# Individual Project

Ring Jewelry Market Analysis on Amazon with Web Scraping

# Background

### amazon

- Amazon.com, Inc. is an American multinational technology company.
- Founded by Jeff Bezos in 1994.
- One of the world's largest e-commerce retailers.

- amazon
- The platform serves millions of customers worldwide with a vast range of products
- Including electronics, fashion, books, jewelry, home goods, and more.

## Objective

- To analyze Amazon global ratings and reviews of ring products.
- To compare popularity across gold (10K, 14K) and silver (925, gold plated) materials.
- To identify top-performing stone types (Diamond, Cubic Zirconia, Labgrown Diamond).
- To find out the leading design styles (Engagement, Eternity band etc.).
- To generate insights that support product development, pricing strategy, and marketing focus.

## **Data Source Details**

### Rings in the jewelry category on Amazon by

- Product ID (ASIN)
- Product Title
- Brand name
- Design style
- Category
- Metal type
- Metal color
- Stone type
- Stone shape
- Size range
- Price
- Customer ratings & global reviews

- 1. Use Python in Jupyter Notebook for Web Scraping
- libraries BeautifulSoup & Pandas
- Create a request and response with User-Agent

```
[1]: import requests
     from bs4 import BeautifulSoup
     import pandas as pd
     import time
     import random
     import re
[2]: urls = {
         f'page{page}': f'https://www.amazon.com/s?k=diamond+rings&page={page}&ref=sr_pg_{page}'
         for page in range(1, 8)
[3]: headers = {
          "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/125.0.0.0 Safari/537.36",
         "Accept-Language": "en-US, en; q=0.9",
         "Accept-Encoding": "gzip, deflate, br",
         "Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8",
          "Connection": "keep-alive",
         "DNT": "1",
          "Upgrade-Insecure-Requests": "1"
[4]: session = requests.Session()
     session.headers.update(headers)
[5]: response = session.get(urls['page1'])
     print(response.status code)
```

#### 2. Create a pattern matching lists from the title

#### keywords:

- Design style
- Categories
- Metal colors
- Metal types
- Stone types
- Shapes



## 3. Function to extract structured info

- Design, Category, Color, Metal, Stone, Shape
- Brand name → Take the first word as brand name and filter out numeric/codes (14k, 1.5ct, etc.)
- Carat value & range, find all format like cts, ctw..

```
[7]: # Extraction function
     def extract info(title):
         title lower = title.lower()
         design = next((d for d in design style if d in title lower), 'N/A')
         category = next((c for c in categories if c in title lower), 'N/A')
         color = next((co for co in metal colors if co in title lower), 'N/A')
         metal = next((m for m in metal types if m in title lower), 'N/A')
         stone = next((g for g in stone types if g in title lower), 'N/A')
         shape = next((s for s in shapes if s in title lower), 'N/A')
         # Brand detection
         brand = title.strip().split()[0] if title.strip() else 'N/A'
         # Check if the first word is numeric or a code (e.g., 14k, 1.5ct, 1/2ct, etc.)
         brand = title.strip().split()[0] if title.strip() else 'N/A'
         if re.match(r'^(d+|d^*..d+|d+/d+))(ct|ctw|carat|k|kt|mm|pcs|cttwdiamond)?$', brand.lower()):
             brand = 'N/A'
          # Carat pattern: range, single value, fractions with units
         carat matches = re.findall(
             r'(\d+(?:\.\d+)?(?:\\d+(?:\.\d+)?)?(?:\s*[--\to]+\s*\d+(?:\.\d+)?(?:\\d+)?)?)?\s*(?:ctw|ct|carat|cttw|ctw))',
             title lower
```

```
carat = ", ".join(dict.fromkeys([m.strip() for m in carat_matches])) if carat_matches else "N/A"

return brand, design, category, color, metal, stone, shape, carat

# Carat range detection
# carat_range = re.search(r'(\d+(?:\.\d+)?|\d+/\d+)\s*(ctw|ct|carat)?[\s\--to]+(\d+(?:\.\d+)?|\d+/\d+)\s*(ctw|ct|carat)?', title_lower)
# if carat_range:
# range_str = f"(carat_range.group(1)) to (carat_range.group(3))"
# return brand, design, category, color, metal, stone, shape, range_str

# Carat individual values
# matches = re.findall(r'(?:(\d+)\s*x\s*)?(\d+(?:\.\d+)?|\d+/\d+)\s*(ctw|ct|carat)', title_lower)
# carat_values = []
# for qty, val, _ in matches:
# carat_values.append(f"{qty} x {val}" if qty else val)

# carat = ", ".join(carat_values) if carat_values else 'N/A'
# return brand, design, category, color, metal, stone, shape, carat
```

#### 4. Create a product list to scrape each pages, included

- Product ID (ASIN)
- Price
- **Star Rating**
- **Global Rating**
- Image & product URL

```
[8]: # List to store product data
      products = []
[9]: # Scrape each page
      for page, url in urls.items():
         try:
              res = session.get(url)
              print(f"Scraping {page} - Status: {res.status code}")
              soup = BeautifulSoup(res.content, "html.parser")
              for item in soup.select(".s-main-slot .s-result-item"):
                  asin = item.get("data-asin", "N/A")
                 title tag = item.select one("h2 span")
                 price tag = item.select one(".a-price .a-offscreen")
                  star rating tag = item.select one(".a-icon-alt")
                  global rating tag = item.select one(".a-size-base.s-underline-text")
                  image elem = item.select one("img.s-image")
                  title = title tag.get text(strip=True) if title tag else "N/A"
                 price = price_tag.get_text(strip=True).replace('$', '') if price_tag else "N/A"
                  star rating = star rating tag.get text(strip=True).split(' out')[0] if star rating tag else "N/A"
                  global rating = global rating tag.get text(strip=True) if global rating tag else "N/A"
                 image url = image elem['src'] if image elem and 'src' in image elem.attrs else "N/A"
                 product url = f"https://www.amazon.com/dp/{asin}" if asin != "N/A" else "N/A"
                  brand, design, category, color, metal, stone, shape, carat = extract info(title)
```







Solid 14k White Gold 4-Prona Petite Twisted Vine 1.0 CT Moissanite Engagement Ring Promise Bridal Ring

4.6\*\*\*\*\* (1K)

\$29399

FREE delivery Wed, Aug 27 to Jon

Add to cart

THELANDA 14k White Gold All Princess Cut Genuine Moissanite or Simulated Diamond Three Stones Engagement Ring 2CT Main Stone 4.6 \*\*\*\*\* (69)

\$36499

FREE delivery Wed, Aug 27 to Hong

Add to cart

14K Solid Gold Side Stones Ring Round/Princess Cut Centerpiece Moissanite Lab Grown Created Diamond, D Color and Flawless... 4.5 \*\*\*\*\* (60)

\$69999

You pay \$629.99 with coupon (some

FREE delivery Wed, Aug 27 to Hong

Add to cart



```
products.append({
            "ASIN": asin,
            "Title": title,
            "Brand Name": brand,
            "Design Style": design,
            "Category": category,
            "Metal Color": color,
            "Metal Type": metal,
            "Stone Type": stone,
            "Stone Shape": shape,
            "Carat": carat,
            "Price": price,
            "Star Rating": star rating,
            "Global Rating": global_rating,
            "Image URL": image url,
            "Product URL": product url
        })
    time.sleep(random.uniform(1, 2))
except Exception as e:
    print(f"Error scraping {page}: {e}")
    continue
```

#### 5) Convert to DataFrame & print

```
[10]: # Optional: Convert to DataFrame
      df = pd.DataFrame(products)
      print(df.head())
                                                                Title Brand Name
               ASIN
                                                                  N/A
                                                                             N/A
      0
         B09L4KQ2BZ FRIENDLY DIAMONDS Lab Grown Diamond Ring For W...
                                                                        FRIENDLY
      2 BODO2GPS8M Le Vian 3/4 or 1 1/2 Carat Chocolate Diamond H...
                                                                              Le
        BODWHNT110 IGI Certified 2 1/2 Carat Emerald Cut Lab Grow...
                                                                             IGI
        BODK495YMS IGI Certified 3 1/4 Carat Oval Lab Grown Diamo...
                                                                             IGI
        Design Style Category Metal Color Metal Type
                                                            Stone Type Stone Shape
                 N/A
                          N/A
                                      N/A
                                                 N/A
                                                                   N/A
                                                                               N/A
           solitaire
                         ring white gold
                                                     lab grown diamond
                                                                               N/A
      2 anniversary
                         ring white gold
                                                14k
                                                               diamond
                                                                             heart
                         ring white gold
                                                     lab grown diamond
      3 anniversary
                                                                           emerald
      4 anniversary
                         ring white gold
                                                14k lab grown diamond
                                                                              oval
                                Price Star Rating Global Rating \
                      Carat
                        N/A
                                 N/A
                                             N/A
      0
                                                           N/A
      1
                 0.50 carat
                               890.00
                                              5.0
                                                             2
                  1/2 carat 5,999.00
                                              N/A
                                                           N/A
          1/2 carat, 2.5 ct 2,499.00
                                             N/A
                                                           N/A
      4 1/4 carat, 3.29 ct 2,499.00
                                             N/A
                                                           N/A
                                                Image URL \
                                                      N/A
      0
      1 https://m.media-amazon.com/images/I/51+7GgG9Xt...
      2 https://m.media-amazon.com/images/I/61gT7fNMj0...
```

```
[11]: # 1. Drop rows without a Title
      df.columns = df.columns.str.strip() # Clean column names
      df cleaned = df.dropna(subset=["Title"]).copy()
      # 2. Fill missing values in feature columns with 'Unknown'
      feature cols = ["Design Style", "Category", "Metal Color", "Metal Type", "Stone Type", "Stone Shape"]
      df cleaned[feature cols] = df cleaned[feature cols].fillna("Unknown")
      # Combine Metal Type and Metal Color into a new column called 'Metal Info'
      df cleaned["Metal Info"] = df cleaned["Metal Type"] + " / " + df cleaned["Metal Color"
      # 3. Clean and convert 'Price' to float
      df cleaned["Price"] = df cleaned["Price"].replace(r"[^\d.]", "", regex=True)
      df cleaned["Price"] = pd.to numeric(df cleaned["Price"], errors="coerce")
      usd to hkd rate = 7.8
      df cleaned["Price HKD"] = (df cleaned["Price"] * usd to hkd rate).round(2)
      # 4. Extract numeric 'Star Rating'
      df cleaned["Star Rating"] = df cleaned["Star Rating"].str.extract(r"(\d+(\.\d+)?)")[0]
      df cleaned["Star Rating"] = pd.to numeric(df cleaned["Star Rating"], errors="coerce")
      # 5. Clean and convert 'Global Rating'
      df_cleaned["Global Rating"] = df_cleaned["Global Rating"].str.replace(",", "")
      df cleaned["Global Rating"] = pd.to numeric(df cleaned["Global Rating"], errors="coerce")
      #6. Clean and convert 'carat'
      df_cleaned["Carat"] = df_cleaned["Carat"].apply(lambda x: f'="{x}"' if isinstance(x, str) and '/' in x else x)
```

#### 6. Cleaned and convert

- Removes all rows without a product title and fill the missing value.
- Combine 'Metal Type' & 'Metal Color' into new column.
- Create new column for 'Price' in USD to 'HKD'
- 'Star Rating' to numeric.
- Cleaned if 'crate' contains a '/' character and a string.

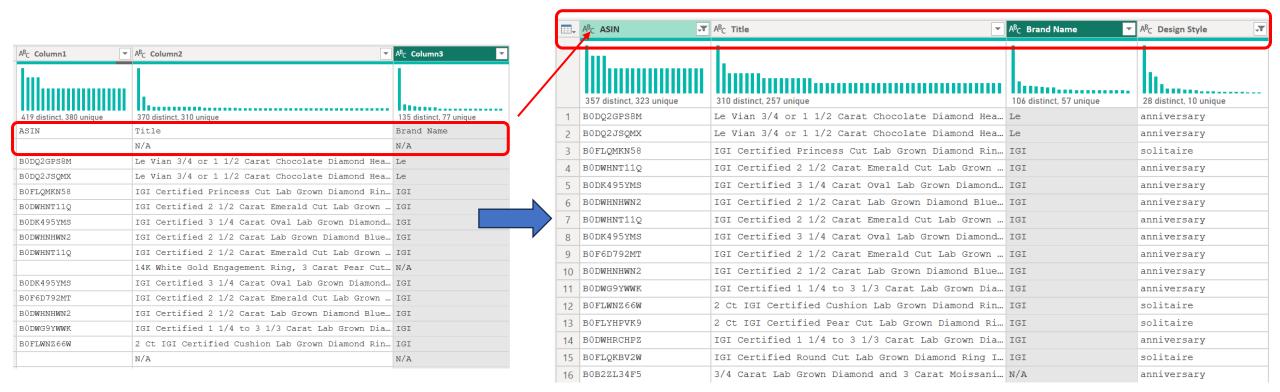
#### 7. Export to dataset csv. file

[12]: # Save cleaned dataset
 df\_cleaned.to\_csv("amazon\_ring\_new.csv", index=False)



### Data Wrangling By Power Query

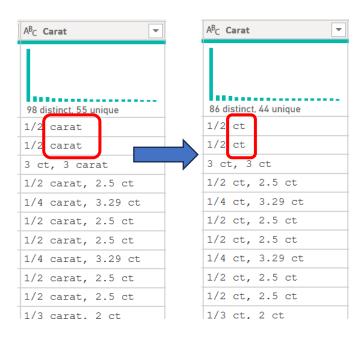
- 1. Filter out 'N/A' in Title column
- 2. Promote the first row as Headers
- 3. Change Type of Price, Star & Global Rating into Decimal number & Whole number
- 4. Remove unnecessary columns (e.g. USD price) and reorder the (HKD price) columns

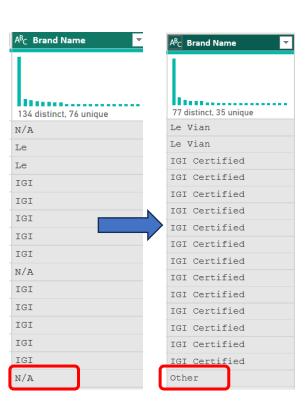


#### Data Wrangling By Power Query

- 4. Use replace value
  - Null value in Star/Global Rating, replace in "0"
  - Replace value "Carat" into "ct"
  - Replace the "N/A" value in Brand Name, Stone Type, Metal type and Color







- Data Wrangling By Power Query
- Create new column for "Product Name".
- 6. Create calculated columns: "Average HK Price", "Average Star Rating", "Total Global Rating".

Product Name	Average Star Rating	Total Global Rating	Average HK Price ▼
925 silver cubic zirconia anniversary ring	3.53	184344	1873.30
925 w/14k gold plated cubic zirconia engagement ring	3.53	184344	1873.30
925 silver cubic zirconia drop ring	3.53	184344	1873.30
925 w/14k gold plated cubic zirconia ring set ring	3.53	184344	1873.30
925 silver cubic zirconia engagement ring	3.53	184344	1873.30
925 silver simulated diamond anniversary ring	3.53	184344	1873.30
925 w/14k gold plated cubic zirconia engagement ring	3.53	184344	1873.30
925 w/18k gold plated simulated diamond eternity band ring	3.53	184344	1873.30
925 silver cubic zirconia engagement ring	3.53	184344	1873.30
925 w/18k gold plated cubic zirconia hip hop ring	3.53	184344	1873.30
925 silver cubic zirconia engagement ring	3.53	184344	1873.30

7. Add New Measure to calculate the average price in "Silver" and "Solid Gold".

581.56
Avg. HK Price (Silver)

3.38K
Avg. HK Price (Solid Gold)

#### Data Visualization

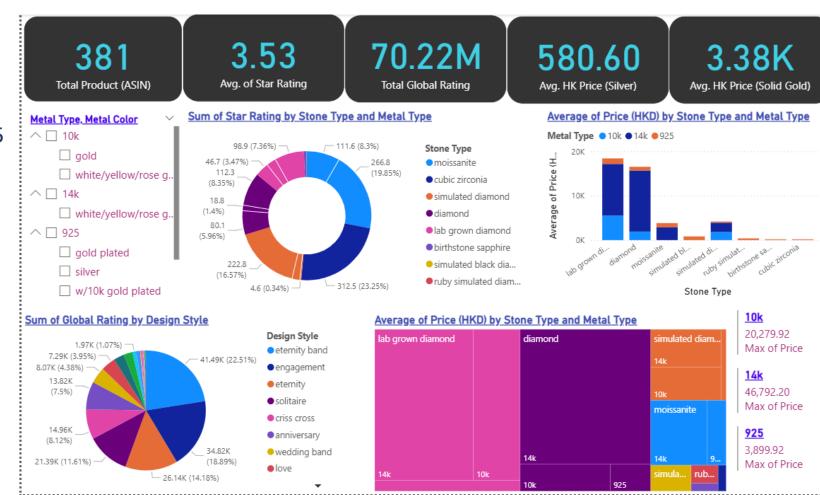
■ **Total Ring Product:** 381pcs

Average Star Rating: 3.53 out of 5

■ Total Global Rating: 70.22M

Silver Average price: \$580.6

Solid Gold Average price: \$3.38K



❖ Data Visualization - Table of Product List with image

#### **Top 5 Ring products (Global Rating)**

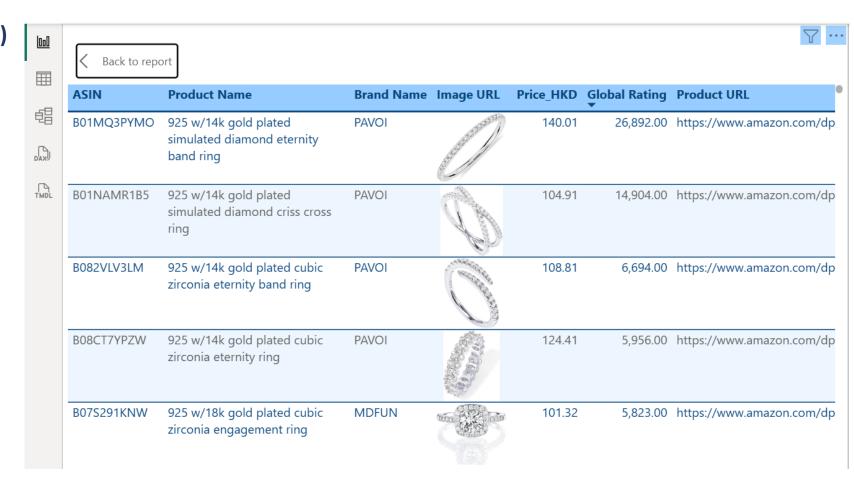
■ **Price range:** HK\$101 to \$140

Metal: 925 silver

Color: 14K or 18K gold plated

Stone: Simulated diamond, Cubic Zirconia

Style: Eternity band ring,
 Crisscross ring,
 Engagement ring



#### **10K** solid gold product:

- 37 products
- Total global review is 6.82M
- **Diamond** has the highest star rating, followed by **lab-grown diamond**.
- Top 3 Design style: Wedding band, Anniversary, 4-prong

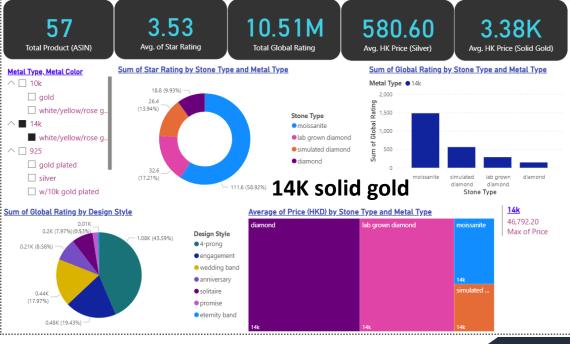
#### **14K solid gold product:**

- **57** products
- Total global review is 10.51M
- Moissanite has the highest star rating, followed by lab-grown diamond.
- Top 3 Design style: 4-prong, Engagement ring, Wedding band

#### **\*** Finding:

- **14K** solid gold products have **more global reviews** (10.51M) than 10K (6.82M).
- Moissanite leads in 14K, while Diamond leads in 10K.
- Lab-grown Diamond ranks second in both 10K and 14K solid gold products.





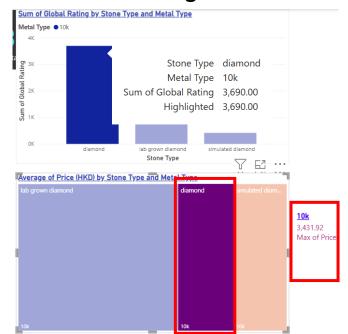
## **❖** Global rating by price w/ stone type 10K solid gold:

- **Diamond** Max. price in HK\$3431, Global rating is 3690.
- Lab-grown diamond Max. price in HK\$20,279, Global rating is 1218.

#### 14K solid gold:

- Moissanite Max. price in HK\$3400, Global rating is 1479.
- **Diamond** Max. price in HK\$46,792, Global rating is 144.
- Lab-grown Diamond Max. price in HK\$19,492,
   Global rating is 287.

#### 10K solid gold



#### 14K solid gold



#### **\*** Finding:

- Price insight: The highest-priced item in 14K is Diamond (HK\$ 46,792), while Moissanite has the highest global rating (1479).
- Price insight: The highest-priced item in 10K is Lab-grown Diamond (HK\$ 20,297), while Diamond has the highest global rating (3690).



### 925 gold plated product:

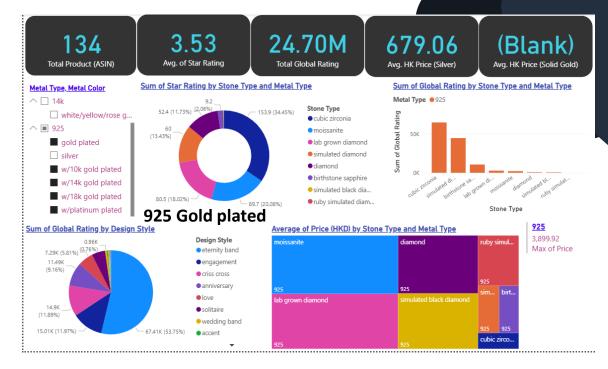
- **134** products
- Total global rating is 24.7M
- Top-rated stone: **Cubic Zirconia**, followed by Moissanite and lab-grown diamond.
- Top 3 Design style: **Eternity band**, Engagement ring, Crisscross ring.

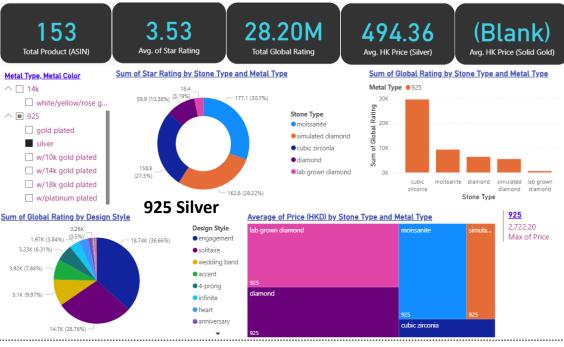
#### 925 silver product:

- **153** products
- Total global rating is 28.2M
- Top-rated stone: **Moissanite**, followed by Stimulated diamond, Cubic Zirconia.
- Top 3 Design style: Engagement ring, Solitaire, Wedding band

#### Finding:

- **925 silver products** have more global reviews (28.2M) than 925 gold plated (24.7M).
- Cubic Zirconia leads in 925 gold plated product, while
   Moissanite lead in 925 silver product by Star rating.





**Key Findings: Design style & Material Popularity** 

#### 925 gold plated product:

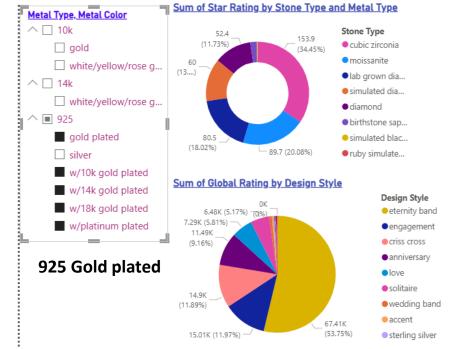
- Eternity Band (53.75%) receives the highest share of global ratings, followed by Engagement Ring (11.97%).
- Stone Type: Cubic Zirconia (34.45%).

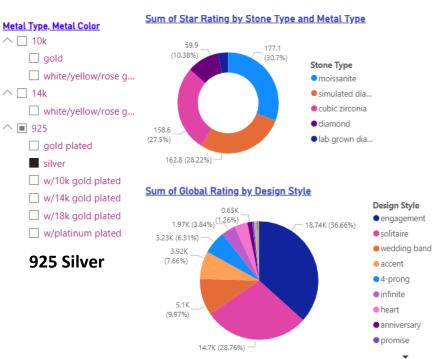
#### 925 silver product:

- Engagement Ring (36.66%) receives the highest share of global ratings, followed by Solitaire (28.76%).
- Stone Type: Moissanite (30.7%) & Simulated diamond (28.22%)

#### **Findings:**

- The **Eternity band (53.75%)** is the most popular design style in 925 gold plated product, while **Engagement ring (36.66%)** leads in 925 silver.
- Cubic Zirconia is the most preferred stone type in 925 gold plated product, while Moissanite leads in 925 silver.





# **Conclusion & Insights**

### 1. Product Popularity

- 925 silver products dominate in total global reviews (28.2M), followed by 925 gold plated (24.7M), showing strong demand for affordable silver jewelry in the market.
- **Solid gold (14K)** products receive more reviews (10.51M) than 10K gold (6.82M), suggesting higher consumer received the quality in 14K gold.

#### 2. Stone Preferences

- Diamond is the top choice in 10K gold products, while Moissanite leads in 14K gold.
- For silver jewelry, Moissanite dominates in 925 silver, while Cubic Zirconia leads in 925 gold plated.
- Lab-grown diamond consistently ranks second or third, indicating growing acceptance.

# **Conclusion & Insights**

### 3. Design Style Trends

- Eternity bands (53.75%) are most popular in 925 gold plated products.
- Engagement rings (36.66%) dominate in 925 silver, while Wedding bands and 4-prong engagement rings remain strong in solid gold.

### 4. Price & Rating Insights

- High price ≠ High rating. For example:
- In 14K gold, Diamonds reach the highest price (HK\$ 46,792) but Moissanite earns the most global ratings.
- In 10K gold, Lab-grown Diamond is highest priced (HK\$ 20,297) but natural Diamond has the most reviews.

# **Conclusion & Insights**

### **Overall Insight:**

- Silver Jewelry (925 silver & gold plated) drives mass-market popularity due to affordability, with Moissanite and Cubic Zirconia being the top stones.
- Solid gold (14K & 10K) attract the premium buyers, where Moissanite shows increasing preference over natural diamonds.
- Eternity bands & Engagement rings remain the strongest universal design trends in the market.