



Corporación Favorita: Time-Series Forecasting

Rafael Velasco Huerta - A01283168

Overview

Overview del problema

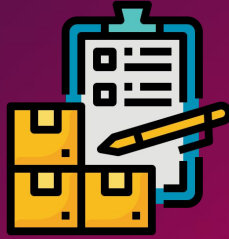
Corporación Favorita es una empresa ecuatoriana de tiendas de autoservicio, inversiones inmobiliarias, generación eléctrica, producción de alimentos y productos de primera necesidad, comercio, y ferretería.

En este caso nos estaremos enfocando en las tiendas de autoservicio, en donde existe la necesidad de predecir las ventas de diferentes familias de productos con el objetivo de que las personas puedan obtener la cantidad de productos que necesitan en el momento en el que lo necesitan.



Implicaciones

Conocer las posibles ventas a futuro puede tener impacto en algunas áreas como:



Compras e Inventarios

Ej. Productos perecederos



Planeación financiera

Ej. Manejo de flujo de efectivo

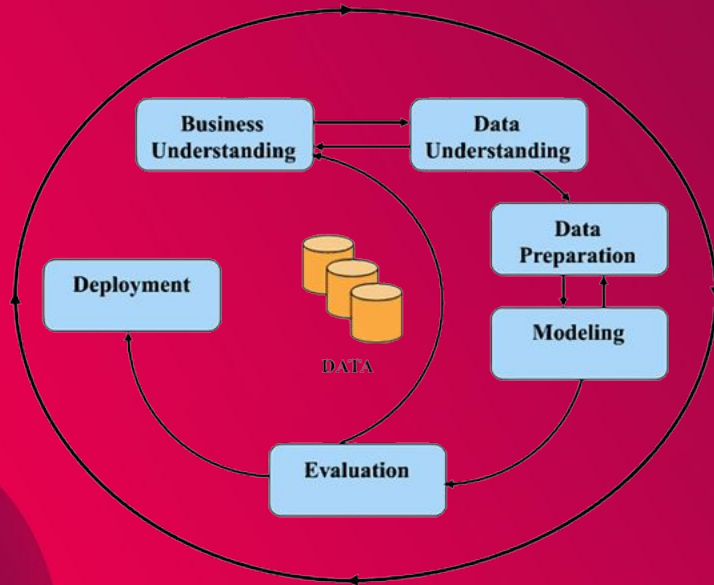


Alineación a objetivos

Ej. Corrección de estrategias

Metodología

Metodología utilizada



CRISP-DM

Business understanding – Qué necesita el negocio?

Data understanding – Qué datos tenemos / están limpios?

Data preparation – Cómo preparamos los datos para modelarlos?

Modeling – Qué técnicas de modelado utilizaremos?

Evaluation – Qué modelo cumple mejor con las necesidades?

Deployment – Cómo podemos darle acceso a las partes interesadas?

Pasos



Análisis Descriptivo

Datasets utilizados

- 1) train.csv
- 2) test.csv
- 3) oils.csv
- 4) transactions.csv
- 5) holiday_events.csv
- 6) stores.csv

```
df_train.head()
```

	id	date	store_nbr	family	sales	onpromotion
0	0	2013-01-01	1	AUTOMOTIVE	0.0	0
1	1	2013-01-01	1	BABY CARE	0.0	0
2	2	2013-01-01	1	BEAUTY	0.0	0
3	3	2013-01-01	1	BEVERAGES	0.0	0
4	4	2013-01-01	1	BOOKS	0.0	0

```
[6]: df_holidays.head()
```

	date	type	locale	locale_name	description	transferred
0	2012-03-02	Holiday	Local	Manta	Fundacion de Manta	False
1	2012-04-01	Holiday	Regional	Cotopaxi	Provincializacion de Cotopaxi	False
2	2012-04-12	Holiday	Local	Cuenca	Fundacion de Cuenca	False
3	2012-04-14	Holiday	Local	Libertad	Cantonizacion de Libertad	False
4	2012-04-21	Holiday	Local	Riobamba	Cantonizacion de Riobamba	False

```
df_stores.head()
```

	store_nbr	city	state	type	cluster
0	1	Quito	Pichincha	D	13
1	2	Quito	Pichincha	D	13
2	3	Quito	Pichincha	D	8
3	4	Quito	Pichincha	D	9
4	5	Santo Domingo	Santo Domingo de los Tsachilas	D	4

```
[5]: df_transactions.head()
```

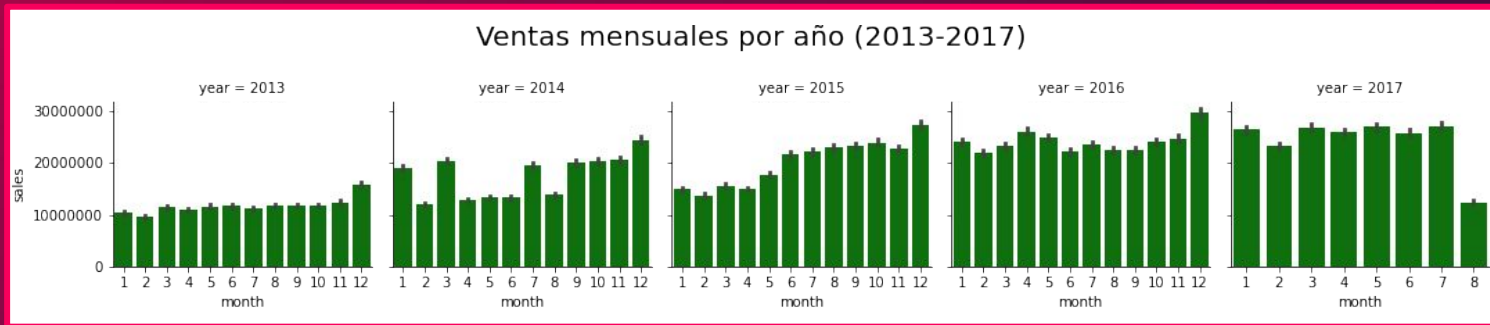
	date	store_nbr	transactions
0	2013-01-01	25	770
1	2013-01-02	1	2111
2	2013-01-02	2	2358
3	2013-01-02	3	3487
4	2013-01-02	4	1922

```
[7]: df_oils.head()
```

	date	dcoilwtico
0	2013-01-01	NaN
1	2013-01-02	93.14
2	2013-01-03	92.97
3	2013-01-04	93.12
4	2013-01-07	93.20

1) Ventas mensuales por año

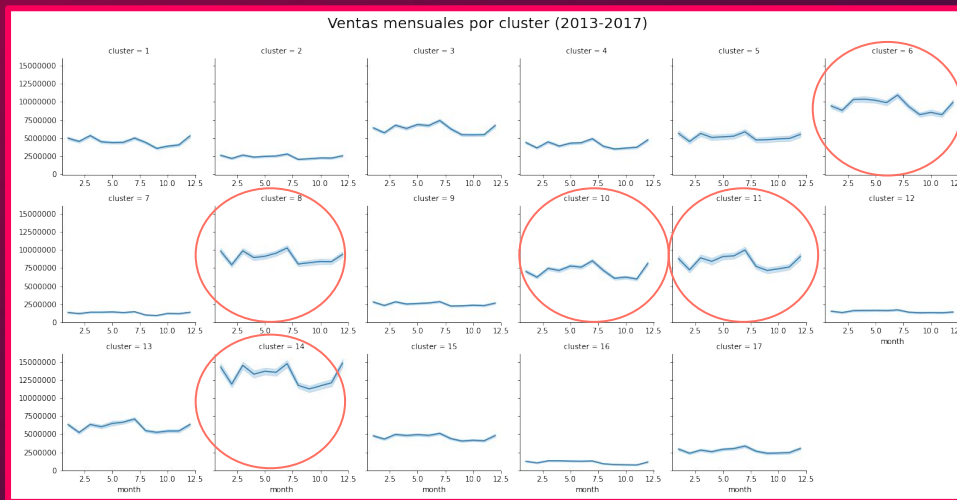
- Las ventas tienen un crecimiento cada año desde el año 2013 hasta el 2017
- **Todas las tiendas** tienen históricamente un **máximo** en el **mes 12, diciembre**
- **Mes** puede ser una columna significativa para la predicción de ventas



Análisis Descriptivo: Hallazgos

2) Ventas mensuales por cluster

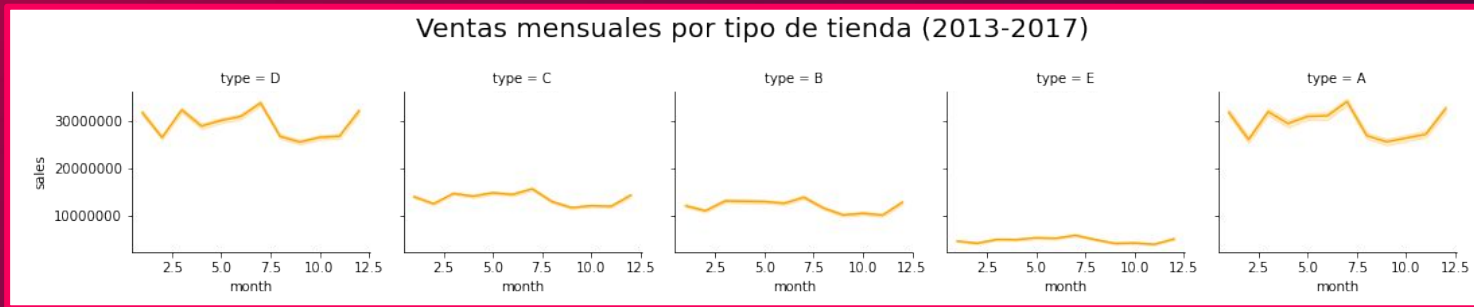
- Los cinco **clústeres** con **mayores ventas mensuales** consistentes durante todo el periodo analizado son el **6, 8, 10, 11 y 14**
- Prácticamente **todas las tiendas** tienen históricamente un **disminución local en agosto**
- **Cluster** puede tener un efecto en la predicción de ventas



Análisis Descriptivo: Hallazgos

3) Ventas mensuales por tipo de tienda

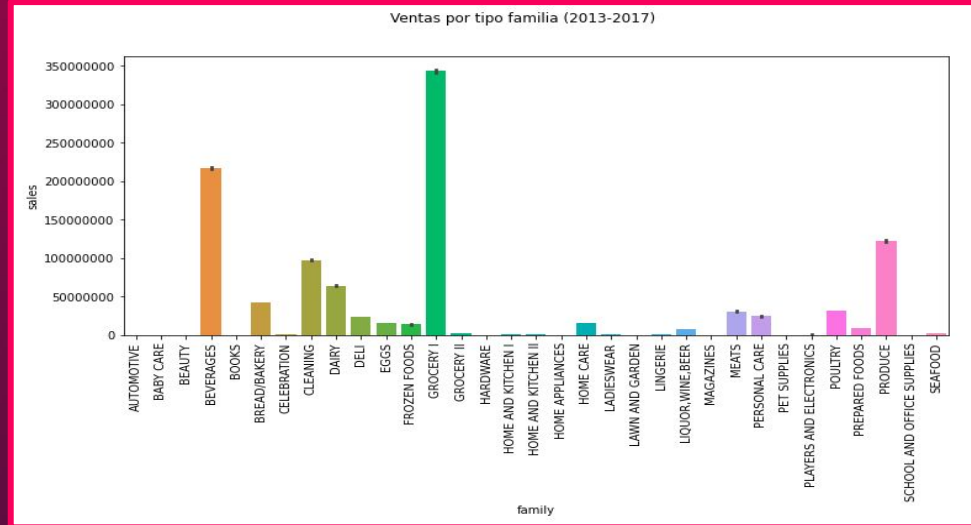
- Las ventas tienen un crecimiento cada año desde el año 2013 hasta el 2017
- **Todas las tiendas** tienen históricamente un **máximo** en el **mes 12, diciembre**
- **Mes** puede ser una columna significativa para la predicción de ventas



Análisis Descriptivo: Hallazgos

4) Ventas mensuales por familia de productos

- Las tres familias de productos con más ventas históricamente son Grocery I, Produce y Beverages
- Familias como Automotive, Baby Care, Beauty, Hardware, Home Appliances, Lawn and Garden, Pet Supplies, Electronics y Seafood tienen ventas mínimas comparadas con las demás



Análisis Estadístico

Análisis Estadístico: Hallazgos

sales	
count	3000888.00000
mean	357.77575
std	1101.99772
min	0.00000
25%	0.00000
50%	11.00000
75%	195.84725
max	124717.00000

	mean	min	max	std	var
year					
2013	216.47953	0.00000	46271.00000	734.07453	538865.41499
2014	322.93979	0.00000	45361.00000	1004.78445	1009591.78862
2015	371.35719	0.00000	40351.46000	1100.27800	1210611.67836
2016	443.79030	0.00000	124717.00000	1292.74033	1671177.56703
2017	480.12447	0.00000	38422.62500	1355.19131	1836543.48409

	mean	min	max	std	var
type					
A	705.878743	0.0	76090.000	1892.700760	3.582316e+06
B	326.739714	0.0	89576.360	977.528999	9.555629e+05
C	197.263301	0.0	45361.000	581.310901	3.379224e+05
D	350.979407	0.0	124717.000	965.728732	9.326320e+05
E	269.121301	0.0	16542.902	761.422519	5.797643e+05

1) Ventas por año:

- La media de ventas estuvo subiendo todos los años desde 2013 en \$216.48 hasta \$480.12 en 2017

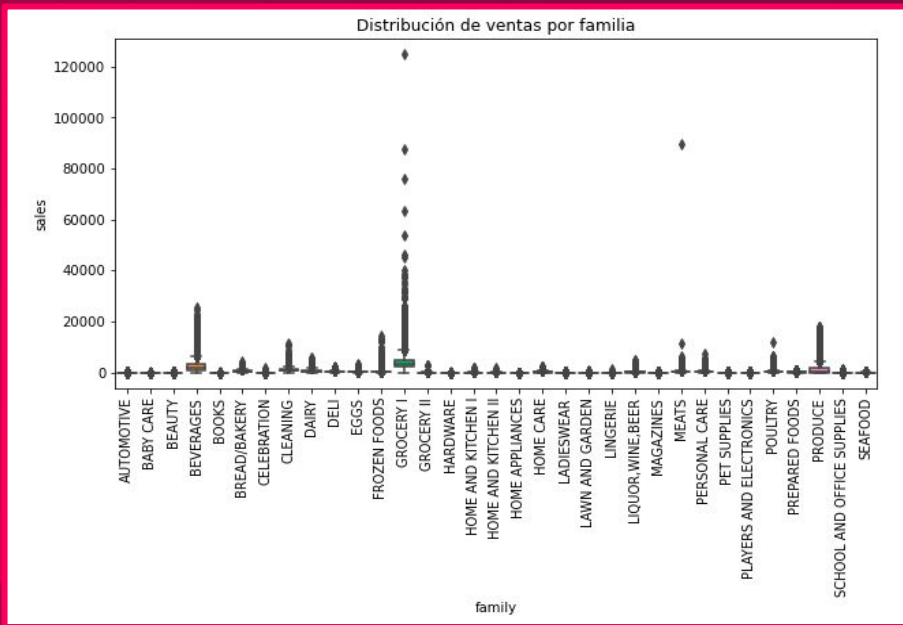
2) Ventas por tipo de tienda

- El tipo de tienda A a pesar de tener la mayor media tiene la mayor varianza
- El tipo de tienda D tiene un récord de ventas con una venta de \$124717

Análisis Estadístico: Hallazgos

Ventas por familia:

- La gran mayoría de familias no tiene ventas mensuales a 20,000
- Grocery y Carnes tienen los récords de ventas mensuales



	mean	min	max	std	var
family					
AUTOMOTIVE	6.10124	0.00000	255.00000	6.19169	38.33703
BABY CARE	0.11053	0.00000	116.00000	0.68106	0.46385
BEAUTY	3.71572	0.00000	136.00000	4.79467	22.98884
BEVERAGES	2385.79315	0.00000	25413.00000	2307.88231	5326320.73551
BOOKS	0.07080	0.00000	61.00000	0.54798	0.30028
BREAD/BAKERY	463.33625	0.00000	4551.29800	368.24637	135605.38669
CELEBRATION	8.37047	0.00000	1783.00000	15.92711	253.67296
CLEANING	1072.41674	0.00000	11377.00000	734.68149	539756.89626
DAIRY	709.15489	0.00000	5636.00000	671.94964	451516.31596
DELI	265.13507	0.00000	2118.32500	210.41707	44275.34481
EGGS	171.42052	0.00000	3110.00000	161.82678	26187.90782
FROZEN FOODS	154.76695	0.00000	14541.55000	324.97130	105606.34880
GROCERY I	3776.97210	0.00000	124717.00000	2874.20884	8261076.48349
GROCERY II	21.58405	0.00000	2931.00000	31.06242	964.87425
HARDWARE	1.13783	0.00000	49.00000	1.63804	2.68317
HOME AND KITCHEN I	20.47034	0.00000	1935.00000	35.96765	1293.67165
HOME AND KITCHEN II	16.72242	0.00000	1568.00000	21.29882	979.61623
HOME APPLIANCES	0.45748	0.00000	15.00000	0.96955	0.94003
HOME CARE	176.19803	0.00000	2504.00000	209.91658	44064.97038
LADIESWEAR	7.16063	0.00000	147.00000	13.94630	194.49933
LAWN AND GARDEN	6.03548	0.00000	712.00000	12.28346	150.88335
LINGERIE	7.18213	0.00000	1059.00000	9.47452	89.76653
LIQUOR, WINE, BEER	85.18782	0.00000	4821.00000	126.55564	16016.32914
MAGAZINES	2.92908	0.00000	79.00000	6.00849	36.10198
MEATS	341.84996	0.00000	89576.36000	455.90850	207852.55964
PERSONAL CARE	270.43251	0.00000	7504.00000	226.51201	51307.68940
PET SUPPLIES	3.92126	0.00000	106.00000	7.65785	58.64270
PLAYERS AND ELECTRONICS	6.18686	0.00000	269.00000	10.54450	111.18642
POULTRY	350.53229	0.00000	12143.20100	400.51163	160409.56682
PREPARED FOODS	96.77020	0.00000	912.56800	104.69514	10961.07324
PRODUCE	1349.35212	0.00000	17850.61500	2186.48133	4780700.61719
SCHOOL AND OFFICE SUPPLIES	2.96160	0.00000	1332.00000	2174594	472.88577
SEAFOOD	22.16319	0.00000	274.44000	32.62763	1064.56208

Feature Engineering

Feature Engineering

Promedio, máximo y mínimo móvil de ventas de cada familia de productos en cada tienda, para:

- **Últimos 23 días** (al predecir a 16 días, se pueden obtener al menos los 7 últimos datos)
- **Últimos 31 días** (promedio mensual)
- 6 columnas en total

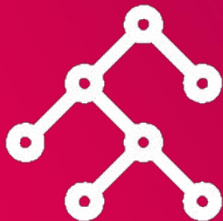
Promedio, máximo y mínimo móvil de ventas de cada familia de productos en cada tienda el último mismo día de la semana (último lunes, último martes, etc.)

- **Últimos 4 mismos días de la semana** (al predecir a 16 días, se pueden obtener al menos los 2 últimos datos)
- **Últimos 5 mismos días de la semana**
- **Últimos 6 mismos días de la semana**
- 9 columnas en total

Ejemplo		sales	weekDay_mean_last4	mean_last23days
Penúltimo registro	1500		1459.34	9530
Último registro	1400		235.88	1239
Primera predicción	-----		1450	1688.84

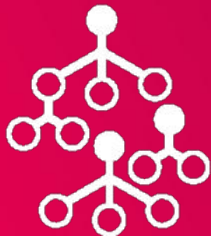
Modelado

Modelos / Algoritmos utilizados



DecisionTree (Regressor)

Una árbol de decisión que combina una secuencia de pruebas a los datos para ajustar su predicción



RandomForest (Regressor)

Una gran cantidad de árboles de decisión que operan como un conjunto

3 pasos:

Creación de
modelo



Ajuste a los
datos



Predicción con
nuevos datos

Experimentos

- **3,000,888 filas** en el set de datos de entrenamiento
- **Alta** cantidad de **procesamiento y cómputo** necesario
- Dado que los **recursos del proyecto** son **limitados**, se optó por realizar **pruebas con entrenamientos de muestras más pequeñas**

1

15,000 últimas filas

2

25,000 últimas filas

3

150,000 últimas filas

4


1,500,000 últimas filas

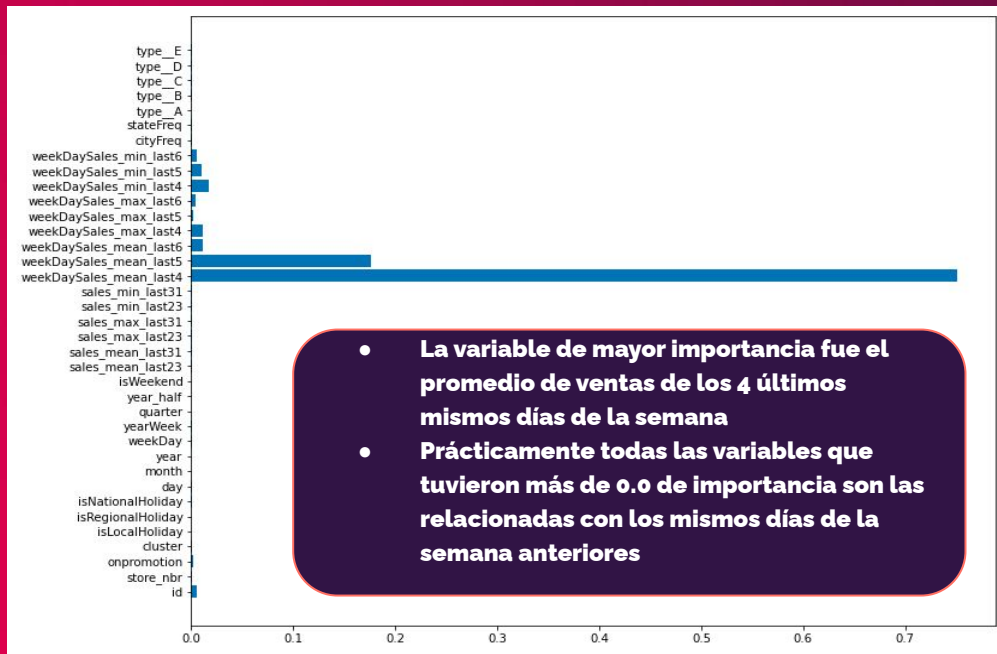
Resultados

Resultados finales: Mean Squared Logarithmic Error

Modelo / Tamaño de la muestra	15,000 filas	25,000 filas	150,000 filas	1,500,000 filas
RandomForest Regressor	0.47824	<u>0.46888</u>	0.47800	Kernel died
DecisionTree Regressor	0.58283	0.55402	0.57101	Kernel died

Mejor Modelo: Random Forest Regressor (25,000 filas)

#	Team	Members	Score	Entries	Last	Code
196	Rafael Velasco Huerta		0.46888	1	1s	



- La variable de mayor importancia fue el promedio de ventas de los 4 últimos mismos días de la semana
- Prácticamente todas las variables que tuvieron más de 0.0 de importancia son las relacionadas con los mismos días de la semana anteriores

Siguientes pasos

Siguientes pasos:

- Evaluar más tipos de modelos y continuar con Feature Engineering para buscar generar más variables relevantes para el modelo
- Buscar plataforma para correr los modelos con más registros/información
- Buscar la manera de automatizar la llegada de información y generar un pipeline que corra el modelo cada cierto periodo de tiempo
- Hacer el modelo un producto y establecer canales y roles de acceso

Bibliografía: Links

Metodología:

- Hotz, N. (2022, April 16). What is CRISP DM? Data Science Process Alliance. Retrieved June 15, 2022, from <https://www.datascience-pm.com/crisp-dm-2/>

Imágenes;

- https://www.freepik.es/vector-gratis/cajero-isometrico-autoservicio_8485241.htm#query=autoservicio&position=18&from_view=search
- https://www.flaticon.es/icono-gratis/inventario_2897785
- <https://www.subpng.com/png-borbyz/download.html>
- https://www.flaticon.es/icono-gratis/metas_3391881

...

Gracias!

.....