

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light mint green. They are both tilted at an angle, with the blue one being more vertical and the green one being more horizontal.

AMAZON BEST SELLER PRODUCTS

By Group 9



About the project

The aim of our project is to conduct an exploratory analysis of Amazon review data in order to gain valuable insights into consumer preferences and/or behaviors.



Components

Data engineering

Data Ingestion

Data transformation

Data modeling

Data serving



ETL Process





Stage 1: API call

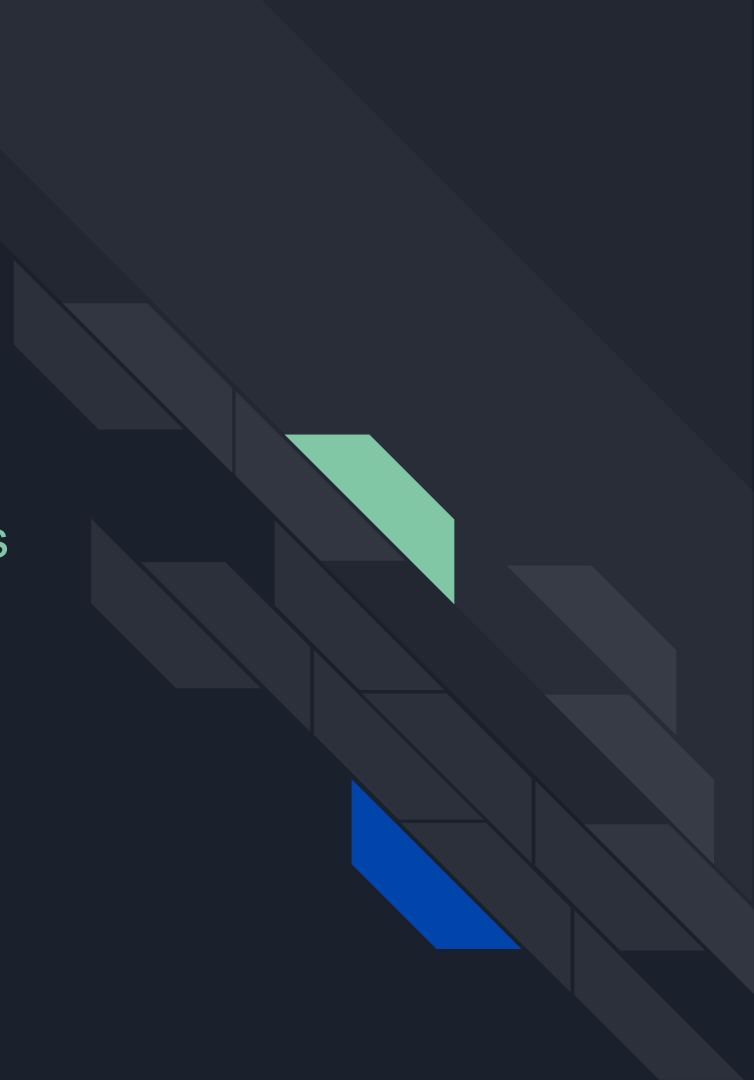
```
# set up the request parameters
params = {
    'api_key': 'api_key',
    'type': 'bestsellers',
    'amazon_domain': 'amazon.ca',
    'category_id': 'bestsellers_videogames',
    'max_page': '5'
}
```

```
.]
```

```
# make the http GET request to Rainforest API
bestsellers_videogames = requests.get('https://api.rainforestapi.com/request', params)
```

- Rainforestapi.com provided the parameters structure and url.
- JSON response

Step 2: JSON Data Load to MongoDB Atlas



Step 3: Transforming the MongoDB Data

```
# define the update
pagination_update = {'$unset': {'pagination': ''}}
```

```
# update the documents
collection.update_many({}, pagination_update)
```

- Dropping the pagination data was successful
- Shortfalls:
 - Using aggregation pipeline on price □ inserting as a new document.
 - Was time constrained.



Step 4: Postgres AWS

- Psycopg2 to connect to Postgres
- Four tables: price, rating, category, book cover types
- Used control flow to iterate over the MongoDB data and insert into Postgres

```
# connect to PostgreSQL
pg_conn = psycopg2.connect(
    dbname='project3',
    user=postgres_user,
    password=postgres_password,
    host='database-1.c9sy44uei6m0.us-east-1.rds.amazonaws.com',
    port='5432'
)

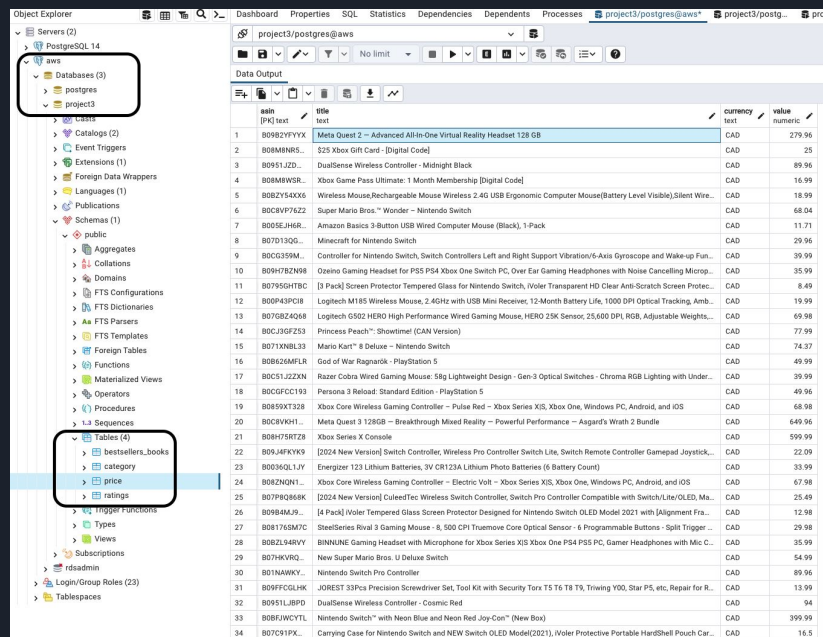
pg_cur = pg_conn.cursor()
```

```
# create a table in PostgreSQL
pg_cur.execute("""
    CREATE TABLE IF NOT EXISTS Price (
        asin TEXT PRIMARY KEY,
        title TEXT,
        currency TEXT,
        value DECIMAL
    )
""")
```

```
# commit the transaction
pg_conn.commit()
```


Step 5: Postgres AWS

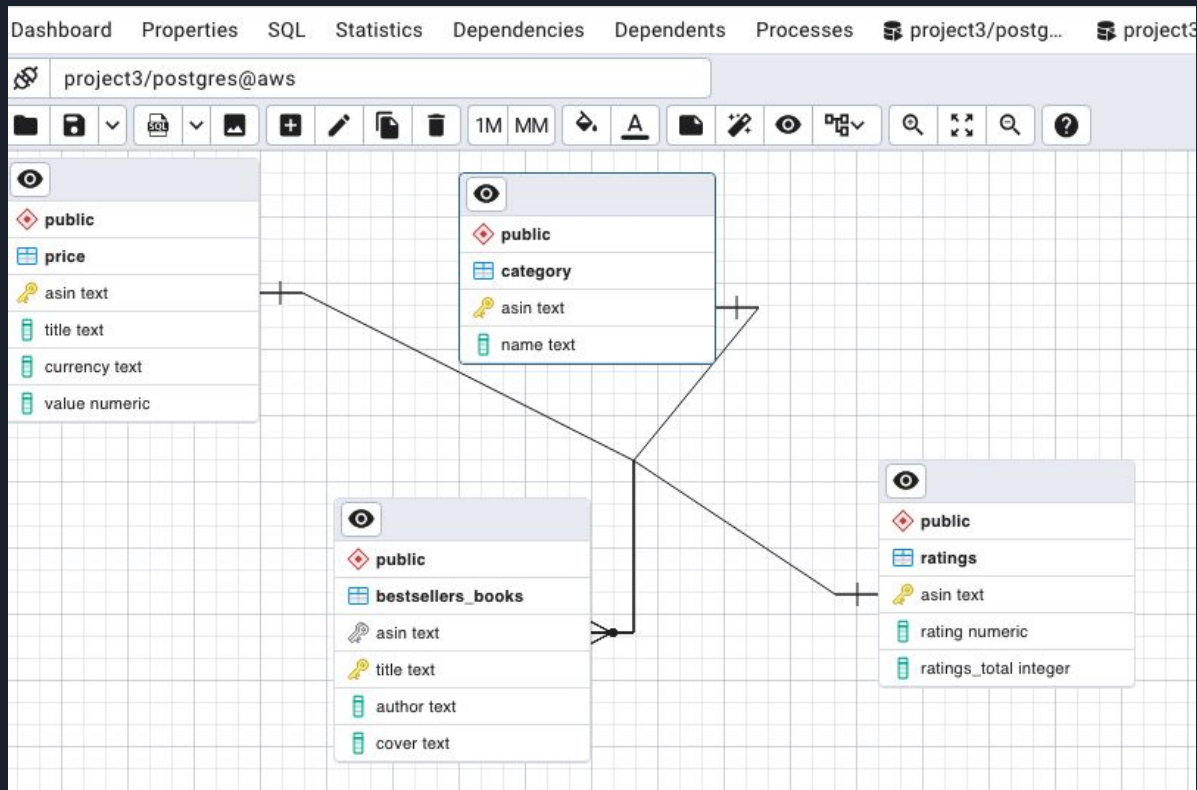
No longer constrained by local Postgres database



The screenshot shows the AWS Data Catalog console. In the left sidebar, the 'public' schema is selected, and the 'bestsellers_books' table is highlighted. The main pane displays a list of data sets with columns: id, title, currency, and value.

id	title	currency	value
1	Meta Quest 2 - Advanced All-in-One Virtual Reality Headset 128 GB	CAD	279.96
2	125 Xbox Gift Card - [Digital Code]	CAD	25
3	DualSense Wireless Controller - Midnight Black	CAD	89.96
4	Xbox Game Pass Ultimate: 1 Month Membership [Digital Code]	CAD	16.99
5	Wireless Mouse, Rechargeable Mouse Wireless 2.4G USB Ergonomic Computer Mouse (Battery Level Visible), Silent Wire...	CAD	18.99
6	Super Mario Bros. "Wonder" - Nintendo Switch	CAD	68.04
7	Amazon Basics 3-Button USB Wired Computer Mouse (Black), 1-Pack	CAD	11.71
8	Minicraft for Nintendo Switch	CAD	29.96
9	Controller for Nintendo Switch, Switch Controllers Left and Right Support Vibration/6-Axis Gyroscope and Wake-up Fun...	CAD	39.99
10	Ozeino Gaming Headset for PS5 PS4 Xbox One Switch PC, Over Ear Gaming Headphones with Noise Cancelling Microph...	CAD	35.99
11	[3 Pack] Screen Protector Tempered Glass for Nintendo Switch, Voler Transparent HD Clear Anti-Scratch Screen Protec...	CAD	8.49
12	Logitech M185 Wireless Mouse, 2.4GHz with USB Mini Receiver, 12-Month Battery Life, 1000 DPI Optical Tracking, Amb...	CAD	19.99
13	Logitech G502 HERO High Performance Wired Gaming Mouse, HERO 25K Sensor, 25,600 DPI, RGB, Adjustable Weights,...	CAD	69.98
14	Princess Peach™: Showtime! (CAN Version)	CAD	77.99
15	Mario Kart™ 8 Deluxe - Nintendo Switch	CAD	74.37
16	God of War Ragnarök - PlayStation 5	CAD	49.99
17	Razer Cobra Wired Gaming Mouse: 58g Lightweight Design - Gen-3 Optical Switches - Chroma RGB Lighting with Under...	CAD	39.99
18	Persona 3 Reload: Standard Edition - PlayStation 5	CAD	49.96
19	Xbox Core Wireless Gaming Controller - Pulse Red - Xbox Series X/S, Xbox One, Windows PC, Android, and iOS	CAD	68.98
20	Meta Quest 3 128GB - Breakthrough Mixed Reality - Powerful Performance - Asgard's Wrath 2 Bundle	CAD	649.96
21	Xbox Series X Console	CAD	599.99
22	[2024 New Version] Switch Controller, Wireless Pro Controller Switch Lite, Switch Remote Controller Gamepad Joystick...	CAD	22.09
23	Energizer 123 Lithium Batteries, 3V CR123A Lithium Photo Batteries (6 Battery Count)	CAD	33.99
24	Xbox Core Wireless Gaming Controller - Electric Volt - Xbox Series X/S, Xbox One, Windows PC, Android, and iOS	CAD	67.98
25	[2024 New Version] CuledTec Wireless Switch Controller, Switch Pro Controller Compatible with Switch/Lite/OLED, Ma...	CAD	25.49
26	[4 Pack] Voler Tempered Glass Screen Protector Designed for Nintendo Switch OLED Model 2021 with [Alignment Fra...	CAD	12.98
27	SteelSeries Rival 3 Gaming Mouse - 8,500 CPI TrueMove Core Optical Sensor - 6 Programmable Buttons - Split Trigger ..	CAD	29.98
28	BINNLUNE Gaming Headset with Microphone for Xbox Series X/S Xbox One PS4 PS5 PC, Gamer Headphones with Mic C...	CAD	35.99
29	New Super Mario Bros. U Deluxe Switch	CAD	54.99
30	Nintendo Switch Pro Controller	CAD	89.96
31	JOREST 33Pcs Precision Screwdriver Set, Tool Kit with Security Torx T5 T6 T8 T9, Triwing Y00, Star P5, etc, Repair for R...	CAD	13.99
32	DualSense Wireless Controller - Cosmic Red	CAD	94
33	Nintendo Switch™ with Neon Blue and Neon Red Joy-Con™ (New Box)	CAD	399.99
34	Carrying Case for Nintendo Switch and NEW Switch OLED Model(2021), Voler Protective Portable HardShell Pouch Car...	CAD	16.5

Postgres ERD



Flask

```
← ↻ ⓘ 127.0.0.1:5000/data
1 {
2   "bestsellers_books": [
3     {
4       "asin": "1639730192",
5       "author": "Sarah J. Maas",
6       "cover": "Paperback",
7       "title": "A Court of Thorns and Roses Paperback Box Set (5 books)"
8     },
9     {
10      "asin": "1338896431",
11      "author": "Dav Pilkey",
12      "cover": "Hardcover",
13      "title": "Dog Man: The Scarlet Shedder: A Graphic Novel (Dog Man #12): From the Creator of Captain Underpants"
14    },
15    {
16      "asin": "0735211299",
17      "author": "James Clear",
18      "cover": "Hardcover",
19      "title": "Atomic Habits: An Easy & Proven Way to Build Good Habits & Break Bad Ones"
20    },
21    {
22      "asin": "0920668364",
23      "author": "Robert Munsch",
24      "cover": "Hardcover",
25      "title": "Love You Forever"
26    },
27    {
28      "asin": "0857197681",
29      "author": "Morgan Housel",
30      "cover": "Paperback",
31      "title": "The Psychology of Money: Timeless lessons on wealth, greed, and happiness"
32    },
33    {
34      "asin": "1635575567",
35      "author": "Sarah J. Maas",
36      "cover": "Paperback",
37      "title": "A Court of Thorns and Roses"
38    },
39    {
40      "asin": "1649374178",
41      "author": "Rebecca Yarros",
42      "cover": "Hardcover",
43      "title": "Iron Flame"
44    },
45    {
46      "asin": "0140280197",
47      "author": "Robert Greene",
48      "cover": "Paperback",
49      "title": "The 48 Laws of Power"
50    },
51    {
52      "asin": "0306832429",
53      "author": "Carleigh Bodrug",
54      "cover": "Hardcover",
55      "title": "PlantYou: Scrappy Cooking: 140+ Plant-Based Zero-Waste Recipes That Are Good for You, Your Wallet, and the Planet"
```

Electronics Data Visualizations and Analysis

```
1 # display dataframe
2 listellers_df
```

✓ 0.0s

	rank	title	rating	ratings_total	price
0	1	All-new Amazon Fire TV Stick 4K streaming devi...	4.6	3909	{'symbol': '\$', 'value': 49.99, 'currency': 'C...
1	2	Amazon Fire TV Stick with Alexa Voice Remote (...)	4.7	46789	{'symbol': '\$', 'value': 41.99, 'currency': 'C...
2	3	All-new Amazon Fire TV Stick 4K Max streaming ...	4.6	2788	{'symbol': '\$', 'value': 64.99, 'currency': 'C...
3	4	Apple EarPods with Lightning Connector - White	4.5	273717	{'symbol': '\$', 'value': 24.94, 'currency': 'C...
4	5	Amazon Kindle Paperwhite (16 GB) - Now with a ...	4.7	7489	{'symbol': '\$', 'value': 159.99, 'currency': 'C...
5	6	TP-Link Tapo Smart Pan/Tilt Indoor Security Ca...	4.5	34000	{'symbol': '\$', 'value': 29.99, 'currency': 'C...
6	7	Wireless Earbuds, Bluetooth Ear Buds 5.3 Mini ...	4.5	2413	{'symbol': '\$', 'value': 28.48, 'currency': 'C...
7	8	Chromecast with Google TV (HD) - Streaming Sti...	4.7	8528	{'symbol': '\$', 'value': 39.98, 'currency': 'C...
8	9	TAGRY Bluetooth Headphones 60H Playback True W...	4.4	62965	{'symbol': '\$', 'value': 39.99, 'currency': 'C...
9	10	Blink Video Doorbell Two-way audio, HD video...	4.0	10949	{'symbol': '\$', 'value': 44.99, 'currency': 'C...
10	11	3rd Gen Remote Control Replacement with Voice ...	4.5	168	{'symbol': '\$', 'value': 13.99, 'currency': 'C...
11	12	Roku Express (Official Roku Product New, 202...	4.6	11174	{'symbol': '\$', 'value': 29.98, 'currency': 'C...
12	13	TP-Link AC1900 WiFi Extender (RE550), Covers U...	4.3	42237	{'symbol': '\$', 'value': 59.97, 'currency': 'C...
13	14	Amazon Fire TV Stick Lite with latest Alexa Vo...	4.6	8958	{'symbol': '\$', 'value': 34.99, 'currency': 'C...
14	15	All-New Blink Outdoor 4 (4th Gen) - Wire-free ...	4.1	1689	{'symbol': '\$', 'value': 77.99, 'currency': 'C...
15	16	Canon PIXMA TS3420 Wireless Inkjet Printer	4.1	4503	{'symbol': '\$', 'value': 59.99, 'currency': 'C...
16	17	soundcore by Anker P20i True Wireless Earbuds,...	4.4	15568	{'symbol': '\$', 'value': 24.99, 'currency': 'C...
17	18	Apple AirPods (2nd Generation)	4.6	19738	{'symbol': '\$', 'value': 168.99, 'currency': 'C...
18	19	Soundcore by Anker Life Q30 Hybrid Active Nois...	4.5	66080	{'symbol': '\$', 'value': 69.99, 'currency': 'C...
19	20	Apple EarPods with USB-C Connector - Microphon...	4.5	273717	{'symbol': '\$', 'value': 25, 'currency': 'CAD'...
20	21	All-new Blink Outdoor 4 (4th Gen) - Wire-free ...	4.1	1689	{'symbol': '\$', 'value': 189.99, 'currency': 'C...
21	22	Pipishell Full Motion TV Wall Mount for 26-65 ...	4.6	7160	{'symbol': '\$', 'value': 49.99, 'currency': 'C...

After calling to API convert to a
Pandas DataFrame

Electronics Data Visualizations and Analysis

```
1 # modify the 'price' column to handle both dictionaries and floats
2 bestsellers_df['price'] = bestsellers_df['price'].apply(lambda x: x['symbol'] + "{:.2f}".format(x['value']) if isinstance(x, dict) else x)
3
4 # display results
5 bestsellers_df
```

✓ 0.0s

	rank	title	rating	ratings_total	price
0	1	All-new Amazon Fire TV Stick 4K streaming devi...	4.6	3909	\$49.99
1	2	Amazon Fire TV Stick with Alexa Voice Remote (...)	4.7	46789	\$41.99
2	3	All-new Amazon Fire TV Stick 4K Max streaming ...	4.6	2788	\$64.99
3	4	Apple EarPods with Lightning Connector - White	4.5	273717	\$24.94
4	5	Amazon Kindle Paperwhite (16 GB) – Now with a ...	4.7	7489	\$159.99
5	6	TP-Link Tapo Smart Pan/Tilt Indoor Security Ca...	4.5	34000	\$29.99
6	7	Wireless Earbuds, Bluetooth Ear Buds 5.3 Mini ...	4.5	2413	\$28.48
7	8	Chromecast with Google TV (HD) - Streaming Sti...	4.7	8528	\$39.98
8	9	TAGRY Bluetooth Headphones 60H Playback True W...	4.4	62965	\$39.99
9	10	Blink Video Doorbell Two-way audio, HD video...	4.0	10949	\$44.99
10	11	3rd Gen Remote Control Replacement with Voice ...	4.5	168	\$13.99
11	12	Roku Express (Official Roku Product New, 202...	4.6	11174	\$29.98
12	13	TP-Link AC1900 WiFi Extender (RE550), Covers U...	4.3	42237	\$59.97
13	14	Amazon Fire TV Stick Lite with latest Alexa Vo...	4.6	8958	\$34.99
14	15	All-New Blink Outdoor 4 (4th Gen) – Wire-free ...	4.1	1689	\$77.99
15	16	Canon PIXMA TS3420 Wireless Inkjet Printer	4.1	4503	\$59.99
16	17	soundcore by Anker P20i True Wireless Earbuds,...	4.4	15568	\$24.99
17	18	Apple AirPods (2nd Generation)	4.6	19738	\$168.99
18	19	Soundcore by Anker Life Q30 Hybrid Active Nois...	4.5	66080	\$69.99

Use the `DataFrame.Apply()`, `Lambda`, `Instances`, `Format` and `Python Replace` to convert the output of the “price” column

Electronics Data Visualizations and Analysis

```
1 # find the index of the highest and lowest priced items
2 highest_priced_index = bestsellers_df['price'].idxmax()
3 lowest_priced_index = bestsellers_df['price'].idxmin()
4
5 # print the title of the highest and lowest priced items
6 print('The highest priced is:', bestsellers_df.loc[highest_priced_index, 'title'])
7 print('The lowest priced item is:', bestsellers_df.loc[lowest_priced_index, 'title'])
```

✓ 0.0s

The highest priced is: JBL Flip Essential 2 - Portable Waterproof Bluetooth Speaker - Gun Metal

The lowest priced item is: 3rd Gen Remote Control Replacement with Voice Function (L5B83G) Fit for Fire Smart TV Cube, 3rd Gen Smart TV, Smart TV Stick

💡 Click here to ask Blackbox to help you code faster

```
1 # find the index of the highest and lowest ratings_total items
2 highest_index = bestsellers_df['ratings_total'].idxmax()
3 lowest_index = bestsellers_df['ratings_total'].idxmin()
4
5 # print the highest and lowest ratings_total items
6 print('The most rated item is:', bestsellers_df.loc[highest_index, 'title'])
7 print('The least rated item is:', bestsellers_df.loc[lowest_index, 'title'])
```

✓ 0.0s

The most rated item is: Apple EarPods with Lightning Connector - White

The least rated item is: 3rd Gen Remote Control Replacement with Voice Function (L5B83G) Fit for Fire Smart TV Cube, 3rd Gen Smart TV, Smart TV Stick

Using idxmin()/idxmax() method and .loc to determine the highest and lowest outcomes.



Electronics Data Visualizations and Analysis

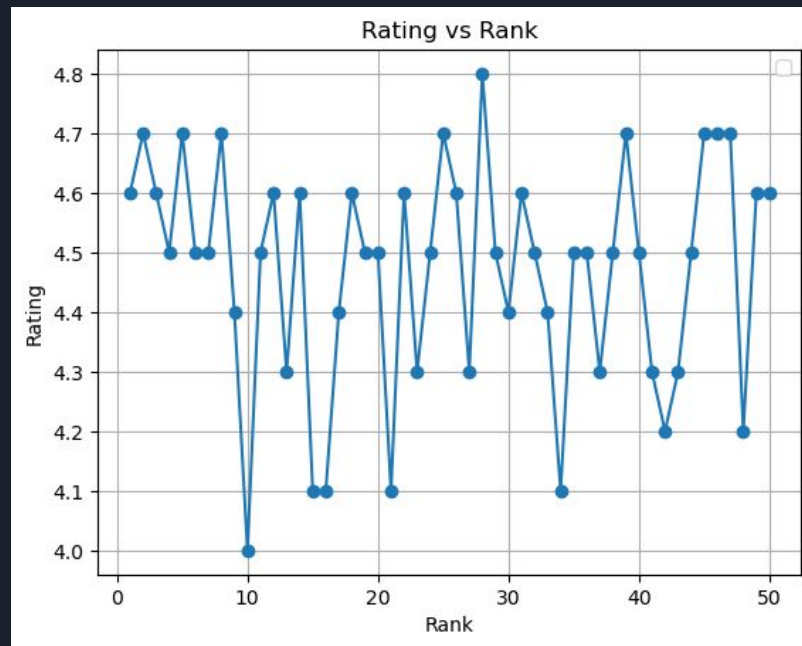
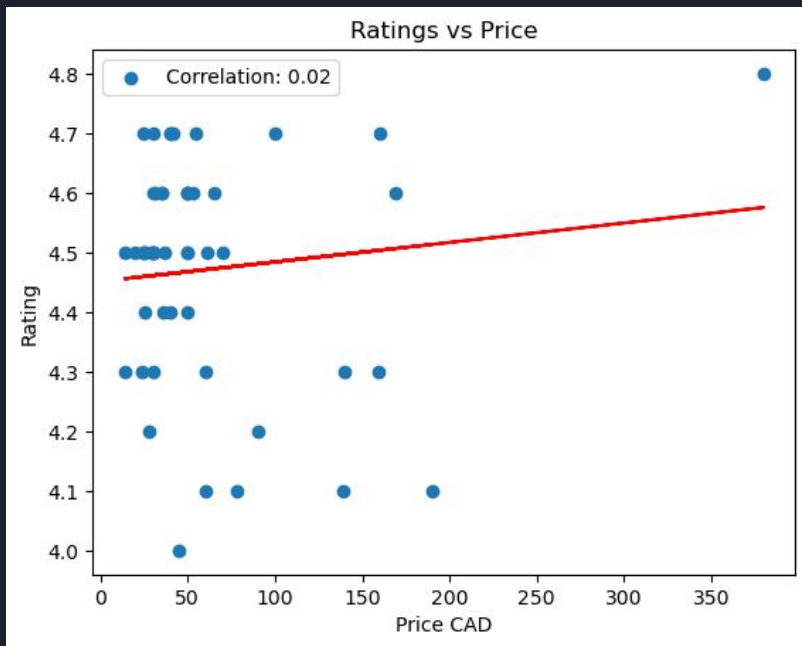
```
1 # find the index of the highest and lowest rated items
2 best_rated_index = bestsellers_df['rating'].idxmax()
3 worst_rated_index = bestsellers_df['rating'].idxmin()
4
5 # print the index of the highest and lowest rated items
6 print('The best rated item is:', bestsellers_df.loc[best_rated_index, 'title'])
7 print('The worst rated item is:', bestsellers_df.loc[worst_rated_index, 'title'])
```

✓ 0.0s

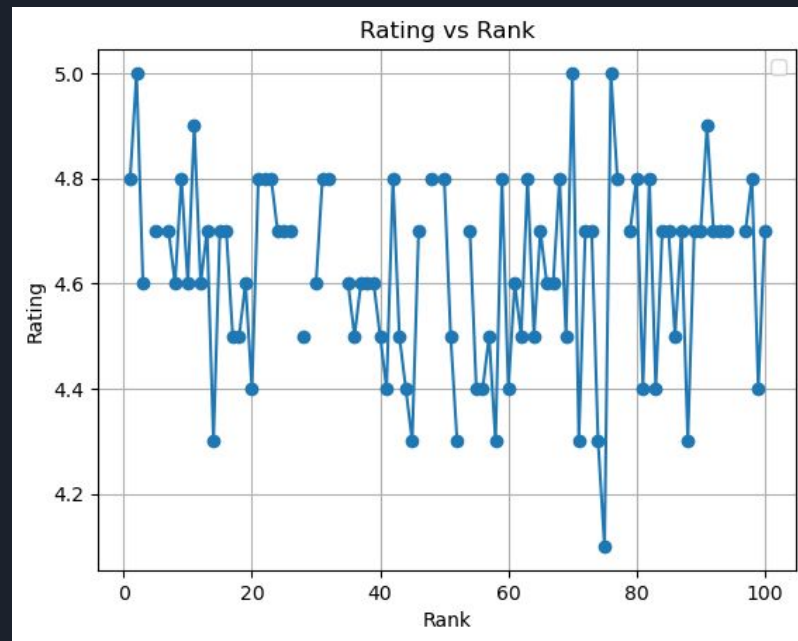
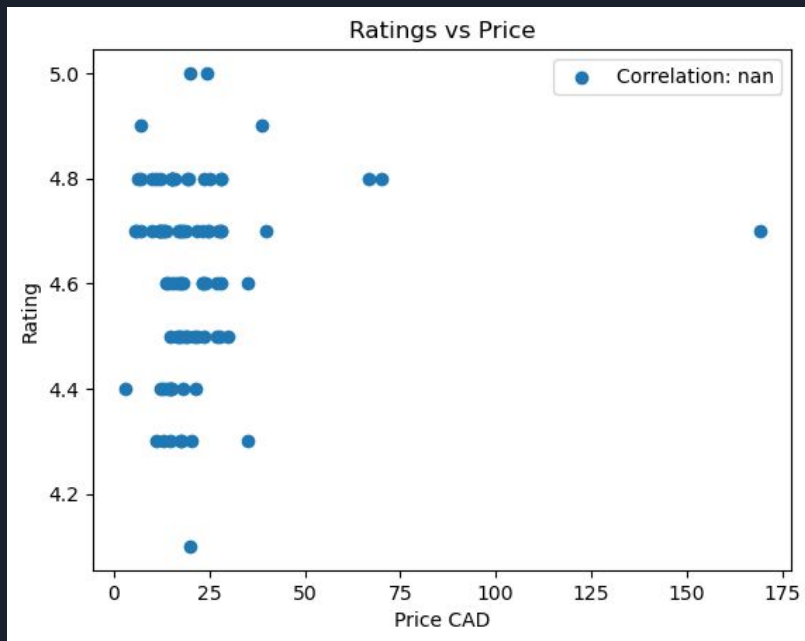
The best rated item is: Apple iPad (9th Generation): with A13 Bionic chip, 10.2-inch Retina Display, 64GB, Wi-Fi, 12MP front/8MP Back Camera, Touch I
The worst rated item is: Blink Video Doorbell | Two-way audio, HD video, motion and chime app alerts, and Alexa enabled – wired or wire-free (Black)

Using `idxmin()/idxmax()` method and `.loc` to determine the highest and lowest outcomes.

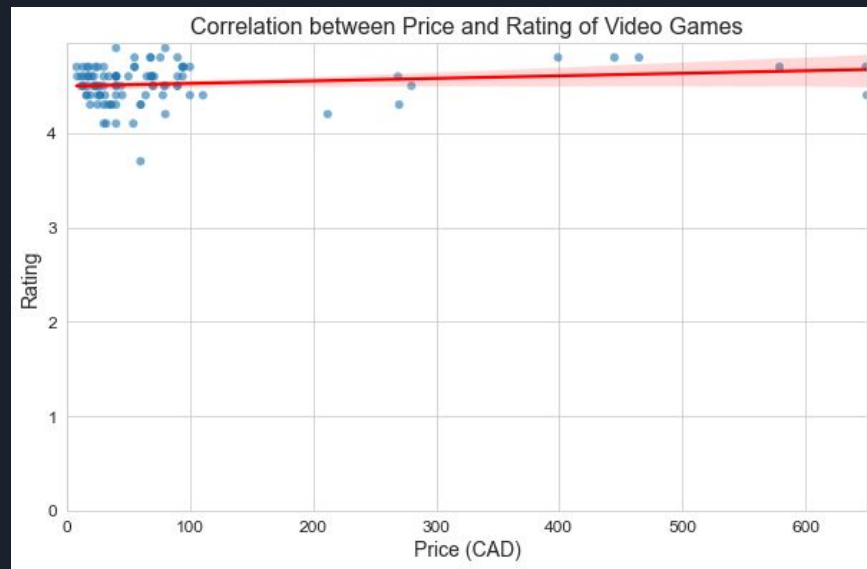
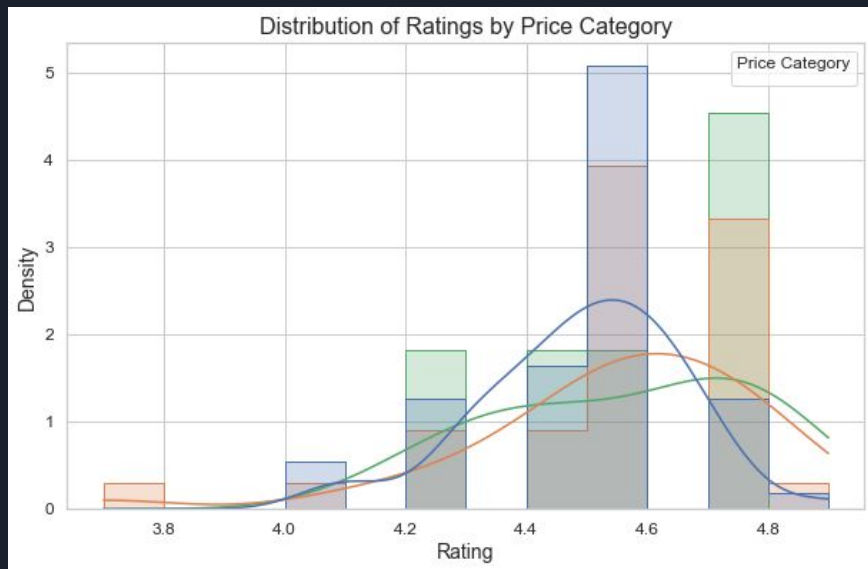
Electronics Data Visualizations and Analysis



Books Data Visualizations and Analysis



Video Game Data Visualizations and Analysis



Thank you

