

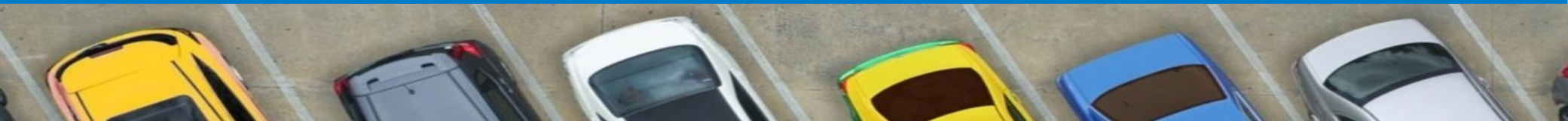


Transferencia de turismos Estudio y modelaje de series temporales

Raúl José de la Cuesta García

- Aprender como estructura la Dirección General de Tráfico la información que pone a disposición
- Preprocesar los datos para adaptarlos
- Analizar el dataset generado de la fuente de datos
- General una modelo para predecir las transferencias futuras
- Representación de los datos predichos en comparativa con los reales

Fuente de datos



Dirección General de Tráfico (DGT)

- El portal de datos estadísticos de la DGT para “satisfacer, de forma ágil, rápida y dinámica, la creciente demanda de información estadística relacionada con la actividad desarrollada”
- Se pueden encontrar datos de las siguientes materias:
- Vehículos
- Conductores
- Accidentes
- Incidencias
- Datos de Tráfico
- Distintivo Ambiental



- https://sedeapl.dgt.gob.es/WEB_IEST_CONSULTA/categoria.faces

Datos: Transferencias de vehículos

- Cambios de titularidad de vehículos.
- Datos originales almacenados mediante la transcripción de documentos administrativos.
- Incluye toda clase de vehículos, incluidos remolques
- Almacenados el lotes por meses
- Archivos de tipo “Ancho fijo” y con codificación ISO-8859-1

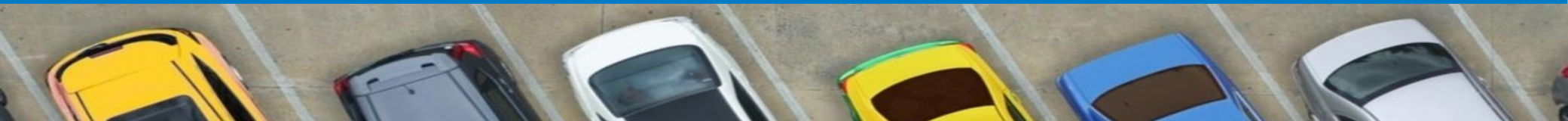
Ficheros de Ancho Fijo

1	30062006021032022CITROEN	66.20	C4 HDI	3VF7LC9HXC74501456	401	1560	11.47	1332	1777	5	0201	SASA20103202237008	ND
	80037274SALAMANCA		5 N										
	LC9HXC												
	01000												
	0												
2	23022022028032022MERCEDES-BENZ		A 250	3WD01770471J077970	400	1991	13.28	0	2025	5	0201	ALMU2010320220473811042019UD	
	B0004102VICAR	165.00	5 170N										
	AG			1505	2025M1	AC	0EURO	6AG	01000				
	000000												
	272915591564M		N0000										
3	29122008001032022SUZUKI		G5XR 600	01032022									
	B0028106PARLA	92.00	2 N	3J51GN7DA472105466	500	599	6.46	178	380	2	0201	M M 20103202228981	ND

	00400												
	0												
	0												
4	06042021001032022AUDI		Q3 SPORTBACK	3WAUZZZF30M1089892	401	1968	13.19	0	2145	5	0201	M M 20103202228721	NXAUD01NQP
	B0028121REDUENA	110.00	5 150N										
	AG			1655	2145M1	AB	0EURO	6AP	01000				
	000000												
	267715921597M		S0000E13 28 29										
5	22032016001032022SKODA		FABIA	01032022									
	B0020069DONOSTIA	77.00	5 95N	3TMBET6NJKGZ151127	401	1422	9.67	0	1620	5	0101	SSSS20103202220011	NDSKO0061A
	A.S.												
	000000			1165	1620M1	AB	0EURO	6W	01000				
	245514631457M												
6	08052009001032022VOLKSWAGEN		PASSAT	01032022									
	R0049021BENAVENTE	103.00	5	3WVWZZ3CZ9P068379	401	1968	13.19	1452	2090	5	0101	ZAZA20103202249600	ND

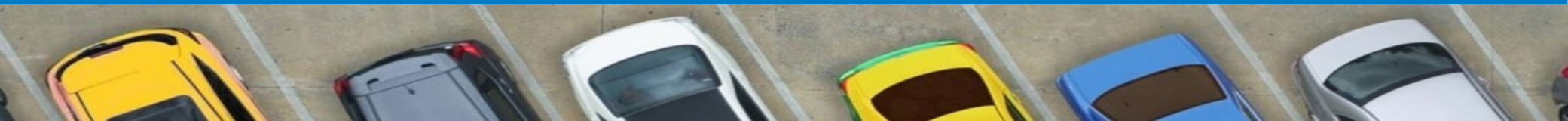
	Nombre Campo	Descripción	Tipo y Longitud	Formato
1	FEC_MATRICULA	Fecha de matriculación del vehículos	CHAR(8)	DATE. Formato: DDMMYYYY
2	COD_CLASE_MAT	Código de clase de matrícula	CHAR(1)	(Ver anexo I)
3	FEC_TRAMITACION	Fecha de tramitación, que se corresponde con la fecha de transferencia del vehículo contenida en los datos de transferencias.	CHAR(8)	DATE. Formato: DDMMYYYY
4	MARCA_ITV	Descripción de la marca del vehículo	CHAR(30)	
5	MODELO_ITV	Modelo del vehículo	CHAR(22)	

Preprocesado de Datos



- Automatizar la descarga de datos (Web Scraper)
- Convertir la codificación de ISO-8859-1 a UTF-8
- Empaquetar los datos distribuidos en meses por años
- Convertir de formato “Ancho fijo” a “CSV”
- Eliminar campos que no son de utilidad (Sin datos o información propia de la DGT sin interés)
- Filtrado de vehículos (solo turismos)

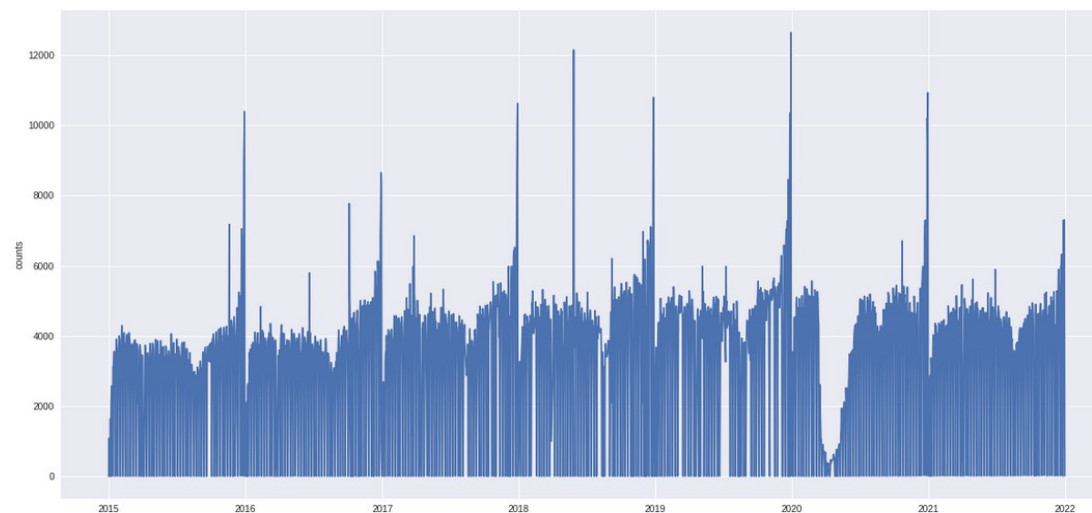
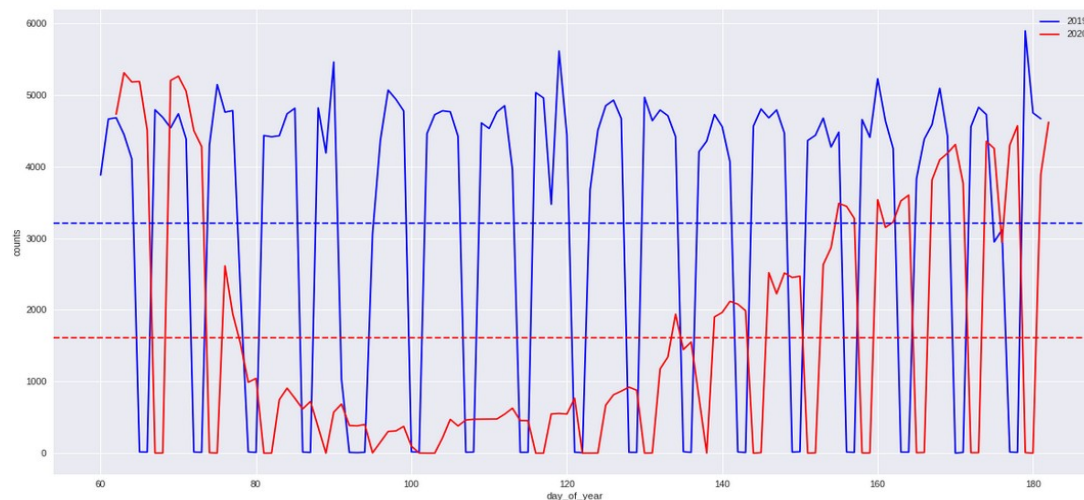
Análisis de Datos



- Los datos son mostrados y manipulados mediante las herramientas de pandas, concretamente usando DataFrame
- Mediante el filtrado y la agrupación podemos aprender de la información y buscar predictores eficientes para el desarrollo de modelos de Machine Learning
- Ver el comportamiento de los posibles predictores a lo largo del tiempo

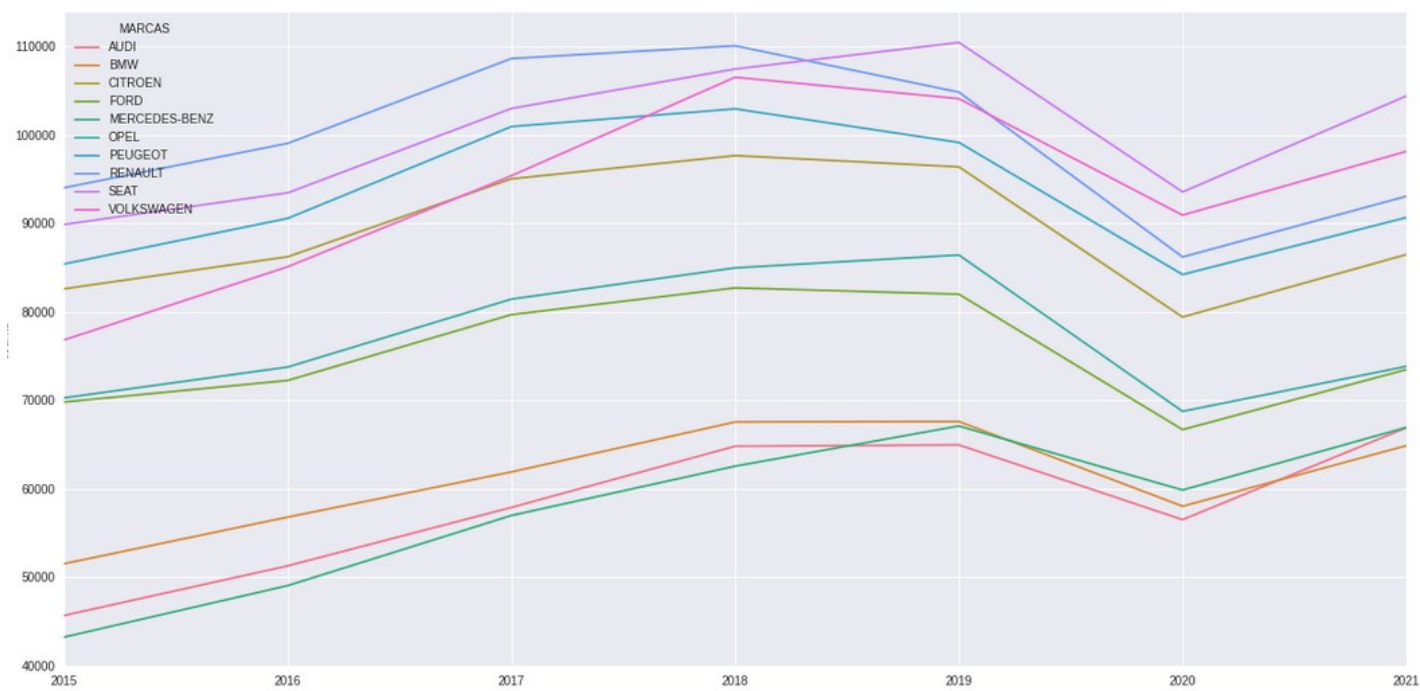
Relación de transferencias por fecha

	FEC_TRAMITE	counts
0	2015-01-01	2
1	2015-01-02	1085
2	2015-01-03	5
3	2015-01-04	45
4	2015-01-05	1631
...
2418	2021-12-27	7293
2419	2021-12-28	6426
2420	2021-12-29	7296
2421	2021-12-30	7306
2422	2021-12-31	15



MARCA_ITV counts		
51	SEAT	702241
47	RENAULT	695955
60	VOLKSWAGEN	657032
45	PEUGEOT	653938
8	CITROEN	623827
44	OPEL	539406
18	FORD	526543
4	BMW	428143
2	AUDI	407864
39	MERCEDES-BENZ	405553

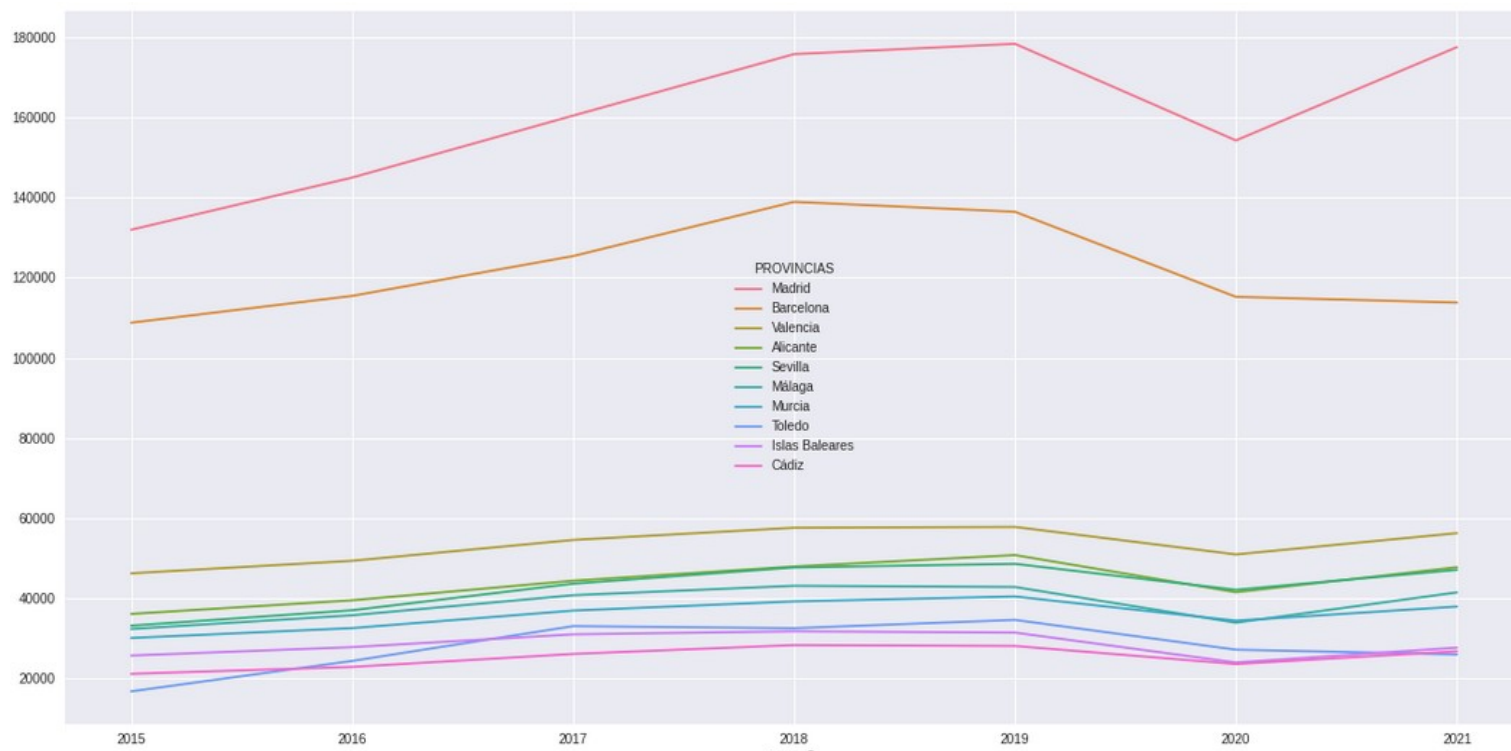
MARCA_ITV year_Q counts			
0	ALFA ROMEO	2015	6120
1	ALFA ROMEO	2016	6144
2	ALFA ROMEO	2017	6476
3	ALFA ROMEO	2018	6638
4	ALFA ROMEO	2019	6613
...
417	VOLVO	2017	12638
418	VOLVO	2018	12988
419	VOLVO	2019	13322
420	VOLVO	2020	11105
421	VOLVO	2021	12034



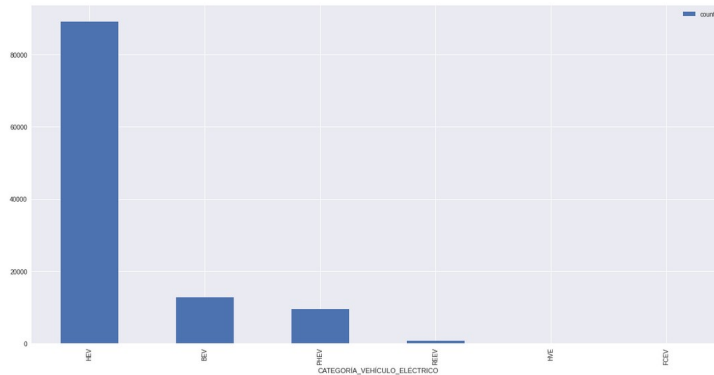
Provincias

ID_MUNICIPIO	counts
27	1123370
7	854043
45	372739
2	307919
40	299313
28	270179
29	251569
6	199348
44	194458
10	176894

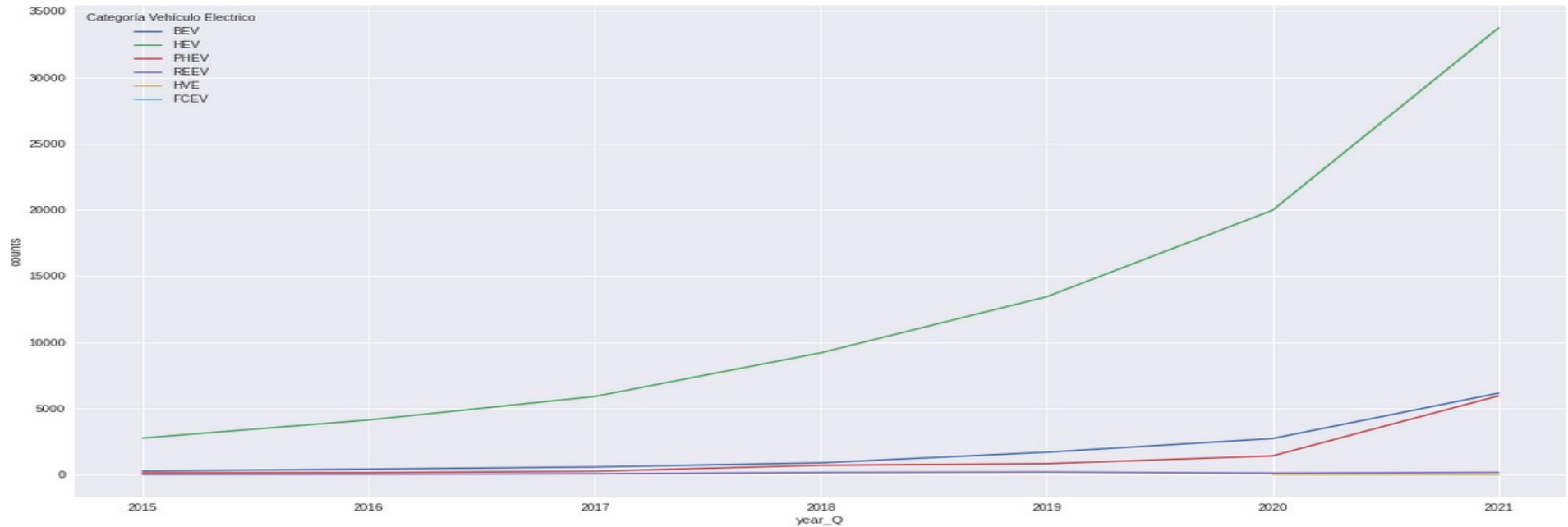
ID_MUNICIPIO	year_Q	counts
193	28	178374
195	28	177530
192	28	175798
191	28	160452
194	28	154258
...
353	51	1621
354	51	1558
356	51	1489
362	52	1441
355	51	1284



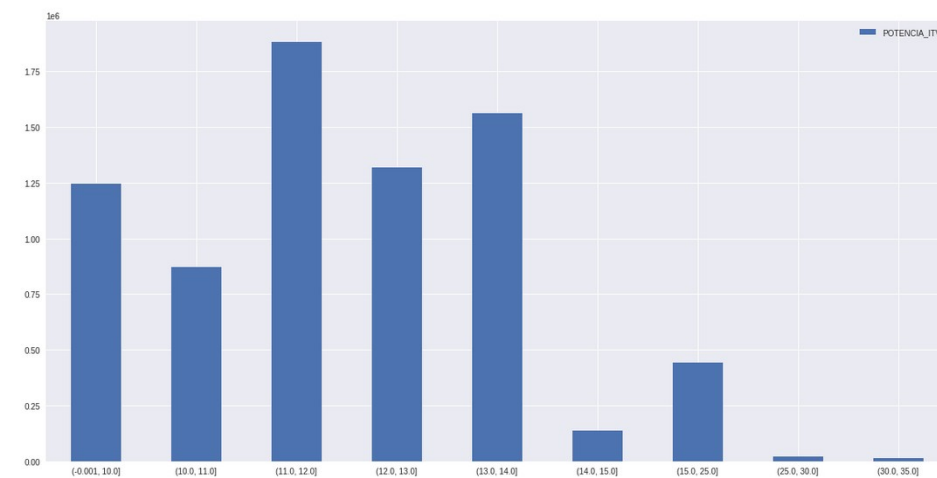
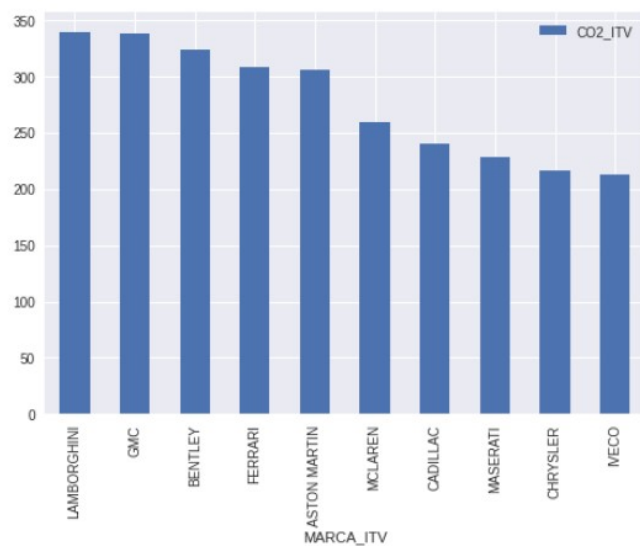
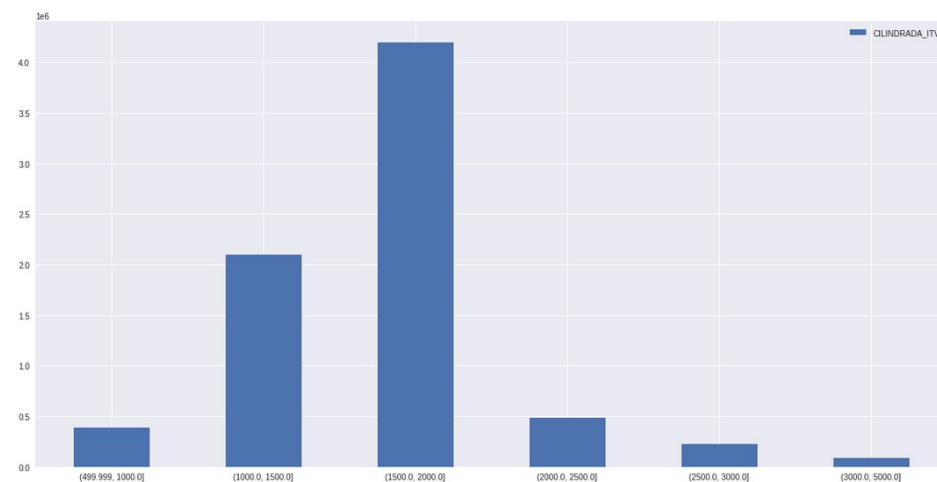
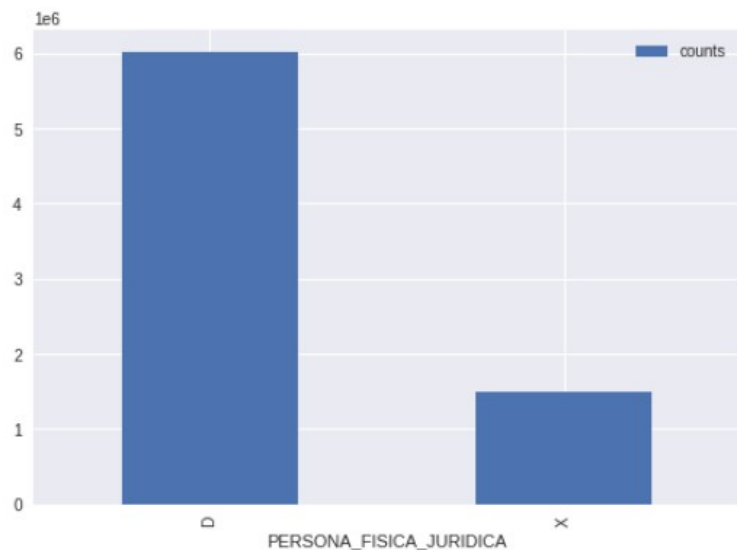
Uso de Eléctricos



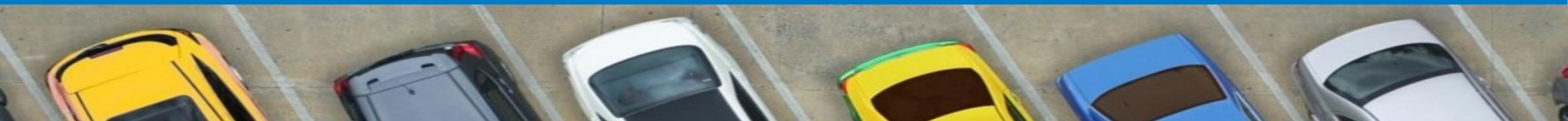
PHEV Eléctrico Enchufable
REEV Eléctrico de Autonomía Extendida
HEV Eléctrico Híbrido
BEV Eléctrico de Batería



Otros Predictores



Modelo



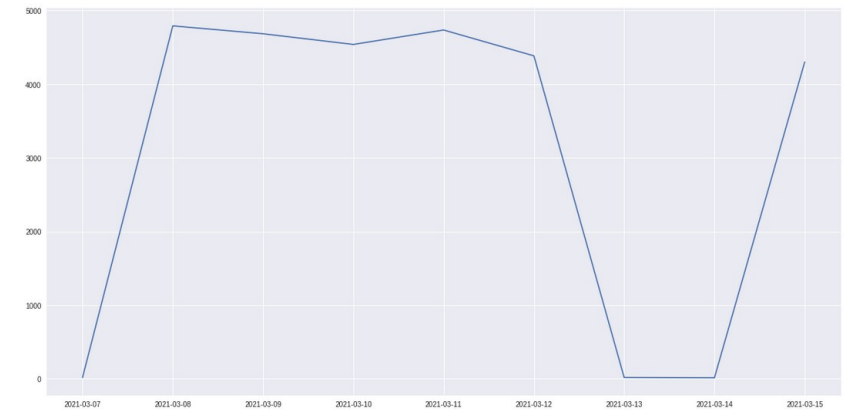
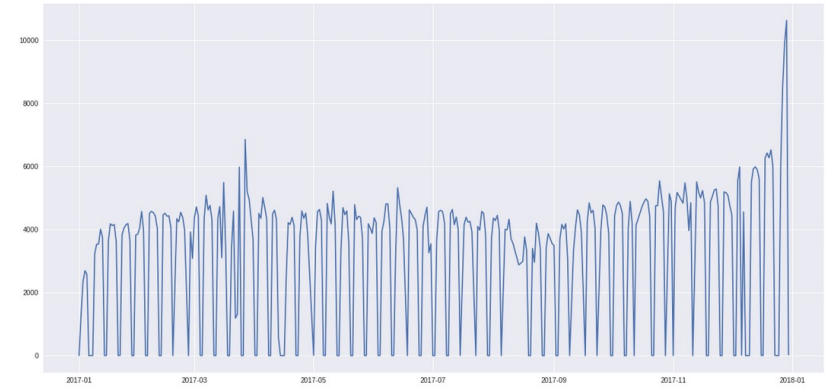
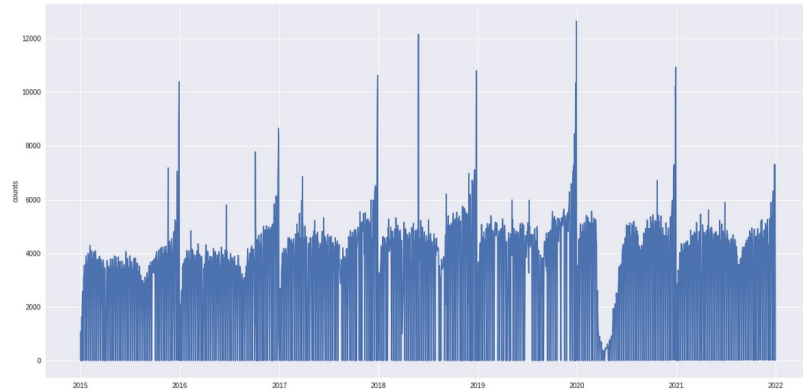
- Algoritmo basados en arboles de decisión
- La predicción es el resultado del promedio de predicciones de todos los árboles
- Aprendizaje supervisado de los datos para generar una serie temporal con las predicciones

¿Por qué he usado RFR y no ARIMA?

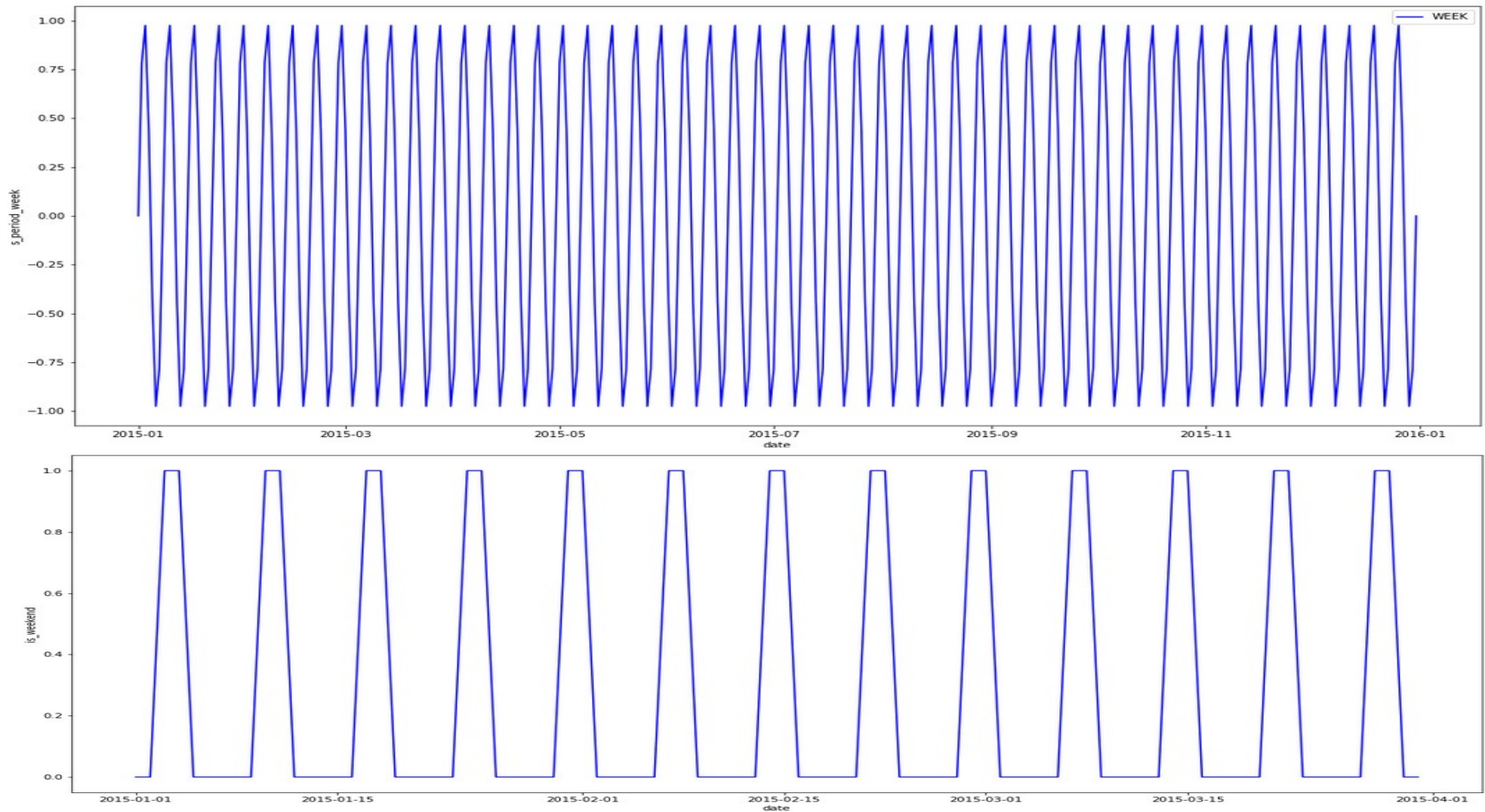
- ARIMA es un modelo estadístico, no es un modelo de Machine Learning
- La idea es mostrar lo aprendido en lo que respecta a los modelos supervisados, aplicando su metodología.
- RF puede usar predictores numéricos y categóricos

- Al querer calcular el número de transacciones, agrupando por fechas, perdemos muchos otros posibles predictores
- Es necesario crear unos nuevos predictores sintéticos para mejorar el aprendizaje de nuestro modelo

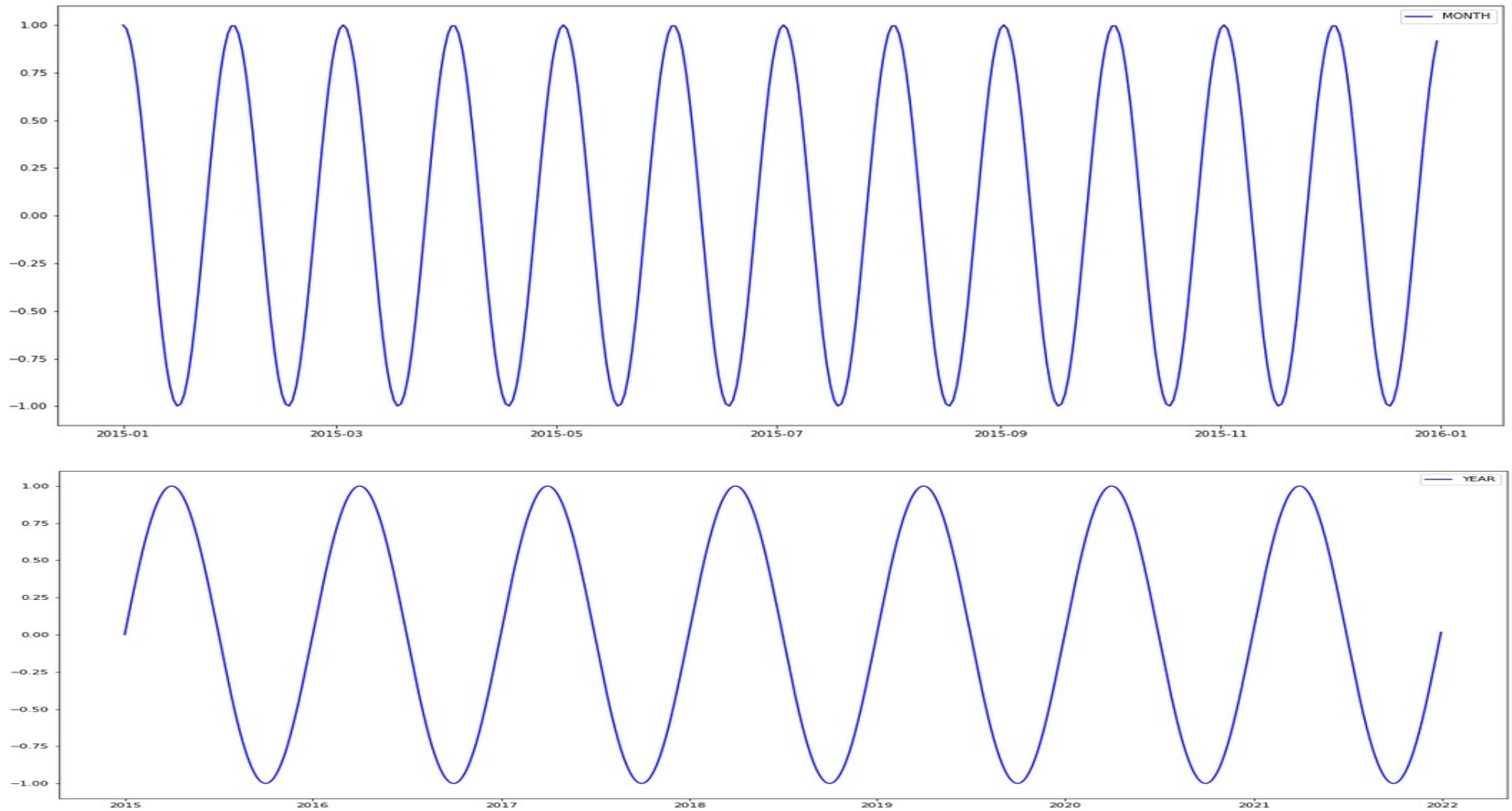
Distribución de las transacciones



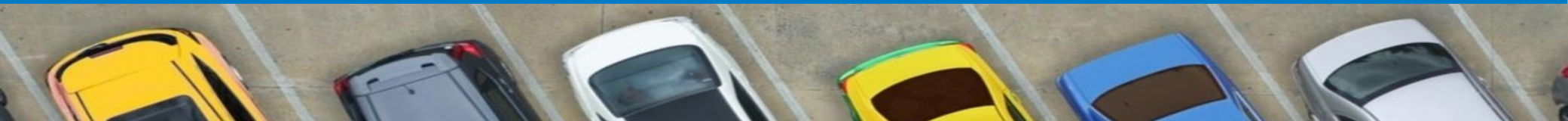
Señales Semanales



Señales mensuales y anuales

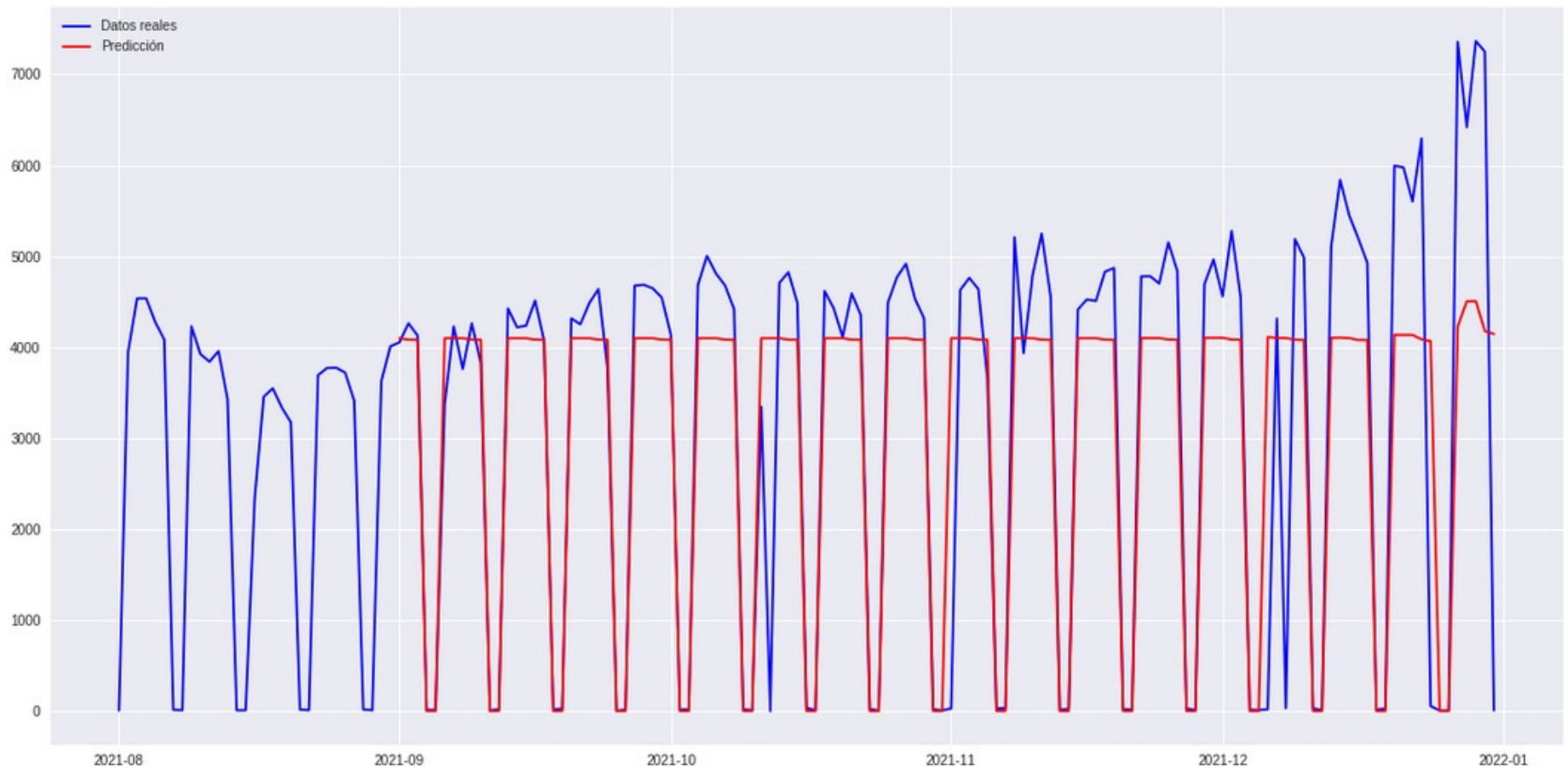


Serie Temporal

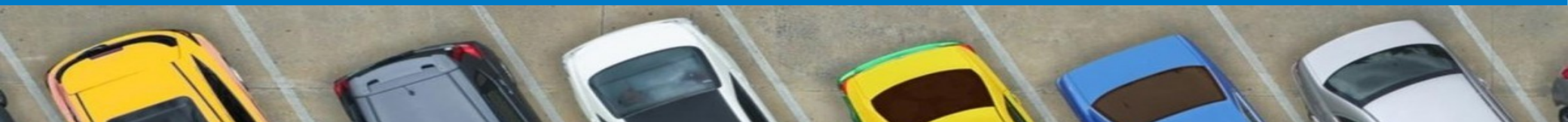


- Luego de realizar mediciones y seleccionar los metadatos más óptimos para nuestro modelo, entrenamos el mismo para obtener los resultados.
- Estos resultados podemos compararlos con los datos de test de manera visual (a parte de las mediciones) para ver que tal se porta nuestro modelo.

Realidad vs Predicción



Queda tanto que hacer...



Mis ideas para continuar

- Aprender mis conocimientos de estadística
- Combinación con otros datasets
- Añadir predictores para crear múltiples series desde el mismo modelo
- Aplicar otros modelos

Gracias

