

# Reporte Proceso Pipeline

David A. Castro S.

2024-08-17

## Contenidos

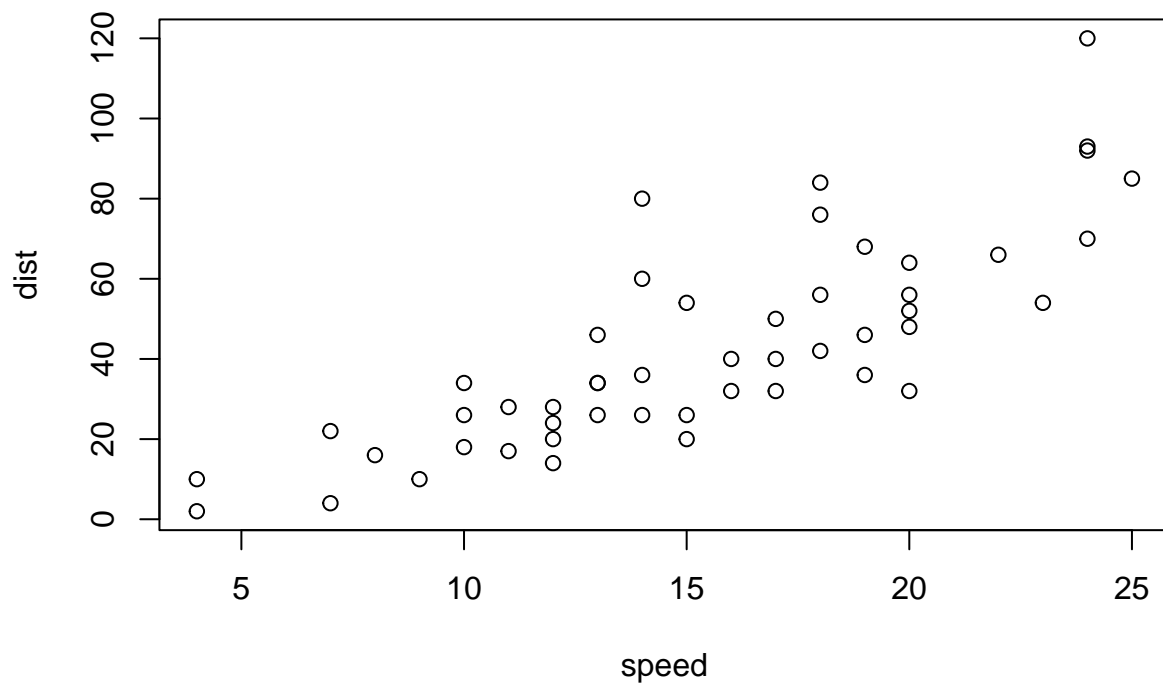
En este reporte se presenta la comparación de métodos de procesamiento **PySpark-SQL** vs **Bash** .

## Análisis de datos

```
# Código R que deseas ejecutar  
summary(cars)
```

```
##      speed      dist  
## Min.   : 4.0    Min.   : 2.00  
## 1st Qu.:12.0    1st Qu.: 26.00  
## Median :15.0    Median : 36.00  
## Mean   :15.4    Mean    : 42.98  
## 3rd Qu.:19.0    3rd Qu.: 56.00  
## Max.   :25.0    Max.    :120.00
```

```
plot(cars)
```



```
plot(mtcars$mpg, mtcars$hp,  
     main = "Relación entre MPG y HP",  
     xlab = "Millas por galón (MPG)",  
     ylab = "Caballos de fuerza (HP)")
```

**Relación entre MPG y HP**

