E-Commerce Price Prophet

Capstone Final Presentation
Austin Chen, Daniel Myers, Kevin Cahillane, Rithvik Sunku, Saurabh Narain
Our Team

- Austin: Subject Matter Expert
- Daniel: Data Engineer
- Kevin: Project Manager
- Rithvik: Infrastructure Engineer
- Saurabh: Dashboard Developer
Shouldn't e-commerce be easier?

WHY?
- Pricing products is time-consuming
- Platforms prioritize profits over fairness

WHAT?
- Buyers face confusion due to price variations
- An unbiased tool could improve fairness
$3 trillion

Estimated revenue from all global e-commerce transactions in 2023

($22.68 trillion)
($1.85 trillion)
“Mercari often recommends a price range for a different product than what I intended to sell”

“The Mercari pricing tool provides lower than expected price ranges”
What does the ideal experience look like for users?

Based on the information you've provided, we believe your product should be priced at around $52.65 with a lower range from $22.11 to an upper range of $75.82.
Buyers and sellers benefit from transparent pricing

**Target User**
- Casual consumer
- Bargain hunter
- Experienced buyer
- First time seller
- Vintage item vendor
- Small business owner

**Trust:** Transparent pricing builds trust, fostering confidence in transactions and e-commerce platforms

**Market Efficiency:** Clear and consistent pricing reduces confusion, streamlining the market for everyone

**Fair Competition:** Sellers at all experience levels can compete more fairly, offering better prices and quality, benefiting consumers
### Our data

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Mercari Price Prediction (Primary)</th>
<th>Amazon Review (Supplementary)</th>
<th>eBay current listings (Secondary)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observations</strong></td>
<td>More than 700k</td>
<td>240 million in original dataset (100k most recent listings utilized)</td>
<td>120k+ observations recorded based on trained cluster characteristics</td>
</tr>
<tr>
<td><strong>Fields</strong></td>
<td>Item title, brand, category, condition, final sales price</td>
<td>Item title, brand, category, condition, final sales price</td>
<td>Item title, brand, category, condition, current listing price</td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td>Kaggle</td>
<td>UCSD data repository</td>
<td>eBay Finding API</td>
</tr>
<tr>
<td><strong>Limitations:</strong></td>
<td>Historical listings collected from 2018-2019</td>
<td>Most recent listings recorded in 2018</td>
<td>Current listings; no final sales price</td>
</tr>
</tbody>
</table>
Data Architecture Diagram

Source data
- Mercari
- Amazon

Model training
- Mixedbread.ai
- embedding

Prediction tuning
- Price range prediction
- Output sanitization
- Ebay Real-time and CPI data

Final Price prediction & evaluation

Data processing and quality check
User input
Output sanitization
Model Pipeline Overview

1. Load Raw Data
2. Create Word Embeddings from Text Features
3. Run Mini-Batch K-Means Clustering
4. Calculate Percentile Ranges on Clusters
5. Multiplier fine tuning with real-time EBay data
6. Output Prediction
Minibatch KMeans algorithm on training data, use cluster assignment to have 30% - 70% quantile ranges for each cluster

Convert item name and description into vectors using MixedBread AI
Multiplier Model for Fine-tuning Price Ranges

Item-specific tuning

Cluster-wide adjustment

The specific item price is fine-tuned by other item characteristics (quality, shipping)

CPI and Ebay data move the lower and upper ranges

This method allows us to create a unique pricing analysis per item
MVP Live Demo

A Mercari-based pricing dashboard that compares similar postings and recommends a price range for sellers.
Results and Evaluation

We used root mean square logarithmic error (RMSLE) as our error metric in evaluating model performance and improvement.

In the end, we achieved an RMSLE of 0.42 (cross-validated).

\[
RMSLE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\log(y_i + 1) - \log(\hat{y}_i + 1))^2}
\]

New Error Function to Account for Skewed Distribution

Transforming Our Data to Account for Heavy Right Skew

Ebay Data Vectorized by Our Model Projected in 3D Space
Benchmarking Against Mercari’s Suggested Prices

To further validate and compare our model, we took an unseen sample from our data and used Mercari’s suggested pricing tool and our tool.
Unbiased and tighter price ranges
Greater transparency for users displaying similar and current listings

More user-friendly yet detailed interface

Price Prediction Range
Based on the information you've provided, we believe your product should be priced at around $44.77 with a lower range from $32.64 to an upper range of $58.97.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force One</td>
<td>The shoes have only been worn twice. Not really my style Size 9 in women's I'm willing to negotiate.</td>
<td>70</td>
</tr>
<tr>
<td>Nike Air Force 1s</td>
<td>It is in good condition. No box</td>
<td>36</td>
</tr>
<tr>
<td>Nike Air Force 1s</td>
<td>Good condition. Weren't worn that much Sz 6.5 in woman's</td>
<td>21</td>
</tr>
<tr>
<td>NIKE SUEDE AIR FORCE ONE</td>
<td>Worn only once. Selling bc I got the wrong size. Imk if u need more detail</td>
<td>108</td>
</tr>
<tr>
<td>Nike Air</td>
<td>Nike Air Force 1 shoes brand new with tag but these had some stains on the inside on the one shoe when I got them so I decided I didn't want them. Shown in last two pics the stain.</td>
<td>80</td>
</tr>
</tbody>
</table>
Future Developments

- Expanding into additional item categories
- Collecting live data to adjust to changes in the market
- More modalities (e.g. images, videos)

![CPI Values Chart](chart.png)
Mission Statement

To optimize e-commerce efficiency and promote price transparency for all.