# Cobify Fuel Election

# 1. Summary:

Cobify has always used SP98 fuel for their cars. Recently, they started using SP95 E10, which is cheaper and more eco-friendly because it approximately contains a 10% ethanol.

However, being eco-friendly might be worse for the company and the planed since it has been observed that cars that use SP95 E10 consume more fuel.

This report aims to test this hypothesis based on a data analysis.

# 2. Input data:

Our input data consists of a .csv that contains the distance (km), the consume (L/100km), the average speed (km/h), the temperature inside the car (°C), the temperature outside (°C), weather conditions and the gas type. Also, I have added a column named external_temp, which possible values are cold or hot based on the outside and inside temperatures.

A very important reminder is that the prices of the fuel are:

**SP98 = 1.459 €/L**

**SP95 E10 = 1.379 €/L**

# 3. Observations:

For some variables, only we have not added the plots for both types of gas, this is because they both look at the average. However, all the plots can on “plots” jupyter notebook in the repository.

## 3.1. Consume/Gas type relation:

As seen in table 1, E10 consumes lightly more than SP98 on average. However, SP98 maximum is bigger than E10 even though its minimum is smaller. We can not make conclusions from this table because if we look at figure 1, we can see that in both cases there are lots of outliers influencing that average.

| | **SP98** | **E10** |
|---|---|---|
| **Mean** | 4.899 | 4.931 |
| **Max** | 12.200 | 10.800 |
| **Min** | 3.300 | 3.700 |
| **Std** | 1.118 | 0.901 |

*Table 1. Consume/gas type statistics.*

Leaving the outliers aside we can see that E10 once again consumes slightly better, but the gas type is not the only variable in this equation. Let’s see what happens in those outliers.

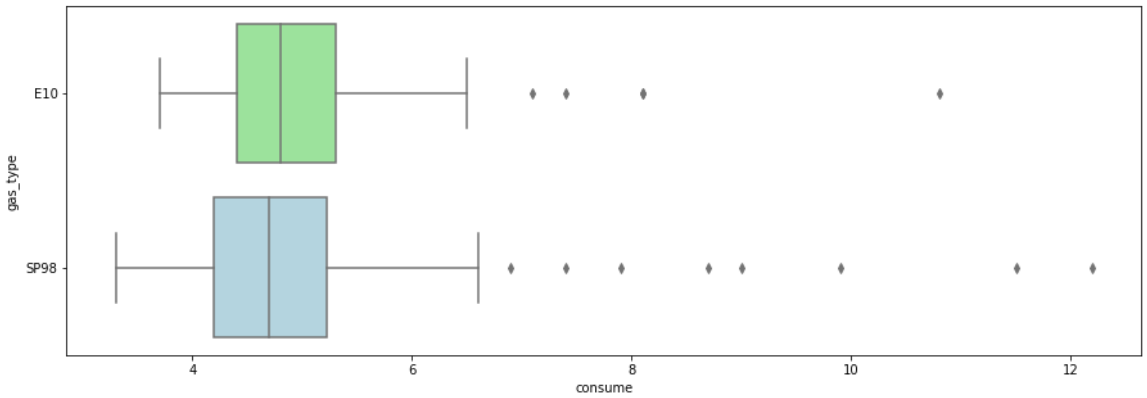


*Figure 1. Boxplot consume/gas type.*

### 3.1.1. Outliers:

In the outliers observed in figure 1, the temperature does not seem to be a determinant variable. However, we can see that consume is bigger when speed is smaller. We'll check that hypothesis later.

SP98 MAX

| | distance | consume | speed | temp_inside | temp_outside | specials | gas_type | AC | rain | sun | refill liters |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **61** | 2.1 | 12.2 | 23 | 21.5 | 9 | NaN | SP98 | 1 | 1 | 0 | NaN |

SP98 MIN

| | distance | consume | speed | temp_inside | temp_outside | specials | gas_type | AC | rain | sun | refill liters |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **44** | 5.4 | 3.3 | 32 | 21.5 | 7 | NaN | SP98 | 0 | 0 | 0 | 37.6 |

E10 MAX

| | distance | consume | speed | temp_inside | temp_outside | specials | gas_type | AC | rain | sun | refill liters |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **147** | 1.7 | 10.8 | 14 | 21.5 | 10 | rain | E10 | 0 | 1 | 0 | NaN |

E10 MIN

| | distance | consume | speed | temp_inside | temp_outside | specials | gas_type | AC | rain | sun | refill liters |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **325** | 16.6 | 3.7 | 49 | 22.0 | 17 | NaN | E10 | 0 | 0 | 0 | 37.7 |
| **337** | 16.0 | 3.7 | 45 | 22.0 | 17 | NaN | E10 | 0 | 0 | 0 | NaN |

*Figure 2. Consume/gas type outliers.*

## 3.2. Consume/speed relation:

Figure 3 confirms that gas consumption is bigger at lower speeds. However, we can see this clear trend with SP98 but it is not that clear with E10.

At higher speeds, such as in a highway it is better to use SP98, in the city it is better to use E10.

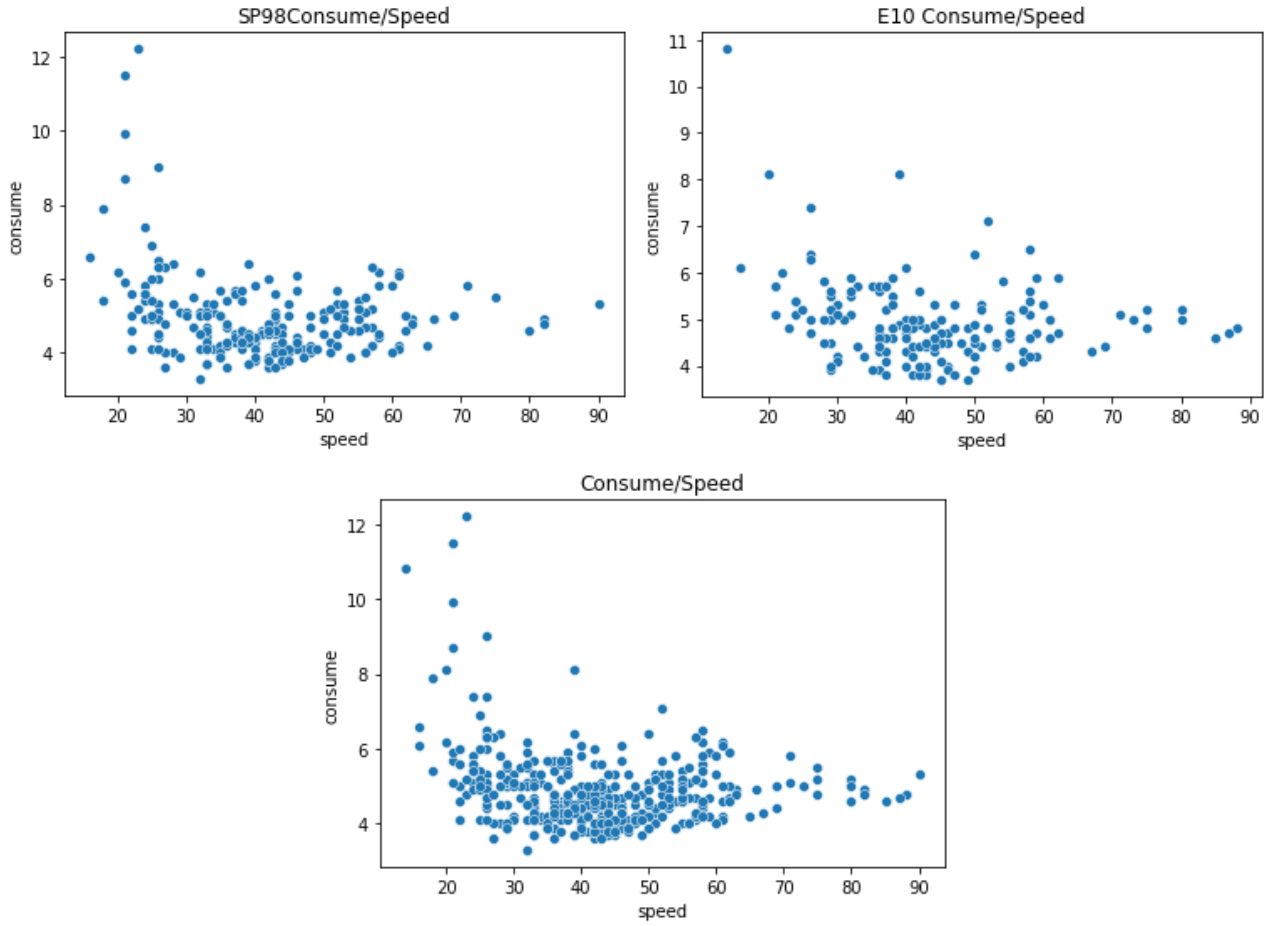


*Figure 3. Consume/speed scatter plot.*

## 3.3. Consume/Air Conditionate relation:

Consume seems to be slightly bigger when AC is ON.ac

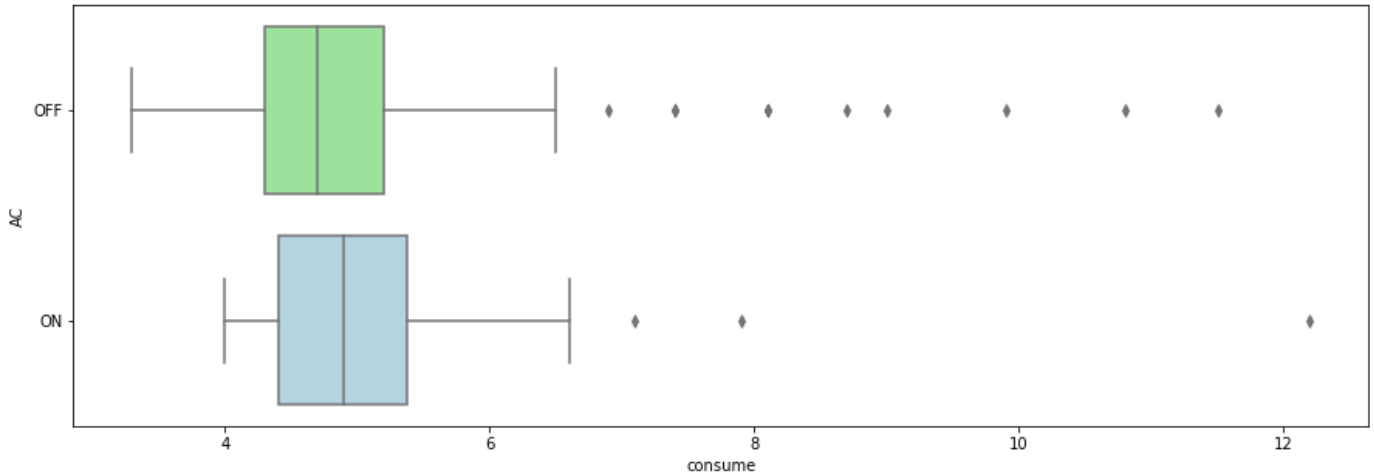


*Figure 4. Consume/AC boxplot.*

## 3.4. Consume/Weather relation:

As seen in figure 5, rainy days consume more fuel. However E10 consumes slightly more on rainy days and SP98 consumes slightly less on sunny days.

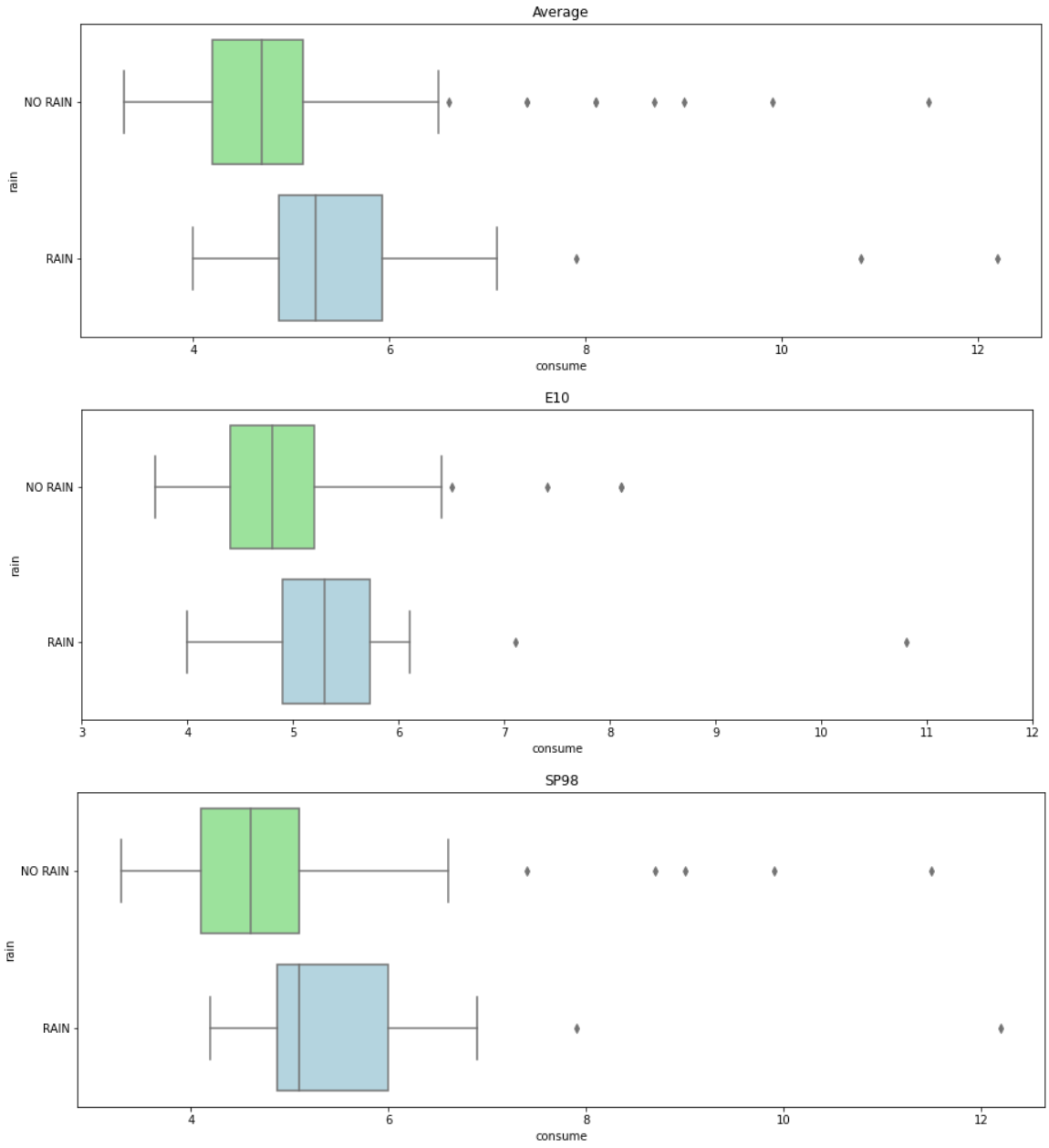


*Figure 5. Consume/Weather boxplots.*

## 3.5. Consume/distance relation:

Smaller distances mean higher consumption. This is probably because smaller gears are used.

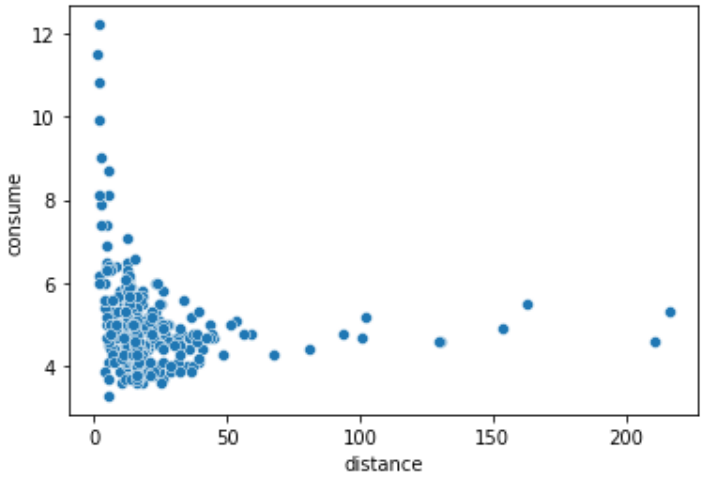


*Figure 6. Consume/distance scatterplot.*

## 3.6. Consume/External Weather relation

Cooler weather means higher consumption. However, when it is hot, E10 performs better.

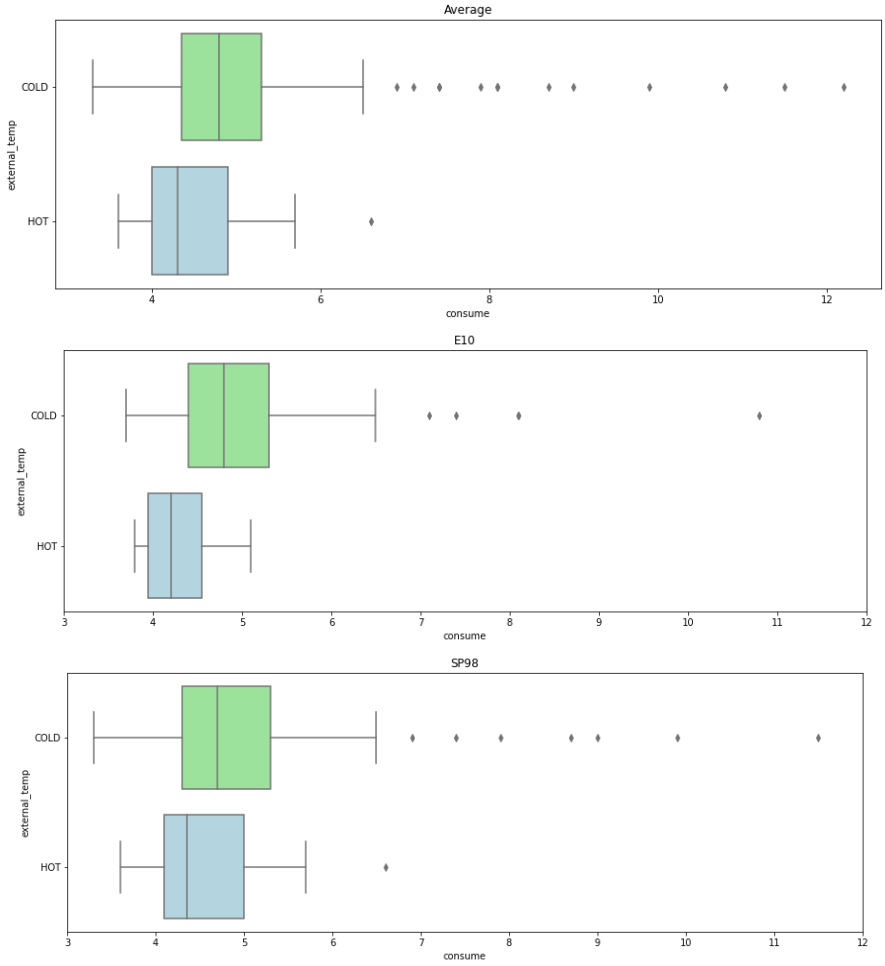


*Figure 7. Consume/External Weather boxplots.*

### 3.7. First Conclusions

In table 2 we can see how each variable affects gas consumption for each gas type. For the weather-related variables, we should have in mind that when compiling this data, SP98 was used during winter and E10 was used during spring.

| | **Conclusion** | **Relevance [1-3]** |
|---|---|---|
| **Distance** | Smaller distances → Higher consume | 3 |
| **Speed** | Highway ride → SP98<br>Cities → E10 | 3 |
| **External temp** | Cold weather → Higher consume in both gas types<br>Hot weather → E10 is better | 3 |
| **Gas type** | E10 → Higher consume | 1 |
| **Rain** | Rainy day → E10<br>Sunny → SP98 | 2 |

*Table 2. First Conclusions.*

We can say that E10:

- **Consumes more** on average.
- It is better in that its rival in the **city**.
- Better than its rival when there is **hot weather**.
- Better than its rival on **rainy days**.

We can say that SP98:

- **Consumes less** on average.
- It is better in that its rival in **highways**.
- Better than its rival on **sunny days**.

A possible solution could be having two types of cars, sending E10 cars for city rides and SP98 for long rides. This way we save fuel, money and we are being more eco-friendly in the city, where air pollution is a real problem.

## 4. Prediction:

In the case we want to know which gas type consumes more under specific conditions, we can test it by using the model contained in the "predictions" jupyter notebook.

This model has been created using H2O automl. It can predict the consume based on the distance, speed, external temperature, gas type and weather conditions of the ride. We achieved an **accuracy of 85.96%** and an **RMSE of 0.3432**.

## 5. Database and API:

All this data has been added to a MongoDB dataset as shown in the figure below.

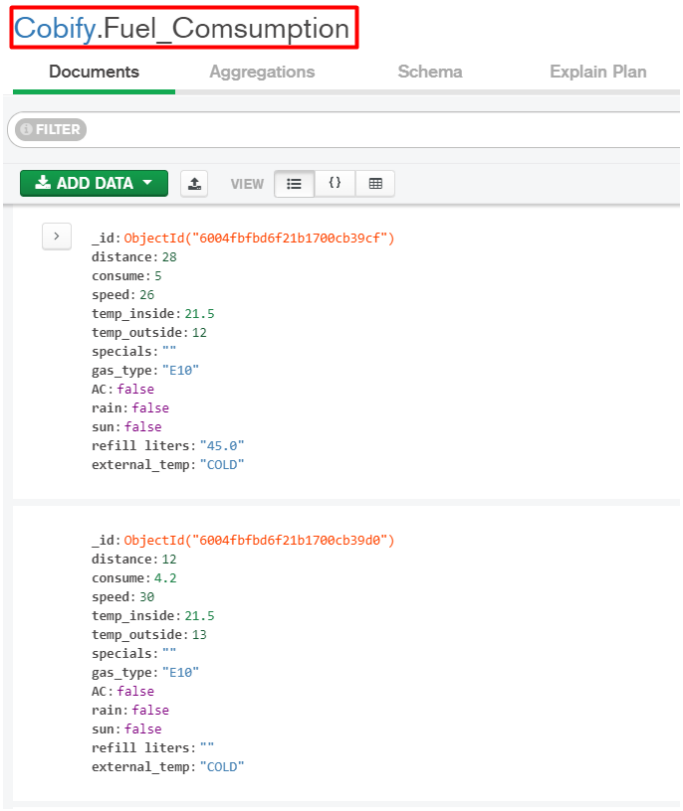


*Figure 8. MongoDB database.*