

Source : www.gia.edu

Étude du prix de diamants

La table de données « diamonds » contient pour 53.940 diamants (taille ronde) différentes caractéristiques. C'est un jeu de données classique qui se trouve dans le package ggplot2 de R.

Le nom des caractéristiques ainsi que leur signification sont indiqués ci-dessous :

carat = le poids du diamant

cut = Qualité de la coupe (Fair, Good, Very Good, Premium, Ideal)

color = couleur de D (meilleure) à J (pire)

clarity = clarté notée de IF (pur), VVS1 et VVS2 (minuscule(s) inclusion(s) très difficilement visible(s) à la loupe), VS1 et VS2 (minuscule(s) inclusion(s) difficilement visible(s) à la loupe), SI1 et SI2 (petite(s) inclusion(s) facilement visible(s) à la loupe) et I1 (inclusion(s) visible(s) à l'œil nu)

depth = pourcentage de profondeur = $z / \text{moyenne}(x,y)$ (exprimé en pourcentage)

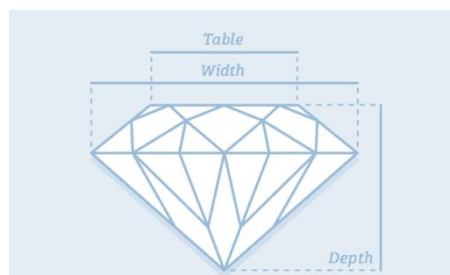
table = table / largeur (exprimé en pourcentage)

price = le prix du diamant en dollars

x = longueur en mm

y = largeur (width) en mm

z = profondeur (depth) en mm



Source : www.diamonds.pro

On souhaite établir un petit guide d'achat qui explique les variations de prix en fonction des caractéristiques. Pour cela on utilisera des représentations graphiques.

Description statistique rapide de chacune des caractéristiques une par une

Illustration(s) graphique(s) des caractéristiques, une par une

Description graphique du lien entre caractéristiques

Lien(s) entre le prix (on prendra le log du prix) et les autres caractéristiques via des graphiques.