

Ciencia de Datos y Reportes de Movilidad Google para Modelizar la Demanda de Combustible

Data Science and Google Mobility Reporting for Fuel Demand Modeling

Irma Noemí No¹, Julián E. Tornillo¹, Guadalupe Pascal¹, Aixa Maldonado¹

¹ Instituto de Investigaciones en Ingeniería Industrial - I4,
Facultad de Ingeniería, Universidad Nacional de Lomas de Zamora
(1832) Buenos Aires, Argentina
{ino, jtornillo, gpascal}@ingenieria.unlz.edu.ar,
maldonadoae0@gmail.com

Resumen. La reciente situación de pandemia mundial impulsó la generación de reportes abiertos de movilidad, iniciativa creada por la empresa Google en apoyo a las políticas sanitarias asociadas al COVID-19. El cambio significativo en el flujo de vehículos durante la situación de pandemia y la variación en el consumo de combustible asociado al transporte y a diversas actividades productivas, requirió la creación de nuevos modelos predictivos relacionados con un conjunto de datos inusuales (por ejemplo, la geolocalización de los conductores). La manipulación y el análisis adecuado de estos datos proporcionan un pronóstico que mejora la previsión de la demanda de combustible asociada al consumo real. En este trabajo analizamos las bases de datos de la venta de combustibles (Nafta y gasoil), disponibles y abiertas en sitios web oficiales e información de la empresa YPF. Los resultados muestran una correlación positiva entre las variables relacionadas a la demanda de estos combustibles y los registros de movilidad de Google, con ciertas particularidades. El lenguaje de programación utilizado para el desarrollo del código de visualización, geoestadística, cálculo predictivo y reportes de la investigación es “R”.

Palabras claves: Demanda de combustible, Movilidad Google, Series de tiempo, Transporte y Logística, Aprendizaje Automático.

Abstract. The recent global pandemic situation prompted the generation of open mobility reports, an initiative created by Google in support of health policies associated with COVID-19. The significant change in the flow of vehicles during the pandemic situation and the variation in fuel consumption associated with transportation and various productive activities required the creation of new predictive models related to an unusual data set (e.g., geolocation of drivers). Proper manipulation and analysis of these data provide a forecast that improves the traditional forecast of fuel demand associated with actual consumption. In this paper we analyze the databases of fuel sales (gasoline and diesel), available and open in official websites and information of the company YPF. The results show a positive correlation between the variables related to the demand for these fuels and the Google mobility records, with some particularities. The

programming language used for the development of the visualization code, geostatistics, predictive calculation and research reports is "R".

Key words: Fuel demand, Google Mobility, Time series, Transportation and Logistics, Machine Learning.

1 Introducción

La empresa YPF y su departamento de analítica avanzada [1], ante el cambio en el comportamiento del mercado interno de combustibles relacionado con una logística restringida impuesta por el decreto 260/2020 de Aislamiento Social Preventivo y Obligatorio (ASPO) en Argentina [2], rápidamente observaron un importante desajuste entre el modelo predictivo de regresión [3] (tradicionalmente utilizado por la empresa para estimar la demanda de combustible) y la realidad vigente en el mercado interno de transportación (Figura 1).



Fig. 1. Representación a escala del volumen de venta real de Gasoil YPF (2020) y su desajuste con respecto al presupuesto anual estimado, las barras verdes representan el volumen de venta real y la línea verde superior "PA" el valor pronosticado por el modelo predictivo tradicional obtenido mediante el utilitario "SAS" (la pérdida por esta diferencia entre la demanda estimada y la venta real es equivalente a 16000 camiones de combustible en 4 meses). La imagen fue generada por la empresa YPF.

Los indicadores utilizados por la fórmula deducida a través de series temporales históricas (con uso del utilitario SAS [4] – ver Figura 2), basadas en variables como la cantidad de Estaciones de servicio por habitante, el ingreso per cápita y la cantidad de vehículos por habitante, entre otras [5], dejaban de tener el peso predictivo esperado y era necesaria la generación de un nuevo esquema que pudiera anticipar la demanda de combustible por actividades logísticas y productivas, con un margen de aproximación aceptable dado el contexto de transportación restringido por las medidas sanitarias vigentes.

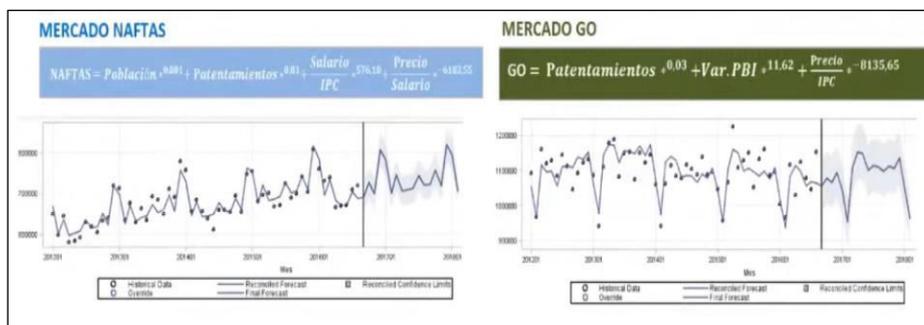


Fig. 2. Fórmulas predictivas tradicionales (YPF – 2020). Sobre el eje x se detalla la fecha (AñoDíaMes: 201201, 201301, etc), El eje y detalla el volumen de venta del combustible en m3, las líneas anteriores a la franja de pronóstico (zona de intervalo de confianza) son ajustes por gráfico de líneas sobre los valores de ventas históricas para el período en cuestión. Las variables que intervienen en las fórmulas siguen los formatos teóricos típicos [5]. La imagen fue generada por la empresa YPF.

La creación de un escenario predictivo adecuado requirió hallar un nuevo conjunto de datos de entrada, que pudieran representar la realidad atípica generada por la nueva situación de circulación restringida. Luego de ensayar diferentes fuentes de datos asociados a la logística de transporte (órdenes de compra, ventas, permisos de circulación) YPF distingue a la movilidad (rastreado por geolocalización) como el input más ajustado a la demanda de combustible por transportación en situación de ASPO (ver Figuras 3 y 4.).

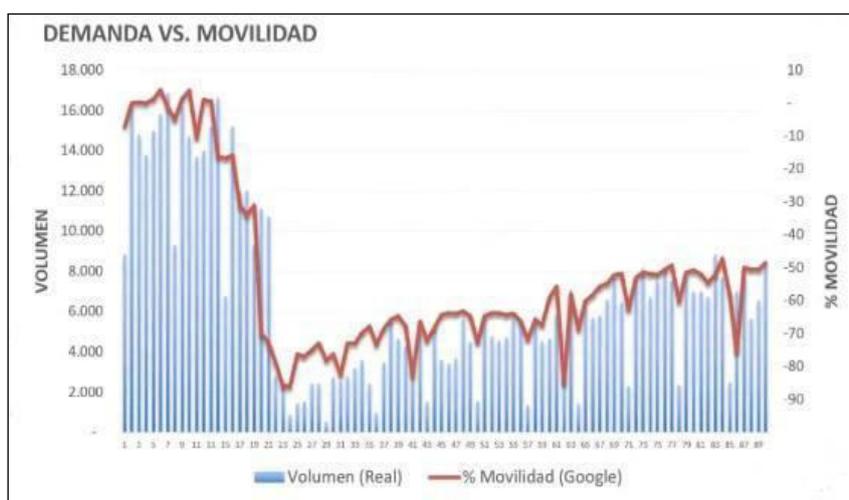


Fig. 3. Ajuste temporal "Demanda de combustible - Movilidad Google". (YPF – 2020). El volumen de venta (m3) corresponde al combustible Gasoil (vista izquierda), a la derecha se observa la variación de movilidad (%) registrada por Google. La variación % es medida con respecto a la movilidad registrada en el período base (del 3/1/20 al 6/2/20). En el eje x se enumeran los días de observación comparativa demanda-movilidad a partir de marzo 2020. La imagen fue generada por la empresa YPF.

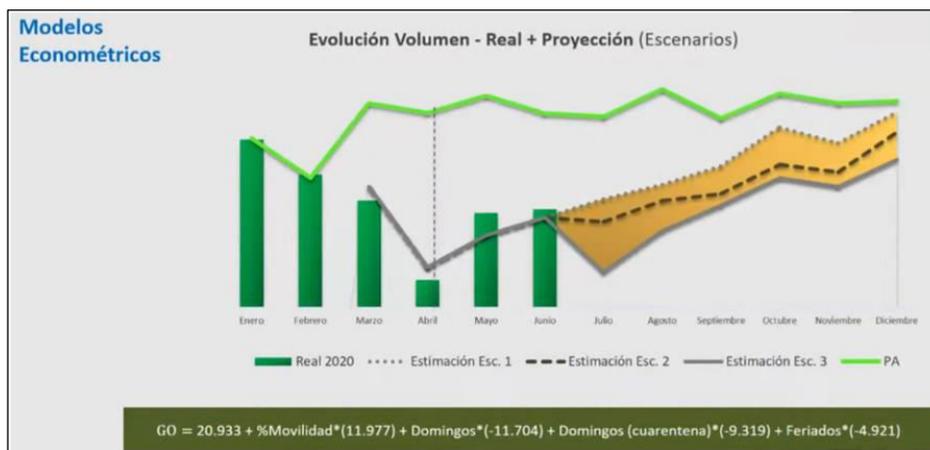


Fig. 4. Modelo econométrico ajustado para el combustible "Gasoil" realizado por la empresa YPF (YPF - 2020). La sigla GO corresponde al combustible Gasoil. Las referencias de la imagen consideran tres escenarios: escenario 1 (optimista), un escenario 2 (pronóstico del modelo), y un escenario 3 (pesimista), ajustados por regresión y sus rangos de confianza. La fórmula indica que la demanda de Gasoil dependerá de la variación % de movilidad a la fecha, los domingos del período (en los cuales baja el expendio de combustible), los domingos que sean parte de la cuarentena (restarán aún más) y los feriados del período (fechas en las cuales también disminuye la demanda de gasoil). La línea verde "PA" es el pronóstico correspondiente al modelo tradicional utilizado anteriormente por la empresa. La imagen y la fórmula han sido generados por la empresa YPF.

Estos resultados pueden validarse y ampliarse mediante el uso de bases de datos oficiales argentinas disponibles en modalidad de datos abiertos [6], realizando los recortes correspondientes a nuevos criterios, productos o empresas de interés, considerando los datos de los Reportes de Movilidad Google [7] como una componente principal del modelo predictivo.

2 Proyecto de Modelización.

En base a los antecedentes mencionados se desarrolla un proyecto de investigación sobre la siguiente hipótesis principal: *La demanda regional de combustible en Argentina puede modelizarse predictivamente mediante técnicas de aprendizaje*

automático con indicadores de geo-movilidad sobre la logística terrestre. La metodología implementada es de tipo cuantitativa con manipulación de bases de datos reales, no experimentales, con énfasis en las fases empírica, analítica y de difusión.

Básicamente se plantean dos objetivos principales para la validación de la hipótesis y un tercero asociado a la reproducibilidad del trabajo realizado:

- Analizar datos correlacionables con el comportamiento logístico terrestre en situación de ASPO argentino y en situación de no restricción (este objetivo supone actividades de extracción, manipulación y análisis de los datos asociados al tema de estudio).
- Proponer un modelo predictivo replicando los análisis de geo-movilidad y consumo/demanda de combustible realizado por la empresa YPF.
- Documentar a través de material reproducible en el entorno RStudio las tareas realizadas y los modelos analizados en lenguaje R.

La planificación básica de las actividades considera los siguientes ítems:

- Descripción y recorte del problema al contexto de investigación: Situación ASPO y poscovid, logística terrestre nacional, consumo interno de combustible de transportación, tipos de combustible y empresas.
- Recolección y preparación de datos: Selección de bibliotecas para su uso en el entorno RStudio, recolección y manipulación de datos abiertos, empresariales YPF y de movilidad Google.
- Utilización de técnicas de análisis de datos, correlación y creación de modelos predictivos (generados por aprendizaje automático) en R sobre los datos obtenidos.
- Creación de mapas con paquetes SIG – R para la visualización del consumo de combustible por transportación logística asociado a los datos de movilidad Google y regionalización de volumen de ventas en diferentes niveles y etapas restrictivas de movilidad territorial.

Se utilizan alternativamente recursos de Excel, Power BI, y/o Python (a elección y por simplificación) para algunas sub tareas a realizar (manipulación de archivos con extensión .csv, filtrados de listas dinámicas y web scraping), que luego son incorporados como parte de scripts y documentos publicables de RMarkdown [8], codificados en el entorno de desarrollo integrado RStudio.

El producto final de la modelización prevé su publicación en modalidad de datos abiertos para permitir la replicación/reproducción de los resultados de la investigación, la obtención de DOI y licencias de autoría.

3 Metodología y Resultados

Se realizan tareas de exploración de datos en bases oficiales argentinas [6] y abiertas de Google [7], manipulación de modelos de aprendizaje automático en el entorno R, explotación de la información correspondiente al modelo desarrollado por YPF para

predecir la demanda de combustible en situación ASPO, pre y poscovid, finalizando con la generación de informes y publicaciones creados con RMarkdown desde el entorno de desarrollo integrado RStudio.

3.1 Acondicionamiento de los Datos.

La metodología empleada es de tipo cuantitativa mediante manipulación de bases de datos reales no experimentales. Las técnicas utilizadas pertenecen a la categoría de aprendizaje automático supervisado (esperando que la variable respuesta de la fórmula predictiva ajuste de manera robusta a la venta real de combustible 2020). Específicamente se utilizaron métodos de regresión lineal múltiple [9][3], este punto se ampliará en el apartado 3.3.

Los datasets descargados desde los sitios oficiales mencionados en el párrafo anterior, permitieron generar en RStudio tablas comparativas, estadísticas y gráficas, desagregadas por fecha, producto, empresa y geolocalización de las plantas de distribución de combustible. El período que abarcan los datos abiertos trabajados se extiende desde el año 2013 hasta agosto del año 2021.

La manipulación realizada sobre las bases datos disponibles se centró primeramente en trabajos de filtrado y resumen (sumarios estadísticos), en función de las observaciones resultantes de cada análisis y de los objetivos del trabajo. Así por ejemplo se detectó que los meses de abril, mayo y junio establecen un salto cuantitativo en los comportamientos de la demanda de combustible en situación de ASPO (año 2020). En la Figura 5 se observa el comportamiento atípico de la demanda en el año 2020, comparado con los mismos meses de otros años.

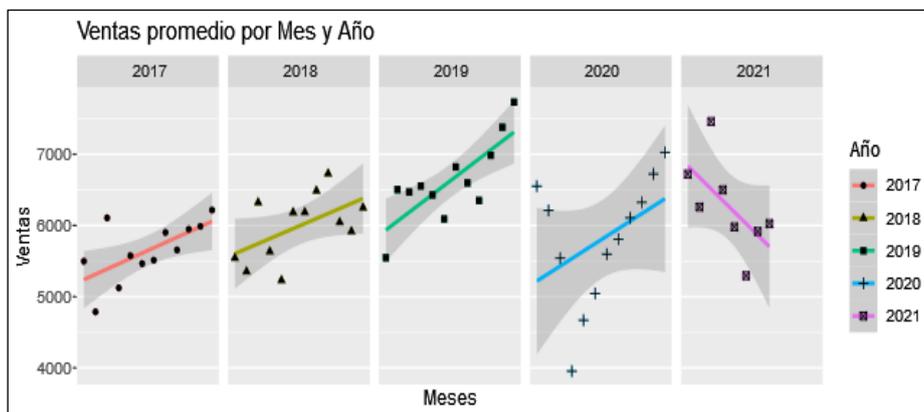


Fig. 5 – Demanda atípica en el año 2020 alcanzando un mínimo inédito. Los valores del eje x corresponden a los meses del año en su orden calendario para cada subgráfico, las ventas son mensuales y corresponden a Nafta y gasoil grados 1 y 2, medidos en m3/100. En cada año se puede observar la recta obtenida por regresión lineal y el intervalo de confianza (generados en R, opción gráfica “stat_smooth(method = "lm)”) sobre el conjunto de puntos: Mes – Cantidad

promedio de combustible vendido. (Fuente: Elaboración propia, versión dinámica en [8], apartado 1.3, pestaña 1.3.2 gráfico inferior). Los datos del año 2021 no se encontraban completos al realizar este estudio.

Estos desajustes sugieren el estudio de series temporales y la generación de nuevos esquemas y fórmulas predictivas, asociadas ahora a las condiciones de restricción o libertad dispuestas sobre la movilidad regional.

Los “Informes de Movilidad Local Google” muestran las tendencias de movilidad por región, categorías de lugares destino (comercios, áreas de ocio, supermercados, farmacias, parques, estaciones de transporte público, lugares de trabajo y zonas residenciales) para cada día del año calendario [7]. Los reportes incluyen los cambios en cada categoría de una región de dos formas distintas:

- Número principal: sirve para comparar la movilidad en la fecha del informe con la movilidad en el día del valor de referencia. Se calcula el dato correspondiente a la fecha del informe (si se dispone de la información) y se muestra un porcentaje positivo o negativo.
- Gráfico de tendencias: muestra los cambios porcentuales durante las 6 semanas anteriores a la fecha del informe.

En los reportes de movilidad Google no se proporciona ninguna información personal identificable sobre la ubicación, los contactos ni los movimientos de ninguna persona en ningún momento. Estos informes se crean a partir de conjuntos de datos agregados y anonimizados procedentes de usuarios que han activado el historial de ubicaciones, el cual está desactivado de forma predeterminada.

La manipulación de las bases de datos de movilidad Google requirió de un gran trabajo de adecuación, compatibilización de denominaciones de Departamentos y Provincias, y la generación de valores promedio para posibilitar su utilización en el pronóstico de la demanda de combustible en Argentina.

Por último, una tercera adecuación de los datos se realizó para lograr la concordancia entre las estructuras cartográficas de Argentina [10] con los datos oficiales de ventas de combustible y los datos de movilidad Google.

3.2 Comportamiento temporal de la venta de combustible y la movilidad Google en Argentina.

Una rápida visualización de la situación atípica registrada en la venta de Nafta de Argentina durante el año 2020 puede observarse en la Figura 6. En este gráfico se han registrado los promedios de los datos oficiales de expendio de Nafta en todo el territorio nacional.

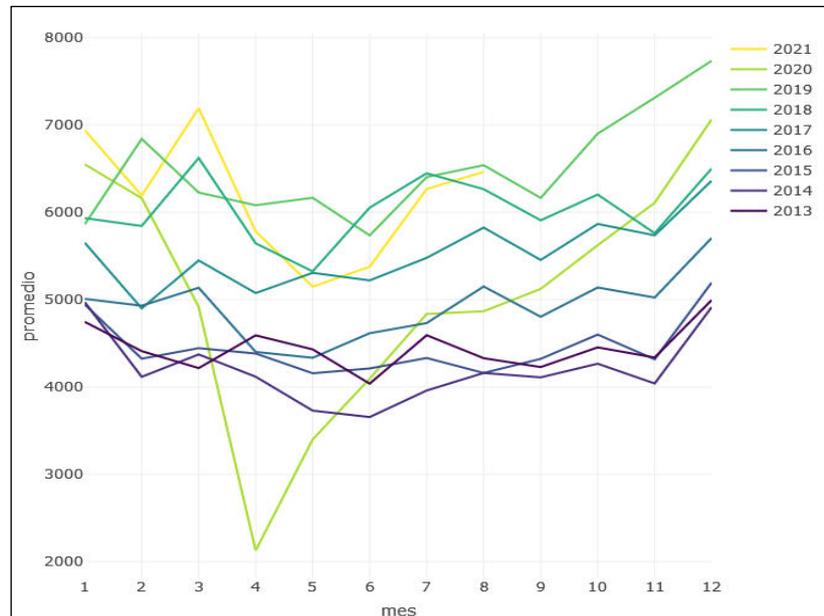


Fig. 6 – Venta sólo del combustible Nafta en Argentina 2013-2021, los valores graficados corresponden al promedio mensual (meses calendario 1 a 12) con unidad de medida $m^3/10^2$ [6]. (Fuente: Elaboración propia, versión dinámica en [8], apartado 2 “Estimaciones de Demanda”).

El comportamiento de la demanda en situación de ASPO y su desfasaje con respecto a los estimadores clásicos también se distribuyeron de manera heterogénea en las diferentes provincias y departamentos, respondiendo a las disposiciones de restricción vigentes en cada zona del país. Como ejemplo, en las Figuras 7 y 8 se puede observar el desfasaje entre las ventas reales y el pronóstico en la provincia de Tucumán en el período ASPO 2020 y su diferencia con la provincia de Santa Fe.

Este comportamiento federalmente heterogéneo con respecto a la diferencia: “venta real versus estimación”, además de responder a diferentes políticas regionales de aislamiento, puede argumentarse desde una lógica de “demanda sectorizada por tipo de expendio” (consumidor final, venta a granel, acopio de proveedores, etc.). Posteriormente se verá que este fenómeno también está relacionado con el tipo de combustible.

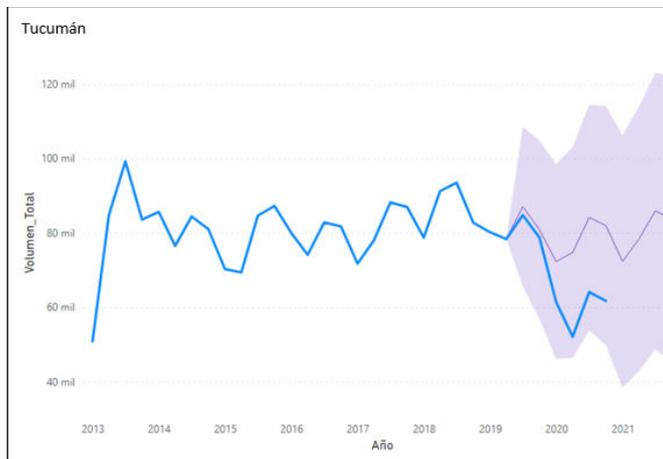


Fig. 7 – Demanda real – pronóstico en Tucumán. El volumen de ventas está expresado en m3. Pronóstico de 10 puntos realizado considerando estacionalidad mensual (período= 12), generado (entrenado) mediante la omisión del último semestre (omisión= 6 datos reales), con la herramienta Power BI, que resume de manera automática a través de un gráfico de líneas (azul) los valores reales de venta mensuales (de Nafta y Gasoil) y estima un escenario pronóstico de demanda (zona sombreada y línea violeta). (Fuente: Elaboración propia).

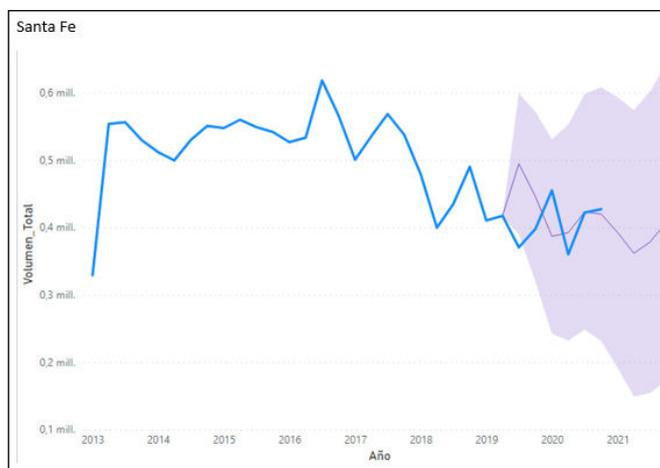


Fig. 8 – Demanda real – pronóstico en Santa Fe. El volumen de ventas está expresado en m3. Pronóstico de 10 puntos realizado considerando estacionalidad mensual (período= 12), generado (entrenado) mediante la omisión del último semestre (omisión= 6 datos reales), con la herramienta Power BI, que resume de manera automática a través de un gráfico de líneas (azul) los valores reales de venta mensuales (de Nafta y Gasoil) y estima un escenario pronóstico de demanda (zona sombreada y línea violeta). (Fuente: Elaboración propia).

El mapeo de las variaciones correspondientes a los indicadores promedio de los reportes de movilidad Google (que pueden observarse dinámicamente en [8] - apartado “Mapas”), también colorean las diferencias regionales (Figura 9) acompañando la atipicidad observada (picos mínimos inéditos) en la demanda y venta de combustibles en Argentina.

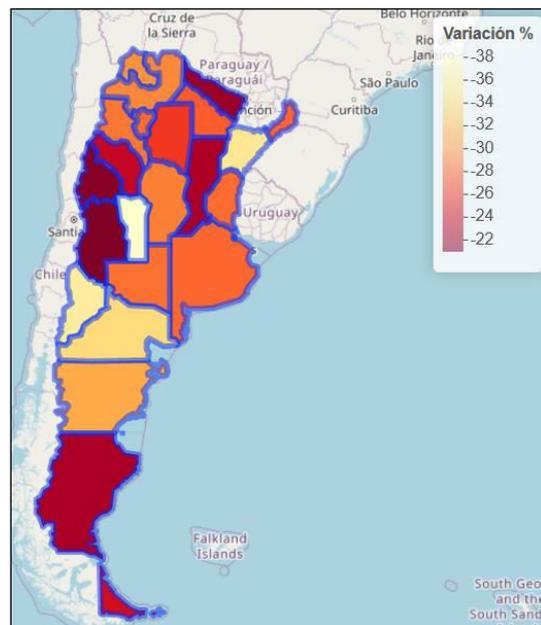


Fig. 9 – Variación promedio de la Movilidad Google – 03/2020. La variación es el promedio de los valores provistos por Google en ese mes para esa provincia sobre los 6 (seis) destinos que considera en sus informes (explicados en los párrafos siguientes). Los valores que figuran en los informes Google son siempre variaciones de movilidad de la fecha en cuestión con respecto al período base (3/1/20 al 6/2/20). (Fuente: Elaboración propia, versión dinámica en [8], apartado 4 “Mapas”).

La interpretación de la variación porcentual reflejada cromáticamente en el mapa de la Figura 9, es la siguiente:

- 1- Los colores más claros corresponden a un gran cambio porcentual, en este caso negativo, reflejando una elevada disminución de movilidad durante el mes de marzo 2020 en esa provincia con respecto al período base considerado por Google (cinco semanas base desde el 3/1/20 al 6/2/20).
- 2- Los valores de variación por provincia fueron calculados como el promedio de los seis (categorías) destinos de movilidad medidos por Google:
 - Comercio minorista y recreación.
 - Supermercado y farmacias.

- Parques.
- Estaciones de transporte.
- Lugares de trabajo.
- Hogares (zonas residenciales).

3- Las diferencias que velozmente se visualizan a través de los niveles de diferenciación cromática se corresponden con la heterogeneidad de las medidas de confinamiento desarrolladas dentro de cada provincia de Argentina.

Fueron utilizadas diferentes herramientas informáticas para visualizar y cuantificar los diagnósticos y pronósticos correspondientes a las series temporales de la demanda de combustible (por tipo, región, empresa) y de movilidad Google.

Los pronósticos estándar creados con las herramientas de Power BI y la biblioteca Prophet de R (Figura 10), permitieron visualizar el desfase entre la estimación de la demanda versus las ventas reales registradas durante la situación de una movilidad restringida, mostrando que varios valores (puntos) del año 2020 se encuentran muy alejados del intervalo de confianza generado.

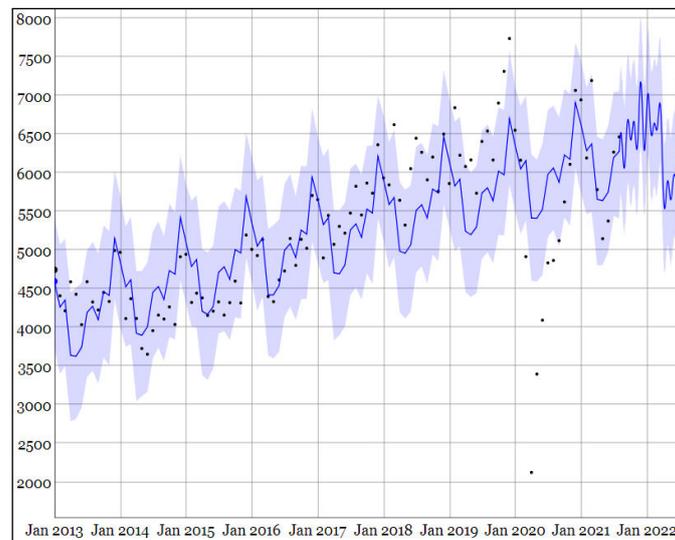


Fig. 10 – Pronóstico con herramienta Prophet de R (línea sólida azul) y ventas reales de Nafta. Los puntos representan los valores promedios por día sobre el total del territorio argentino graficados en m³. En el eje x la denominación “Jan” corresponde a la expresión anglosajona del mes de “Enero”. La herramienta Prophet es útil a nivel visual, y no sigue la modelización algebraica final que hemos propuesto. (Fuente: Elaboración propia, versión dinámica en [8] apartado “Estimaciones de demanda”, 2.1 “Naftas”).

El comportamiento de la serie temporal creada a partir de los reportes de movilidad de Google también demuestra la existencia de estos patrones atípicos en el año 2020 (ver Figura 11), principalmente en los meses de abril, mayo y junio (exactamente los mismos meses que fueron mencionados en párrafos anteriores como altamente diferenciales en la demanda de combustible con respecto a los años anteriores).

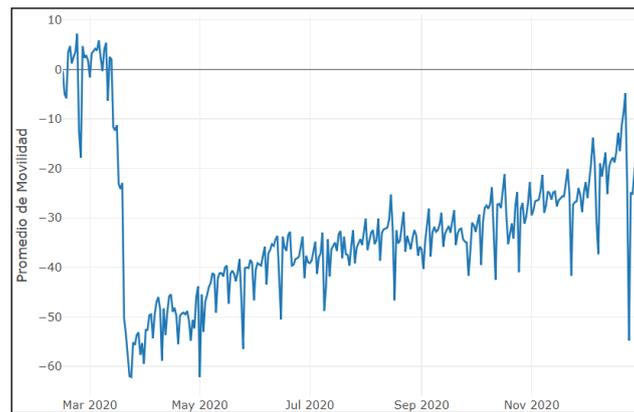


Fig. 11 – Evolución promedio de Google Mobility Trends (Reportes de Movilidad Google). Los valores expresados en el eje “y” son el promedio diario de los valores % provistos por Google en cada fecha sobre los 6 (seis) destinos (categorías) que se consideran en sus informes, para el total del territorio argentino. (Fuente: Elaboración propia, versión dinámica en [8] apartado 3 “Movilidad Google”).

3.3 Correlación Demanda – Movilidad Google.

Ante la singularidad del comportamiento de los datos observados, se analiza la existencia de correlación a través de la función “cor” de R utilizando el método “Pearson”, y posteriormente se utiliza la biblioteca “corrplot” para lograr una visualización sobre los resultados observados. Se halló una correlación positiva (de 0.8) entre las variables movilidad Google y la venta de Nafta consideradas. Aportamos a continuación una visualización conjunta para la comprensión de este comportamiento de similitud (Figura 12).

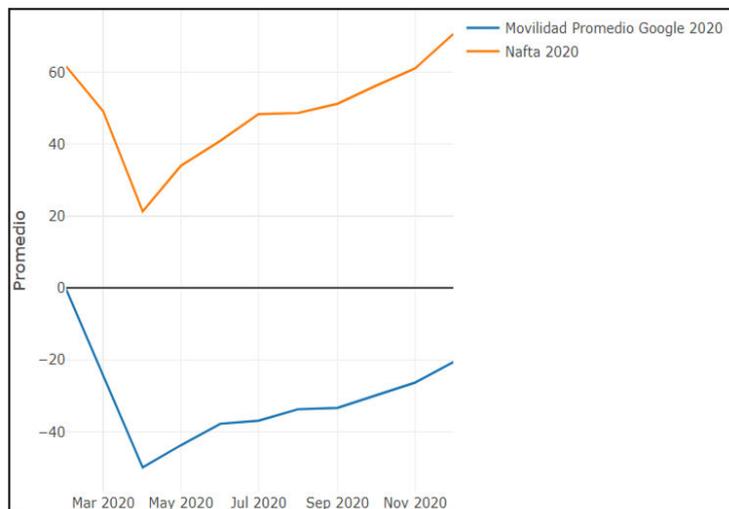


Fig. 12 – Comportamiento de la Movilidad promedio Google y la demanda de Nafta 2020. Los valores de la movilidad Google están expresados en variación porcentual (promedio de las 6 categorías destino informadas por Google) y los valores de la Nafta en promedio mensual sobre el territorio argentino en $m^3/10^4$. (Fuente: Elaboración propia, versión dinámica en [8], apartado 3.1 -observación: leyenda corregida según fe de erratas en repositorio de origen propio-).

Con respecto al Gasoil, la correlación entre la variable movilidad y demanda es baja (~ 0.3) graficada en la Figura 13, una posible interpretación de este valor es la no desagregación del destino final del producto (considerando que las ventas al agro y a la industria se realizan a granel o por acopio y no se ven entonces correlacionadas fuertemente con la movilidad registrada en los reportes de Google).

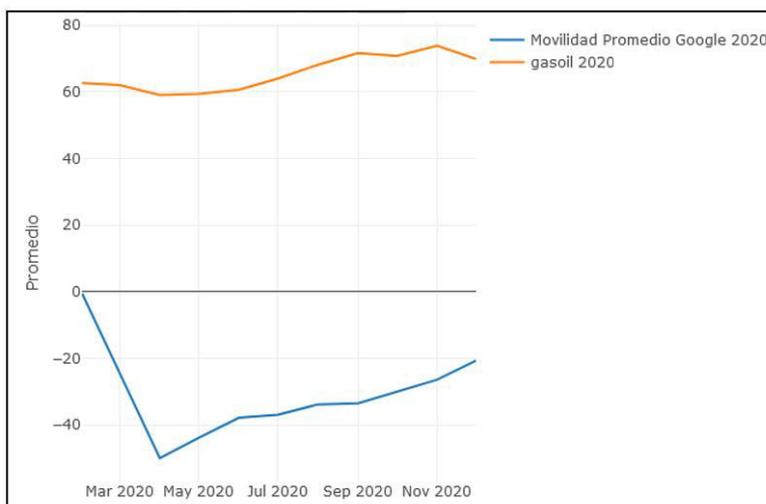


Fig. 13 – Comportamiento de la Movilidad promedio Google y la demanda de Gasoil 2020. Los valores de la movilidad Google están expresados en variación porcentual (promedio de las 6 categorías destino informadas por Google) y los valores de la Gasoil en promedio mensual sobre el territorio argentino en $m^3/10^4$. (Fuente: Elaboración propia, versión dinámica en [8], apartado 3.1 -observación: leyenda corregida según fe de erratas en repositorio de origen propio-).

Trabajando en la corroboración y el ajuste de una fórmula predictiva mediante técnicas de aprendizaje automático de modelado lineal múltiple en R [9], se considera la influencia del promedio de movilidad Google de manera diferenciada en los días de semana, fines de semana y feriados, por lo tanto, se ha estudiado la incidencia del día de la semana y la demanda de combustible asociada (ver Figura 14) para intentar su incorporación en una posible expresión matemática.

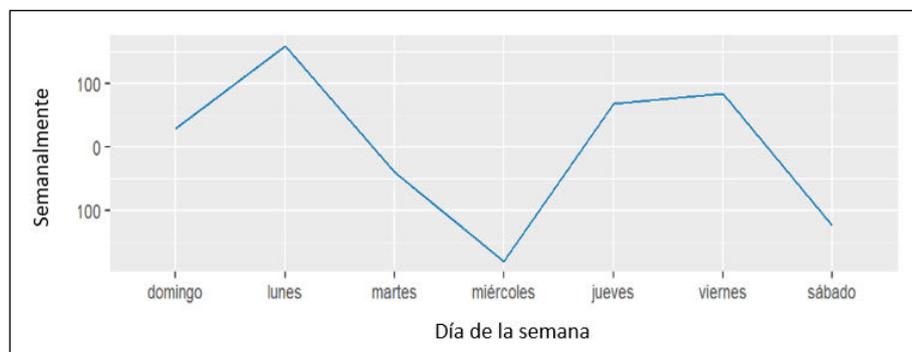


Fig. 14 – Comportamiento de la demanda de Nafta según el día de la semana. Este gráfico se obtiene a través de la biblioteca “Prophet” de R, que desde su propio modelo predictivo permite visualizar componentes de periodicidad considerando todos los datos cargados. En este caso sobre todas las ventas de Nafta volcadas por fecha en el territorio nacional (2013-2021) vemos que el lunes parece ser el día de mayor peso en la venta de Nafta y los miércoles el menos frecuente o de menor impacto en volumen. (Fuente: Elaboración propia). Más allá de la importancia comercial de este dato, el modelo predictivo puede arrancar desde cualquier día, YPF eligió el domingo.

La empresa YPF formuló un modelo considerando la individualización del domingo como un parámetro del cálculo predictivo (ver figura 4) correspondiendo a un valor de partida para la estimación del período a subsiguiente, de acuerdo con la automatización implementada para la captura y manipulación de datos. Los volúmenes de venta se miden en m^3 .

Sin embargo, sin contar con los datos desagregados, exactos y específicos correspondientes al volumen de ventas de combustible de cada día del año en nuestra base de datos obtenida bajo la modalidad de datos abiertos (en la cual los organismos oficiales concentran la carga de los volúmenes de venta a principio de mes), la robustez predictiva se pierde, por lo cual recurrimos a construir nuevas variables y a ensayar diferentes modelos predictivos.

Finalmente, la fórmula (1) de estimación de demanda de Nafta en Argentina con énfasis en la variable de movilidad promedio Google por regresión lineal múltiple, demostró el mejor ajuste dentro de las regresiones lineales múltiples ($r^2= 0.8768879$ y $AIC=463.1654$ =mínimo). El nivel de granularidad alcanzado según la frecuencia de información de las ventas en modalidad de datos abiertos oficiales es mensual.

$$D = 41593.5 + 0.84293*VENTA_{ant} + 4228.18*MOVILIDAD_{dif} \quad (1)$$

Donde:

D = Demanda del combustible (m3)

VENTA_{ant} = Venta de combustible del mes anterior (m3)

MOVILIDAD_{dif} = Diferencia entre los porcentuales de variación de movilidad (escalar) (el coeficiente 4228.18 son los m3 promedio, que se ganan o pierden mensualmente por cada punto de diferencia de movilidad según este modelo y los datos disponibles en modalidad abierta).

Los porcentuales de variación de movilidad (variable “MOVILIDAD”) se obtuvieron como promedio de los 6 destinos (categorías) y fecha en todo el territorio nacional, y son calculados respecto del período base del reporte (el valor de referencia de cada día de la semana propuesto por Google es la mediana correspondiente al período de 5 semanas base desde el 3/1/20 al 6/2/20). La variable MODALIDAD_{dif} se calcula en base a la diferencia de la variable MOVILIDAD en períodos mensuales consecutivos.

Los modelos de regresión lineal múltiple probados se obtuvieron entrenando sobre el total de datos del año 2020 y los dos primeros meses del año 2021, y realizando pruebas de ajuste sobre el semestre siguiente del año 2021. La función utilizada para generar la modelización es “lm” del lenguaje R. Se testearon 9 (nueve) modelos diferentes creados considerando: las variables originales (fecha, movilidad, ventas) y otras creadas a partir de ellas (movilidad anterior, diferencia de movilidad, $\ln(\text{abs}(\text{movilidad}))$, etc.). Las variables fueron propuestas a partir de ciertas observaciones gráficas del comportamiento ventas - movilidad y fueron analizadas mediante matrices de correlación. De todos los modelos propuestos, el óptimo ha sido el formulado en (1), modelo al cual también se ha llegado mediante el camino alternativo de modelización generado por la función “step” de R, aplicado sobre el mismo conjunto de variables del problema. ([8]: apartado 5. “Fórmulas”)

El ajuste del modelo para asegurar la satisfacción de la demanda de combustible es por ahora limitado (Figura 15); consideramos que podría mejorarse a futuro contando con una mayor historia de datos y probando nuevos modelos (no lineales).

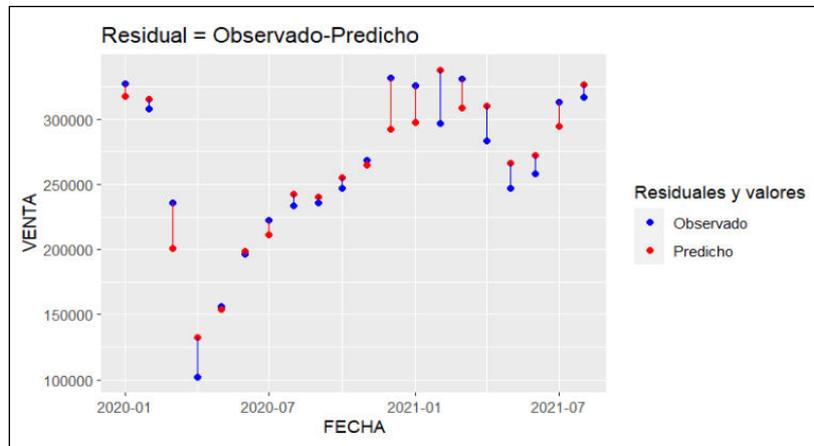


Fig. 15 – Gráfico de ventas mensuales de Nafta (medidos en m3) desde datos abiertos oficiales de estaciones de servicio en el territorio nacional, y valor estimado por la fórmula pronóstico. Los segmentos en azul corresponden a la satisfacción de la demanda (la predicción supera a la venta real). (Fuente: Elaboración propia).

4 Conclusiones

Se ha realizado una importante labor de adecuación sobre los datos abiertos, hasta lograr la generación de bases de datos ordenadas y consistentes para el problema. Estas adecuaciones serán de utilidad para el abordaje de problemáticas similares sobre la estimación de demandas georreferenciadas a corto plazo (por ejemplo, en industrias asociadas al turismo, entre otras).

El trabajo realizado hasta el momento corrobora la existencia de una importante correlación (superior a 0.8) entre las series temporales de los indicadores de movilidad Google y las correspondientes a la demanda de Nafta en Argentina.

Desde el análisis estadístico también se desprende que una aproximación no lineal podría generar alguna mejora predictiva para la variable “ventas de Nafta”, por lo cual se recomienda el ensayo de nuevos modelos de aprendizaje automático, apartándose de la linealidad múltiple propuesta por YPF.

La baja correlación con respecto a la relación demanda global de gasoil - movilidad Google mencionada anteriormente, se interpreta a través de las características propias del expendio de este combustible, el cual mayormente es destinado al agro, la industria y la transportación logística de gran porte. A pesar de ello, la empresa YPF logró estimar sus ventas de gasoil mediante técnicas de desagregación diferencial por destino de producto (técnica que se puede ensayar a futuro si se obtienen dichos datos desagregados).

Es interesante observar que las variables de movilidad interprovincial explican fenómenos atípicos de desabastecimiento y ciertos desajustes de oferta/demanda,

dejando en evidencia la necesidad de la generación de modelos dinámicos formulados con la incorporación de nuevas variables construidas sobre datos en tiempo real (disponibles en la nube). Como ejemplo de esta afirmación podemos mencionar la situación argentina de desabastecimiento de Gasoil [11], generada por fenómenos migratorios limítrofes (y también por acopio), basada en la gran diferencia existente en los precios de venta regionales para este combustible (ver Figura 16).

Contando con datos abiertos y disponiendo de valiosos recursos tecnológicos y de procesamiento, la modelización de pronósticos ajustados que funcionen como gemelos sobre diferentes realidades económicas, industriales y de negocios seguramente se podrán formular en nuevos proyectos de investigación de líneas similares a la presente.

Por último, la empresa Google Inc. ha anunciado la permanencia de los datos de movilidad en formato público, pero determinó como fecha límite de carga y actualización de estos datos el 15 de octubre de 2022.



Fig. 16 – Comportamiento atípico de demanda en gasoil argentino (junio, 2022) [11]

Referencias

1. YPF. (2020). YPF-Ruta. Sitio Web. Consultado el 01/04/2023: <https://www.ypf.com/productosyservicios/YPF-Ruta/Paginas/home.html>

2. Ministerio de Salud. (2020). Decreto de Necesidad y Urgencia (PEJ). *DNU 260/2020*. Recuperado el 01/04/2023 de: <https://www.argentina.gob.ar/coronavirus/dnu>
3. Hernández F., Usuga, O., Mazo M. (2023). Modelos de Regresión con R. Edición digital Bookdown [s.n]. Recuperado el 01/04/2023 de: https://fhernanb.github.io/libro_regresion/index.html
4. SAS. (s.f.). Forecast Server. Sitio Web. Consultado el 01/04/2023: https://www.sas.com/es_ar/software/forecast-server.html
5. Coloma, G. (1998). Análisis de comportamiento del mercado argentino de combustibles líquidos. *XXXIII Reunión de la Asociación Argentina de Economía Política*, Mendoza. Recuperado el 01/04/2023 de: <http://cdi.mecon.gov.ar/bases/doc/aaep/cong/98/coloma.pdf>
6. Datos Argentina (s.f.). *Jefatura de Gobierno, Argentina Unida*. Sitio Web. Consultado en el portal de origen (ME -Secretaría de Energía) el 01/06/2022 desde: <https://datos.gob.ar/>
7. Google. (s.f.). Informes de Movilidad. Sitio Web. Consultado el 01/04/2023: <https://www.google.com/covid19/mobility/>
8. GitHub Inc. (s.f.). Plataforma de desarrollo colaborativo. Repositorio propio recuperado el 01/04/2023 desde: <https://github.com/irmanoemino/demandacombustible1>, Publicado en formato html en: <https://irmanoemino.github.io/demandacombustible1/>
9. Faraway, J. J. (2016). *Extending the Linear Model with R*. Chapman and Hall/CRC. Londres.
10. IGN. (s.f.). Instituto Geográfico Nacional. SIG Argentina. Consultado el 01/04/2023: <https://www.ign.gob.ar>
11. Bronstein, M. (2022). Crisis del gasoil: en algunas provincias triplicaron la venta a vehículos extranjeros por la diferencia en los precios. Infobae. Recuperado el 01/04/2023 de: <https://www.infobae.com/economia/2022/06/21/crisis-del-gasoil-en-algunas-provincias-triplicaron-la-venta-a-vehiculos-extranjeros-por-la-diferencia-en-los-precios/>