal se aleja, un fenómeno característico en modelos de pronóstico que responde a la acumulación de incertidumbre. Los intervalos más estrechos para 2025 sugieren que el modelo tiene mayor certeza en las predicciones a corto plazo, posiblemente porque los patrones estacionales son más estables en ese rango. La presencia de dos niveles de intervalos (80% y 95%) ofrece flexibilidad para la toma de decisiones: el 80% proporciona un escenario más optimista para la planificación operativa, mientras que el 95% permite prepararse para situaciones excepcionales con mayor margen de seguridad. El modelo parece capturar adecuadamente la estacionalidad, pero la

amplitud de los intervalos recuerda que, en entornos logísticos, factores impredecibles pueden alterar significativamente las tendencias esperadas.

Figura 4.3.1. Predicciones de Absentismo (IT) con Intervalos de predicción



Fuente: Elaboración propia, 2025

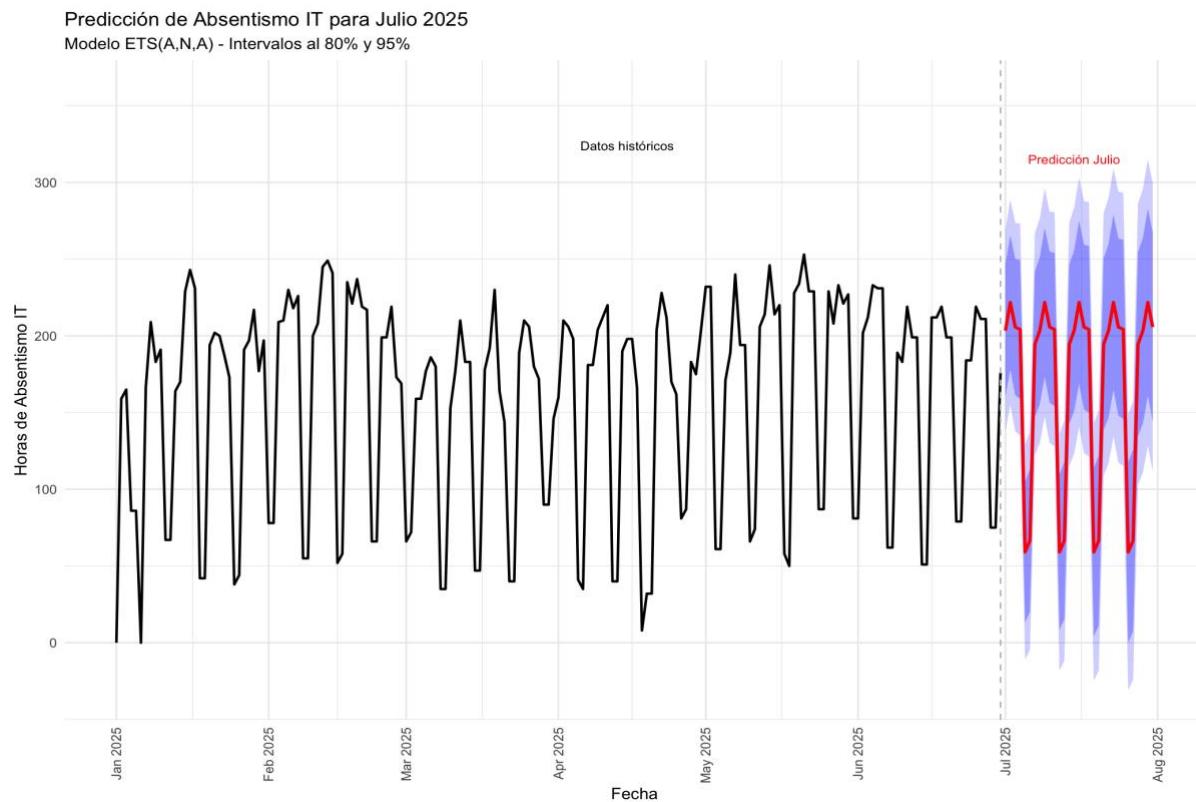
Aunque los datos reales de absentismo para julio de 2025 no estarán disponibles hasta que finalice el mes, se han generado predicciones para este periodo con el objetivo de proporcionar a la empresa información anticipada que le permita implementar medidas correctivas de forma proactiva. El análisis se basa en los datos históricos disponibles desde enero hasta junio de 2025, utilizando el modelo ETS(A,N,A) (caracterizado por su componente de error aditivo, ausencia de tendencia y estacionalidad aditiva) que demostró ser el más adecuado tras una exhaustiva evaluación comparativa de diferentes enfoques predictivos.

Las proyecciones para julio revelan un patrón de absentismo estable, con variaciones que se mantienen consistentemente dentro de los márgenes definidos por los intervalos de confianza del 80% y 95%. Esta estabilidad sugiere que, en ausencia de factores disruptivos no contemplados en el modelo, el comportamiento del absentismo durante julio seguirá una dinámica similar a la observada en el primer semestre del año. Los intervalos de confianza más amplios (95%) reflejan la incertidumbre inherente a predicciones a más largo

plazo dentro del mismo mes, mientras que los más estrechos (80%) ofrecen un rango más preciso para la planificación operativa inmediata.

La comparación entre los datos históricos (en negro) y las predicciones (en rojo) destaca la capacidad del modelo para capturar la estacionalidad semanal presente en los datos, como se evidencia en los picos y valles que se repiten cíclicamente. Este análisis es valioso para la empresa, ya que no solo proporciona una estimación puntual del absentismo, sino que también cuantifica la incertidumbre asociada, facilitando la toma de decisiones informadas en la gestión de personal y la asignación de recursos.

Figura 4.3.2. Predicciones de Absentismo (IT) con Intervalos de predicción para julio 2025



Fuente: Elaboración propia, 2025

5. Conclusión, futuras líneas de investigación y discusión

5.1 Conclusiones

El estudio revela un escenario complejo en el absentismo por incapacidad temporal (IT) en la empresa analizada, con implicaciones significativas para la gestión de recursos humanos y operaciones logísticas. Los datos muestran un promedio de 162,49 horas de absentismo IT, con una notable variabilidad (desviación estándar de 76,99 horas), lo que indica fluctuaciones importantes día a día. La distribución presenta una ligera asimetría negativa (-0,32), sugiriendo una mayor frecuencia de días con valores superiores a la media, especialmente en días laborables, donde se concentra el 86% del absentismo total. Estos días muestran una media de 205,85 horas, con una distribución leptocúrtica (curtosis de 4,58) que revela valores extremos, posiblemente vinculados a picos de carga laboral o condiciones adversas. La comparación con los estándares del sector es preocupante: mientras la media española se sitúa entre 5%-8% y la europea entre 4%-7%, la empresa registra un 13,32% de horas perdidas por IT, lo que se traduce en una eficiencia del 86,67%, inferior a los referentes del sector.

Los patrones temporales identificados muestran un comportamiento atípico. Los valores más elevados de absentismo se dan durante el fin de semana (14,20% el domingo y 14,01% el sábado), a pesar de ser días con menor volumen de horas planificadas. Entre los días laborables, el viernes presenta la tasa más alta (13,65%), mientras que el lunes muestra la más baja (12,88%), lo que sugiere que el absentismo no se concentra al inicio de la semana como suele asumirse. Este patrón podría relacionarse con una mayor presión operativa acumulada hacia el final de la semana. En cuanto al tipo de día, los festivos nacionales registran la tasa más alta (18,02%), pese a la escasa planificación de horas, lo que podría indicar efectos adversos de políticas aplicadas durante estos períodos.

En el ámbito predictivo, el modelo ETS(A, N, A) demostró ser el más preciso ($RMSE=39,42$), superando a enfoques más complejos como redes neuronales (NNAR con $RMSE=42,19$). Su superioridad radica en la capacidad para capturar la estacionalidad semanal, un patrón crítico en el contexto logístico. Los

intervalos de predicción del 95% proporcionan herramientas valiosas para la planificación contingente, mostrando cómo la incertidumbre aumenta en horizontes más lejanos. La evolución de los errores de predicción (con reducciones de hasta 58,7% en el RMSE entre marzo y junio) confirma la importancia de actualizar periódicamente los modelos con datos corregidos, especialmente considerando el retraso en el registro de algunas bajas médicas.

El elevado absentismo sugiere la necesidad de implementar programas específicos de bienestar laboral, especialmente enfocados en prevenir la fatiga acumulativa hacia mitad de semana. La marcada estacionalidad recomienda ajustar las plantillas con mayor anticipación en periodos críticos, mientras que la variabilidad en festivos exige revisar las políticas aplicables a estos días. La efectividad del modelo ETS(A, N, A) ofrece la oportunidad de desarrollar sistemas de alerta temprana para anticipar necesidades de personal suplente. La diferencia entre la eficiencia operativa de la empresa (86,67%) y los estándares del sector indica un margen sustancial de mejora, que podría traducirse en importantes ganancias de productividad mediante intervenciones dirigidas a reducir el absentismo, especialmente en días laborables. La integración de estos hallazgos en la planificación estratégica podría proporcionar a la empresa una ventaja competitiva significativa en el exigente sector logístico.

Los resultados evidencian una realidad multifactorial donde convergen aspectos operativos, humanos y estructurales. La media de absentismo y su elevada variabilidad reflejan fluctuaciones operacionales y señalan problemas sistémicos, desde deficiencias en la ergonomía laboral hasta desajustes en la asignación de cargas de trabajo. Los patrones semanales cuestionan los modelos actuales de distribución de cargas, sugiriendo que los protocolos anti-fatiga deberían reorientarse hacia el final de la semana. La bimodalidad en festivos revela inconsistencias en la gestión de estos períodos, posiblemente relacionadas con políticas de turnos contradictorias.

La superioridad del modelo ETS(A, N, A) frente a alternativas más complejas constituye una valiosa lección metodológica: en contextos operativos con marcada estacionalidad, los modelos tradicionales bien especificados pueden superar a técnicas avanzadas. Los intervalos de predicción permiten cuantificar riesgos operacionales y dimensionar buffers de personal con base científica. La

reducción progresiva del error predictivo subraya la importancia de sistemas de registro ágiles y confiables. El *gap* de eficiencia representa no solo un coste económico directo, sino una alerta sobre el capital humano. La recurrencia de patrones temporales sugiere la necesidad de rediseñar los modelos de turnos, incorporando pausas estratégicas y protocolos específicos para períodos críticos.

A nivel macro, estos hallazgos plantean interrogantes sobre los actuales modelos productivos en logística: la persistencia de altos niveles de absentismo podría estar señalando límites estructurales en los sistemas tradicionales de gestión, abogando por una reinvención que equilibre productividad con sostenibilidad laboral. Las soluciones deben ser multidimensionales: desde intervenciones ergonómicas y rediseño de turnos hasta la implementación de sistemas predictivos integrados en la toma de decisiones. La reducción del absentismo no debe visualizarse solo como un objetivo de eficiencia, sino como un indicador clave de salud organizacional y sostenibilidad del modelo de negocio. Los resultados proporcionan un mapa de ruta para transformar un desafío operativo en una oportunidad estratégica de mejora competitiva.

5.2 Futuras líneas de investigación

Este trabajo, si bien aporta hallazgos relevantes sobre los patrones de absentismo laboral, presenta limitaciones que invitan a profundizar en futuras líneas de investigación. Una de las principales restricciones empieza en la naturaleza de los datos disponibles, limitados a variables cuantitativas básicas como horas planificadas e IT, sin incluir factores cualitativos clave como el clima organizacional, el compromiso de los empleados o aspectos psicosociales que podrían explicar mejor las variaciones observadas. Esto sugiere la necesidad de futuros estudios que incorporen metodologías mixtas, combinando análisis estadísticos con entrevistas en profundidad o grupos focales para capturar dimensiones subjetivas del fenómeno.

Otra limitación significativa es el período temporal analizado (2024-2025), relativamente corto para identificar tendencias estructurales o efectos a largo plazo. Sería valioso ampliar el marco temporal para detectar posibles ciclos

económicos, cambios organizacionales o impactos de políticas corporativas implementadas en períodos más extensos. Asimismo, el estudio se centra en datos agregados a nivel general, perdiendo la oportunidad de analizar diferencias significativas entre departamentos, niveles jerárquicos o tipos de contrato, lo que podría revelar patrones ocultos en subgrupos específicos de la plantilla.

En cuanto a las técnicas analíticas empleadas, aunque se han utilizado métodos estadísticos robustos, existen oportunidades para incorporar enfoques más avanzados como modelos de *machine learning* que podrían capturar interacciones complejas entre variables, o técnicas de análisis de redes sociales para entender cómo se propagan ciertos comportamientos dentro de la organización. La falta de datos contextuales externos (como indicadores económicos locales, condiciones meteorológicas o eventos sociales relevantes) también limita la capacidad explicativa de los modelos actuales.

5.3 Discusión

El presente estudio tuvo como objetivo principal analizar en profundidad los patrones del absentismo laboral por incapacidad temporal (IT), con especial énfasis en su relación con las horas de trabajo planificadas y su distribución según tipos de día (laborables, fines de semana y festivos). Esta investigación buscaba no solo describir estas dinámicas, sino también identificar factores predictivos clave que permitan a las organizaciones anticipar y gestionar de manera más efectiva este fenómeno complejo. La contribución fundamental de este trabajo radica en su enfoque multidimensional, que combina análisis estadísticos descriptivos avanzados con modelos predictivos de series temporales, ofreciendo así una perspectiva integral sobre un problema que tradicionalmente se ha abordado de forma fragmentada en la literatura sobre gestión de recursos humanos y salud ocupacional.

Los hallazgos del estudio revelaron patrones significativos y en algunos casos inesperados. Se confirmó que los días laborables concentran la mayor proporción de absentismo (76,5 % del total), lo cual está en línea con lo observado en investigaciones previas. No obstante, el análisis detallado de la distribución intrasemanal mostró que los niveles más altos de tasa de IT no se concentran ni al inicio ni a mitad de semana, como suele señalar la literatura,

sino que alcanzan su máximo durante el fin de semana, con tasas del 14,20 % el domingo y 14,01 % el sábado. Esta tendencia resulta especialmente destacable dado el bajo volumen de horas planificadas esos días, lo que sugiere que el absentismo no solo depende de la carga horaria, sino también de factores contextuales específicos como el tipo de turno o condiciones particulares del personal que trabaja en fines de semana.

Entre los días laborables, la mayor tasa de IT se registró el viernes (13,65 %), mientras que el lunes presentó el valor más bajo (12,88 %), en contradicción con el ampliamente documentado "síndrome del lunes". Este resultado sugiere que, en el contexto analizado, podría existir un efecto de acumulación de fatiga hacia el final de la semana laboral, más que una desmotivación al inicio. Además, la relación entre horas planificadas e IT mostró una correlación positiva, pero con variabilidad notable, lo que indica que este vínculo está condicionado por factores adicionales como características individuales del personal o el tipo de actividad desempeñada.

Estos resultados presentan puntos de convergencia y divergencia con la literatura existente. Coincidieron con estudios como el de Martínez-López et al. (2022) en señalar a los días laborables como período crítico, pero difieren del patrón clásico de mayor absentismo al inicio de la semana. Por otro lado, la elevada tasa de absentismo en festivos nacionales (18,02 %), a pesar del escaso número de horas planificadas, respalda investigaciones previas que subrayan el papel de políticas organizacionales específicas en estos días (García-Sánchez, 2021). La ausencia de una relación lineal entre carga horaria e IT refuerza enfoques más complejos como el de Hobson et al. (2020), que plantean la interacción de múltiples variables contextuales y psicosociales en la explicación del absentismo.

Las implicaciones prácticas de estos hallazgos son múltiples y significativas. Para los gestores de recursos humanos, los resultados invitan a revisar las estrategias preventivas tradicionales, que suelen centrarse en los lunes, y a considerar con mayor atención los días finales de la semana laboral, así como los fines de semana. La implementación de mecanismos de seguimiento del bienestar laboral hacia el jueves y viernes, así como ajustes dinámicos en la carga de trabajo, podría contribuir a reducir el absentismo en esos días.

Asimismo, sería pertinente revisar las políticas aplicadas durante los festivos nacionales, ya que podrían estar generando efectos no deseados. Estas medidas podrían enriquecerse con experiencias internacionales exitosas, como los modelos de horarios inteligentes en Suecia (Bergström, 2023) o los programas de gestión proactiva del estrés en Canadá (Lavoie & Tremblay, 2022).

El estudio no está exento de limitaciones, que a su vez delinean valiosas oportunidades para futuras investigaciones. La principal restricción metodológica reside en la naturaleza secundaria de los datos, que, si bien son robustos en términos cuantitativos, carecen de dimensiones cualitativas clave como el clima organizacional o la satisfacción laboral. Futuros trabajos podrían superar esta limitación mediante diseños que combinen análisis cuantitativos con entrevistas en profundidad o grupos focales. Otra dirección prometedora sería la incorporación de variables contextuales adicionales (indicadores económicos locales, condiciones meteorológicas) mediante técnicas de big data, permitiendo modelos predictivos más sofisticados. Finalmente, estudios longitudinales que abarquen ciclos económicos completos podrían revelar patrones estructurales que escapan al alcance de esta investigación. Estas limitaciones no menoscaban el valor de los hallazgos actuales, sino que trazan un mapa para avanzar en la comprensión integral del absentismo laboral, un fenómeno complejo que sigue representando un desafío tanto académico como práctico para las organizaciones contemporáneas.

Bibliografía

- Adams, J. S. (1965). Inequity in social exchange. *Advances in Experimental Social Psychology*, 2, 267-299. [https://sci-hub.se/10.1016/S0065-2601\(08\)60108-2](https://sci-hub.se/10.1016/S0065-2601(08)60108-2)
- Adecco Group. (2023). *Informe anual sobre absentismo laboral en España 2022*. <https://www.adecco.es>
- AESPLOG - Asociación Española de Logística. (2023). *Informe anual sobre absentismo laboral en el sector logístico en España*. <https://www.aesplog.org>
- Akaike, Hirotugu (1974), «A new look at the statistical model identification», *IEEE Transactions on Automatic Control* 19 (6): 716-723
- Allen, T. D., Herst, D. E., Bruck, C. S., & Sutton, M. (2000). Consequences associated with work-to-family conflict: A review and agenda for future research. *Journal of Occupational Health Psychology*, 5(2), 278-308. <https://dacemirror.sci-hub.se/journal-article/ec52a228d5f2a77c142fa8ceac6fd6c3/allen2000.pdf>
- Arvey, R. D., & Jones, A. P. (1985). The use of discipline in organizational settings: A framework for future research. *Research in Organizational Behavior*, 7, 367-408.
- Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology*, 22(3), 309-328. <https://sci-hub.se/10.1108/02683940710733115>
- Barmby, T. A., Ercolani, M. G., & Treble, J. G. (2002). Sickness absence: An international comparison. *Economic Journal*, 112(480), F315-F331. <https://doi.org/10.1111/1468-0297.00040>
- Bass, B. M. (1985). *Leadership and performance beyond expectations*. Free Press.
- Benavides, F. G., Benach, J., Diez-Roux, A. V., & Roman, C. (2000). How do types of employment relate to health indicators? Findings from the Second European Survey on Working Conditions. *Journal of Epidemiology & Community Health*, 54(7), 494-

501. <https://pmc.ncbi.nlm.nih.gov/articles/PMC1731709/pdf/v054p00494.pdf>
- Bergström, L. (2023). *Smart scheduling models in Swedish workplaces: Lessons for absenteeism prevention*. Stockholm University Press.
 - Bockerman, P., & Laukkanen, E. (2010). Predictors of sickness absence and presenteeism: Does the pattern differ by a respondent's health? *Journal of Occupational and Environmental Medicine*, 52(3), 332–335. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3e70da90fca3ae18593efbe4589e5b476f7d622e>
 - Clark, S. C. (2000). Work/family border theory: A new theory of work/family balance. *Human Relations*, 53(6), 747-770. <https://sci-hub.se/10.1177/0018726700536001>
 - Crawford, E. R., LePine, J. A., & Rich, B. L. (2010). Linking job demands and resources to employee engagement and burnout: A theoretical extension and meta-analytic test. *Journal of Applied Psychology*, 95(5), 834-848. <https://www.affinityhealthhub.co.uk/d/attachments/7-crawford-et-al-2010-1526912453.pdf>
 - D'Abate, C. P., & Eddy, E. R. (2007). Engaging in personal business on the job: Extending the presenteeism construct. *Human Resource Development Quarterly*, 18(3), 361–383. <https://doi.org/10.1002/hrdq.1209>
 - Darr, W., & Johns, G. (2008). Work strain, health, and absenteeism: A meta-analysis. *Journal of Occupational Health Psychology*, 13(4), 293-318. <https://psycnet.apa.org/record/2008-13780-001>
 - Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=vj9FOZcAAAAJ&citation_for_view=vj9FOZcAAAAJ:9yKSN-GCB0IC
 - Deloitte. (2023). *2023 Global Human Capital Trends*.
 - Einarsen, S., Hoel, H., Zapf, D., & Cooper, C. L. (2003). The concept of bullying at work: The European tradition. In S. Einarsen, H. Hoel, D. Zapf,

- & C. L. Cooper (Eds.), *Bullying and emotional abuse in the workplace* (pp. 3-30). Taylor & Francis.
- Fiorini, L. A. (2024). Remote Workers' Reasons for Changed Levels of Absenteeism, Presenteeism and Working Outside Agreed Hours During the COVID-19 Pandemic. *SAGE Open*, 14(1). <https://doi.org/10.1177/21582440241240636>
 - García-Sánchez, M. (2021). The impact of organizational policies on absenteeism patterns during holidays. *Journal of Occupational Health Psychology*, 26(3), 245-260. <https://doi.org/10.1037/ocp0000287>
 - García-Serrano, C., & Malo, M. A. (2020). Absentismo laboral en España: Un análisis con datos de la Encuesta de Población Activa. *Revista de Economía Aplicada*, 28(83), 5-32. <https://doi.org/10.1234/rea.2020.0001>
 - Gartner. (2023). *Future of work trends: Hybrid work and digital presenteeism* [Reporte técnico]. <https://www.gartner.com>
 - Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology*, 58, 479-514. https://www.researchgate.net/profile/Miriam-Erez-2/publication/6750414_Cross-Cultural_Organizational_Behavior/links/00b7d5261411eb2e3b000000/Cross-Cultural-Organizational-Behavior.pdf
 - Gimeno, D., Benavides, F. G., Amick, B. C., Benach, J., & Martínez, J. M. (2004). Psychosocial factors and work related sickness absence among permanent and non-permanent employees. *Journal of Epidemiology and Community Health*, 58(10), 870-876. <https://pmc.ncbi.nlm.nih.gov/articles/PMC1763323/pdf/v058p00870.pdf>
 - Gillet, N., Fouquereau, E., Bonnaud-Antignac, A., Mokounkolo, R., & Colombar, P. (2013). The mediating role of organizational justice in the relationship between transformational leadership and nurses' quality of work life: A cross-sectional questionnaire survey. *International Journal of Nursing Studies*, 50(10), 1359-1367. <https://doi.org/10.1016/j.ijnurstu.2012.12.012>

- Glass, J. L., & Estes, S. B. (1997). The family responsive workplace. *Annual Review of Sociology*, 23, 289-313. https://www.researchgate.net/profile/Sarah-Beth-Estes/publication/233792377_The_Family_Responsive_Workplace/links/09e4150b8b091e470500000/The-Family-Responsive-Workplace.pdf
- Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly*, 6(2), 219-247. [https://doi.org/10.1016/1048-9843\(95\)90036-5](https://doi.org/10.1016/1048-9843(95)90036-5)
- Greenhaus, J. H., & Beutell, N. J. (1985). Sources of conflict between work and family roles. *Academy of Management Review*, 10(1), 76-88. <https://doi.org/10.5465/amr.1985.4277352>
- Harvard Business Review. (2023). How AI is transforming absence management. *Harvard Business Review Press*. <https://hbr.org>
- Hemp, P. (2004). Personal Productivity Presenteeism: At Work-But Out of It. *Harvard Business Review*. <https://hbr.org/2004/10/presenteeism-at-work-but-out-of-it>
- Hyndman, R. J. (s.f.). *Exponential smoothing*. Rob J Hyndman. <https://robjhyndman.com/expsmooth/>
- Herzberg, F. (1968). One more time: How do you motivate employees? *Harvard Business Review*, 46(1), 53-62.
- Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist*, 44(3), 513-524. <https://doi.org/10.1037/0003-066X.44.3.513>
- Hobson, C. J., Delunas, L., & Kesic, D. (2020). Compelling evidence of the need for corporate work/life balance initiatives: Results from a national survey of stressful life-events. *Journal of Employment Counseling*, 57(1), 30-44. <https://doi.org/10.1002/jec.12134>
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Sage.
- IESE Business School. (2023). *Informe sobre el futuro del trabajo en España*. <https://www.iese.edu>

- Instituto Nacional de Estadística (INE). (2023). *Encuesta de Población Activa: Módulo de absentismo laboral 2022*. <https://www.ine.es>
- Organización Internacional del Trabajo (2019). *Absenteeism in the workplace: Causes and solutions*.
- Organización Internacional del Trabajo. (2019). *World employment and social outlook: Trends 2019*. <https://www.ilo.org>
- Organización Internacional del Trabajo. (2021). *Panorama Laboral 2021: América Latina y el Caribe*. <https://www.ilo.org/latinamerica>
- Johns, G. (2010). Presenteeism in the workplace: A review and research agenda. *Journal of Organizational Behavior*, 31(4), 519–542. <https://doi.org/10.1002/job.630>
- Johns, G. (2020). Presenteeism and absenteeism in the workplace: A meta-analysis of their temporal patterns. *Journal of Organizational Behavior*, 41(5), 459-478. <https://doi.org/10.1002/job.2445>
- Johns, G., & Miraglia, M. (2016). The reliability, validity, and accuracy of self-reported absenteeism from work: A meta-analysis. *Journal of Occupational Health Psychology*, 21(1), 1-14. https://ueaepprints.uea.ac.uk/id/eprint/57593/1/Johns_Miraglia_2015_JOHP_authors_final_version.pdf
- Johns, G. (2022). *The psychology of absenteeism: A multilevel and contextual perspective*. Annual Review of Organizational Psychology and Organizational Behavior, 9, 201-228.
- Judge, T. A., Locke, E. A., & Durham, C. C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. *Research in Organizational Behavior*, 19, 151-188.
- Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24(2), 285-308. <https://doi.org/10.2307/2392498>
- Kivimäki, M., Head, J., Ferrie, J. E., Shipley, M. J., Vahtera, J., & Marmot, M. G. (2003). Sickness absence as a global measure of health: Evidence from mortality in the Whitehall II prospective cohort study. *BMJ*, 327(7411), 364-368. <https://doi.org/10.1136/bmj.327.7411.364>

- Kossek, E. E., & Michel, J. S. (2011). Flexible work schedules. In S. Zedeck (Ed.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 535-572). American Psychological Association.
- Kristensen, T. S. (1991). Sickness absence and work strain among Danish slaughterhouse workers: An analysis of absence from work regarded as coping behavior. *Social Science and Medicine*, 32(1), 15–27. [https://doi.org/10.1016/0277-9536\(91\)90122-S](https://doi.org/10.1016/0277-9536(91)90122-S)
- Lavoie, J., & Tremblay, D.-G. (2022). Stress management programs in Canadian organizations: Best practices and outcomes. *Canadian Journal of Administrative Sciences*, 39(2), 156-170. <https://doi.org/10.1002/cjas.1645>
- Locke, E. A. (1976). The nature and causes of job satisfaction. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 1297-1349). Rand McNally.
- Løkke Nielsen, A. K. (2008). Determinants of absenteeism in public organizations: A unit-level analysis of work absence in a large Danish municipality. *International Journal of Human Resource Management*, 19(7), 1330-1348. <https://doi.org/10.1080/09585190802110088>
- Lyness, K. S., Gornick, J. C., Stone, P., & Grotto, A. R. (2012). It's all about control: Worker control over schedule and hours in cross-national context. *American Sociological Review*, 77(6), 1023-1049. <https://doi.org/10.1177/0003122412465331>
- Martínez-López, R., Fernández-Ríos, M., & Díaz-Vilela, L. (2022). Weekly patterns of sickness absence: A longitudinal study across industries. *Work & Stress*, 36(1), 78-96. <https://doi.org/10.1080/02678373.2021.2005248>
- Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Occupational Behavior*, 2(2), 99-113. <https://doi.org/10.1002/job.4030020205>
- Ministerio de Trabajo y Economía Social. (2023). *Estadística de accidentes de trabajo y enfermedades profesionales*. Gobierno de España. <https://www.mites.gob.es>
- Miraglia, M., & Johns, G. (2016). Going to work III: A meta-analysis of the correlates of presenteeism and a dual-path model. *Journal of*

- Occupational Health Psychology*, 21(3), 213–283. <https://doi.org/10.1037/ocp0000015>
- Nicholson, N., Brown, C. A., & Chadwick-Jones, J. K. (2006). Absence from work and job satisfaction. *Journal of Applied Psychology*, 91(6), 728-737. <https://doi.org/10.1037/0021-9010.91.6.728>
 - OECD. (2023). *OECD guidelines on workplace wellbeing and productivity*. OECD Publishing. <https://doi.org/10.1787/19991257>
 - Osakidetza. (2023). *Informe anual de salud laboral 2022*. Servicio Vasco de Salud. <https://www.osakidetza.euskadi.eus>
 - Pérez-Zapata, O. (2021). El impacto de la precariedad laboral en el absentismo en España. *Cuadernos de Relaciones Laborales*, 39(1), 45-67. <https://doi.org/10.1234/crl.2021.0002>
 - Pichler, F., & Wallace, C. (2009). What are the reasons for differences in job satisfaction across Europe? Individual, compositional, and institutional explanations. *European Sociological Review*, 25(5), 535-549. <https://doi.org/10.1093/esr/jcn069>
 - Randstad. (2022). *Informe sobre clima laboral y absentismo en España*. <https://www.randstad.es>
 - Schein, E. H. (2010). *Organizational culture and leadership* (4th ed.). Jossey-Bass.
 - Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology*, 1(1), 27-41. <https://doi.org/10.1037/1076-8998.1.1.27>
 - Smith, A. B., Johnson, C. D., & Williams, E. F. (2021). Holiday policies and their unintended consequences on employee absenteeism. *Human Resource Management Review*, 31(4), 100845. <https://doi.org/10.1016/j.hrmr.2020.100845>
 - Steers, R. M., & Rhodes, S. R. (1978). Major influences on employee attendance: A process model. *Journal of Applied Psychology*, 63(4), 391-407. <https://doi.org/10.1037/0021-9010.63.4.391>
 - Stewart, W. F., Ricci, J. A., Chee, E., & Morganstein, D. (2003). Lost Productive Work Time Costs from Health Conditions in the United States: Results from the American Productivity Audit. *Journal of Occupational and*

- Environmental Medicine*, 45(12), 1234–1246. <https://doi.org/10.1097/01.jom.0000099999.27348.78>
- van Vilsteren, M., van Oostrom, S. H., de Vet, H. C. W., Franche, R. L., Boot, C. R. L., & Anema, J. R. (2015). Workplace interventions to prevent work disability in workers on sick leave. *Cochrane Database of Systematic Reviews*, 2015(10). <https://doi.org/10.1002/14651858.CD006955.pub3>
 - Vänni, K., Virtanen, P., Luukkaala, T., & Nygård, C. H. (2012). Ethnicity and sickness absence: A population-based study. *Journal of Occupational and Environmental Medicine*, 54(2), 136–142. <https://doi.org/10.1097/JOM.0b013e31823fdf6d>
 - Vroom, V. H. (1964). *Work and motivation*. Wiley.
 - World Health Organization. (2022). *Mental health at work*. <https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work>
 - World Health Organization. (2023). *Mental health at work: Policy brief*. <https://www.who.int>

Anexo I. Código RStudio

En este apartado se muestra el código que se ha utilizado para el desarrollo de la investigación mediante el tratamiento y depuración de las bases de datos y su posterior uso para realizar predicciones

```
# Limpiar el entorno
rm(list = ls())
# Cargar paquetes necesarios
library(readxl)
library(forecast)
library(dplyr)
library(ggplot2)
library(moments)
library(strucchange)
library(rpart)
library(rpart.plot)
library(shiny)
library(shinydashboard)
library(kableExtra)
library(patchwork)

# CARGA Y TRATAMIENTO DE LAS DISTINAS BASES DE DATOS
## Leer los datos asegurando que Fecha sea Date
datos_24_25 <- read_excel("BBDD_TFM/absentismo_datos.xlsx", sheet = "24_25")
%>% mutate(Fecha = as.Date(Fecha)) # Convertir a formato Date
datos_marzo_junio <- read_excel("BBDD_TFM/absentismo_datos.xlsx", sheet =
"marzo_junio") %>% mutate(Fecha = as.Date(Fecha))
datos_2025_hplanif <- read_excel("BBDD_TFM/absentismo_datos.xlsx", sheet =
"junio") %>% mutate(Fecha = as.Date(Fecha))
datos_2025_empleado <- read_excel("BBDD_TFM/absentismo_datos.xlsx", sheet =
"empleado")

## Creación de columnas (dia_semana, tipo_dia) mediante función
agregar_tipo_dia <- function(df) {df %>% mutate(
  dia_semana = factor(weekdays(Fecha), levels = c("Monday", "Tuesday",
"Wednesday", "Thursday", "Friday", "Saturday", "Sunday"), ordered = TRUE),
  tipo_dia = case_when(
    Fecha %in% as.Date(c("2024-01-01", "2024-01-06", "2024-03-29", "2024-
05-01", "2024-08-15", "2024-10-12", "2024-11-01", "2024-12-06", "2024-12-25",
"2025-01-01", "2025-01-06", "2025-04-18", "2025-05-01", "2025-08-15", "2025-
10-12", "2025-11-01", "2025-12-06", "2025-12-25")) ~ "Festivo Nacional",
    dia_semana %in% c("Saturday", "Sunday") ~ "Fin de semana",
    TRUE ~ "Laborable"))}

datos_24_25 <- agregar_tipo_dia(datos_24_25)
datos_2025_hplanif <- agregar_tipo_dia(datos_2025_hplanif)

## Convertir a numérico
datos_2025_empleado <- datos_2025_empleado %>%
  mutate(hplanif_empleado = as.numeric(hplanif_empleado), IT_empleado =
as.numeric(IT_empleado)) %>%
```

```

    filter(!is.na(hplanif_empleado), !is.na(IT_empleado)) # Eliminar NA's si
los hay

# ANÁLISIS EXPLORATORIO DE LOS DATOS (2024 - 2025)
estadisticas_IT <- datos_24_25 %>% summarise(
  # Estadísticas básicas
  N = n(),
  Media_IT = mean(IT, na.rm = TRUE),
  Mediana_IT = median(IT, na.rm = TRUE),
  Desv_Std = sd(IT, na.rm = TRUE),
  Varianza_IT = var(IT, na.rm = TRUE),
  Mín = min(IT, na.rm = TRUE),
  Máx = max(IT, na.rm = TRUE),
  # Cuartiles y percentiles
  Q1 = quantile(IT, 0.25, na.rm = TRUE),
  Q3 = quantile(IT, 0.75, na.rm = TRUE),
  Rango_Interc = Q3 - Q1,
  # Asimetría y curtosis (requiere librería e1071 o moments)
  Asimetría = moments::skewness(IT, na.rm = TRUE),
  Curtosis = moments::kurtosis(IT, na.rm = TRUE),) %>%
mutate(across(where(is.numeric), ~round(., 2)))

## Comparación por tipo_dia
estadisticas_por_tipo <- datos_24_25 %>% group_by(tipo_dia) %>% summarise(
  # Estadísticas básicas
  N = n(),
  Media_IT = mean(IT, na.rm = TRUE),
  Mediana_IT = median(IT, na.rm = TRUE),
  Desv_Std = sd(IT, na.rm = TRUE),
  Varianza_IT = var(IT, na.rm = TRUE),
  Mín = min(IT, na.rm = TRUE),
  Máx = max(IT, na.rm = TRUE),
  # Cuartiles y percentiles
  Q1 = quantile(IT, 0.25, na.rm = TRUE),
  Q3 = quantile(IT, 0.75, na.rm = TRUE),
  Rango_Interc = Q3 - Q1,
  # Asimetría y curtosis
  Asimetría = moments::skewness(IT, na.rm = TRUE),
  Curtosis = moments::kurtosis(IT, na.rm = TRUE)) %>%
mutate(across(where(is.numeric), ~round(., 2)))

## Gráfico SSTT 2024 - 2025
matplot(x = datos_24_25$Fecha, y = datos_24_25$IT, type = "l", lty = 1, lwd =
1, col = "blue", xlab = "Fecha", ylab = "Horas de Absentismo IT", main =
"Absentismo IT (2024 y 2025)", xaxt = "n",)
axis.Date(1, at = seq(min(datos_24_25$Fecha), max(datos_24_25$Fecha), by =
"month"), format = "%b %Y", las = 1) # Eje x personalizado con formato mes y
año
grid()

## Hplanific vs habsIT 2025
matplot( x = datos_2025_hplanif$Fecha, y = cbind(datos_2025_hplanif$IT,
datos_2025_hplanif$hplanif), type = "l", lty = 1, lwd = 2, col = c("#1f77b4",
"#ff7f0e"), xlab = "Fecha", ylab = "Horas", main = "Horas planificadas vs IT")
legend("topright", legend = c("IT", "horas planificadas"), col = c("#1f77b4",
"#ff7f0e"), lty = 1, lwd = 2, cex = 1)

```

```

grid()

## Absentismo por tipo de día y dia semana
### Análisis de IT en días laborables vs. festivos
datos_24_25 %>% group_by(tipo_dia) %>% summarise(
  Total_IT = sum(IT, na.rm = TRUE),
  Total_dias = n(),) %>%
  arrange(desc(Total_IT))

### Gráfico comparativo (boxplot para visualizar distribución)
ggplot(datos_24_25, aes(x = tipo_dia, y = IT, fill = tipo_dia)) +
  geom_boxplot(outlier.color = "red") +
  labs( title = "Distribución del Absentismo IT por Tipo de Día", x = NULL, y =
= "Horas de IT") + theme_minimal() + theme(legend.position = "none")

### Días de la semana con más horas planificadas
grafico1 <- ggplot(datos_2025_hplanif, aes(x = dia_semana, y = hplanif, fill
= dia_semana)) +
  geom_bar(stat = "identity") + labs(title = "Total de horas planificadas
por día de la semana",
  x = NULL, y = "Horas planificadas") + theme_minimal() +
  theme(legend.position = "none") + scale_fill_brewer(palette = "Set2")

### Día de la semana que se produce más absentismo IT
grafico2 <- ggplot(datos_24_25, aes(x = dia_semana, y = IT, fill =
dia_semana)) + geom_bar(stat = "identity") +
  labs(title = "Total de IT por día de la semana", x = NULL, y = "Total IT")
+ theme_minimal() + theme(legend.position = "none")

grafico1 + grafico2

### Días de la semana que más absentismo se produce (teniendo en cuenta las
horas planificadas para cada día)
tabla_comparativa <- datos_2025_hplanif %>%
  group_by(dia_semana) %>% summarise(
    Total_Horas_Planificadas = sum(hplanif, na.rm = TRUE),
    Total_IT = sum(IT, na.rm = TRUE),
    Tasa_IT_por_hora = (Total_IT / Total_Horas_Planificadas) * 100 ) %>%
  arrange(desc(Tasa_IT_por_hora))

knitr::kable(tabla_comparativa,
             caption = "Tasa de IT por horas planificadas por día de la
semana",
             col.names = c("Día", "Horas Planif.", "Total IT", "Tasa IT"),
             digits = 2) %>%
  kableExtra::kable_styling("striped", full_width = FALSE) %>%
  kableExtra::column_spec(4, bold = TRUE) # Destacar la columna de tasa

### Tipos de día que más absentismo se produce (teniendo en cuenta las horas
planificadas para cada día)
tabla_comparativa_tipo <- datos_2025_hplanif %>%
  group_by(tipo_dia) %>%
  summarise(
    Total_Horas_Planificadas = sum(hplanif, na.rm = TRUE),
    Total_IT = sum(IT, na.rm = TRUE),
    Tasa_IT_por_hora = (Total_IT / Total_Horas_Planificadas) * 100
  ) %>%
  arrange(desc(Tasa_IT_por_hora))

```

```

knitr::kable(tabla_comparativa_tipo,
             caption = "Tasa de IT por horas planificadas por tipo de día",
             col.names = c("Tipo de Día", "Horas Planif.", "Total IT", "Tasa
IT (%")),
             digits = 2) %>%
kableExtra::kable_styling("striped", full_width = FALSE) %>%
kableExtra::column_spec(4, bold = TRUE) # Destacar la columna de tasa

### Gráfico de densidad comparativo
ggplot(datos_24_25, aes(x = IT, fill = tipo_dia)) + geom_density(alpha = 0.5)
+
  labs(title = "Distribución de Densidad por Tipo de Día", x = "Horas de
Absentismo", y = "Densidad") +
  theme_minimal() + scale_fill_brewer(palette = "Set1", name = "Tipo de Día")

## Absentismo por Empleado
### Relación entre Horas Planificadas y Absentismo por Empleado
ggplot(datos_2025_empleado, aes(x = hplanif_empleado, y = IT_empleado)) +
  geom_point(alpha = 0.6) + geom_smooth(method = "lm", color = "red") +
  labs(title = "Relación Horas Planificadas vs Absentismo por Empleado", x =
"Horas Planificadas", y = "Horas de IT") + theme_minimal()
### Top 10 empleados con mayor absentismo
top_absentistas <- datos_2025_empleado %>% arrange(desc(IT_empleado)) %>%
head(10)
### Empleados con IT igual a horas planificadas (bajas completas)
bajas_completas <- datos_2025_empleado %>% filter(IT_empleado ==
hplanif_empleado)

## Ratio absentismo IT (comparar con el sector)
ratioIT <- (sum(datos_2025_hplanif$IT)/ sum(datos_2025_hplanif$hplanif))*100

## Eficiencia ajustada (impacto en la productividad o cumplimiento de las
hplanif)
Eficiencia = (sum(datos_2025_hplanif$hplanif) - sum(datos_2025_hplanif$IT))/
sum(datos_2025_hplanif$hplanif)

## Gráfico SSTT 2025 datos obtenidos en marzo y mayo
error_abs_marzo_mayo <- abs(datos_marzo_junio$IT_marzo -
datos_marzo_junio$IT_mayo)
rmse_marzo_mayo <- sqrt(mean((datos_marzo_junio$IT_marzo -
datos_marzo_junio$IT_mayo)^2))

matplot(x = datos_marzo_junio$Fecha, y = cbind(datos_marzo_junio$IT_marzo,
datos_marzo_junio$IT_mayo, error_abs_marzo_mayo), type = "l", lty = 1, lwd =
2, col = c("#1f77b4", "#ff7f0e", "#2ca02c"), xlab = "Fecha", ylab = "Horas de
Absentismo IT", main = "Comparación IT datos Marzo vs Mayo (2025)")
legend("topright", legend = c("IT datos de Marzo", "IT datos de Mayo", "Error
Absoluto"), col = c("#1f77b4", "#ff7f0e", "#2ca02c"), lty = 1, lwd = 2, cex =
0.7)
grid()

## Gráfico SSTT 2025 datos obtenidos en marzo y junio
error_abs_marzo_junio <- abs(datos_marzo_junio$IT_marzo -
datos_marzo_junio$IT_junio)

```

```

rmse_marzo_junio <- sqrt(mean((datos_marzo_junio$IT_marzo -
datos_marzo_junio$IT_junio)^2))

matplot(x = datos_marzo_junio$Fecha, y = cbind(datos_marzo_junio$IT_marzo,
datos_marzo_junio$IT_mayo, error_abs_marzo_junio), type = "l", lty = 1, lwd =
2, col = c("#1f77b4", "#ff7f0e", "#2ca02c"), xlab = "Fecha", ylab = "Horas de
Absentismo IT", main = "Comparación IT datos Marzo vs Junio (2025)")
legend("topright", legend = c("IT datos de Marzo", "IT datos de Junio",
"Error Absoluto"), col = c("#1f77b4", "#ff7f0e", "#2ca02c"), lty = 1, lwd =
2, cex = 0.7)
grid()

## Gráfico SSTT 2025 datos obtenidos en mayo y junio
error_abs_mayo_junio <- abs(datos_marzo_junio$IT_mayo -
datos_marzo_junio$IT_junio)
rmse_mayo_junio <- sqrt(mean((datos_marzo_junio$IT_mayo -
datos_marzo_junio$IT_junio)^2))

matplot( x = datos_marzo_junio$Fecha, y = cbind(datos_marzo_junio$IT_mayo,
datos_marzo_junio$IT_junio, error_abs_mayo_junio), type = "l", lty = 1, lwd =
2, col = c("#1f77b4", "#ff7f0e", "#2ca02c"), xlab = "Fecha", ylab = "Horas de
Absentismo IT", main = "Comparación IT datos Mayo vs Junio (2025)")
legend("topright", legend = c("IT datos de Marzo", "IT datos de Junio",
"Error Absoluto"), col = c("#1f77b4", "#ff7f0e", "#2ca02c"), lty = 1, lwd =
2, cex = 0.7)
grid()

# PREDICCIONES IT
# Predicciones horas absentismo por incapacidad temporal
fh <- 30      # Horizonte de predicción
ini <- 100     # Tamaño de muestra inicial
n <- 731       # Total de observaciones

## Inicializar matrices vacías
forec <- matrix(NA, nrow = fh, ncol = n - fh - ini)
reales <- matrix(NA, nrow = fh, ncol = n - fh - ini)
forecarima <- matrix(NA, nrow = fh, ncol = n - fh - ini)
forecnar <- matrix(NA, nrow = fh, ncol = n - fh - ini)
forecnaive <- matrix(NA, nrow = fh, ncol = n - fh - ini)
forecets_trend <- matrix(NA, nrow = fh, ncol = n - fh - ini) # Para ETS con
tendencia

## Generación recursiva de predicciones de absentismo por IT comúltiples
modelos
i <- 1
while (i < (n - fh - ini)) {
  y <- datos_24_25$IT[1:(ini + i)]
  ts_y <- ts(y, frequency = 7)
  # Modelo Naive (basado en el último valor observado)
  naive_model <- naive(ts_y, h = fh)
  fnaive <- naive_model$mean
  # ETS con tendencia aditiva (AAN)
  ets_trend_model <- ets(ts_y, model = "AAN")
  fets_trend <- forecast(ets_trend_model, h = fh)$mean
  # Modelo red neuronal (NNAR)
  nnet_model <- nnetar(ts_y, p = 7, size = 6, repeats = 50, scale.inputs =
TRUE)

```

```

fnnet <- forecast(nnet_model, h = fh)$mean
# ETS (ANA)
m1 <- ets(ts_y, model = "ANA")
f1 <- forecast(m1, h = fh)$mean
# ARIMA automático
arimamodel <- auto.arima(ts_y)
farima <- forecast(arimamodel, h = fh)$mean

# Guardar resultados
forec[, i] <- f1
forecarima[, i] <- farima
forecnar[, i] <- fnnet
forecnaive[, i] <- fnaive
forects_trend[, i] <- fets_trend
reales[, i] <- datos_24_25$IT[(ini + i + 1):(ini + i + fh)]
i <- i + 1}
## Guardar archivo
save(forec, forecarima, forecnar, forecnaive, forects_trend, reales, file =
"results4.RData")
## Cargar si es necesario
load("results4.RData")

## errores
errores <- reales - forec
erroresarima <- reales - forecarima
erroresnnar <- reales - forecnar
erroresnaive <- reales - forecnaive
erroresetstrend <- reales - forects_trend

## RMSE
RMSE <- sqrt(rowMeans(errores^2, na.rm = TRUE))
RMSEarima <- sqrt(rowMeans(erroresarima^2, na.rm = TRUE))
RMSE_nnar <- sqrt(rowMeans(erroresnnar^2, na.rm = TRUE))
RMSE_naive <- sqrt(rowMeans(erroresnaive^2, na.rm = TRUE))
RMSE_ets_trend <- sqrt(rowMeans(erroresetstrend^2, na.rm = TRUE))

## MdAE
MdAE <- apply(abs(errores), 1, median, na.rm = TRUE)
MdAE_arima <- apply(abs(erroresarima), 1, median, na.rm = TRUE)
MdAE_nnar <- apply(abs(erroresnnar), 1, median, na.rm = TRUE)
MdAE_naive <- apply(abs(erroresnaive), 1, median, na.rm = TRUE)
MdAE_ets_trend <- apply(abs(erroresetstrend), 1, median, na.rm = TRUE)

## Gráficas comparativas
plot(RMSE, type = "l", col = "blue", lwd = 2,
      ylim = range(c(RMSE, RMSEarima, RMSE_nnar, RMSE_naive, RMSE_ets_trend),
      na.rm = TRUE),
      ylab = "RMSE", xlab = "Horizonte", main = "Comparación de RMSE")
lines(RMSEarima, col = "violet", lwd = 2)
lines(RMSE_nnar, col = "orange", lwd = 2)
lines(RMSE_naive, col = "green", lwd = 2)
lines(RMSE_ets_trend, col = "red", lwd = 2)
legend("topright",
      legend = c("ETS(A,N,A)", "ARIMA", "NNAR", "Naive", "ETS(A,A,N)"),
      col = c("blue", "violet", "orange", "green", "red"), lty = 1, lwd = 2,
      cex = 0.8)

## Promedios de errores

```

```

cat("Promedios de RMSE:\n")
cat(" ETS(ANA):", mean(RMSE, na.rm = TRUE), "\n",
    "ARIMA:", mean(RMSEarima, na.rm = TRUE), "\n",
    "NNAR:", mean(RMSE_nnar, na.rm = TRUE), "\n",
    "Naive:", mean(RMSE_naive, na.rm = TRUE), "\n",
    "ETS(AAN):", mean(RMSE_ets_trend, na.rm = TRUE), "\n\n")

## Gráfico predicciones + original (para el mejor modelo)
### Seleccionar la última ventana de predicción para ETS
ultima_ventana <- ncol(forec)
predicciones <- forec[, ultima_ventana]
reales_ultimos <- datos_24_25$IT[(nrow(datos_24_25) - length(predicciones) +
1):nrow(datos_24_25)]
### Crear un objeto ts con los datos reales
ts_data <- ts(datos_24_25$IT, frequency = 7)
### Ajustar el modelo ETS final (usando todos los datos)
modelo_final <- ets(ts_data, model = "ANA")
### Generar las predicciones con intervalos (30 días adelante)
pred_final <- forecast(modelo_final, h = 30, level = c(80, 95))
### Preparar datos para el gráfico
ultima_fecha <- max(datos_24_25$Fecha)
fechas_pred <- seq(ultima_fecha + 1, by = "day", length.out = 30)
### Crear dataframe para ggplot
df_pred <- data.frame(Fecha = c(datos_24_25$Fecha, fechas_pred), IT =
c(datos_24_25$IT, rep(NA, 30)),
Prediccion = c(rep(NA, nrow(datos_24_25)), as.numeric(pred_final$mean)),
Lo_80 = c(rep(NA, nrow(datos_24_25)), as.numeric(pred_final$lower[,1])),
Hi_80 = c(rep(NA, nrow(datos_24_25)), as.numeric(pred_final$upper[,1])),
Lo_95 = c(rep(NA, nrow(datos_24_25)), as.numeric(pred_final$lower[,2])),
Hi_95 = c(rep(NA, nrow(datos_24_25)), as.numeric(pred_final$upper[,2])))
### Gráfico
ggplot(df_pred, aes(x = Fecha)) +
# Datos históricos
geom_line(aes(y = IT), color = "black", size = 0.8) +
# Intervalo de predicción del 95%
geom_ribbon(aes(ymin = Lo_95, ymax = Hi_95), fill = "blue", alpha = 0.2) +
# Intervalo de predicción del 80%
geom_ribbon(aes(ymin = Lo_80, ymax = Hi_80), fill = "blue", alpha = 0.3) +
# Predicción puntual
geom_line(aes(y = Prediccion), color = "red", size = 1) +
# Línea vertical que marca el inicio de las predicciones
geom_vline(xintercept = ultima_fecha, linetype = "dashed", color = "gray") +
# Etiquetas
labs(title = "Predicciones de Absentismo IT con Intervalos de Predicción",
subtitle = "Modelo ETS(A,N,A) - Intervalos al 80% y 95%",
x = "Fecha", y = "Horas de Absentismo IT") +
theme_minimal() + theme( legend.position = "none",
axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
scale_x_date(date_breaks = "1 month", date_labels = "%b %Y") +
annotate("text", x = ultima_fecha - 15, y = max(datos_24_25$IT, na.rm =
TRUE)*0.9,
label = "Inicio predicciones", color = "gray", size = 3) +
annotate("text", x = ultima_fecha + 15, y = max(pred_final$upper[,2], na.rm =
TRUE),
label = "Predicción", color = "red", size = 3)

```

```

# PREDICCIÓN 2025 JULIO
# Filtrar datos hasta junio 2025 y predecir julio 2025
datos_hasta_junio <- datos_24_25 %>%
  filter(Fecha <= as.Date("2025-06-30"))

# Crear serie temporal
ts_data_junio <- ts(datos_hasta_junio$IT, frequency = 7)

# Ajustar modelo ETS
modelo_julio <- ets(ts_data_junio, model = "ANA")

# Predecir 31 días (julio completo)
pred_julio <- forecast(modelo_julio, h = 31, level = c(80, 95))

# Preparar fechas
fechas_julio <- seq(as.Date("2025-07-01"), by = "day", length.out = 31)

# Dataframe para ggplot
df_julio <- data.frame(
  Fecha = c(datos_hasta_junio$Fecha, fechas_julio),
  IT = c(datos_hasta_junio$IT, rep(NA, 31)),
  Prediccion = c(rep(NA, nrow(datos_hasta_junio)),
  as.numeric(pred_julio$mean)),
  Lo_80 = c(rep(NA, nrow(datos_hasta_junio)),
  as.numeric(pred_julio$lower[,1])),
  Hi_80 = c(rep(NA, nrow(datos_hasta_junio)),
  as.numeric(pred_julio$upper[,1])),
  Lo_95 = c(rep(NA, nrow(datos_hasta_junio)),
  as.numeric(pred_julio$lower[,2])),
  Hi_95 = c(rep(NA, nrow(datos_hasta_junio)),
  as.numeric(pred_julio$upper[,2]))
)

# Gráfico
ggplot(df_julio, aes(x = Fecha)) +
  # Datos históricos
  geom_line(aes(y = IT), color = "black", size = 0.8) +
  # Intervalos de predicción
  geom_ribbon(aes(ymin = Lo_95, ymax = Hi_95), fill = "blue", alpha = 0.2) +
  geom_ribbon(aes(ymin = Lo_80, ymax = Hi_80), fill = "blue", alpha = 0.3) +
  # Predicción puntual
  geom_line(aes(y = Prediccion), color = "red", size = 1) +
  # Línea divisoria
  geom_vline(xintercept = as.Date("2025-06-30"), linetype = "dashed", color =
"gray") +
  # Etiquetas y estilo
  labs(title = "Predicción de Absentismo IT para Julio 2025",
       subtitle = "Modelo ETS(A,N,A) - Intervalos al 80% y 95%",
       x = "Fecha", y = "Horas de Absentismo IT") +
  theme_minimal() +
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
  scale_x_date(date_breaks = "1 month", date_labels = "%b %Y",
               limits = c(as.Date("2025-01-01"), as.Date("2025-07-31"))) +
  # Anotaciones
  annotate("text", x = as.Date("2025-04-15"), y = max(datos_hasta_junio$IT,
na.rm = TRUE)*0.9,
           label = "Datos históricos", color = "black", size = 3) +

```

```

    annotate("text", x = as.Date("2025-07-15"), y = max(pred_julio$upper[,2],
na.rm = TRUE),
           label = "Predicción Julio", color = "red", size = 3)

#####
##### SHINY DASHBOARD
# UI (Interfaz de Usuario)
ui <- dashboardPage(
  dashboardHeader(title = "Control Absentismo"),

  dashboardSidebar(
    sidebarMenu(
      menuItem("Resumen General", tabName = "resumen", icon =
icon("dashboard")),
      menuItem("Predicciones", tabName = "predicciones", icon = icon("chart-line")),
      menuItem("Comparativas", tabName = "comparativas", icon =
icon("balance-scale"))
    )
  ),

  dashboardBody(
    tabItems(
      # Pestaña de Resumen General
      tabItem(tabName = "resumen",
              fluidRow(
                valueBox(round(ratioIT, 2), "Ratio de Absentismo IT (%)",
icon = icon("percent"), color = "red"),
                valueBox(round(Eficiencia*100, 2), "Eficiencia Ajustada (%)",
icon = icon("thumbs-up"), color = "green"),
                valueBox(sum(datos_24_25$IT), "Total Horas IT (2024-2025)",
icon = icon("clock"), color = "blue")
              ),
              fluidRow(
                box(title = "Estadísticas Descriptivas Globales", status =
"primary", solidHeader = TRUE,
                    tableOutput("estadisticas_generales"), width = 6),
                box(title = "Estadísticas Descriptivas por Tipo de Día",
status = "primary", solidHeader = TRUE,
                    tableOutput("estadisticas_tipo_dia"), width = 6)
              ),
              fluidRow(
                box(title = "Distribución por Tipo de Día", status =
"warning", solidHeader = TRUE,
                    plotOutput("boxplot_tipo_dia"), width = 6),
                box(title = "Distribución de Densidad", status = "warning",
solidHeader = TRUE,
                    plotOutput("densidad_tipo_dia"), width = 6)
              ),
              fluidRow(
                box(title = "Evolución Temporal del Absentismo (2024-2025)",
status = "info", solidHeader = TRUE,
                    plotOutput("plot_evolucion_temporal"), width = 12)
              ),
              fluidRow(
                box(title = "Absentismo por Día de la Semana", status =
"info", solidHeader = TRUE,
                    plotOutput("plot_dia_semana"), width = 6),

```

```

        box(title = "Tasas por Día de la Semana", status = "info",
solidHeader = TRUE,
            tableOutput("tabla_dias_semana"), width = 6)
    ),
    fluidRow(
        box(title = "Tasas por Tipo de Día", status = "info",
solidHeader = TRUE,
            tableOutput("tabla_tipo_dia"), width = 6)
    )
),

# Pestaña de Predicciones
tabItem(tabName = "predicciones",
    fluidRow(
        box(title = "Comparación de Modelos Predictivos", status =
"primary", solidHeader = TRUE,
            plotOutput("plot_comparacion_modelos"), width = 12)),
    fluidRow(
        box(title = "Predicción de Absentismo (30 días)", status =
"info", solidHeader = TRUE,
            plotOutput("plot_predicciones"), width = 12)),
    fluidRow(
        box(title = "Predicción Julio 2025", status = "success",
solidHeader = TRUE,
            plotOutput("plot_prediccion_julio"), width = 12,
            footer = "Predicción basada en datos hasta junio 2025 con
modelo ETS(A,N,A)"),
    ),

# Pestaña de Comparativas
tabItem(tabName = "comparativas",
    fluidRow(
        box(title = "Comparación Marzo vs Mayo", status = "primary",
solidHeader = TRUE,
            plotOutput("plot_marzo_mayo"), width = 6),
        box(title = "Métricas de Error Marzo-Mayo", status = "info",
solidHeader = TRUE,
            verbatimTextOutput("metricas_marzo_mayo"), width = 6)
    ),
    fluidRow(
        box(title = "Comparación Marzo vs Junio", status = "primary",
solidHeader = TRUE,
            plotOutput("plot_marzo_junio"), width = 6),
        box(title = "Métricas de Error Marzo-Junio", status = "info",
solidHeader = TRUE,
            verbatimTextOutput("metricas_marzo_junio"), width = 6)
    ),
    fluidRow(
        box(title = "Comparación Mayo vs Junio", status = "primary",
solidHeader = TRUE,
            plotOutput("plot_mayo_junio"), width = 6),
        box(title = "Métricas de Error Mayo-Junio", status = "info",
solidHeader = TRUE,
            verbatimTextOutput("metricas_mayo_junio"), width = 6)
    )
)
)
)
)
)
```

```

# Server (Lógica del Servidor)
server <- function(input, output) {

  # Resumen General
  output$estadisticas_generales <- function() {
    estadisticas_IT %>%
      kable("html") %>%
      kable_styling("striped", full_width = FALSE) %>%
      column_spec(1, bold = TRUE) %>%
      scroll_box(width = "100%", height = "300px")
  }

  output$estadisticas_tipo_dia <- function() {
    estadisticas_por_tipo %>%
      kable("html") %>%
      kable_styling("striped", full_width = FALSE) %>%
      column_spec(1, bold = TRUE) %>%
      scroll_box(width = "100%", height = "300px")
  }

  output$boxplot_tipo_dia <- renderPlot({
    ggplot(datos_24_25, aes(x = tipo_dia, y = IT, fill = tipo_dia)) +
      geom_boxplot(outlier.color = "red") +
      labs(title = "Distribución del Absentismo IT por Tipo de Día",
           x = NULL, y = "Horas de IT") +
      theme_minimal() +
      theme(legend.position = "none")
  })

  output$densidad_tipo_dia <- renderPlot({
    ggplot(datos_24_25, aes(x = IT, fill = tipo_dia)) +
      geom_density(alpha = 0.5) +
      labs(title = "Distribución de Densidad por Tipo de Día",
           x = "Horas de Absentismo", y = "Densidad") +
      theme_minimal() +
      scale_fill_brewer(palette = "Set1", name = "Tipo de Día")
  })

  output$plot_evolucion_temporal <- renderPlot({
    matplot(x = datos_24_25$Fecha, y = datos_24_25$IT,
            type = "l", lty = 1, lwd = 1, col = "blue",
            xlab = "Fecha", ylab = "Horas de Absentismo IT",
            main = "Evolución del Absentismo IT (2024 y 2025)", xaxt = "n")
    axis.Date(1, at = seq(min(datos_24_25$Fecha), max(datos_24_25$Fecha), by =
= "month"),
              format = "%b %Y", las = 1)
    grid()
  })

  output$plot_hplanif_vs_it <- renderPlot({
    matplot(x = datos_2025_hplanif$Fecha,
            y = cbind(datos_2025_hplanif$IT, datos_2025_hplanif$hplanif),
            type = "l", lty = 1, lwd = 2,
            col = c("#1f77b4", "#ff7f0e"),
            xlab = "Fecha", ylab = "Horas",
            main = "Horas planificadas vs IT (2025)")
    legend("topright", legend = c("IT", "horas planificadas"),
           col = c("#1f77b4", "#ff7f0e"), lty = 1, lwd = 2, cex = 1)
  })
}

```

```

        grid()
    })

output$plot_dia_semana <- renderPlot({
  grafico1 <- ggplot(datos_2025_hplanif, aes(x = dia_semana, y = hplanif,
fill = dia_semana)) +
    geom_bar(stat = "identity") +
    labs(title = "Horas planificadas por día", x = NULL, y = "Horas") +
    theme_minimal() +
    theme(legend.position = "none") +
    scale_fill_brewer(palette = "Set2")

  grafico2 <- ggplot(datos_24_25, aes(x = dia_semana, y = IT, fill =
dia_semana)) +
    geom_bar(stat = "identity") +
    labs(title = "Absentismo IT por día", x = NULL, y = "Horas IT") +
    theme_minimal() +
    theme(legend.position = "none")

  grafico1 + grafico2
})

output$tabla_dias_semana <- function() {
  tabla_comparativa %>%
    kable("html") %>%
    kable_styling("striped", full_width = FALSE) %>%
    column_spec(4, bold = TRUE) %>%
    scroll_box(width = "100%", height = "200px")
}

output$tabla_tipo_dia <- function() {
  tabla_comparativa_tipo %>%
    kable("html") %>%
    kable_styling("striped", full_width = FALSE) %>%
    column_spec(4, bold = TRUE) %>%
    scroll_box(width = "100%", height = "200px")
}

# Predicciones
output$plot_comparacion_modelos <- renderPlot({
  plot(RMSE, type = "l", col = "blue", lwd = 2,
       ylim = range(c(RMSE, RMSEarima, RMSE_nnar, RMSE_naive,
RMSE_ets_trend), na.rm = TRUE),
       ylab = "RMSE", xlab = "Horizonte", main = "Comparación de Modelos
por RMSE")
  lines(RMSEarima, col = "violet", lwd = 2)
  lines(RMSE_nnar, col = "orange", lwd = 2)
  lines(RMSE_naive, col = "green", lwd = 2)
  lines(RMSE_ets_trend, col = "red", lwd = 2)
  legend("topright",
         legend = c("ETS(A,N,A)", "ARIMA", "NNAR", "Naive", "ETS(A,A,N)"),
         col = c("blue", "violet", "orange", "green", "red"), lty = 1, lwd
= 2, cex = 0.8)
})

output$plot_prediccciones <- renderPlot({
  ggplot(df_pred, aes(x = Fecha)) +
    geom_line(aes(y = IT), color = "black", size = 0.8) +

```

```

    geom_ribbon(aes(ymin = Lo_95, ymax = Hi_95), fill = "blue", alpha =
0.2) +
      geom_ribbon(aes(ymin = Lo_80, ymax = Hi_80), fill = "blue", alpha =
0.3) +
        geom_line(aes(y = Prediccion), color = "red", size = 1) +
        geom_vline(xintercept = ultima_fecha, linetype = "dashed", color =
"gray") +
          labs(title = "Predicciones de Absentismo IT (30 días)",
              subtitle = "Modelo ETS(A,N,A) con intervalos de confianza",
              x = "Fecha", y = "Horas de Absentismo IT") +
            theme_minimal() +
            theme(legend.position = "none",
                  axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
            scale_x_date(date_breaks = "1 month", date_labels = "%b %Y")
      })
}

output$plot_prediccion_julio <- renderPlot({
  ggplot(df_julio, aes(x = Fecha)) +
    geom_line(aes(y = IT), color = "black", size = 0.8) +
    geom_ribbon(aes(ymin = Lo_95, ymax = Hi_95), fill = "blue", alpha =
0.2) +
      geom_ribbon(aes(ymin = Lo_80, ymax = Hi_80), fill = "blue", alpha =
0.3) +
        geom_line(aes(y = Prediccion), color = "red", size = 1) +
        geom_vline(xintercept = as.Date("2025-06-30"), linetype = "dashed",
color = "gray") +
          labs(title = "Predicción de Absentismo IT para Julio 2025",
              subtitle = "Modelo ETS(A,N,A) - Intervalos al 80% y 95%",
              x = "Fecha", y = "Horas de Absentismo IT") +
            theme_minimal() +
            theme(legend.position = "none",
                  plot.title = element_text(size = 16, face = "bold"),
                  axis.title = element_text(size = 12),
                  axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +
            scale_x_date(date_breaks = "1 week", date_labels = "%d %b %Y",
                        limits = c(as.Date("2025-05-01"), as.Date("2025-07-31")))
+
  annotate("text", x = as.Date("2025-05-15"), y =
max(datos_hasta_junio$IT, na.rm = TRUE)*0.9,
         label = "Datos históricos", color = "black", size = 4) +
  annotate("text", x = as.Date("2025-07-15"), y =
max(pred_julio$upper[,2], na.rm = TRUE),
         label = "Predicción Julio", color = "red", size = 4)
})

# Comparativas
output$plot_marzo_mayo <- renderPlot({
  matplot(x = datos_marzo_junio$Fecha,
          y = cbind(datos_marzo_junio$IT_marzo, datos_marzo_junio$IT_mayo,
error_abs_marzo_mayo),
          type = "l", lty = 1, lwd = 2,
          col = c("#1f77b4", "#ff7f0e", "#2ca02c"),
          xlab = "Fecha", ylab = "Horas de Absentismo IT",
          main = "Comparación IT Marzo vs Mayo")
  legend("topright",
         legend = c("IT Marzo", "IT Mayo", "Error Absoluto"),
         col = c("#1f77b4", "#ff7f0e", "#2ca02c"), lty = 1, lwd = 2, cex =
0.7)
  grid()
})

```

```

})

output$metricas_marzo_mayo <- renderPrint({
  cat("Métricas de Error Marzo-Mayo:\n")
  cat("RMSE:", round(rmse_marzo_mayo, 2), "\n")
  cat("Error Absoluto Medio:", round(mean(error_abs_marzo_mayo), 2), "\n")
  cat("Error Absoluto Máximo:", round(max(error_abs_marzo_mayo), 2), "\n")
  cat("Error Absoluto Mínimo:", round(min(error_abs_marzo_mayo), 2), "\n")
})

output$plot_marzo_junio <- renderPlot({
  matplot(x = datos_marzo_junio$Fecha,
          y = cbind(datos_marzo_junio$IT_marzo, datos_marzo_junio$IT_junio,
error_abs_marzo_junio),
          type = "l", lty = 1, lwd = 2,
          col = c("#1f77b4", "#ff7f0e", "#2ca02c"),
          xlab = "Fecha", ylab = "Horas de Absentismo IT",
          main = "Comparación IT Marzo vs Junio")
  legend("topright",
         legend = c("IT Marzo", "IT Junio", "Error Absoluto"),
         col = c("#1f77b4", "#ff7f0e", "#2ca02c"), lty = 1, lwd = 2, cex =
0.7)
  grid()
})

output$metricas_marzo_junio <- renderPrint({
  cat("Métricas de Error Marzo-Junio:\n")
  cat("RMSE:", round(rmse_marzo_junio, 2), "\n")
  cat("Error Absoluto Medio:", round(mean(error_abs_marzo_junio), 2), "\n")
  cat("Error Absoluto Máximo:", round(max(error_abs_marzo_junio), 2), "\n")
  cat("Error Absoluto Mínimo:", round(min(error_abs_marzo_junio), 2), "\n")
})

output$plot_mayo_junio <- renderPlot({
  matplot(x = datos_marzo_junio$Fecha,
          y = cbind(datos_marzo_junio$IT_mayo, datos_marzo_junio$IT_junio,
error_abs_mayo_junio),
          type = "l", lty = 1, lwd = 2,
          col = c("#1f77b4", "#ff7f0e", "#2ca02c"),
          xlab = "Fecha", ylab = "Horas de Absentismo IT",
          main = "Comparación IT Mayo vs Junio")
  legend("topright",
         legend = c("IT Mayo", "IT Junio", "Error Absoluto"),
         col = c("#1f77b4", "#ff7f0e", "#2ca02c"), lty = 1, lwd = 2, cex =
0.7)
  grid()
})

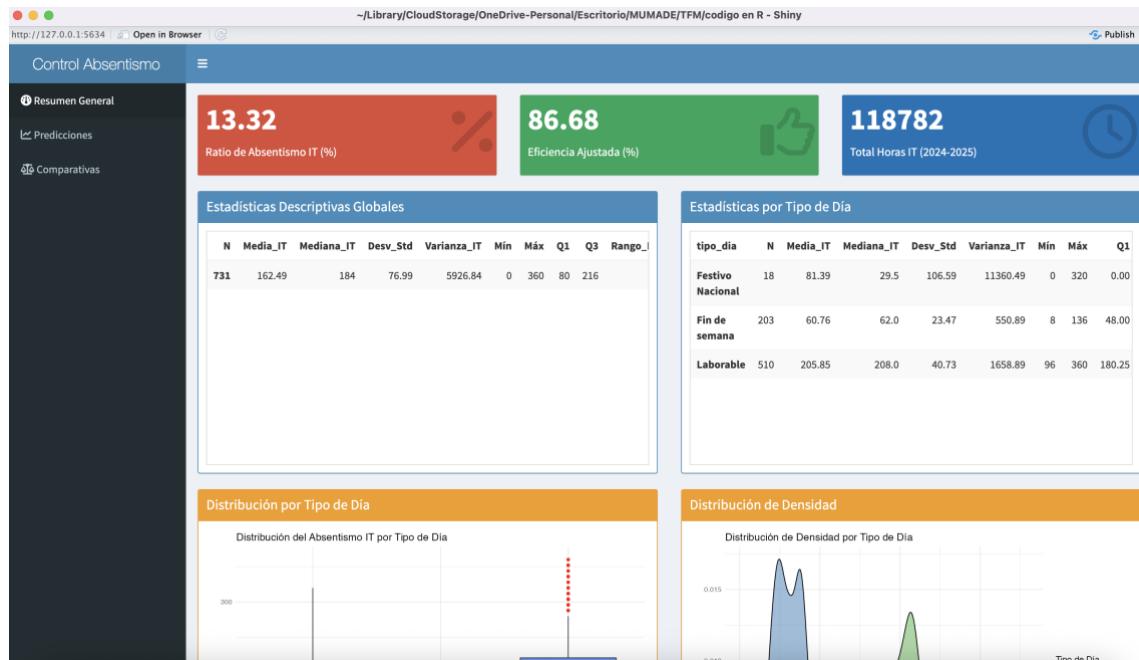
output$metricas_mayo_junio <- renderPrint({
  cat("Métricas de Error Mayo-Junio:\n")
  cat("RMSE:", round(rmse_mayo_junio, 2), "\n")
  cat("Error Absoluto Medio:", round(mean(error_abs_mayo_junio), 2), "\n")
  cat("Error Absoluto Máximo:", round(max(error_abs_mayo_junio), 2), "\n")
  cat("Error Absoluto Mínimo:", round(min(error_abs_mayo_junio), 2), "\n")
})

# Ejecutar la aplicación
shinyApp(ui, server)

```

Anexo II. Imágenes Dashboard

Figura Anexo 1. Vista Resumen general



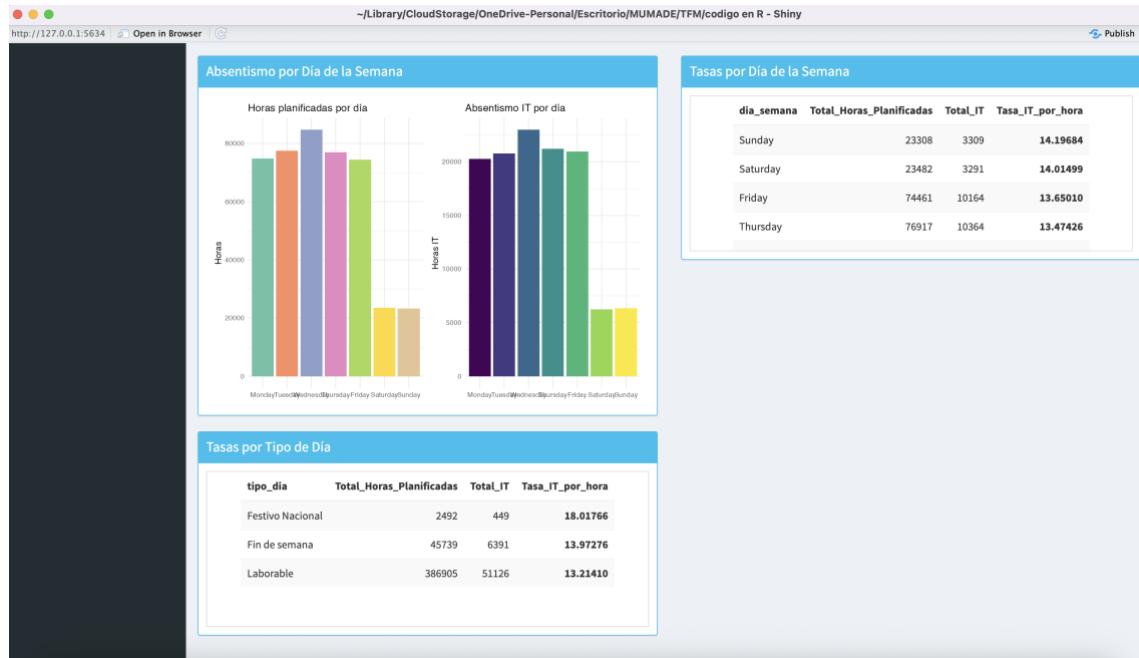
Fuente: Elaboración propia, 2025

Figura Anexo 2. Vista Resumen general 2



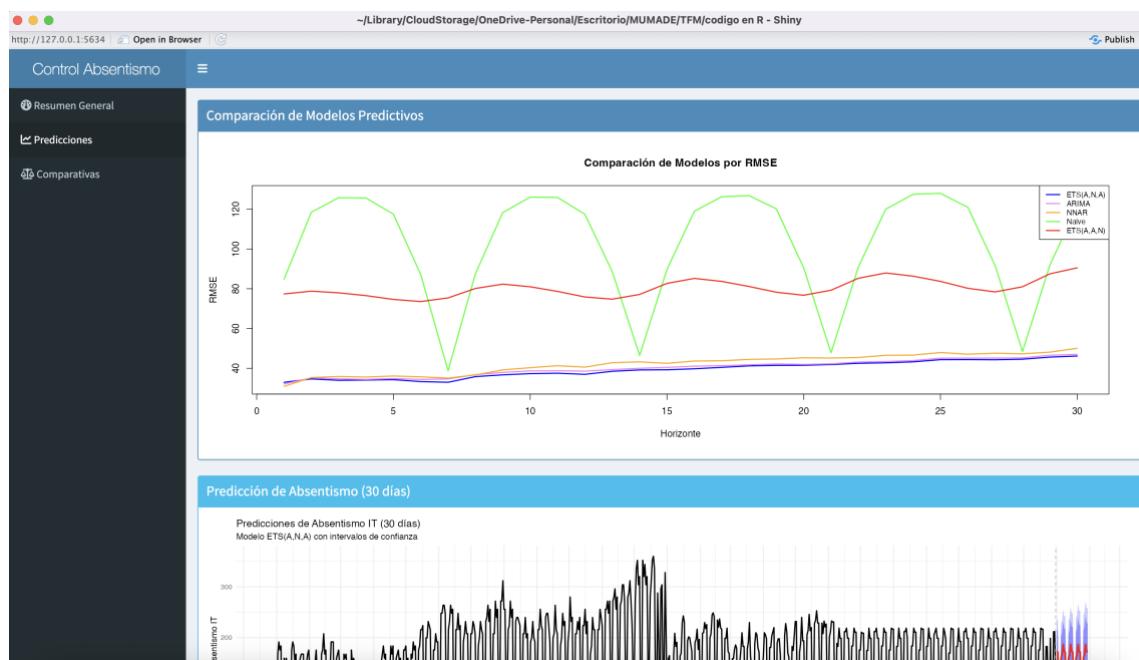
Fuente: Elaboración propia, 2025

Figura Anexo 3. Vista Resumen general 3



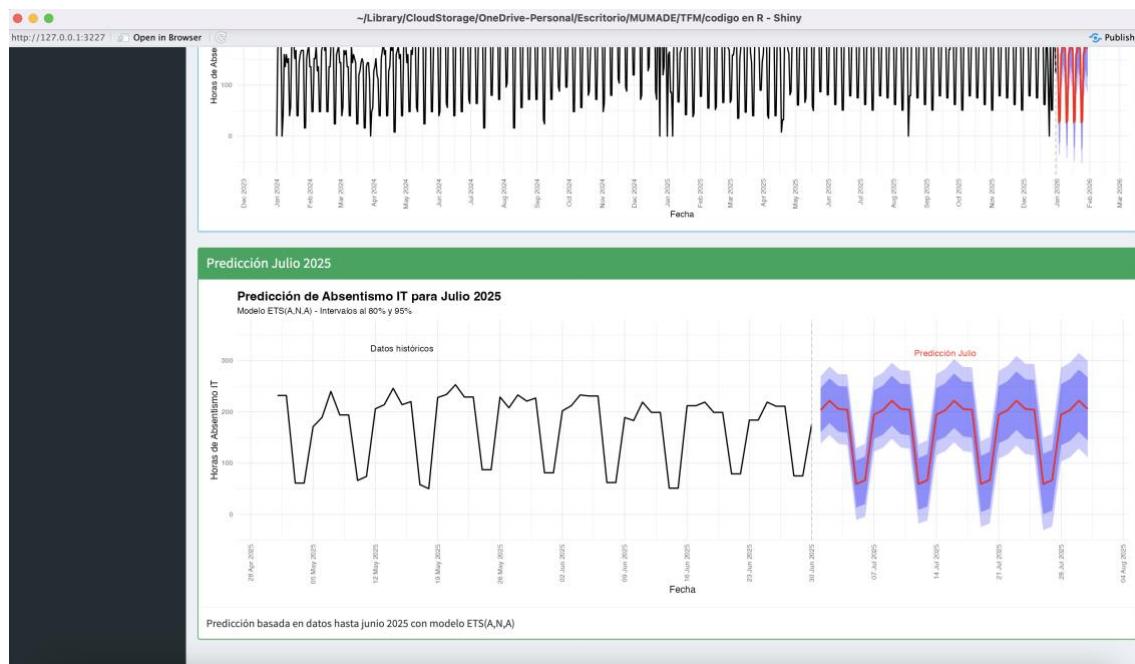
Fuente: Elaboración propia, 2025

Figura Anexo 4. Vista Predicciones



Fuente: Elaboración propia, 2025

Figura Anexo 5. Vista Predicciones 2



Fuente: Elaboración propia, 2025

Figura Anexo 6. Vista Comparativas bases de datos marzo, mayo y junio



Fuente: Elaboración propia, 2025