1. What is the source of your dataset?

(Provide website/API/source link)

- ⇨ We used web scraping to scrap real time data form amazon
  link : https://www.amazon.in/

2. Why did you choose this dataset for your problem statement?

(Relevance and clarity)

- ⇨ Amazon operates one of the world's largest third-party marketplaces. This environment is the primary ecosystem for Grey Market activity, where authorized goods are sold outside of intended distribution channels. The dataset provides a realistic representation of price variations and unauthorized seller behavior necessary to train a detection model.

- ⇨ Illegal or grey market listings often display specific anomalies in metadata. Amazon datasets typically include:
- ⇨ Seller Information: Allows for the identification of unverified or high-volume suspicious entities.
- ⇨ Pricing History: Essential for detecting "price dumping," a hallmark of grey market sales.
- ⇨ Product Descriptions & Images: Provides the raw data needed for NLP and Computer Vision models to identify counterfeit indicators or tampered packaging.

- ⇨ 3. Diversity of Product Categories
- ⇨ Grey market activity is not uniform; it varies significantly between electronics, luxury goods, and pharmaceuticals. Amazon’s broad catalog allows the model to be trained across multiple categories, ensuring the AI solution is robust and scalable rather than overfitted to a single industry.

- ⇨ 4. Real-World Complexity
- ⇨ Detecting illegal sales requires distinguishing between legitimate discounts and illicit activity. Amazon data includes "noisy" variables like seasonal sales and lightning deals, forcing the AI to develop higher precision in identifying true negatives (legitimate sales) versus true positives (illegal activity).

3. How was the data collected?
(Web scraping, API, open dataset, etc.)

⇨ Web scraping