

Cahier des charges

Solution d'étude et de prédiction des prix de pierres précieuses

Atelier Benoit

2025-2026, version : 1_12/01/2026

1. Table des matières

1. Table des matières	2
2. Contexte et Objectifs	3
I. Problématique	3
II. Objectifs	3
A. Étude de prix de pierres précieuses	3
B. Modélisation Prédictive	3
C. Interface	4
3. Périmètre	4
III. Pierres à surveiller	4
IV. Sources de données (Cibles)	4
D. Diamants (principalement)	4
E. Pierres de couleur	5
V. Données à récupérer	5
4. Contraintes Techniques et Juridiques	6
VI. Techniques :	6
F. Langage et Outils	6
G. Hébergement	6
VII. Juridiques :	6
H. Respect du robots.txt	6
I. Charge serveur	6
J. Usage des données	6
5. Stockage et Restitution des Données	8
VIII. Format de sortie	8
IX. Outil de prédiction des prix	8
X. Visualisation	8
6. Planning et Livrables	9
XI. Planning	9
XII. Livrables attendus	9

2. Contexte et Objectifs

I. Problématique

Comment passer de données brutes à une prédition tarifaire ?

Le marché des pierres précieuses est un marché volatile et assez complexe. Beaucoup de paramètres rentrent en jeu lors d'un état des lieux des prix (par exemple, les 4C pour les diamants).

Ainsi, ce projet de web scraping cherche à concevoir un système d'aide à la décision. La collecte de données sera automatisée grâce à des outils de web scraping afin de constituer une base de données fiable, indispensable à l'élaboration d'un modèle algorithmique de prédition des prix.

II. Objectifs

L'objectif du projet est de développer une solution logicielle capable d'automatiser la collecte de données sur un panel de 10 à 15 sources web, de consolider ces informations dans une base de données structurée.

Celle-ci permettra non seulement d'avoir un état des lieux fiable de la situation actuelle, mais aussi d'implémenter un algorithme prédictif de ces prix.

Cet outil devra fournir une estimation de prix (fourchette) fiable pour une pierre donnée, assortie d'un score de confiance.

Tout ceci sera également accompagné d'une IHM permettant la visualisation de toutes ces données.

A. Étude de prix de pierres précieuses

- Identification des sources : Définir une liste de 10 à 15 sources de données sur internet
- Créer de la base de données de prix avec les données nécessaires pour chaque type de pierre.
- Nettoyage et Consolidation des données : Normalisation des données, traitement des doublons et des valeurs aberrantes, mise en place d'un outil pour visualiser les données.

B. Modélisation Prédictive

- Développement d'un modèle de prédition capable d'estimer les prix de marché. Le modèle doit fournir une fourchette de prix, pour refléter la volatilité des prix du marché.
- Validation : Générer un score de confiance associé à chaque donnée.

C. Interface

- Ajouter une IHM (Interface Homme Machine).

3. Périmètre

III. Pierres à surveiller

- Diamants
- Rubis
- Saphirs
- Émeraudes
- Spinelles
- Grenats
- Tourmalines
- Aigue-Marines
- Zircons
- Tanzanites

IV. Sources de données (Cibles)

D. Diamants (principalement)

- **Blue Nile** (bluenile.com) | Référence mondiale, énorme base de données
- **Nivoda** (nivoda.com) | Achat auprès de plusieurs fournisseurs du monde entier, nécessite une connexion
- **James Allen** (jamesallen.com) | Très riche en données.
- **77 Diamonds** (77diamonds.com) | Acteur européen basé à Londres / Berlin, prix potentiellement plus pertinents pour le marché européen.

E. Pierres de couleur

- **GemSelect** (gemselect.com) | Gros catalogue avec des fiches techniques détaillées.
- **BitoneStone** (bitonestone.com)
- **Gem-plus** (gems-plus.com)

- **Frediani** (frediani.fr)
- **Gemfrance** (gemfrance.com)
- **Voillot** (maximevoillotgems.com)

V. Données à récupérer

- Type de pierre
- Forme
- Poids (carats)
- Taille
- Couleur
- Pureté
- Traitement
- Origine
- Coupe
- Certificat
- Image
- Prix
- Date de relevé du prix

4. Contraintes Techniques et Juridiques

VI. Techniques :

F. Langage et Outils

Utilisation de python 3.14 comme langage de scripting, FireFox avec extension Web Scraper, Pandas (pour agencer les données dans des tableaux).

Stockage : base de données (MySQL).

G. Hébergement

Les scripts et données seront hébergées sur un drive (ex : Google Drive).

VII. Juridiques :

H. Respect du robots.txt

Le script doit vérifier les droits d'accès avant chaque lancement sur la page robots.txt afin d'assurer du respect des conditions d'utilisation.

I. Charge serveur

Ne pas surcharger les serveurs cibles : par exemple, ne faire tourner le script que lors des heures creuses si nécessaire, ne pas envoyer trop de requête aux sites.

J. Usage des données

Ne pas utiliser les données à des fins commerciales directes ou publiques :

Usage strictement interne (Back-office) : Les données collectées serviront uniquement d'outil d'aide à la décision (définition de la stratégie d'achat/vente). Elles ne devront jamais être exposées publiquement.

Interdiction de revente de la base de données : Le fichier de données généré est la propriété exclusive de l'entreprise dans le cadre de sa veille. Il est formellement interdit de vendre, louer ou transférer ce fichier à des tiers (autres bijoutiers, courtiers en données).

Pas d'automatisation tarifaire directe ("Dynamic Pricing" aveugle) : Le script ne doit pas modifier automatiquement les prix de vente de la bijouterie en se basant sur les concurrents sans validation humaine. L'outil est consultatif, pas décisionnel (pour éviter des erreurs de prix catastrophiques dues à un mauvais scraping).

Respect de la Propriété Intellectuelle (Contenu éditorial) :

Seules les **données factuelles** (Prix, Carat, Poids, Certification) sont extraites.
Il est interdit de récupérer et de réutiliser les créations intellectuelles des sites cibles : descriptions produits rédigées ("copywriting"), ou logos des concurrents.

5. Stockage et Restitution des Données

VIII. Format de sortie

Insertion des données dans une base de données MySql.

IX. Outil de prédition des prix

Développement d'un algorithme permettant d'estimer une fourchette de prix (min/max) pour une pierre donnée en comparant ses caractéristiques (poids, pureté, couleur) avec l'historique de la base de données, assorti d'un score de confiance sur la fiabilité du résultat.

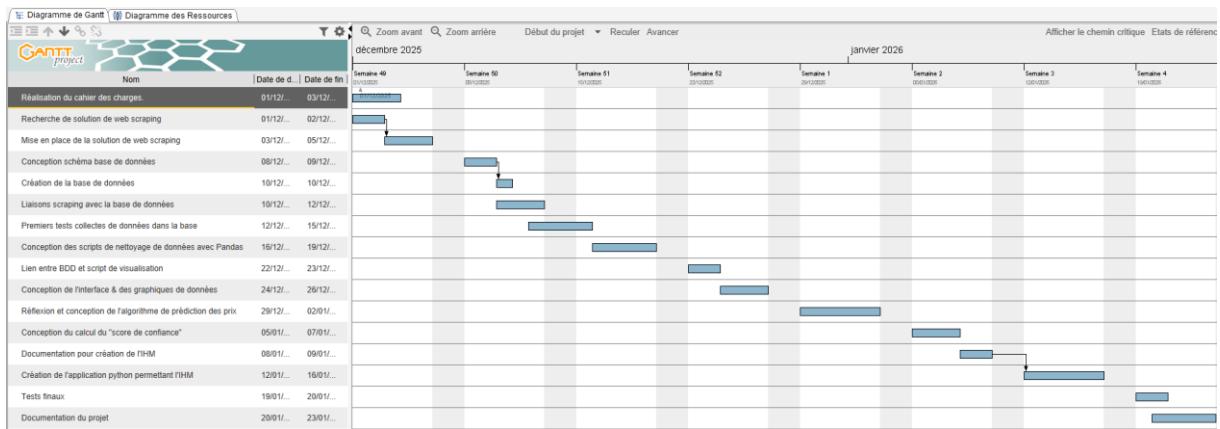
X. Visualisation

Création d'une application locale en python permettant la visualisation des données présentes dans la base.

Création d'une interface pour visualiser la prédition avec un certains nombres de critères renseignés.

6. Planning et Livrables

XI. Planning



XII. Livrables attendus

- Les scripts Python commenté (prédiction, extraction de données, visualisation de données...)
- Le fichier requirements.txt
- Un fichier Excel, CSV ou un dump d'une base de données généré contenant un échantillon récupéré lors du scraping.
- Une documentation technique d'installation, et de modifications.