

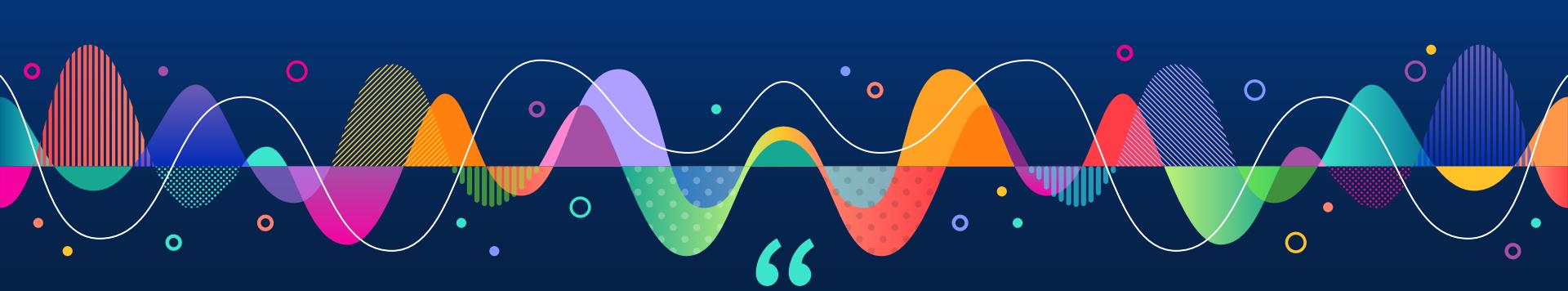
HKTV Mall Web Scraping

Team 3
Henk
Peter
King
Rachel

- ▷ Aim
- ▷ Business Value
- ▷ Data Collection
- ▷ Data Preprocessing
- ▷ Analysis
- ▷ Challenges and Next Steps
- ▷ Conclusion

Aim





“

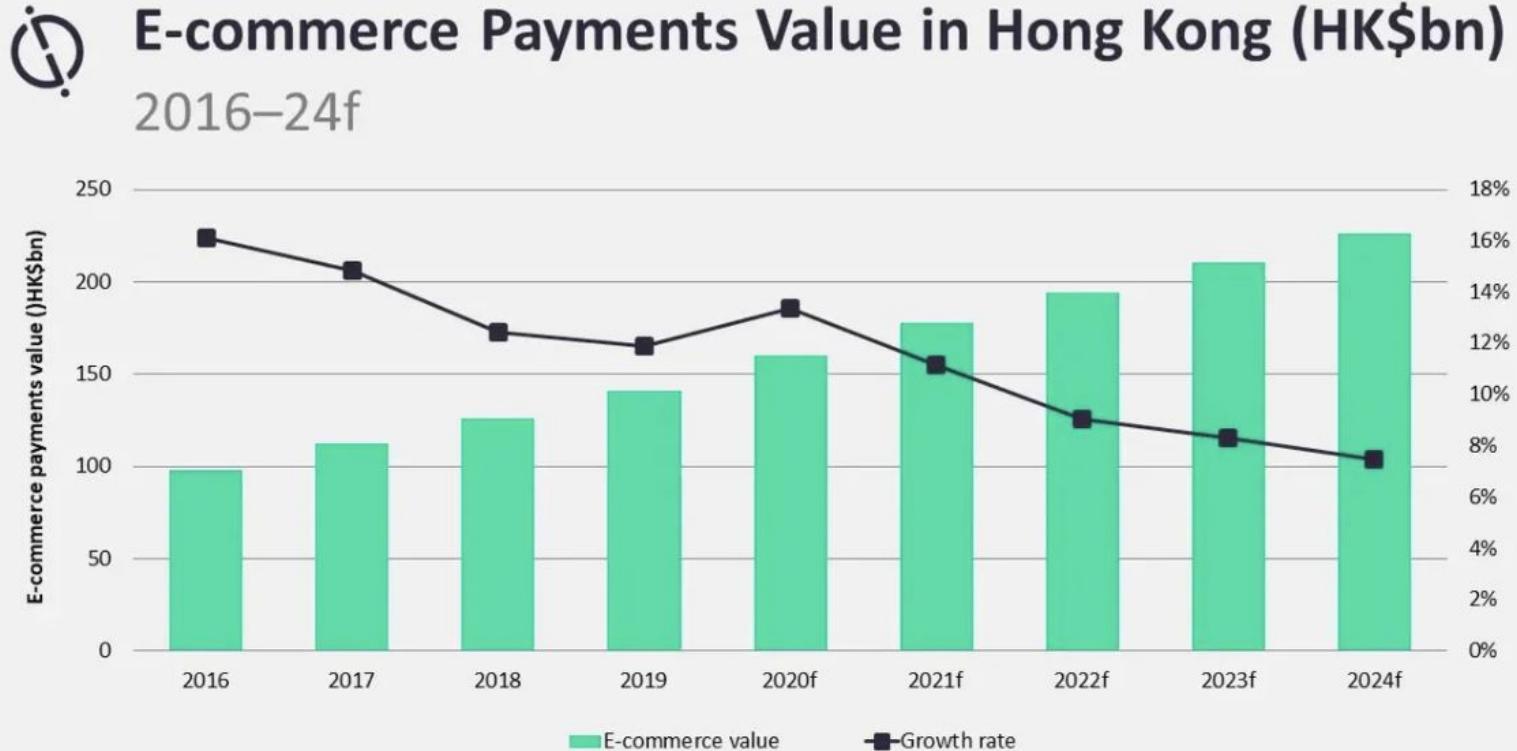
Help Merchants to stand out from the competition , and recommend what products you can sell on HKTVMall.



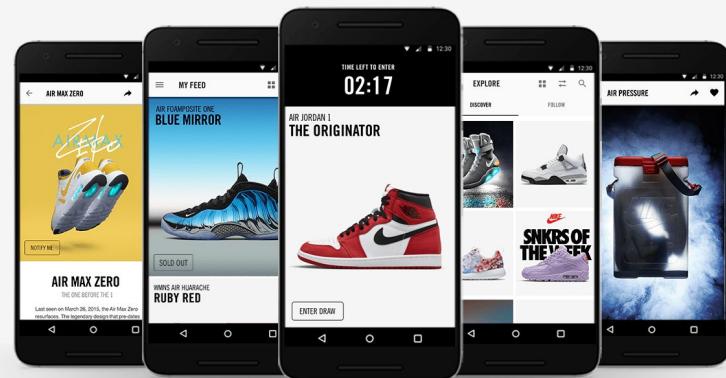
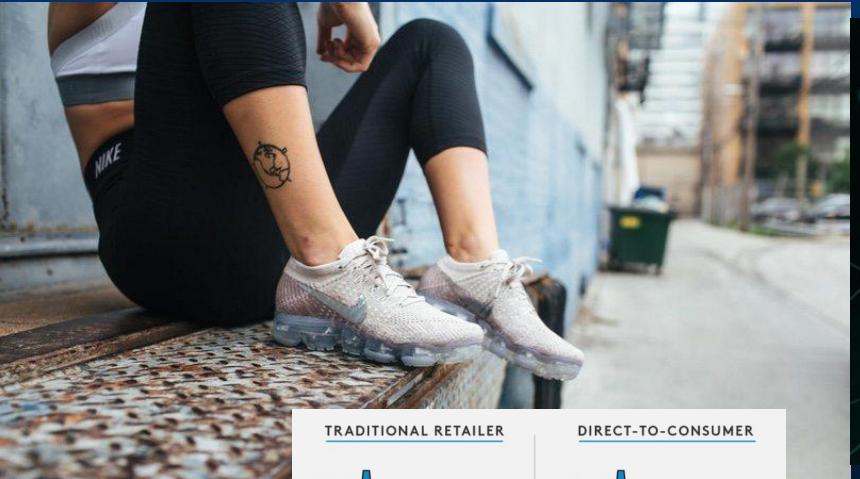
Business Value



Why E-commerce?



Data Science in Retail



HKTVMall

Global Rank ⓘ
Worldwide

▲ 9,050

Country Rank ⓘ
Hong Kong

▲ 43

Category Rank ⓘ
E commerce and Shopping > E com... ⚪

2

Rank ⓘ	Website ⓘ	Category ⓘ
1	 openrice.com	E commerce and Shopping > E commerce and Shopping
2	 hktvmall.com	E commerce and Shopping > E commerce and Shopping
3	 hkgolden.com	E commerce and Shopping > E commerce and Shopping
4	 fortress.com.hk	E commerce and Shopping > E commerce and Shopping
5	 yohohongkong.com	E commerce and Shopping > E commerce and Shopping

Data Collection



Target Data



Libraries Used

- ▷ BeautifulSoup, Selenium
- ▷ Pandas, NumPy, re
- ▷ Seaborn, Matplotlib

Categories:

1. Mobile Phone
2. Tablet
3. Headphone
4. Laptop
5. Smart Watch

Scraping

```
<!-- react-text: 22 -->
<!-- /react-text -->
</div>
▼<div class="info-wrapper" style="display: flex; align-items: center; justify-content: space-between; padding: 10px; margin-bottom: 10px;"> == $0
  ▼<div class="upper-wrapper" style="height: 94px;">
    ▶<div class="brand-product-name">...</div>
    ▶<div class="salesNumber-container">...</div>
    ▶<div class="review-info">...</div>
  </div>
  ▼<div class="lower-wrapper" style="height: 92px;">
    ▶<div class="price-label">...</div>
    ▶<div class="delivery-label-container">...</div>
    ▶<a class="store-name-label crown" target="_blank" href="/hktv/en/s/H7518001">...</a>
  </div>
  ▶<div class="bottom-button-wrapper">...</div>
</div>
▶<div class="price-dropdown-wrapper">...</div>
<a href="main/Whiteplace-Limited/s/H7518001/Home-Appliances/Electrical-Appliance%84%82%E7%A3%85-%E9%9B%BB%E5%AD%90%E7%A3%85/p/H7518001_S_XMSCALE2" target="_blank" rel="noopener"></a>
</div>
</span>
▶<span class="product-brief-wrapper">...</span>
<span class="product-brief-wrapper">...</span>
```

```
while True:
    driver.get(driver.current_url)
    subhtml = driver.page_source
    soup = BeautifulSoup(subhtml, "html.parser")
    nextlink = driver.find_element_by_class_name('next-btn')
    ad_find = 'none' not in driver.find_element_by_class_name("CrazyAd").get_attribute('style')

    for i in soup.find_all(class_="upper-wrapper"):

        product_name = i.find(class_="brand-product-name").text

        try:
            sales_number=i.find(class_="salesNumber-container").text
        except:
            sales_number= 'Null'

        try:
            ratings=i.find(class_="star_container")['data-rating']
        except:
            ratings='Null'

        try:
            review_numbers=i.find(class_="review-number").text
        except:
            review_numbers='Null'

        df=df.append({"Product Name":product_name, "Sales Volume":sales_number, "Rating":ratings,
```

Data Preprocessing



Cleaning Data



- ▷ Extract and clean brand names
- ▷ Clean Chinese characters using re.UNICODE flag
- ▷ Clean and convert type for “Rating”, “Number of Review”, “Price” columns
- ▷ Drop unrelated products

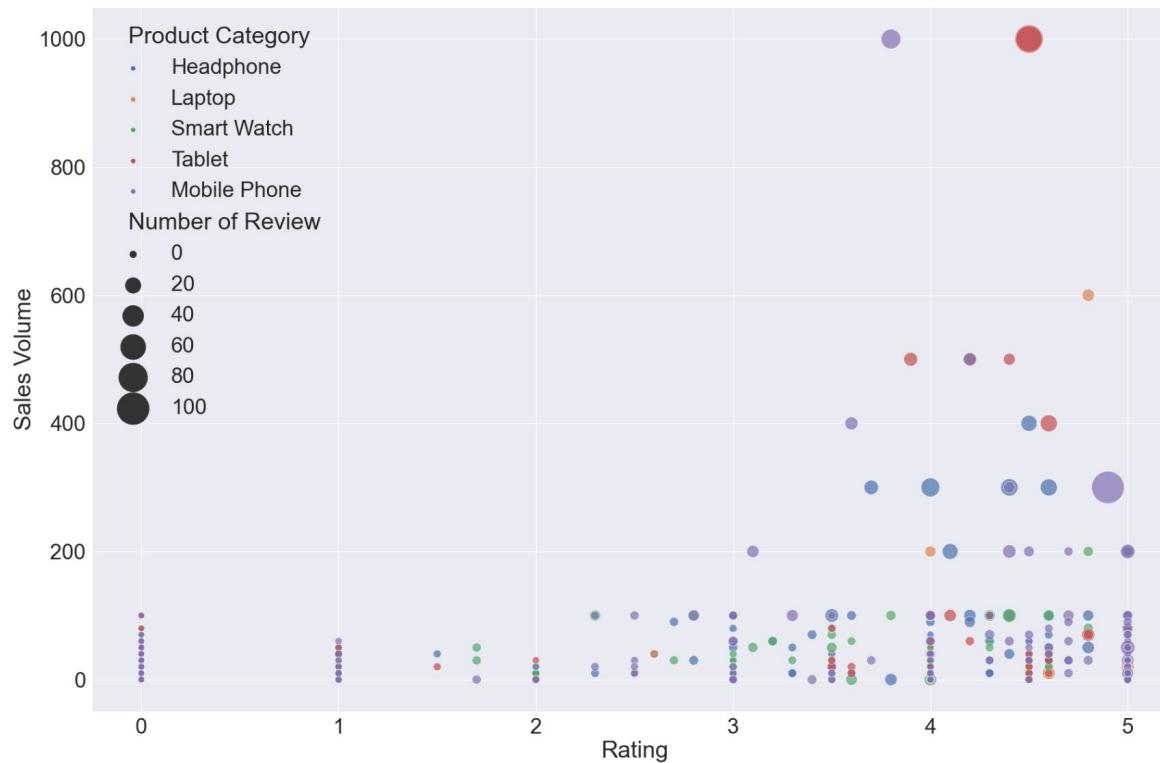
Final DataFrame

	Brand	Product Name	Sales Volume	Product Category	Rating	Number of Review	Price
0	APPLE	AirPods Max - Noise Cancelling Over-Ear Headph...	10	Headphone	0.0	0	3699
1	Bose	Sleepbuds II	10	Headphone	0.0	0	1999
2	Samsung	SAMSUNG Original Headphones EG920	400	Headphone	4.5	22	89
3	Turtle Beach	Stealth 700 Gen 2 Wireless Headset - Xbox Seri...	20	Headphone	0.0	0	1099
4	AWEI	A780BL Bluetooth Headphones Foldable Noise Can...	200	Headphone	4.1	21	188
...
3463	Samsung	Galaxy Z Flip3 5G (8GB+256GB) Phantom Black S...	0	Mobile Phone	0.0	0	8298
3464	Samsung	Galaxy Z Flip3 5G (8GB+256GB) Green Smartphon...	0	Mobile Phone	0.0	0	8298
3465	Samsung	Galaxy Z Flip3 5G (8GB+256GB) Lavender Purple...	0	Mobile Phone	0.0	0	8298
3466	Samsung	Galaxy Z Flip3 5G (8GB+256GB) Cream White Sma...	0	Mobile Phone	0.0	0	8298
3467	Xiaomi	Redmi Note10 5G 128GB Gray Smartphone 【CMHK rec...	0	Mobile Phone	0.0	0	1599

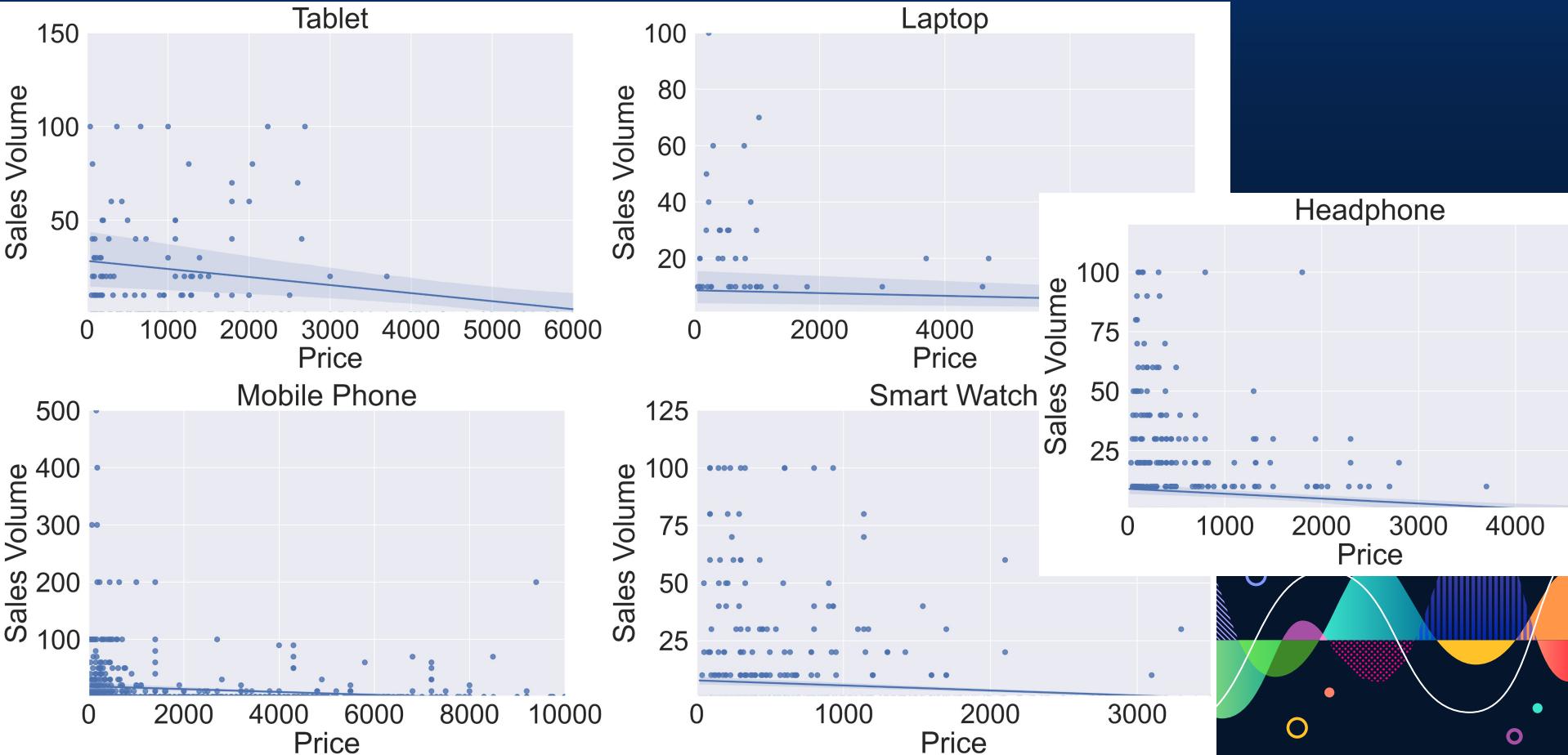
Analysis



General Overview of Data



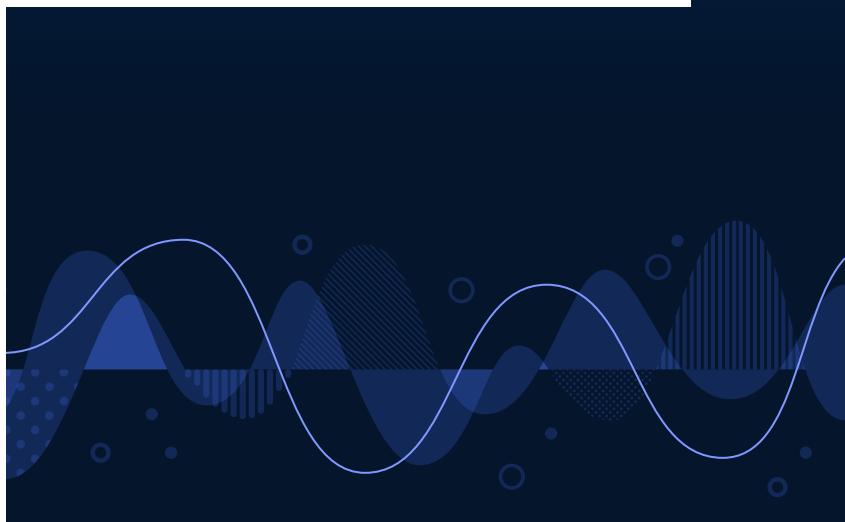
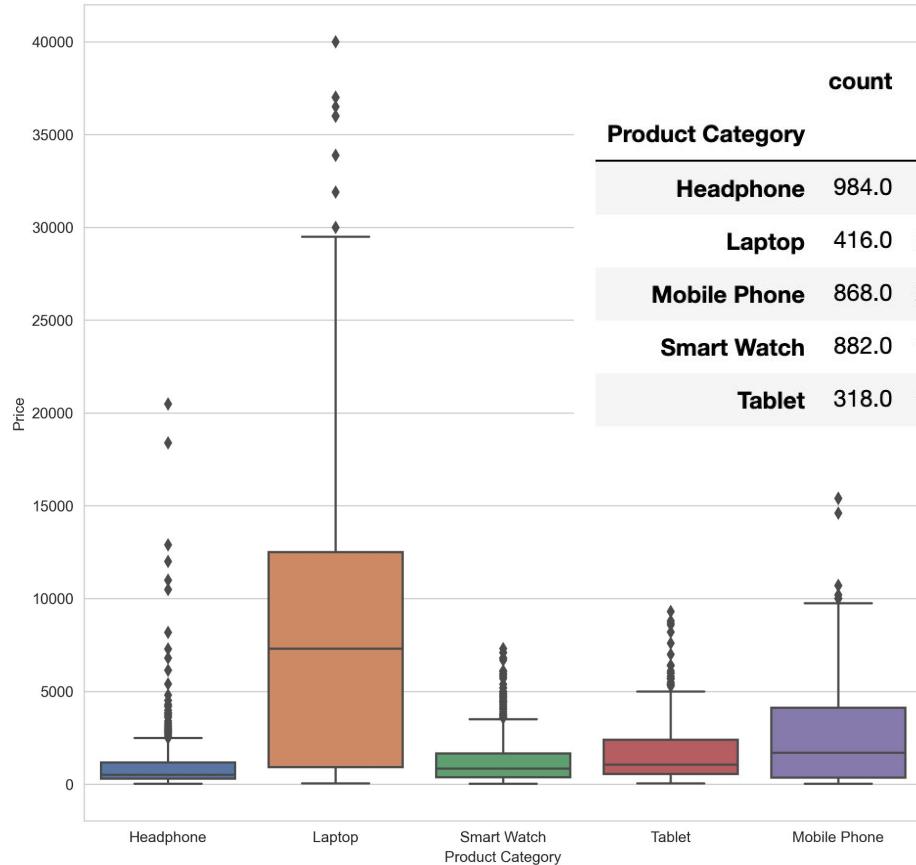
Sales Volume vs Price for 5 Categories

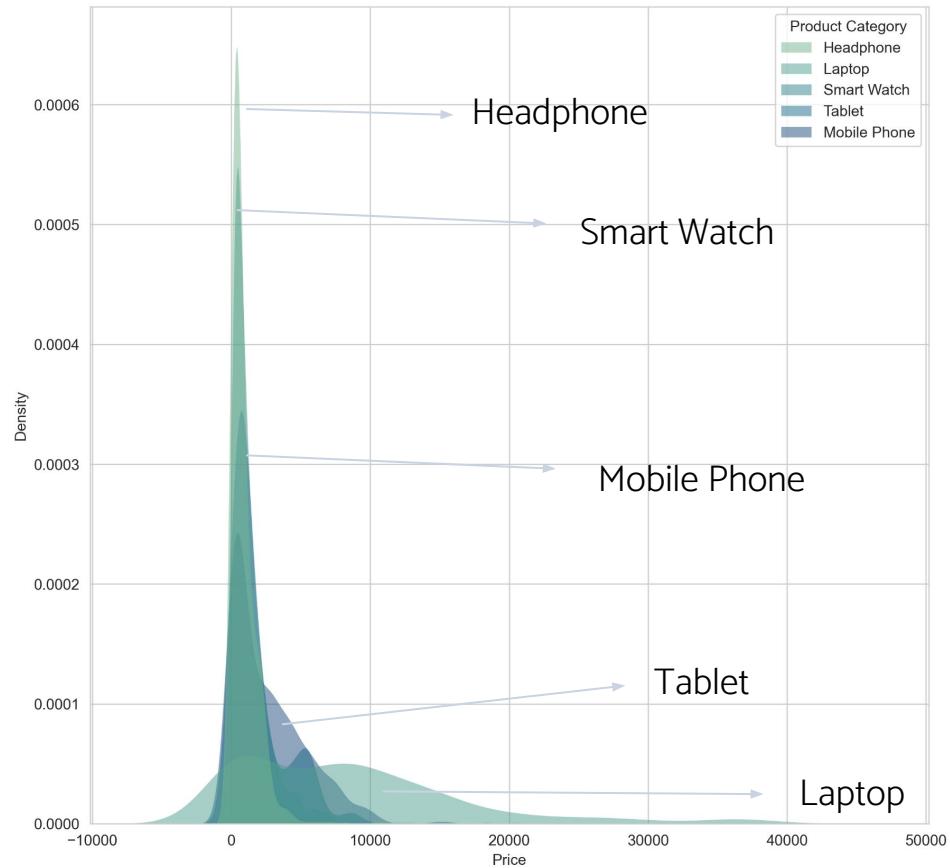


Percentage of Items having no sale for Brand has >50 items in the site

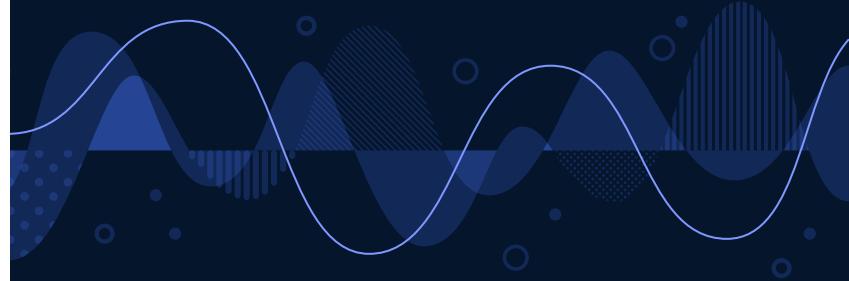


Category vs. Price

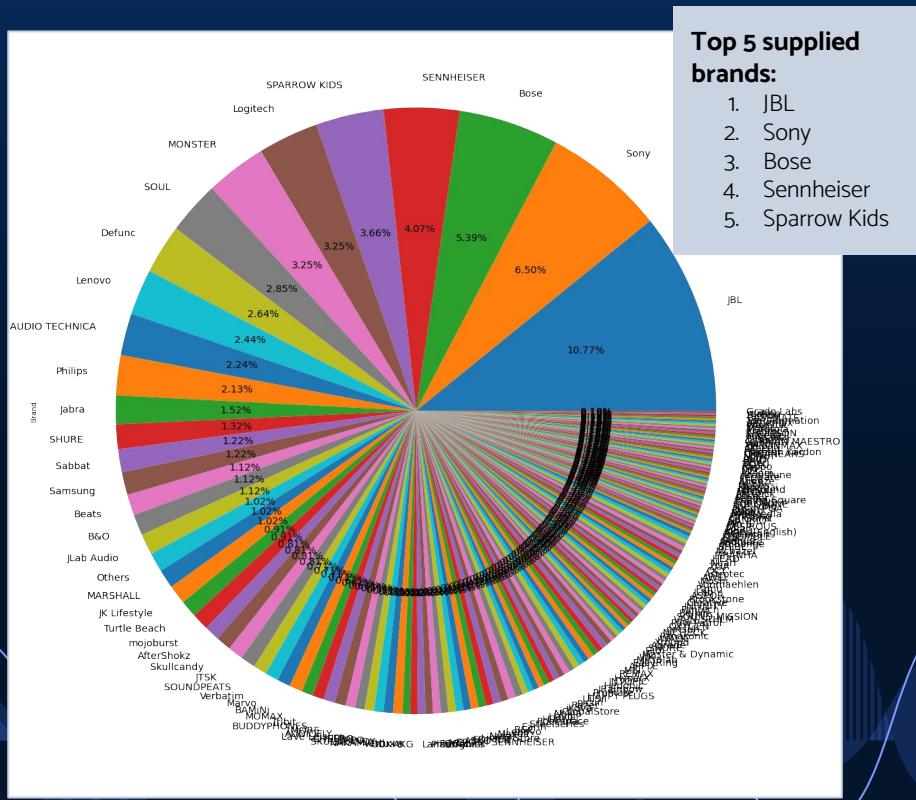
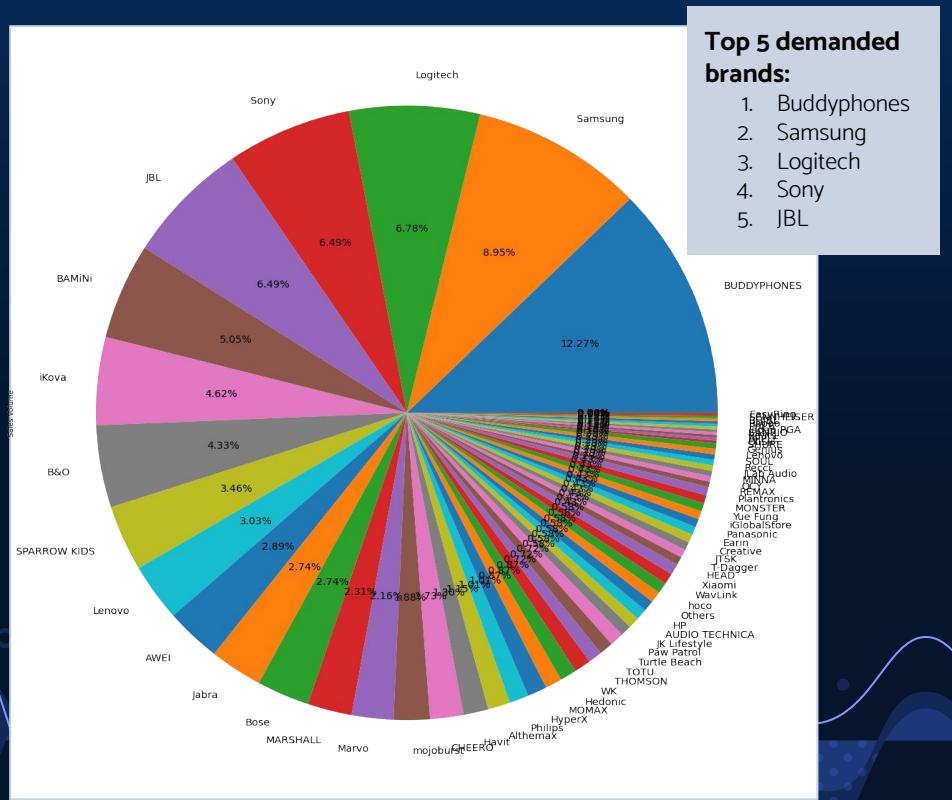




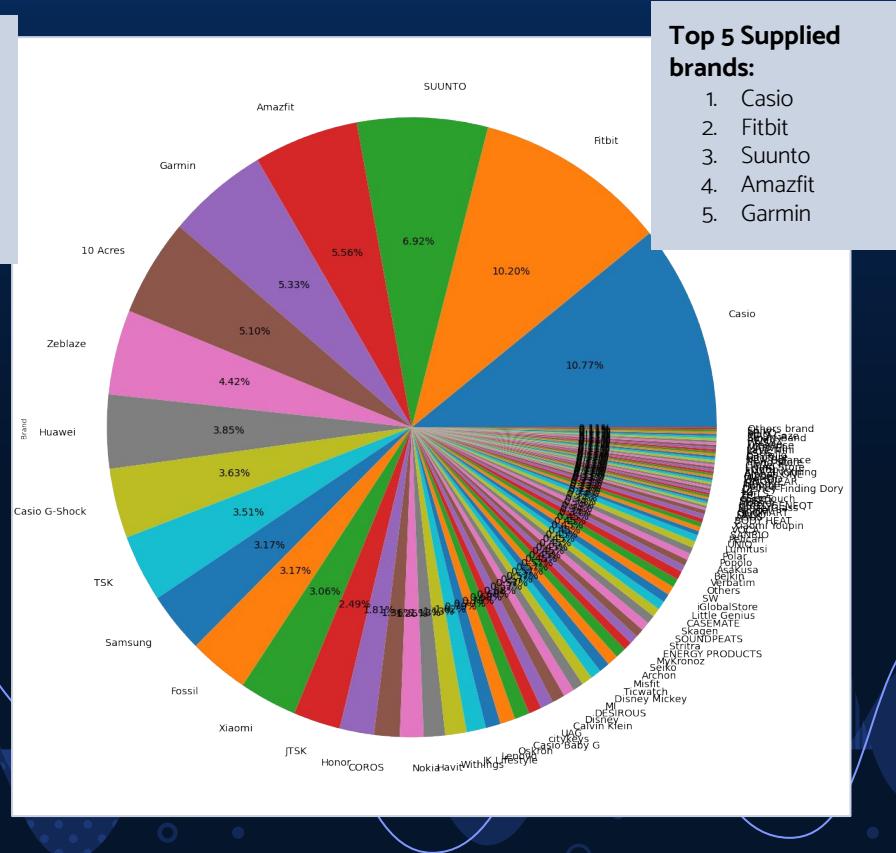
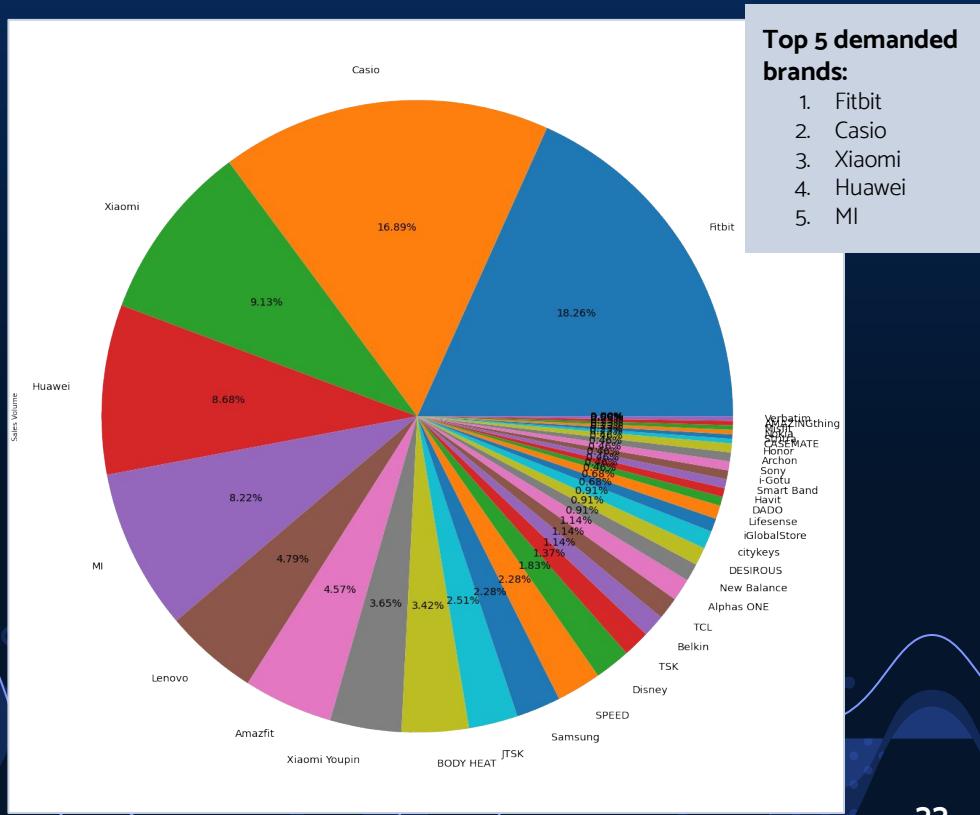
Price vs. Density



Demand vs Supply (Headphones)



Demand vs Supply (Smart Watch)



Challenges



Challenges



- ▷ Crazy Ads stops us from clicking the next page button on the scraping stage

All Home Appliances > Electrical Appliances > Television & Audio

Filtering Results: 4938 items Sort by: Relevance

SAMSUNG

Pre-order Offer
Gifts worth up to \$3,000

HKTVMall Privilege
Galaxy Buds Live*
Trade-in Rebate \$800

Silicon Cover (Green)
Galaxy Store \$400 Coupon

1 /83 Pages Next Page

30天保養
Mi Box S
Taiwanese Rice
EXTRA 12% OFF

Free
UBOX
UBOX
Free: A
200+ S
Galaxy Z Fold3 5G

Image simulated.
The S Pen/Fold Edition sold separately and is only compatible with Z Fold3 5G.
*Customers are responsible for the payment of deductible \$1,155.
*Upgrade the gift of Galaxy Buds Live to Galaxy Buds Pro with \$150.
Promotion till 9 Sep 2021. Subject to limited quota while stocks last, terms & conditions apply.
Please refer to the official Samsung website for details.

- ▷ Sellers can choose which categories (e.g. toothbrush/remote cars in tablet category, smart watch categorized with phone accessories)

Category filtering: All Gadgets & Electronics > Gadgets & Electronics > Tablet & E-Book Reader [422]

Filtering Results: 422 Items Sort by: Relevance

Product	Rating	Price
Xiaomi - [香港行貨] 小米體脂計2 電子磅	★★★★★ (13)	\$159.00
Philips - Sonicare S (4pcs) Sensitive Standard brush...	★★★★★ (1)	\$258.00
PIPO - N2 4G Tablet - 10.1" / 八核心 / 4+64GB /...	★★★★★ (1)	\$999.00

- ▷ No standard way to name products and brands (can include “Fast delivery” or “1 Year Warranty” in name)

Brand	Count
日本熱銷	13
月の光	25
長城電子	5
韓國熱賣	1
飛利浦	1

438 ASUS 華碩

1333 ASUS 華碩

TCL - 32S6500 32" Google Play Android™ TV..

TCL 特約零售商 享6年原廠保養

400+ Sold

★★★★★ (27)

送掛牆架 及 HDMI線

\$1,780.00

- ▷ Sales Volume doesn't clearly represent exact sale numbers

Conclusion



Recommendations for sourcing (Headphones) :



> Buddyphones

Demand: **12.27%** (Ranked highest of all brands)

Supply: **0.06%**

Specific products that sold well:

- Buddyphones infight kids headphones (Blue)
- Explore foldable (blue : BP-ex-fd-blue-01-K)



Recommendations for sourcing (Smart Watch) :



> Lenovo

Demand: **4.79%** (Ranked 6th highest of all brands)

Supply: **0.79%**

Specific products that sold well:

- HX03F Smart Fitness Band Colorful Display



Other Advices:



- Wrong labelling might affect sales when it doesn't show up in the correct category
- The percentage of 'no sales' is high in general, even for the big brands, this could be due to the shopping pattern of customers, as people tend to shop for electronic products in-store

THANKS!

