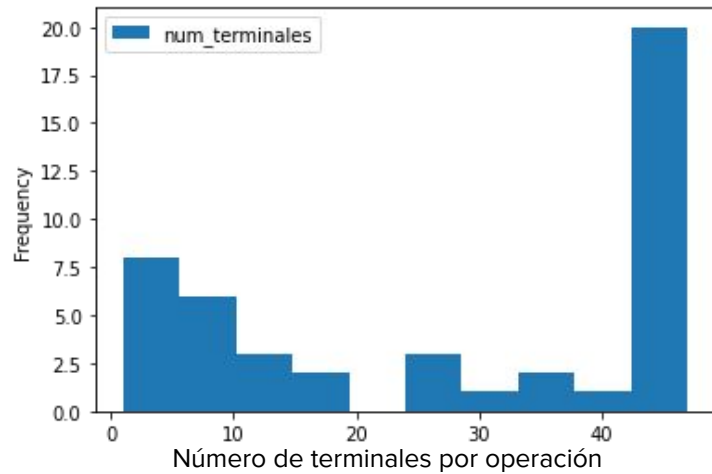
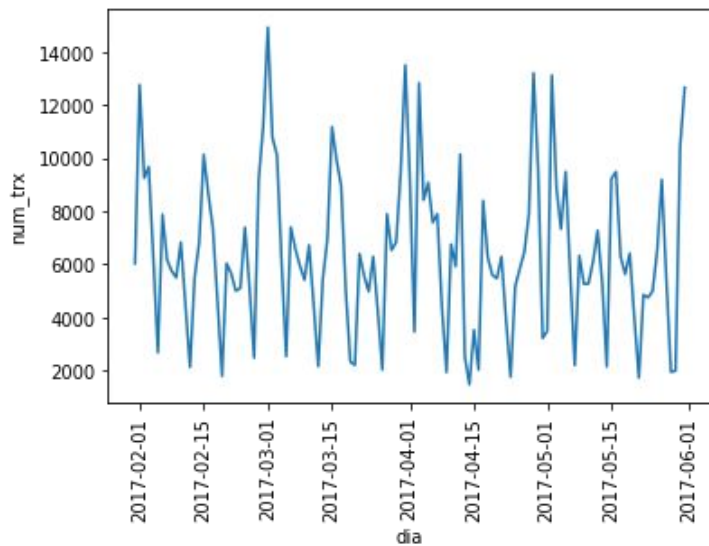


Pronóstico de transacciones

ADL

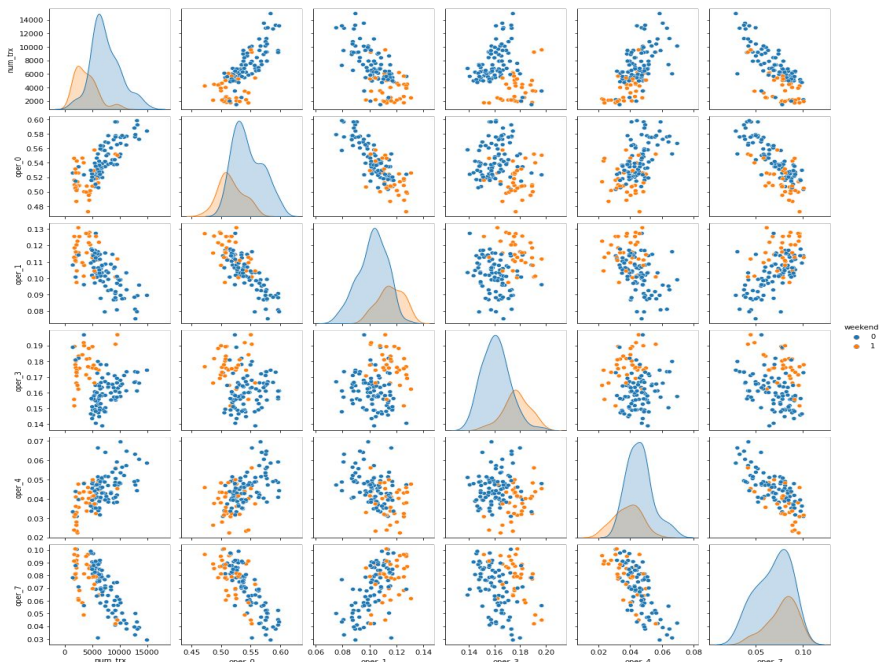
Trabajamos con 775.466 transacciones en un lapso 4 meses



- El número de trx no presenta tendencia de crecimiento o decrecimiento.
- El promedio de trx diarias, durante los fines de semana es 3813 y entre semana es 7422.

- En total hay 46 tipos de operaciones diferentes.
- En total hay 47 terminales diferentes.
- Hay 15 operaciones que utilizan los 47 terminales que realizan el 98.6% de las trx.

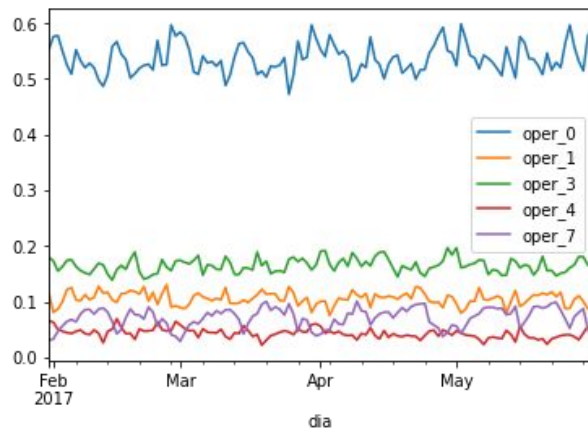
El Top 5 de operaciones que más trx realizan, tienen el 92% de la participación



- Cuando el share de la operación 0 aumenta el share de la operación 1 y 7 disminuye.
- Cuando el share de la operación 4 aumenta el share de la operación 7 disminuye.
- Cuando el share de la operación 0 aumenta el share de la operación 4 también aumenta.

Podemos apreciar la relación que existe entre la participación del top 5 de operaciones:

- Las operaciones 0 y 4 presentan un decrecimiento de participación los fines de semana.
- Las operaciones 1, 3 y 7 presentan un crecimiento de participación los fines de semana.



Se propuso proyectar el número de transacciones diarias usando 4 diferentes enfoques de modelación

- Algo importante a notar, es que solo contamos con 106 días de información en nuestra data de entrenamiento.
- La información de test, será los últimos 15 días.
- Puede ser muy poca información para un modelo de series de tiempo o Machine Learning, sin embargo, se intentará ajustar un modelo de series de tiempo (ARIMA) y también se formulará como un problema de regresión con el fin de probar una Regresión Lineal y un Random Forest.
- En cuanto al modelo baseline, consiste en que el pronóstico será el número de transacciones realizadas en la semana anterior.

$$\hat{y}_t = y_{t-7}$$

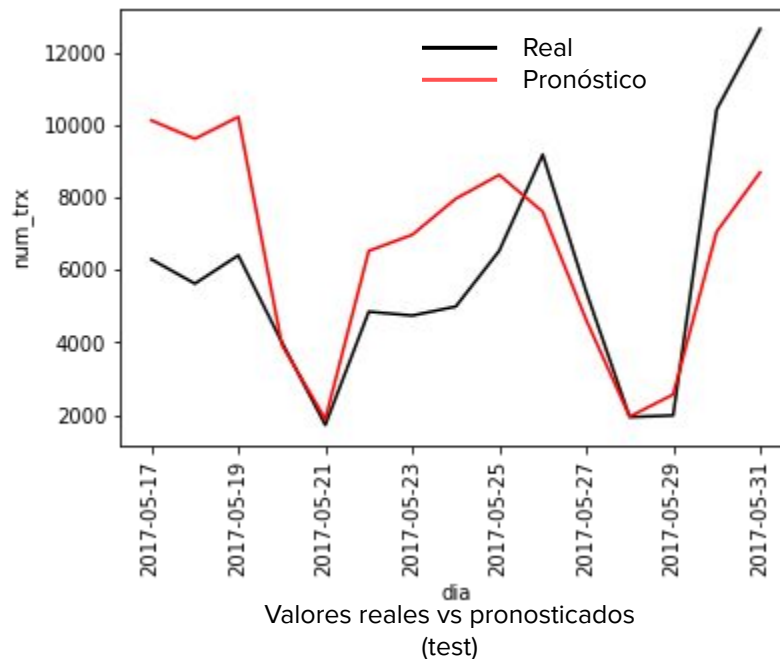
El mejor modelo fue el Random Forest, con un MAPE de 33% y un RMSE de 2541 transacciones, evaluado en la base de test.

Modelo	MAPE	RMSE
Baseline	42%	2584
ARIMA	56%	2793
Regresión Lineal	38%	2531
Random Forest	33%	2541

Las variables utilizadas en el enfoque de regresión fueron:

- Rezago de orden 1
- Rezago de orden 7
- Sí el día es fin de semana o no (incluyendo festivo)

Dado que el modelo **Random Forest** presenta el **menor** MAPE y su RMSE es muy similar a la regresión lineal, fue seleccionado como **mejor modelo** dentro del champion challenge.



Para la proyección por operación y terminal se utilizó el modelo ganador y la participación promedio diaria de cada operación en los últimos dos semanas

Operación	MAPE	RMSE
Operación 0	36%	1551
Operación 1	25%	190
Operación 3	36%	433
Operación 4	47%	149
Operación 7	18%	114

Se utilizó el mismo enfoque para realizar la proyección por terminal, en este caso considerando la participación de cada terminal dentro de cada operación.

Una de las ventajas más importantes de este modelo, es la escalabilidad y fácil mantenimiento en producción.

weekend	oper	idTerminal	share_term	dia	pred_oper	pred_term
0	oper_0	1774	0.036610	2017-05-17	5559.443109	203.529390
0	oper_0	1774	0.036610	2017-05-18	5281.494005	193.353765
0	oper_0	1774	0.036610	2017-05-19	5613.064789	205.492462
0	oper_0	1774	0.036610	2017-05-22	3582.894192	131.168582
0	oper_0	1774	0.036610	2017-05-23	3827.375026	140.118945
0	oper_0	1774	0.036610	2017-05-24	4372.911132	160.090843
0	oper_0	1774	0.036610	2017-05-25	4734.283935	173.320583
0	oper_0	1774	0.036610	2017-05-26	4175.888505	152.877910
0	oper_0	1774	0.036610	2017-05-30	3872.813500	141.782433
0	oper_0	1774	0.036610	2017-05-31	4771.517249	174.683683
0	oper_0	1908	0.033508	2017-05-17	5559.443109	186.284391
0	oper_0	1908	0.033508	2017-05-18	5281.494005	176.970944
0	oper_0	1908	0.033508	2017-05-19	5613.064789	188.081132
0	oper_0	1908	0.033508	2017-05-22	3582.894192	120.054698
0	oper_0	1908	0.033508	2017-05-23	3827.375026	128.246699
0	oper_0	1908	0.033508	2017-05-24	4372.911132	146.526382
0	oper_0	1908	0.033508	2017-05-25	4734.283935	158.635169
0	oper_0	1908	0.033508	2017-05-26	4175.888505	139.924599
0	oper_0	1908	0.033508	2017-05-30	3872.813500	129.769239
0	oper_0	1908	0.033508	2017-05-31	4771.517249	159.882774

Muestra de la tabla de salida

Próximos pasos...

- Con el modelo propuesto, obtenemos métricas aceptables, para mejorar las métricas podemos apalancarnos en la construcción de features adicionales, como tarifas de las transacciones.
- Desplegar la solución desarrollada.
- Evaluar métricas por terminal.
- Evaluar el ajuste de una red neuronal recurrente, la poca cantidad de datos podría no hacer funcionar este enfoque, sin embargo se podría hacer el ejercicio de evaluar.
- Se podría pensar en construir un modelo por operación.
- Un modelo por operación y terminal podría volverse muy complejo de mantener y desplegar en producción, adicionalmente sería muy variable las estimaciones debido al bajo volumen de transacciones que hay por terminal.

Gracias !!
