

Predicción de afluencia de pasajeros para mejorar la gestión del transporte urbano en Montevideo

Entregado como requisito para la obtención del título de Máster en Big Data

Christian Vera – 146159

Verónica González- 245393

Tutor: Sergio Yovine

2025

Declaración de autoría

Nosotros, Christian Vera y Verónica González declaramos que el trabajo que se presenta en esa obra es de nuestra propia mano. Podemos asegurar que:

- La obra fue producida en su totalidad mientras realizábamos Trabajo Final del Máster en Big Data;
- Cuando hemos consultado el trabajo publicado por otros, lo hemos atribuido con claridad;
- Cuando hemos citado obras de otros, hemos indicado las fuentes. Con excepción de estas citas, la obra es enteramente nuestra;
- En la obra, hemos acusado recibo de las ayudas recibidas;
- Cuando la obra se basa en trabajo realizado conjuntamente con otros, hemos explicado claramente qué fue contribuido por otros, y qué fue contribuido por nosotros;
- Ninguna parte de este trabajo ha sido publicada previamente a su entrega, excepto donde se han realizado las aclaraciones correspondientes.



Christian Vera

21-04-2025



Verónica González

21-04-2025

Agradecimientos

A nuestras familias y amigos por el apoyo incondicional, a nuestro tutor por su apoyo constante, a la Universidad ORT Uruguay y a la Intendencia de Montevideo por brindarnos los datos para el desarrollo del trabajo y a sus técnicos por la disposición ante consultas realizadas.

Resumen

Este proyecto propone el desarrollo de un modelo predictivo para estimar la cantidad de pasajeros por línea de ómnibus en Montevideo, viendo el impacto de la lluvia, utilizando datos históricos de ocupación de los ómnibus y registros de precipitaciones proporcionados por la Intendencia de Montevideo.

El objetivo principal es optimizar la asignación de la flota de ómnibus, mejorando la eficiencia del servicio y reduciendo los tiempos de espera de los usuarios. Para ello, se aplican técnicas de análisis de datos y modelos de aprendizaje automático (en adelante, Machine Learning o simplemente ML por sus siglas en inglés) incluyendo enfoques de series temporales y redes neuronales MLP (Multi Layer Perceptron) con el fin de capturar patrones históricos y variables meteorológicas relevantes. Se espera que los resultados obtenidos ofrezcan un sustento sólido para la toma de decisiones de las autoridades, impulsando una planificación de transporte más efectiva y contribuyendo a la satisfacción de los pasajeros.

Palabras clave

Modelo Predictivo, Machine Learning, Series Temporales, Redes Neuronales MLP, Patrones históricos.

Indice

1. INTRODUCCIÓN	7
2. MARCO HISTÓRICO	7
3. CASO DE ESTUDIO	8
4. METODOLOGÍA	9
5. TRANSFORMACIÓN DE LOS DATOS	9
5.1. Integración de fuentes de datos	9
5.2 Limpieza de datos	10
5.3 Transformación de variables	10
5.4 Selección de variables relevantes	10
5.5 Transformación temporal: nivel horario	11
5.6 Representación de variables cíclicas	11
6. TRANSFORMACIÓN DE LOS DATOS	11
6.1 Análisis Exploratorio	11
6.1.1 Análisis de la ocupación promedio	11
6.1.2 Comparación entre días hábiles, sábados, domingos y feriados	12
6.1.3 Impacto de las precipitaciones en la ocupación	13
6.1.4 Análisis de datos de precipitaciones	14
6.1.5 Observaciones adicionales del análisis exploratorio	15
6.2 Modelado Inicial	15
6.2.1 Descripción inicial del proceso	15
6.2.2 Modelado de series temporales SARIMAX	16
6.2.3 Primer modelo de red neuronal simple (MLP)	16
6.2.4 Conclusiones de la primera etapa	17
6.3 Construcción del modelo MLP mejorado	17
6.3.1 Decisiones de diseño para la construcción del modelo	18
6.3.2 Reformulación del enfoque y preparación de los datos	19
6.3.3 Primer modelo basado en la arquitectura del paper	19
6.3.4 Optimización de hiperparámetros y construcción del modelo mejorado	20
6.3.5 Resultados finales del modelo optimizado	21
6.4 Evaluación general y sugerencias de mejoras	23
6.4.1 Limitaciones observadas	23
6.4.2 Sugerencias de mejoras futuras	23
7. ACCESO A LOS RECURSOS DEL PROYECTO	24
8. REFERENCIAS BIBLIOGRÁFICAS	25
ANEXO 1	26
ANEXO 2	27

1. INTRODUCCIÓN

En la última década, la inteligencia artificial (IA) y el aprendizaje automático (Machine Learning, ML) se han convertido en componentes fundamentales para la toma de decisiones en organizaciones tanto públicas como privadas. Su capacidad para identificar patrones complejos y realizar predicciones precisas ha impulsado su adopción en sectores como la industria, la salud, las finanzas y la gestión urbana.

El crecimiento en la disponibilidad de datos y la capacidad de procesamiento computacional ha motivado a gobiernos e instituciones a incorporar modelos predictivos que optimicen recursos y mejoren la eficiencia de los servicios. En este marco, se plantea la presente investigación, cuyo objetivo es desarrollar un modelo de predicción de la cantidad de pasajeros en el sistema de transporte público de Montevideo, incorporando información climática (lluvia) como variable clave. Esto permitirá a la Intendencia, ajustar la flota en función de las condiciones climáticas adversas, mejorando la experiencia de los usuarios y optimizando la asignación de recursos. Cabe destacar que no se encontraron investigaciones o herramientas que aborden este tema en Uruguay.

Los datos utilizados en este estudio fueron brindados por la Intendencia de Montevideo, más allá de que los mismos son de carácter público y se encuentran disponibles en su portal de datos abiertos. Como expresa la Intendencia, “de esa manera pueden ser procesados por cualquier ciudadano, investigador, periodista, universidades y organizaciones nacionales y extranjeras” (Intendencia de Montevideo, 2014). Para este trabajo, se contó además con la colaboración directa del organismo mediante una solicitud de acceso a la información pública, que proporcionó un conjunto de datos ya preprocesado y acotado al período comprendido entre el 1/01/2024 y el 31/12/2024, lo que facilitó su análisis y aplicación al modelo predictivo.

2. MARCO HISTÓRICO

La inteligencia artificial (IA) como disciplina formal comenzó a tomar forma en la década de 1950, impulsada por avances en lógica matemática, neurociencia y computación. El hito fundacional más reconocido es la conferencia de Dartmouth en 1956, donde se propuso que “cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede ser descrito con suficiente precisión como para que una máquina pueda simularlo” [1]. Desde entonces, la IA ha atravesado distintas etapas de desarrollo, alternando entre períodos de entusiasmo y estancamiento, hasta llegar a enfoques modernos que integran el aprendizaje automático y el procesamiento masivo de datos como pilares fundamentales.

A partir de la década de 1990, el enfoque estadístico comenzó a consolidarse como uno de los principales motores del avance de la inteligencia artificial. El aprendizaje automático (Machine Learning, ML) emergió como un subcampo clave, utilizando modelos probabilísticos y técnicas estadísticas que permitieron a los sistemas aprender directamente a partir de los datos, sin necesidad de reglas explícitas [1]. En sus primeras aplicaciones, estos modelos se centraban en tareas supervisadas como la clasificación y la regresión, utilizando algoritmos relativamente simples en comparación con los enfoques actuales.

Con el crecimiento exponencial del poder computacional y la disponibilidad de grandes volúmenes de datos a partir de la década de 2000, se produjo un impulso decisivo en el desarrollo de modelos de aprendizaje más sofisticados. Entre estos, destacan las redes neuronales profundas y los métodos de aprendizaje por refuerzo, que lograron avances significativos en áreas como el reconocimiento de voz, la visión por computadora y la robótica [1].

La inteligencia artificial, en su concepción general, comprende todos aquellos sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, el aprendizaje, la percepción y la toma de decisiones [1]. Dentro de este campo, el aprendizaje automático representa una subdisciplina enfocada en el desarrollo de algoritmos que permiten a las máquinas mejorar su desempeño en tareas específicas a partir de datos, sin ser programadas explícitamente [2]. A su vez, el aprendizaje profundo (Deep Learning, DL) constituye un subcampo dentro del ML, caracterizado por el uso de redes neuronales profundas que permiten modelar patrones altamente complejos en datos de gran volumen y variedad.

La relación entre estas disciplinas puede representarse conceptualmente como conjuntos anidados, donde el aprendizaje profundo se encuentra dentro del aprendizaje automático, que a su vez forma parte del campo más amplio de la inteligencia artificial. Lo mencionado se ilustra en la Figura 1.

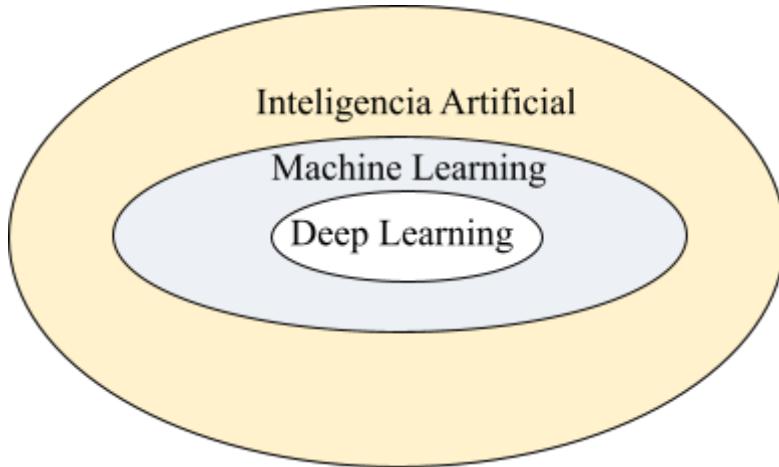


Figura 1 - Diagrama relación IA/ML/DL

En los últimos años, los avances en Deep Learning han impulsado el desarrollo de aplicaciones de Inteligencia Artificial Generativa (GenAI), cuyo objetivo es la creación de contenido original como textos, imágenes, música o videos. GenAI utiliza modelos basados en redes neuronales profundas, como transformers o redes generativas adversariales (GANs). Entre los ejemplos más representativos se encuentran ChatGPT para generación de texto y DALL·E para generación de imágenes. Si bien GenAI no constituye un subcampo técnico separado, representa una de las aplicaciones más transformadoras surgidas del Deep Learning reciente.

3. CASO DE ESTUDIO

El transporte público en Montevideo constituye un componente esencial de la movilidad urbana y social. Según datos del Ministerio de Transporte y Obras Públicas (MTOP) [3], hasta el año 2018, aproximadamente un 46 % de los trabajadores del departamento utilizaban el transporte colectivo para desplazarse a sus lugares de trabajo, con variaciones mínimas respecto a años anteriores.

A nivel general, la Encuesta de Movilidad del Área Metropolitana de Montevideo realizada en 2016 estimó que se realizaban aproximadamente 4,2 millones de viajes diarios en la ciudad. De ese total, un 25,2 % se efectuaban en ómnibus y un 32 % en automóvil particular, evidenciando una tendencia creciente hacia el uso del transporte privado. Además, la encuesta indicó que la duración promedio de un viaje en transporte público era de 46 minutos, mientras que en vehículo particular era de 21 minutos. Esta diferencia significativa en los tiempos de desplazamiento influye directamente en la elección modal de los usuarios, especialmente entre quienes disponen de la posibilidad económica de optar por medios privados.

En los últimos años, el debate sobre el futuro del sistema de transporte público de Montevideo ha cobrado relevancia, evidenciando múltiples desafíos estructurales. Según un análisis publicado en el diario *El Observador* por Varela y Olid [4], estos desafíos incluyen la disminución en la frecuencia de ciertas líneas, la falta de priorización del transporte público en la infraestructura vial, y la falta de integración efectiva entre las zonas periféricas y los corredores principales de transporte, lo que ha derivado en impactos negativos principalmente sobre los sectores más vulnerables de la población.

El mismo informe destaca que la planificación ineficiente de rutas, las barreras tecnológicas y la falta de coordinación institucional han dificultado la modernización del sistema. Asimismo, se señala que experiencias exitosas en otras ciudades latinoamericanas, como Bogotá y Santiago de Chile, demuestran que la incorporación de tecnologías, esquemas de gestión integrados y planificación basada en datos ha permitido mejorar la eficiencia y la equidad del transporte público.

Montevideo forma parte del Sistema de Transporte Metropolitano (STM), que integra también a los departamentos de Canelones y San José. Dentro del departamento de Montevideo, la red de transporte está compuesta por 92 líneas urbanas, 44 líneas locales de cercanía y 5 líneas diferenciales semidirectos, operadas por diversas empresas concesionarias.

En este contexto, disponer de herramientas que permitan anticipar la demanda de pasajeros y adaptar dinámicamente la oferta de servicios resulta clave para optimizar el funcionamiento del sistema. El presente trabajo propone el desarrollo de un modelo de predicción de la cantidad de pasajeros, utilizando datos históricos de movilidad y registros de precipitaciones, con el objetivo de contribuir a una mejor planificación operativa de la flota de ómnibus en días de lluvia, promoviendo así un transporte público más eficiente, equitativo y sostenible.

4. METODOLOGÍA

El presente trabajo se desarrolló a partir de datos proporcionados por la Intendencia de Montevideo, correspondientes a los registros de transporte público y precipitaciones meteorológicas para el período comprendido entre el 1 de enero y el 31 de diciembre de 2024. Si bien los datos de transporte forman parte del portal de datos abiertos de la Intendencia, para esta investigación se contó con la colaboración directa del organismo, que facilitó un conjunto preprocesado y acotado específicamente para el análisis requerido.

La metodología seguida para el desarrollo del modelo predictivo se compone de las siguientes etapas principales:

- Recolección y comprensión de los datos:

Se analizaron las estructuras de los datasets recibidos, identificando las variables relevantes para el objetivo del proyecto. Los datos incluían información horaria de ocupación de ómnibus, líneas de transporte, fechas, horarios y registros de precipitación acumulada.

- Preparación de los datos:

Se realizaron tareas de limpieza, tales como la eliminación de registros inconsistentes y la corrección de valores atípicos evidentes. Asimismo, se llevó a cabo la transformación de variables necesarias para su posterior uso en el modelo predictivo.

- Selección de variables:

A partir del análisis exploratorio inicial, se determinaron las variables más relevantes para la predicción de la cantidad de pasajeros, en escenarios con y sin lluvia. Se consideraron factores como el día de la semana, la hora del día, la línea de transporte y el volumen de precipitaciones acumuladas.

- Construcción del modelo:

Se evaluaron distintos algoritmos de aprendizaje supervisado. La elección final se basó en la capacidad de los modelos para capturar patrones complejos en los datos temporales y en su rendimiento durante la fase de validación.

- Evaluación del modelo:

La performance del modelo fue evaluada mediante métricas de error comunes en problemas de regresión, tales como el error absoluto medio (MAE) y el error cuadrático medio (RMSE), considerando tanto la precisión global como su desempeño específico en condiciones de lluvia.

Las etapas de transformación de los datos y la selección detallada de las variables utilizadas se describen en profundidad en la sección siguiente.

5. TRANSFORMACIÓN DE LOS DATOS

El proceso de preparación de los datos fue una etapa fundamental en el desarrollo del modelo predictivo, dado que la calidad de los datos disponibles impacta directamente en la performance del modelo. A continuación, se describen las principales transformaciones realizadas sobre los conjuntos de datos utilizados.

5.1 Integración de fuentes de datos

Los datos utilizados provienen de dos fuentes principales: registros de transporte público (líneas de ómnibus, horarios y ocupación) y registros meteorológicos (precipitaciones acumuladas cada 15 minutos).

En el caso de los datos de transporte, los registros indicaban la ocupación de los ómnibus en intervalos de 15 minutos. Para su posterior uso en el modelo, se realizó una agregación horaria, calculando el promedio de ocupación correspondiente a cada hora cerrada a partir de los cuatro registros de cada hora.

Respecto a los datos de precipitaciones, estos consistían en registros acumulados cada 15 minutos provenientes de 15 estaciones meteorológicas ubicadas en diferentes zonas de Montevideo. Para adecuarlos al análisis, se aplicaron las siguientes transformaciones:

- Conversión de la variable de fecha a formato datetime y establecimiento de la fecha como índice.
- Agrupación de los registros por estación y resampleo en intervalos de una hora, sumando las precipitaciones acumuladas en cada estación.
- Cálculo del promedio horario de precipitaciones entre todas las estaciones meteorológicas.

La decisión de utilizar el promedio de precipitaciones entre todas las estaciones se basó en dos consideraciones fundamentales. En primer lugar, Montevideo es una ciudad de tamaño relativamente pequeño, donde no se registran diferencias significativas en los niveles de precipitación entre distintas zonas en la mayoría de los eventos climáticos. En segundo lugar, el uso de un valor promedio simplificó el proceso de integración de las fuentes de datos, evitando la necesidad de asignar estaciones específicas a cada recorrido de ómnibus, lo que hubiera implicado complejidad adicional sin un beneficio proporcional en la precisión del modelo.

Finalmente, se extrajeron los atributos de día y hora cerrada a partir de las marcas de tiempo, permitiendo el emparejamiento posterior de los datos meteorológicos con los registros de transporte.

5.2 Limpieza de datos

Se realizaron tareas de limpieza consistentes en:

- Eliminación de registros incompletos o inconsistentes, como horarios fuera del rango esperado o datos faltantes en campos clave.
- Corrección de inconsistencias de formato en variables de fecha y hora.
- Normalización de nombres de líneas y paradas para asegurar su homogeneidad.

5.3 Transformación de variables

Se llevaron a cabo diversas transformaciones para adecuar las variables a los requerimientos del modelo predictivo:

- **Generación de nuevas variables temporales:** A partir de la fecha y hora original, se extrajeron atributos como el día de la semana, la hora cerrada y el mes, con el fin de capturar patrones de comportamiento de la demanda a lo largo del tiempo.
- **Agregación de precipitaciones:** Se trabajó a nivel horario, asociando a cada registro de ocupación de transporte el promedio de precipitación de esa hora.
- **Discretización de niveles de precipitación:** Para algunos modelos exploratorios, se evaluó la creación de categorías de precipitación (por ejemplo, sin lluvia, lluvia leve, lluvia moderada, lluvia intensa) a partir de umbrales predefinidos.

5.4 Selección de variables relevantes

La variable objetivo seleccionada para el desarrollo del modelo fue la **ocupación promedio** de los ómnibus, medida en un intervalo horario determinado. El modelo predictivo fue diseñado para estimar este valor en función de las condiciones temporales y climáticas.

En cuanto a las variables de entrada, tras un análisis exploratorio de correlaciones y comportamientos observados, se seleccionaron las siguientes como insumos principales:

- Línea de ómnibus
- Día de la semana
- Hora cerrada
- Precipitación acumulada
- Categoría de precipitación (cuando se evaluó discretización)

Adicionalmente, durante la etapa de definición del conjunto de datos de entrenamiento, se utilizó la variable **ocupación máxima** como criterio de filtrado. Esta variable permitió identificar y seleccionar aquellas líneas de transporte que presentaban una mayor ocupación en al menos un momento del día, concentrando el análisis en aquellas rutas de mayor relevancia operativa.

5.5 Transformación temporal: nivel horario

Los datos originales de ocupación de ómnibus y de precipitaciones meteorológicas se encontraban registrados en intervalos de 15 minutos. Para alinear la granularidad de los datos con el objetivo del modelo (realizar predicciones horarias), se realizó un proceso de agregación.

En el caso del transporte, se calculó el promedio de ocupación de cada hora a partir de los cuatro registros correspondientes. Para las precipitaciones, se sumaron los valores de los cuatro registros de 15 minutos y luego se calculó el promedio entre las diferentes estaciones meteorológicas para cada hora.

Cada observación del conjunto de datos final representa así la ocupación promedio de una línea de ómnibus y el promedio de precipitación correspondiente en una hora específica. Esta decisión metodológica permitió una mejor consistencia entre los datos de entrada y la salida esperada del modelo.

5.6 Representación de variables cíclicas

Algunas variables temporales presentan una naturaleza cíclica inherente, como las horas del día, los meses del año y los días de la semana. Para capturar esta ciclicidad de manera efectiva, se aplicó una transformación utilizando funciones seno y coseno, técnica comúnmente utilizada en el preprocesamiento de variables temporales en modelos de machine learning. Concretamente, se crearon dos nuevas variables para cada atributo cíclico:

- Para la hora del día, se generaron las variables hora_sin y hora_cos.
- Para el mes del año, se generaron las variables mes_sin y mes_cos.
- Para el día de la semana, se generaron las variables dia_semana_sin y dia_semana_cos.

Estas variables permiten representar la naturaleza cíclica de la información de manera continua, evitando discontinuidades artificiales (por ejemplo, entre la hora 23 y la hora 0), y mejorando así la capacidad del modelo para aprender patrones temporales.

6. RESULTADOS

6.1 Análisis Exploratorio

6.1.1 Análisis de la ocupación promedio

El análisis exploratorio se realizó utilizando los datos correspondientes al período del 1 de enero de 2024 al 31 de diciembre de 2024. La información disponible incluía el detalle de la ocupación promedio de los ómnibus, desglosado por línea, fecha y franja horaria.

La ocupación promedio se calcula a partir de los registros del sistema STM (Sistema de Transporte Metropolitano), que contabiliza los pasajeros en el momento de su subida a las unidades. Sin embargo, no se dispone de información exacta sobre los descensos, dado que no existe un registro automático de bajada. Para estimar la ocupación promedio, se utilizan matrices de bajada y subida, que combinan la información histórica de ascensos con modelos de asignación de bajadas en función de las paradas y los horarios estimados.

Se detectaron registros en los datos de ocupación donde los valores eran menores a cero, producto de los cálculos de estimación mencionados. Ante esta situación, y siguiendo las indicaciones del organismo, se procedió a setear esos valores como cero.

Al analizar la distribución de la ocupación promedio, se puede observar una clara asimetría a la derecha, indicando que la mayoría de los viajes tienen una ocupación baja, mientras que existen algunos casos con valores significativamente más altos. La moda se encuentra en valores bajos, reflejando que la mayoría de los ómnibus viajan con ocupación reducida. Además, la cola larga hacia la derecha sugiere que en ciertos momentos o líneas específicas, la ocupación puede ser mucho mayor.

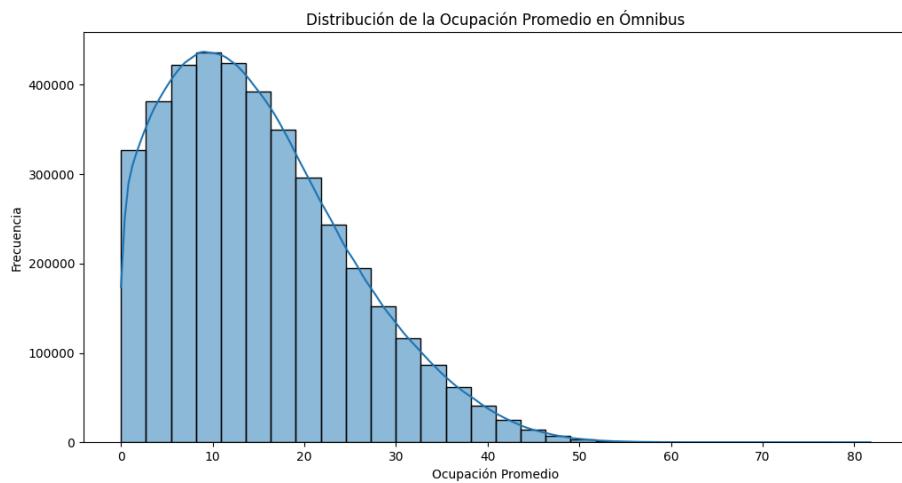


Figura 2- Distribución de la Ocupación Promedio en Ómnibus

6.1.2 Comparación entre días hábiles, sábados, domingos y feriados

El dataset proporciona una clasificación de los días en cinco categorías:

- Días hábiles
- Sábados
- Domingos
- Feriados laborables
- Feriados no laborables

Se observó que la **ocupación promedio** es más alta en días hábiles en comparación con fines de semana y feriados. Los datos mostraron mayor dispersión en días hábiles y feriados laborables, reflejando momentos de alta demanda (outliers de ocupación elevada). Por el contrario, los feriados no laborables presentaron la menor ocupación promedio y la menor dispersión, coincidiendo con la reducción en la frecuencia de servicio en esos días.

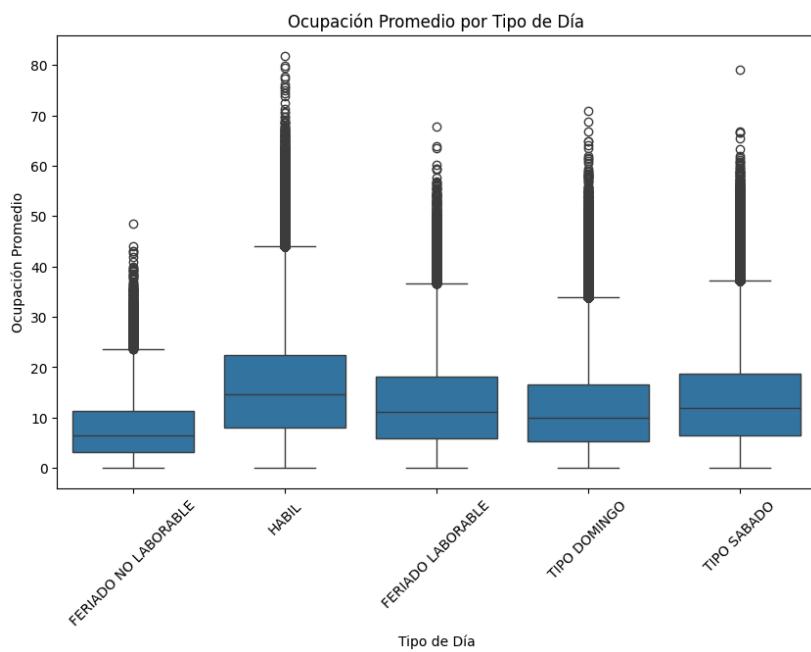


Figura 3- Boxplot Ocupación Promedio por Tipo de Día

6.1.3 Impacto de las precipitaciones en la ocupación

El gráfico presentado a continuación, junto con el análisis numérico correspondiente, refleja para cada día de la semana la ocupación promedio en escenarios con lluvia y sin lluvia, permitiendo identificar cómo varía la demanda del servicio de transporte frente a condiciones climáticas adversas.

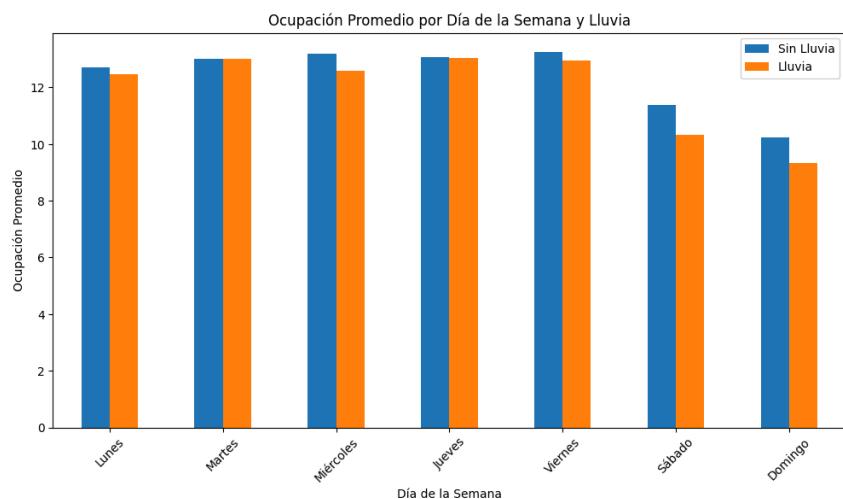


Figura 4- Ocupación por día de la Semana con y sin lluvia

En términos generales, se observa que la ocupación promedio sin lluvia tiende a ser ligeramente mayor que en días lluviosos.

Durante los días laborables (de lunes a viernes), los valores de ocupación promedio se sitúan en torno a los 13 pasajeros por ómnibus, con un rango que varía aproximadamente entre 12,7 (lunes) y 13,3 (viernes) en ausencia de precipitaciones, mientras que en días con lluvia los valores fluctúan entre 12,5 y 13,0 pasajeros.

El viernes se consolida como el día de mayor ocupación promedio, alcanzando aproximadamente 13,26 pasajeros sin lluvia y 12,94 con lluvia.

En contraste, los fines de semana muestran niveles de ocupación promedio más bajos.

Los sábados presentan una ocupación de aproximadamente 11,37 pasajeros en condiciones secas, disminuyendo a 10,32 pasajeros cuando hay lluvias.

Los domingos registran valores cercanos a 10,24 sin lluvia y alrededor de 9,32 bajo precipitaciones.

Esto implica diferencias de prácticamente un punto completo en la ocupación promedio durante los fines de semana, algo mayores que las observadas en los días laborables, donde las diferencias oscilan entre casi 0 y 0,60 pasajeros, según el día.

Una posible explicación para la menor variación durante los días laborales radica en la existencia de obligaciones, como el trabajo o los estudios, que incentivan a las personas a movilizarse incluso en condiciones climáticas desfavorables.

Por el contrario, durante los fines de semana, quienes pueden evitar salir tienden a permanecer en sus hogares si el clima no es favorable, lo que incrementa la brecha de ocupación entre escenarios de lluvia y ausencia de lluvia.

No obstante, la dinámica semanal sigue estando determinada por el comportamiento laboral típico de lunes a viernes, con un descenso significativo en la ocupación los sábados y domingos, y un efecto de la lluvia más marcado en esos días.

De cualquier forma, en términos absolutos, estas variaciones en la ocupación promedio se mantienen dentro de un rango moderado.

Si bien se observan tendencias claras, especialmente los fines de semana, la diferencia real entre la ocupación de los días laborables y de los fines de semana no resulta extremadamente pronunciada.

Finalmente, se analizó la correlación entre la ocupación promedio y las precipitaciones totales, obteniéndose un valor bajo de **-0,07**. Este resultado refuerza la idea de que la cantidad de lluvia no tiene un impacto directo fuerte sobre la ocupación general del transporte público en Montevideo.

6.1.4 Análisis de datos de precipitaciones

Tal como se detalló en la sección de Transformaciones de los Datos, se trabajó con el promedio de precipitaciones horarias obtenidas a partir de los registros de 15 estaciones meteorológicas distribuidas en Montevideo (las estaciones que tenían datos en este año).

A partir de estos datos, se analizó la cantidad de observaciones correspondientes a escenarios de lluvia y sin lluvia a lo largo de los meses del año 2024.

La siguiente tabla resume la distribución:

Mes	Sin lluvia	Con lluvia	Total
1	633	56	689
2	618	78	696
3	576	167	743
4	508	208	716
5	573	146	719
6	603	117	720
7	693	51	744
8	586	158	744
9	641	79	720
10	684	60	744
11	609	110	719
12	616	116	732

Del análisis mensual se observa que, durante 2024, los meses de marzo, abril y agosto registraron un mayor número de observaciones con lluvia, mientras que los meses de enero, julio y octubre presentaron la menor proporción de registros lluviosos.

Estos resultados sugieren una tendencia temporal en la distribución de las precipitaciones a lo largo del año analizado, con una mayor frecuencia de eventos de lluvia en los meses correspondientes al otoño y a finales del invierno.

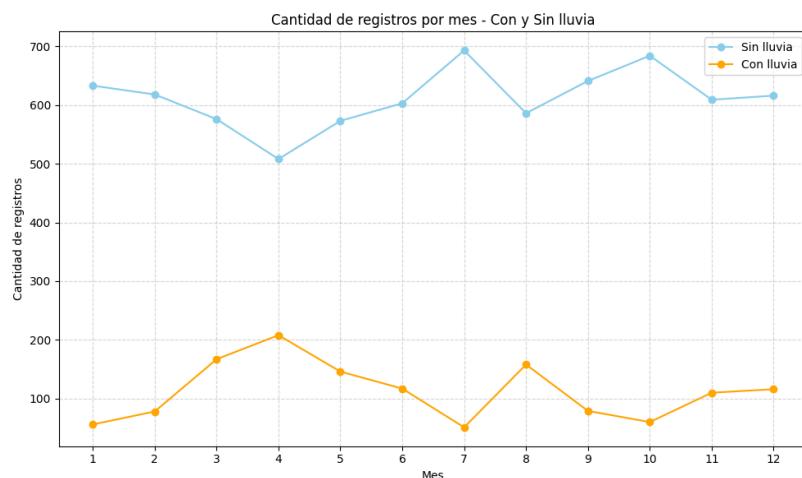


Figura 5- Cantidad de registros por mes: Con y Sin lluvia

6.1.5 Observaciones adicionales del análisis exploratorio

Durante el análisis exploratorio de los datos, surgieron algunas observaciones relevantes que complementan los hallazgos ya mencionados.

En primer lugar, se identificó que los valores de ocupación promedio presentan una distribución altamente asimétrica hacia la derecha. Esto indica que la mayoría de los viajes de ómnibus tienen baja ocupación, con un número relativamente reducido de casos donde la ocupación promedio es significativamente más elevada. Esta característica sugiere la presencia de valores atípicos (outliers) en la distribución, correspondientes a momentos o recorridos específicos de alta demanda.

Por otra parte, al analizar la relación entre la cantidad de precipitaciones y la ocupación promedio, se observó que el coeficiente de correlación entre ambas variables resultó ser de aproximadamente -0,07. Este valor tan bajo refuerza la idea de que, en términos generales, la presencia de lluvia no tiene un impacto directo fuerte sobre la cantidad de pasajeros transportados.

Si bien se identificaron ligeras diferencias en la ocupación promedio en escenarios con y sin lluvia (especialmente durante los fines de semana), estas variaciones no son lo suficientemente pronunciadas como para afirmar una fuerte dependencia climática en el comportamiento general de uso del transporte público en Montevideo.

Por último, cabe destacar que el conjunto de datos, tanto de transporte como de precipitaciones, muestra una cobertura adecuada durante el año analizado, sin ausencias de información que comprometieran los análisis posteriores o la construcción de los modelos predictivos.

6.2 Modelado Inicial

6.2.1 Descripción inicial del proceso

En una primera instancia, se trabajó con datos correspondientes a los meses de Setiembre, Octubre y Noviembre de 2024, con el objetivo de realizar un análisis exploratorio preliminar y probar enfoques iniciales de modelado.

Durante esta etapa se realizaron las primeras pruebas utilizando dos metodologías diferentes:

- Análisis de series temporales mediante el modelo SARIMAX.
- Construcción de un modelo de red neuronal MLP (Multi-Layer Perceptron) sencillo.

La disponibilidad de solo tres meses de datos resultó rápidamente insuficiente para obtener conclusiones robustas, motivo por el cual se procedió a solicitar y utilizar la totalidad del año 2024 para las etapas siguientes.

En estas primeras pruebas, se planteó el problema como una predicción de ocupación promedio diaria. Posteriormente, y a raíz de observaciones realizadas sobre la variabilidad intradía de los datos, se decidió reformular el objetivo de predicción hacia una ventana horaria, transformando adecuadamente los datasets para trabajar con registros agregados por hora.

Una vez incorporados los datos completos del año 2024 y definida la nueva unidad temporal, se reestructuró el análisis y el modelado, como se detalla en las secciones siguientes.

6.2.2 Modelado de series temporales SARIMAX

Como primera aproximación, se construyó un modelo basado en SARIMAX para la predicción de la ocupación promedio diaria.

Se realizaron diferentes análisis exploratorios de series temporales:

- Gráfico de autocorrelación (ACF):
Se observó que el lag 0 presenta un valor de 1, como es habitual, dado que representa la autocorrelación consigo mismo. En los lags siguientes, la mayoría de los valores se mantuvieron dentro de la franja azul de confianza, indicando una falta de autocorrelación significativa. Aunque algunos picos superaron marginalmente los límites, no se identificaron tendencias ni ciclos claros en la serie.
- Gráfico de autocorrelación parcial (PACF):
Se corroboró la ausencia de autocorrelaciones significativas en los primeros 40 lags, reforzando la idea de que

el comportamiento de la serie es altamente aleatorio, probablemente influenciado por factores exógenos como las precipitaciones, cuya dinámica no es lineal ni fácilmente predecible a partir de registros pasados.

Posteriormente, se ajustó un modelo SARIMAX incorporando como variables exógenas:

- Precipitación instantánea promedio (instant_precipitation_avg).
- Precipitación acumulada 15 minutos (ac_15_avg).
- Precipitación acumulada 24 horas (ac_1440_avg).
- Tipo de día (hábil, sábado, domingo, feriado laborable, feriado no laborable).

El resumen estadístico del ajuste mostró que las variables de precipitaciones no eran estadísticamente significativas ($p>0.05$), mientras que las variables relacionadas al tipo de día resultaron altamente significativas ($p\approx 0.000$). Esto sugiere que el tipo de día tiene un impacto más importante en la ocupación promedio que las precipitaciones.

Al comparar las predicciones del modelo con los datos reales, se observó que, si bien la tendencia general era capturada, el modelo presentaba dificultades para predecir los picos de alta ocupación y mantenía errores sistemáticos, especialmente en días atípicos.

Estas limitaciones llevaron a reconsiderar el enfoque.

6.2.3 Primer modelo de red neuronal simple (MLP)

En paralelo al ajuste del modelo SARIMAX, se desarrolló una red neuronal simple de tipo MLP (Multi-Layer Perceptron) como otra aproximación inicial.

Para esta primera etapa, se entrenó un modelo por líneas de ómnibus, seleccionando cinco líneas aleatorias y construyendo un modelo independiente para cada una.

La arquitectura del MLP se definió considerando las siguientes características:

- Número de capas: 4 capas densas.
- Dimensiones: 64 neuronas en la primera capa, 32 en la segunda, 16 en la tercera, y 1 en la capa de salida.
- Funciones de activación: ReLU en las capas ocultas y función lineal en la capa de salida.
- Regularización: Dropout de 0.2 en las capas intermedias.
- Optimizador: Adam.
- Función de pérdida: (MSE, Mean Squared Error).
- Métrica de evaluación: MAE (Mean Absolute Error).
- Parámetros de entrenamiento: batch_size=32, epochs=500, learning_rate=0.001.

Para la validación, se utilizó un validation_split=0.2, es decir, un 20% de los datos de entrenamiento fue reservado para validación interna durante el entrenamiento.

Adicionalmente, el dataset de test se conformó seleccionando muestras aleatorias del total de los datos disponibles, garantizando la evaluación del modelo en observaciones no vistas durante el entrenamiento.

Esta configuración básica se inspiró en el enfoque presentado en el artículo científico [4], donde se aplican redes neuronales para predecir la demanda de transporte público considerando condiciones meteorológicas.

Los resultados obtenidos mostraron que, en los gráficos de entrenamiento, se observó que el modelo convergía a valores estables de error antes de completar las 500 epochs, sugiriendo la posibilidad de optimizar el entrenamiento reduciendo la cantidad de iteraciones. En los gráficos de comparación entre valores reales y predichos, el modelo mostró un comportamiento aceptable, capturando las tendencias generales de la ocupación aunque con cierta pérdida de precisión en los picos extremos.

Se concluyó que el modelo MLP simple era una aproximación prometedora, pero que requería ajustes en su arquitectura y en los datos de entrada para mejorar su capacidad de generalización.

6.2.4 Conclusiones de la primera etapa

Finalizada esta primera instancia de pruebas, se decidió avanzar en el trabajo sobre la base de redes neuronales MLP, dado que:

- Presentaban una mejor adaptación a la complejidad del problema que los modelos SARIMAX.
- Permitían incorporar múltiples variables simultáneamente (precipitaciones, tipo de día, horario, etc.).
- Resultaban más adecuados para capturar relaciones no lineales entre las variables de entrada y la ocupación promedio.

Además, se definió la reformulación del objetivo de predicción a nivel horario, con el fin de reflejar de manera más realista el comportamiento de la demanda de transporte público, considerando las variaciones intradía.

A partir de allí, se planificaron ajustes en la estructura del modelo MLP, en el preprocesamiento de datos y en los criterios de evaluación, los cuales se describen en las secciones siguientes.

6.3 Construcción del modelo MLP mejorado

6.3.1 Decisiones de diseño para la construcción del modelo

Previo a la construcción final del modelo predictivo, se tomaron una serie de decisiones clave que permitieron delimitar correctamente el alcance del trabajo y ajustar el enfoque de modelado a las necesidades específicas del problema.

Selección de líneas de ómnibus

La primera definición importante fue determinar sobre cuántas y cuáles líneas de ómnibus se entrenaría el modelo.

Aunque los datos disponibles incluían todas las líneas de la ciudad, se consideró que, debido a limitaciones prácticas de alcance y capacidad computacional, no era viable abordar el conjunto completo.

Se decidió trabajar finalmente con seis líneas seleccionadas siguiendo una estrategia de representatividad de diferentes casuísticas:

- Se identificaron las líneas con mayor ocupación promedio, resultando las líneas con código 504 y 20.
- Se calcularon las líneas con mayores valores máximos de ocupación puntual, seleccionándose los códigos 497 y 2, que representan casos con picos extremos en ciertos horarios o días.
- Se seleccionaron además dos líneas adicionales de manera aleatoria, eligiéndose los códigos 484 (línea que ya se venía utilizando en análisis previos) y 23.

Inicialmente se evaluó también la posibilidad de incluir líneas con menor ocupación promedio, pero se descartó esta opción dado que estas líneas correspondían a servicios especiales o esporádicos (como refuerzos en eventos), cuya dinámica particular podría introducir ruido innecesario en el modelo.

Posteriormente, mediante el cruce con la tabla auxiliar de líneas, se obtuvo la correspondencia de la línea real:

Código de línea	Línea	Descripción
504	124	Mayor promedio
20	546	Mayor promedio
484	L12	Random
23	100	Random
497	G	Mayor máximo de ocupación
2	404	Mayor máximo de ocupación

Definición de la ventana temporal de predicción

La segunda gran definición fue establecer la ventana temporal sobre la cual se realizaría la predicción.

Inicialmente, se había trabajado con predicciones a nivel de día completo (ventanas de 24 horas).

Sin embargo, se constató que esta escala no era adecuada para reflejar la realidad del transporte urbano, donde la ocupación puede variar fuertemente a lo largo de las horas del día, especialmente en horarios pico.

Se decidió, por tanto, trabajar con ventanas de una hora, lo cual permitía capturar de manera más precisa las variaciones horarias.

Aunque se consideró la posibilidad de reducir aún más la ventana (por ejemplo, a cada 15 minutos), se optó finalmente por mantener la escala horaria, balanceando la necesidad de detalle temporal con la capacidad de procesamiento disponible y el alcance de este trabajo.

Definición sobre el uso de modelos recursivos

Otra definición técnica relevante fue decidir si el modelo debía ser recursivo, es decir, si debía utilizar como entrada sus propias predicciones anteriores.

Dado que la ocupación de los ómnibus puede variar abruptamente incluso en períodos cortos (especialmente en horarios pico), se concluyó que un modelo recursivo sólo tendría sentido en ventanas de tiempo muy pequeñas (por ejemplo, 5 o 15 minutos).

Al trabajar finalmente con ventanas de una hora, se determinó que no era conveniente utilizar un modelo recursivo, ya que incorporar el valor predicho de la hora anterior podría introducir ruido y dificultar la adaptación a cambios abruptos de demanda.

Definición del esquema de división de datos: entrenamiento, validación y test

Por último, se definió el esquema de división del conjunto de datos para las etapas de entrenamiento, validación y prueba.

En una primera instancia, se había planteado realizar una división aleatoria.

Sin embargo, tras un análisis más detallado de la distribución de registros con y sin lluvia por mes (analizado en el apartado de exploración de datos climáticos), se tomó la decisión de dividir los datos de forma temporal.

El mes de abril fue seleccionado para conformar el conjunto de test, dado que presentaba una distribución relativamente balanceada entre registros de días con lluvia y sin lluvia. El resto de los meses del año fue utilizado para entrenamiento y validación del modelo.

Se consideró que esta estrategia era suficiente para evaluar el modelo en escenarios diversos sin necesidad de aplicar técnicas de data augmentation u otros métodos de balanceo de clases.

6.3.2 Reformulación del enfoque y preparación de los datos

A partir de la decisión de realizar predicciones a nivel horario, se realizaron las transformaciones necesarias para adaptar los datos a esta nueva unidad temporal.

Cada observación pasó a representar la ocupación promedio de una línea de ómnibus en una hora cerrada específica, vinculada además con la información sobre las precipitaciones correspondientes a ese período.

Las variables utilizadas como insumos del modelo incluyeron:

- Hora cerrada (de 0 a 23).
- Día de la semana (lunes a domingo).
- Mes (enero a diciembre).
- Precipitación acumulada en la hora.
- Tipo de día (hábil, sábado, domingo, feriado laborable, feriado no laborable), representado mediante variables binarias.

Adicionalmente, para capturar de forma adecuada la naturaleza cíclica de ciertas variables temporales, se aplicaron transformaciones seno y coseno a: hora cerrada, día de la semana y mes del año. Esto permitió representar adecuadamente la continuidad circular de estos atributos, evitando discontinuidades artificiales entre extremos (por ejemplo, entre la hora 23 y la hora 0, o entre diciembre y enero).

La variable objetivo definida para el modelo fue la ocupación promedio horaria de las líneas seleccionadas.

Una vez estructurado el conjunto de datos, se realizó la división entre entrenamiento, validación y test, reservando el mes de abril de 2024 para evaluación final y utilizando el resto de los datos para entrenamiento y validación mediante una partición interna.

Adicionalmente, se aplicó una normalización mediante StandardScaler sobre la variable de precipitación acumulada en la hora, estandarizando sus valores a una distribución con media cero y desviación estándar uno. Esta transformación se implementó con el objetivo de homogeneizar la escala de las variables de entrada, facilitando el proceso de entrenamiento del modelo MLP y mejorando la estabilidad de la optimización.

6.3.3 Primer modelo basado en la arquitectura del paper

En una primera instancia, el modelo MLP se construyó replicando la arquitectura propuesta en el estudio realizado sobre la red de transporte de los Países Bajos.

Esta arquitectura inicial se definió de la siguiente manera:

- Número de capas: 4 capas densas.
- Dimensiones: 64 neuronas en la primera capa, 32 en la segunda, 16 en la tercera, y 1 en la capa de salida.
- Funciones de activación: ReLU en las capas ocultas y función lineal en la capa de salida.
- Regularización: Dropout de 0.2 en las capas intermedias.
- Optimizador: Adam.
- Función de pérdida: (MSE, Mean Squared Error).
- Métrica de evaluación: MAE (Mean Absolute Error).
- Parámetros de entrenamiento: batch_size = 32, learning_rate = 0,001, epochs = 500.

Además, se mejoró la flexibilidad del modelo adaptando la salida para que contemplara simultáneamente todas las líneas seleccionadas, permitiendo una ampliación futura sin necesidad de modificar la arquitectura.

El conjunto de datos de entrenamiento abarcó todos los meses del año 2024, exceptuando abril, que fue reservado íntegramente para testeo.

Dentro del conjunto de entrenamiento, se reservó un 20% de los datos para validación (validation_split=0.2).

La evaluación inicial de este primer modelo arrojó las siguientes métricas de error (MAE y MSE) por línea de ómnibus, diferenciando entre condiciones de lluvia y sin lluvia:

Código de Línea	Línea	MAE (lluvia)	MAE (sin lluvia)	MSE (lluvia)	MSE (sin lluvia)
2	404	2.118	2.582	8.558	12.717
20	546	3.822	3.349	28.296	23.799
23	100	2.453	2.290	11.967	9.328
484	L12	1.914	1.722	9.697	9.305
497	G	3.202	3.113	14.755	17.322
504	404	1.472	1.587	15.718	16.331

Estos resultados mostraron un desempeño aceptable en términos generales, aunque revelaron la necesidad de mejorar la capacidad del modelo para capturar los picos horarios, especialmente bajo condiciones de lluvia y en líneas con alta variabilidad de ocupación.

6.3.4 Optimización de hiperparámetros y construcción del modelo mejorado

Ante las oportunidades de mejora identificadas, se implementaron dos estrategias principales:

- Early Stopping, para detener automáticamente el entrenamiento cuando no se observaban mejoras en la métrica de validación.
- Optimización Bayesiana de hiperparámetros, que permitió buscar combinaciones más eficientes de:
 - Número de neuronas por capa.
 - Tasas de Dropout.
 - Tasa de aprendizaje.

El proceso de optimización derivó en cambios significativos en la arquitectura del modelo, que quedó definida finalmente de la siguiente manera:

- Número de capas densas: 4.
- Dimensiones de las capas: 67, 49, 100 y 1 neurona respectivamente.
- Funciones de activación: ReLU en las capas ocultas y lineal en la capa de salida.
- Regularización: Dropout capa 1 de 0,4354.
- Regularización: Dropout capa 2 de 0,1784.
- Regularización: Dropout capa 3 de 0,1181.
- Optimizador: Adam, con tasa de aprendizaje reducida a 0,0033.

Esta nueva arquitectura buscó mejorar la capacidad del modelo para capturar patrones no lineales más complejos, y controlar el riesgo de sobreajuste asociado al incremento en la profundidad de la red.

Tras completar el proceso de optimización, se realizó una nueva evaluación sobre el conjunto de test, obteniéndose las siguientes métricas:

Código de Línea	Línea	MAE (lluvia)	MAE (sin lluvia)	MSE (lluvia)	MSE (sin lluvia)
2	404	2.160	2.478	7.811	11.525
20	546	3.571	3.132	24.677	21.416
23	100	2.410	2.171	11.636	8.130
484	L12	1.957	1.806	9.648	9.534
497	G	3.074	2.857	14.074	14.666
504	404	2.135	1.586	21.024	15.913

6.3.5 Resultados finales del modelo optimizado

Una vez optimizado el modelo MLP mediante técnicas de ajuste de hiperparámetros y early stopping, se procedió a la evaluación final utilizando el conjunto de test correspondiente al mes de abril de 2024.

El desempeño global del modelo se evaluó diferenciando entre días con y sin lluvia.

Se obtuvo un MAE promedio de 2,55 pasajeros en días sin lluvia y un MAE de 2,34 pasajeros en días con lluvia.

Estas diferencias son mínimas, lo cual indica que el modelo se adapta adecuadamente a ambos escenarios climáticos sin una pérdida de desempeño relevante.

Teniendo en cuenta que la ocupación promedio se encuentra en torno a 10–15 pasajeros por ómnibus, un error de aproximadamente 2 a 2,5 pasajeros es considerado aceptable para fines de planificación operativa.

A continuación, se presentan las visualizaciones que ilustran el comportamiento del modelo.

Comparación global: ocupación real vs. predicha

La Figura 6 muestra, mediante boxplots y dispersión de puntos, la comparación de los valores reales y predichos en días con y sin lluvia.

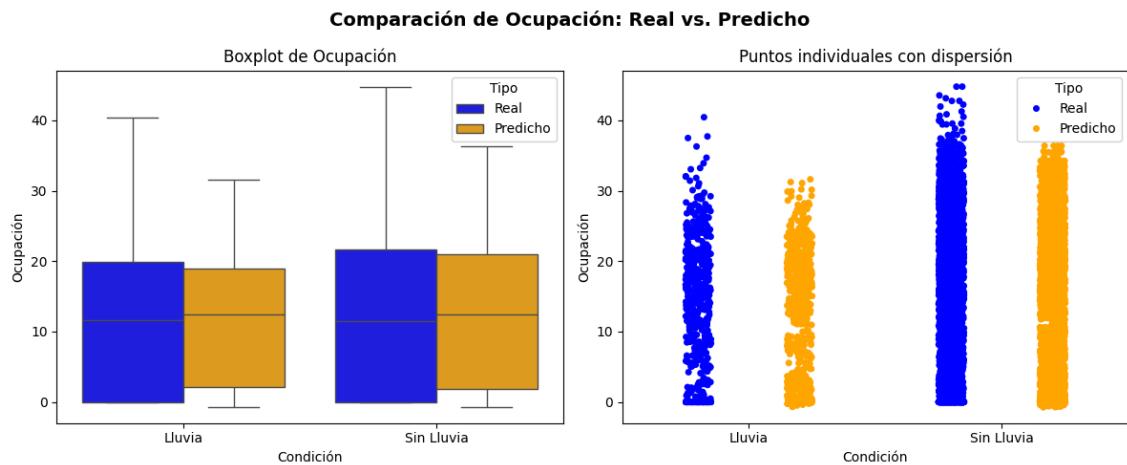


Figura 6- Comparación de ocupación real vs. predicha por condición climática

El modelo logra seguir de manera muy precisa la tendencia central de los datos en ambos escenarios. Se observa una ligera tendencia a suavizar los valores extremos, especialmente en días secos donde existen picos más pronunciados de ocupación. La diferencia entre mediana, cuartiles y amplitud entre ocupación real y predicha es muy pequeña, indicando una excelente generalización.

Distribución de ocupación por hora del día

La Figura 7 muestra la distribución de ocupación a lo largo de las 24 horas del día durante el mes de abril.

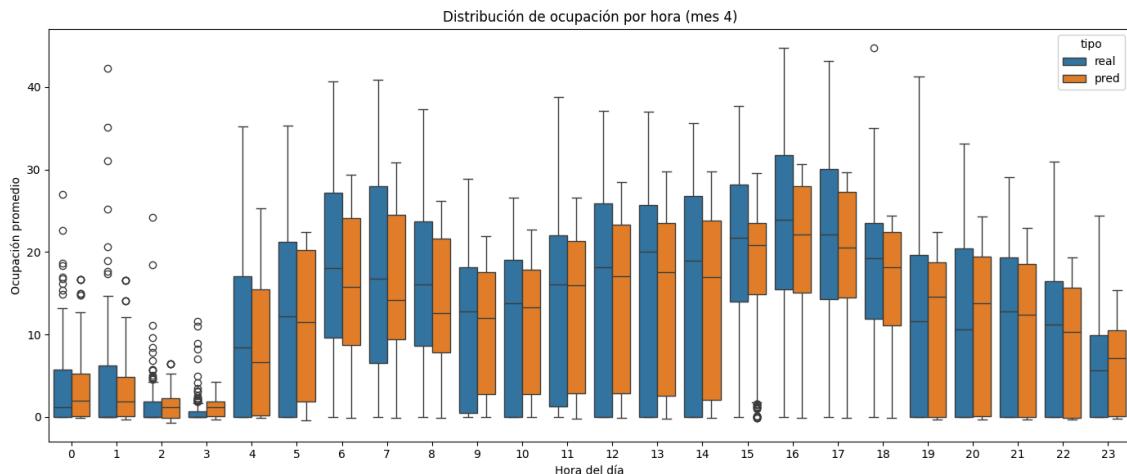


Figura 7- Distribución horaria de ocupación real y predicha

El modelo reproduce correctamente los picos horarios de mayor demanda, tanto en la mañana como en la tarde. Se capturan adecuadamente las horas de menor flujo de pasajeros. Nuevamente, la principal diferencia se manifiesta en los máximos extremos, donde la predicción tiende a ser más conservadora.

Predicción horaria en días concretos: líneas específicas

Con el objetivo de analizar el comportamiento del modelo en situaciones concretas, se seleccionaron dos líneas de ómnibus y se graficaron las predicciones por hora para un día sin lluvia y un día con lluvia.

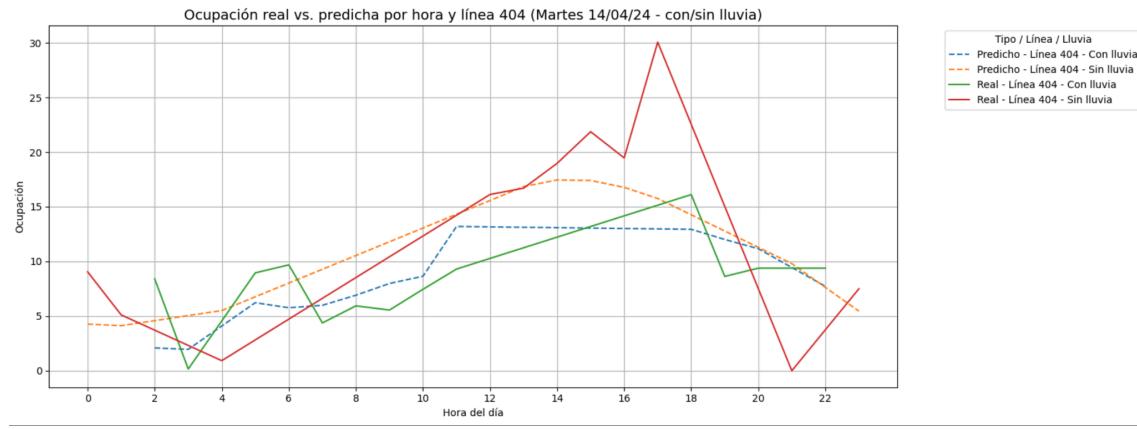


Figura 8- Predicción horaria para línea 404

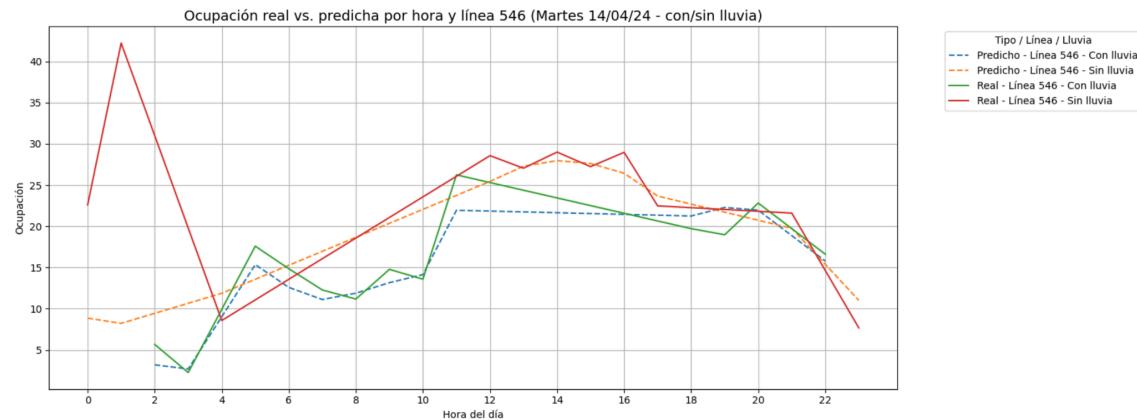


Figura 9- Predicción horaria para línea 546

En ambos casos, el modelo sigue de manera adecuada la evolución diaria de la ocupación, identificando correctamente los horarios de mayor y menor afluencia. En los días sin lluvia, donde existen mayores niveles de ocupación, el modelo tiende a suavizar los máximos, mientras que en los días de lluvia la predicción es aún más precisa debido a la menor dispersión de datos.

La comparación visual reafirma que el modelo es robusto en términos de tendencias generales, aunque conserva cierta limitación para capturar valores extremos.

Conclusiones finales de resultados

En resumen, el modelo final demuestra:

- Alta capacidad para predecir ocupación horaria en diferentes condiciones climáticas.
- Error absoluto promedio muy bajo considerando el contexto de ocupación típico de Montevideo.
- Reproducción correcta de los patrones horarios de movilidad, capturando los momentos de mayor y menor uso del transporte.
- Una pequeña tendencia a suavizar picos extremos, inherente a las características de la arquitectura utilizada.

Estos resultados consolidan la robustez y aplicabilidad del modelo, proporcionando una herramienta predictiva efectiva para apoyar la planificación y optimización del sistema de transporte público.

Aplicando este modelo a todas las líneas de ómnibus la Intendencia de Montevideo en su Departamento de Sistema Metropolitano, podrá usar el mismo para predecir la ocupación que tendrán los ómnibus para así poder gestionar y exigir a las empresas un mejor servicios en las distintas líneas, ya sea con más frecuencias para evitar servicios ómnibus con exceso de pasajeros y esperas en las paradas, ayudando así a un mejor Sistema de Transporte.

6.4 Evaluación general y sugerencias de mejoras

6.4.1 Limitaciones observadas

Si bien el modelo desarrollado mostró un desempeño robusto en la predicción de la ocupación promedio horaria en diferentes condiciones climáticas, se identificaron algunas limitaciones inherentes al enfoque adoptado.

En particular, se observó una ligera tendencia del modelo a suavizar los valores extremos de ocupación, especialmente en horarios pico o en líneas caracterizadas por una alta variabilidad. Este comportamiento es habitual en arquitecturas de tipo MLP utilizadas para tareas de regresión, donde la predicción tiende a aproximarse a los valores promedio, limitando su capacidad para capturar eventos de demanda súbita.

Además, el entrenamiento realizado sobre un subconjunto reducido de líneas de ómnibus, y utilizando datos de un único año, acotó el rango de situaciones climáticas y de comportamiento de la demanda que el modelo pudo aprender.

6.4.2 Sugerencias de mejoras futuras

A partir de los resultados obtenidos y del análisis de las limitaciones observadas, se proponen las siguientes líneas de mejora para futuros desarrollos:

Ampliación del alcance del modelo:

Entrenar el modelo utilizando la totalidad de las líneas de ómnibus disponibles permitiría capturar una mayor diversidad de patrones de ocupación y enriquecer la capacidad de generalización del sistema.

Esto requeriría contar con una mayor capacidad de procesamiento computacional para gestionar el volumen de datos incrementado.

Utilización de históricos de varios años:

Si bien en este proyecto se trabajó exclusivamente con datos correspondientes al año 2024, la Intendencia de Montevideo cuenta con registros históricos de ocupación y precipitaciones de más de una década.

Incorporar esta información permitiría entrenar modelos más robustos, capaces de aprender patrones de largo plazo y adaptarse mejor a variaciones estacionales o a cambios estructurales en el comportamiento de los usuarios.

Mejorar la predicción de extremos de ocupación:

Desarrollar estrategias específicas para capturar eventos de alta ocupación de corta duración (como los picos en horarios punta) podría mejorar aún más la precisión del modelo en escenarios críticos.

Esto podría lograrse mediante el ajuste de la función de pérdida para penalizar errores en valores extremos, o explorando arquitecturas de redes más adecuadas para series temporales complejas, como redes recurrentes (RNN) o LSTM.

Estas mejoras no sólo aumentarían la precisión del modelo, sino que también ampliarían su aplicabilidad como herramienta de soporte a la planificación y gestión dinámica del transporte público en Montevideo.

7. ACCESO A LOS RECURSOS DEL PROYECTO

Con el objetivo de facilitar la revisión y replicación de los experimentos realizados, se pone a disposición el repositorio de trabajo utilizado en este proyecto, el cual contiene:

- Notebooks de procesamiento de datos.
- Scripts de entrenamiento de modelos.
- Resultados y gráficos generados.
- Documentación adicional.

El repositorio se encuentra disponible en la siguiente dirección:

trabajo-final-maestria-bigdata-predicción-afluencia-pasajeros-mejorar-transporte-montevideo – GitHub Repository

URL:

<https://github.com/ChristianVera49/trabajo-final-maestria-bigdata-prediccion-afluencia-pasajeros-mejorar-transporte-montevideo.git>

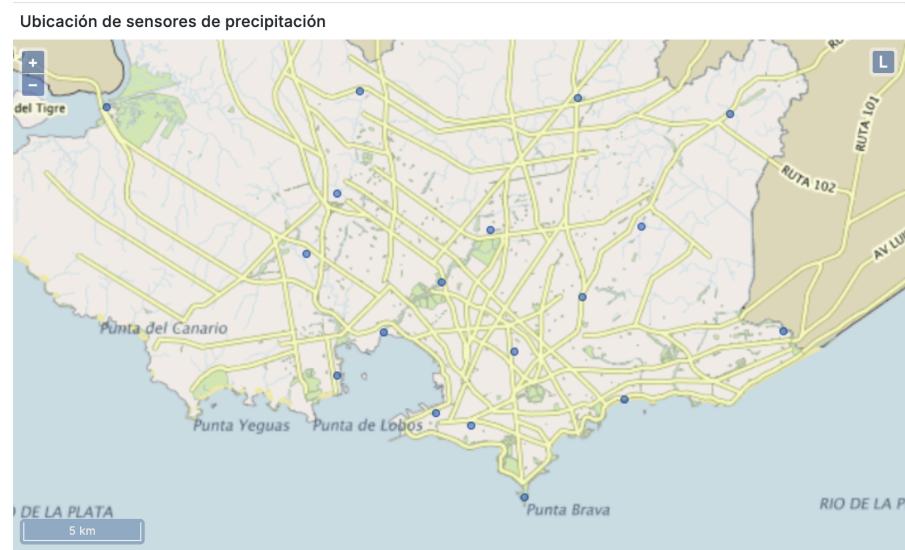
8. REFERENCIAS BIBLIOGRÁFICAS

- [1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson, 2016.
- [2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [3] Ministerio de Transporte y Obras Públicas (MTOP), "URB01 - Medio de transporte para ir a trabajar por departamento," Observatorio Nacional de Infraestructura, Transporte y Logística, Montevideo, Uruguay. [En línea]. Disponible en: https://observatorio.mtop.gub.uy/transporte_urbano.php
- [4] D. Varela and O. Olid, "Transporte público urbano: ¿qué se puede aprender de otros proyectos en la región?", *El Observador*, Montevideo, Uruguay, 24-Mar-2025. [Online]. Available: <https://www.elobservador.com.uy/opinion/transporte-publico-urbano-que-se-puede-aprender-otros-proyectos-la-region-n5991188>
- [5] Z. H. Farahmand, K. Gkiotsalitis, y K. T. Geurs, "Predicting bus ridership based on the weather conditions using deep learning algorithms," *Transportation Research Interdisciplinary Perspectives*, vol. 19, 2023, p. 100833.
- [6] OpenAI, "ChatGPT," OpenAI, San Francisco, CA, USA, 2024. [En línea]. Disponible en: <https://openai.com/chatgpt>. [Accedido: abril de 2025]

ANEXO 1: Estaciones meteorológicas y sus datos

La Intendencia de Montevideo cuenta con 18 estaciones meteorológicas en Montevideo.

El siguiente mapa muestra sus ubicaciones



Dichas estaciones son: Capurro, Carrasco, Casavalle, Centro, Cerro, Chacarita, Ciudad Vieja, Colón, Curva de Maroñas, La Blanqueada, La Esperanza, Lezica, Malvín, Manga, Paso de la Arena, Prado, Punta Carreta y Santiago Vázquez

Esta información es pública en los tableros del Observatorio Ambiental de la Intendencia de Montevideo en tiempo real, y también se encuentra su histórico en el portal de Datos Abierto (CKAN) de la Intendencia de Montevideo y AGESIC. A ellos se puede acceder desde los siguientes link

Observatorio Ambiental, Precipitaciones:

https://graf.montevideo.gub.uy/graf/d/i0P_HCDnz/precipitaciones-y-ubicacion-geografica?orgId=2&from=now-24h&to=now&timezone=browser

Datos Abierto (CKAN) de la Intendencia de Montevideo:

<https://ckan.montevideo.gub.uy/dataset/informacion-pluviometrica>

Como se dijo en el documento si bien existen estaciones en todo Montevideo, para este trabajo se utilizó el promedio de las mismas dado que la geografía de Montevideo es muy chica para pensar grandes cambios de precipitaciones de un extremo de la ciudad a otro.

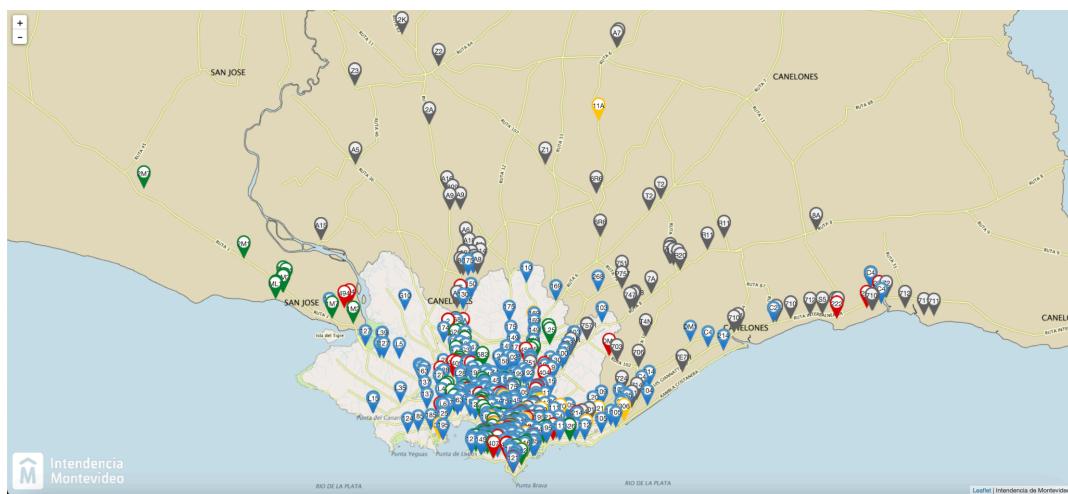
Si bien estos datos son abiertos a la comunidad, se utilizaron datos procesados que fueron brindados por el equipo técnico de la Intendencia de Montevideo.

La estructura de los datos brindados por la I.M. son:

nombre,
fecha,instant_precipitation,ac_10,ac_15,ac_20,ac_25,ac_30,ac_60,ac_120,ac_180,ac_360,ac_720,ac_1080,ac_1440,ani
o_mes

ANEXO 2: Sistema de Transporte Metropolitano y líneas de ómnibus elegidos para el modelo

La Intendencia de Montevideo tiene su página donde muestra los ómnibus en tiempo real como se observa a continuación

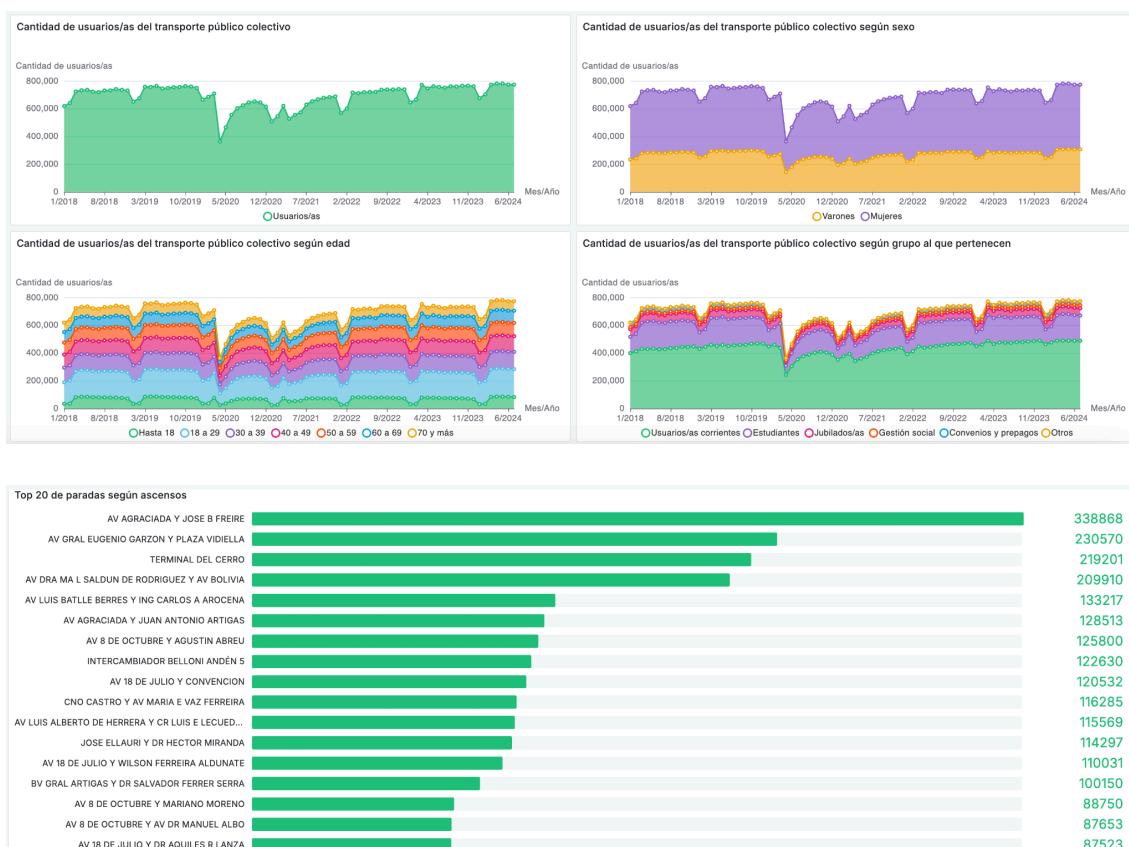


Se puede observar su recorrido, tiempo, líneas etc, incluso seguir un ómnibus que deseé ver

El link a la Web en tiempo real es: <https://www.montevideo.gub.uy/buses/mapaBuses.html>

También cuenta con un tablero del Observatorio del STM de la Intendencia de Montevideo donde se pueden observar datos importantes como cantidad de usuarios totales, separados por rango de edad, paradas más pobladas, etc.

Ejemplo de los tableros que se pueden encontrar en la misma:



Link al tablero del Observatorio del STM de la Intendencia de Montevideo:
<https://graf.montevideo.gub.uy/graf/d/fWBNRLUVk/viajes-stm?orgId=2>

Por último también mencionar que en su Portal de Datos Abiertos de la Intendencia de Montevideo, se encuentran varios datos abiertos vinculados al STM que pueden ser muy interesantes

Portal de Datos Abiertos

Inicio / Conjuntos de datos

Publicadores

- Departamento de Mov... 2
- Tránsito y transporte 2
- Tecnología de la In... 1

Categorías

- Datos abiertos 5
- Movilidad 4
- Transparencia Insti... 1

Etiquetas

- STM 5
- Movilidad 3
- destino 1
- Horarios 1
- origen 1
- Paradas de ómnibus 1
- Recarga 1

Formatos

- TXT 3
- CSV 2
- CSV ZIP 2
- csv zip 1
- ZIP 1

Conjuntos de datos

STM

5 conjuntos de datos encontrados para "STM"

Ordenar por: Relevancia

Locales de recarga del STM
Lista de los locales de recarga del STM, sus coordenadas geográficas e información vinculada a cada local, tal como se muestra en el sitio web de la IM . Más información Mapa...

Horarios de ómnibus urbanos, por parada - STM
Este juego de datos contiene los horarios de ómnibus del transporte colectivo urbano de Montevideo para cada una de las paradas. Estos son los horarios teóricos estimados en los...

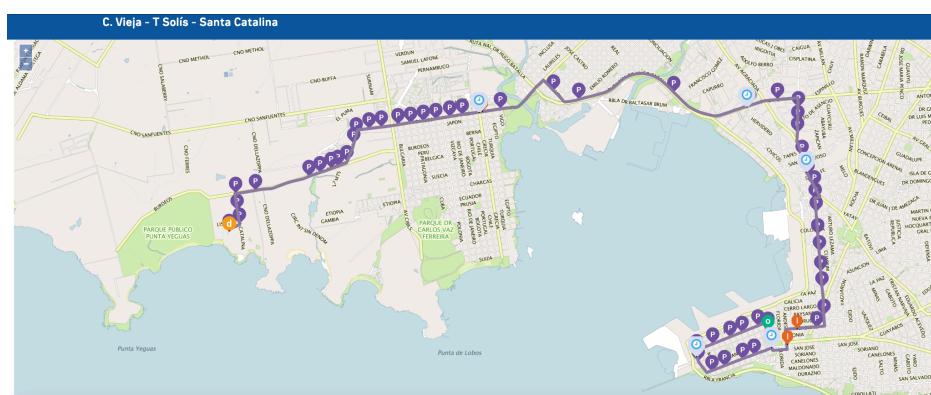
Viajes realizados en los ómnibus del Sistema de Transporte Metropolitano - STM
Este conjunto de datos contiene todos los viajes realizados en las líneas de transporte colectivo urbano de Montevideo, por empresa operadora, línea, variante, día y hora....

Encuesta Origen-Destino Montevideo
Se ponen a disposición los datos de las encuestas origen destino realizadas en los años 2009 y 2016. Los juegos de datos consisten en un archivo comprimido dentro del cual se...

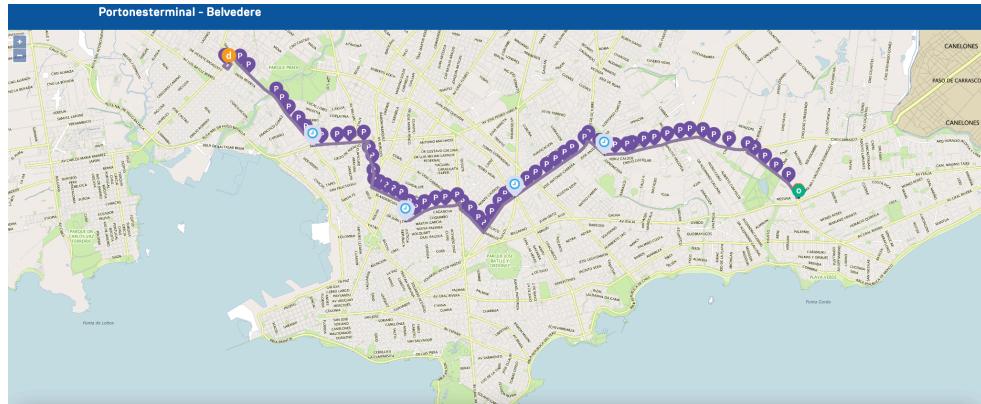
El link al Portal de Datos Abiertos de la Intendencia de Montevideo es: <https://ckan.montevideo.gub.uy/dataset/>

Ahora se procede a mostrar las líneas elegidas y explicada en el documento las líneas elegidas fueron: 124, 546, L12, 100, G, 404 . A continuación se muestra el recorrido, con sus paradas, de cada una de las líneas seleccionada:

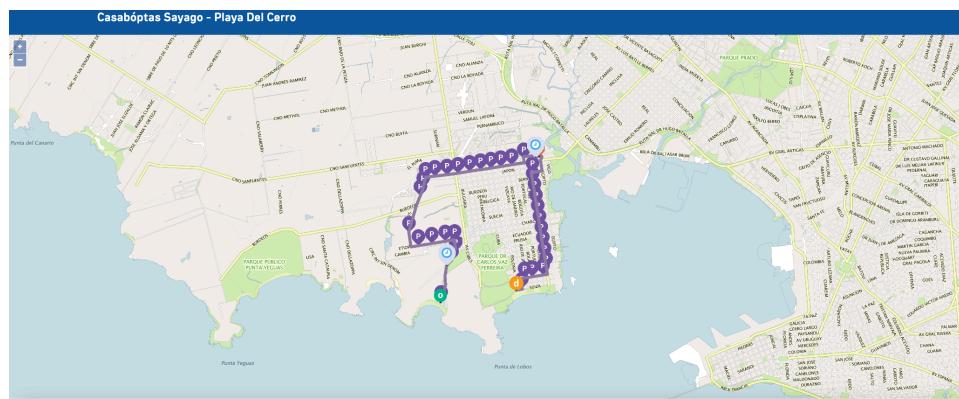
Línea 124



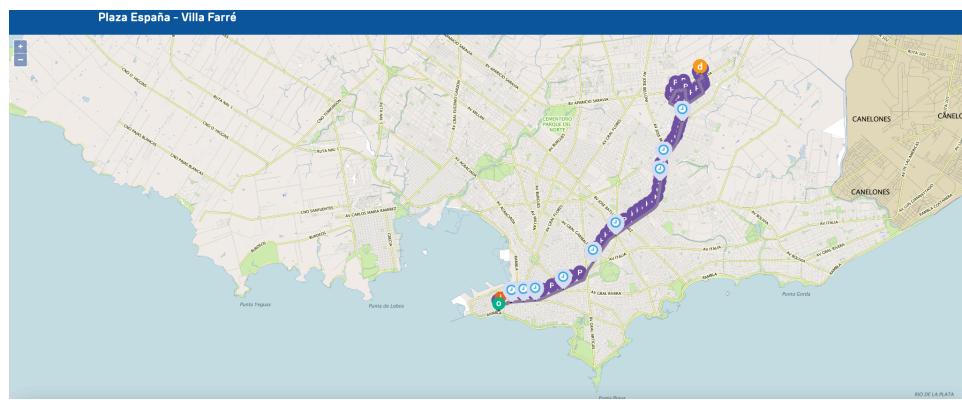
Línea 546



Línea L12



Línea 100



Línea G

