
Documentación Metodológica

Forecasting Ventas inmobiliarias

Jesús María - Miraflores

Julio 2022



Propuesta exclusiva para:



1. Objetivo

El presente documento tiene como objetivo exponer el desarrollo metodológico de los modelos de predicción de venta de inmuebles elaborados durante la consultoría brindada al equipo de ASEI.

2. Alcance

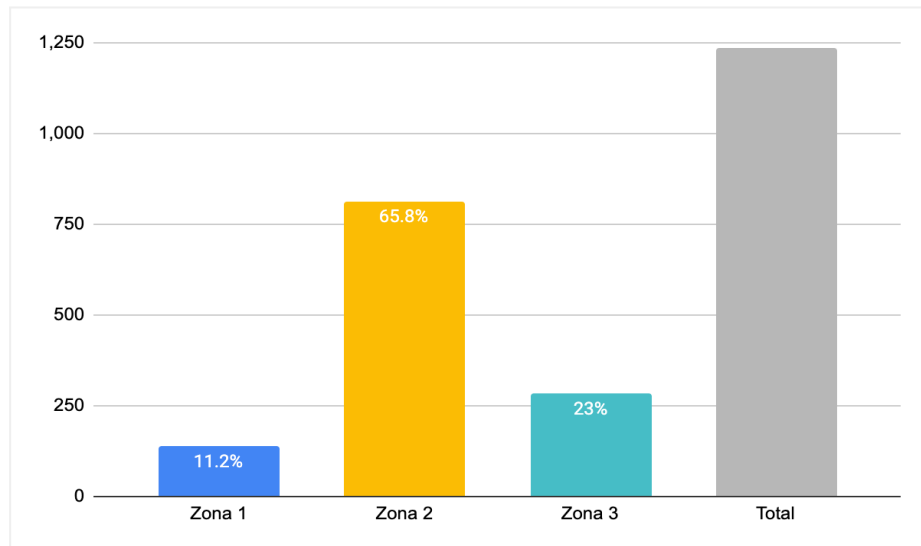
Los modelos construidos coberturan dos distritos de Lima metropolitana: **Jesús María** y **Miraflores** los cuales son divididos en zonas (1, 2 y 3) bajo el esquema de segmentación propuesta por el equipo de ASEI.

3. Análisis de segmentación

Como parte del proceso de evaluación se analizaron cada una de las series temporales brindadas por ASEI en cada una de las zonas por distritos:

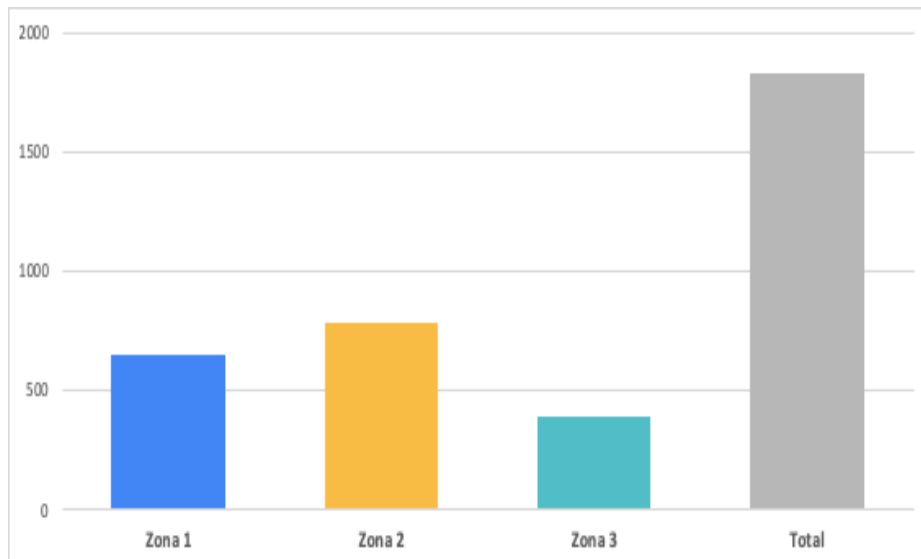
Miraflores

Distribución por Zona



Jesús María

Distribución por Zona



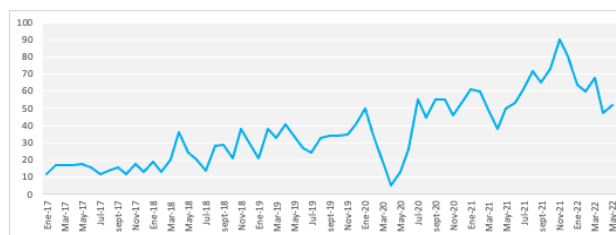
Periodo de análisis: Abr2021-Abr2022

Los gráficos de distribución demuestran que existen diferencias importantes en la cantidad de ventas colocadas por zonas siendo predominante en ambos distritos las colocaciones en la zona 2.

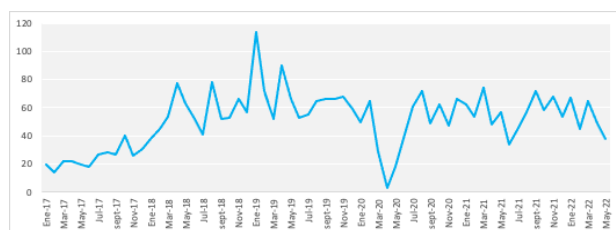
Asimismo se evaluó que la tendencia variaba de forma diferente en cada uno de las zonas de modelamiento por lo cual se tomó la decisión de modelar las series temporales por separado.

Las series totales reflejan el comportamiento de la zona con mayor materialidad por lo cual **será necesario trabajar un modelo por cada zona por distrito.**

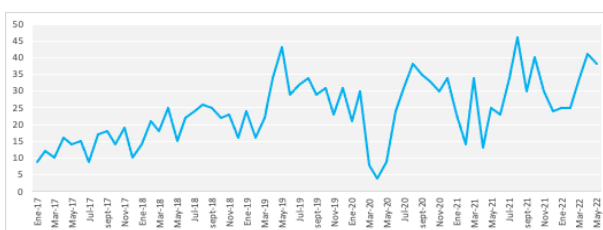
Zona 1



Zona 2



Zona 3



4. Información utilizada

El proceso de modelamiento incluyó dos principales fuentes de información compuestas por información brindada por el equipo de ASEI además de una segunda fuente conformada por data de series macroeconómicas extraída de la fuente del BCR.

Información ASEI

- Ventas mensuales totales y por zona (Jesus María y Miraflores)
- Precios de oferta de inmuebles totales y por zona (Jesus María y Miraflores)
- Precios de venta de inmuebles totales y por zona (Jesus María y Miraflores)
- Fechas de ferias inmobiliarias

Información BCR

- Tasa de interés de fondeo
- Tasa de interés activa

- Índice de precios inmueble
- Índice de precios por m2 en distritos medios
- Índice de precios por m2 en distritos altos*
- Índice de precios por m2 en 12 distritos
- Índice de precios de consumidor Lima met. (alimentos & energía)
- Producción de energía Lima
- Importación de bienes de consumo duradero
- índice de coyuntura de energía
- índice de coyuntura de energía sin minas
- índice de coyuntura de consumo de energía
- Variación porcentual demanda interna
- Variación porcentual del PBI*
- Variación porcentual del PBI - sector construcción*
- índice de precios de inflación subyacente de bienes
- Importaciones de materiales de construcción

- Índice de precios de importaciones
- Expectativa de PBI
- Tipo de cambio bancario mensual
- Indicador de variación mensual del tipo de cambio

Variables con Alto potencial

Identificamos **variables con alto valor para potenciar los modelos**, evaluaremos a profundidad las interacciones que se pueden tener para predecir nuestras series de interés.

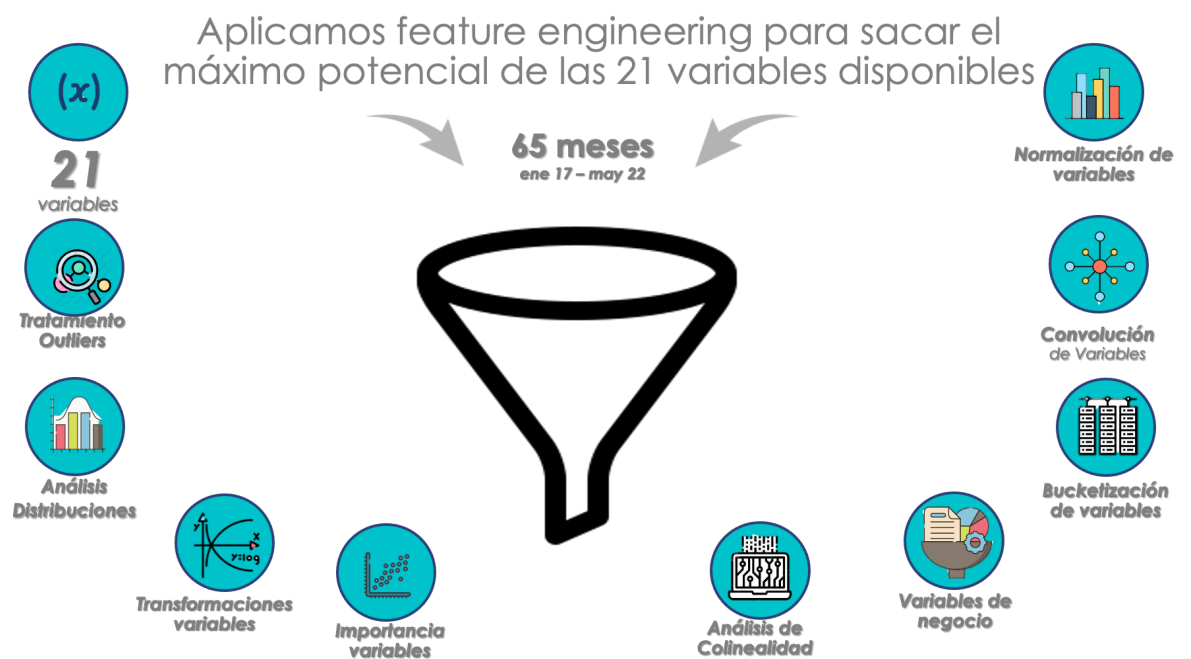
Target: Venta total
índice correlación

Índice de precios de importaciones	
Índice de precios del consumidor alimentos y energía (Lima)	
Índice de precios de inmuebles en distritos de medio valor	
Oferta total del mes previo	
Tipo de cambio bancario mensual	

--	--

Feature Engineering

Las variables identificadas pasaron sobre procesos de feature engineering para potenciar su valor predictivo.

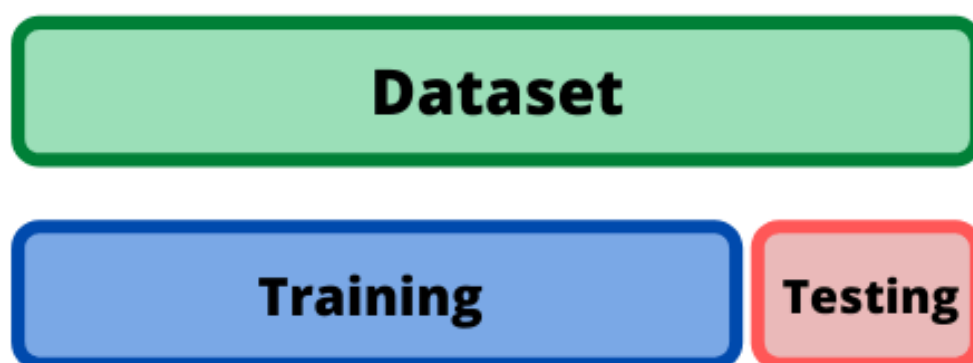


El featurring consistió en la creación de variables derivadas del tipo:

- Promedios 3, 6, y 12 últimos meses de cada variable
- Ratios promedio 3/6, 6/12 meses de cada variable
- Estandarización de las variables (tomando medias y desviaciones de periodo train)

Definición muestral

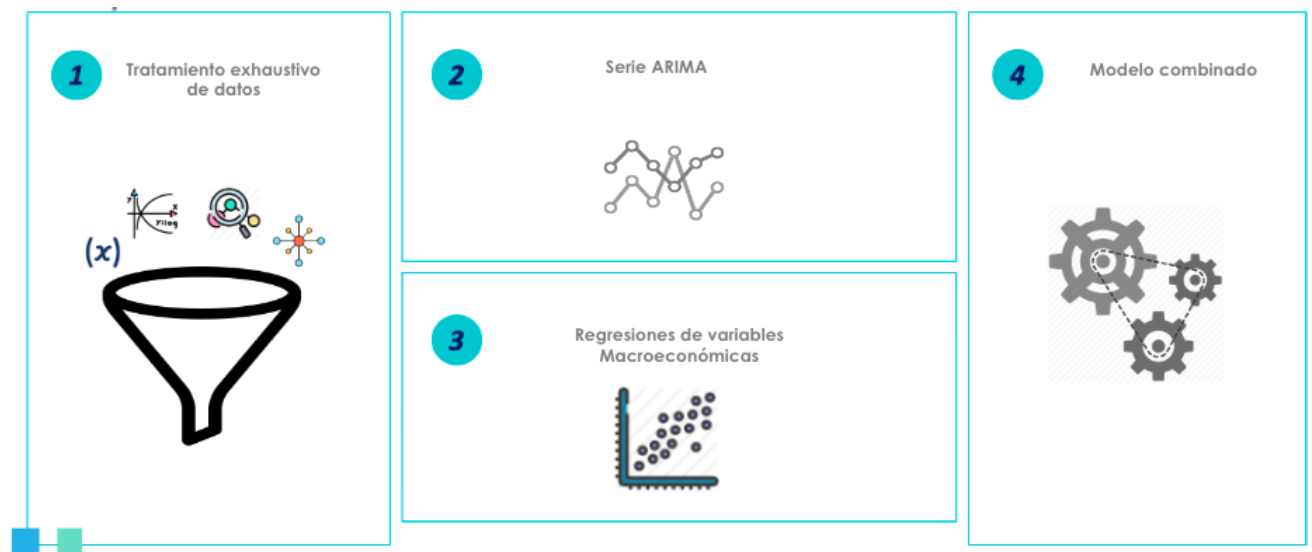
Para probar de forma adecuada el funcionamiento del modelo trabajamos con un enfoque muestral OOT, el cual sugiere reservar los últimos periodos de la data para probar el desempeño del modelo.



Los modelos tomarán una ventana OOT de los 2 últimos meses de información disponible brindada por ASEI a la fecha de ejecución del modelo.

5. Proceso de Modelamiento

El desarrollo metodológico se dará de forma estratégica en 3 etapas:



La etapa 1 expuesta en el apartado anterior tendrá por finalidad extraer el máximo provecho de las posibles interacciones y patrones de la data.

La etapa 2 consiste en desarrollar un modelo ARIMA basado en el aprendizaje de la misma serie temporal tomando como input la misma serie y su comportamiento histórico.

La etapa 3 desarrollará un modelo de regresión lineal donde nuestro sistema aprenderá de las

interacciones y elasticidades que tienen los componentes macroeconómicos sobre la venta de inmuebles.

La etapa 4 finalmente tendrá como objetivo ensamblar los modelos desarrollados en la etapa 2 y 3. En los modelos de Miraflores se usó la metodología de ensamble por medias y en Jesús María se utilizaron modelos var (Vectores autorregresivos) para ensamblar los modelos.

6. Desarrollo de modelos

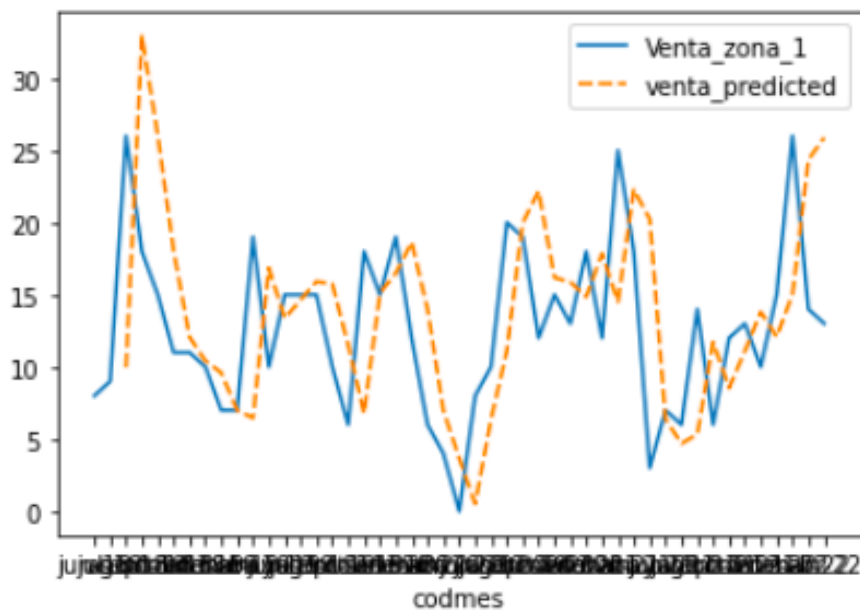
Miraflores Zona 1

MZ1 - ARIMA

El modelo Arima y sus componentes p , d y q se basaron en aspectos de pruebas de hipótesis basados en pruebas de dicky-fuller y de diferenciación desarrollados en los notebooks de trabajo además de contrastar los resultados con el modelamiento de la serie y los p -values obtenidos para cada grado de autoregresión y estacionariedad.

modelo resultado:

ARIMA Model Results						
Dep. Variable:	D2.Venta_zona_1	No. Observations:	43			
Model:	ARIMA(1, 2, 1)	Log Likelihood	-143.792			
Method:	css-mle	S.D. of innovations	6.517			
Date:	Thu, 21 Jul 2022	AIC	297.585			
Time:	16:18:02	BIC	306.391			
Sample:	2	HQIC	300.832			
	coef	std err	z	P> z	[0.025	0.975]
const	1.0562	1.257	0.840	0.406	-1.408	3.520
Venta_zona_1_lag2_avg3	-0.0840	0.102	-0.824	0.415	-0.284	0.116
ar.L1.D2.Venta_zona_1	-0.2812	0.148	-1.894	0.066	-0.572	0.010
ma.L1.D2.Venta_zona_1	-1.0000	0.064	-15.505	0.000	-1.126	-0.874
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-3.5561	+0.0000j	3.5561	0.5000		
MA.1	1.0000	+0.0000j	1.0000	0.0000		



MZ1 - REGRESIÓN MACRO

El modelo de regresión macro contiene las siguientes

variables y coeficientes correspondientes:

OLS Regression Results						
=====						
Dep. Variable:	Venta_zona_1	R-squared:	0.307			
Model:	OLS	Adj. R-squared:	0.253			
Method:	Least Squares	F-statistic:	5.751			
Date:	Thu, 21 Jul 2022	Prob (F-statistic):	0.00234			
Time:	17:12:43	Log-Likelihood:	-129.54			
No. Observations:	43	AIC:	267.1			
Df Residuals:	39	BIC:	274.1			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

intercept	12.7242	0.789	16.135	0.000	11.129	14.319
precio_oferta_zon1_rat3to6_std	-2.2126	0.788	-2.809	0.008	-3.806	-0.619
exp_PBI_avg3_std	1.2573	0.793	1.586	0.121	-0.346	2.861
flg_feia_avg3_std	1.8231	0.789	2.312	0.026	0.228	3.418
=====						
Omnibus:	3.670	Durbin-Watson:	1.614			
Prob(Omnibus):	0.160	Jarque-Bera (JB):	3.280			
Skew:	0.671	Prob(JB):	0.194			
Kurtosis:	2.834	Cond. No.	1.18			

Cabe mencionar que algunos p-values superan los valores de 0.05 (confianza de 95%) bajo decisión experta pues se consideró adecuado para el modelo considerar algunos componentes sacrificando significancia estadística para mejorar la robustez que puede presentar el modelo frente a cambios sistémicos del país.

MZ1 - COMBINACIÓN

La combinación de modelos en esta zona se realizó bajo la metodología de ensamble por medias que consiste en promediar el resultado de ambos modelos para la estimación final.

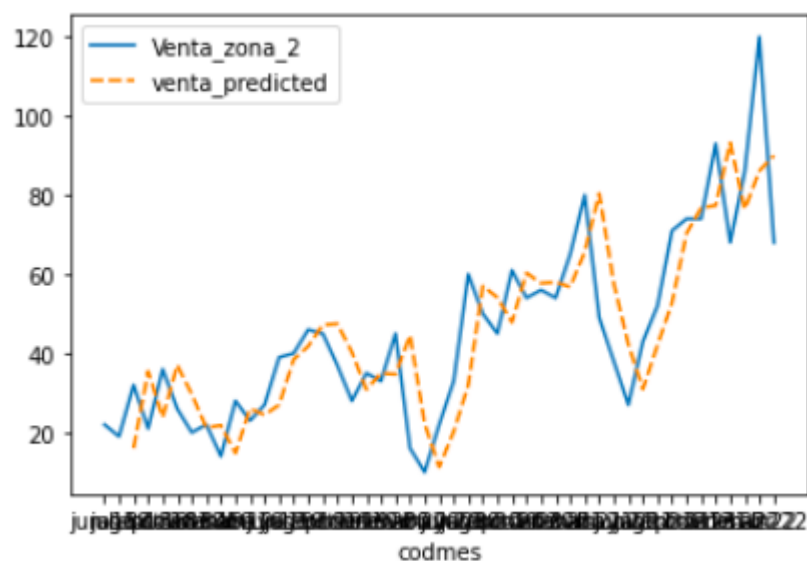
Miraflores Zona 2

MZ2 - ARIMA

El modelo Arima y sus componentes p, d y q se basaron en aspectos de pruebas de hipótesis basados en pruebas de dicky-fuller y de diferenciación desarrollados en los notebooks de trabajo además de contrastar los resultados con el modelamiento de la serie y los p-values obtenidos para cada grado de autoregresión y estacionariedad.

modelo resultado:

ARIMA Model Results						
Dep. Variable:	D2.Venta_zona_2	No. Observations:	43			
Model:	ARIMA(1, 2, 1)	Log Likelihood	-172.267			
Method:	css-mle	S.D. of innovations	12.663			
Date:	Sun, 24 Jul 2022	AIC	354.534			
Time:	05:52:51	BIC	363.340			
Sample:	2	HQIC	357.781			
	coef	std err	z	P> z	[0.025	0.975]
const	-0.0173	0.839	-0.021	0.984	-1.662	1.627
Venta_zona_2_lag2_avg3	0.0021	0.022	0.096	0.924	-0.040	0.044
ar.L1.D2.Venta_zona_2	-0.1969	0.151	-1.302	0.200	-0.493	0.099
ma.L1.D2.Venta_zona_2	-1.0000	0.062	-16.210	0.000	-1.121	-0.879
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-5.0797	+0.0000j	5.0797	0.5000		
MA.1	1.0000	+0.0000j	1.0000	0.0000		



MZ2 - REGRESIÓN MACRO

El modelo de regresión macro contiene las siguientes variables y coeficientes correspondientes:

OLS Regression Results						
=====						
Dep. Variable:	Venta_zona_2	R-squared:	0.739			
Model:	OLS	Adj. R-squared:	0.719			
Method:	Least Squares	F-statistic:	36.78			
Date:	Sun, 24 Jul 2022	Prob (F-statistic):	1.87e-11			
Time:	05:53:20	Log-Likelihood:	-161.27			
No. Observations:	43	AIC:	330.5			
Df Residuals:	39	BIC:	337.6			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

intercept	42.7099	1.651	25.868	0.000	39.370	46.050
exp_FBI_avg3_std	5.9631	1.657	3.600	0.001	2.612	9.314
ind_prec_cons_lima_met_alim_ener_avg3_std	15.6759	1.684	9.307	0.000	12.269	19.083
flg_feia_avg3_std	3.8777	1.647	2.354	0.024	0.546	7.209
=====						
Omnibus:	1.317	Durbin-Watson:	1.517			
Prob(Omnibus):	0.518	Jarque-Bera (JB):	1.011			
Skew:	-0.373	Prob(JB):	0.603			
Kurtosis:	2.915	Cond. No.	1.17			

Cabe mencionar que algunos p-values superan los

valores de 0.05 (confianza de 95%) bajo decisión experta pues se consideró adecuado para el modelo considerar algunos componentes sacrificando significancia estadística para mejorar la robustez que puede presentar el modelo frente a cambios sistémicos del país.

MZ2 - COMBINACIÓN

La combinación de modelos en esta zona se realizó bajo la metodología de ensamble por medias que consiste en promediar el resultado de ambos modelos para la estimación final.

Miraflores Zona 3

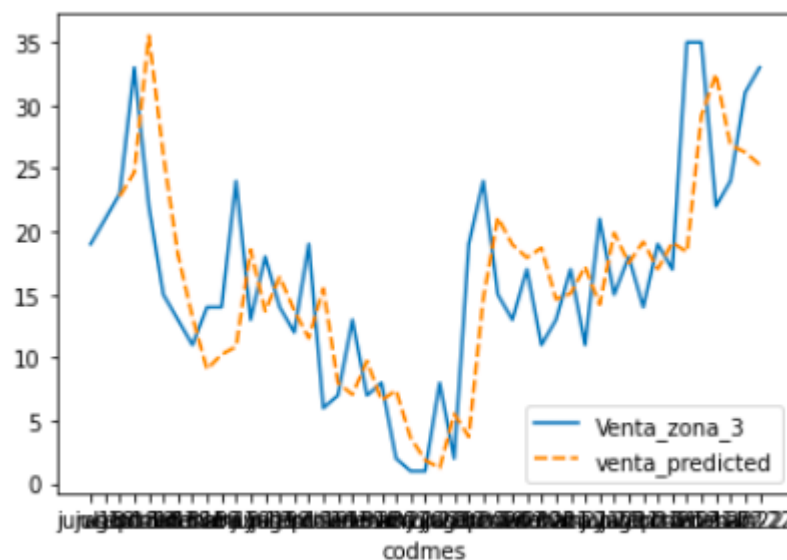
MZ3 - ARIMA

El modelo Arima y sus componentes p , d y q se basaron en aspectos de pruebas de hipótesis basados en pruebas de dicky-fuller y de diferenciación desarrollados en los notebooks de trabajo además de contrastar los resultados con el modelamiento de la serie y los p-values obtenidos

para cada grado de autoregresión y estacionariedad.

modelo resultado:

ARIMA Model Results						
Dep. Variable:	D2.Venta_zona_3	No. Observations:	43			
Model:	ARIMA(3, 2, 1)	Log Likelihood	-143.487			
Method:	css-mle	S.D. of innovations	6.417			
Date:	Sun, 24 Jul 2022	AIC	300.974			
Time:	19:48:05	BIC	313.303			
Sample:	2	HQIC	305.521			
	coef	std err	z	P> z	[0.025	0.975]
const	0.4113	0.353	1.164	0.252	-0.282	1.104
Venta_zona_3_lag2_avg3	-0.0281	0.028	-1.019	0.315	-0.082	0.026
ar.L1.D2.Venta_zona_3	-0.4224	0.155	-2.729	0.010	-0.726	-0.119
ar.L2.D2.Venta_zona_3	-0.2348	0.178	-1.319	0.195	-0.584	0.114
ar.L3.D2.Venta_zona_3	-0.0782	0.172	-0.455	0.652	-0.415	0.259
ma.L1.D2.Venta_zona_3	-1.0000	0.063	-15.872	0.000	-1.123	-0.877
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-0.1331	-2.1579j	2.1620	-0.2598		
AR.2	-0.1331	+2.1579j	2.1620	0.2598		
AR.3	-2.7361	-0.0000j	2.7361	-0.5000		
MA.1	1.0000	+0.0000j	1.0000	0.0000		



MZ3 - REGRESIÓN MACRO

El modelo de regresión macro contiene las siguientes variables y coeficientes correspondientes:

OLS Regression Results						
=====						
Dep. Variable:	Venta_zona_3	R-squared:	0.440			
Model:	OLS	Adj. R-squared:	0.413			
Method:	Least Squares	F-statistic:	16.51			
Date:	Wed, 13 Jul 2022	Prob (F-statistic):	5.14e-06			
Time:	20:12:44	Log-Likelihood:	-143.74			
No. Observations:	45	AIC:	293.5			
Df Residuals:	42	BIC:	298.9			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

intercept	15.5556	0.911	17.082	0.000	13.718	17.393
ind_prec_inf_suby_bienes_avg3_std	4.8866	0.935	5.226	0.000	3.000	6.774
precio_venta_total_rat3to6_std	-1.3525	0.935	-1.446	0.155	-3.239	0.534
=====						
Omnibus:	0.002	Durbin-Watson:	1.373			
Prob(Omnibus):	0.999	Jarque-Bera (JB):	0.133			
Skew:	0.011	Prob(JB):	0.936			
Kurtosis:	2.734	Cond. No.	1.19			

Cabe mencionar que algunos p-values superan los valores de 0.05 (confianza de 95%) bajo decisión experta pues se consideró adecuado para el modelo considerar algunos componentes sacrificando significancia estadística para mejorar la robustez que puede presentar el modelo frente a cambios sistémicos del país.

MZ3 - COMBINACIÓN

La combinación de modelos en esta zona se realizó bajo la metodología de ensamble por medias que consiste en promediar el resultado de ambos modelos para la estimación final.

Jesús María Zona 1

JMZ1 - ARIMA

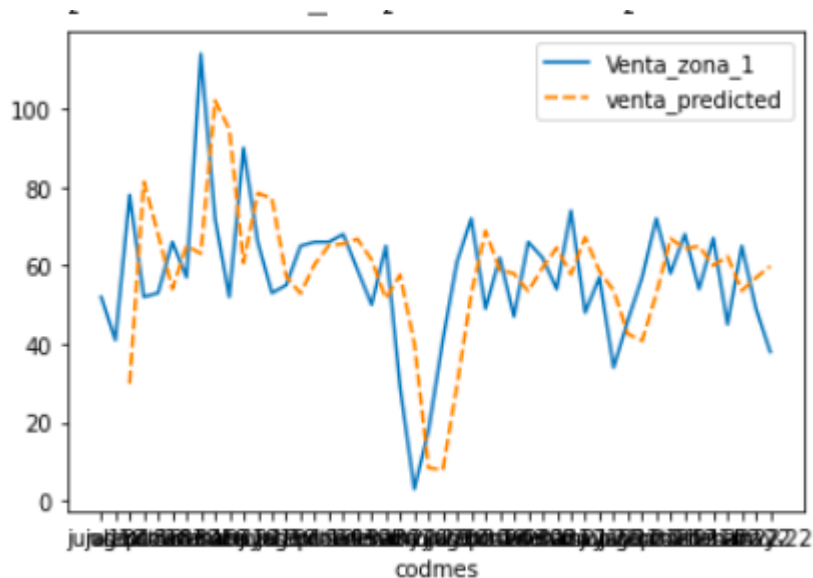
El modelo Arima y sus componentes p, d y q se basaron en aspectos de pruebas de hipótesis basados en pruebas de dicky-fuller y de diferenciación desarrollados en los notebooks de trabajo además de contrastar los resultados con el modelamiento de la serie y los p-values obtenidos para cada grado de autoregresión y estacionariedad.

modelo resultado:

ARIMA Model Results						
=====						
Dep. Variable:	D2.Venta_zona_1	No. Observations:	44			
Model:	ARIMA(1, 2, 1)	Log Likelihood	-194.030			
Method:	css-mle	S.D. of innovations	18.885			
Date:	Mon, 25 Jul 2022	AIC	398.060			
Time:	07:08:49	BIC	406.981			
Sample:	2	HQIC	401.369			
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	0.8765	2.350	0.373	0.711	-3.730	5.483
Venta_zona_1_lag3_avg3	-0.0158	0.041	-0.383	0.704	-0.097	0.065
ar.L1.D2.Venta_zona_1	-0.3925	0.138	-2.835	0.007	-0.664	-0.121
ma.L1.D2.Venta_zona_1	-0.9998	0.060	-16.729	0.000	-1.117	-0.883
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	-2.5480	+0.0000j	2.5480	0.5000		
MA.1	1.0002	+0.0000j	1.0002	0.0000		



JMZ1 - REGRESIÓN MACRO

El modelo de regresión macro contiene las siguientes variables y coeficientes correspondientes:

OLS Regression Results						
=====						
Dep. Variable:	Venta_zona_1	R-squared:	0.215			
Model:	OLS	Adj. R-squared:	0.177			
Method:	Least Squares	F-statistic:	5.614			
Date:	Mon, 25 Jul 2022	Prob (F-statistic):	0.00700			
Time:	07:09:16	Log-Likelihood:	-182.87			
No. Observations:	44	AIC:	371.7			
Df Residuals:	41	BIC:	377.1			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

intercept	57.6329	2.418	23.839	0.000	52.751	62.515
ind_prec_inm_rat3to6_std	7.6093	2.511	3.030	0.004	2.538	12.680
ind_prec_inf_suby_bienes_avg6_std	5.9194	2.597	2.279	0.028	0.674	11.165
=====						
Omnibus:	2.788	Durbin-Watson:	1.630			
Prob(Omnibus):	0.248	Jarque-Bera (JB):	1.898			
Skew:	0.191	Prob(JB):	0.387			
Kurtosis:	3.943	Cond. No.	1.38			

Cabe mencionar que algunos p-values superan los valores de 0.05 (confianza de 95%) bajo decisión

experta pues se consideró adecuado para el modelo considerar algunos componentes sacrificando significancia estadística para mejorar la robustez que puede presentar el modelo frente a cambios sistémicos del país.

JMZ1 - COMBINACIÓN

La combinación de modelos en esta zona se realizó bajo la metodología de vectores autorregresivos la cual combina ambos resultados como parte de un modelo ARIMA con las mismas consideraciones del modelo anterior.

ARIMA Model Results						
Dep. Variable:	D2.Venta_zona_1	No. Observations:	42			
Model:	ARIMA(1, 2, 1)	Log Likelihood	-183.694			
Method:	css-mle	S.D. of innovations	18.178			
Date:	Mon, 25 Jul 2022	AIC	379.389			
Time:	07:09:30	BIC	389.815			
Sample:	2	HQIC	383.210			
	coef	std err	z	P> z	[0.025	0.975]
const	1.2984	5.381	0.241	0.811	-9.249	11.846
venta_predicted	-0.0777	0.043	-1.789	0.082	-0.163	0.007
macro_pred	0.0539	0.107	0.504	0.618	-0.156	0.264
ar.L1.D2.Venta_zona_1	-0.3945	0.146	-2.695	0.011	-0.681	-0.108
ma.L1.D2.Venta_zona_1	-1.0000	0.069	-14.457	0.000	-1.136	-0.864
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-2.5351	+0.0000j	2.5351	0.5000		
MA.1	1.0000	+0.0000j	1.0000	0.0000		

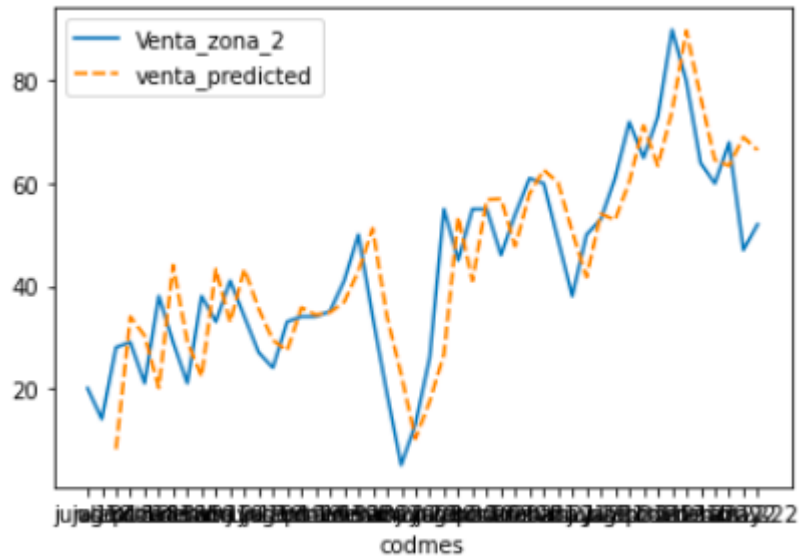
Jesús María Zona 2

JMZ2 - ARIMA

El modelo Arima y sus componentes p, d y q se basaron en aspectos de pruebas de hipótesis basados en pruebas de dicky-fuller y de diferenciación desarrollados en los notebooks de trabajo además de contrastar los resultados con el modelamiento de la serie y los p-values obtenidos para cada grado de autoregresión y estacionariedad.

modelo resultado:

ARIMA Model Results						
Dep. Variable:	D2.Venta_zona_2	No. Observations:	44			
Model:	ARIMA(3, 2, 1)	Log Likelihood	-167.138			
Method:	css-mle	S.D. of innovations	10.270			
Date:	Mon, 25 Jul 2022	AIC	348.276			
Time:	07:53:22	BIC	360.765			
Sample:	2	HQIC	352.907			
	coef	std err	z	P> z	[0.025	0.975]
const	0.3422	0.617	0.555	0.582	-0.866	1.551
Venta_zona_2_lag2_avg3	-0.0085	0.016	-0.541	0.592	-0.039	0.022
ar.L1.D2.Venta_zona_2	-0.0313	0.151	-0.207	0.837	-0.328	0.265
ar.L2.D2.Venta_zona_2	-0.1906	0.149	-1.276	0.210	-0.483	0.102
ar.L3.D2.Venta_zona_2	-0.0977	0.155	-0.631	0.532	-0.401	0.206
ma.L1.D2.Venta_zona_2	-1.0000	0.062	-16.005	0.000	-1.122	-0.878
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	0.5192	-1.7761j	1.8504	-0.2047		
AR.2	0.5192	+1.7761j	1.8504	0.2047		
AR.3	-2.9886	-0.0000j	2.9886	-0.5000		
MA.1	1.0000	+0.0000j	1.0000	0.0000		



JMZ2 - REGRESIÓN MACRO

El modelo de regresión macro contiene las siguientes variables y coeficientes correspondientes:

OLS Regression Results						
=====						
Dep. Variable:	Venta_zona_2	R-squared:	0.311			
Model:	OLS	Adj. R-squared:	0.259			
Method:	Least Squares	F-statistic:	6.013			
Date:	Mon, 25 Jul 2022	Prob (F-statistic):	0.00176			
Time:	07:53:45	Log-Likelihood:	-182.43			
No. Observations:	44	AIC:	372.9			
Df Residuals:	40	BIC:	380.0			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

intercept	44.0168	2.418	18.205	0.000	39.130	48.903
var_porc_demanda_interna_avg3_std	9.1532	2.610	3.506	0.001	3.877	14.429
flg_feia_avg3_std	5.0141	2.443	2.053	0.047	0.077	9.951
exp_PBI_avg3_std	7.1417	2.598	2.748	0.009	1.890	12.393
=====						
Omnibus:	1.494	Durbin-Watson:	0.597			
Prob(Omnibus):	0.474	Jarque-Bera (JB):	1.446			
Skew:	0.351	Prob(JB):	0.485			
Kurtosis:	2.456	Cond. No.	1.55			
=====						

Cabe mencionar que algunos p-values superan los valores de 0.05 (confianza de 95%) bajo decisión

experta pues se consideró adecuado para el modelo considerar algunos componentes sacrificando significancia estadística para mejorar la robustez que puede presentar el modelo frente a cambios sistémicos del país.

JMZ2 - COMBINACIÓN

La combinación de modelos en esta zona se realizó bajo la metodología de vectores autorregresivos la cual combina ambos resultados como parte de un modelo ARIMA con las mismas consideraciones del modelo anterior.

ARIMA Model Results						
Dep. Variable:	D.Venta_zona_2	No. Observations:	43			
Model:	ARIMA(0, 1, 3)	Log Likelihood	-156.913			
Method:	css-mle	S.D. of innovations	8.982			
Date:	Mon, 25 Jul 2022	AIC	327.826			
Time:	07:53:53	BIC	340.155			
Sample:	1	HQIC	332.373			
	coef	std err	z	P> z	[0.025	0.975]
const	-3.8318	3.066	-1.250	0.219	-9.841	2.177
venta_predicted	-0.0667	0.077	-0.866	0.392	-0.218	0.084
macro_pred	0.1814	0.136	1.336	0.190	-0.085	0.447
ma.L1.D.Venta_zona_2	-0.2815	0.170	-1.653	0.107	-0.615	0.052
ma.L2.D.Venta_zona_2	-0.4959	0.206	-2.405	0.021	-0.900	-0.092
ma.L3.D.Venta_zona_2	-0.2226	0.156	-1.430	0.161	-0.527	0.082
Roots						
	Real	Imaginary	Modulus	Frequency		
MA.1	1.0000	-0.0000j	1.0000	-0.0000		
MA.2	-1.6141	-1.3740j	2.1197	-0.3878		
MA.3	-1.6141	+1.3740j	2.1197	0.3878		

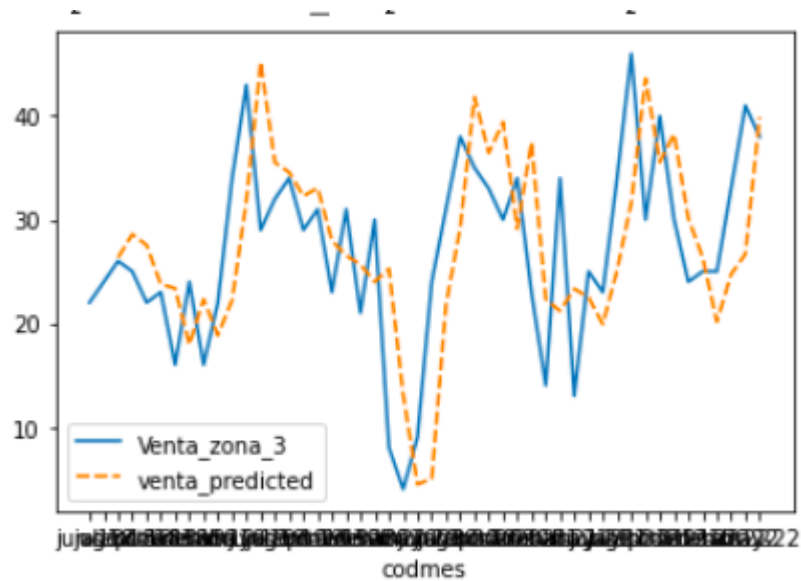
Jesús María Zona 3

JMZ3 - ARIMA

El modelo Arima y sus componentes p, d y q se basaron en aspectos de pruebas de hipótesis basados en pruebas de dicky-fuller y de diferenciación desarrollados en los notebooks de trabajo además de contrastar los resultados con el modelamiento de la serie y los p-values obtenidos para cada grado de autoregresión y estacionariedad.

modelo resultado:

ARIMA Model Results						
Dep. Variable:	D2.Venta_zona_3	No. Observations:	44			
Model:	ARIMA(3, 2, 3)	Log Likelihood	-157.895			
Method:	css-mle	S.D. of innovations	7.875			
Date:	Mon, 25 Jul 2022	AIC	333.790			
Time:	08:06:28	BIC	349.848			
Sample:	2	HQIC	339.745			
	coef	std err	z	P> z	[0.025	0.975]
const	1.6228	1.505	1.078	0.288	-1.328	4.573
Venta_zona_3_lag2_avg3	-0.0621	0.058	-1.079	0.288	-0.175	0.051
ar.L1.D2.Venta_zona_3	-2.2183	0.338	-6.569	0.000	-2.880	-1.556
ar.L2.D2.Venta_zona_3	-1.5675	0.525	-2.985	0.005	-2.597	-0.538
ar.L3.D2.Venta_zona_3	-0.3347	0.224	-1.493	0.144	-0.774	0.105
ma.L1.D2.Venta_zona_3	1.0000	0.390	2.567	0.015	0.236	1.764
ma.L2.D2.Venta_zona_3	-1.0000	0.124	-8.062	0.000	-1.243	-0.757
ma.L3.D2.Venta_zona_3	-1.0000	0.389	-2.574	0.014	-1.762	-0.238
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	-1.0762	-0.1496j	1.0866	-0.4780		
AR.2	-1.0762	+0.1496j	1.0866	0.4780		
AR.3	-2.5304	-0.0000j	2.5304	-0.5000		
MA.1	1.0000	-0.0000j	1.0000	-0.0000		
MA.2	-1.0000	-0.0000j	1.0000	-0.5000		
MA.3	-1.0000	+0.0000j	1.0000	0.5000		



JMZ3 - REGRESIÓN MACRO

El modelo de regresión macro contiene las siguientes variables y coeficientes correspondientes:

OLS Regression Results						
=====						
Dep. Variable:	Venta_zona_3	R-squared:	0.374			
Model:	OLS	Adj. R-squared:	0.330			
Method:	Least Squares	F-statistic:	8.381			
Date:	Mon, 25 Jul 2022	Prob (F-statistic):	0.000176			
Time:	08:06:48	Log-Likelihood:	-153.75			
No. Observations:	46	AIC:	315.5			
Df Residuals:	42	BIC:	322.8			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

intercept	26.5652	1.056	25.153	0.000	24.434	28.697
ind_prec_inm_avg3_std	3.7640	1.083	3.475	0.001	1.578	5.950
ind_prec_inf_suby_bienes_rat3to6_std	2.7606	1.187	2.326	0.025	0.365	5.156
prod_ener_lima_avg6_std	-2.5381	1.179	-2.153	0.037	-4.917	-0.159
=====						
Omnibus:	1.024	Durbin-Watson:	1.525			
Prob(Omnibus):	0.599	Jarque-Bera (JB):	0.330			
Skew:	0.039	Prob(JB):	0.848			
Kurtosis:	3.407	Cond. No.	1.61			

Cabe mencionar que algunos p-values superan los valores de 0.05 (confianza de 95%) bajo decisión

experta pues se consideró adecuado para el modelo considerar algunos componentes sacrificando significancia estadística para mejorar la robustez que puede presentar el modelo frente a cambios sistémicos del país.

JMZ3 - COMBINACIÓN

La combinación de modelos en esta zona se realizó bajo la metodología de vectores autorregresivos la cual combina ambos resultados como parte de un modelo ARIMA con las mismas consideraciones del modelo anterior.

ARIMA Model Results						
Dep. Variable:	D2.Venta_zona_3	No. Observations:	42			
Model:	ARIMA(3, 2, 3)	Log Likelihood	-145.050			
Method:	css-mle	S.D. of innovations	6.630			
Date:	Mon, 25 Jul 2022	AIC	310.100			
Time:	08:07:34	BIC	327.477			
Sample:	2	HQIC	316.470			
	coef	std err	z	P> z	[0.025	0.975]
const	2.7887	1.226	2.275	0.030	0.386	5.191
venta_predicted	-0.1239	0.044	-2.838	0.008	-0.209	-0.038
macro_pred	0.0197	0.024	0.828	0.413	-0.027	0.066
ar.L1.D2.Venta_zona_3	-0.7102	0.198	-3.592	0.001	-1.098	-0.323
ar.L2.D2.Venta_zona_3	0.4195	0.310	1.352	0.186	-0.189	1.027
ar.L3.D2.Venta_zona_3	0.2918	0.207	1.406	0.169	-0.115	0.698
ma.L1.D2.Venta_zona_3	-0.9990	0.183	-5.468	0.000	-1.357	-0.641
ma.L2.D2.Venta_zona_3	-1.0000	0.143	-7.015	0.000	-1.279	-0.721
ma.L3.D2.Venta_zona_3	0.9990	0.182	5.503	0.000	0.643	1.355
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	1.5524	+0.0000j	1.5524	0.0000		
AR.2	-1.3296	+0.0000j	1.3296	0.5000		
AR.3	-1.6603	+0.0000j	1.6603	0.5000		
MA.1	-1.0000	+0.0000j	1.0000	0.5000		
MA.2	1.0000	+0.0000j	1.0000	0.0000		
MA.3	1.0010	+0.0000j	1.0010	0.0000		

7. Conclusiones y usos de modelos

En función a los resultados obtenidos se plantearon dos alternativas de uso de los modelos.

Uso mensual: Se basa en el output de los modelos y estima cuánto será la venta de inmuebles mensual en un horizonte no mayor a 6 meses (por recomendación del modelador). En el escenario de los modelos en semáforo amarillo y rojo se necesitará un monitoreo trimestral del modelo para poder decidir si mantenerlo en despliegue. Finalmente, en un escenario óptimo se esperaría que un modelo en verde pueda predecir dentro de los márgenes esperados por 1 año a partir de la fecha de construcción del modelo.

Uso de trimestres acumulados: Se basa en el output de los modelos acumulados de tres meses y estima cuánto será la venta de inmuebles trimestral en un horizonte no mayor a 6 meses (por recomendación del modelador). En el escenario de los modelos en semáforo amarillo y rojo se necesitará un monitoreo trimestral del modelo para poder decidir si mantenerlo en despliegue. Finalmente, en un escenario óptimo se esperaría que un modelo en verde pueda predecir dentro de los márgenes esperados por 1 año a partir de la fecha de construcción del modelo.

Leyenda de semáforos de modelos



Se recomienda el uso del modelo, basándonos en los buenos indicadores de aciertos con las bandas de predicción y el R2 del modelo.



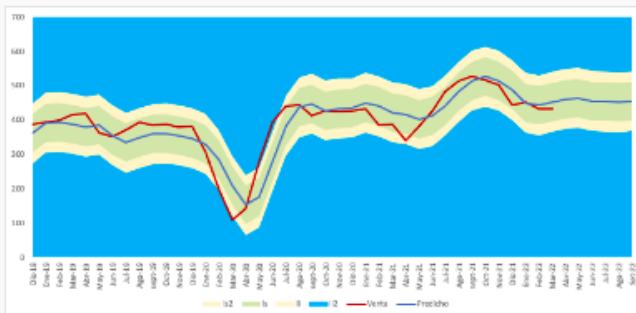
Se recomienda el uso bajo monitoreo, debido a que pese a tener indicadores aceptables el modelo podría perder capacidad predictiva debido a la volatilidad del segmento.



No se recomienda su uso, debido a que el modelo presenta indicadores pobres e inestables debido a la alta volatilidad del segmento.



Jesús María Agrupación trimestral



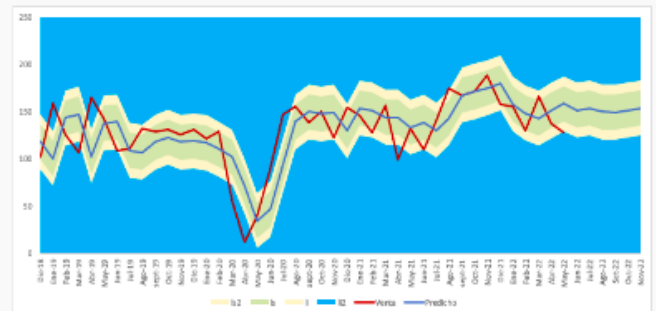
Ajuste del modelo
R2: 76.3%

92.5%
De predicciones correctas

	I2	I1	Pred	S1	S2	Real
<u>Abr22</u>	365	397	452	508	540	431



Jesús María Predicción Mensual



Ajuste del modelo
R2: 43.3%

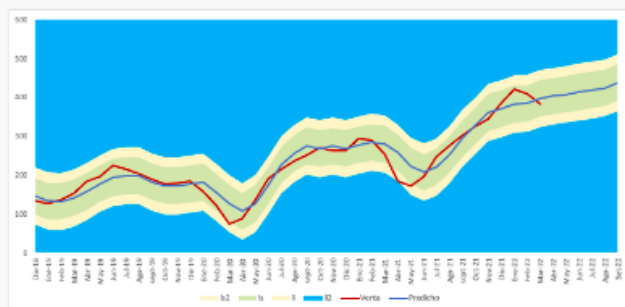
75.0%
De predicciones correctas

	I2	I1	Pred	S1	S2	Real
<u>Abr22</u>	114	124	143	161	172	166



Miraflores

Agrupación trimestral



Ajuste del modelo

R2: 92.1%

97.5%

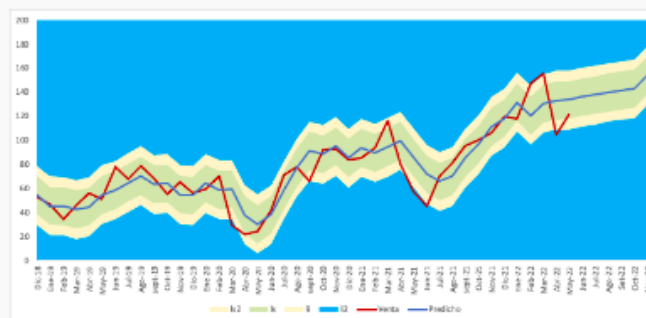
De predicciones correctas

	I2	I1	Pred	S1	S2	Real
Abr22	323	350	397	444	471	382



Miraflores

Predicción Mensual



Ajuste del modelo


















R2: 79.9%

85.0%

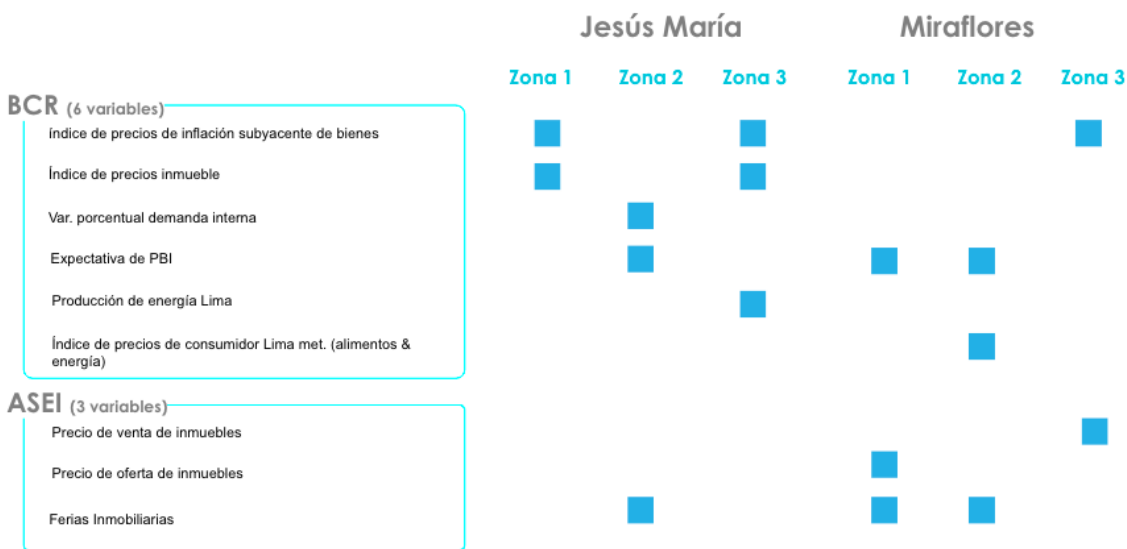
De predicciones correctas

	I2	I1	Pred	S1	S2	Real
Abr22	109	118	134	149	158	121

Resumen de modelos

		Zona 1	Zona 2	Zona 3	Total
 Jesús María	Agrupación mensual	R2: 30.4% Ac: 38.6% 	R2: 70.5% Ac: 78.6% 	R2: 27.8% Ac: 54.8% 	R2: 43.3% Ac: 75.0% 
	Agrupación trimestral	R2: 51.8% Ac: 83.3% 	R2: 87.9% Ac: 95.0% 	R2: 56.2% Ac: 87.5% 	R2: 76.3% Ac: 92.5% 
	Agrupación mensual	R2: 27.5% Ac: 59.1% 	R2: 71.6% Ac: 79.5% 	R2: 53.1% Ac: 75.0% 	R2: 79.9% Ac: 85.0% 
	Agrupación trimestral	R2: 57.1% Ac: 92.7% 	R2: 89.6% Ac: 97.6% 	R2: 84.7% Ac: 100% 	R2: 92.1% Ac: 97.5% 

Resumen de variables para implementación



Finalmente como recomendaciones finales se deja expuesta la recomendación de utilizar acumulaciones trimestrales por sobre las estimaciones mensuales además de agrupaciones por distrito por sobre las desagregaciones por Zona. Esto con la finalidad de mantener los niveles de precisión de estimación lo más robustos y confiables posible.

8. Referencias bibliográficas

Time Series Analysis: Forecasting and Control;Wiley Series in Probability and Statistics - Box, George E. P., Jenkins, Gwilym M., Reinsel, Gregory C.

Time Series Analysis and its Applications - Robert H. Shumway; David S. Stoffer

Practical Time Series Analysis - Aileen Nielsen

Analysis of Financial Time Series - Ruey S. Tsay

Applied Time Series Analysis - Torence Mills

Forecasting: principles and practice - Rob J Hyndman and George Athanasopoulos