

Zachary Allen
961 WGU Street, Suite T2
College, UT, 27182

Ruby C. Hartcut
Head of Product Development for Gemstones
FBay Retail
123 Website Street
Arpa, NY, 31415

Dear Mrs. Ruby Hartcut,

Attached is a proposal to automatically produce pricing estimates for diamonds listed on FBay. My team has prepared this document over the last week, testing the basic premise against our records of past sales. We carry a combined 50 years of experience in developing production-ready predictive models for corporations. The proposed estimation technique gives excellent accuracy, scales well, and runs on efficient, mass-produced datacenter hardware, keeping projected development costs at \$2,250.

Preliminary market surveys suggest these estimates are trusted by consumers, and increase their loyalty to our site, brand, and partnered retailers.

We expect this proposal, if implemented, will lead FBay to rapidly become one of the top 5 gemstone marketplaces in the next year.

If you have any questions about this proposal, please contact me at +0(123)456-7890.

Sincerely,

Zach Allen
FBay Product Designer

Zachary Allen

SID 001091806
C961 T2

In The Loupe

2022 August 18th

loupe

noun; (OED) a small magnifying glass used by jewelers and watchmakers.

OVERVIEW

Consumer confidence in the fairness of luxury pricing and sourcing is at an all time high, but price-shopping for gemstones remains a rare endeavor among new 18-25s. FBay can lead the online diamond market by keeping customers In The Loupe, providing accurate estimates of fair pricing for diamonds on our site. **Evaluator: The hypothesis of the project is that if diamonds are priced based on their measurable features (what can be seen 'In The Loupe'), then those prices can be predicted efficiently and accurately from only those measurable features.**

GOALS

- Develop and verify a diamond price prediction model with 95%+ accuracy and real-time responsiveness.
- Lay the groundwork for future value, including:
 - Increased customer satisfaction, confidence, and loyalty.
 - Positive media attention towards FBay and our innovation.
 - Increased revenue in one of our largest-margin product categories.

SPECIFICATIONS

Based on preliminary studies, a Decision Tree model will be fitted to our diamond price data we store on [kaggle.com](https://www.kaggle.com). This price data describes many of the features our customers care about, such as cut, color, quality, and carat, enabling us to produce price predictions integrated with existing search query tools.

These price predictions must be completed with >95% accuracy. In order to support the live query features we offer our F+ Loyalty Program customers, the model must handle 99.9% of queries in 50ms or less, and be able to serve >50,000 queries per second 99.9% of the time.

PROJECT METHODOLOGY AND PLAN

Evaluator: This section describes both the methodologies and plan I will follow to develop the data product.

This project's planned implementation will follow the DIKW (Data, Information, Knowledge, Wisdom) process. The DIKW process focuses on creating consistent, reusable value at every stage of a data pipeline's development, by methodically transforming Data into Information, then to Knowledge, and finally to Wisdom. The DIKW process is a mainstay of my team, and has helped us get the most out of our 50 years experience developing these pipelines for business needs.

1. Data
 - a. Diamond sale data will be loaded into a Python Jupyter Notebook.
2. Information
 - a. Data will be cleaned based on having zero or empty values.
 - b. Outliers will be removed if they are far outside the range of other values.
 - c. Data will be explored to identify artifacts, trends, biases, and distributions.
 - i. If any are found which remove the data's integrity, return to step 1.
3. Knowledge
 - a. Correlations between features will be investigated.
 - b. Additional inspection will be performed to ensure data has been cleaned and organized correctly.
 - i. If any issues arise, return to step 2.
 - c. A Decision Tree will be fit to the data, using the SKLearn Python library.
4. Wisdom
 - a. The model will be used to make predictions about diamond price.
 - b. The model will be evaluated for accuracy, throughput, and latency.
 - c. These measures will be used to inform whether 'In The Loupe' phase 2 should proceed.
 - d. This model's predictions could be used by customers to make decisions about whether to buy a diamond at a particular price.

MILESTONES

Data Preparation

(1-3h)

Investigate trends, outliers, and artifacts in the data. Clean, format, and organize the data to best support the price prediction model.

Data Analysis

(2h)

Investigate correlations between features. Identify the features which are necessary or redundant for prediction, based on correlation. Finalize the features required for the prediction model.

Model Fit and Evaluation

(1h)

Fit a Decision Tree model to the price data. Evaluate the model for accuracy, throughput, and latency. Demonstrate the model interactively.

Report Development and Documentation

(4h)

Develop reports describing the performance of the model and recommendations for future development. Develop documentation for the data pipeline, model evaluation, and how to access the model for further development and testing.

FUNDING

Based on the milestone estimates, the total development cost is estimated at \$2,250 in salary + overhead plus \$0 in compute.

DATA ACCESS AND USAGE

Our sales data is publicly available, which permits us to set internal guidelines for security and privacy. To protect any intellectual property developed during this phase of the project, all development will be done on company-approved hardware and software. Evaluator: The nature of this data does not require any privacy or ethical considerations.