



Store Sales - Time Series Forecasting

Por Javier Martínez y Jorge López

GENERAL OVERVIEW



States



Stores



Months



Products



TARGET	MEASURE	ALGORITHM	
Predicción de Ventas de la Empresa "Corporación Favorita" entre 2017/08/15 y 2017/08/30	RMSLE (Error Log Cuadratico Medio)	LightGBM "Light Gradient Boosting Machine"	0.46566

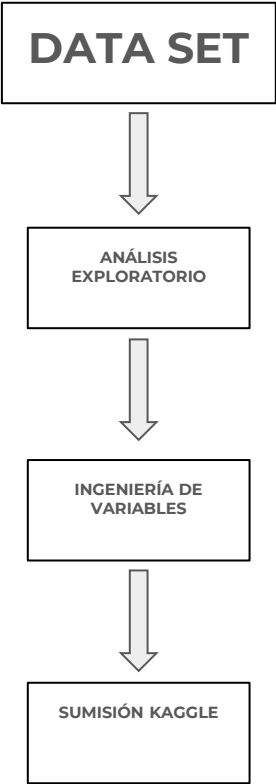
TRAIN

STORES

OIL

TRANSACTIONS

HOLIDAYS

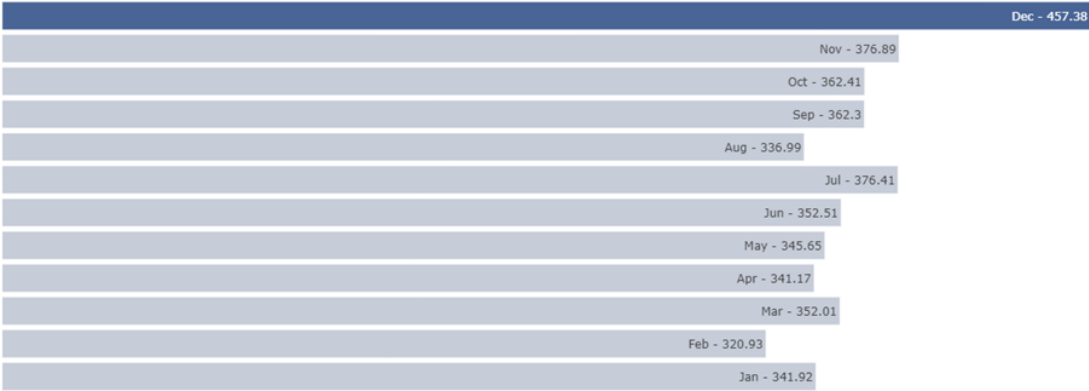


DATA SET
TRAIN

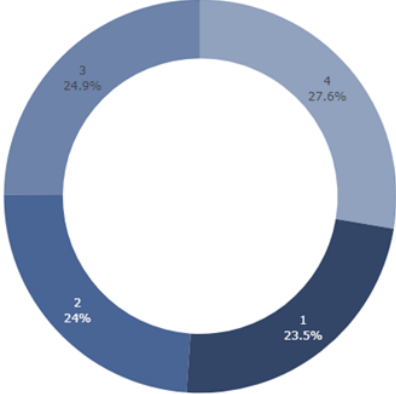
TRAIN - ANALISIS EXPLORATORIO - Average Sales Analysis

Average Sales Analysis

Month wise Avg Sales Analysis



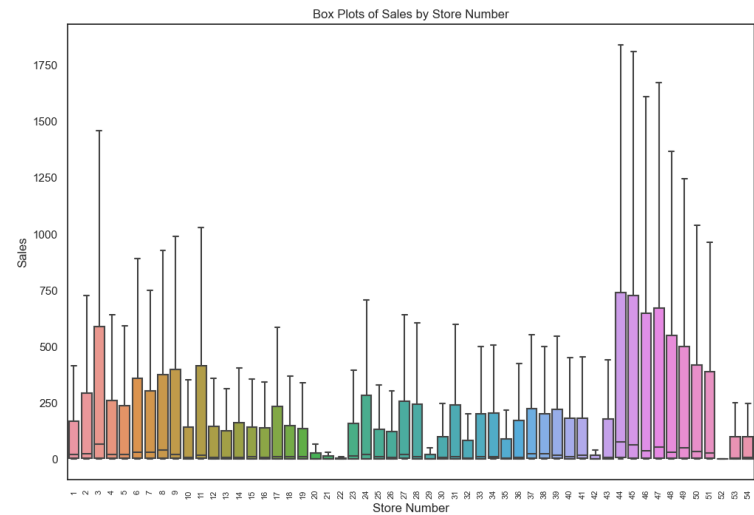
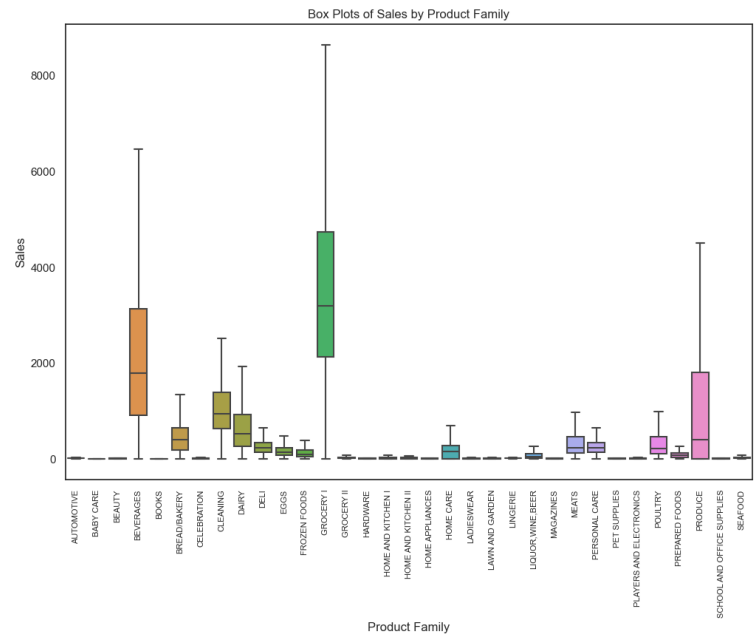
Quarter wise Avg Sales Analysis



Week wise Avg Sales Analysis



TRAIN - ANÁLISIS EXPLORATORIO - Sales by Product Family & Store

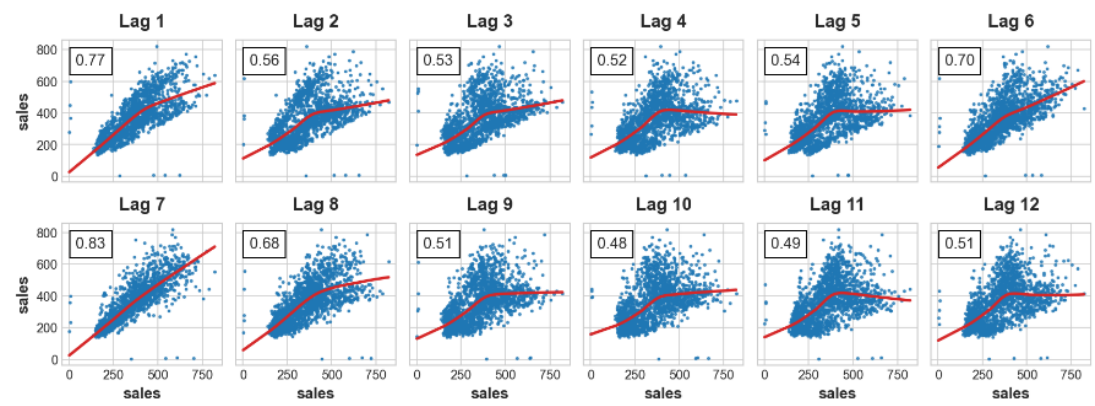


CONCLUSIÓN

Distribución por familia muy diferente

Separaremos el DATASET por familias en el Entrenamiento & Sumisión

LAG FEATURES - FIGURAS DE RETARDO [8]

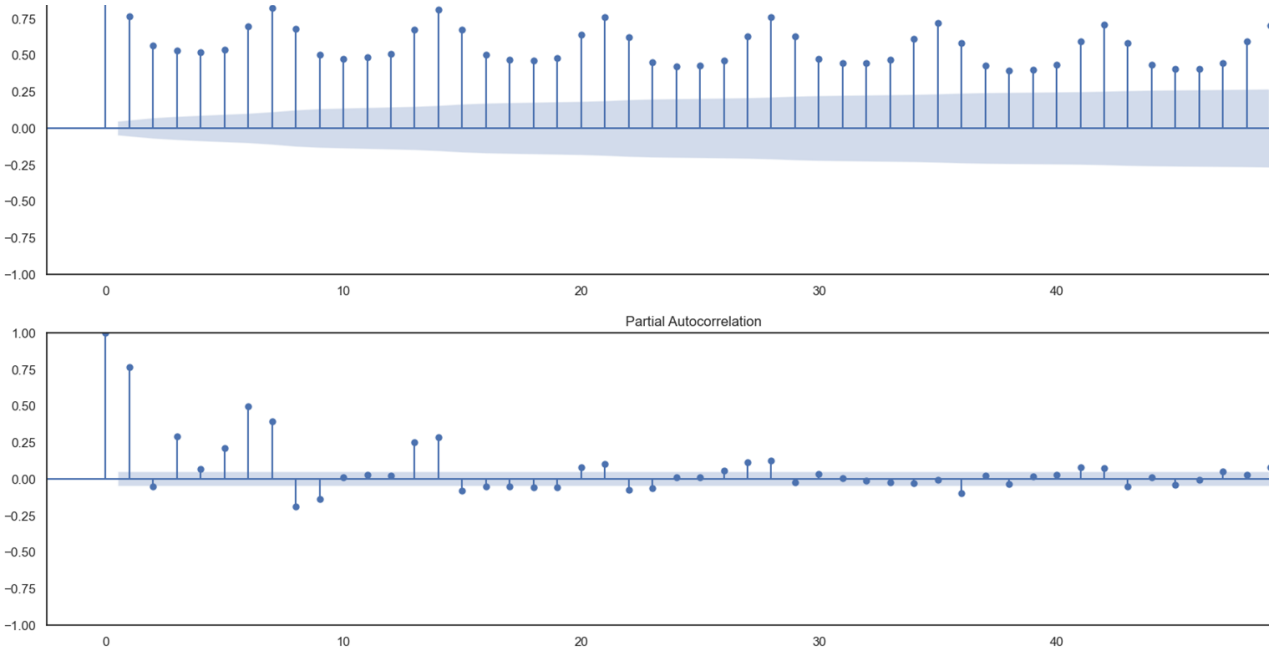


TIME RELATED FEATURES [15]



Autocorrelación

La autocorrelación nos brinda una ventana hacia la estructura interna de las series temporales. Al **comprender cómo los valores pasados afectan los valores presentes**, podemos **identificar patrones clave y tendencias** que son esenciales para la toma de decisiones informada en diversos campos, desde finanzas hasta operaciones comerciales.



TRAIN - Ingeniería de Variables- Figuras de tiempo

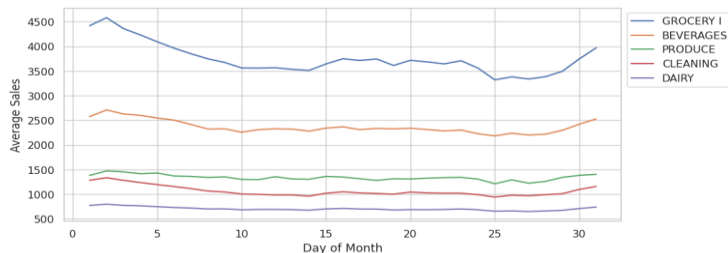
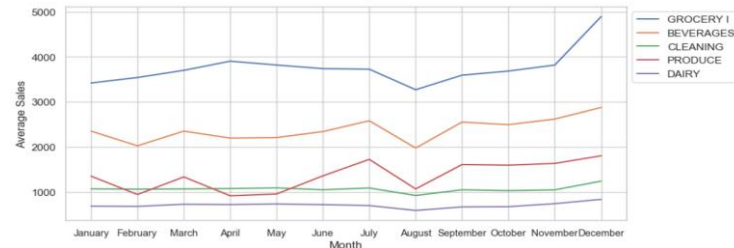
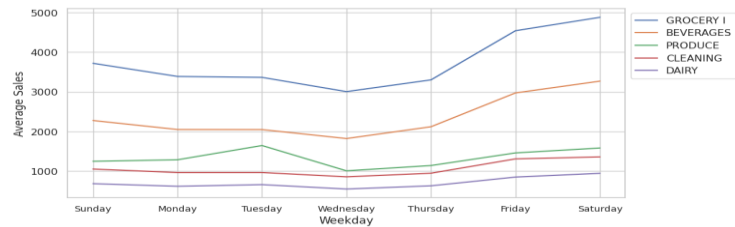
Patrones Estacionales y Temporales: Las variables como mes, día del mes, día de la semana y trimestre pueden ayudar a capturar patrones estacionales y temporales en los datos. Por ejemplo, ciertos meses o días de la semana pueden tener patrones de comportamiento distintivos.

Días Laborables y Fines de Semana: Creamos la variable `is_wknd` (es fin de semana) útil para diferenciar entre días laborables y fines de semana, ya que los patrones de comportamiento pueden variar significativamente entre estos dos tipos de días.

Inicio y Fin de Períodos: Creamos Variables como `is_month_start`, `is_month_end`, `is_quarter_start`, `is_quarter_end`, `is_year_start` y `is_year_end`. Pueden ser útiles para modelar cambios en el comportamiento al comienzo o al final de diferentes períodos de tiempo, como el mes, el trimestre o el año.

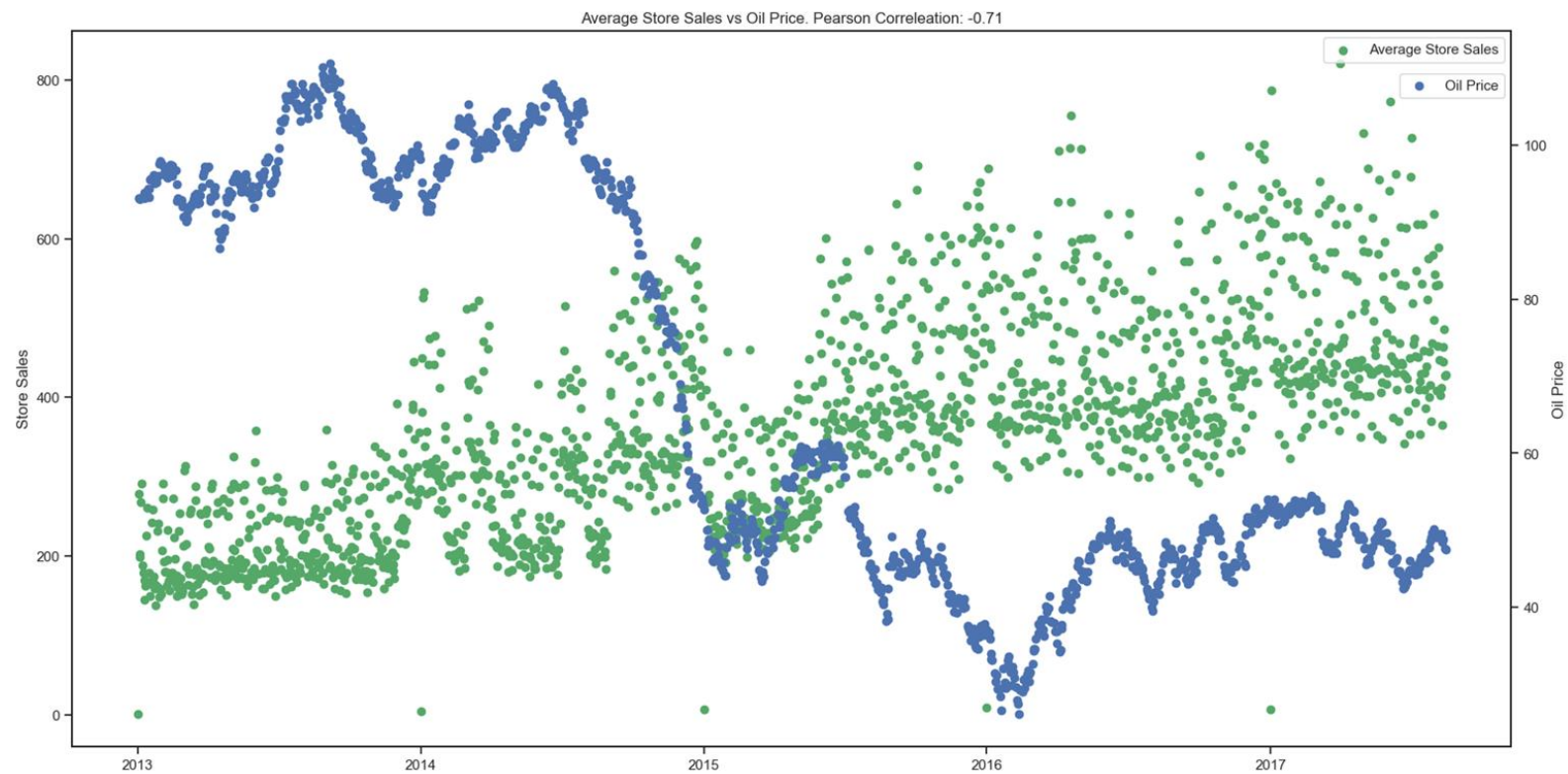
Estacionalidad Mensual y Trimestral: La variable `season` puede ayudar a modelar la estacionalidad mensual y trimestral, dividiendo el año en diferentes estaciones como invierno, primavera, verano y otoño.

*Creamos en total 15 va



DATA SET
OIL

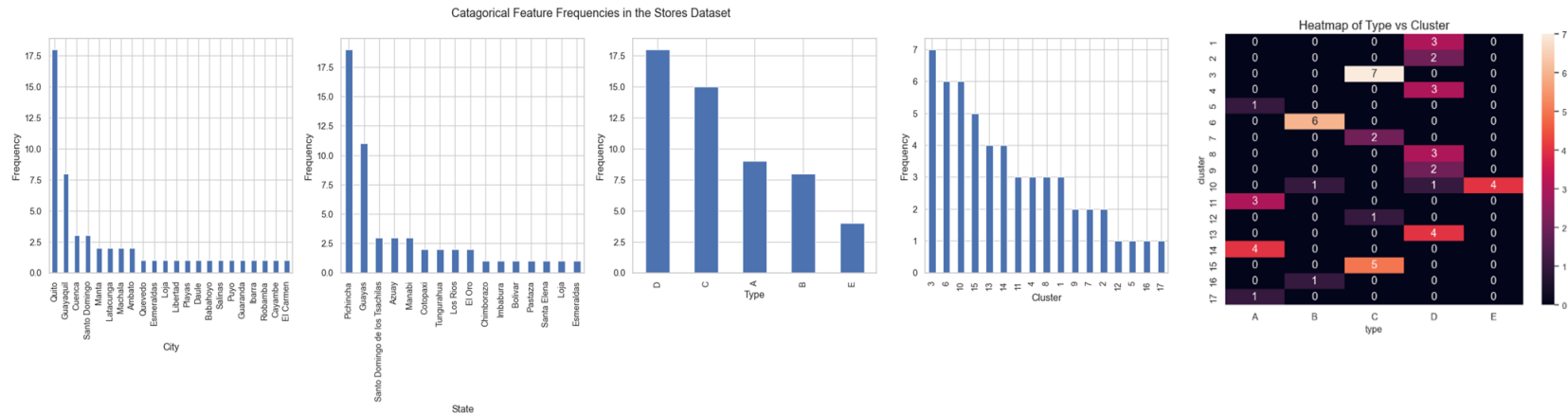
OIL - ANÁLISIS EXPLORATORIO & INGENIERÍA DE VARIABLES



43 Valores Nulos
=> Utilizamos una media Móvil de 5 días para completar los datos faltantes

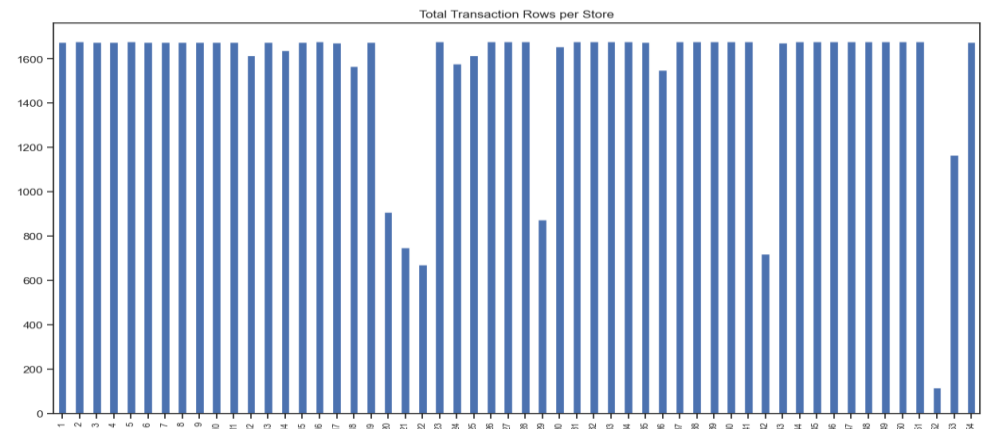
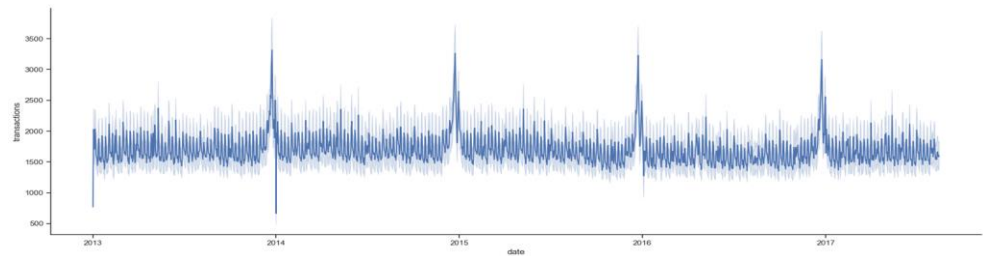
DATA SET STORES

STORES - ANÁLISIS EXPLORATORIO & INGENIERÍA DE VARIABLES



Realizaremos un OneHotEncoding para los distintos predictores categóricos dejando como resultado un DATASET de 54x64

DATA SET TRANSACTIONS

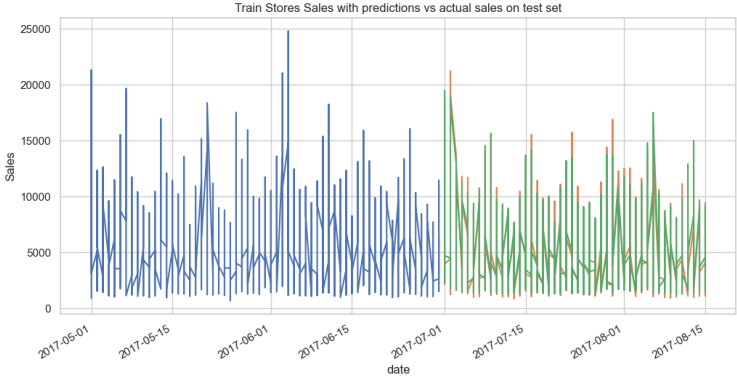


Ingeniería de Variables

- Estudio de Autocorrelación de la serie
- Creación de Lag Features (8)

SUMISIÓN - Sumisión 1

- En el Problema se explica un Terremoto que tuvo lugar el 16 de Marzo de 2016 -> Decidimos eliminar 4 semanas a partir del mencionado Evento
- Usando **LGBMRegressor** hacemos una predicción de los datos del dataframe de test y los comparamos con los que tenemos. (Lazy Predict)



NUMBER	RESULTADO
Sumisión 1	0.47079

DATA SET
HOLIDAYS

1) Problemas con el DATASET -> Limpieza

- Fiestas nacionales han sido transferidas.
- Eventos repetidos
- ¿Cuál es el ámbito de las fiestas? Puede ser regional, nacional o local. Hay que dividirlos según el ámbito.
- Otros

2) One Hot Encoding

- 50 Features

3) Contraste de Hipótesis:

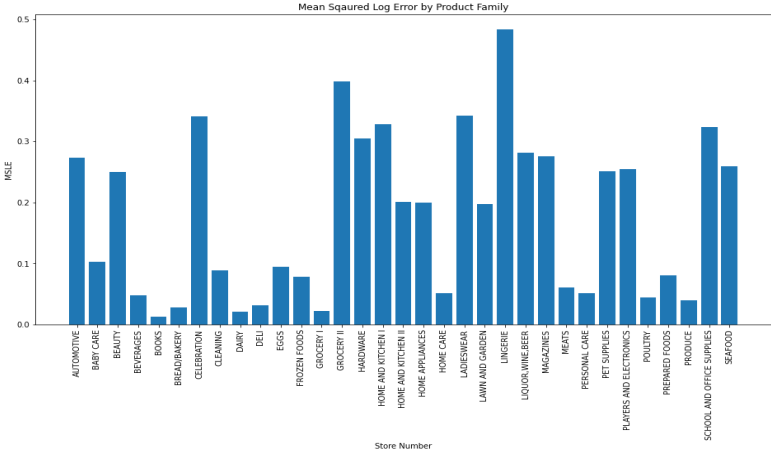
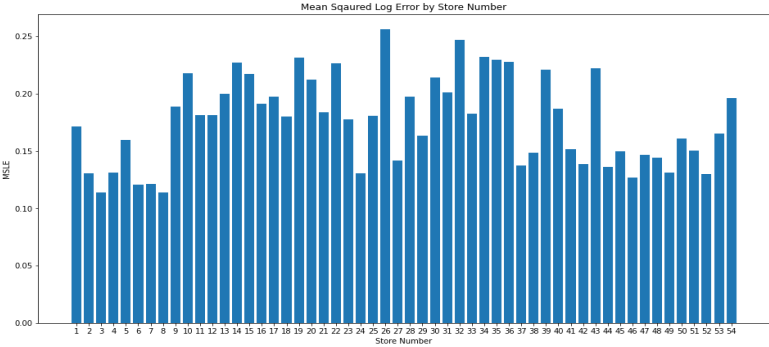
- H0: La Media de las Ventas son iguales para ese Evento/Fiesta (M1=M2) -> 17/50
- H1: Las Media de las Ventas NO son iguales para ese Evento/Fiesta (M1!=M2)

NUMBER	RESULTADO
Sumisión 1	0.47079
Sumisión 2	0.47129
Sumisión 3	0.46776

	Feature	Test Type	AB Hypothesis	p-value	Comment	GroupA_mean	GroupB_mean	GroupA_median	GroupB_median
0	events_Black_Friday	Non-Parametric	Reject H0	5.740308e-24	A/B groups are not similar!	363.360708	354.392622	17.000000	10.0
0	events_Cyber_Monday	Non-Parametric	Reject H0	2.150031e-31	A/B groups are not similar!	436.220250	354.263819	17.000000	10.0
0	events_Dia_de_la_Madre	Non-Parametric	Reject H0	9.085187e-03	A/B groups are not similar!	348.536505	354.425769	9.000000	10.0
0	events_Futbol	Non-Parametric	Reject H0	7.094536e-79	A/B groups are not similar!	310.965144	354.769187	5.000000	10.0
0	events_Terremoto_Manabi	Non-Parametric	Reject H0	0.000000e+00	A/B groups are not similar!	494.904072	351.798883	23.000000	10.0
0	holiday_national_binary	Non-Parametric	Reject H0	1.935148e-10	A/B groups are not similar!	424.447188	351.040370	11.000000	10.0
0	holiday_local_binary	Non-Parametric	Reject H0	2.329166e-12	A/B groups are not similar!	424.755295	354.129947	14.000000	10.0
0	holiday_regional_binary	Non-Parametric	Reject H0	1.430984e-03	A/B groups are not similar!	229.725442	354.450566	6.000000	10.0
0	holiday_national_Batalla_de_Pichincha	Non-Parametric	Fail to Reject H0	1.317803e-01	A/B groups are similar!	391.726745	354.298364	12.000000	10.0
0	holiday_national_Carnaval	Non-Parametric	Fail to Reject H0	3.663501e-01	A/B groups are similar!	332.882573	354.535820	10.000000	10.0
0	holiday_national_Dia_de_Difuntos	Non-Parametric	Reject H0	6.005582e-30	A/B groups are not similar!	431.034644	354.182412	16.000000	10.0

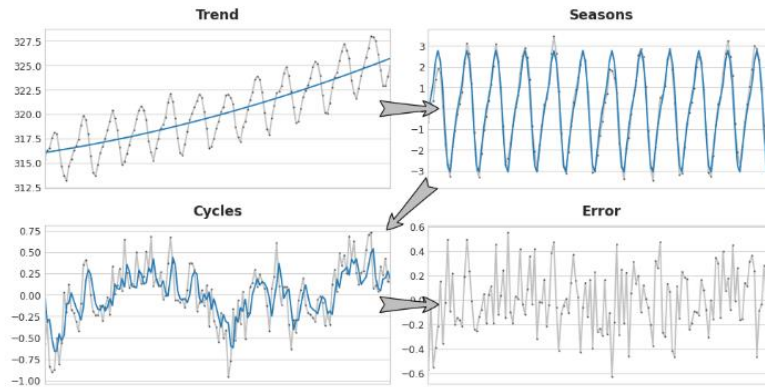
- Para esta sumisión, decidimos **igualar a cero todas las ventas de productos sobre tiendas que históricamente han sido cero**. Podrían haber sido productos eliminados de la tienda.

NUMBER	RESULTADO
Sumisión 4	0.46566



ACCIONES A ESTUDIAR

- **Promotion:** Agregar "Promotion" como Leading Indicator.
Hemos estado utilizando los Lags pero no hemos incorporado Leads para dar indicadores adelantados
- **LACS:** Problema con LAG-16.
Hemos Utilizado como primer Lag, Lag-16. Podríamos hacer un modelo en que la primera predicción se coloque como Lag-1 de la siguiente y así sucesivamente para la estructura de la sumisión 2017/08/15 - 2017/08/31
- **Desestacionalizar las series:** Para cada familia



GRACIAS !!