



VAMOS POR UN CAFÉ?...

Mónica Ramírez Bernal



MÓNICA RAMÍREZ BERNAL

- *Ingeniera de Sistemas. Bogotá, Colombia*
- *Data Scientist en Everis Colombia*
- *Certificada Data Scientist - Metis Data Science Academy*
- *Me gusta leer, ir a cine, viajar, ir a Meetups y el café!*



VAMOS POR UN CAFÉ?...

Mónica Ramírez Bernal



AGENDA

- *Introducción*
- *Contexto*
- *Buscar los Datos*
- *Buscar el Mejor Modelo*
- *Buscar las Mejores Variables*
- *Conclusiones*



CONTEXTO



- *Segundo Productor de Café en el Mundo*
- *3% del área total = 10% Café Mundial!*
- *El café solo crece con un clima y en una época en particular (May-Oct)*
- *Tasa de intercambio entre Peso Colombiano y Dólar: TRM*

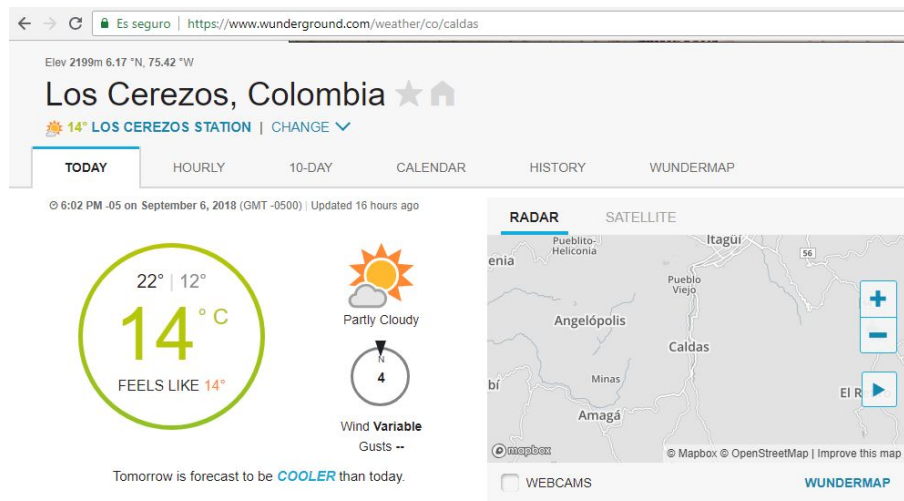
BUSCAR LOS DATOS

- *Producción de Café (Histórica)*
- *Producción de Café en países competidores*
- *Datos Económicos*
- *Clima Quindío, Colombia*



BUSCAR LOS DATOS

- *Producción de Café (Histórica)*
- *Producción de Café en países competidores*
- *Datos Económicos*
- *Clima Quindío, Colombia*



DATA SCRAPING

- *Beautiful Soup*
- *Selenium*
- *Scrapy*

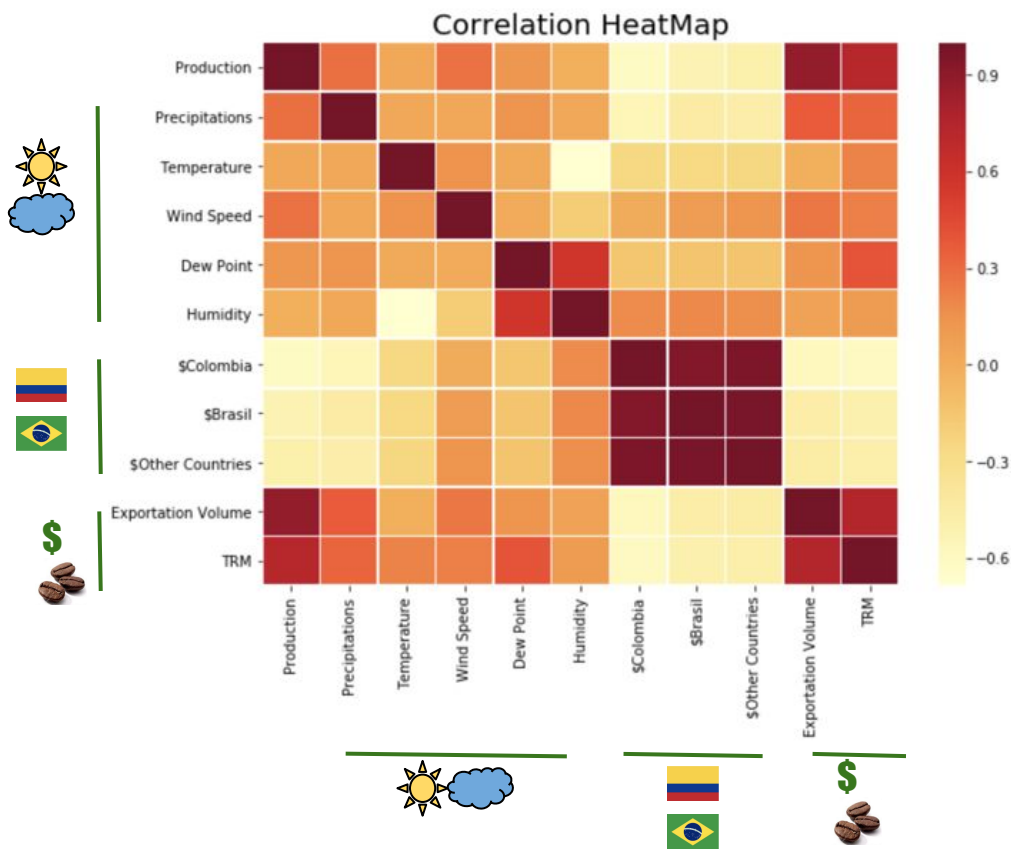
CLIC AQUÍ 4 Dias Crucero Cuba | 3 Dias Miami Beach 5 Dias Orlando por persona

Daily Observations

Time	Temperature (° F)			Dew Point (° F)			Humidity (%)			Wind Speed (mph)			Pressure	
Jan	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
1	73	69	64	66	66	66	100	-	100	7	-	0	30.16	-
2	74	64	64	0	0	0	0	-	0	4	-	0	30.15	-
3	77	64	64	0	0	0	0	-	0	4	-	0	30.12	-
4	83	74	64	70	66	64	100	-	49	7	-	0	30.15	-
5	81	72	63	70	66	63	100	-	57	5	-	0	30.2	-
6	81	73	65	72	67	65	100	-	53	6	-	0	30.14	-
7	81	72	62	68	66	63	100	-	51	6	-	0	30.11	-
8	82	72	62	69	66	63	100	-	48	6	-	0	30.12	-
9	72	67	62	68	65	63	100	-	78	7	-	0	30.14	-
10	81	70	60	66	63	61	100	-	45	7	-	0	30.12	-
11	80	70	60	68	65	60	100	-	52	9	-	0	30.12	-
12	78	71	64	70	66	64	100	-	63	5	-	0	30.14	-

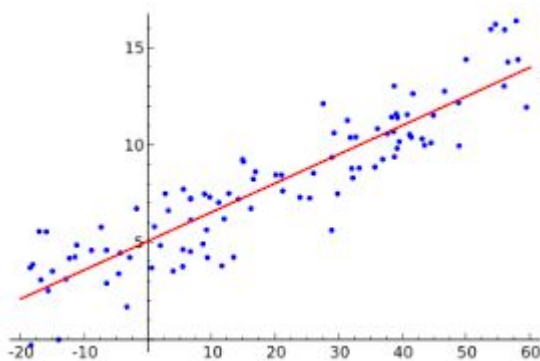
```
# Average Temperature: position 2 from each row of _ngcontent-c8 in _ngcontent-c8 table
avg_tmp = response.xpath(
    '//table[@id="_ngcontent-c8"]/tbody/tr/td[@id="_ngcontent-c8"]/text()'
).extract()[c+2]
```


ENTONCES, QUÉ HACE CRECER EL CAFÉ?



BUSCAR EL MEJOR MODELO

- Separar los datos entre set de Entrenamiento (70%) y Pruebas (30%)
- Ejecutar un primer Modelo (OLS)



	coef	std err	t	P> t	[0.025	0.975]
avg_precip	862.3188	559.620	1.541	0.128	-256.345	1980.983
avg_tmp	-35.7624	52.471	-0.682	0.498	-140.651	69.126
avg_wind	112.2885	54.855	2.047	0.045	2.634	221.943
dew_point	52.1720	63.235	0.825	0.413	-74.232	178.576
avg_humidity	-12.1280	17.816	-0.681	0.499	-47.741	23.485
Intercept	859.7245	3086.128	0.279	0.781	-5309.357	7028.805

↓

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$



PRIMER ANALISIS

VARIABLES:

Precipitaciones

Temperatura

Velocidad del viento

Humedad

Rocío

RESULTADOS:

R2 = 0.27

Adj R2 = 0.20

Skew = 0.72

P-Values [0.039 - 0.745]

R2-test = 0.148

VARIABLES CATEGORICAS

Fecha Producción	Producción	Temperatura ...	Mes
2015-02-01	515	73 ...	Febrero
2001-11-01	629	76 ...	Septiembre
...
2010-11-01	603	74 ...	Noviembre

Fecha Producción	Producción	Mes[Enero]	Mes[Febrero] ...	Mes[Noviembre]
2015-02-01	515	0	1 ...	0
2001-11-01	629	0	0 ...	0
...
2010-12-01	603	0	0 ...	0



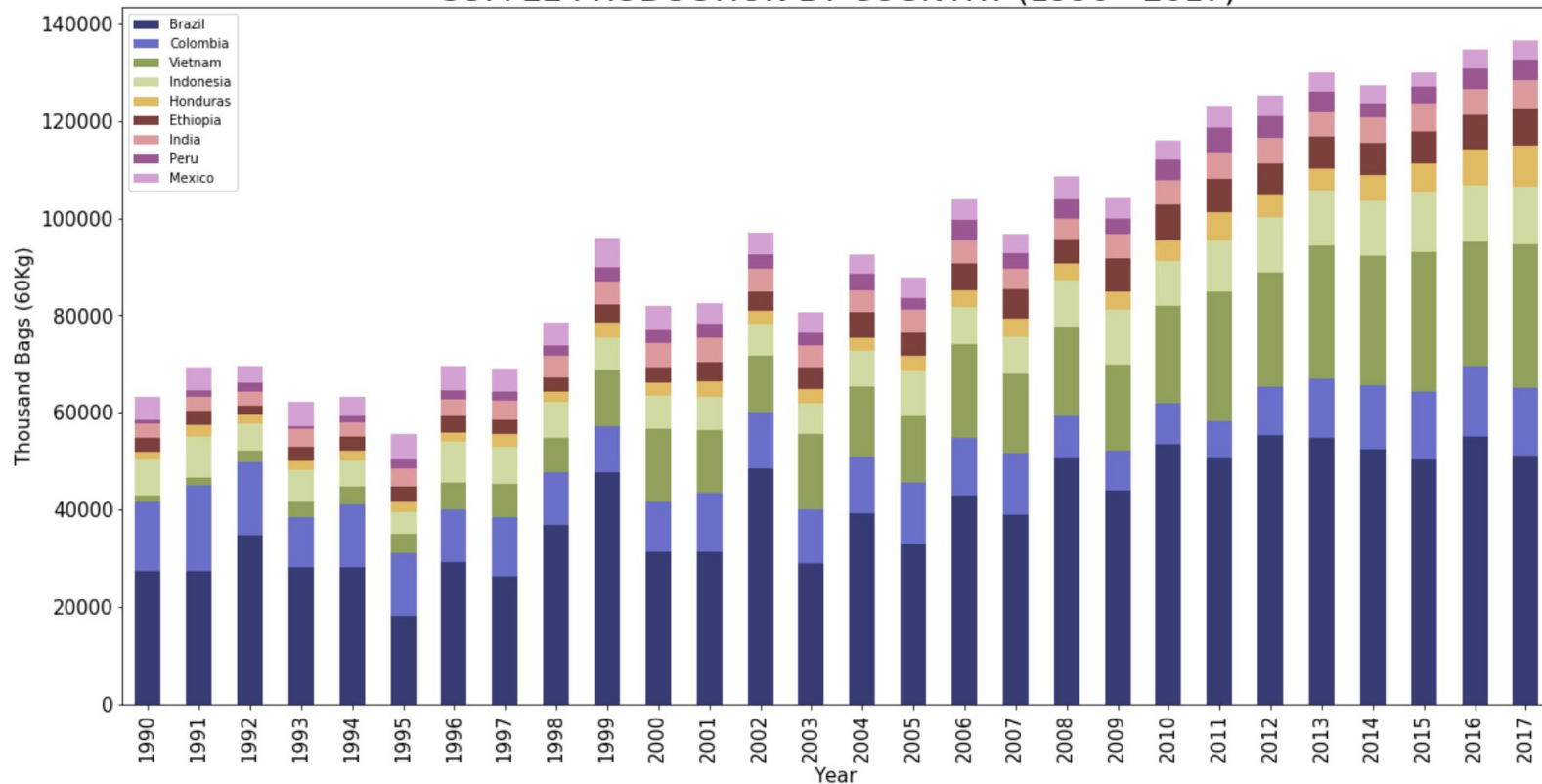
ANALISIS CLIMA + MES

	coef	std err	t	P> t	[0.025	0.975]
avg_precip	1338.7589	550.696	2.431	0.019	233.190	2444.327
avg_tmp	-38.1641	53.153	-0.718	0.476	-144.873	68.545
avg_wind	114.3527	55.903	2.046	0.046	2.123	226.583
dew_point	147.8855	65.828	2.247	0.029	15.729	280.042
avg_humidity	-21.2452	19.217	-1.106	0.274	-59.825	17.335
month[T.August]	313.2312	155.334	2.017	0.049	1.386	625.077
month[T.December]	641.7043	145.625	4.407	0.000	349.349	934.059
month[T.February]	232.2729	135.291	1.717	0.092	-39.336	503.882
month[T.January]	308.6742	170.989	1.805	0.077	-34.600	651.949
month[T.July]	450.3257	151.367	2.975	0.004	146.443	754.208

R2 = 0.50
Adj R2 = 0.33
Skew = 0.49
P-Values [0.002 - 0.759]
R2-test = 0.634

QUIERES OTRA TAZA?

COFFEE PRODUCTION BY COUNTRY (1990 - 2017)



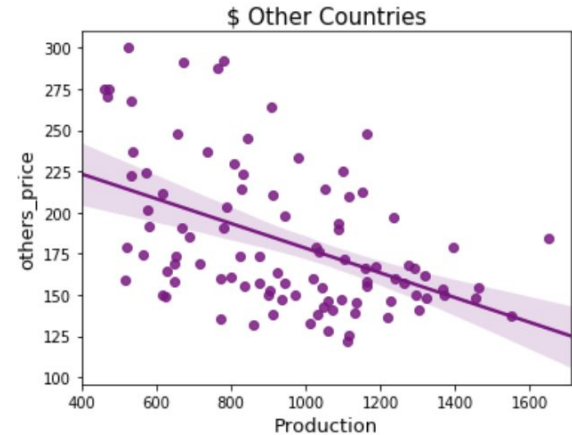
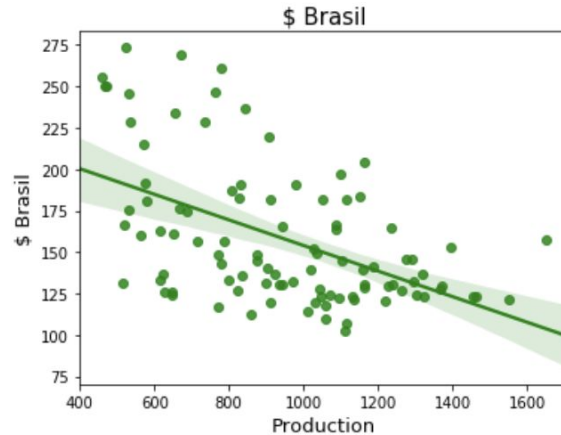
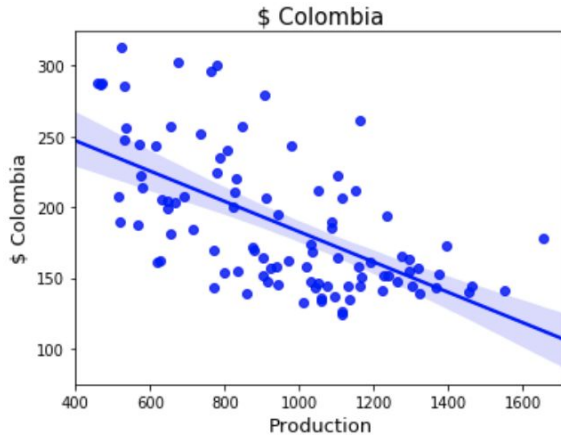
ANALISIS POR PAÍSES

Producción

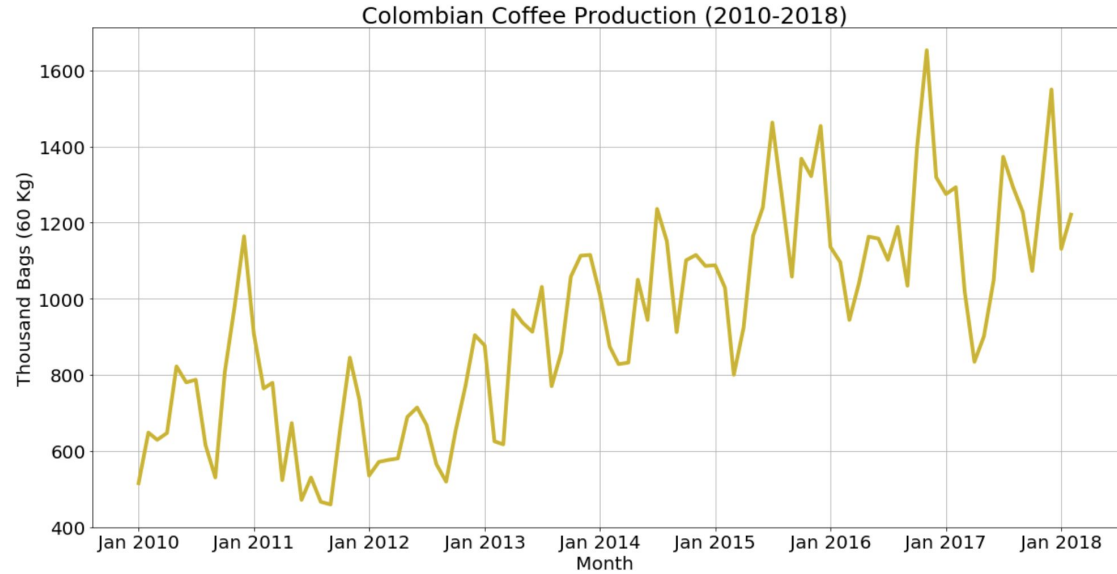
$R^2 = 0.94$
 $Adj\ R^2 = 0.93$
 $Skew = -0.79$
 $P\text{-Values} [0.011 - 0.553]$
 $R^2\text{-test} = 0.71$

Precio del café

$R^2 = 0.824$
 $Adj\ R^2 = 0.810$
 $Skew = -0.018$
 $P\text{-Values} [0.001 - 0.749]$
 $R^2\text{-test} = 0.703$



ES TIEMPO PARA UN CAFÉ



ES TIEMPO PARA UN CAFÉ



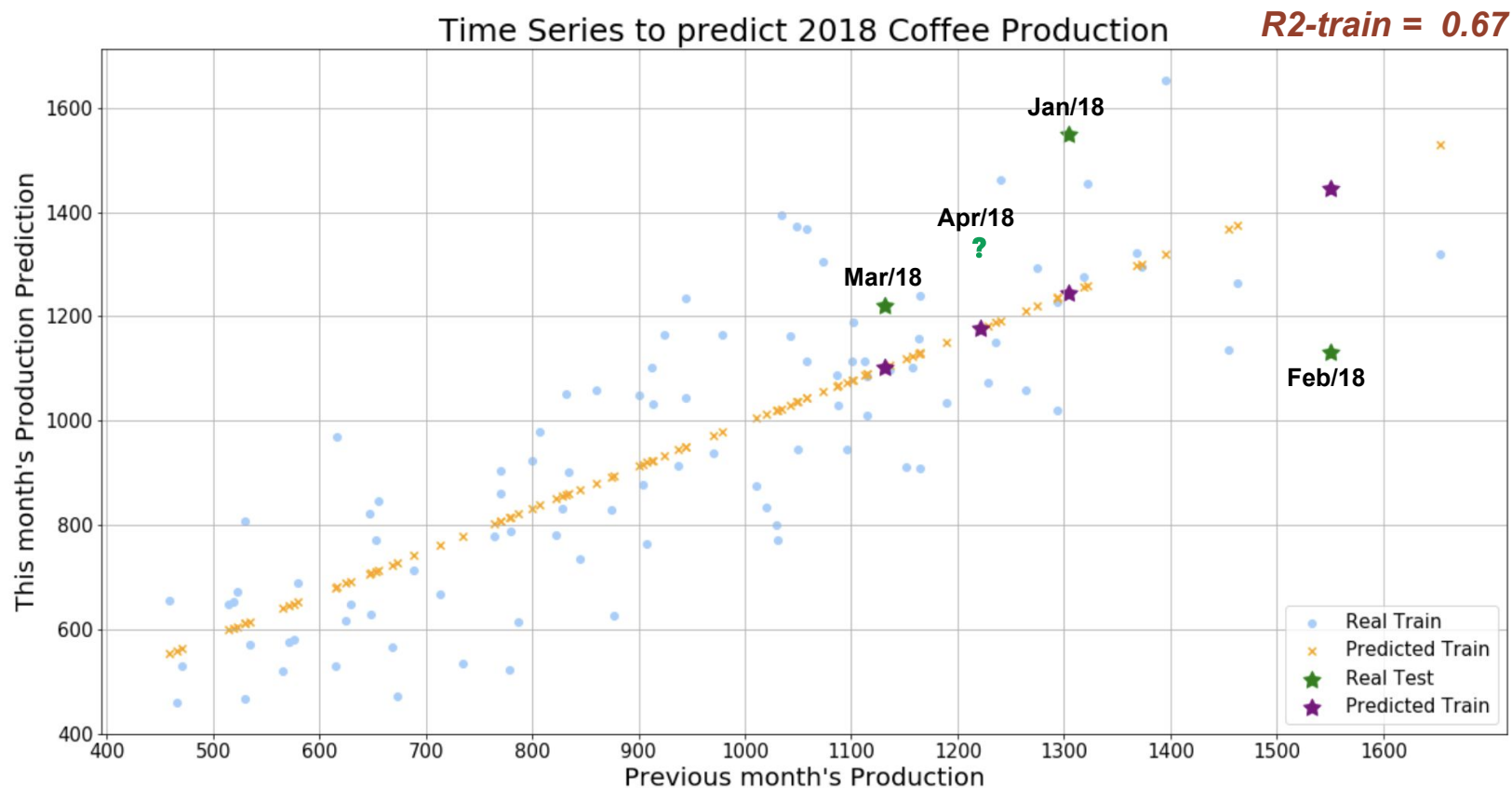
	month_dt	Produccion	prev_month
1	2010-02-01	648.0	515.0
2	2010-03-01	629.0	648.0
3	2010-04-01	647.0	629.0
4	2010-05-01	822.0	647.0
5	2010-06-01	780.0	822.0

Incluyendo solo un mes atrás:

R2-test = 0.679

It's Time for Coffee

SERIES DE TIEMPO

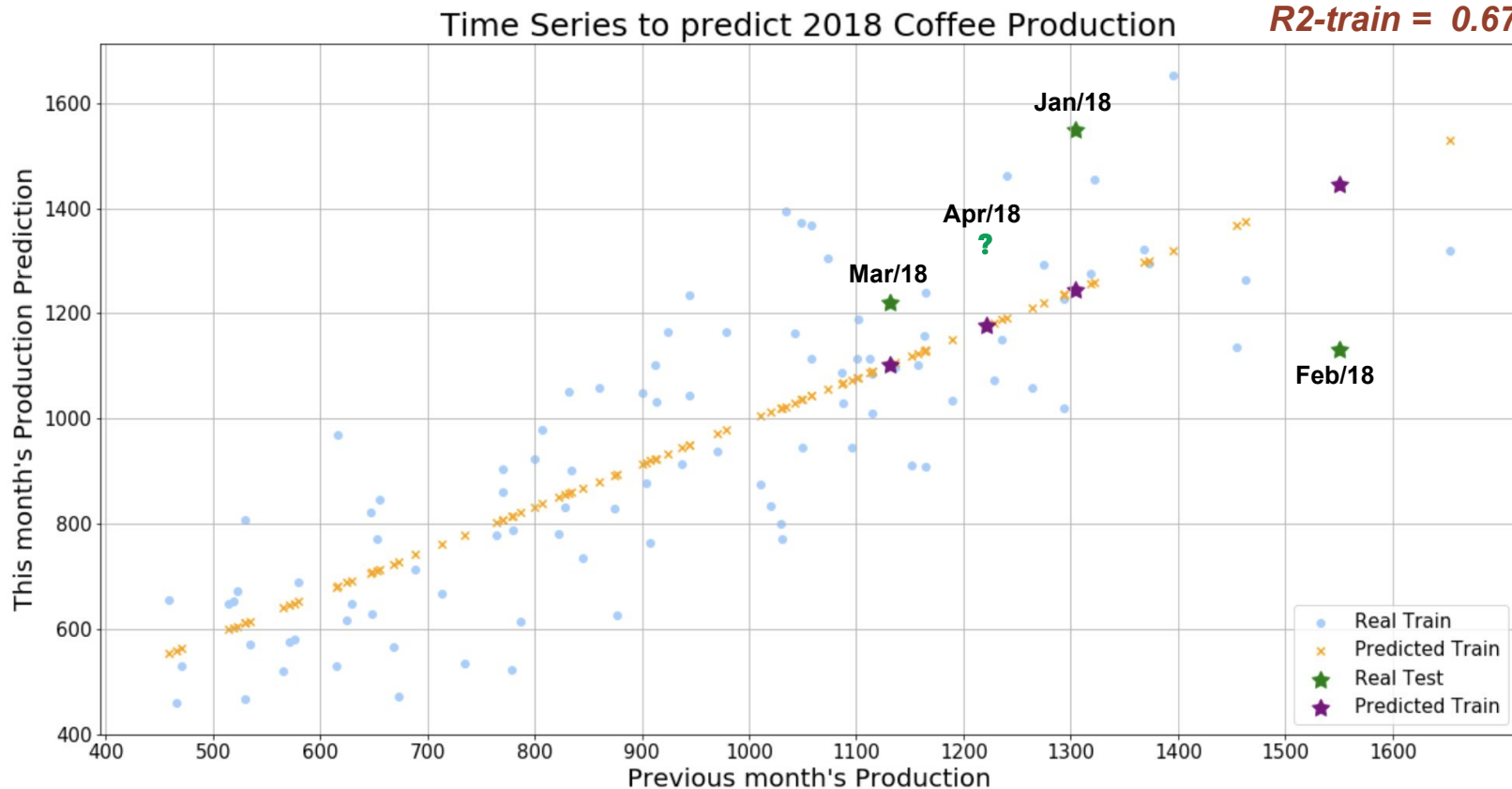


SERIES DE TIEMPO

Valor Estimado Abr-2018: 1.176

Valor Real Abr-2018: 1.037

R²-train = 0.67

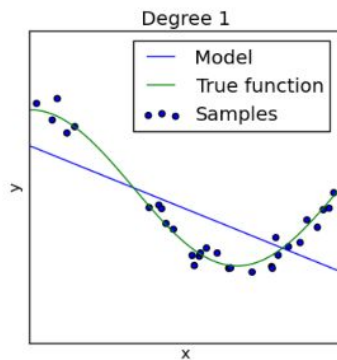


BUSCAR LAS MEJORES VARIABLES

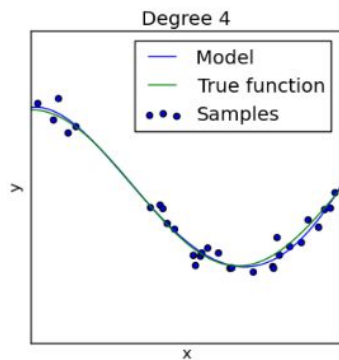
REGULARIZACIÓN:

Mejor Ajuste, Penalizar Complejidad

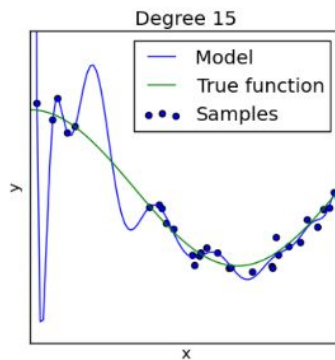
Underfitting



Just Right



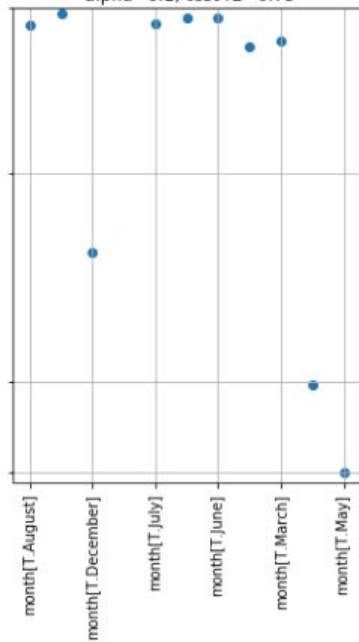
Overfitting



BUSCAR LAS MEJORES VARIABLES

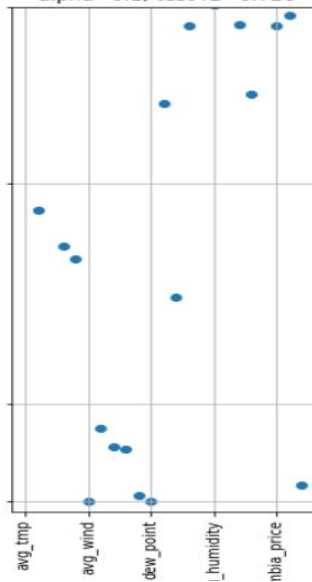
$$\lambda \sum_{j=1}^k \beta_j^2$$

Ridge with
alpha=0.1, test r2=0.75



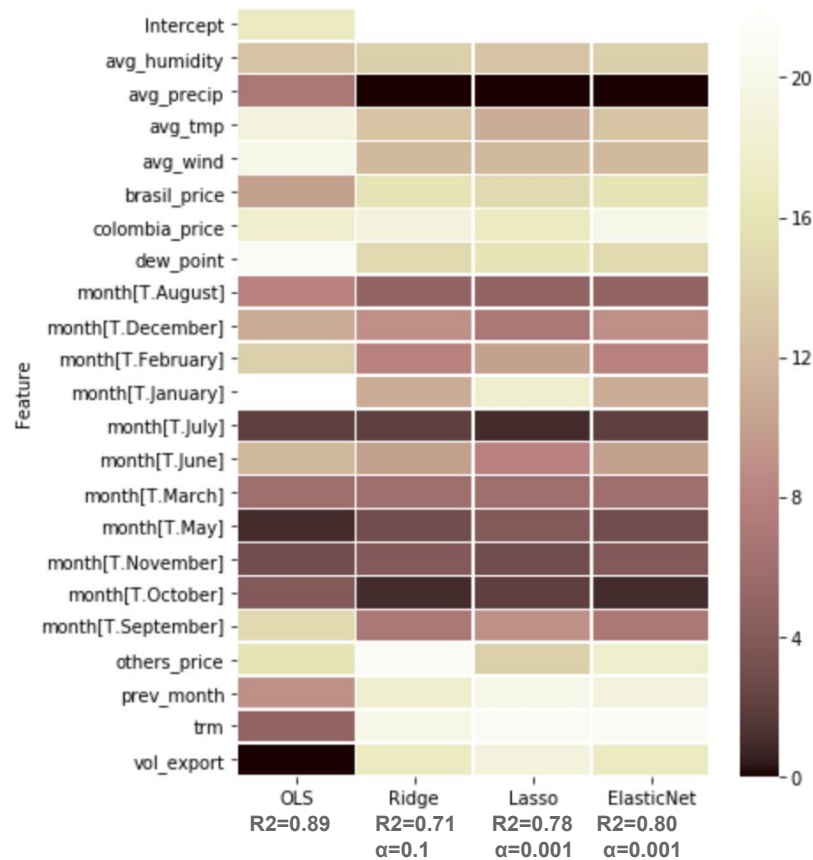
$$\lambda \sum_{j=1}^k |\beta_j|$$

Lasso with
alpha=0.1, test r2=0.726



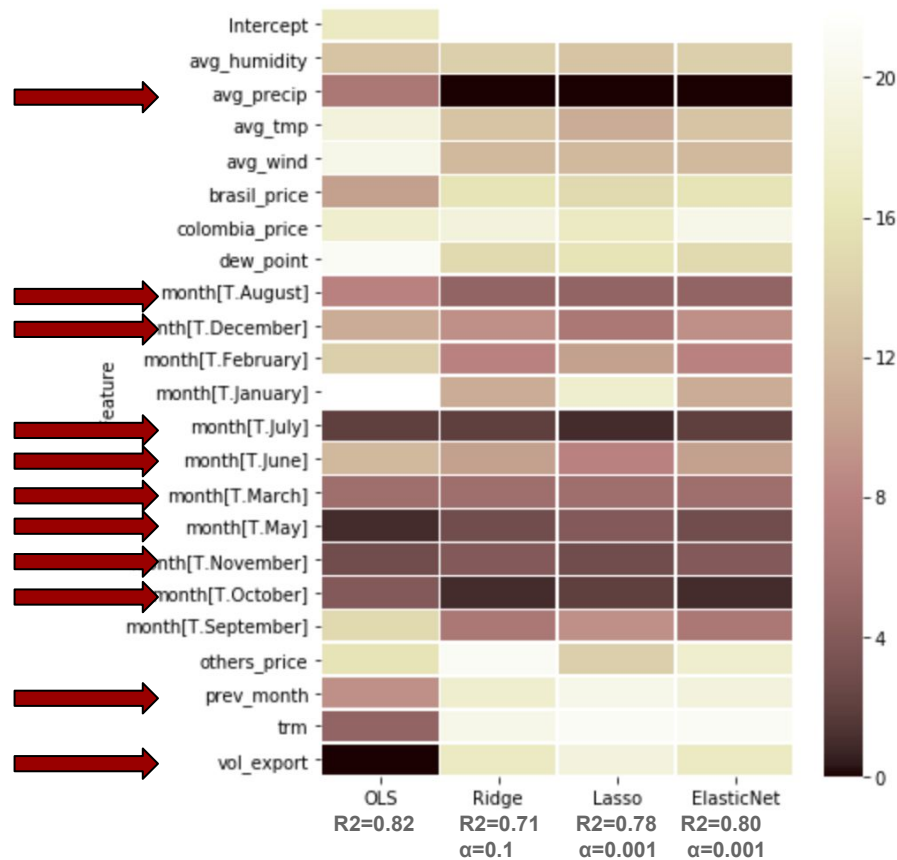


VAMOS TODOS POR UN CAFÉ





VAMOS TODOS POR UN CAFÉ





MODELO FINAL

R2 = 0.885

Adj R2 = 0.864

Skew = -0.15

R2-test = 0.783

	coef	std err	t	P> t
avg_precip	-378.4855	258.340	-1.465	0.148
month[T.August]	-1.4343	43.652	-0.033	0.974
month[T.December]	27.2085	45.625	0.596	0.553
month[T.July]	179.7751	61.335	2.931	0.005
month[T.June]	111.0806	43.326	2.564	0.013
month[T.March]	-112.6883	54.096	-2.083	0.042
month[T.May]	107.0420	42.930	2.493	0.016
month[T.November]	96.2914	49.520	1.944	0.057
month[T.October]	113.2358	45.682	2.479	0.016
vol_export	1.1348	0.061	18.673	0.000
Intercept	-52.4104	54.933	-0.954	0.344



CONCLUSIONES

- *La única variable climática que afecta la Producción del Café es la lluvia*
- *La Producción puede verse afectada por mayor o menor producción en el mes anterior*
- **Siguientes pasos:** *Añadir más variables*
- **Siguientes pasos:** *Análisis de Tiempo en otras variables*

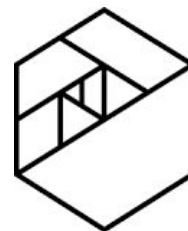
FUENTES

- *Weather Underground (www.wunderground.com)
Weather Data from Quindio, Caldas and Risaralda,
Colombia (Retrieved 18-04-2018)*
- *Colombian Coffee Growers Organization
(www.federaciondecafeteros.org) Coffee Information
and Photos (Retrieved 19-04-2018 - 26-06-2018)*
- *International Coffee Organization (www.ico.org)
Coffee Information (Retrieved 19-04-2018)*



DATA SCIENCE BOOTCAMP

- *12 semanas inmersivas*
- *Proceso de Selección*
- *Prewrite*
- *Metodología: Clases + Proyectos*
- *Vida Profesional*



METIS



Mónica Ramírez Bernal

Blog: monitarb.github.io

Projects GitHub Repo:

https://github.com/monitarb/first_projects/tree/master/P2_ColombianCoffeePrediction

2018

