

# Dados de Automóveis

Dados na Ciência, Gestão e Sociedade



André Silvestre N°104532 | Diogo Catarino N°104745  
Eduardo Silva N°104943 | Francisco Gomes N°104944

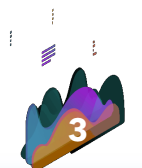




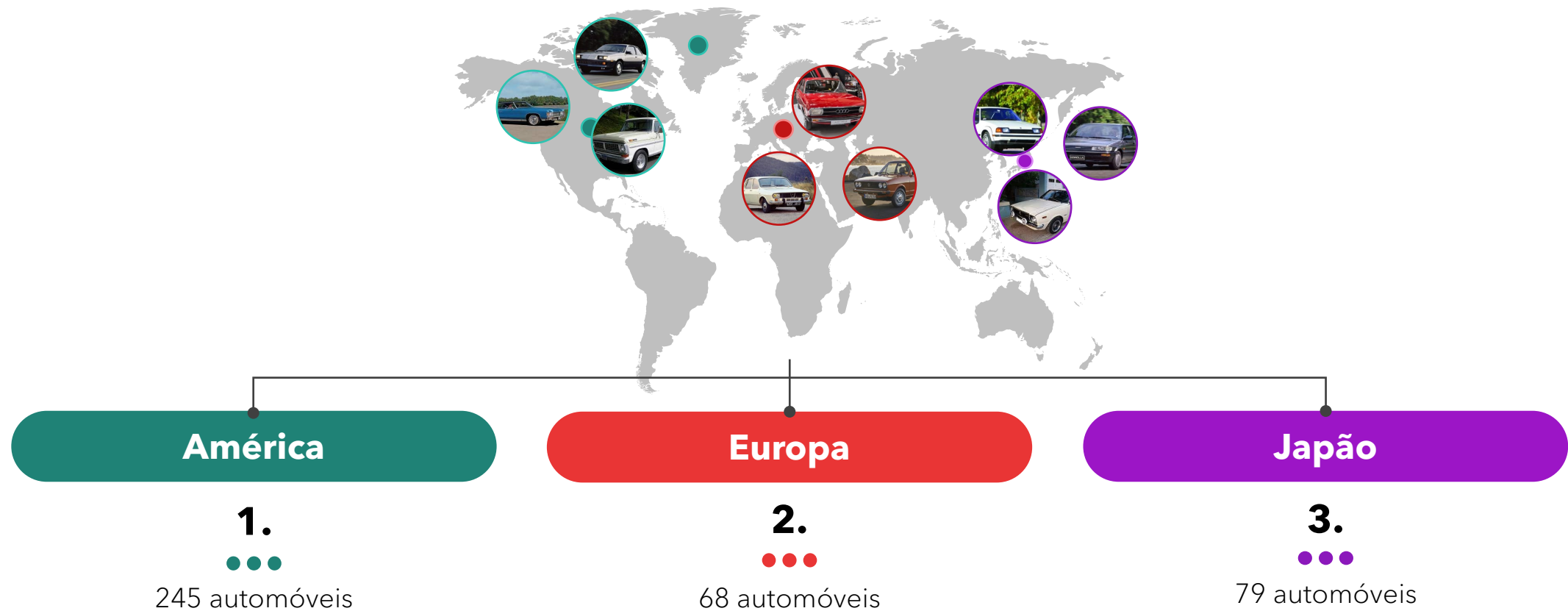
## Problema a desenvolver

Com este trabalho procuramos compreender de que modo o consumo dos automóveis, expresso em **MPG**, está relacionado com as características intrínsecas aos mesmos.

# VARIÁVEIS EM ESTUDO



# ORIGEM DOS AUTOMÓVEIS



# Análise Exploratória de Dados

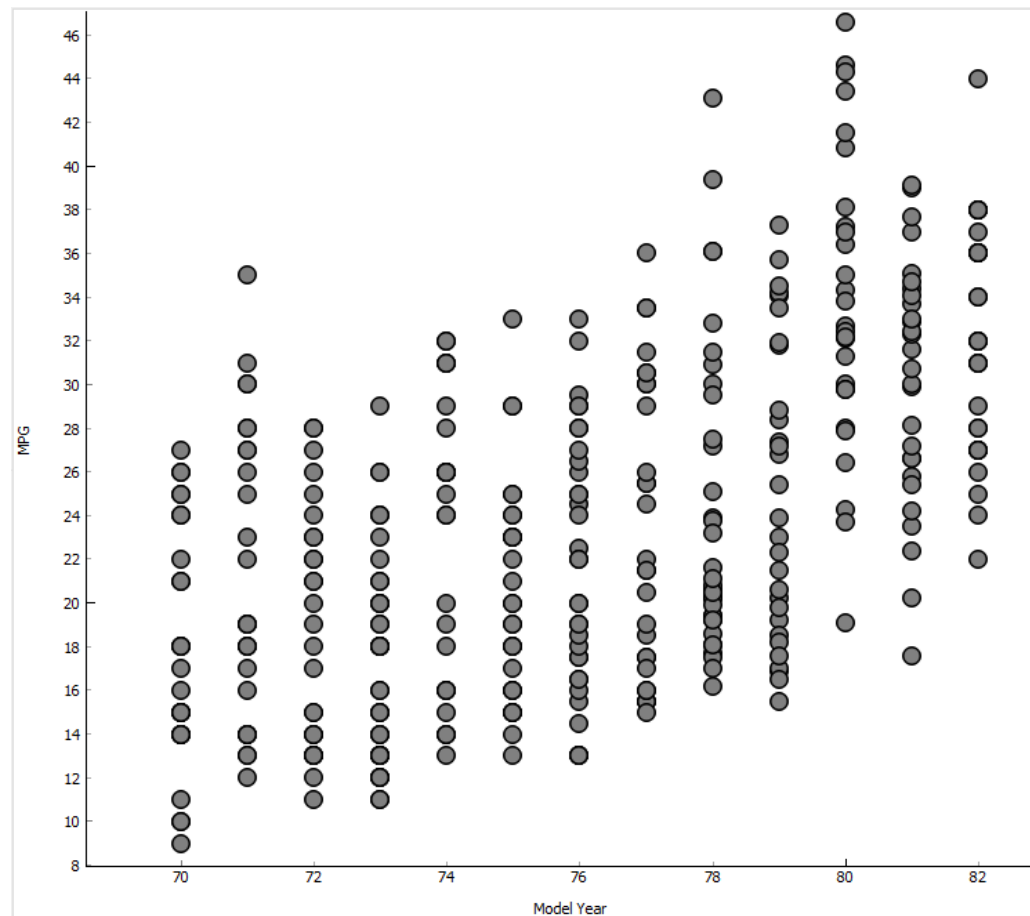


Gráfico 1 | Gráfico da evolução do *MPG* em função dos *Anos dos modelos*



# Análise Exploratória de Dados

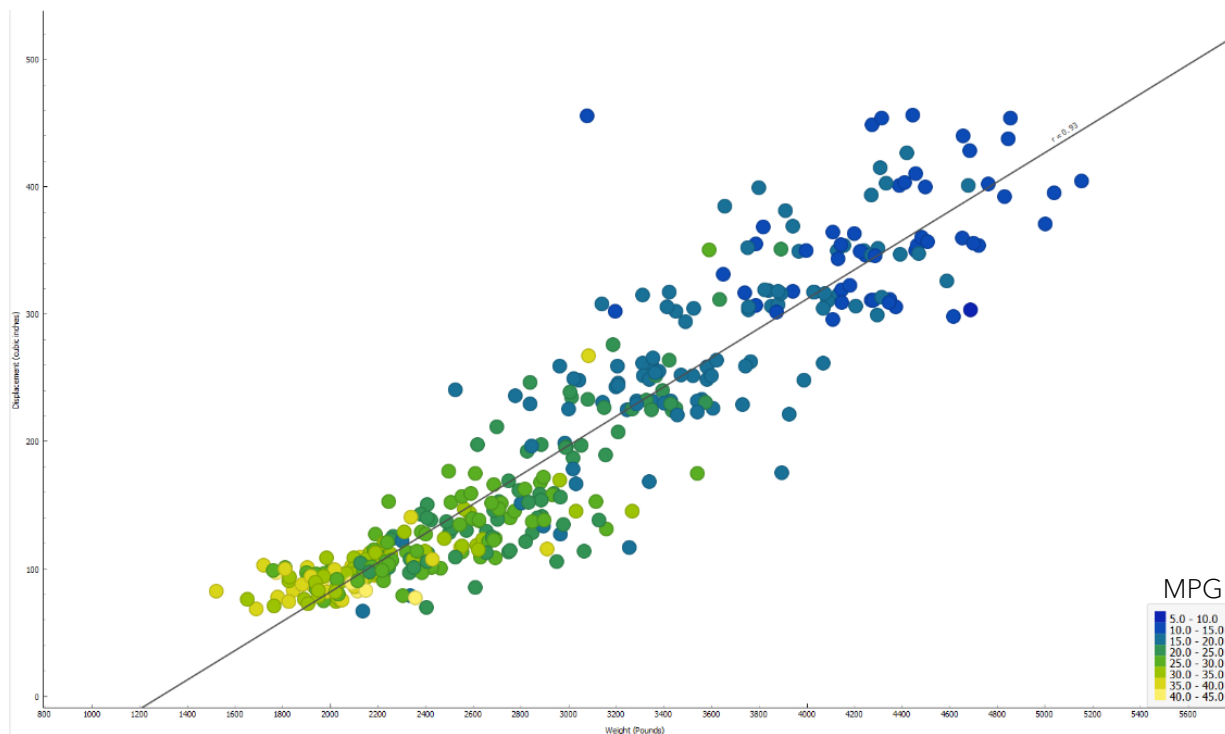


Gráfico 2 | Relação entre o *Peso*, em Kg, a *Cilindrada*, em  $inch^3$ , e o MPG.

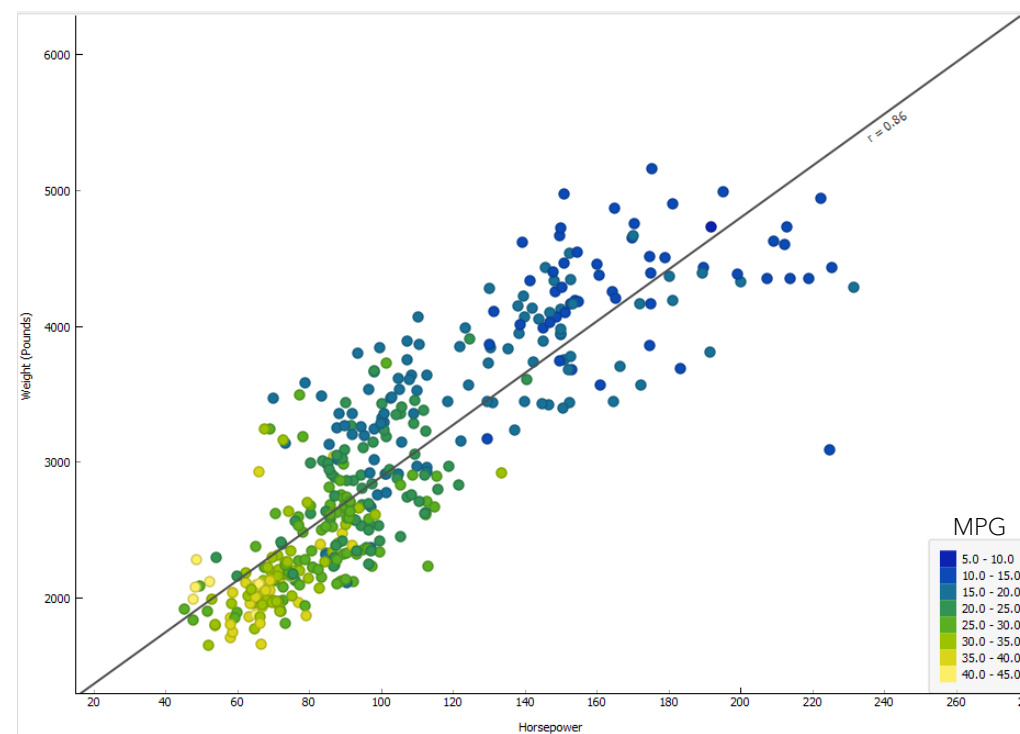


Gráfico 3 | Relação entre os *Cavalos*, o *Peso*, em *pounds*, e o MPG.

# Análise Exploratória de Dados

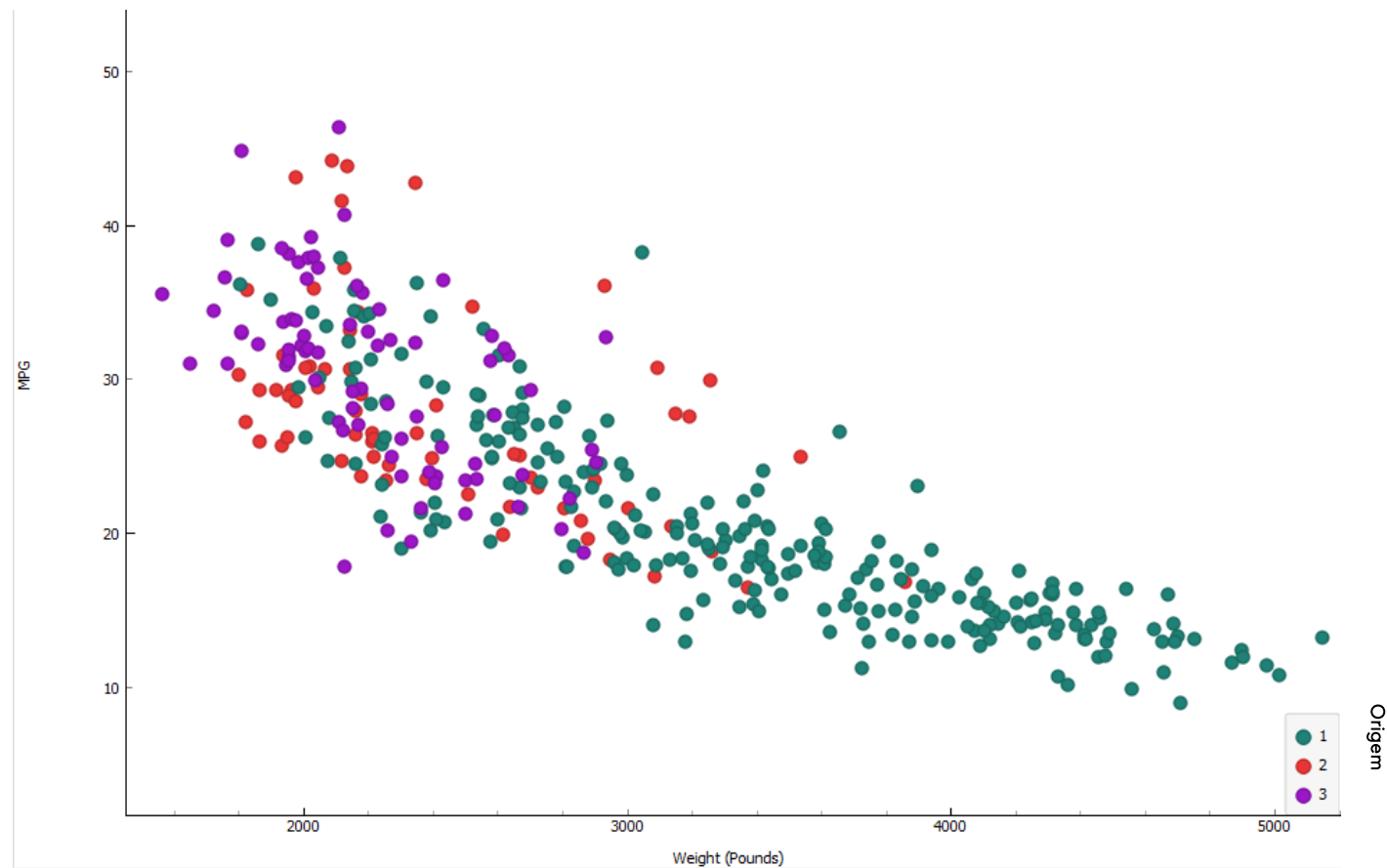


Gráfico 4 | Relação entre os *Peso*, em *pounds*, o *MPG*, e a *Origem* dos automóveis.



# Modelo

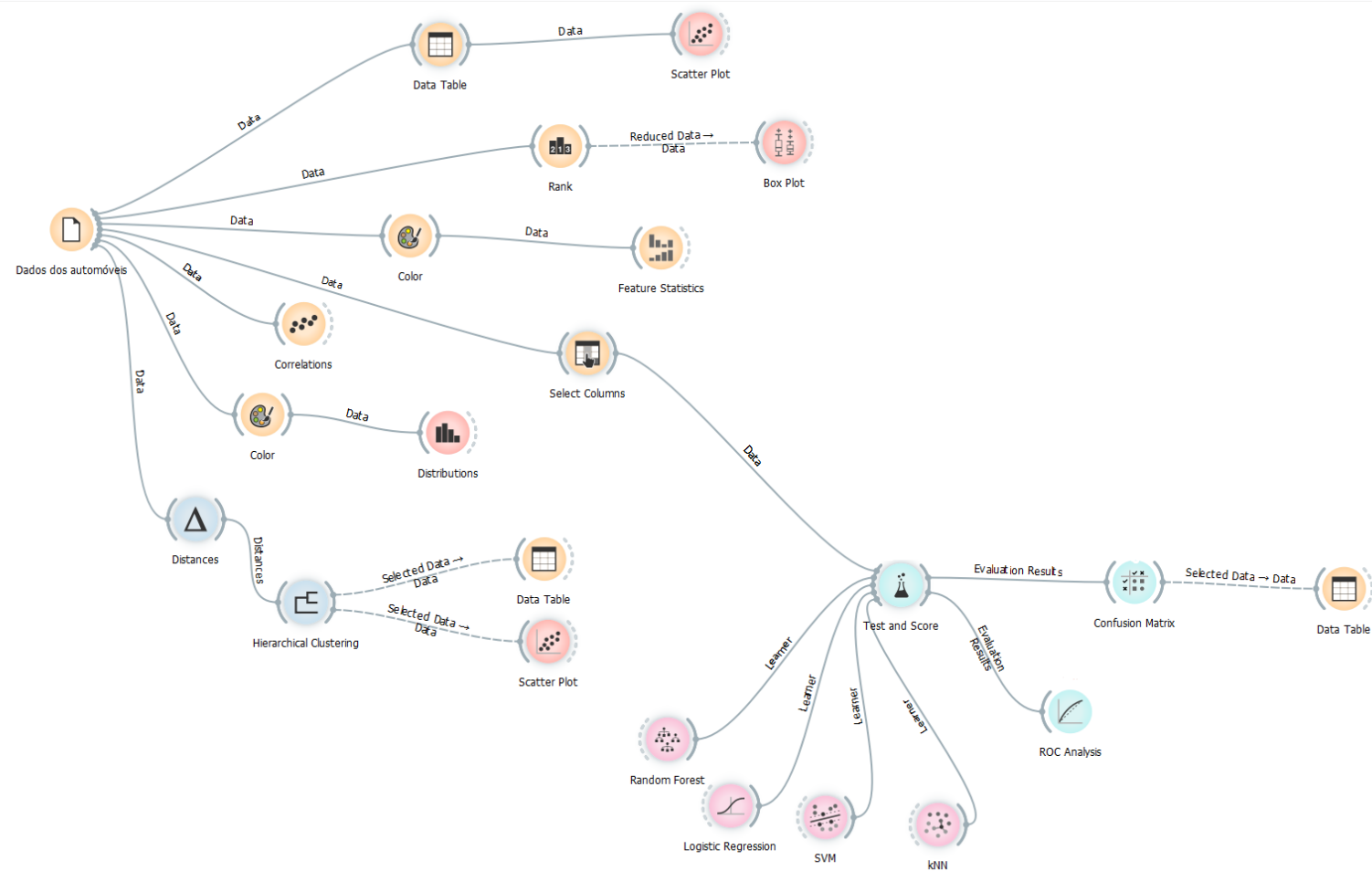
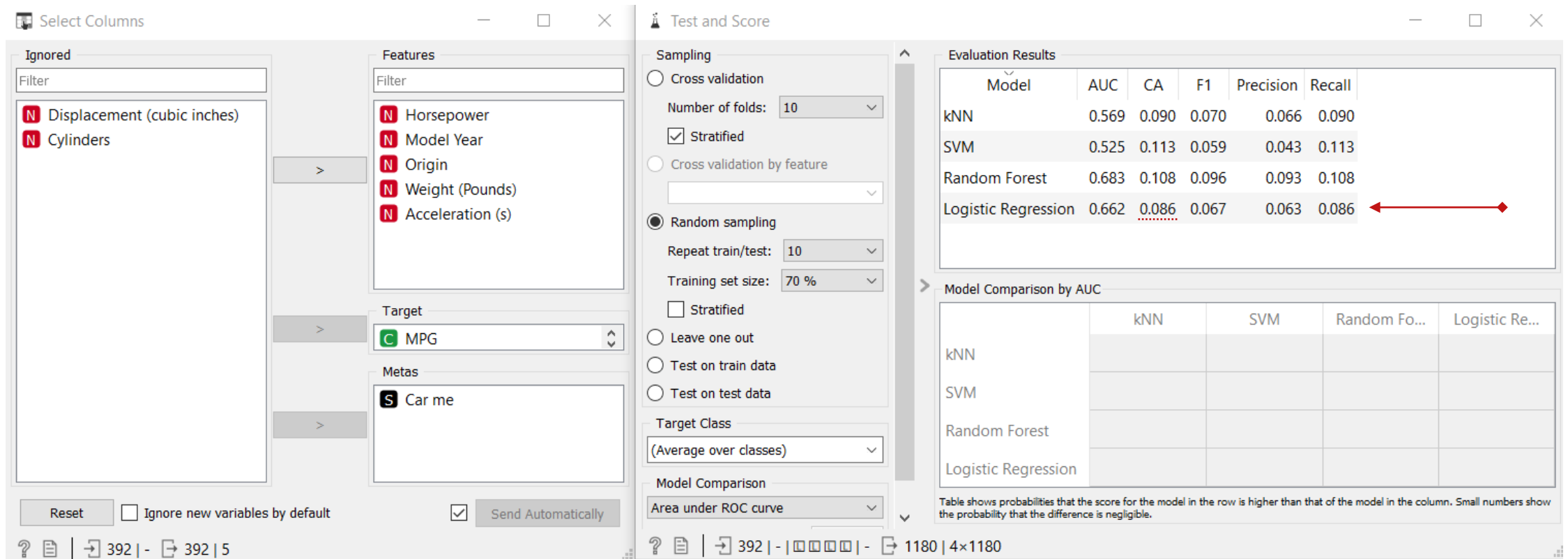


Figura 1 | Fluxo criado do *Orange*.



# Logistic Regression



The screenshot displays the Orange data mining software interface, specifically the 'Test and Score' window. The 'Features' list includes Horsepower, Model Year, Origin, Weight (Pounds), and Acceleration (s). The 'Target' is set to MPG. The 'Evaluation Results' table shows the performance of various models, with Logistic Regression having the highest AUC (0.662) and being highlighted with a red arrow. The 'Model Comparison by AUC' table shows that Logistic Regression is the best model.

Model	AUC	CA	F1	Precision	Recall
kNN	0.569	0.090	0.070	0.066	0.090
SVM	0.525	0.113	0.059	0.043	0.113
Random Forest	0.683	0.108	0.096	0.093	0.108
Logistic Regression	0.662	0.086	0.067	0.063	0.086

	kNN	SVM	Random Fo...	Logistic Re...
kNN				
SVM				
Random Forest				
Logistic Regression				

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

Figura 2 | Melhor modelo com as respectivas Colunas criado do *Orange*.



# Conclusão



Ainda que o algoritmo apresentasse baixa *Calculation Accuracy*, foi possível extrair correlações em relação ao consumo como propusemos responder.