

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

Profesor: Sarahí Aguilar González

Fecha de entrega: May 24, 2022

Ciclo: 1222

Nombre del proyecto: Price of gasoline

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

Data Science Final Project

Abstract- The purpose of this document is to show the development that was carried out for the final project of the Data Science course. It aims to answer a research question through the implementation of a learning model, using a public domain dataset. It will seek to explain in a clear and simple way the social implications of the topics covered, as well as to corroborate whether the research question was answered satisfactorily, using data analysis tools.

Key words—Learning model, dataset, research question, social implications

I. INTRODUCTION

Filling up the gas tank in our country can be a luxury for many people, since the price given at gas stations has reached levels not seen in years.

Even so, Mexico is one of the countries that pays one of the lowest prices in the world for gasoline, with an average of 22.97 Mexican pesos per liter (1.08 dollars), while the world average is 27.58 pesos (1.29 dollars).

This price difference is caused by the different taxes and subsidies that each country imposes on gasoline, since they all have access to the same oil prices in the international market. It is evident that paying lower gasoline prices, compared to the rest of the world, must have its counterpart, and in this case it affects public finances, but how much?

The Mexican government stopped charging the IEPS in its entirety, while at the same time increasing the extra stimulus for importers and suppliers, managing to keep gasoline prices at the top of the scale. This policy has as repercussions losses of up to almost 18 billion pesos in a single week; in 2021 alone, subsidies cost Mexico 104 billion pesos, equivalent to the amount budgeted for national security in 2022.

Although this policy benefits the citizens, it ends up affecting the country's economy in the long run, and this is why the issue of gasoline prices caught our attention.

Focusing now on the situation in Mexico, the price of gasoline is determined by certain factors:

- price per molecule
- quality adjustment
- logistics costs (transportation, storage, marketing)
- profit in service stations (supply and demand).

However, we were curious as to whether there were other factors that determine the price of gasoline, since it is no surprise that there are gas stations where the price difference is notorious, even within the same area. Due to the above, the research question of our project is the following: *Is there a*

relationship between the price of gasoline and certain sociodemographic and socioeconomic variables?

To answer the question, the impact of the following variables and their relationship with the price of gasoline will be analyzed:

Independent:

1. number of inhabitants per municipality
2. location of gas stations
3. extension of the territory
4. socioeconomic level per municipality
 - a. poverty
 - b. extreme poverty
 - c. vulnerability due to social deprivation
 - d. educational deprivation
 - e. lack of access to health services
 - f. lack of access to social security
 - g. housing quality and space deprivation
 - h. lack of basic services in housing

Dependents:

1. gasoline price

For the above, the following data sources will be used:

1. INEGI
2. CONEVAL
3. Mexican Postal Service
4. CRE
5. Geocoding API

II. DEVELOPMENT

The social implications of the price of gasoline and the factors that determine it are extremely relevant, since gasoline is the main fuel for citizens to transport themselves in private cars, and if the price is too high, they will have to resort to other forms of transportation, such as public transportation or transportation services in apps, such as Uber or Didi.

It is true that most Mexicans do not have the possibility of owning their own car, so they choose to travel by public transportation; but it is no surprise that the security in these is much lower due to the large amount of harassment, assaults and even murders that usually occur in these.

To give an example, if a person does not consider it viable to travel by private car due to the cost of gasoline, he/she will choose to travel by public transportation, and it is true that it also benefits the environment by reducing gas emissions, but it is not always the best option considering the scenario mentioned above. And yes, Mexico does its best to keep the

cost low compared to other countries, but as mentioned in the introduction, a lot of economic resources are spent that could be allocated to other areas.

Evidently, this is a very complex issue that requires several studies to find the best solution and seek to benefit both citizens and the country itself.

Once we have the research question and the data sources with which we will work, there is only one detail missing prior to the analysis of the data, which is to propose the model to be developed.

It will be a supervised learning model, since we have a dependent variable, being in this case the price of gasoline.

Prediction will be prioritized, since our objective is to calculate the price of gasoline from the most significant variables.

Finally, it is a regression problem, since the price of gasoline is a numerical variable.

For the preliminary analysis of the data, we first downloaded two data sources in XML format from CRE, where the first one provided the name of each gas station along with its location in coordinates, and the second one the price of regular and premium gasoline; both related according to the ID of the gas station.

Then we exported the XML to Excel, in order to visualize them in Pandas through a DataFrame.

From this point on, all the pre-processing and data cleaning was carried out, which took quite some time due to the irregularities in the data from the different sources; for example, municipalities that did not contain accents or lacked data in certain columns.

The variables were plotted in order to observe which ones had the greatest impact on the price of gasoline in order to answer the research question.

For training, we opted for a regression model, and by means of the *ECM*, to determine the performance measure.

An *ECM* value of 1.66 was obtained, which is not a very low value and suggests that the variables do not have sufficient weight to determine the price of gasoline.

III. CONCLUSIONS

We can conclude that the independent variables chosen were not relevant enough to influence the price of gasoline, since external and internal factors have the most weight.

Several things could have been improved, such as investigating more independent variables and analyzing which ones were more relevant. Also, more research could have been done on the reason for the difference in prices between delegations or municipalities. In the future, a solution could be sought so that Mexicans do not pay so much for gasoline while the country does not lose so much, perhaps distributing

resources in a more efficient way, although it would be a very complicated goal to achieve.

IV. SIMILAR INVESTIGATIONS

The idea of predicting the cost of some good, and more so if it is one that is consumed on a regular basis, is a rather interesting and attractive idea for those familiar with data management; so some similar projects were investigated in order to give an overview of where to go.

First, there is a project called "ESTIMATION AND PREDICTION OF THE PRICE OF GASOLINE IN MEXICO FROM THE EVANS PRICE ADJUSTMENT MODEL", where second semester undergraduate engineering students from Tec de Monterrey sought to obtain a prediction of the price of Magna, through a mathematical model based on two pillars: the Evans price adjustment model, which translates into a first order differential equation, and the linear regression adjustment of supply and demand using information generated by PEMEX.

There is another project called "Natural Gas Spot Price Prediction Using Artificial Neural Network", which seeks to predict the price of natural gas with variables that have a direct or indirect impact, such as meteorological and economic data, etc., related to climate, with an artificial neural network based on historical prices.

Finally, there is a Master's thesis on Data Mining and Business Intelligence entitled "Study of the competition of fuel prices in Spain through Game Theory and Machine Learning", where they focused on different Machine Learning techniques to predict the price of fuel at gas stations in Spain, and then compared them to conclude which of these models is the one that best predicts with respect to its database; it should be noted that it is a fairly comprehensive and informative work.

V. REFERENCES

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