

**Materia:** Gestión de Análisis y Diseño de Comercialización (COM145)

**Profesor:** Sarahí Aguilar González

**Fecha de entrega:** 24 de mayo de 2022

**Ciclo:** 1222

**Nombre del proyecto:** Precio de la Gasolina

Miembros del Equipo		
ID	Nombre	Carrera
0217787	Arias Cortina Carlos	ITISI
0213359	Cabrera Ábrego Raúl Andrés	ITISI

Rúbricas				
ID	2-social		7-knowledge	
	D	C	A	JI

# Proyecto Final Data Science

**Abstract-** El presente documento tiene como objetivo mostrar el desarrollo que se llevó a cabo para el proyecto final de la materia de Gestión y Análisis del Conocimiento. Este tiene como fin responder a una pregunta de investigación mediante la implementación de un modelo de aprendizaje, empleando un conjunto de datos de dominio público. Se buscará explicar de forma clara y sencilla las implicaciones sociales de los temas tratados, así como corroborar si se logró responder de manera satisfactoria la pregunta de investigación; empleando herramientas de análisis de datos.

**Índice de Términos—Modelo de aprendizaje, conjunto de datos, pregunta de investigación, implicaciones sociales**

## I. INTRODUCCIÓN

Llenar el tanque de gasolina en nuestro país puede suponer para muchas personas un lujo, ya que el precio dado en las gasolineras ha llegado a niveles no vistos en años.

Aún así, México es de los países que paga uno de los precios más bajos en el mundo por la gasolina, siendo un promedio de 22.97 pesos mexicanos por litro (1.08 dólares); mientras que el promedio a nivel del mundo es de 27.58 pesos (1.29 dólares).

Esta diferencia de precios es ocasionada por los distintos impuestos y subsidios que cada país impone a la gasolina, ya que todos tienen acceso a los mismos precios del petróleo en el mercado internacional. Es evidente el hecho de que pagar la gasolina a precios más bajos, en comparación con el resto del mundo, debe tener su contraparte, y en este caso afecta a las finanzas públicas, pero ¿qué tanto?

El gobierno mexicano dejó de cobrar en su totalidad el IEPS, a la vez que aumentó el estímulo extra para importadores y proveedores, logrando mantener a tope la subida de los precios de la gasolina. Esta política tiene como repercusiones pérdidas de hasta casi 18,000 millones de pesos en una sola semana; tan solo en 2021 los subsidios le costaron a México 104,000 millones de pesos, equivalente a lo presupuestado para la seguridad nacional en 2022.

Si bien esta política beneficia a la ciudadanía, termina afectando a la larga la economía del país, y es por esto que nos llamó la atención el tema del precio de la gasolina.

Centrándonos ahora en la situación de México, el precio de la gasolina se determina a partir de ciertos factores:

- precio de la molécula
- ajuste por calidad
- costos logísticos (transporte, almacenamiento, comercialización)
- ganancia en estaciones de servicio (oferta y demanda)

No obstante, nos surgió la curiosidad sobre si existían otros factores que determinaran el precio de la gasolina, ya que no es sorpresa que hay gasolineras donde la diferencia del precio es notoria, inclusive dentro de la misma zona. Debido a lo

anterior, la pregunta de investigación de nuestro proyecto es la siguiente: *¿Existe una relación entre el precio de la gasolina y ciertas variables sociodemográficas y socioeconómicas?*

Para responder a la pregunta, se analizará el impacto de las siguientes variables y su relación con el precio de la gasolina:

Independientes:

1. número de habitantes por municipio
2. ubicación de las gasolineras
3. extensión del territorio
4. nivel socioeconómico por municipio
  - a. pobreza
  - b. pobreza extrema
  - c. vulnerabilidad por carencia social
  - d. carencia por rezago educativo
  - e. carencia por acceso a los servicios de salud
  - f. carencia por acceso a la seguridad social
  - g. carencia por calidad y espacios de la vivienda
  - h. carencia por servicios básicos en la vivienda

Dependientes:

1. precio de la gasolina

Para lo anterior, se hará uso de las siguientes fuentes de datos:

1. INEGI
2. CONEVAL
3. Servicio Postal Mexicano
4. CRE
5. Geocoding API

## II. DESARROLLO

Las implicaciones sociales sobre el precio de la gasolina y los factores que lo determinan son de suma relevancia, ya que la gasolina es el combustible principal para que los ciudadanos puedan transportarse en automóviles particulares, y si el precio es muy elevado, tendrán que recurrir a otras formas de traslado, como puede ser el transporte público o servicios de transporte en aplicaciones, como puede ser Uber o Didi.

Es cierto que gran parte de los mexicanos no tienen posibilidad de tener un automóvil propio, por lo que optan por trasladarse en transporte público; pero no es sorpresa que la seguridad en éstos es mucho menor debido a la gran cantidad de acoso, asaltos e inclusive asesinatos que suelen ocurrir en estos.

Por poner un ejemplo, si una persona no considera viable viajar en automóvil privado debido al costo de la gasolina, optará por viajar en transporte público, y es cierto que también beneficia al medio ambiente por la reducción de gases, más no

es siempre la mejor opción considerando el panorama que se mencionó anteriormente. Y sí, México hace lo posible por mantener el costo bajo a comparación de otros países, pero como se mencionó en la introducción, se gastan muchos recursos económicos que podrían destinarse a otras áreas.

Evidentemente es un tema muy complejo al cual se deben realizar varios estudios para buscar la mejor solución y buscar beneficiar tanto a los ciudadanos como al país en sí.

Una vez que se cuenta con la pregunta de investigación y las fuentes de datos con las que se trabajará, sólo falta un detalle previo al análisis de los datos, que es plantear el modelo a desarrollar.

Será un modelo de aprendizaje supervisado, ya que contamos con una variable dependiente, siendo en este caso el precio de la gasolina.

Se priorizará la predicción, ya que nuestro objetivo es calcular el precio de la gasolina a partir de las variables más significativas.

Por último, es un problema de regresión, ya que el precio de la gasolina es una variable numérica.

Para el análisis preliminar de los datos, primero se descargaron dos fuentes de datos en formato XML de CRE, donde el primero brindaba el nombre de cada estación de gasolina junto con su localización en coordenadas, y el segundo el precio de la gasolina regular y premium; ambas relacionadas según el ID de la gasolinera.

Después exportamos los XML a Excel, para poder visualizarlos en Pandas mediante un DataFrame.

A partir de aquí se llevó a cabo todo el preprocesamiento y limpieza de los datos, lo cual llevó bastante tiempo debido a las irregularidades que hay en los datos de las distintas fuentes; por ejemplo, municipios que no contenían acentos o carecían de datos en ciertas columnas.

Se graficaron las variables con el fin de observar cuáles tenían mayor repercusión con el precio de la gasolina para lograr responder a la pregunta de investigación.

Para el entrenamiento, optamos por un modelo de regresión, y mediante el ECM, determinar la medida de rendimiento.

Se obtuvo un valor de ECM de 1.66, por lo cual no es un valor muy bajo y da a entender que las variables no tienen un peso suficiente para determinar el precio de la gasolina.

### III. CONCLUSIONES

Podemos concluir que las variables independientes elegidas no tuvieron la relevancia suficiente para influir en el precio de la gasolina, ya que los factores externos e internos son los que tienen más peso.

Se pudieron mejorar varias cosas, como investigar más variables independientes y analizar cuáles tenían mayor relevancia. También se pudo haber investigado más a fondo el porqué la diferencia de precios entre delegaciones o municipios. Se podría a futuro buscar una solución para que los mexicanos no paguen tanto por la gasolina a la vez que el país no salga perdiendo tanto, tal vez distribuir los recursos de

una manera más eficiente, aunque sería una meta muy complicada de alcanzar.

### IV. INVESTIGACIONES SIMILARES

La idea de predecir el costo de algún bien, y más si es uno que se consume de forma regular, es una idea bastante interesante y atractiva para aquellos familiarizados con el manejo de los datos; por lo que se investigaron algunos proyectos similares con el fin de dar un panorama de hacia dónde ir.

Primero, existe un proyecto de nombre “*ESTIMACIÓN Y PREDICCIÓN DEL PRECIO DE LA GASOLINA EN MÉXICO A PARTIR DEL MODELO DE AJUSTE DE PRECIOS DE EVANS*”, donde estudiantes de segundo semestre de licenciatura en ingeniería del Tec de Monterrey buscaban obtener una predicción del precio de la Magna, a través de un modelo matemático basado sobre dos pilares: el modelo de ajuste de precios de Evans, que se traduce en una ecuación diferencial de primer orden, y el ajuste por regresión lineal de la oferta y demanda utilizando la información generada por PEMEX.

Existe otro proyecto de nombre “*Natural Gas Spot Price Prediction Using Artificial Neural Network*”, en el cual se busca predecir el precio del gas natural con variables que impacten de forma directa o indirecta, tales como datos meteorológicos, económicos, etc; relacionados al clima, con una red neuronal artificial a partir de los precios históricos.

Por último, existe un trabajo de fin de máster de Minería de Datos e Inteligencia de Negocio con título “*Estudio de la competencia de los precios de los carburantes en España mediante Teoría de Juegos y Machine Learning*”, donde se centraron en distintas técnicas de Machine Learning para predecir el precio del carburante de las gasolineras de España, y posteriormente las compararon para concluir cuál de esos modelos es el que mejor predice con respecto a su base de datos; cabe destacar que es un trabajo bastante amplio e informativo.

### V. BIBLIOGRAFÍA

1. Franco, V. (n.d). ¿Cómo se determina el precio de la gasolina?  
Sodexo.  
<https://blog.sodexo.com.mx/blog/que-factores-influyen-para-determinar-el-precio-final-gasolina-o-diesel>.
2. Becerra, J., & Becerra, J. (2022, Marzo 22). Gasolina, ¿cuál es la diferencia entre subsidio y estímulo fiscal?  
EL CEO.  
<https://elceo.com/politica/cual-es-la-diferencia-entre-subsidio-y-estimulo-fiscal-para-gasolina/#:%7E:text=>

Contener%20el%20precio%20de%20la,en%20este%  
20caso%2C%20la%20gasolina

3. Expansión. (2022, Abril 21). ¿Quiénes ganan y quiénes pierden con el subsidio a la gasolina? <https://expansion.mx/economia/2022/03/30/subsidio-gasolina-mexico-2022>
4. Expansión. (2022a, March 15). Gasolina mexicana, entre las más baratas del mundo, pero ¿a qué costo? <https://expansion.mx/economia/2022/03/15/gasolina-mexicana-mas-barata-mundo-a-que-costo>
5. Prado, C. (2019). ESTIMACIÓN Y PREDICCIÓN DEL PRECIO DE LA GASOLINA EN MÉXICO A

PARTIR DEL MODELO DE AJUSTE DE PRECIOS DE EVANS. Acta Latinoamericana de Matemática Educativa. Retrieved May 22, 2022, from <http://funes.uniandes.edu.co/14013/1/Prado2019Estimacion.pdf>

6. Pablos, B. (2017). Estudio de la competencia de los precios de los carburantes en España mediante Teoría de Juegos y Machine Learning. Universidad Complutense Madrid. Retrieved May 23, 2022, from [https://eprints.ucm.es/id/eprint/45868/1/TFM\\_Beatriz\\_Escribano\\_Pablos.pdf](https://eprints.ucm.es/id/eprint/45868/1/TFM_Beatriz_Escribano_Pablos.pdf)