Comienzo modificando el feature **FUEL** de la siguiente forma:

```
df["FUEL"].replace({"gasoline": .70, "kerosene": .30, "lpg" : .90, "thinner":  .50}, inplace=True)
```

Los valores asignados corresponden al valor de octanaje de cada producto en una escala de 0 a 130, gasolina = 94-95 , queroseno = 46, gpl = 103-115, y thinner fue asignado guiandome de la “experiencia”. Los valores de **df[FUEL]** corresponden a una aproximación del valor **OCTANAJE / 130** para cada clase de gasolina.

PCA:

Index	Componentes PCA:1	Componentes PCA:2	Componentes PCA:3	Componentes PCA:4	Componentes PCA:5
SIZE	0	0	0.036413	0.036413	0.036413
FUEL	6.0979e-19	4.044e-16	0.999337	0.999337	0.999337
DISTANCE	0.997452	1.00422	1.00422	1.05874	1.10425
DESIBEL	0.0364217	0.268895	0.268895	1.18272	1.51371
AIRFLOW	0.0613384	0.102911	0.102911	0.4497	1.38471
FREQUENCY	0.000856711	0.972547	0.972547	1.17672	1.29559

Este es el aporte de cada atributo al explicar el dataset en diferentes componentes principales, vemos que SIZE no realiza aporte significativo entre **PCA:1** y **PCA:5**, pero **DESIBEL**, **FUEL** y **AIRFLOW** van dando mayor explicabilidad del dataset a medida que aumentamos el número de componentes, los resultados obtenidos usando un árbol de decisión predeterminado fue:

```
('DESIBEL', 'AIRFLOW', 'FUEL')
--- 0.008532285690307617 seconds ---
Precisión: 0.9040983606557377
Exactitud: 0.8750238868717752
Recall: 0.8400609291698401
F1 SCORE : 0.8709040663245163


Todas las columnas excepto SIZE
--- 0.014019012451171875 seconds ---
Precisión: 0.8965936739659367
Exactitud: 0.8713930823619339
Recall: 0.8410041841004184
F1 SCORE : 0.867909715407262
```

La variable en segundos corresponde al tiempo de entrenamiento de cada árbol con solo las variables mencionadas arriba del mismo. Tenemos el mismo comportamiento, en ambos casos e incluso mejor para el primer entrenamiento que solo usa **DESIBEL**, **AIRFLOW** y **FUEL**.

Usando **SIZE** él en segundo entrenamiento:

```
('DESIBEL', 'AIRFLOW', 'FUEL')
--- 0.007978916168212891 seconds ---
Precisión: 0.9018255578093306
Exactitud: 0.8763615516911906
Recall: 0.8458904109589042
F1 SCORE : 0.8729628902415088


Todas las columnas
--- 0.013963699340820312 seconds ---
Precisión: 0.9597855227882037
Exactitud: 0.9594878654691381
Recall: 0.9590508993494068
F1 SCORE : 0.9594180704441041
```

Aumenta considerablemente el valor del **F1 SCORE**

Casos especiales:

```
Todas las columnas excepto DESIBEL
--- 0.013961791992187 seconds ---
Precisión: 0.9584905660377359
Exactitud: 0.9596789604433403
Recall: 0.9617569102612646
F1 SCORE : 0.9601209601209602
```

CASO 1

```
Todas las columnas excepto FREQUENCY
--- 0.011998176574707031 seconds ---
Precisión: 0.927584919746174
Exactitud: 0.9256640550353525
Recall: 0.9272388059701493
F1 SCORE : 0.927411830565404
```

CASO 2

```
Todas las columnas
--- 0.012936830520629883 seconds ---
Precisión: 0.9673024523160763
Exactitud: 0.9656029046436079
Recall: 0.9628051142967842
F1 SCORE : 0.9650485436893204
```

CASO 3

De las imágenes anteriores concluimos que el aporte de **FREQUENCY** no es significativo e incluso en la sección de **PCA** vemos que no hubo cambio en la cantidad de información que aportaba a medida que se aumentaba la cantidad de componentes. También existe estrecha relación entre **FREQUENCY** y **DESIBEL**, para este caso **DESIBEL** aporta mucho más. La diferencia entre las métricas del **CASO 1** y **CASO 3** son mínimas y su tiempo de ejecución es similar, por lo cual descartar el atributo **FREQUENCY** sería una opción.