

## Fiche de Projet Python [Analyse de données]

### I. Description du projet

**1. Titre :** Analyse d'un dataset représentant les diamants vendus par la société Brilliant Earth.

**2. Data Set :**

<https://www.kaggle.com/miguelcorraljr/brilliant-diamonds/version/1>

### 3. Description des colonnes de data set

**ID :** ID du diamant fourni par Brilliant Earth.

**URL :** url de la page contenant les détails du diamant.

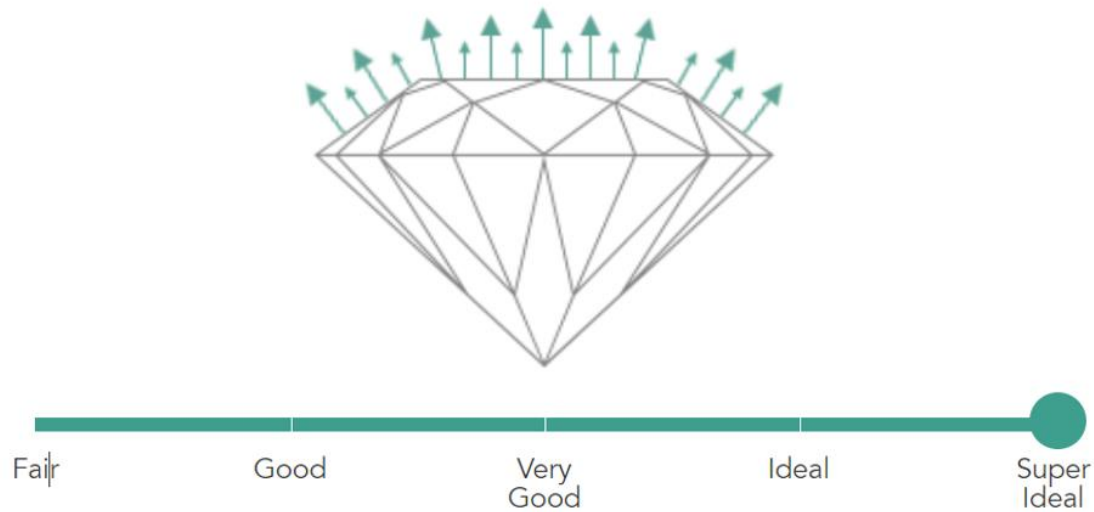
**Shape :** La forme fait référence à l'aspect extérieur général du diamant.



**Price :** Prix en dollar américain.

**Carat :** Unité de mesure utilisée pour décrire le poids d'un diamant.

**Cut :** La coupe se réfère aux angles et aux proportions d'un diamant. La taille d'un diamant, sa profondeur, sa largeur, et l'uniformité des facettes déterminent sa beauté. L'habileté avec laquelle un diamant est taillé détermine dans quelle mesure il reflète et réfracte la lumière.



### Color :

Couleur naturelle ou absence de couleur visible dans un diamant, selon l'échelle de grade GIA (Gemological Institute of America)

J : Presque incolore. Couleur légèrement détectée, une bonne valeur.

I : Presque incolore. Couleur légèrement détectée, une bonne valeur.

H : Presque incolore. Couleur perceptible par rapport aux diamants de meilleures qualités, mais offre une excellente valeur.

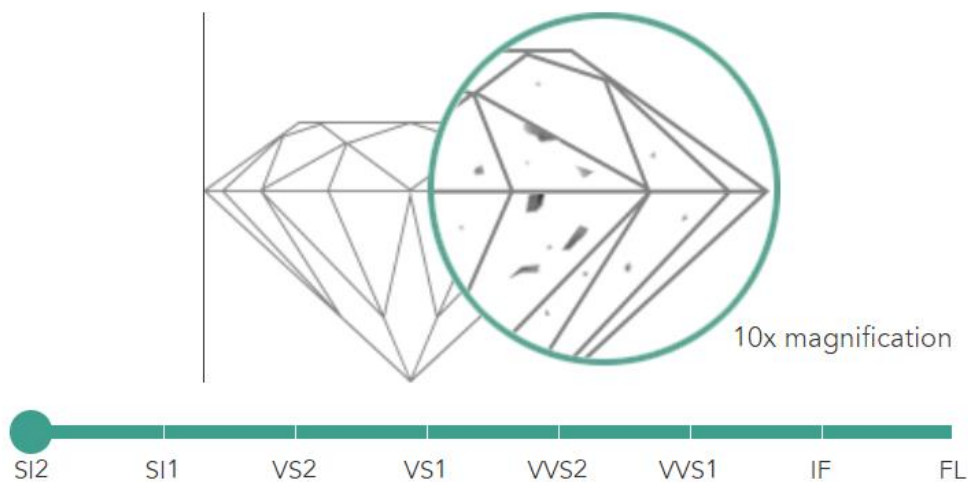
G : Presque incolore. Couleur perceptible par rapport aux diamants de meilleures qualités, mais offre une excellente valeur.

F : Incolore. Légère couleur détectée par un expert gemmologue, mais toujours considérée comme une qualité « incolore », un diamant de haute qualité.

E : Incolore. Seules de minuscules traces de couleur peuvent être détectées par un gemmologue expert, un diamant rare de haute qualité.

D : Absolument incolore ou blanc glacé. La plus haute qualité de couleur, extrêmement rare et la plus chère.

**Clarity :** La clarté d'un diamant se réfère à la présence d'impuretés sur et dans la pierre.



**Report :**

Certificat Diamant ou rapport de classement fourni par un laboratoire de gemmologie indépendant.

**Type :** Diamants naturels ou de laboratoire.

**Date\_fetched :** Les données ont été récupérées du site officiel de « Brilliant Earth » qui est une entreprise de vente de bijoux par un scraper. Cette colonne, alors, présente la date de récupération des données.

**4. Objectifs et Questions de l'analyse**

L'achat d'un diamant peut être frustrant et coûteux. L'analyse des données de l'entreprise de vente de diamants (Brilliant earth) aide les clients à connaître toutes les caractéristiques des pièces précieuses sans avoir besoin de consulter toutes les pages du site web de la société. Ainsi le choix du diamant sera plus facile. Aussi la société aura la possibilité de bien contrôler leurs articles.

**II. Travail réalisé****Lien drive du projet :**

[https://drive.google.com/file/d/1Kxn8as-EA0t6qV8J4\\_9rE1fKPGeUkSQ/view?usp=sharing](https://drive.google.com/file/d/1Kxn8as-EA0t6qV8J4_9rE1fKPGeUkSQ/view?usp=sharing)

**1. Description des méthodes d'analyse et d'apprentissage adoptées**

Supervised learning

**2. bibliothèques Python importées**

Pandas, Numpy, Matplotlib, seaborn

**3. Les modules implémentés**

matplotlib.pyplot

**Pour chaque module, il faut illustrer les différentes fonctions implémentées**

- `Plt.title()` : pour donner un titre au figure.
- `plt.tight_layout()` : pour ajuster automatiquement les paramètres de subplot pour donner un remplissage spécifié.
- `plt.grid()` : pour configurer les lignes de la grille.

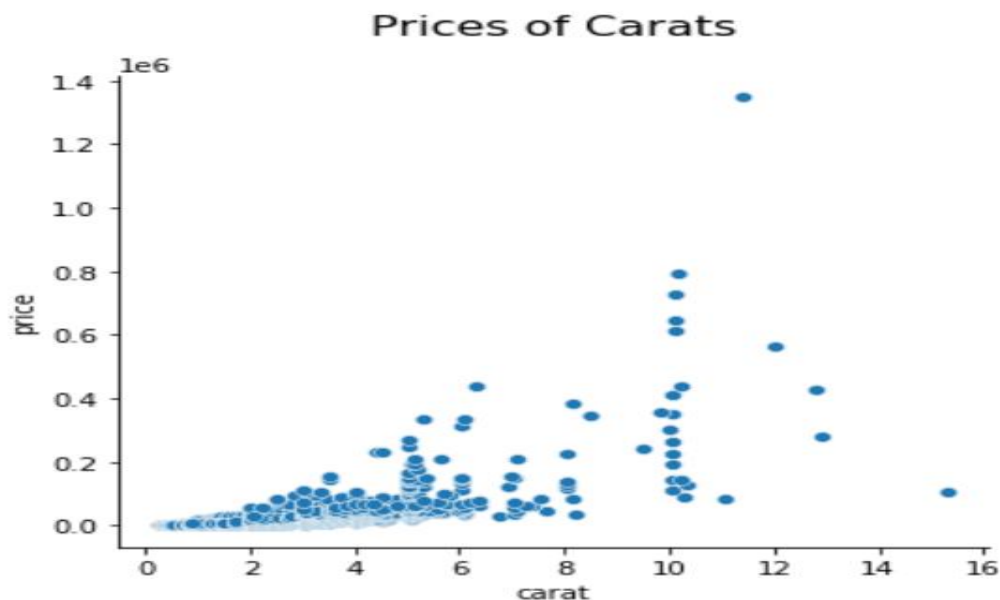
**4. Les résultats obtenus**

**Les résultats obtenus peuvent être présentés par des figures (courbes) et ils doivent être analysés.**

### Résultat 1 :

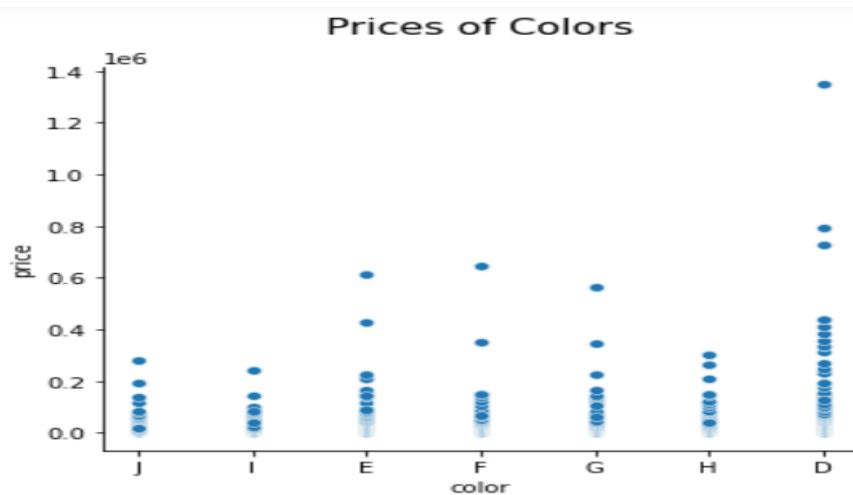
La courbe ci-dessous présente le prix des carats de diamants. Les carats sont supérieurs à 0 et inférieurs à 16. On remarque que la majorité des pièces ont un carat inférieur ou égal à 6.

Les prix sont entre 400 dollars et 1600 dollars.



### Résultat 2 :

La courbe ci-dessous présente le prix des couleurs de diamants. Selon l'échelle de grade GIA, on a 7 couleurs : J, I, E, F, G, H, D. La majorité des pièces sont de couleur D. La pièce la plus chère est de couleur D.

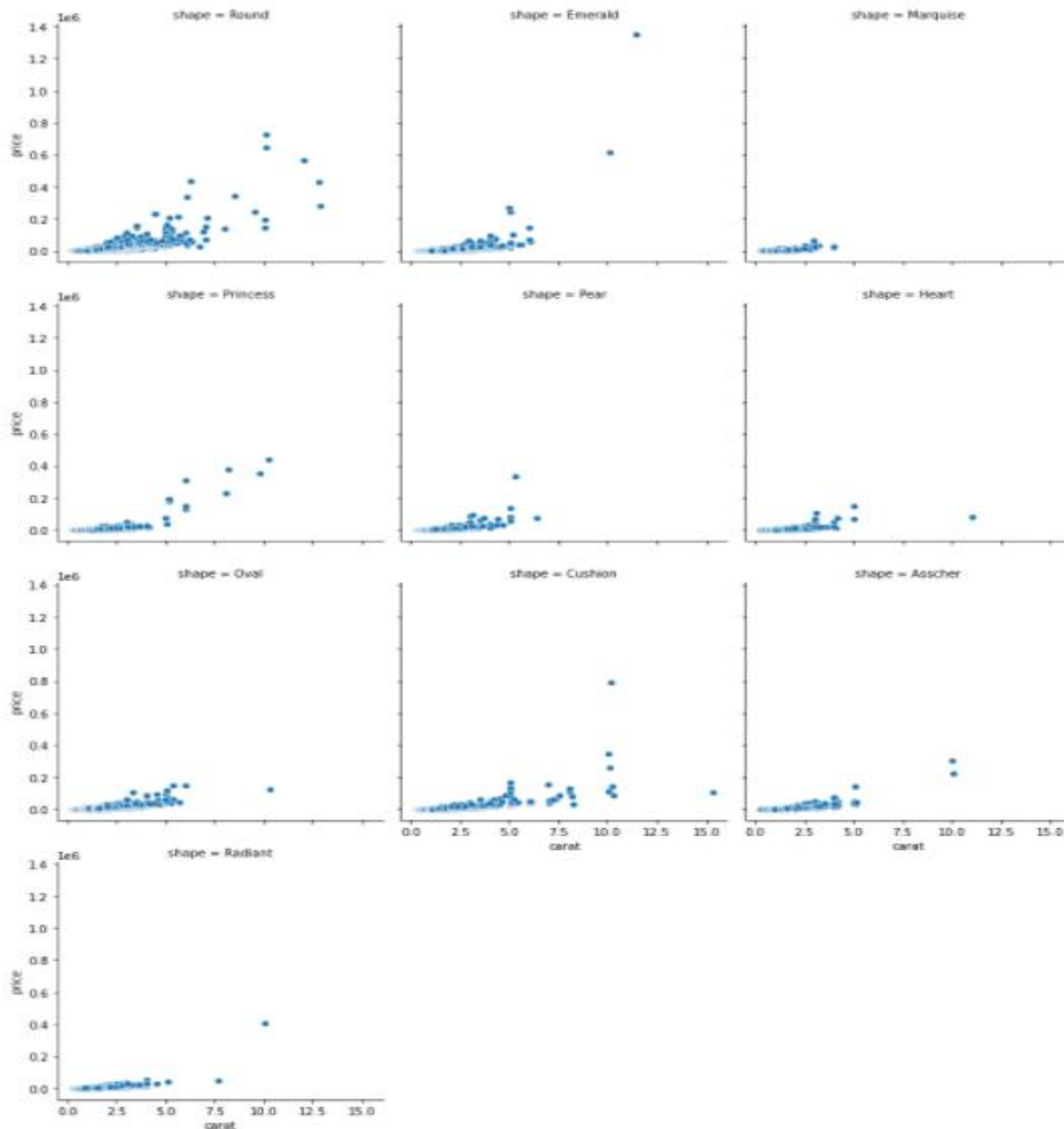


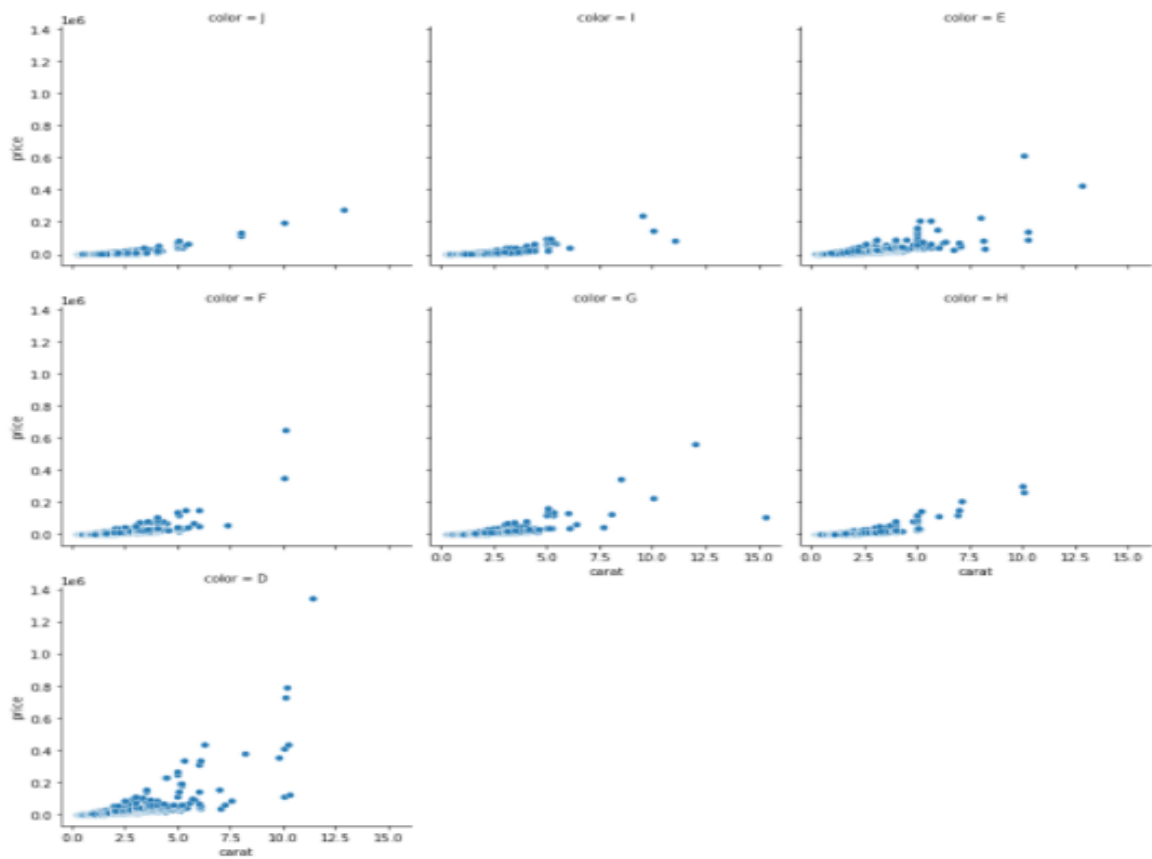
### Résultat 3 :

Les courbes ci-dessous présentent le prix des carats de chaque forme (Shape).

La plupart des pièces sont de forme : Round.

La pièce la plus chère est de forme : Emerald



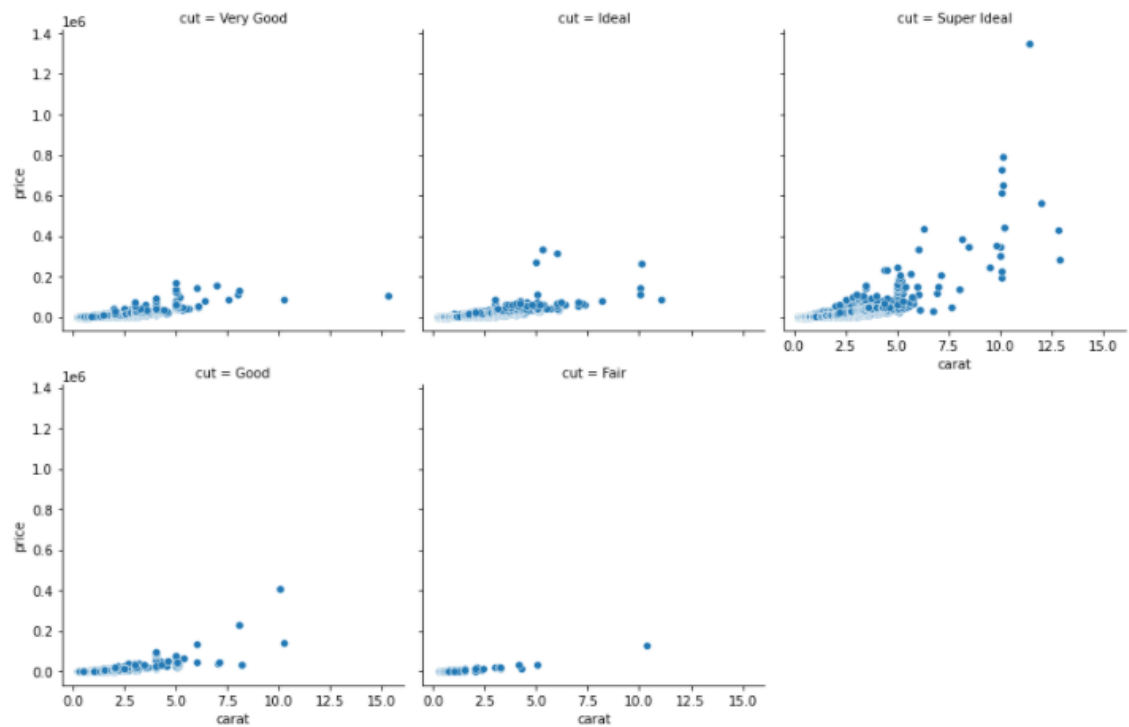


## Résultat 5 :

Les courbes ci-dessous présentent le prix des carats de chaque cut.

Une minorité des pièces sont de cut : Fair.

La pièce la plus couteuse est de cut : Super Ideal.

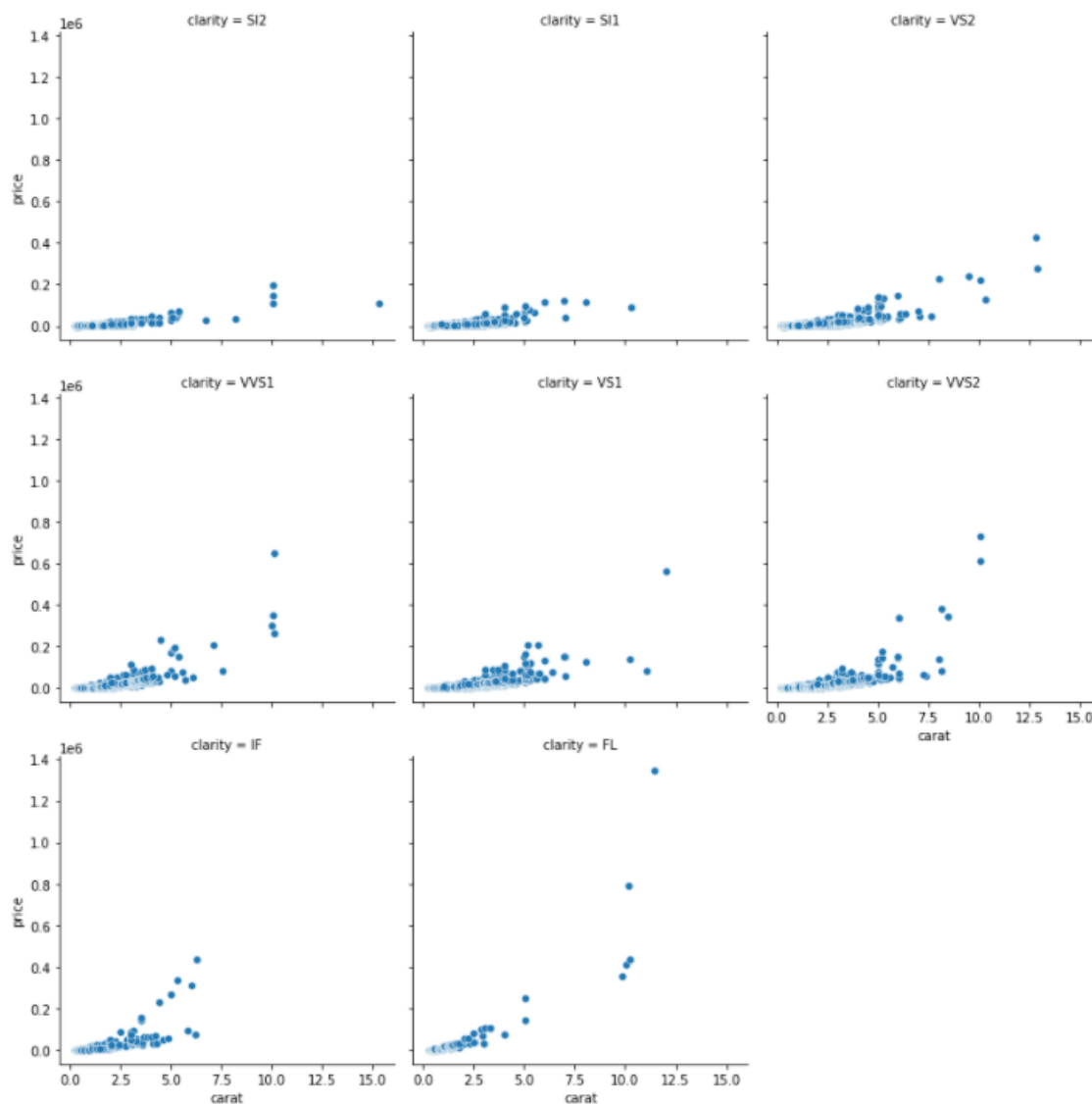


## Résultat 6 :

Les courbes ci-dessous présentent le prix des carats de chaque clarity.

Une minorité des pièces sont de clarity : SI2.

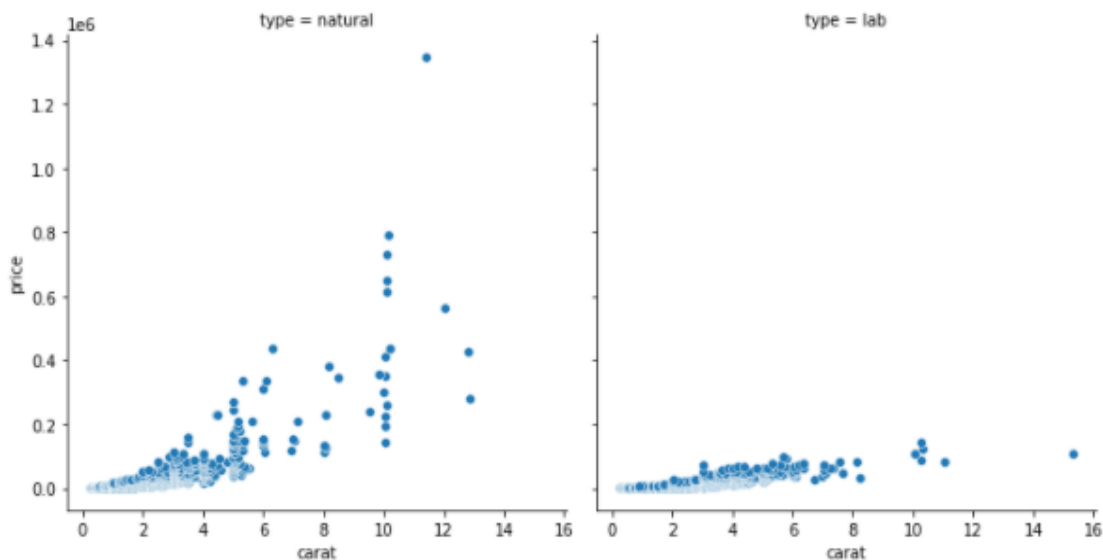
La pièce la plus couteuse est de clarity : FL.



### Résultat 7 :

Les courbes ci-dessous présentent le prix des carats de chaque type (naturel ou de laboratoire).

La plupart des pièces sont naturelles. Elles sont plus coûteuses que celles fabriquées en laboratoire.



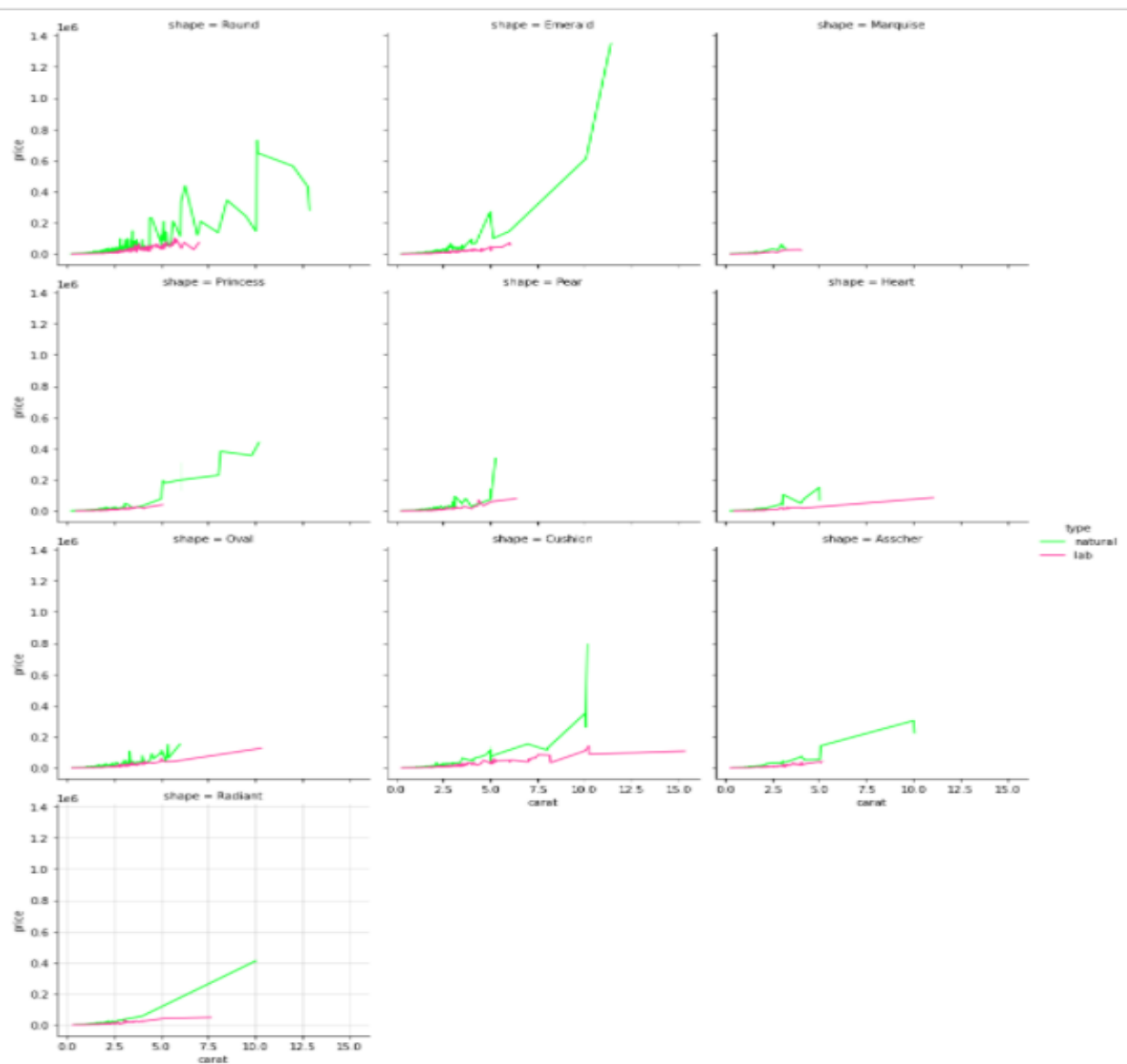
### Résultat 8 :

Les courbes ci-dessous font la différence entre le prix des carats de toutes les formes des pièces naturels et de laboratoire.

Les courbes vertes présentent les pièces naturelles. Et les courbes roses présentent les pièces de laboratoires.

Pour toutes les formes, les pièces naturelles sont plus nombreuses et plus coûteuses.



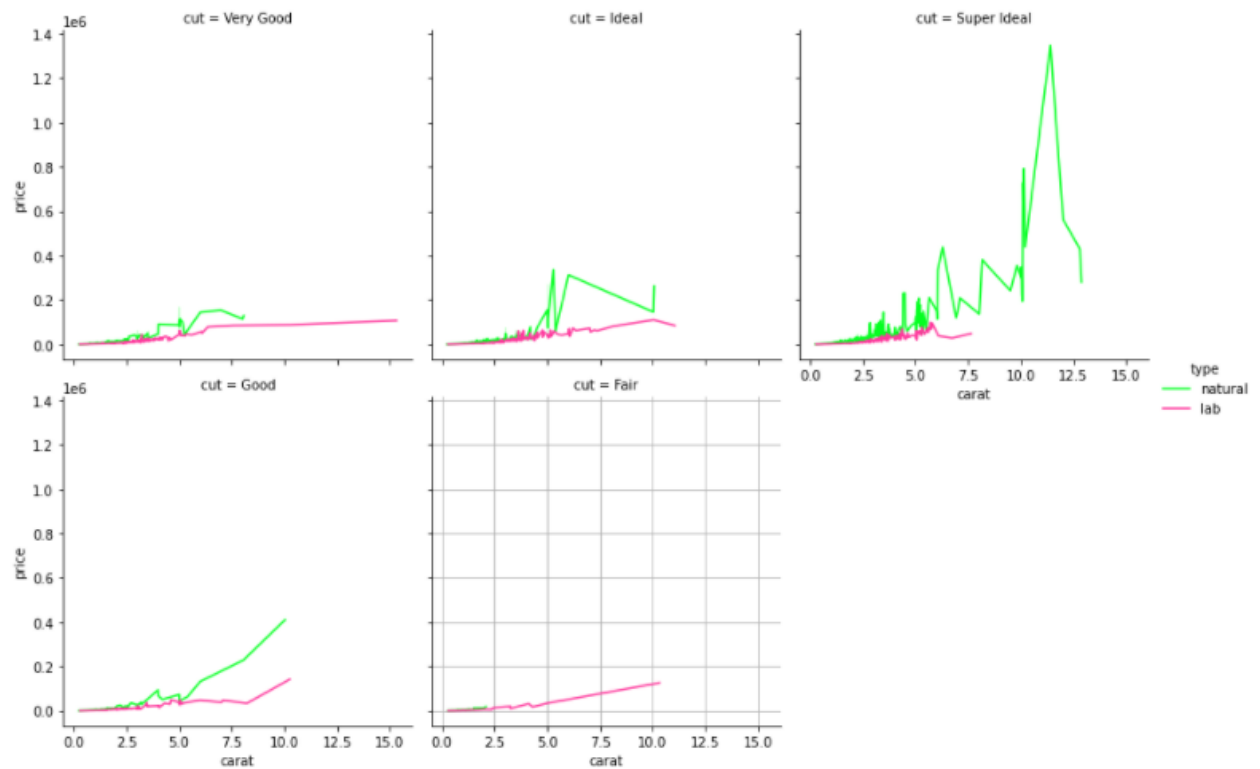


## Résultat 9 :

Les courbes ci-dessous font la différence entre le prix des carats de toutes les cuts des pièces naturels et de laboratoire.

Les courbes vertes présentent les pièces naturelles. Et les courbes roses présentent les pièces de laboratoires.

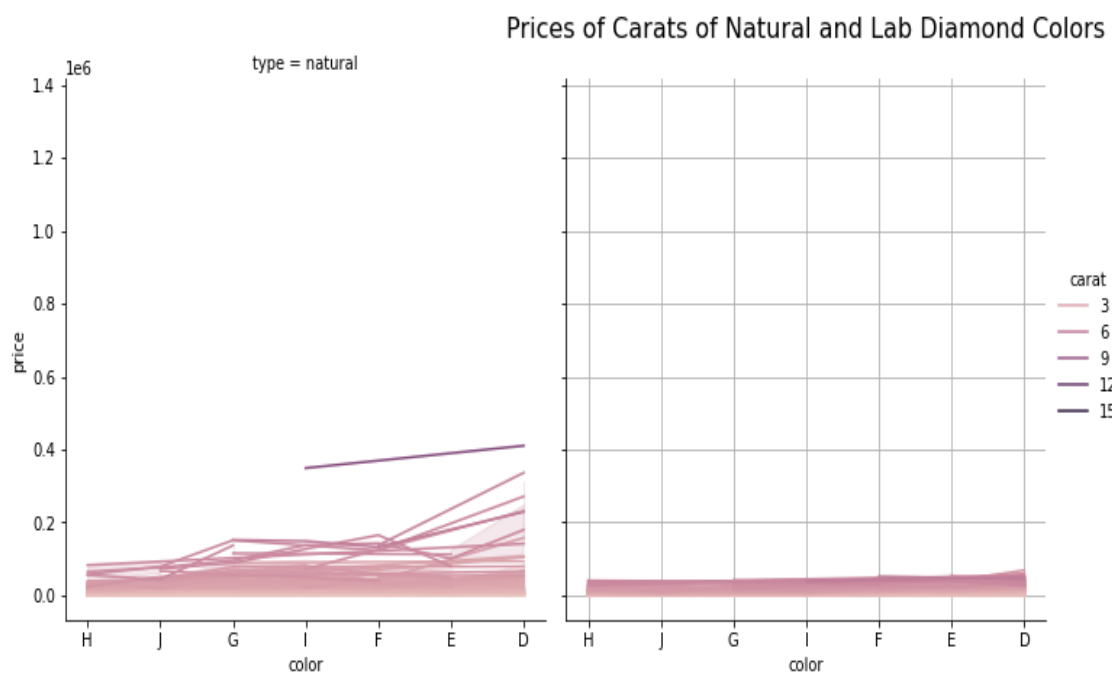
Pour toutes les cuts, les pièces naturelles sont plus nombreuses et plus coûteuses.



## Résultat 10 :

Les courbes ci-dessous font la différence entre le prix des couleurs des pièces naturels et de laboratoire.

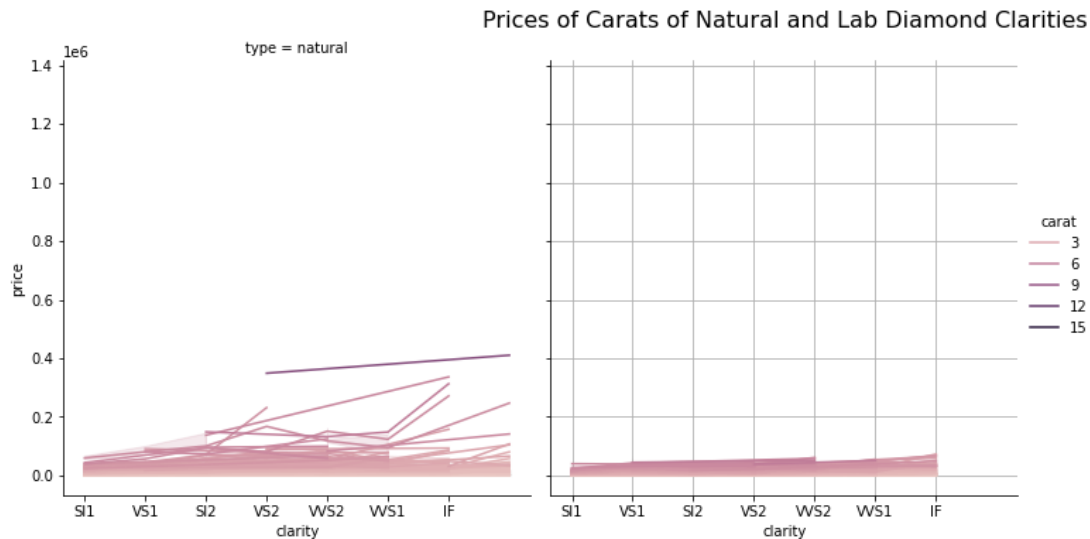
Pour toutes les couleurs, les pièces naturelles sont plus nombreuses et plus coûteuses.



### Résultat 11 :

Les courbes ci-dessous font la différence entre le prix des clarity des pièces naturels et de laboratoire.

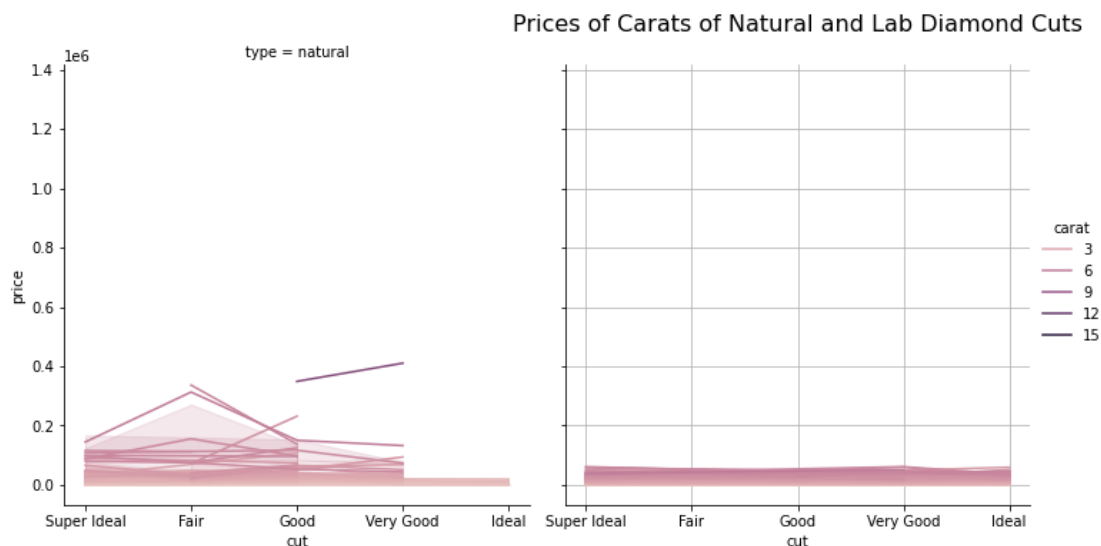
Pour toutes les clarity, les pièces naturelles sont plus nombreuses et plus coûteuses.



### Résultat 12 :

Les courbes ci-dessous font la différence entre le prix des cuts des pièces naturels et de laboratoire.

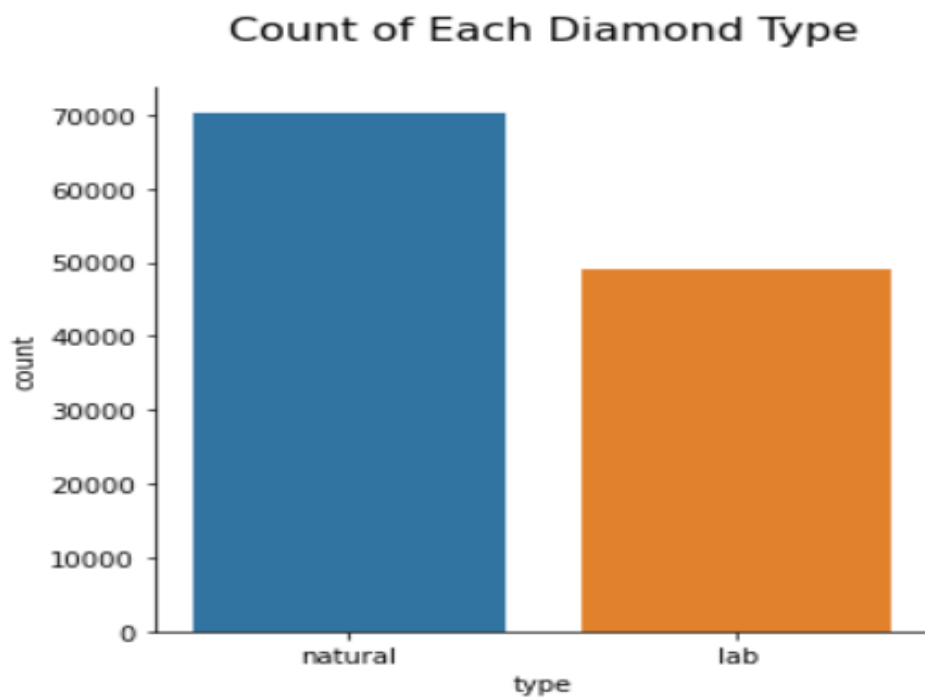
Pour toutes les cuts, les pièces naturelles sont plus nombreuses et plus coûteuses.



### Résultat 13 :

La courbe ci-dessous présente le nombre des diamants naturels et de laboratoire.

On a 70000 pièces naturelles et 50000 pièces de laboratoires.



#### Résultat 14 :

La courbe ci-dessous présente le nombre des diamants ayant comme Report : GIA, HRD, IGI, GCAL.

Les pièces de Report GIA sont plus nombreuses que celles de IGI.

Les pièces de Report IGI sont plus nombreuses que celles de GCAL.

La minorité des pièces sont de Report HRD.

