

**Universidad Nacional de Córdoba - Facultad de Matemática, Astronomía, Física y Computación**

**Diplomatura en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones**

**Análisis y Visualización de Datos - 2019**

**Práctico I - Estadística**

Mentor:  
Lucas Grando

Equipo:  
Natalia Moreno  
Juan Fabro

# **Mentoría**

## **Predicción de Fallas en Camiones de Minería**

Inspeccionamos un dataset que contiene muestras de aceite extraídas a diferentes componentes para un tipo específico de camión de los utilizados en minería.

Las muestras se envían a un laboratorio que analiza los diferentes componentes químicos de las mismas (Hierro, Cromo, presencia de Agua, etc), luego de analizarlas emiten un informe indicando el estado de la muestra, lo que permite a los analistas encontrar posibles fallas en los equipos, previniendo futuras roturas. Los resultados posibles son tres:

- 1 - Bien
- 2 - Regular
- 3 - Mal

Dentro de las características más importantes que se encuentran en la muestra, se nos solicitó atención en las siguientes variables:

- Horas Funcionamiento: Indica la cantidad de horas de funcionamiento del camión
- Horas del Aceite: Representa la cantidad de horas de utilización del aceite (debido a que a mayor horas de uso del aceite, el mismo comienza a degradarse)
- St: Presencia de Hollín en el Aceite
- Al: Presencia de Aluminio en el Aceite
- Fe: Presencia de Hierro en el Aceite
- Si: Presencia de Silicio en el Aceite
- Na: Presencia de Sodio en el Aceite
- Visco: Viscosidad del aceite

El dataset cuenta con otros features que si bien, a priori podrían no ser tan relevantes, podrían ser utilizados para encontrar algún patrón no conocido

## Informe del estudio exploratorio del dataset

### 1.1

La muestra está compuesta por 18.469 resultados de laboratorio, para 87 camiones, los cuales tienen 11 componentes que utilizan aceite. Estos componentes son los siguientes:

- Convertidor
- Diferencial Trasero
- Mando Final TD
- Mando Final TI
- Masa Derecha
- Masa Izquierda
- Motor
- Radiador
- Sist de Dirección
- Sist. Hidráulico
- Transmisión

Observamos las principales variables de la muestra:

	Horas Funcionamiento	Horas del Aceite	St	Al	Fe	Si	Na	VISCO
count	18469.000000	17707.000000	2931.000000	17818.000000	17820.000000	17818.000000	17818.000000	17611.000000
mean	37053.585955	1000.688391	21.852610	0.777332	17.144147	4.955410	18.343136	17.290938
std	10765.690236	1270.912998	20.670388	1.229572	24.489269	24.170705	500.655499	8.439637
min	106.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	32502.000000	308.000000	0.000000	0.000000	5.000000	2.000000	3.000000	11.000000
50%	40116.000000	585.000000	24.000000	0.100000	10.000000	3.000000	4.000000	14.900000
75%	44704.000000	1226.000000	35.000000	1.000000	21.000000	5.000000	6.000000	24.200000
max	79990.000000	20208.000000	131.000000	37.100000	709.000000	1721.000000	35352.000000	341.700000

Esto nos permite decir que el promedio de funcionamiento de los camiones se encuentran en las 37.053 horas para esta muestra. Podemos identificar un vehículo con 79.990 horas (vehículo 1402) y otro con 106 horas (vehículo 2060). El promedio de utilización de aceite se encuentra en las 1.000 hs con una mediana de 585 horas.

También podemos decir que la pieza más controlada es el Motor, seguida por el Sistema Hidráulico, Transmisión y la que menos chequeos recibe es el Radiador:

Componente	Equipo count
Convertidor	1789.0
Diferencial Trasero	1823.0
Mando Final TD	1656.0
Mando Final TI	1662.0
Masa Derecha	1695.0
Masa Izquierda	1691.0
Motor	2174.0
Radiador	574.0
Sist de Dirección	1726.0
Sist. Hidráulico	1841.0
Transmisión	1838.0

## 1.2

Pasamos a graficar media, mediana y desviación estándar de los principales características en estudio del aceite agrupandolas por componentes con el objetivo de buscar algún patrón entre las variables.

Partimos considerando las variables de horas:

Componente	Horas Funcionamiento							
	count	mean	std	min	25%	50%	75%	max
Convertidor	1789.0	36998.072666	10851.475985	256.0	32456.00	40199.0	44675.00	51153.0
Diferencial Trasero	1823.0	36896.695557	10777.150591	256.0	32422.50	40031.0	44539.00	51153.0
Mando Final TD	1656.0	36901.275362	10831.748277	256.0	32402.00	40007.0	44548.25	79990.0
Mando Final TI	1662.0	36946.679904	10740.831293	256.0	32431.75	40037.5	44573.25	51153.0
Masa Derecha	1695.0	36976.085546	10750.986529	256.0	32422.50	40048.0	44585.00	51153.0
Masa Izquierda	1691.0	37001.237138	10735.386766	256.0	32422.50	40085.0	44616.50	51153.0
Motor	2174.0	36878.517479	10836.947169	106.0	32633.00	39896.0	44659.25	51153.0
Radiador	574.0	39824.977352	9687.960305	4284.0	34468.50	43387.0	46712.50	57715.0
Sist de Dirección	1726.0	36935.951333	10876.215622	256.0	32299.50	40151.0	44700.00	51153.0
Sist. Hidráulico	1841.0	37030.670831	10831.991076	256.0	32456.00	40310.0	44725.00	51153.0
Transmisión	1838.0	37091.756801	10650.210641	256.0	32747.50	40097.0	44541.75	51153.0

Componente	Horas del Aceite							
	count	mean	std	min	25%	50%	75%	max
Convertidor	1733.0	636.529717	345.056400	0.00	321.00	559.0	843.0	3370.0
Diferencial Trasero	1769.0	1277.054833	852.976627	2.00	569.00	1127.0	1788.0	5904.0
Mando Final TD	1604.0	1314.650249	883.848008	0.00	584.75	1181.0	1830.0	6424.0
Mando Final TI	1610.0	1304.600621	864.514858	41.00	579.25	1161.5	1828.5	5904.0
Masa Derecha	1642.0	445.361145	187.719891	84.00	279.00	428.0	549.0	1439.0
Masa Izquierda	1637.0	443.210751	187.118780	9.00	278.00	422.0	549.0	1413.0
Motor	2056.0	296.896401	127.096038	0.00	250.00	279.0	342.0	976.0
Radiador	417.0	5637.158273	4534.458406	0.00	1445.00	5151.0	8155.0	16576.0
Sist de Dirección	1673.0	1337.911536	938.268142	61.00	559.00	1096.0	1871.0	4471.0
Sist. Hidráulico	1783.0	1319.707426	1052.493407	0.34	532.00	1108.0	1865.0	20208.0
Transmisión	1783.0	638.569826	358.503274	0.00	321.00	561.0	845.0	4444.0

La presencia de hollín en el aceite:

St									
	count	mean	std	min	25%	50%	75%	max	
Componente									
Convertidor	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Diferencial Trasero	376.0	0.002660	0.051571	0.0	0.0	0.0	0.0	1.0	
Mando Final TD	345.0	0.002899	0.053838	0.0	0.0	0.0	0.0	1.0	
Mando Final TI	345.0	0.002899	0.053838	0.0	0.0	0.0	0.0	1.0	
Masa Derecha	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Masa Izquierda	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Motor	1865.0	34.341555	15.573263	0.0	25.0	32.0	41.0	131.0	
Radiador	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Sist de Dirección	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Sist. Hidráulico	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Transmisión	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

Aluminio

Al									
	count	mean	std	min	25%	50%	75%	max	
Componente									
Convertidor	1787.0	0.963234	1.039929	0.0	1.00	1.0	1.0	28.0	
Diferencial Trasero	1822.0	0.095445	0.410604	0.0	0.00	0.0	0.0	7.0	
Mando Final TD	1655.0	0.069305	0.252198	0.0	0.00	0.0	0.0	3.0	
Mando Final TI	1662.0	0.078821	0.349692	0.0	0.00	0.0	0.0	9.0	
Masa Derecha	1695.0	0.128319	1.155690	0.0	0.00	0.0	0.0	35.0	
Masa Izquierda	1688.0	0.093780	0.299758	0.0	0.00	0.0	0.0	3.1	
Motor	2118.0	1.770066	1.355686	0.0	1.00	2.0	2.0	37.1	
Radiador	15.0	10.720000	2.315846	7.5	9.35	10.4	11.2	15.5	
Sist de Dirección	1726.0	0.996234	0.692772	0.0	1.00	1.0	1.0	5.0	
Sist. Hidráulico	1819.0	2.157944	1.582278	0.0	1.00	2.0	3.0	34.7	
Transmisión	1831.0	0.971709	0.871189	0.0	1.00	1.0	1.0	22.0	

Hierro

Fe									
	count	mean	std	min	25%	50%	75%	max	
Componente									
Convertidor	1787.0	5.505596	8.560878	0.0	3.00	4.3	6.000	227.0	
Diferencial Trasero	1823.0	22.336753	13.806464	0.0	12.15	19.0	28.000	96.0	
Mando Final TD	1655.0	22.793897	13.963497	3.0	13.00	19.0	29.000	96.0	
Mando Final TI	1662.0	22.887304	13.984084	3.0	13.00	19.0	29.475	96.0	
Masa Derecha	1695.0	23.956519	38.397371	0.6	6.00	11.0	27.000	374.0	
Masa Izquierda	1688.0	30.228791	37.895078	0.6	7.00	16.0	39.000	357.0	
Motor	2119.0	22.516564	15.465287	0.0	13.00	18.7	27.000	128.0	
Radiador	15.0	12.306667	16.178887	0.1	3.10	4.4	16.550	62.1	
Sist de Dirección	1726.0	1.871553	1.273451	0.0	1.00	2.0	2.000	32.0	
Sist. Hidráulico	1819.0	14.312314	37.069066	0.0	5.00	8.0	13.000	709.0	
Transmisión	1831.0	5.676461	8.859478	0.0	3.85	4.8	6.000	228.0	



## Silicio

Si									
	count	mean	std	min	25%	50%	75%	max	
Componente									
Convertidor	1787.0	3.640627	2.517350	0.0	3.00	3.0	4.0	53.0	
Diferencial Trasero	1822.0	2.910154	2.239511	0.0	1.00	2.0	4.0	21.1	
Mando Final TD	1655.0	2.888036	2.181432	0.0	1.40	2.0	4.0	20.0	
Mando Final TI	1662.0	2.901745	2.201103	0.0	1.60	2.0	4.0	19.0	
Masa Derecha	1695.0	4.678053	7.561246	0.0	2.00	3.0	5.0	123.0	
Masa Izquierda	1688.0	5.023460	6.953884	0.0	2.00	3.0	5.0	57.0	
Motor	2118.0	6.231067	68.481294	0.0	2.00	3.0	4.0	1721.0	
Radiador	15.0	39.520000	39.193644	13.2	17.75	25.2	35.7	146.0	
Sist de Dirección	1726.0	3.996466	2.250978	0.0	3.00	4.0	5.0	21.8	
Sist. Hidráulico	1819.0	12.681748	6.011311	0.0	9.00	11.0	15.0	123.0	
Transmisión	1831.0	3.670071	2.428791	0.0	3.00	3.0	4.0	53.0	

## Sodio

Na									
	count	mean	std	min	25%	50%	75%	max	
Componente									
Convertidor	1787.0	2.513822	1.636638	0.0	1.8	2.0	3.0	19.0	
Diferencial Trasero	1822.0	5.772393	1.740392	0.1	5.0	5.1	6.5	19.0	
Mando Final TD	1655.0	5.790030	1.733402	0.1	5.0	5.1	6.5	20.0	
Mando Final TI	1662.0	5.790975	2.131765	0.1	5.0	5.0	6.4	56.0	
Masa Derecha	1695.0	4.589853	1.393374	0.0	4.0	4.0	5.0	19.0	
Masa Izquierda	1688.0	5.258057	6.312212	0.0	4.0	4.0	5.0	88.0	
Motor	2118.0	17.467469	103.289313	0.0	3.0	5.0	7.0	2600.0	
Radiador	15.0	14610.333333	9434.967058	4256.0	7561.0	13944.0	18655.0	35352.0	
Sist de Dirección	1726.0	3.069235	2.174369	0.0	2.0	3.0	4.0	50.0	
Sist. Hidráulico	1819.0	5.386751	5.657373	0.0	3.0	4.0	7.0	96.0	
Transmisión	1831.0	2.577226	1.744503	0.0	2.0	2.0	3.0	20.0	

Y por último la Viscosidad del Aceite:

VISCO									
	count	mean	std	min	25%	50%	75%	max	
Componente									
Convertidor	1766.0	11.105170	0.635523	0.0	10.80	11.0	11.30	14.1	
Diferencial Trasero	1804.0	23.837034	2.163549	0.0	23.65	24.3	24.90	26.8	
Mando Final TD	1643.0	23.912124	1.921640	0.0	23.70	24.3	24.90	26.6	
Mando Final TI	1646.0	23.910923	1.904884	0.0	23.70	24.3	24.90	26.8	
Masa Derecha	1678.0	24.590942	11.015474	0.0	23.70	24.2	25.00	341.7	
Masa Izquierda	1671.0	24.574722	10.849432	0.0	23.70	24.3	24.97	337.2	
Motor	2090.0	14.590799	0.698541	7.0	14.40	14.6	14.80	37.7	
Radiador	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Sist de Dirección	1708.0	10.906534	0.786144	0.0	10.80	10.9	11.20	21.6	
Sist. Hidráulico	1796.0	6.514761	0.577938	0.0	6.30	6.6	6.70	13.8	
Transmisión	1809.0	11.111316	0.655656	0.0	10.80	11.0	11.30	14.2	

La mayor presencia de Sodio se da en el Radiador, lo cual parece lógico por ser el componente que trabaja principalmente con líquido refrigerante (factiblemente conteniendo un porcentaje de agua). El componente que en segundo lugar presenta valores relacionados al agua, aunque en mucha menor medida es el motor, lugar donde no debería detectarse la presencia de este elemento, o bien de detectarse sería un indicador de alguna problemática.

También se observan las mayores concentraciones de Hollín en el motor, y de forma marcadamente menor en el Diferencial, Mando Final TD y Mando Final TI. El resto de los componentes no presentan valores para este elemento.

La mayor presencia de Aluminio se da en el Radiador, factiblemente este valor se encuentre influenciado por los componentes propios del sistema de refrigeración. Mientras que el Hierro se presenta de forma más o menos pareja en casi todos los componentes, y de forma mucho más marcada en el motor.

En cuanto al Silicio podría esperarse que se encuentre en mayores concentraciones tanto el Radiador como en el Sistema Hidráulico, dado que es un componente elemento ligado a la arena.

Los resultados observados en general, para la media y mediana, parecen acordes al tipo de componente y elemento químico en cada caso, ya que al producirse desgaste de los mismos con el uso, van generando los diferentes residuos químicos que detecta el laboratorio. Con lo cual sería razonable graficarlos en el tiempo como lo haremos más adelante en este mismo informe.

En principio estos valores no permiten obtener una medida de correlación entre las variables, más adelante también profundizaremos con otro tipo de gráficos que nos permitan encontrar relaciones y medir la mismas en caso de encontrarse.

### 1.3

Si bien el dataset presenta valores nulos en sus filas, en algunos componentes no parece deberse al hecho que los vehículos no siempre son muestreados en todos sus componentes, sino que parecen ser el resultado natural de la relación entre el elemento químico en cuestión y el tipo de componente analizado.

Así es de esperarse que para el Hollín (St) existan valores nulos en los componentes: Convertidor, Radiador, Masa Derecha, Masa Izquierda, Sistema de Dirección, Sistema Hidráulico y Transmisión. Esto sería interesante poder confirmarlo con el experto del dataset o dominio de negocio.

En estos escenarios sería factible completar con valores 0 (cero) los campos donde no hay información relevada (Null).

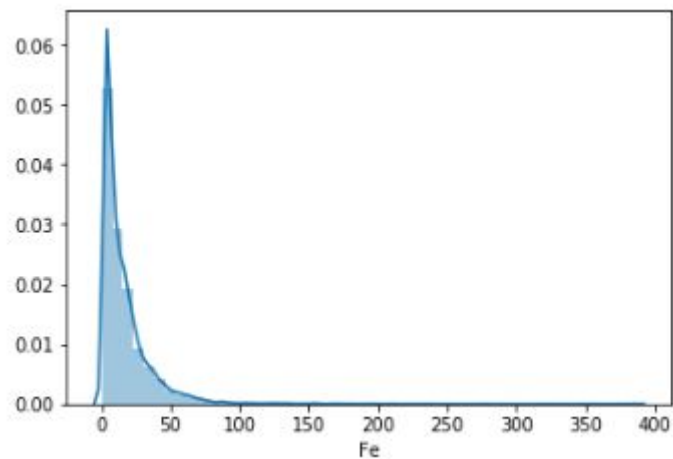
Para otro lado para los casos en que efectivamente existen valores nulos debido a que no siempre se realizan pruebas para todos los componentes, considerando que estos representan un valores relativamente pequeño respecto de tamaño de la muestra, por ejemplo 55 casos nulos de Fe en una muestra de 18.469 valores muestrales, podría decir que la supresión de estos resultados no afectará nuestros indicadores. Los valores nulos totales, representan el 15



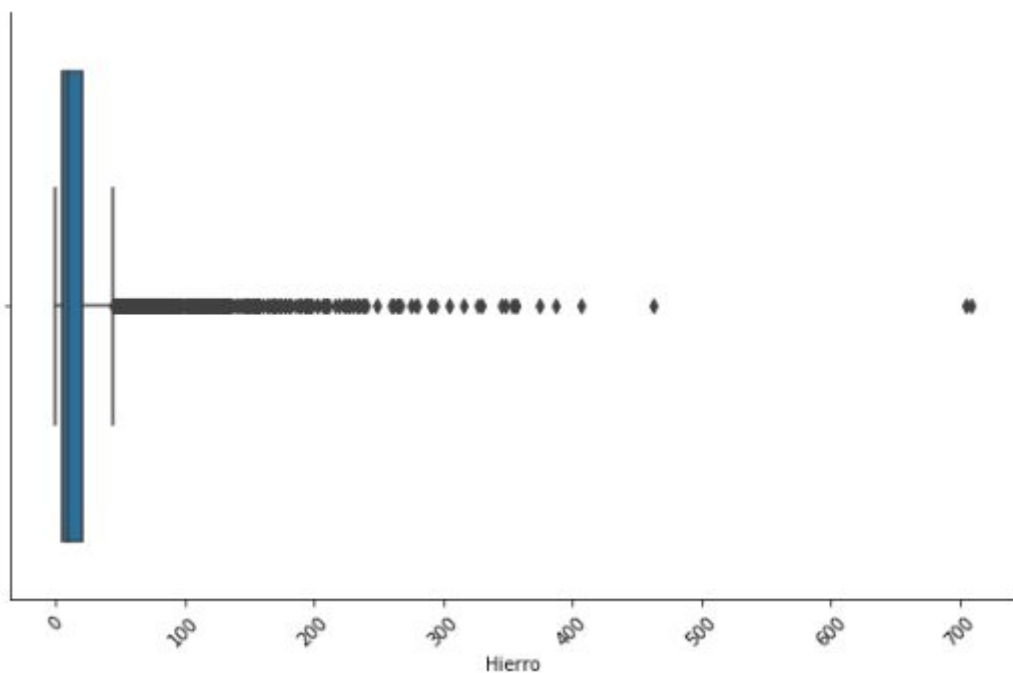
% de los datos de la muestra. (Este punto puede ser revisado luego de la materia Exploración y Curación de Datos)

#### 1.4

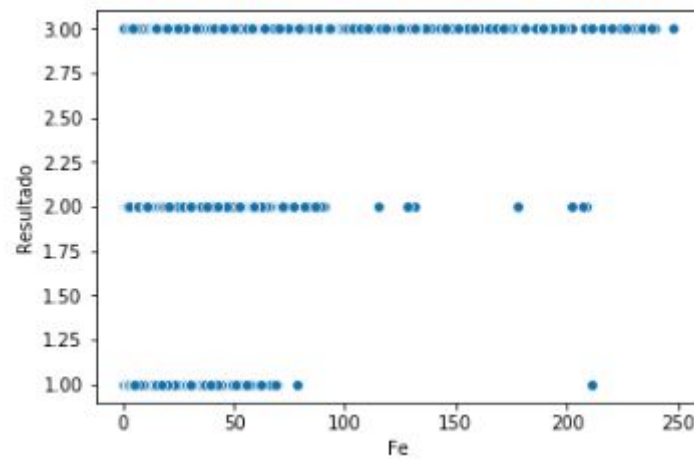
Una de las características más importantes en la muestra es la presencia de Hierro (Fe) en casi todos los componentes. El cual comentamos en el apartado 1.2 parece ser un elemento químico que está presente en casi todos los componentes de una forma más o menos homogénea. Si graficamos la distribución de probabilidad, parece representar una distribución exponencial:



Tenemos algunos indicios de la existencia de outliers, esto es, valores extraños en la muestra. Si observamos discriminado por componentes, nuestros outliers para el Hierro tienen la siguiente forma:



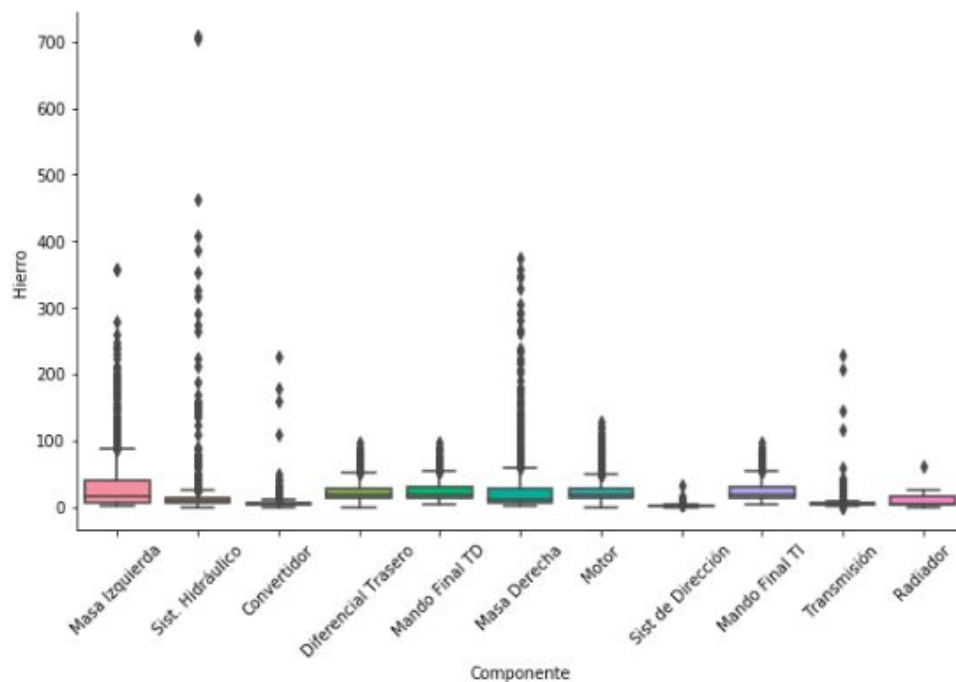
Utilizamos un gráficos de puntos buscando correlación entre estas variables, eliminando los valores raros, sin poder de momento establecer una relación directa entre la presencia de Hierro y el Resultado del laboratorio:



El cálculo de correlación de Spearman con un valor de 0.38 tampoco es lo suficientemente fuerte como para afirmar una correlación, sin embargo tampoco podemos afirmar que las variables sean totalmente independientes entre sí.

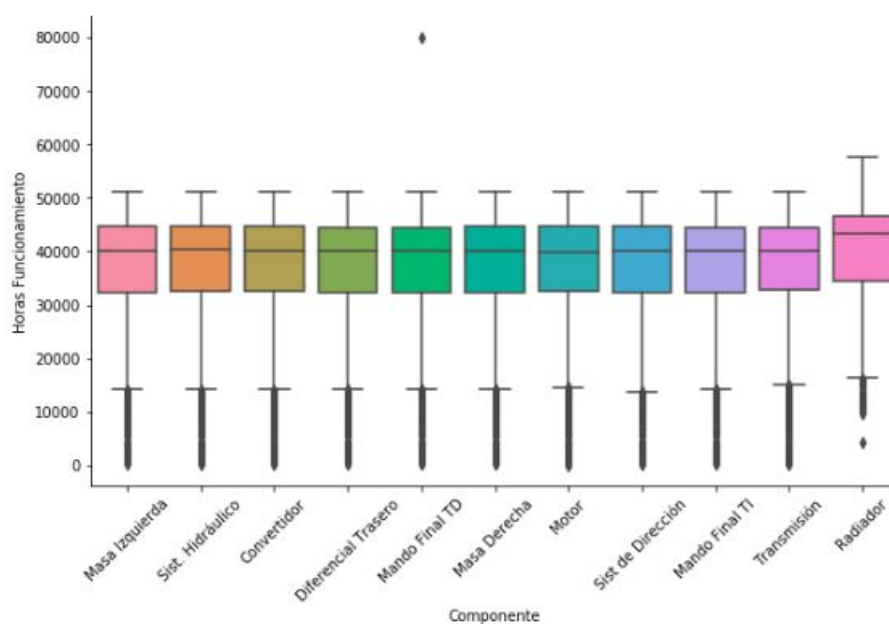
Revisando puntualmente los outliers nos encontramos con los siguientes casos:

- El Hierro presenta 4 valores raros por encima de un valor Fe de 400 y se da para el Sistema Hidráulico en 4 vehículos:



	Equipo	Componente	Id	Resultado	Horas Funcionamiento	Horas del Aceite	St	Al	Fe	Si	Na	VISCO	
	27	1356	Sist. Hidráulico	155972	3	27624	187.0	NaN	1.0	407.0	13.0	19.0	7.0
	7159	1395	Sist. Hidráulico	280961	3	43242	587.0	NaN	5.0	705.0	20.0	3.0	6.7
	12954	1405	Sist. Hidráulico	275521	3	40048	424.0	NaN	1.0	709.0	14.0	4.0	6.6
	13365	1407	Sist. Hidráulico	225504	3	38458	379.0	NaN	1.0	462.0	8.0	10.0	6.4

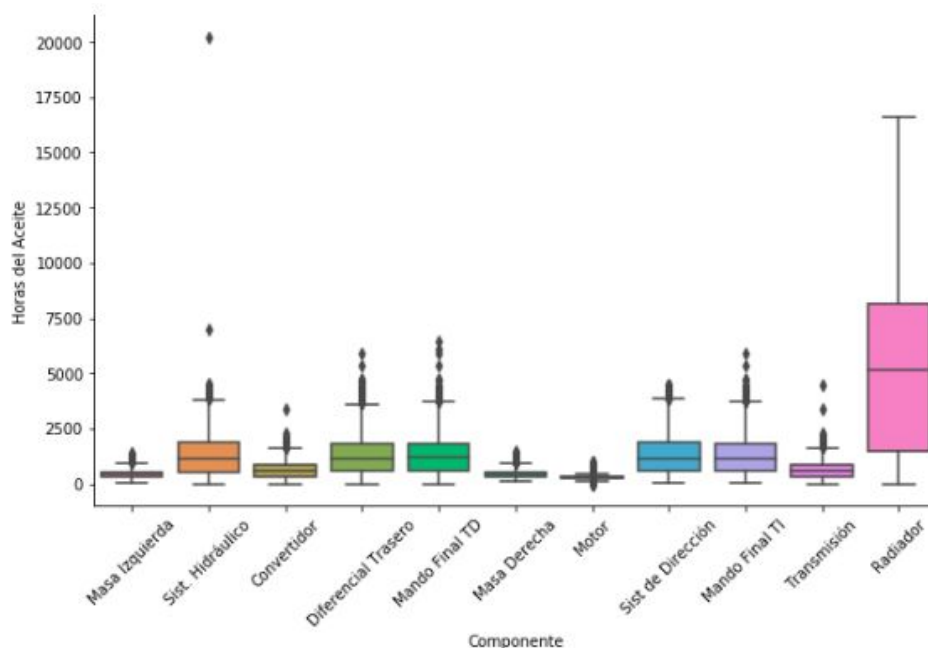
- Tenemos un valor atípico en las horas de funcionamiento, un vehículo (1402) donde se revisó el Mando Final TD y se marcan 79.990 hs:



```
dataset[dataset['Horas_Funcionamiento'] > 60000]
```

	Equipo	Componente	Id	Resultado	Horas_Funcionamiento	Horas_del_Aceite	B	Nit	Oxi	Sul	...	P	Zn	Ag	Ti	VISCO	TBN	TAM
12297	1402	Mando Final TD	248433	2	79990	1197.0	89.0	4.0	9.0	12.0	...	370.0	106.0	0.0	0.0	21.9	NaN	NaN

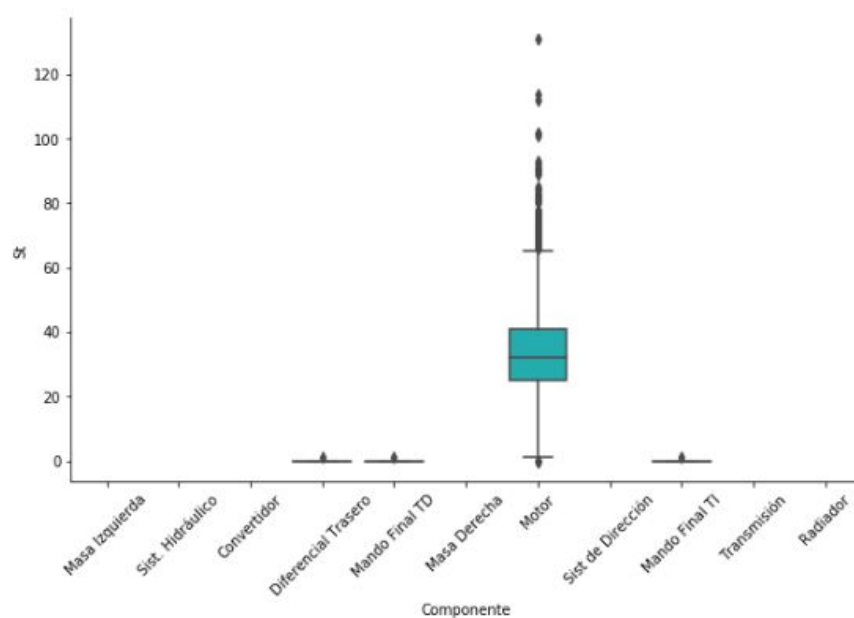
- En las horas de aceite detectamos un vehiculo con 31.873 hs en el Sist. Hidráulico:



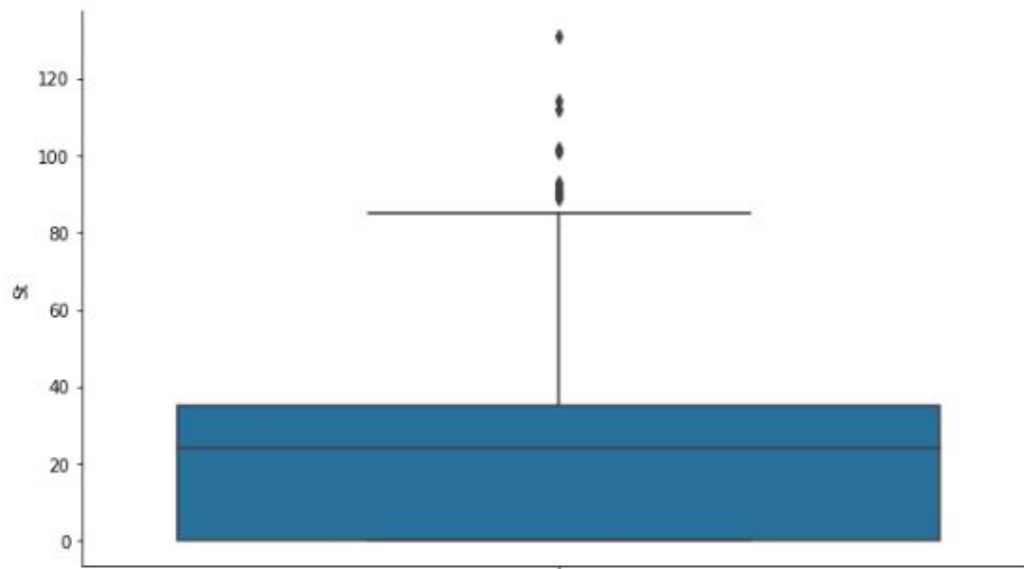
```
dataset[dataset['Horas_del_Aceite'] > 17000]
```

	Equipo	Componente	Id	Resultado	Horas_Funcionamiento	Horas_del_Aceite	B	Nit	Oxi	Sul	...	P	Zn	Ag	Ti	VISCO	TBN	T.
17790	1426	Sist. Hidráulico	229563	2	31873	20208.0	0.0	NaN	NaN	NaN	...	964.0	1188.0	0.0	0.0	6.8	NaN	N

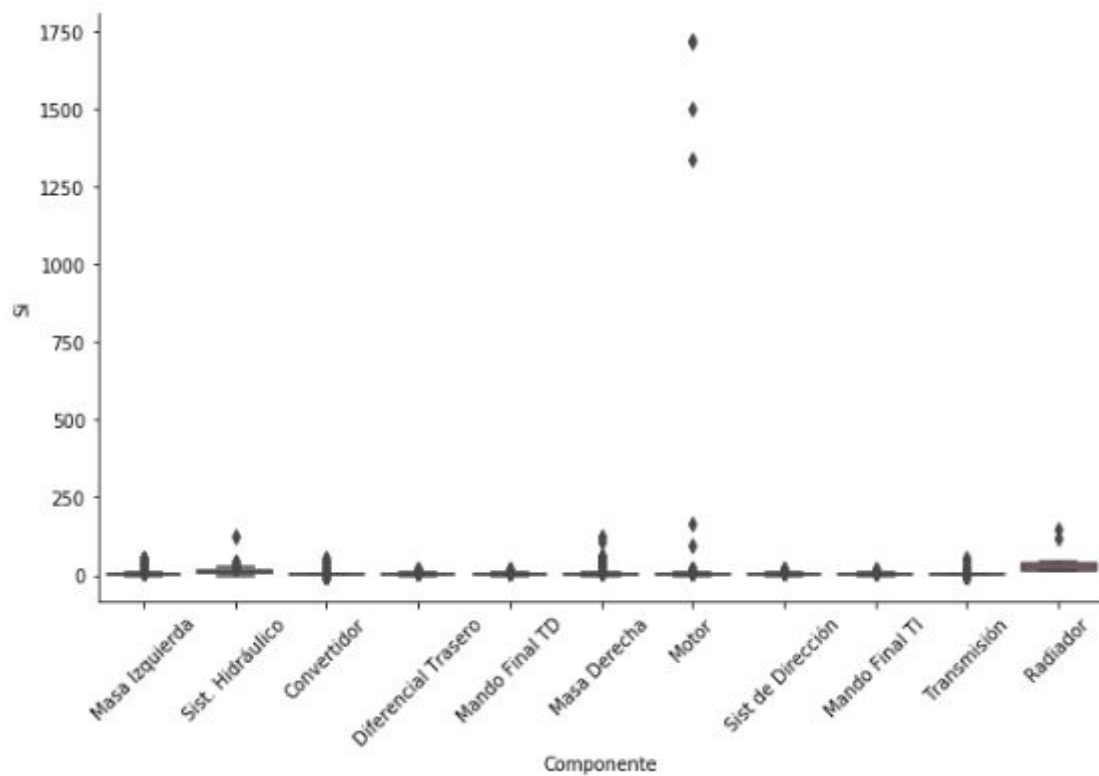
- El Hollín tiene valores atípicos por encima de los 80 y la mayor distribución se encuentra en el motor:



Observamos más de cerca el límite de los valores fuera de rango:



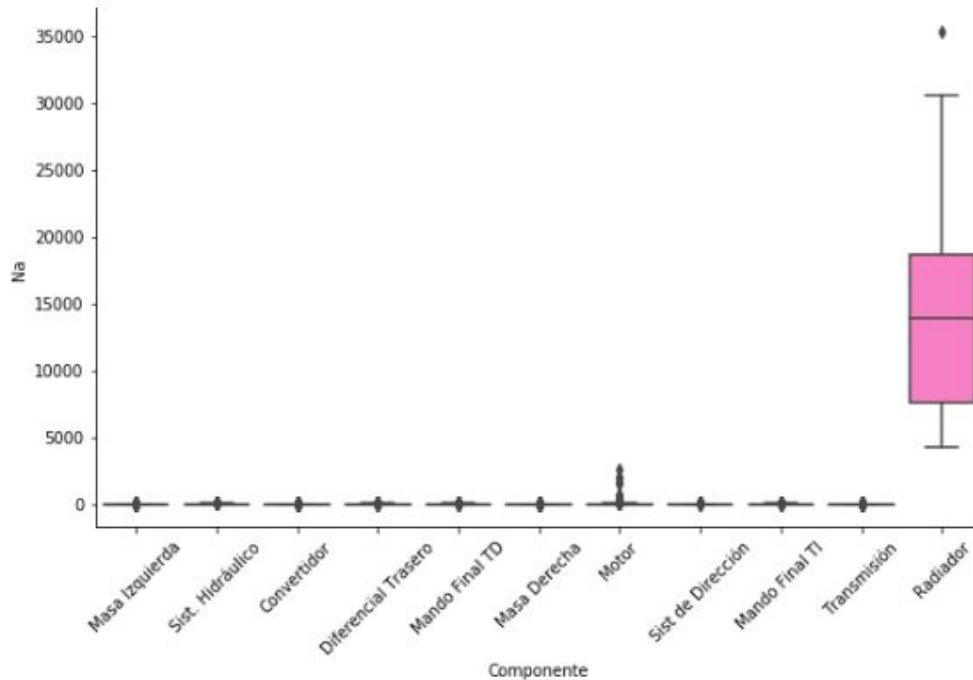
- El cuanto al Silicio tenemos 4 valores extraños por encima de (Si) 1.000 en el motor del equipo 1834





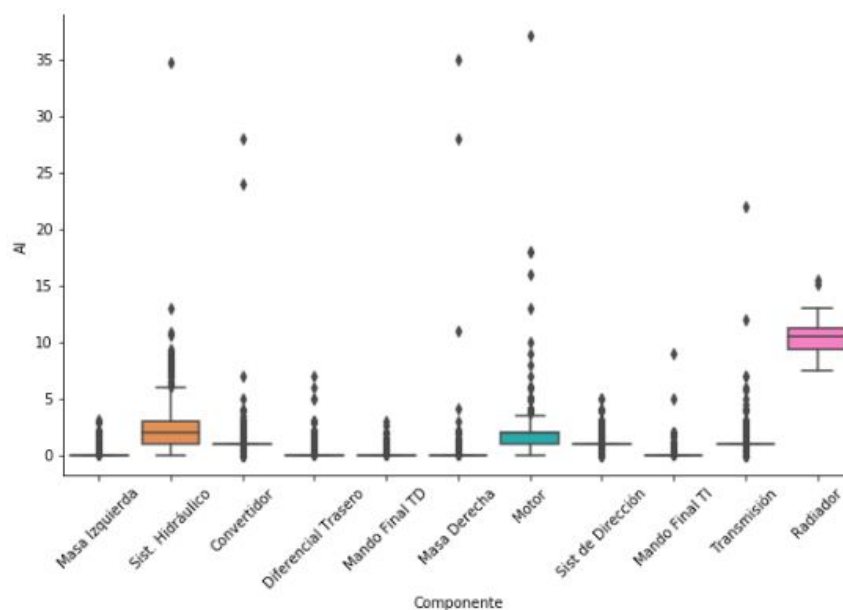
Equip	Componente	Id	Resultad	Horas Funcionamien	Horas del Acei	Si
1834	Motor	269441	2	17829	215	1714
1834	Motor	270353	2	17868	254	1721
1834	Motor	273136	2	18057	443	1501
1834	Motor	274912	3	18148	443	1338

- En equipo 1946 con 2 valores atípicos de Sodio para el Radiador del equipo 1946:



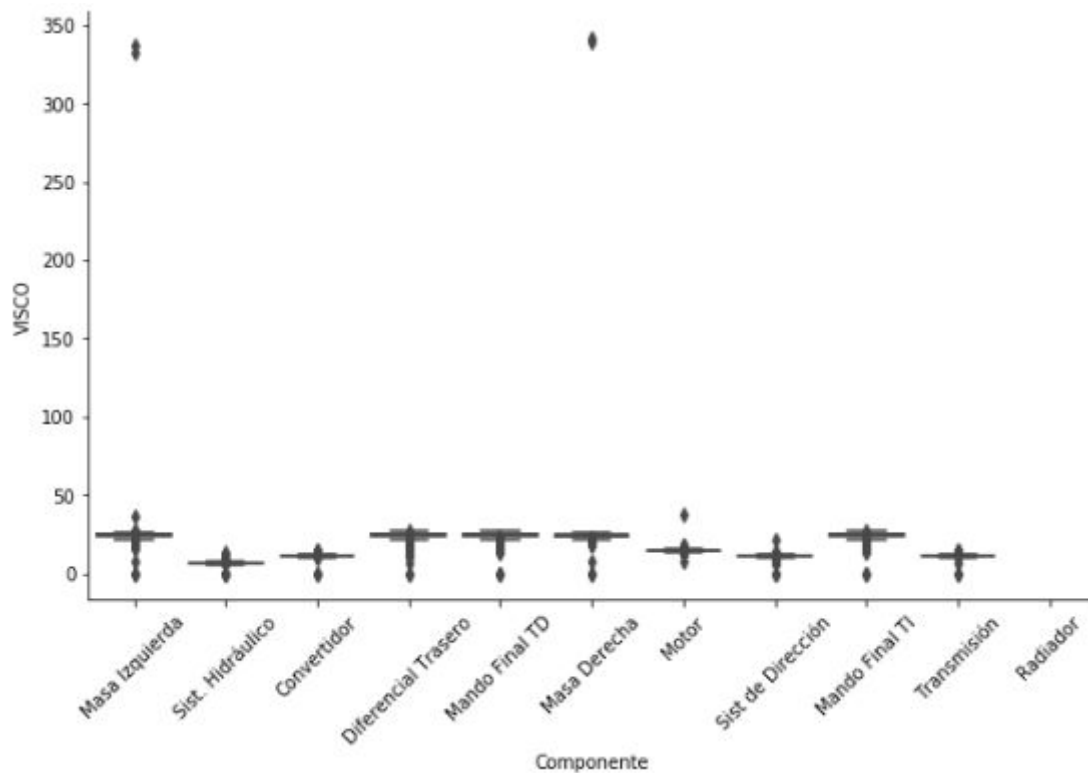
Equip	Componente	Id	Resultad	Horas Funcionamien	Horas del Acei	Na
1946	Radiador	256389	3	34233	4018	35352
1946	Radiador	256392	3	48007	3653	30556

- El Aluminio presenta outliers de forma más genérica por encima del valor 15, detectandose 12 equipos con esta característica



Equipo	Componente	Id	Resultado	Horas Funcionamiento	Horas del Aceite	AI
1355	Masa Derecha	278700	3	22349	399	28
1355	Masa Derecha	284180	3	22729	380	35
1360	Sist. Hidráulico	253201	2	48274	3169	34.7
1366	Motor	259186	3	46019	154	18
1372	Motor	191921	3	40116	275	18
1366	Motor	258493	3	45977	112	16
1391	Convertidor	203879	1	41492	279	28
1391	Convertidor	211524	3	42042	829	24
1391	Transmisión	196572	3	40951	852	22
1418	Motor	255029	2	33305	254	37.1
1946	Radiador	256388	3	48576	2247	15.1
1946	Radiador	256390	3	33305	5833	15.5

- Por último tenemos 4 valores atípico en dos equipos 1359 y 1357 para la Viscosidad por encima del valor 300:



```
dataset[dataset['VISO'] > 300]
```

	Equipo	Componente	Id	Resultado	Horas_Funcionamiento	Horas_del_Aceite	B	Nit	Oxi	Sul	...	P	Zn	Ag	Ti	VISCO	TBN	TAN
83	1356	Masa Izquierda	155977	1	27624	663.0	NaN	NaN	NaN	NaN	...	411.0	31.0	0.0	0.0	337.2	NaN	NaN
97	1356	Masa Derecha	155976	1	27624	663.0	NaN	NaN	NaN	NaN	...	411.0	20.0	0.0	0.0	339.6	NaN	NaN
124	1357	Masa Derecha	156981	1	27876	519.0	NaN	NaN	NaN	NaN	...	441.0	22.0	0.0	0.0	341.7	NaN	NaN
125	1357	Masa Izquierda	156986	1	27876	519.0	NaN	NaN	NaN	NaN	...	437.0	20.0	0.0	0.0	332.3	NaN	NaN

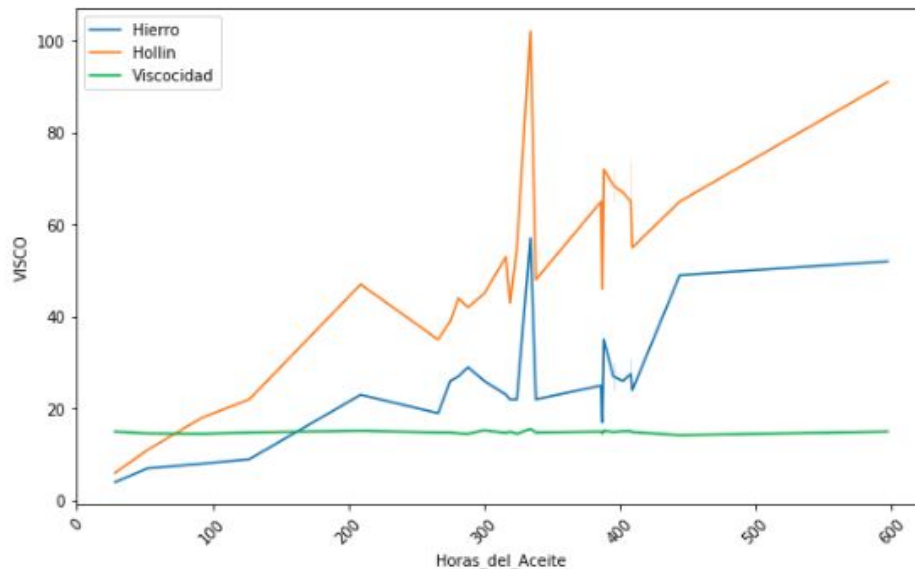
4 rows × 42 columns

## 2.1

En esta sección revisaremos casos específicos buscando detectar algún indicador de relevancia.

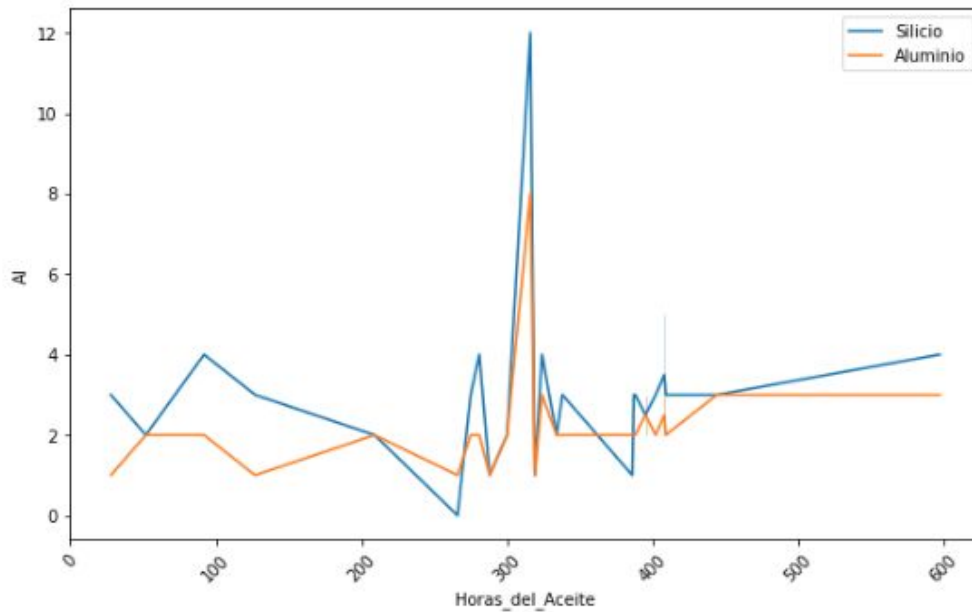
En primer lugar tomamos 3 motores al azar y graficamos la evolución de al menos 3 características en el tiempo. Primeramente y por indicaciones de un experto graficamos Hierro, Hollín y Viscosidad

Motor 1395:



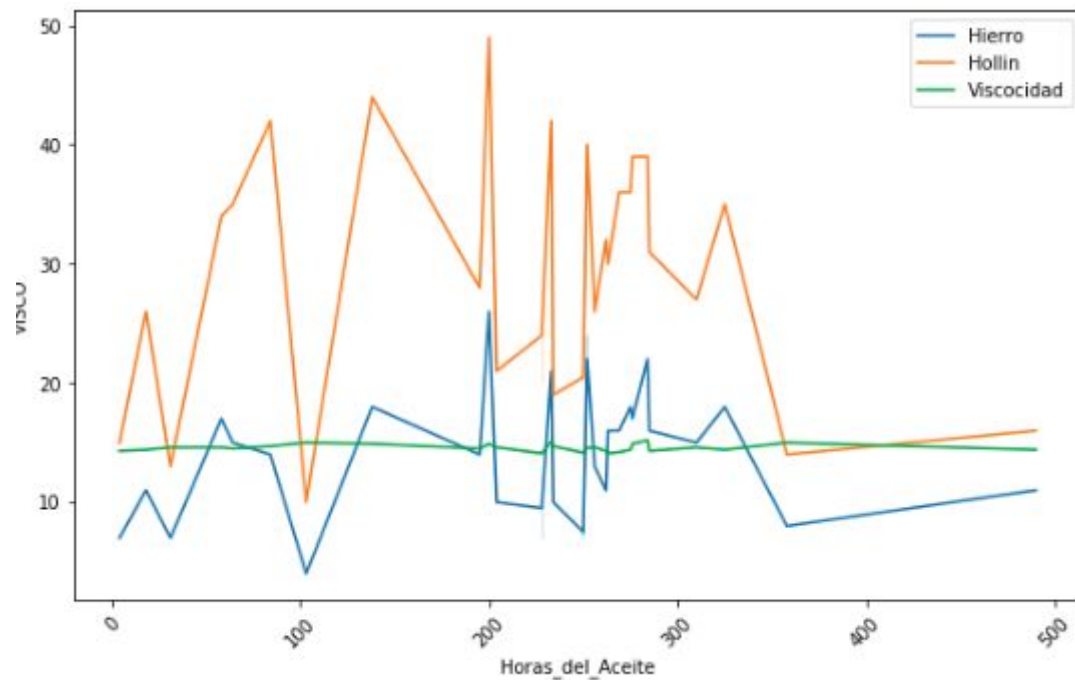
Para las tres características mencionadas para el motor 1395 se muestra una tendencia creciente en el tiempo y con una fluctuación similar para el Hierro y Hollín, presentando un pico pasando las 300 horas de aceite.

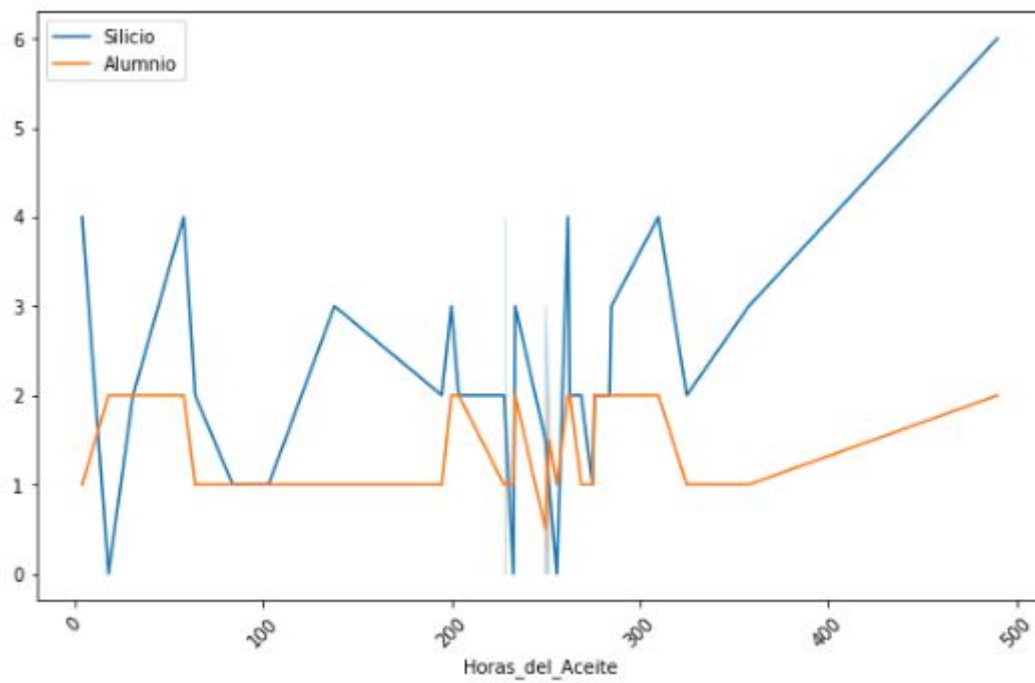
Adicionalmente comparamos el Silicio con el Aluminio, con una aparente tendencia en su comportamiento medido en las horas de uso del aceite, aunque estas fluctuaciones parecen menos acompañadas que las anteriores variables:



### Motor 1427

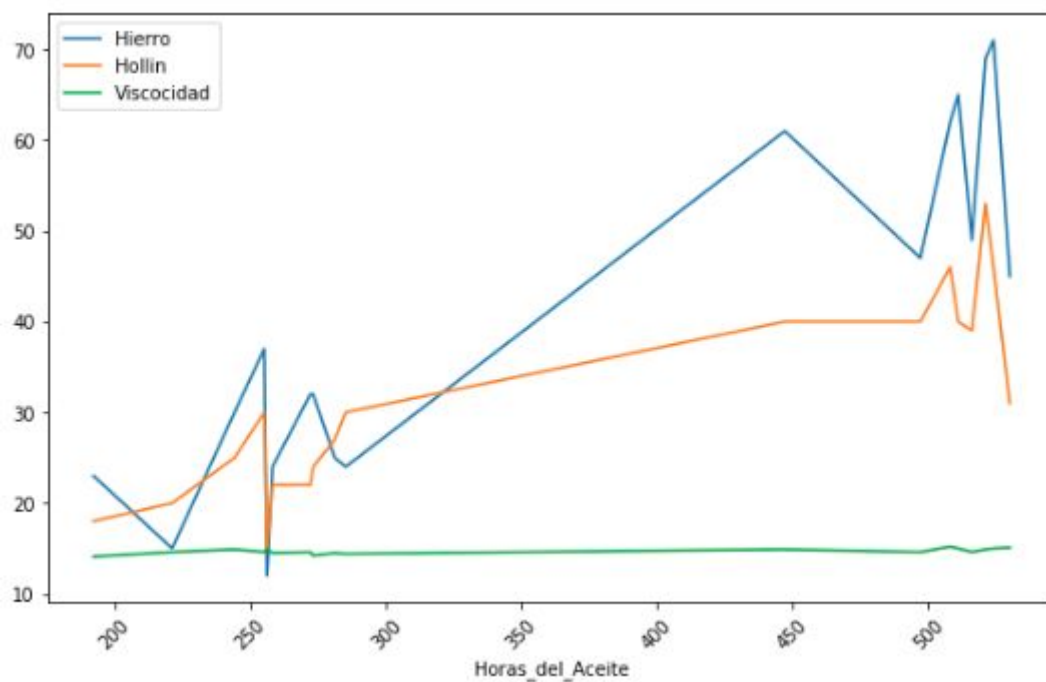
Algo similar ocurre con el motor 1427, las variables parecen presentar un comportamiento hermanado entre las variables Hierro - Hollín y Silicio - Aluminio, como se puede observar en las siguientes gráficas:



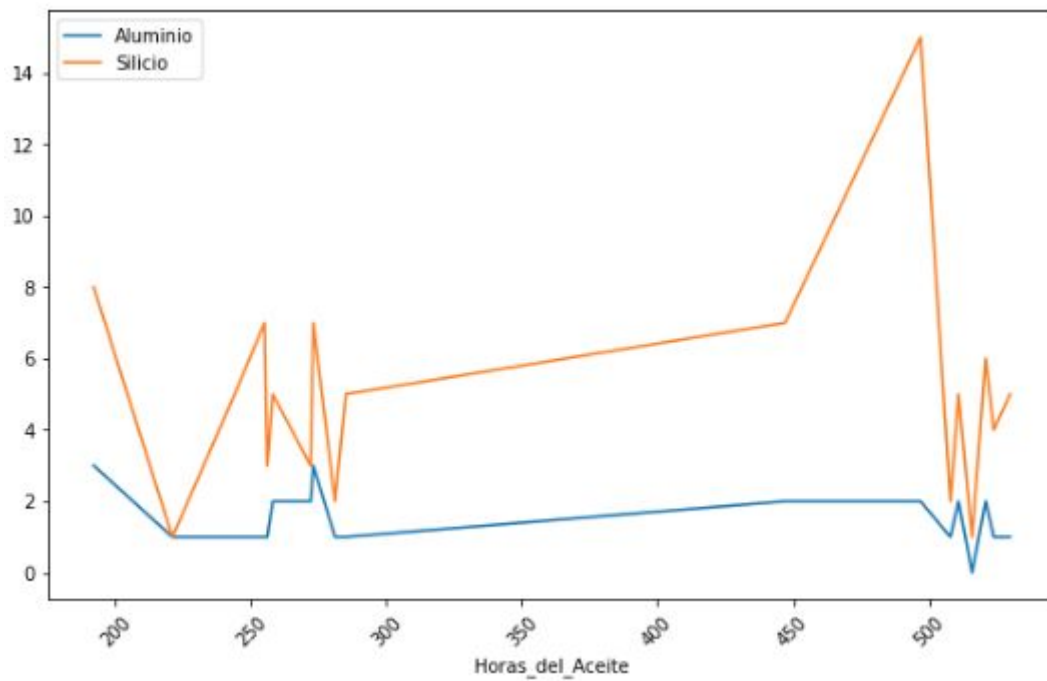


Motor 2061

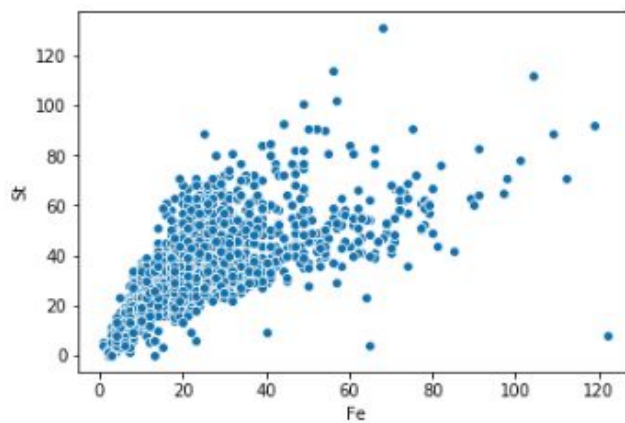
Mismas observaciones que en los motore previos, motivos por lo cual, podríamos intentar algún gráfico y cálculos que correlacionen estas variables:



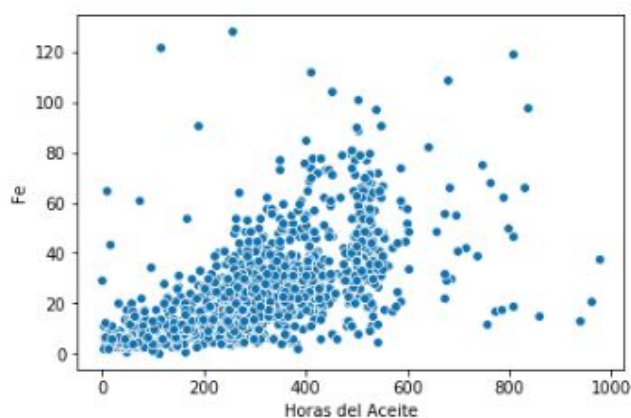




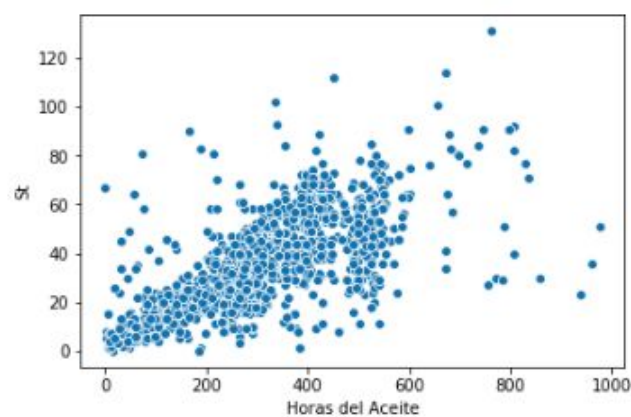
Intentamos establecer una correlación para estos valores, en los motores en general, basado en los gráficos previos. Pasamos a los gráficos de puntos, además calculamos la medida de correlación utilizando Spearman, ya que se trata de variables continuas que no tienen una distribución normal:



	Fe	St
Fe	1.000000	0.704958
St	0.704958	1.000000



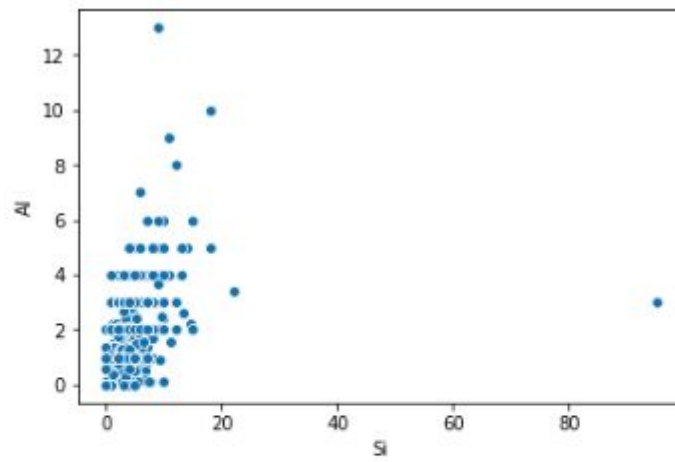
	Horas del Aceite	Fe
Horas del Aceite	1.000000	0.640501
Fe	0.640501	1.000000



	Horas del Aceite	St
Horas del Aceite	1.000000	0.659333
St	0.659333	1.000000

Encontramos una correlación positiva entre la presencia de Hierro y Hollín con un valor de 0.70, que no se presenta tan marcada entre estos dos elementos químicos y las Horas de Aceite en el caso del motor.

En cuanto al Silicio y el Aluminio la correlación es mucho más débil:



```
motores[['Al', 'Si']].corr(method='spearman')
```

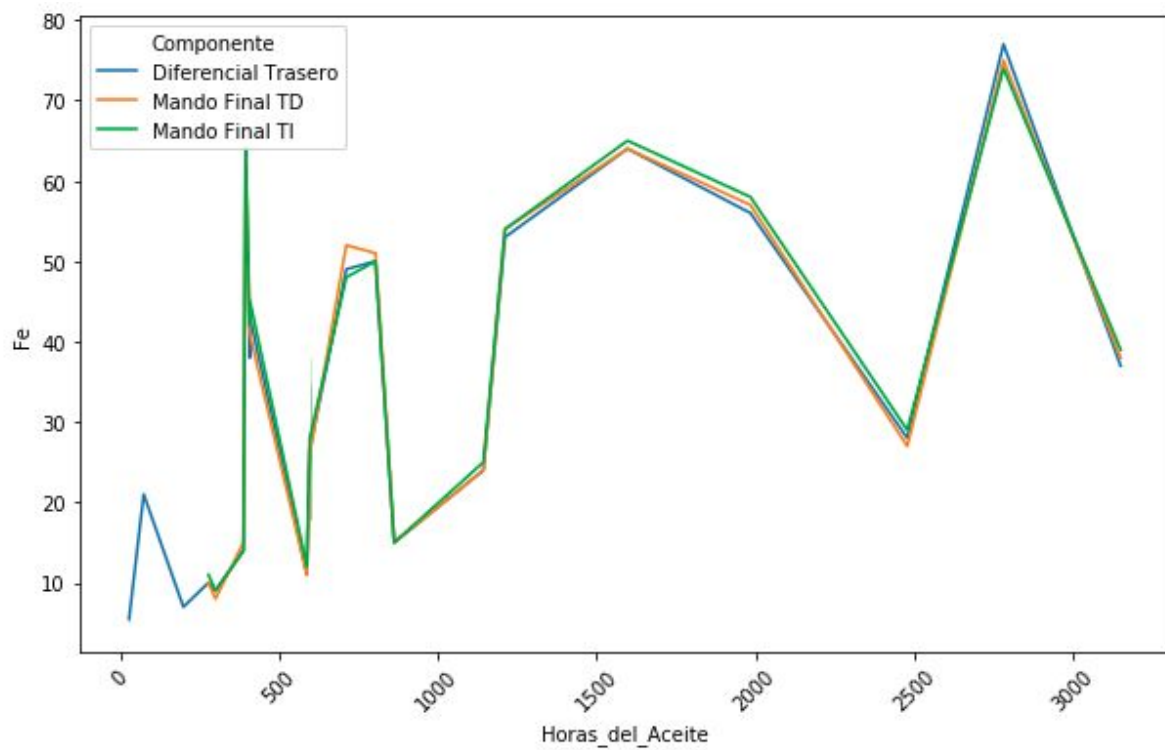
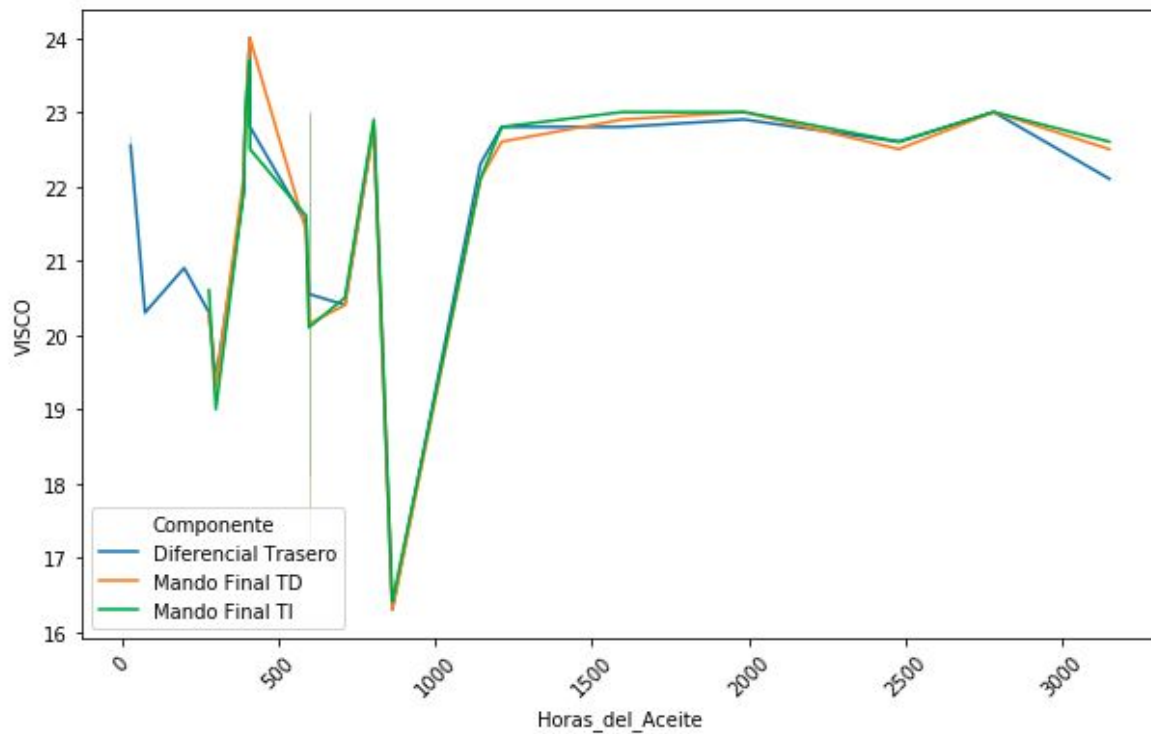
	Al	Si
Al	1.000000	0.362485
Si	0.362485	1.000000

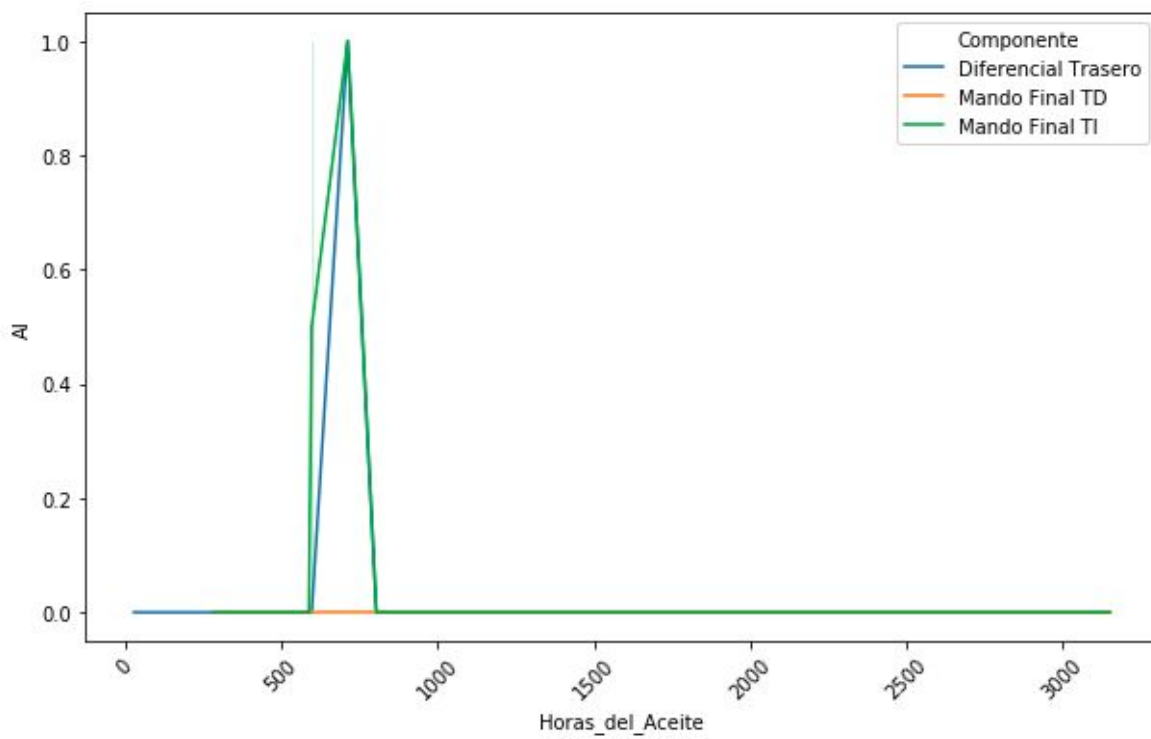
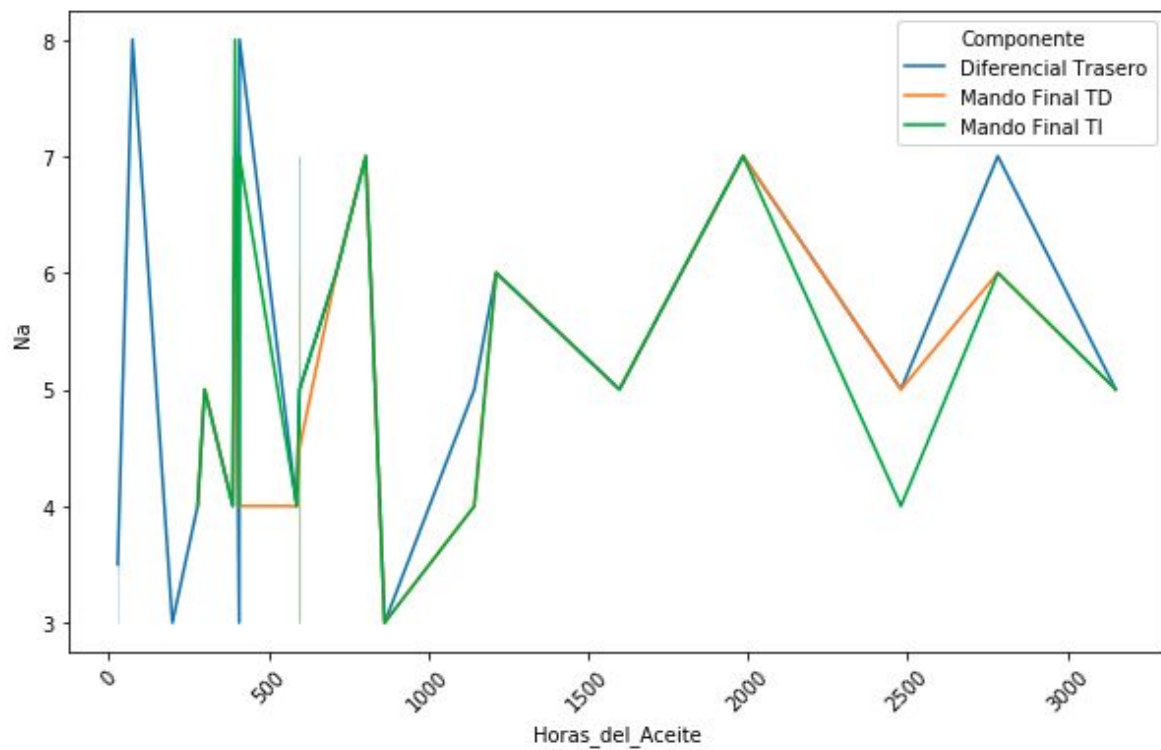
No podemos asegurar que las variables están relacionadas, aunque tampoco podemos asegurar que sean totalmente independientes.

## 2.2

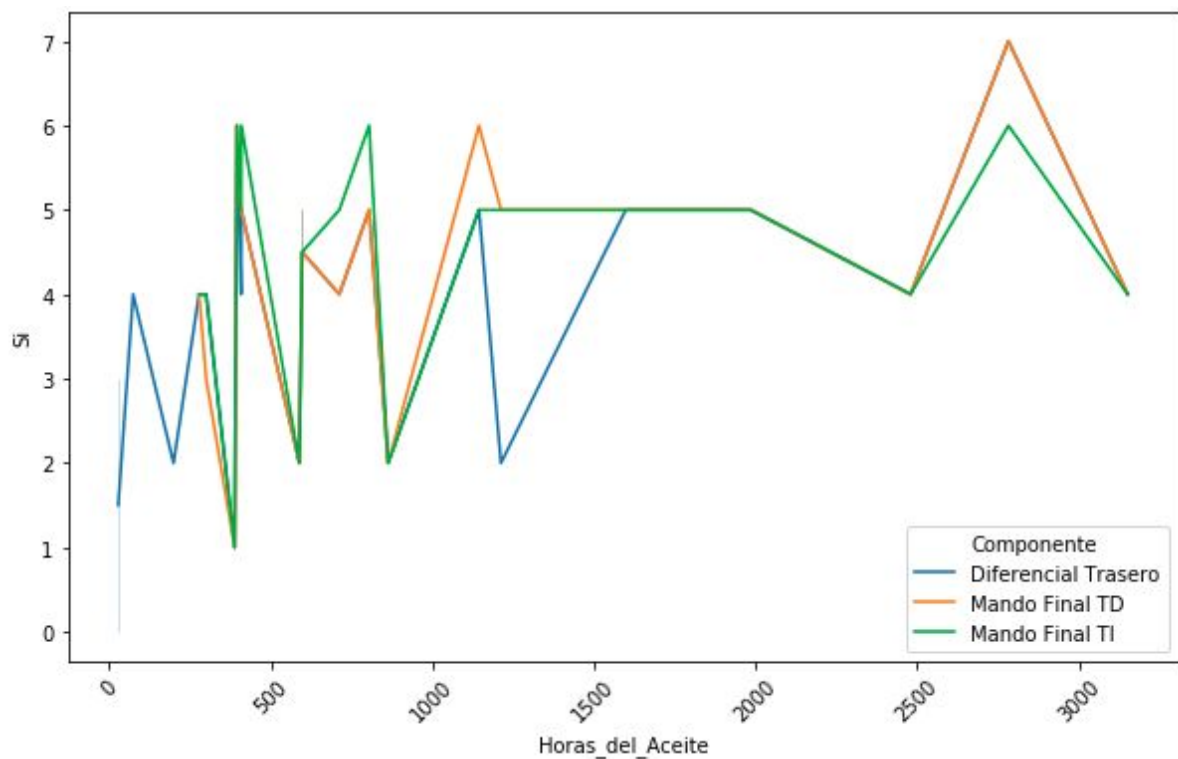
Si centramos nuestra atención en un solo equipo, analizamos algunos componentes y observamos el comportamiento de los elementos químicos principales:

Diferencial Trasero, Mando Final TD, Mando Final TI

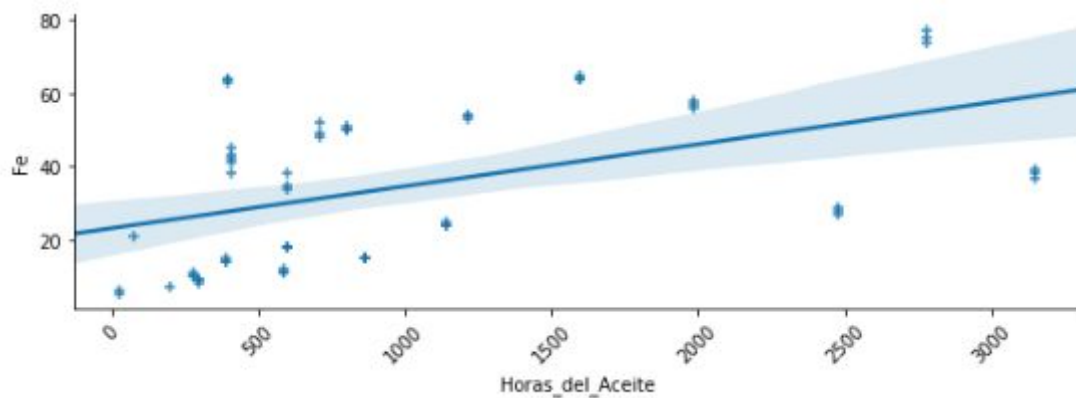








Observamos que en estos componentes los elementos químicos tienen un comportamiento casi idéntico en el tiempo. Si medimos la correlación de los diferentes elementos químicos con las Horas del Aceite para estos tres componentes:



```
equipol395_3comp[['Horas_del_Aceite', 'Fe']].corr(method='spearman')
```

	Horas_del_Aceite	Fe
Horas_del_Aceite	1.000000	0.602664
Fe	0.602664	1.000000

```
equipol395_3comp[['Horas_del_Aceite', 'Al']].corr(method='spearman')
```

	Horas_del_Aceite	Al
Horas_del_Aceite	1.0	0.0
Al	0.0	1.0

```
equipol395_3comp[['Horas_del_Aceite', 'Si']].corr(method='spearman')
```

	Horas_del_Aceite	Si
Horas_del_Aceite	1.000000	0.349961
Si	0.349961	1.000000

```
equipol395_3comp[['Horas_del_Aceite', 'Na']].corr(method='spearman')
```

	Horas_del_Aceite	Na
Horas_del_Aceite	1.000000	0.232837
Na	0.232837	1.000000

```
equipol395_3comp[['Horas_del_Aceite', 'St']].corr(method='spearman')
```

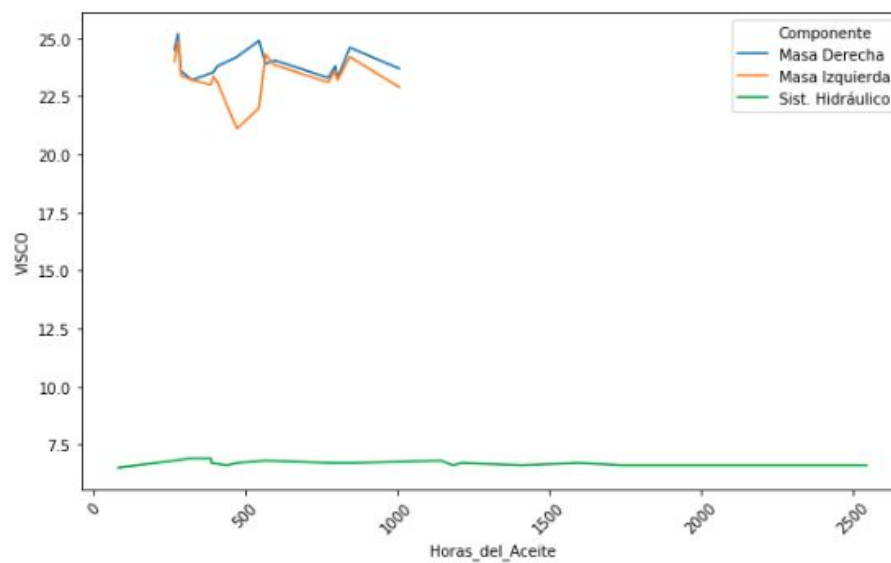
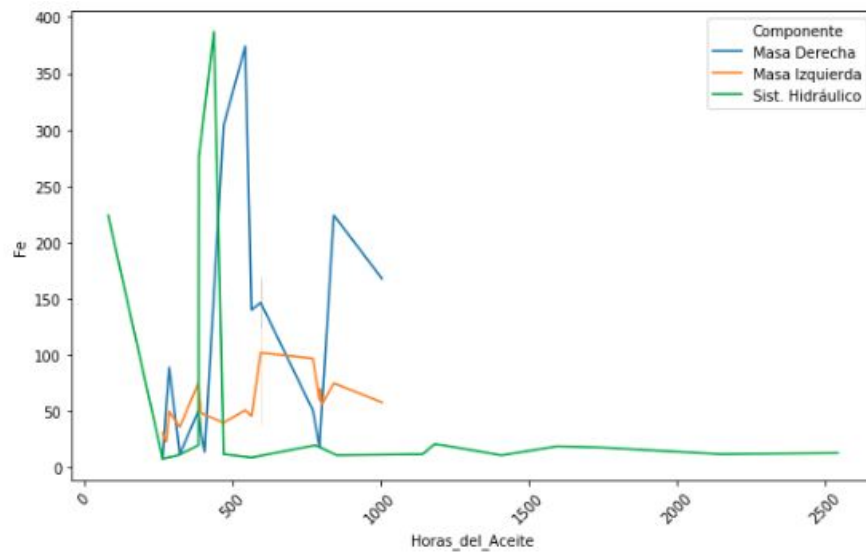
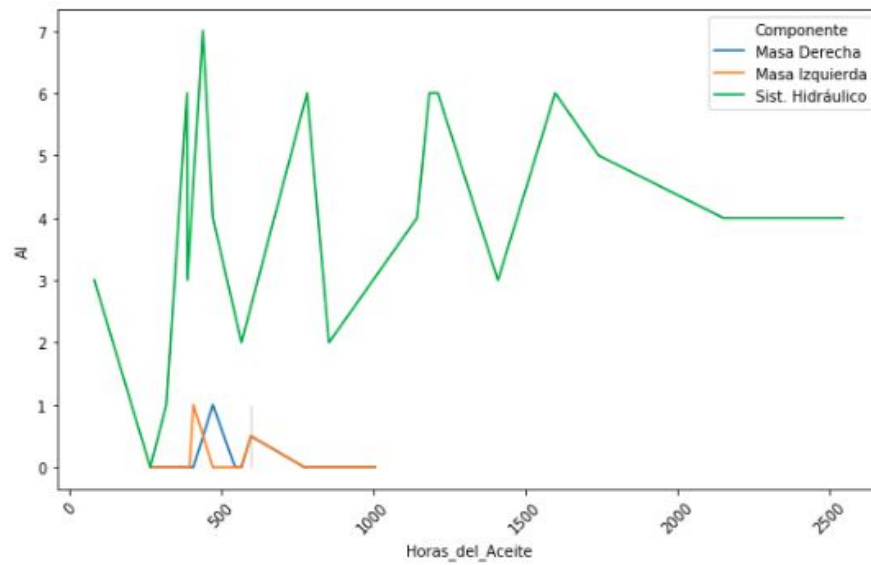
	Horas_del_Aceite	St
Horas_del_Aceite	1.0	NaN
St	NaN	NaN

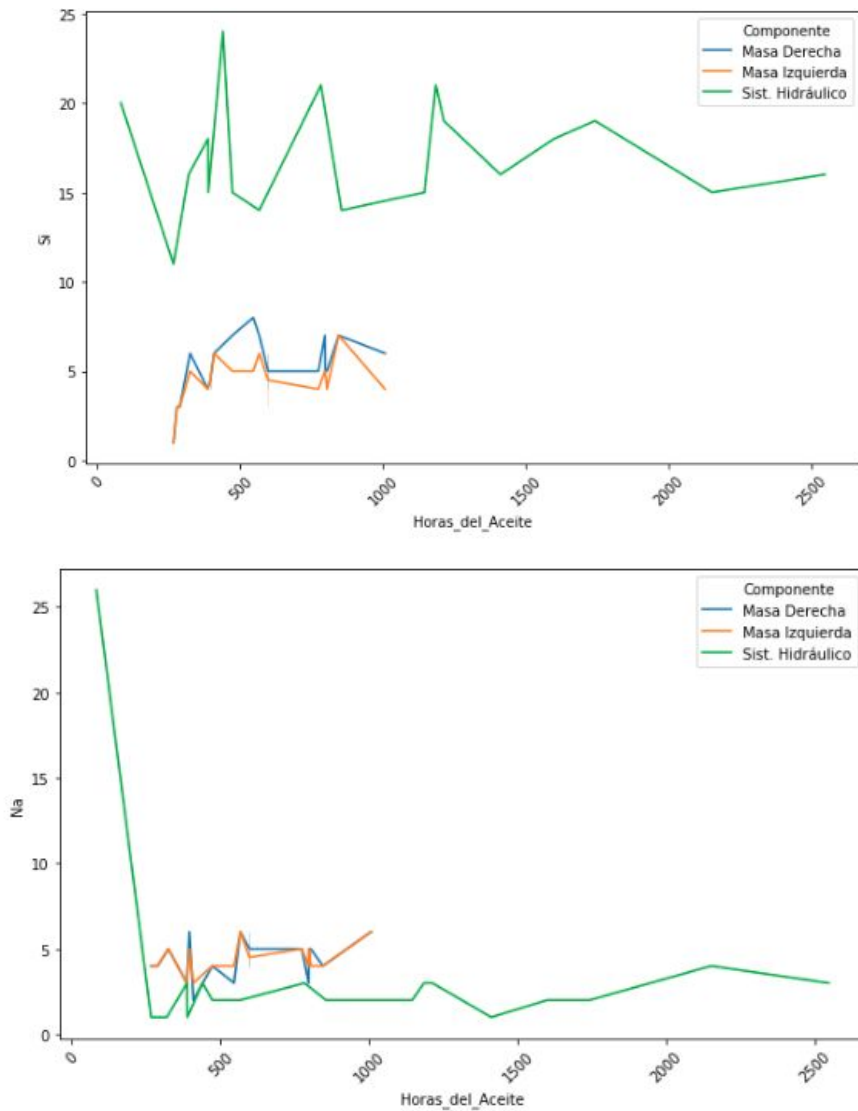
```
equipol395_3comp[['Horas_del_Aceite', 'VISCO']].corr(method='spearman')
```

	Horas_del_Aceite	VISCO
Horas_del_Aceite	1.000000	0.298079
VISCO	0.298079	1.000000

Podemos encontrar la correlación más fuerte entre el Hierro y las Horas de Aceite con un valor de 0.60

### Masa Derecha, Masa Izquierda, Sist. Hidráulico





Observamos que Masa Derecha y Masa Izquierda tienen un comportamiento similar en el tiempo. Si analizamos los elementos químicos en relación a las Horas de Aceite para estos componentes, las correlaciones entre elementos químicos son más bien débiles que para los componentes anteriores. Sin embargo por primera vez encontramos una correlación negativa entre las Horas de Aceite y la Viscosidad:

```
equipol395_3otros[['Horas_del_Aceite', 'Fe']].corr(method='spearman')
```

	Horas_del_Aceite	Fe
Horas_del_Aceite	1.000000	-0.075385
Fe	-0.075385	1.000000

```
equipol395_3otros[['Horas_del_Aceite', 'Al']].corr(method='spearman')
```

	Horas_del_Aceite	Al
Horas_del_Aceite	1.000000	0.339465
Al	0.339465	1.000000

```
equipol395_3otros[['Horas_del_Aceite', 'Si']].corr(method='spearman')
```

	Horas_del_Aceite	Si
Horas_del_Aceite	1.000000	0.405103
Si	0.405103	1.000000

```
equipol395_3otros[['Horas_del_Aceite', 'Na']].corr(method='spearman')
```

	Horas_del_Aceite	Na
Horas_del_Aceite	1.000000	-0.095692
Na	-0.095692	1.000000

```
equipol395_3otros[['Horas_del_Aceite', 'St']].corr(method='spearman')
```

	Horas_del_Aceite	St
Horas_del_Aceite	1.0	NaN
St	NaN	NaN

```
equipol395_3otros[['Horas_del_Aceite', 'VISCO']].corr(method='spearman')
```

	Horas_del_Aceite	VISCO
Horas_del_Aceite	1.000000	-0.271508
VISCO	-0.271508	1.000000

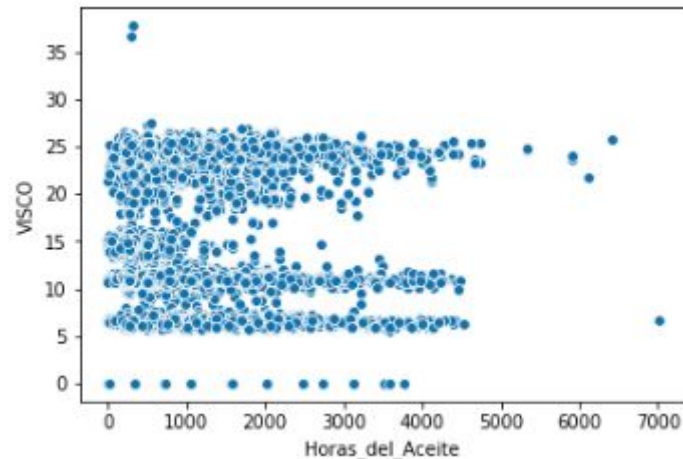


### 3.3

Del análisis previo observamos una correlación positiva entre el Hierro y el Hollín en un valor de 0.70 y en menor medida entre las Horas de Aceite y el Hierro en 0.64; Horas de Aceite y Hollín en 0.66.- Estos valores son válidos para los motores.

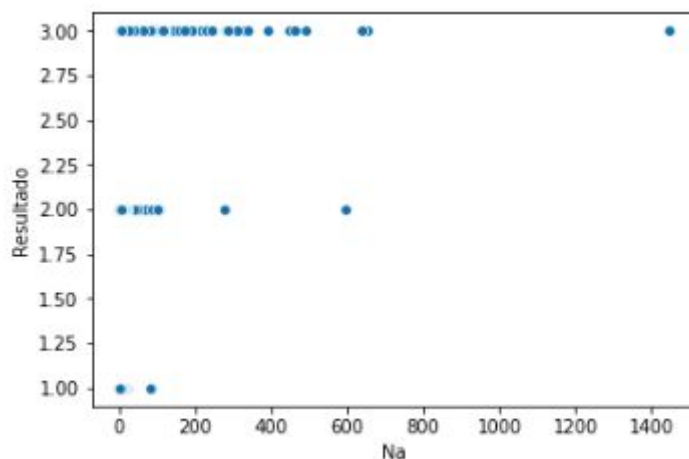
Además encontramos una correlación positiva de 0.60 entre las Horas de Aceite y Hierro para el caso del Diferencial Trasero, Mandos Final TD y Mando Final TI.

Sin embargo no podemos establecer una correlación entre la Viscosidad y las Horas de uso del Aceite:



	Horas_del_Aceite	VISCO
Horas_del_Aceite	1.000000	-0.068636
VISCO	-0.068636	1.000000

Teníamos indicaciones por parte de los expertos de una posible relación entre el Sodio (Na) y el Resultado de laboratorio que tampoco podemos establecer esta correlación:



	Na	Resultado
Na	1.000000	0.184733
Resultado	0.184733	1.000000