

1. Introducción

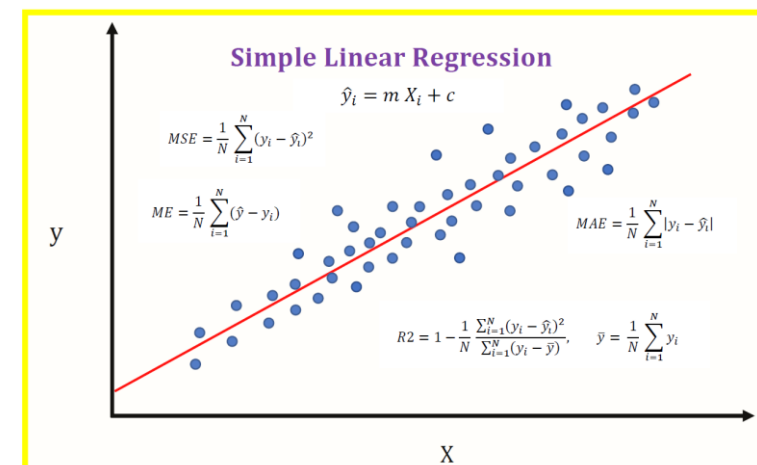
La Tarjeta Alimentar es un instrumento del Plan Argentina contra el Hambre, una política integral que impulsa la Nación en articulación con las provincias, y los municipios. El programa está orientado a garantizar a las familias el acceso a la canasta básica alimentaria. Permite comprar todo tipo de alimentos, a excepción de bebidas alcohólicas.

Está dirigido a madres y padres con hijos e hijas de hasta 6 años de edad que reciben la Asignación Universal por Hijo (AUH). El método de uso del dinero disponible es a través de una tarjeta entregada por el Banco Nación, de la cual no se permite extraer dinero en efectivo ni realizar transferencias.

2. Objetivo y Métodos

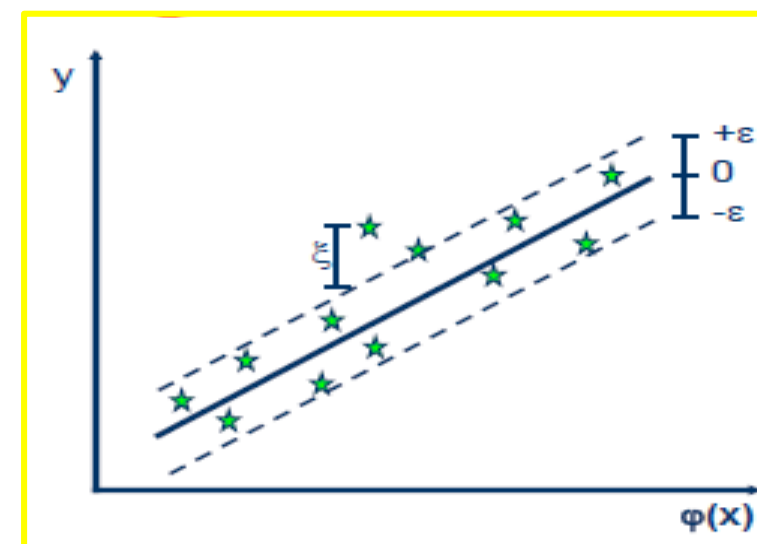
El objetivo que tuvo este trabajo fue el de entender la base de datos subida al portal de datos abiertos de la nación (1) y, a través de modelos de aprendizaje supervisado, entender si, con los datos provistos, se podía armar un modelo predictivo de suscripciones por día al programa. Los modelos utilizados fueron Regresión Lineal y Support Vector Regression (SVR).

Regresión Lineal:



Por medio de una función lineal se buscará minimizar la suma de residuos al cuadrado (MSE). Es así que se generará una función decisión como la que se puede observar en rojo a nuestra izquierda ($y' = mX_1 + c$). Las observaciones o muestras son los puntos azules, graficados en dos dimensiones x/y. (2)

Support Vector Regression (SVR):



Funciona bajo la misma lógica del modelo de clasificación de Support Vector Machine, solo que en este caso busca determinar un margen/radio (Épsilon) como función de costo y trata de que todas las muestras estén dentro del margen. Al igual que la regresión, construye una función lineal. (3)

(1) Titulares de la Tarjeta AlimentAR (2020). Recuperado de <https://datos.gob.ar/dataset/desarrollo-social-titulares-tarjeta-alimentar>

3. Análisis Exploratorio de Datos

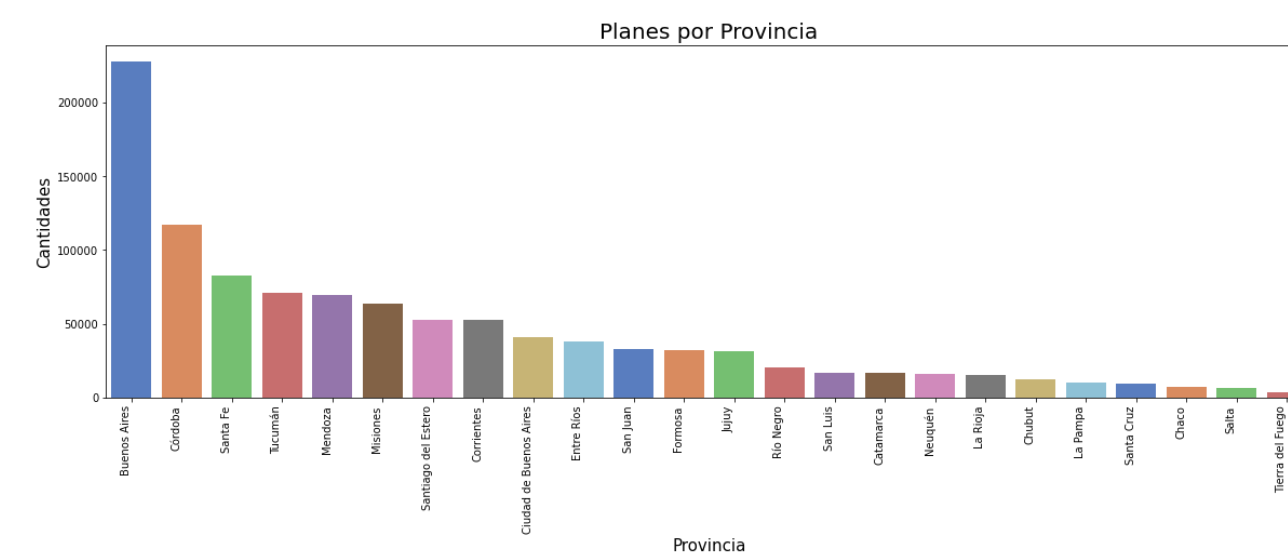
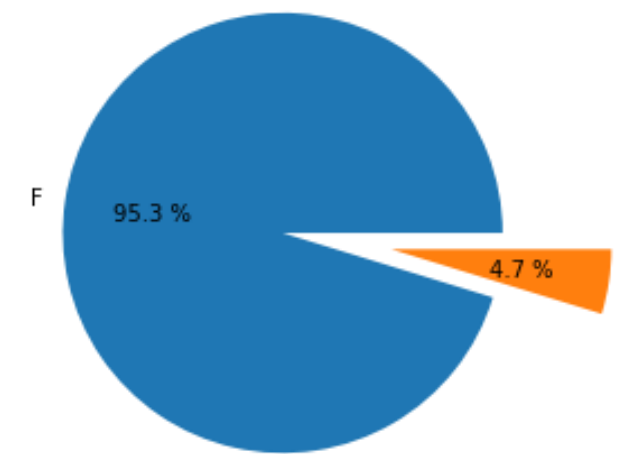
La base de datos "Tarjeta AlimentAR" contiene todos los registros únicos de personas que hicieron uso del derecho de obtención de este plan. Los mismos son aproximadamente 1.2 millones de personas en todo el país, clasificados por provincia, departamento, fecha de alta, tiempo de permanencia en el plan y monto liquidado por cada mes.

A fines de depurar la base de datos para poder analizar correctamente los datos se procedió a:

- Verificar registros duplicados (los cuales no se encontraron)
- Encontrar valores nulos (había una columna de "fecha_baja" completamente vacía)
- Agregar una columna de monto total liquidado por persona.

El primer hallazgo importante dentro de nuestro *DataSet* fue que el 95% de las personas adherentes al plan pertenecen al sexo femenino, mientras que el restante 5% al masculino.

Es decir, dentro de una familia tipo, es la mujer quien se acerca a retirar la tarjeta para el grupo familiar.

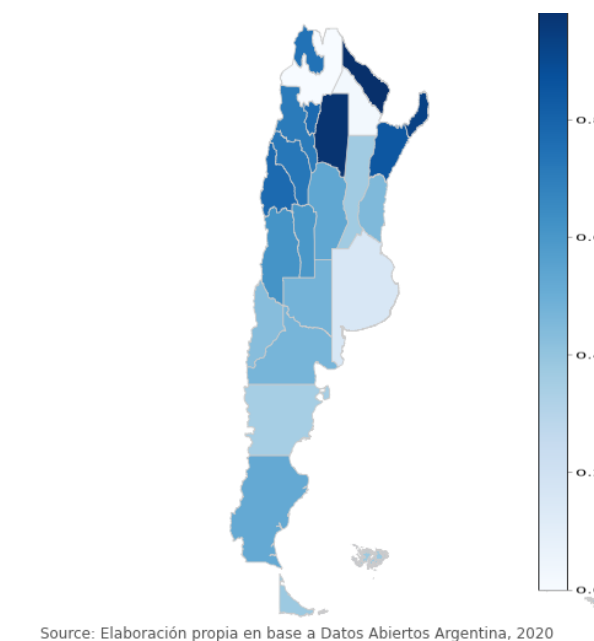
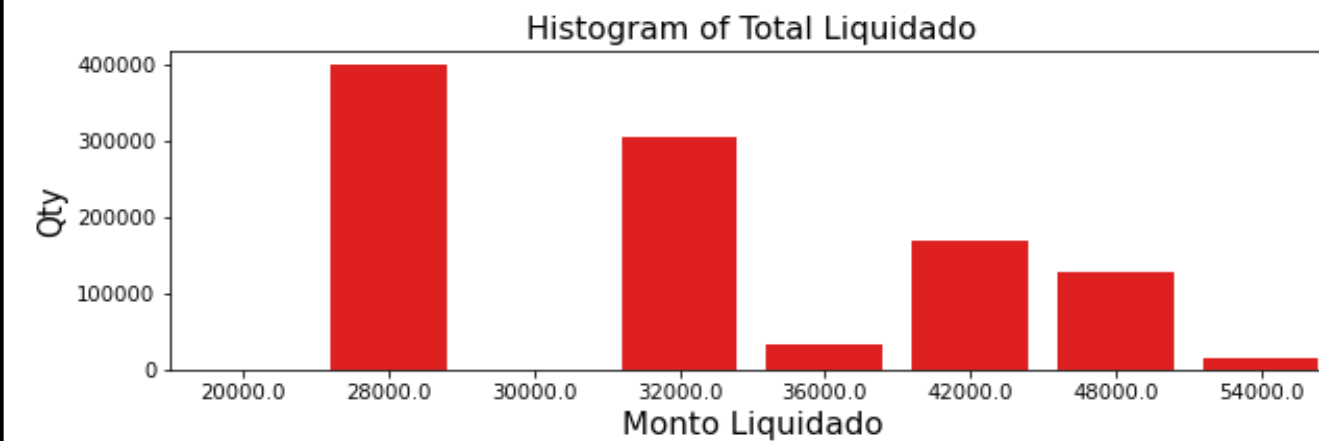


El segundo hallazgo del *DataSet* es que las tres provincias con mayor cantidad de planes otorgados son Buenos Aires (s/Capital Federal), Córdoba y Santa Fe, aunque, veremos que la incidencia en estas provincias no es de las más altas.

(2) Obi Tayo, Benjamin (2020). Recuperado de <https://medium.com/towards-artificial-intelligence/how-to-i-calculate-accuracy-for-regression>

(3) Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer Science + Business Media.

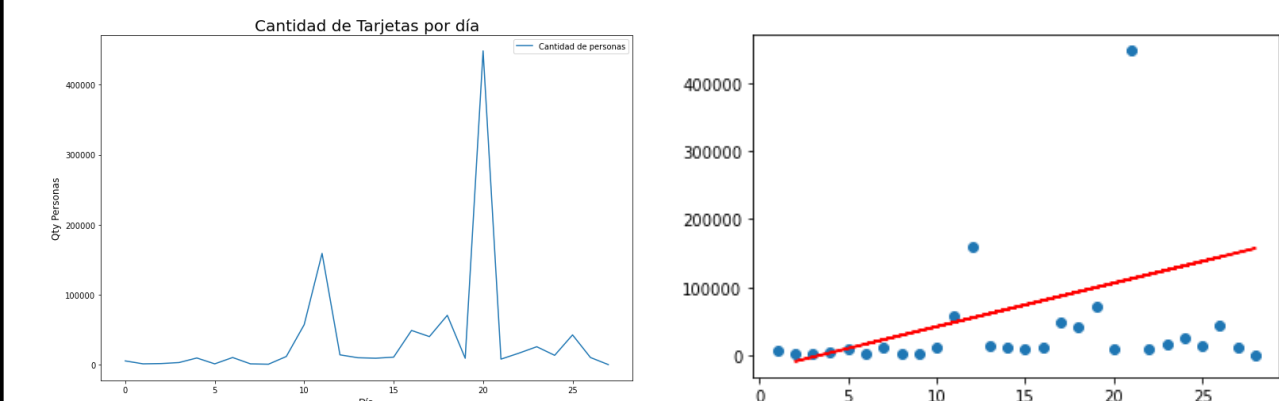
El tercer hallazgo fue que el monto total cobrado que más se repite en el *DataSet* es de \$28.000 ARS, mientras que pocas familias cobraron el máximo visto en el análisis (\$54.000 ARS). El valor medio cobrado es de \$34.000 ARS



Finalmente, el cuarto hallazgo del *DataSet* fue a través de un "Heatmap" identificando la incidencia del plan AlimentAR en cada provincia. Como se puede observar, las 3 provincias con mayor cantidad de planes no son las de mayor incidencia vs cantidad total de habitantes. Las 3 con mayor incidencia son Santiago del Estero, Formosa y Misiones, en azul más oscuro.

Source: Elaboración propia en base a Datos Abiertos Argentina, 2020

4. Resultados y Conclusiones



Lo que se hizo fue identificar la cantidad de personas por día inscriptas, para poder armar un modelo predictivo de potenciales nuevos adherentes al plan. Lo que podemos ver es que en el par días/cantidad de personas hay valores muy distintos entre sí, generando varios *Outliers* (el día con mayor incidencia de inscriptos son los principios de mes).

En el segundo gráfico se observa la regresión generada por el modelo lineal, donde vemos que simplifica demasiado el problema, no obteniendo una generalización acorde a lo buscado. También se trabajó con el modelo SVR y se quitó del *DataSet* los *Outliers*, pero por esto último se identificó que representan el 45% de los inscriptos, por lo que no sería correcto hacerlo.