

***INTELIGÊNCIA ARTIFICIAL***

**UA02 / LABORATÓRIO # 2**

**ASSUNTO**: EDA -EXPLORATORY DATA ANALYSIS

**Materiais de Apoio**

**Site oficial do Scikit Learn:**

**scikit-learn.org/**

***Dataset usado:*** *CarPrice\_Assignment*

**# Importação de Bibliotecas**

import pandas as pd

import seaborn as sns

**# Carga de Dados (no Kaggle)**

df = pd.read\_csv("../input/car-price-assignment/CarPrice\_Assignment.csv",sep=",")

**# Compreensão Inicial dos Dados**

df.head()

df.info()

df.isnull().sum().sum()

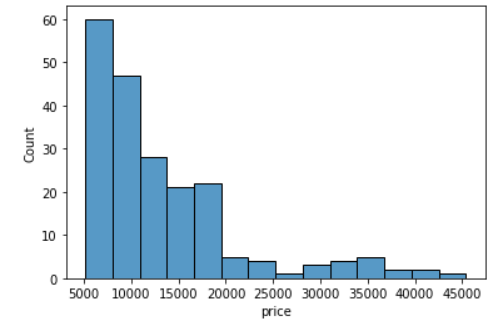
df.describe()

print(df['fuelsystem'].value\_counts())

print(df['enginetype'].value\_counts())

**# Distribuição de Variável Target**

fig = sns.histplot(data=df, x = 'price', kde=True)



**# Contagem e gráfico de Valores Categóricos**

df\_categorical = df.select\_dtypes(include = 'object').columns

for i in df\_categorical:

fig = sns.histplot(data=df, x = i)

**Explodiu, com sobreposições!**

**Como obter o gráfico abaixo?**

Alta dimensionalidade:

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Somente uma dimensão: “enginelocation”

fig = sns.histplot(data=df, x = 'enginelocation')

Gráfico, Gráfico de barras

Descrição gerada automaticamente

**# Boxplot para Verificação de Outliers**

fig = sns.boxplot(data=df, y = 'price')

**# Mapa de Calor para Correlação (HeatMap)**

sns.heatmap(df.corr())

**Como fazer um gráfico que compare tendências de dois dados numéricos com alta correlação, positiva ou negativa?**

**# Exclusão da Coluna “cad\_ID”**

df.drop('car\_ID', axis = 1, inplace = True)

**Bom Trabalho!!**