

CS 732 Data Visualization: A4 Project

Electronic Product analysis

Urja Srivastava (MT2020037)
Akshay Gudiyawar (MT2020137)
Bhashkar Yadav (MT2020136)

1 Problem statement

This project is about implementing a visual analytics workflow. We create different visualizations of different aspects of the dataset, and all of them are stitched together in the workflow.
Tools used - Tableau and Tableau Prep

2 Dataset description

The dataset we have used is Datafiniti Electronic Products and Pricing dataset. This is a list of over 7,000 electronic products with pricing information across 10 unique fields provided by Datafiniti's Product Database. The dataset also includes the brand, category, merchant, name, source, and more.

3 Data Pre-processing/Modelling - By Urja

I have pre-processed and modeled the data using Tableau Prep.

3.1 Grouping

In many columns of the dataset multiple values shared the exact same meaning. Thus such values can be grouped together under one name. In prices.availability column "In stock", "available" and "yes" have the same meaning, thus these values are grouped together, similarly values "no", "out of stock" "not available" have the same meaning.

Similarly I have performed grouping of attributes under different columns which had the exact same meaning. These columns are

1. prices.availability - product availability related data, example - out of stock , in stock , only special order.
2. prices.condition - product condition related data, example - new, pre owned or refurbished.
3. prices.shipping - shipping options and charges
4. categories - product categories, example - electronic furniture, electronics , carplay etc

TABLE COLUMNS

id ⓘ
prices_amountmax ⓘ
prices_amountmin ⓘ
prices_availability ⓘ
prices_condition ⓘ
prices_currency ⓘ
prices_dateseen ⓘ
prices_issale ⓘ
prices_merchant ⓘ
prices_shipping ⓘ
prices_sourceurls ⓘ
asins ⓘ
brand ⓘ
categories ⓘ
dateadded ⓘ
dateupdated ⓘ
ean ⓘ
imageurls ⓘ
keys ⓘ
manufacturer ⓘ
manufacternumber ⓘ
name ⓘ
primarycategories ⓘ
sourceurls ⓘ
upc ⓘ 2
weight ⓘ

Figure 1: Dataset Columns

Average Product Prices grouped By Brand

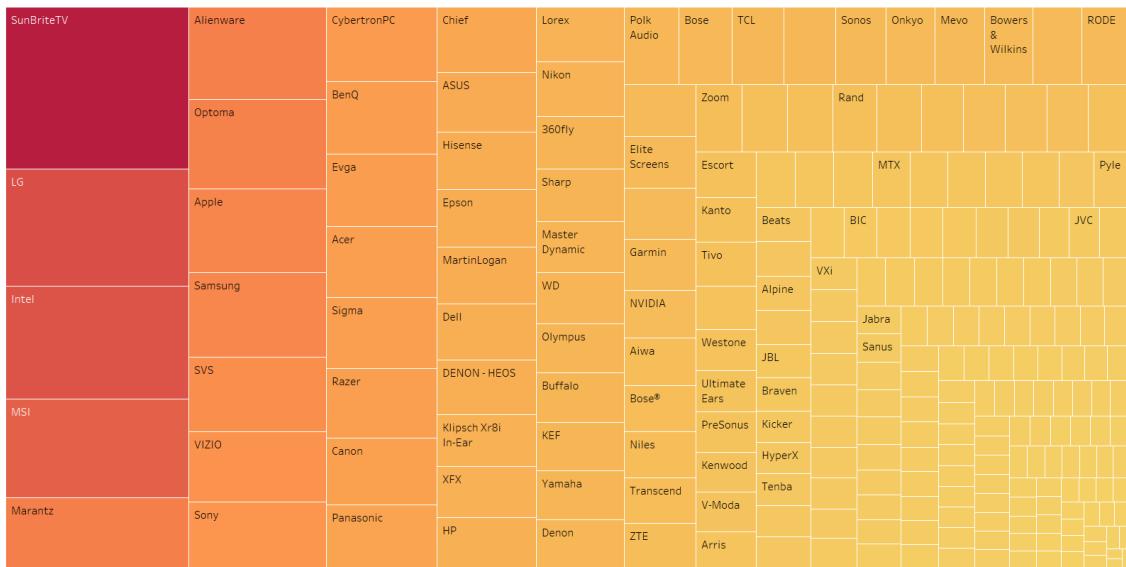


Figure 2: Price Analysis- by Urja

3.2 Calculated Fields

I have created three new features based on my initial analysis of data. These features proved to be very useful in further analysis of the dataset, especially price related analysis.

1. $\text{prices.discount} = ([\text{prices.amountmax}-\text{prices.amountmin}]/\text{prices.amountmax}) * 100$
2. prices.website – website names which are selling the products are extracted from the prices.sourceurls by splitting the column url values.
3. prices.difference – difference between prices.amountmax and prices.amountmin .

3.3 Handling missing data

The column ean is removed from the dataset as its mostly null, in other columns nulls are left as it is to avoid any bias.

4 Analysis- By Urja

4.1 Price Analysis

Products by brands SunbriteTv, LG, Intel and MSI have the highest average product prices. During sale LG and Samsung products have the highest average price difference.

Average Product prices on Walmart decrease during Sale. Other competitors do not offer any discounts during sale on the electronic products listed on their website.

Average product prices for products added in the year 2017, fall the most. Average product prices for products added in the year 2017 are also higher than other years.

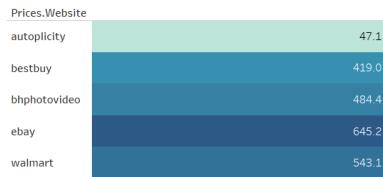
Average product prices decrease only for new products. Pre-owned and refurbished products

Average product prices in Sale grouped by Brand



Figure 3: Price Analysis- by Urja

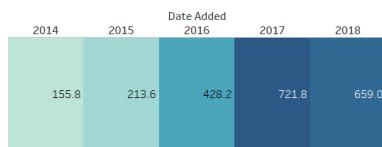
Average product prices grouped by listing website



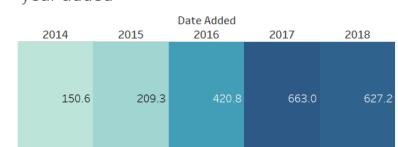
Average product prices on Sale grouped by listing website



Average product prices grouped by date added



Average product prices on Sale grouped by year added



Average product prices goes down for Walmart during Sale.

Average product prices for the year 2017 shows the most decrease during Sale.

Figure 4: Price Analysis- by Urja

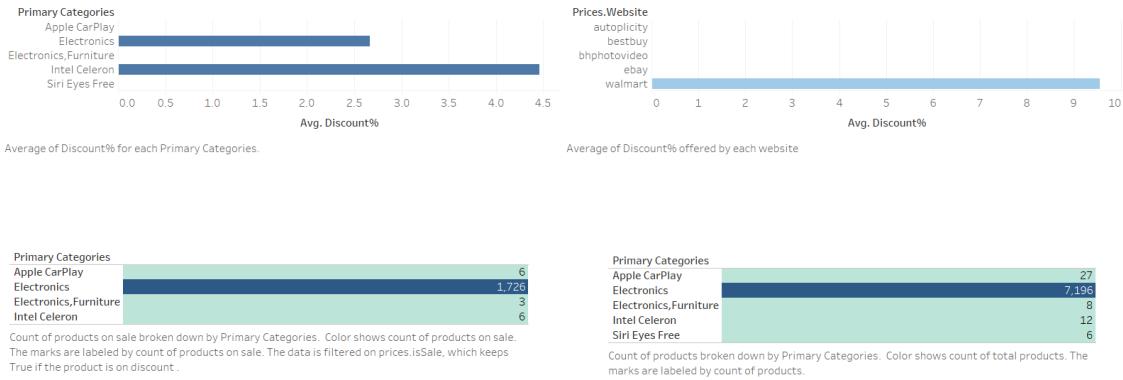


Average product prices of only new products decreases during Sale. Average product prices of pre owned and refurbished do not change.
Average product prices of Out of Stock products is less than that of In stock products.

Figure 5: Price Analysis- by Urja



Figure 6: Price Analysis- by Urja



Only Walmart offers discounts on sale products. Majority of products on sale are electronics and intel products.

Out of 7196 electronics products, 1726 are on sale with an average discount of 2.66%.

Figure 7: Product Analysis- by Urja

show no decrease in their average prices.

Out of stock products average prices are lesser compared to in stock products.

4.2 Product Analysis

Only Walmart offers discounts on sale products. Majority of products on sale are electronics and intel products. Out of 7196 electronics products, 1726 are on sale with an average discount of 2.66%.

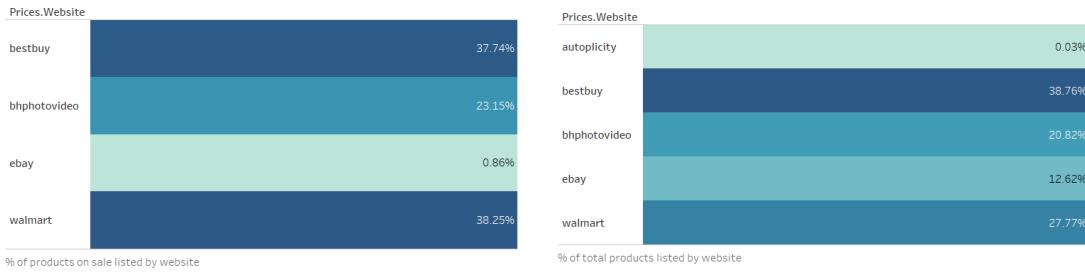
Walmart and Bestbuy have almost equal share in the number of products on sale. Bestbuy however has a bigger catalogue of products overall and has listed 38.76% of the total products. Walmart has a bigger sale compared to other competitors, listing 38.25% of the total products on sale.

Only ebay is offering pre-owned products on sale. Bestbuy is offering both new and refurbished products on sale. Walmart and bhphotovideo offers only new products on sale.

Ebay has free shipping on 80.55% of its products. Ebay levies varying shipping charges depending on the product being shipped. Walmart levies standard shipping charges on 32.44% of its products and varying charges on 34.62% of its products.

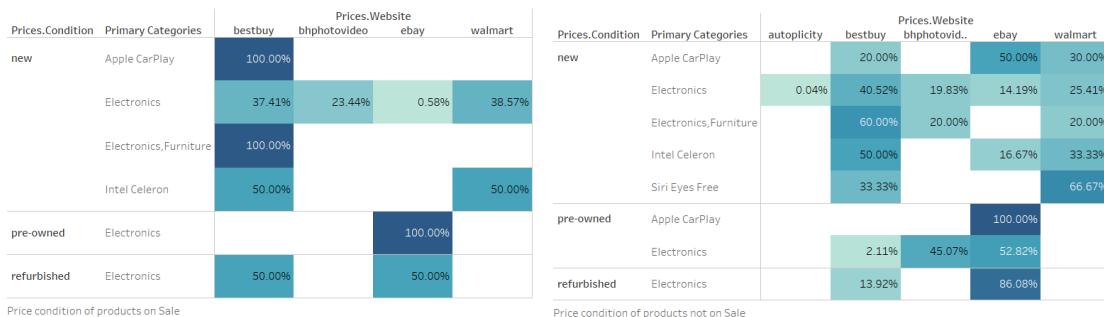
Bestbuy maintains the biggest catalogue with majority of its products in stock. Walmart has the highest share of out of stock products.

Majority of products on all websites were added in 2017 and 2015. Some of the products added in 2015 and 2018 are available for special orders in sale. Majority of the product catalogue on sale was added in the year 2017.



Walmart and Bestbuy have almost equal share in the number of products on sale.
 Bestbuy however has a bigger catalogue of products overall and has listed 38.76% of the total products.
 Walmart has a bigger sale compared to other competitors, listing 38.25% of the total products on sale.

Figure 8: Product Analysis- by Urja



Only ebay is offering pre-owned products on sale.
 Bestbuy is offering both new and refurbished products on sale.
 Walmart and bhphotovideo offers only new products on sale.

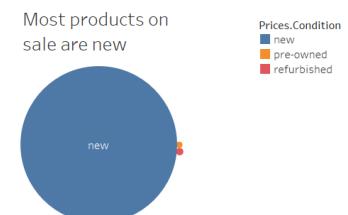


Figure 9: Product Analysis- by Urja

Shipping charged by Websites



Ebay has free shipping on 80.55% of its products.

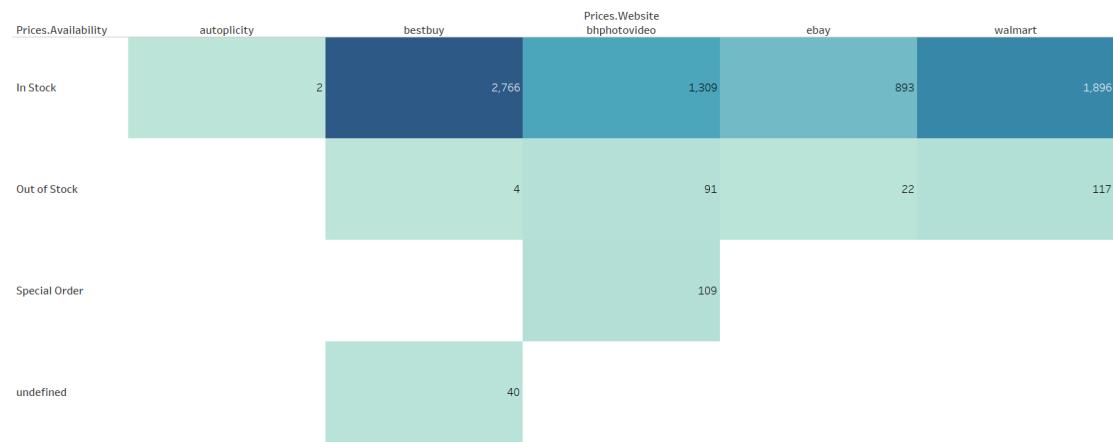
Ebay levies varying shipping charges depending on the product being shipped.

Walmart levies standard shipping charges on 32.44 % of its products and varying charges on 34.62% of its products.

..

Figure 10: Product Analysis- by Urja

Availability of Listed Products



Bestbuy maintains the biggest catalogue with majority of its products in stock.

Walmart has the highest share of out of stock products.

Figure 11: Product Analysis- by Urja

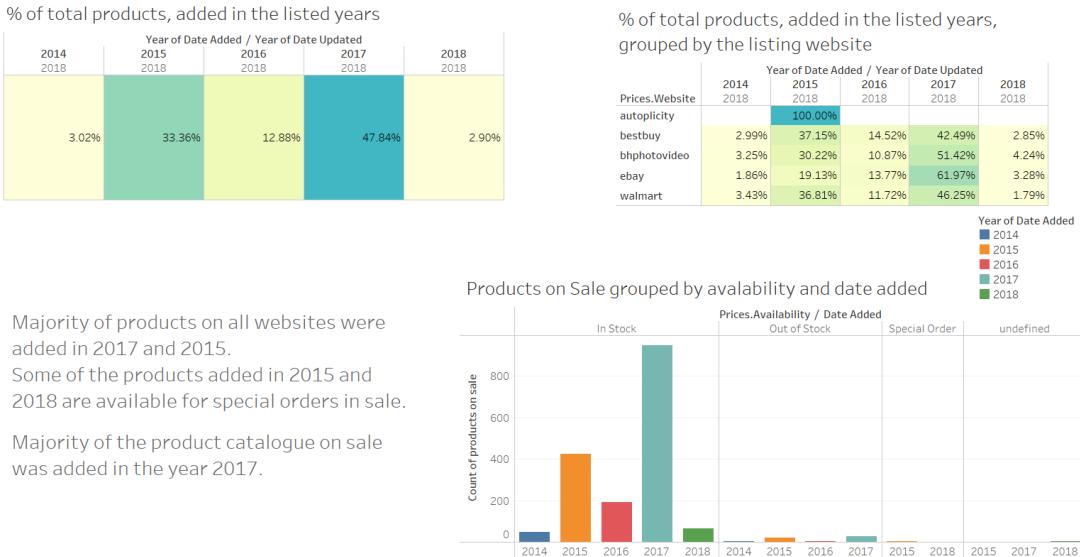


Figure 12: Product Analysis- by Urja

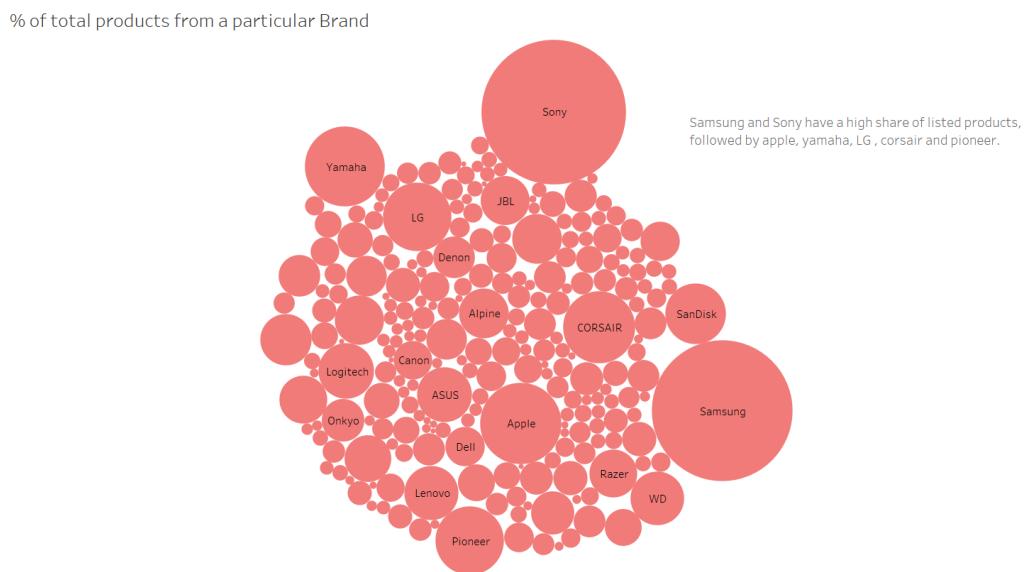


Figure 13: Brand analysis- by Urja

% share of a Brand in total products on sale, along with average discount offered on the brand products.		
Brand	% of Total Count of prices.isSale along Brand	Avg. Discount%
Samsung	17.06%	18.58
Sony	14.53%	8.99
LG	4.31%	20.29
Apple	4.19%	0.25
Yamaha	3.96%	0.00
Pioneer	3.27%	11.91
SanDisk	2.24%	0.80
Lenovo	2.07%	0.52
Alpine	1.84%	6.27
Kenwood	1.78%	6.44
Panasonic	1.49%	5.38
Razer	1.49%	7.65
Elite Screens	1.49%	11.74
CORSAIR	1.49%	17.57
Dell	1.44%	1.09
ASUS	1.44%	1.49
Alienware	1.38%	1.10
Canon	1.03%	2.81
Olympus	1.03%	3.94
Denon	0.98%	5.51
JBL	0.86%	3.21
Sennheiser	0.86%	10.73
Onkyo	0.80%	5.53
Logitech	0.75%	1.92
House of Marley	0.75%	14.03
V-Moda	0.75%	15.60
360fly	0.75%	30.30
Microsoft	0.69%	3.87
Sanus	0.63%	38.13
Insignia	0.57%	0.00
WD	0.57%	4.95
Grace Digital	0.57%	19.22
Power Acoustik	0.57%	32.73
Jensen	0.57%	48.56
Actiontec	0.52%	18.23
Netgear	0.46%	6.54
Acer	0.46%	9.91

Figure 14: Brand analysis- by Urja

% share of a Brand in total products on sale, along with average discount offered on the brand products.			Avg. Discount%
Brand	F	% of Total Count of prices.isSale along Brand	
Viper		0.06%	71.53
Jaybird		0.23%	62.70
M-Audio		0.17%	52.21
Jensen		0.57%	48.56
iLive		0.17%	48.31
Pyle Pro		0.17%	47.78
Polk Audio		0.17%	43.08
Lowepro		0.06%	41.44
Russound		0.29%	39.41
Braven		0.23%	38.83
Sanus		0.63%	38.13
Optoma		0.23%	36.51
Power Acoustik		0.57%	32.73
Sound Design		0.17%	30.58
360fly		0.75%	30.30
Cobra Electronics		0.06%	29.08
StarTech.com		0.29%	27.65
CLARITY-TELECOM		0.23%	26.49
Belkin Inc.		0.23%	25.60
AOC		0.34%	25.11
Linksys		0.40%	23.35
ZTE		0.29%	23.20
Planet Audio		0.06%	22.76
GoPro		0.17%	21.69
ViewSonic		0.46%	21.37
Midland		0.06%	21.35
Rand McNally		0.06%	21.25
RCA		0.06%	21.24
mophie		0.11%	20.66
LG		4.31%	20.29
TCL		0.17%	19.44
Grace Digital		0.57%	19.22
Turtle Beach		0.11%	18.71
ECOXGEAR		0.40%	18.66
Samsung		17.06%	18.58
Pyle		0.40%	18.26
Actiontec		0.52%	18.23

Figure 15: Brand analysis- by Urja

Sale Products Catalogue of Websites with Average Sale Discount			
Prices.We..	Brand	% of Total Count of prices.isSale along Brand	Avg. Discount%
walmart	Viper	0.15%	71.53
	Jaybird	0.60%	62.70
	M-Audio	0.45%	52.21
	Jensen	1.50%	48.56
	iLive	0.45%	48.31
	Pyle Pro	0.45%	47.78
	Polk Audio	0.45%	43.08
	Lowepro	0.15%	41.44
	Russound	0.75%	39.41
	Braven	0.60%	38.83
	Sanus	1.05%	39.91
	Optoma	0.60%	36.51
	Power Acoustik	1.35%	36.37
	Sound Design	0.45%	30.58
	360fly	1.65%	35.81
	Cobra Electronics	0.15%	29.08
	StarTech.com	0.60%	34.57
	CLARITY-TELECOM	0.45%	35.32
	Belkin Inc.	0.30%	51.19
	AOC	0.90%	25.11
	Linksys	0.90%	27.24
	ZTE	0.60%	28.99
	Planet Audio	0.15%	22.76
	GoPro	0.15%	65.08
	ViewSonic	1.05%	24.43
	Midland	0.15%	21.35
	Rand McNally	0.15%	21.25
	RCA	0.15%	21.24
	mophie	0.15%	41.31
	LG	6.61%	34.58
	TCL	0.30%	29.17
	Grace Digital	1.05%	27.45
	Turtle Beach	0.30%	18.71
	ECOXGEAR	0.75%	26.12
	Samsung	24.62%	33.65
	Pyle	1.05%	18.26
	Actiontec	1.20%	20.51

Figure 16: Brand analysis- by Urja

Sale Products Catalogue of Websites with Average Sale Discount			
Prices.We..	Brand	% of Total Count of prices.isSale along Brand	Avg. Discount%
walmart	Samsung	24.62%	33.65
	Sony	12.61%	27.09
	Apple	0.15%	18.31
	SanDisk	0.45%	10.41
	iSimple	0.30%	15.26
	LG	6.61%	34.58
	Razer	1.80%	16.58
	Pioneer	3.30%	30.85
	Sennheiser	1.20%	20.12
	Logitech	0.15%	25.01
	Netgear	0.45%	17.43
	Zmodo	0.30%	13.95
	Nikon	0.15%	25.00
	Lenovo	0.15%	18.70
	Panasonic	0.75%	27.99
	Alpine	0.60%	50.13
	Kenwood	0.90%	33.25
	Canon	0.45%	16.87
	Elite Screens	2.25%	20.35
	Dell	0.45%	9.09
	ASUS	0.45%	12.39
	Olympus	0.60%	17.72
	CORSAIR	2.10%	32.64
	Alienware	0.30%	13.15
	Denon	0.60%	23.43
	JBL	0.30%	24.04
	V-Moda	1.20%	25.35
	360fly	1.65%	35.81
	Onkyo	0.45%	25.79
	House of Marley	1.05%	26.05
	Sanus	1.05%	59.91
	Microsoft	0.45%	15.49
	WD	0.75%	9.90
	TP-Link	0.45%	26.76
	Grace Digital	1.05%	27.45
	Power Acoustik	1.35%	36.37
	Jensen	1.50%	48.56

Figure 17: Brand analysis- by Urja

Most out of stock products belong to the brands offering highest discounts

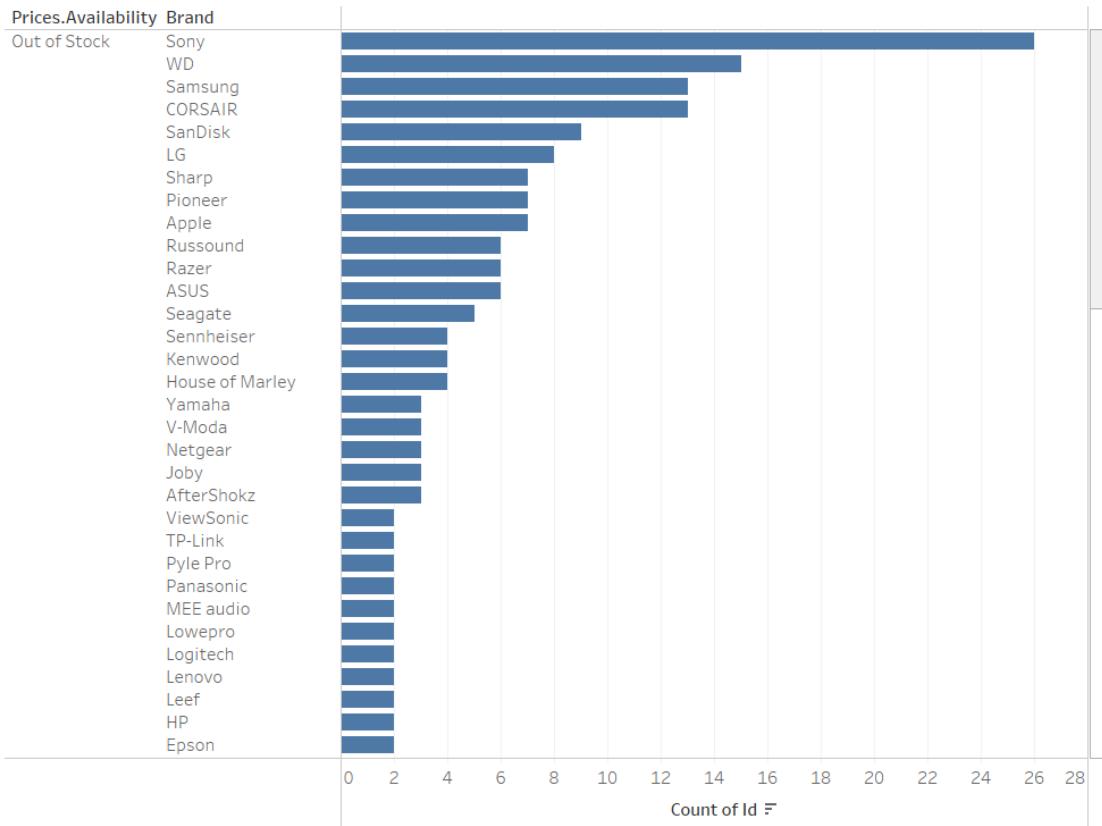


Figure 18: Brand analysis- by Urja

4.3 Brand analysis

Samsung and Sony have a high share of listed products, followed by apple, yamaha, LG , corsair and pioneer.

Fig 14 shows brand sorted by highest product share in the total listed products. Fig 15 shows brand sorted by highest average discount. Fig 16 shows details of brands listed by Walmart, it is sorted by brands with the highest average discount. Fig 17 also shows details of brands listed by Walmart, it is sorted by brands with the highest total product percentage. Most out of stock products belong to the brands offering the biggest average price drops.

5 Conclusion

Visual analysis by Urja - I have explored three different aspects of the dataset namely price, products and brands. I have analyzed each of these aspects in great detail and presented all of my results above.

6 Data Pre-processing/Modelling - By Akshay

The product dataset collected from datafiniti website was in bad condition (noisy) and required fare amount of cleaning. I used Tableau prep builder tool to perform data cleaning and modelling. The process of preparation is described below in detail:

To get a sense of how to proceed with cleaning, I first analysed the schema of dataset with basic python/pandas exploration, and prepared a chart as shown in Fig 19. It helped in understanding the abstract relation between the column features, the datatype of each features, and which are necessary for further analysis. By, this we could also get information on each columns are related in meaning which could be used to fill null values. The unique categories provide info on which values could be merged together if they contain redundant info (eg: "No", "0", values for stock can be merged into "Out of stock" value).

might help in merging/re-arranging

unique Categories

correlation?

merge features based on market segment

Column	Null values?	unique Categories	Type	description	Needed?
id	--	835	hash like	20 chars	?
prices.amountmax	--	2519	Float	Max price listed	
prices.amountmin	--	2668	Float	Min price listed	
Prices.availability	--	14 (merged)	text (low to remain)	Availability types	
prices.condition	--	11 (pre)	Text(keyword) noisy text	cond. of prod when being sold @ this price	
prices.currency	--	2	int	USD, CAD uniform inference?	
prices.dateseen	--	3834 (pre)	Date/timestamp	List of dates when this price was seen	
prices.issale	--	2	bool	Is prod on sale/discount?	
prices.merchant	--	693	text pre - lowercase, stem words	Merchant/website name which is selling @ this price	
prices.shipping	Yes (2972)	74	text + int (pre) - merge catg.	Shipping terms	
price.sourceurls	--	4323	text	List of urls where this price was seen	?
asins	--	835	Alphanumeric id	Amazon identifier for product	?
brand	--	257	text (pre - text uniform)	Brand name of prod	
categories	--	803	text/text/... break it down - sub catg.	List of categories	
dateadded	--	799	date	Date @ which prod was added to DB	
dateupdated	--	737	date	Most recent date this prod price was updated	
ean	Yes (5706)	118	int (code)	European article no.. barcode like for each prod	?
imageurls	--	835	url - can extract source info	List of urls for this prod	?
keys	--	835	hash/text	Internal identifier for DB	No
manufacturer	Yes (4014)	198	text	Manufacturer of prod	
manufacturer number	--	835	alphanumeric	Model no of this prod	?
name	--	828	text (pre)	Product's name	
primary categories	--	5	text *taxonomy avail. in website?	List of categories which prod belongs	
source urls	Yes (233)	811	url demography loc info?	List of urls to gen data for this prod	?
upc	--	392	int (exp format) to uniform unit	Universal prod code.. barcode like	?
weight	--	587	Int + text(unit)	Weight of prod	

Figure 19: Preliminary analysis on dataset schema

6.1 Reducing dimension of data (Grouping)

First of all, unnecessary columns - id, asin (Amazon id), keys (internal id for DB), manufacturer no and upc (barcode) were removed as they don't have any use for further analysis. Since the columns contained values implying the same meaning, the following categories were merged:

Column name	Grouping
Availability	Instock + True + Yes + x available Out of stock + Sold + No + False
Condition	New, Refurbished - merged lower, uppercase words Used + pre-owned
Shipping	Free + Free Delivery + Free Shipping Free after X = Free for orders > x amount Standard + USD X values Cost categories grouped by value range (eg: Cost < 10, Cost 10-15)
Primary categories	Furniture, Automotive = products having respective description
sourceurls	similar urls grouped by common chars eg: reviews.bestbuy.com + www.bestbuy.com amazon.com + amazon.de + amazon.ca
Manufacturer	duplicates grouped based on common chars
Brand	merged similar fields (earphones, earpods, headphones)

6.2 Knowledge discovery (Calculating new features)

The given dataset contained some features such as date.added, day.updated, price min and max etc. But we could still extract latent knowledge from these fields to improve our visual analytics. I calculated the following new features by using info of existing fields:

- Day of the week: This was calculated from date field (separate day fields for both day added & updated). This could provide important info on whether there was more activity on weekends. Similarly I calculated the "month" field from given dates.
- Time to update - This infers the difference between date.added and date.updated. I first subtracted the dates and this difference value was divided by 7 to get data-points in weeks. The reason behind creating this feature was to identify if there is any relation between time it took to update the price (say shelf-time for the product) and sale of that product on sale/discount.

- Discount %: calculated based on price-min and price-max columns. First with subtraction we obtain the price-difference, then by dividing it by price-max & multiplying by 100, we get discount % of that product. This field is valid only for those items which were on sale. Non-sale products had discount zero as values in this field.
- Region of origin: This field is based on currency unit information. The dataset contained currency info on US & Canadian dollars, so the reasoning was that the product could have originated from the corresponding region/country. This feature could be useful for plotting geographical map plots for the products.

6.3 Handling Null values

Only few columns contained nulls in large percentage. Others which had very few null rows were excluded from dataset. Instead of filling nulls by measures such as mode/median of those columns (which could introduce unwanted bias into analysis), I tried to fill the nulls by using info of their other corresponding column fields as explained below:

- Condition: To fill nulls in this column, I checked for keywords such as "new", "used" in their corresponding description field and filled them with those values.
- Manufacturer: This column contained about 20% null values. I observed that the manufacturer and brand features were closely correlated (most of the brand names were manufactured by same company - Apple brand manufactured by Apple itself). So, I filled the null values of manufacturer column with corresponding brand values. There were some exceptions such as eg: "360fly" brand was manufactured by "Audiovoc". But since this should be manually checked and filled, this wasn't done for all the values of manufacturer field.

6.4 Miscellaneous

The dataset had a "categories" column with values in text description. It often contained categories as a mix of multiple values eg: "Electronics, Computer, Computer Accessories, Memory, HDD". It seemed we could benefit by splitting it into different sub-categories instead of using it as a single category field. Since it had ',' char as delimiter, it was easier to split it into different feature columns as sub-categories.

The price_url column was split on "://" & "/" delimiters to extract just the website names from urls. This will split the url into values such as "https", "website.com", "id=x" etc separately. We can just take the website name and discard other values. Similar thing was done to image.url column also.

Some values of "categories" contained long noisy text descriptions, which were grouped with proper categories. Eg: "6.5 Car speakers, for Best bass 135 db ..." merged with "Car speakers" category.

7 Analysis -By Akshay

The main objective of my analysis was to identify the factors influencing the seller to sell a product on discount/sale. The arguments for the analysis which would act as mini-objectives

are as follows:

- To correlate the discount % to features such as "time to update", which was a calculated new feature, depicting how long the product was kept at the same price without any stock change (I assumed this mostly applies to In-stock categories)
- Identifying the role of sub-category of product (Usually some electronic sub-categories are more likely to be on sale in special events, I thought of analysing this by extracting month information from date field and checking it against US/Canada event calendar)
- Variation of discount across different vendors (Do some websites/merchants provide more discount compared to their competitors?)
- Role of Condition of the product (Seller's strategy to sell used/refurbished products)
- Determining if the shipping terms change when the product goes on sale. Eg: Changing shipping from standard rate to Free to lure customers
- Are some brands more open to provide discounts than other?
- Determining if sales are more on weekend days
- Analysing price timeline (change in price across Quarter of the year) filtered on different sub-categories and brands

I tried to analyse the above points in the below sections, although some statements cannot be properly inferred as the dataset doesn't have other vital information such no of products sold, stock changes in diff periods of time etc.

7.1 Analysis of Price Update-time of products

To check how long sellers keep products at same price before changing it, a bar chart was plotted along with average time line as shown in Fig 20. This average line acts as a warning for products being kept in stock for long time without being sold. I am excluding out of stock category here, assuming it being indirectly inferring that the product was sold. The Inferences from this plot (Fig 20) & Fig 21, which is treemap (size of rectangles indicating magnitude of update time) for the same are as follows:

The Average time a product is kept at the same price before providing discount is 88.2 weeks. Product Categories such as Musical, Audio, Speakers, Office are among the longest being kept at the same price. This could mean that these products are not easily sold or maybe some items are becoming obsolete such as speakers and musical, which are somewhat replaced by modern digital devices. Items such as GPUs, VR Gaming devices, Mobiles/iPhones have shortest shelf time - indicating they are sold relatively quicker and the demand in consumer market is more.

These plots are interactive and can be analysed by changing the filters applied on them. Here we are applying stock availability as filter. Fig 22 shows the visualization for special orders.

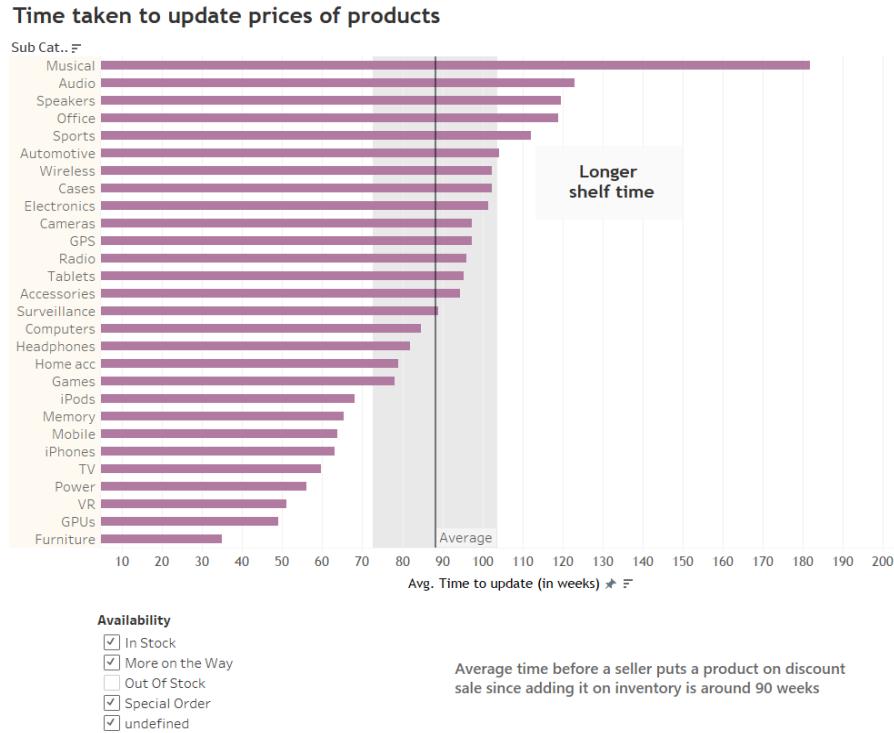


Figure 20: Price Update time across sub-categories (Filtered on stock availability)

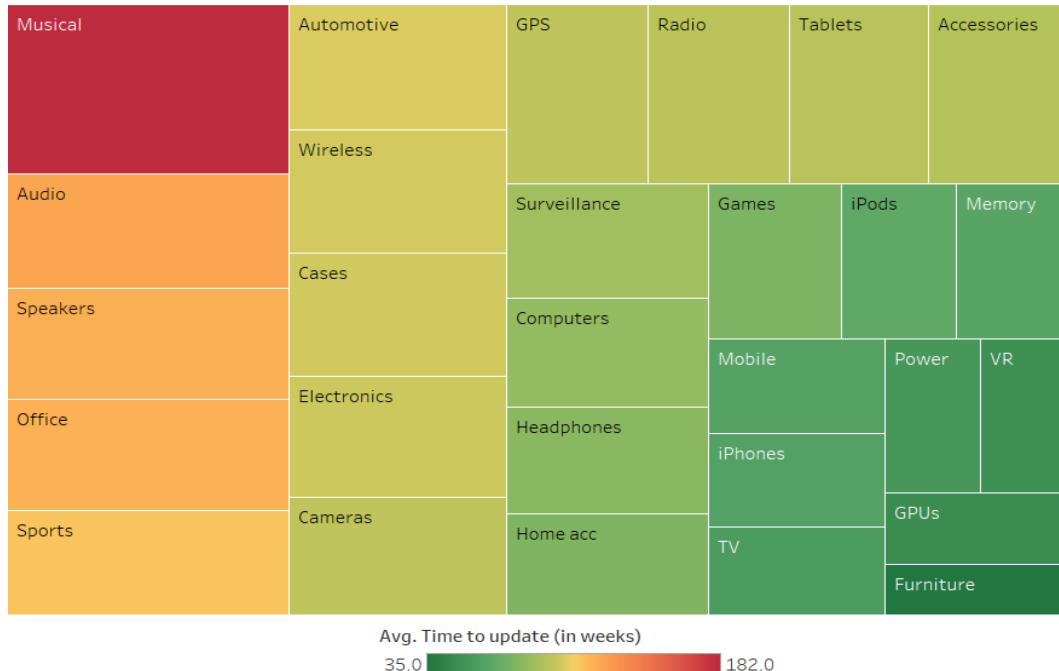


Figure 21: Avg Price Update time across sub-categories

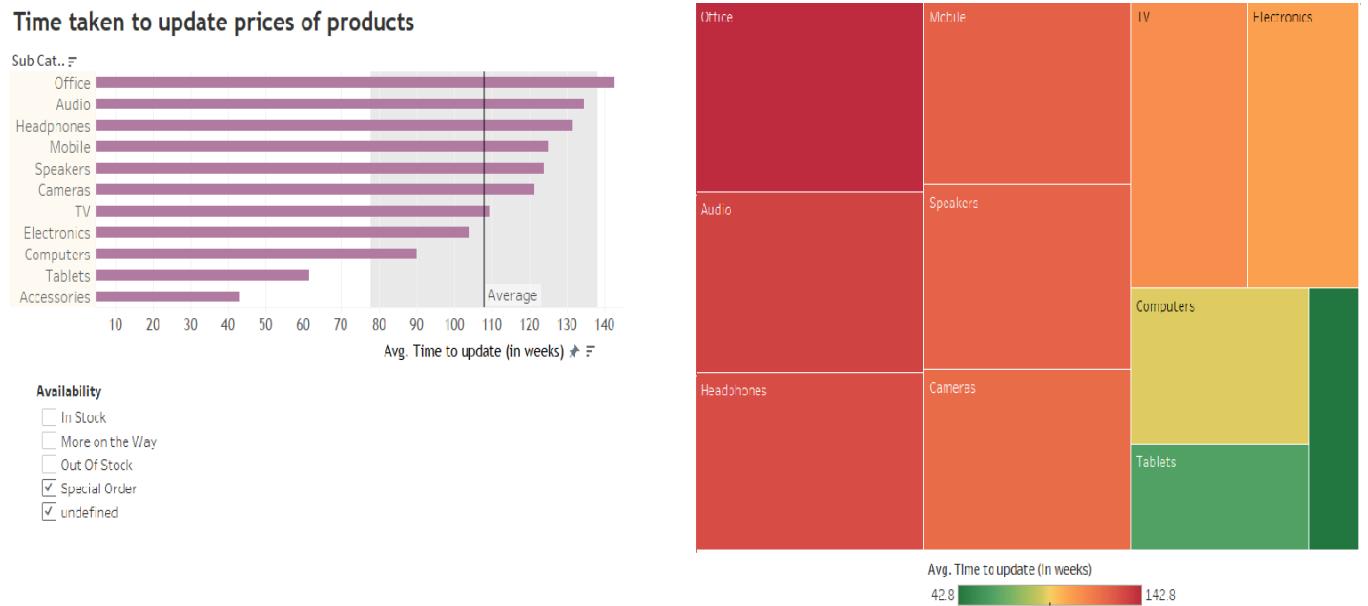


Figure 22: Avg Price Update time for special orders

7.2 Sale/Discount statistics

In this section we can analyse the change of discount price according to different factors such as brand, availability, condition, vendor. These plots are interactive so we can dynamically change these filters to visualize specific info of required fields. There is a discount % adjuster, by which we can see only the products which have certain range of discount.

By selecting dropdown lists of shipping and brand, we can selectively analyse discount of those categories only eg: "discounted products of apple which have shipping as free and from vendor bestbuy, condition as refurbished". With these plots we can check the % count of products on sale & also the price difference for each category when it is on sale.

By Fig 23, the inference was as follows - TVs had the largest number of products on sale at 39.6%, followed by Computers with them having 20.2% of their total products on sale. Coming to price, TVs had highest average price of USD 1787 on sale, followed by Cameras with average price of USD 657. The products whose price didn't change much even when on sale are - Memory devices, Radio, Power and Cases.

It was also found out that almost all products which were on sale are of New condition. This can be seen in Fig 24, where only computers and speakers are on sale when checked on other conditions excluding "new".

Only TVs and Home accessories formed the highest range of discount % from (15%) as shown in Fig 25. Expedited (Fast) delivery shipping made up significant part of on sale products. I didn't find much difference between shipping terms whether the product is on sale or not.

An interesting pattern observed when we selected only some prominent brands such as

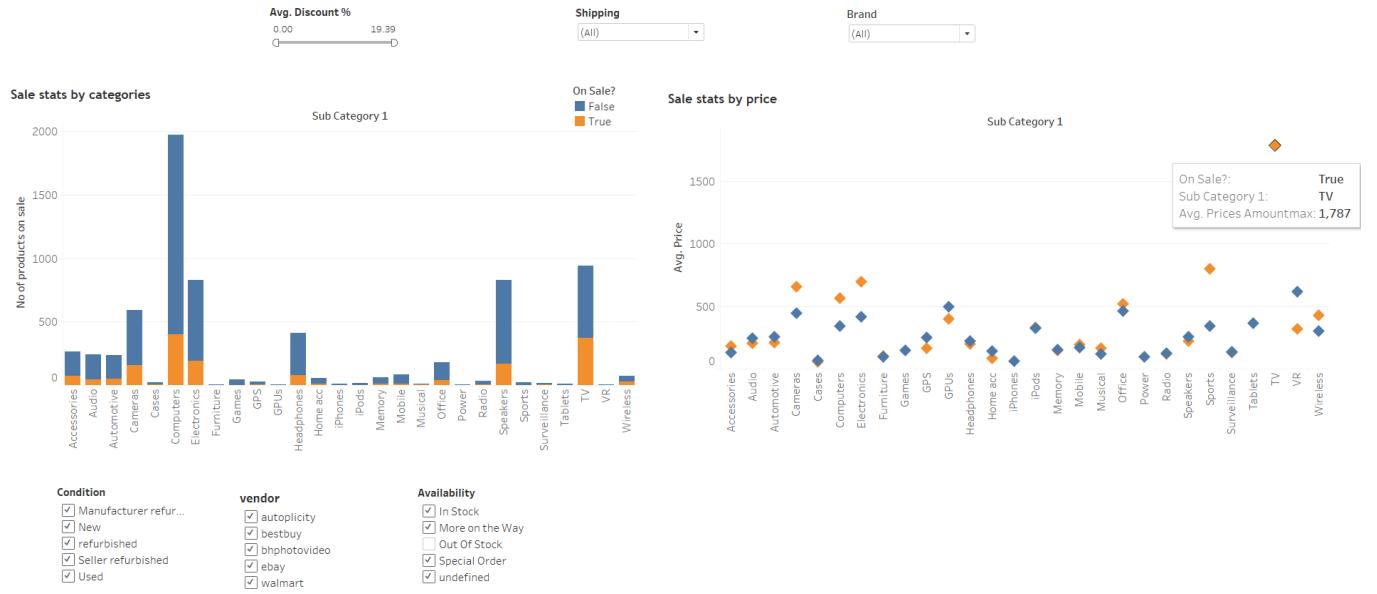


Figure 23: Sale statistics across all brands and vendors (excluding out-of-stock)



Figure 24: Sale statistics - excluding new conditioned products

Apple, Samsung, LG etc, the avg price of sale was higher than the price of product when it wasn't on sale (in Fig 26). This behaviour was observed in some categories generally too, but it was more prominent in popular brands.

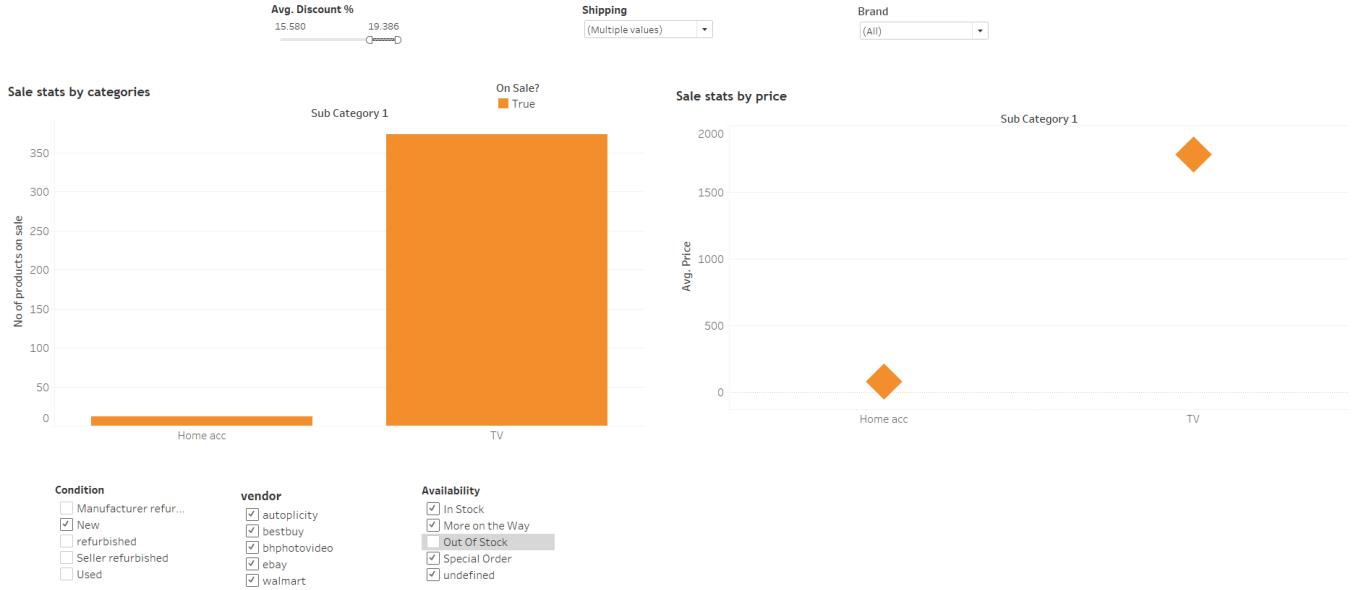


Figure 25: Filtered on high range of discount

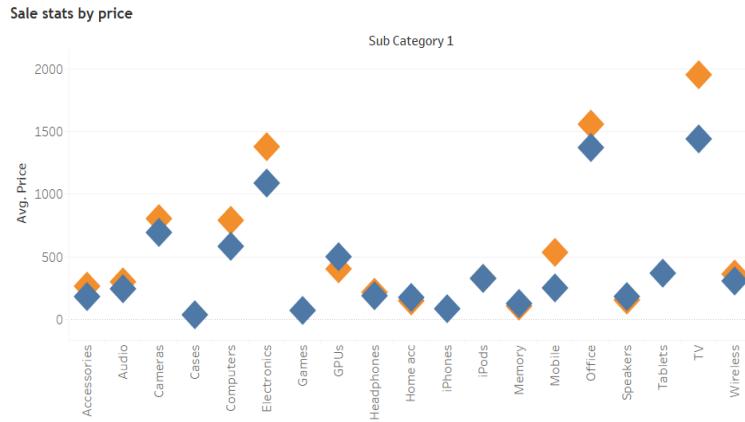


Figure 26: Sale statistics of popular brands

7.3 Discount across Days & item condition

The highest discount% of sale occurred on Saturday with avg. discount% of 3.8%, followed by monday at 3.24% (Fig 27). The lowest being on sunday at 1.35%. This is based on the assumption that "date.updated" field corresponds to the date on which sale was issued.

I also analysed the avg price (not to be confused with discount) on different item condition. It was found that products with "New" condition had highest avg price at USD 468.4 (across all categories) as expected, 19% greater than refurbished ones. The dataset contained ambiguous values of refurbished values, as it contained manufacturer, seller refurbished along with another separate refurbished category (although it should have been the super-set of the both former values). It was found that seller refurbished items had the least avg. price at 157.4 USD.



Figure 27: Avg discount on days of the week

Manufacturer refurbished fared a little better at USD 186.1. This is depicted in Fig 28.



Figure 28: Avg price across different item conditions

7.4 Price timeline analysis

This section consists analysis of price trends wrt different time periods, and also price update timeline for different brands.

7.4.1 Avg.Price v/s Yearly Quarters

The plot shown in Fig 29 also contains trend-lines, summarizing general price trend in each year. We can see that the price trend is upwards in all the years i.e it starts from lower price in

Q1 to gradually increasing to highest in Q4, except in 2017. Year 2017 seems to be an anomaly as prices touched all time high of USD 1056 in Q2 across all product categories.

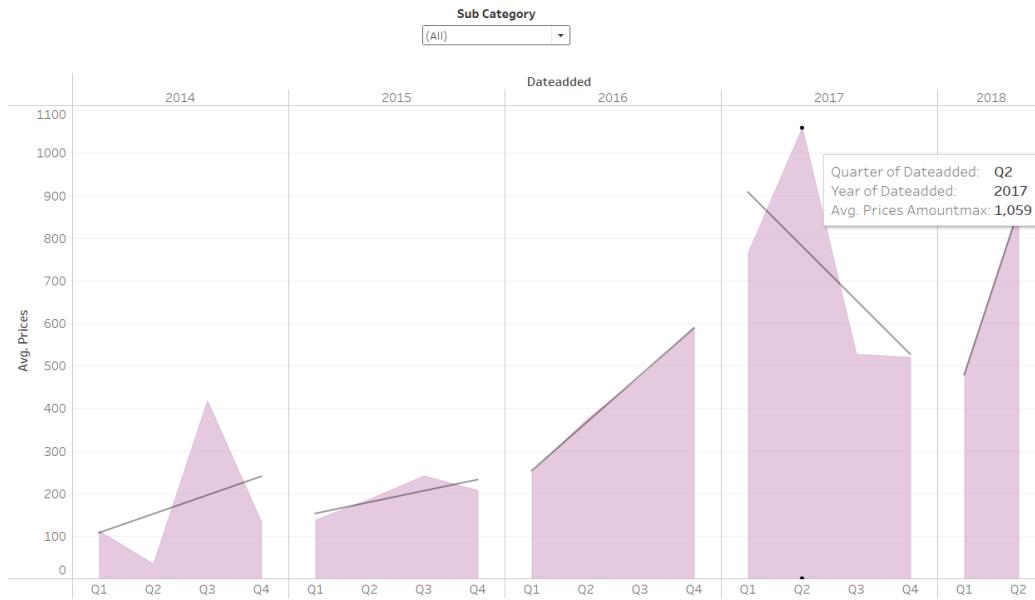


Figure 29: Price-trend across all categories

Camera products had a different pattern as in Fig 30, reaching highest price in 2018 year as opposed to behaviour of other categories.

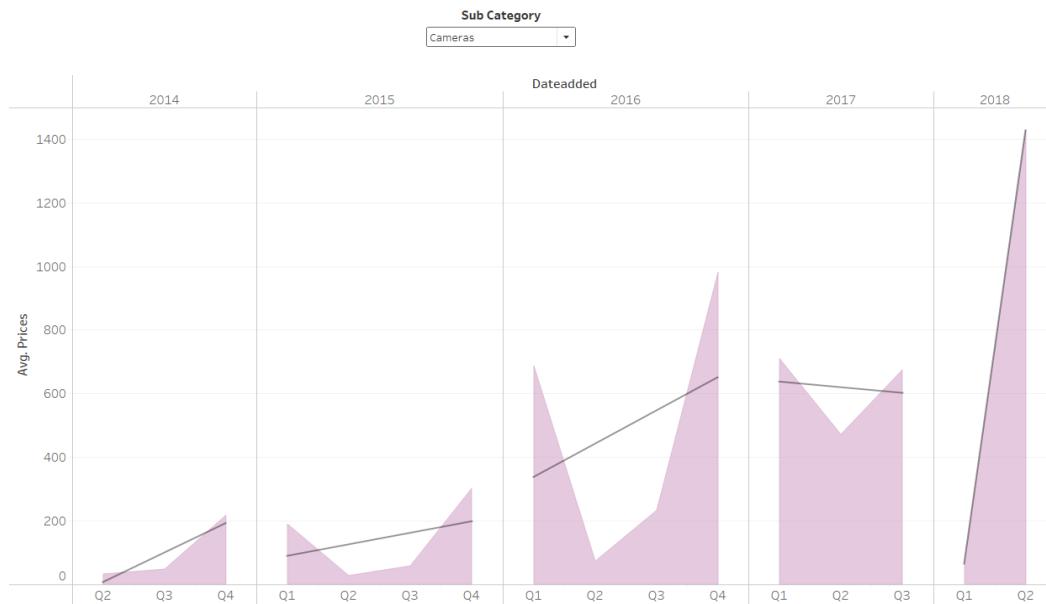


Figure 30: Price-trend of camera products

7.4.2 Price update timeline for different brands

I used a modified version of line chart to visualize timeline between date-added & updated features for different brands. Each band refers to a certain brand depicting year added, year updated and avg time it took to update in weeks. This is shown in Fig 31 (left), it looks clumsy when visualized together for all brands, but it provides a useful info when filtered on individual entities - Fig 31 (right). Here it shows the Denon brand products were added in 2015 and price updated in 2017 at avg update time of 149 weeks across all products of that brand

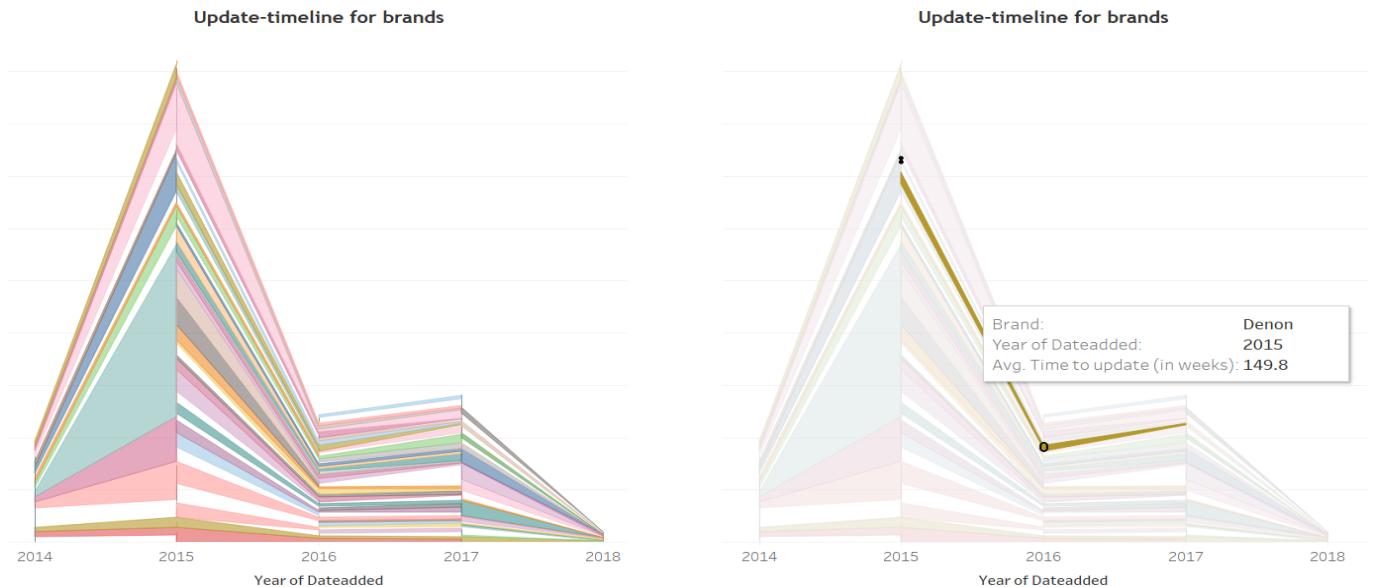


Figure 31: Update timeline (Each colored-band refers to a specific brand)

Fig 32 shows the analysis of specific category (here filtered on headphones). We can see that Fig 32 (a) shows the timeline of adding and updating of prices of different brands in headphones. Fig 32 (b) shows various brands of headphones with prices, this plot is particularly helpful in quickly identifying which brand is costliest in particular years. Here Sennheiser and AudioQuest brands are relatively costlier than their competitors.

Fig 33 shows the visualization of different computer brands. We can see an eg of Dell computers products, which were first added in 2016 and last price updated in 2018. Similarly we can see visualizations for different cameras and TV brands in Fig 34.

7.5 Summary (Inferences) -By Akshay

I have provided detailed inferences in above sub-sections, but I will try to summarize the findings and correlate them to the objectives which I mentioned at the beginning of my analysis. We found that categories such as Musical, Audio, Speakers, Office items were kept longer without price change or giving it on sale. And on the opposite spectrum Gaming/VR, TVs,

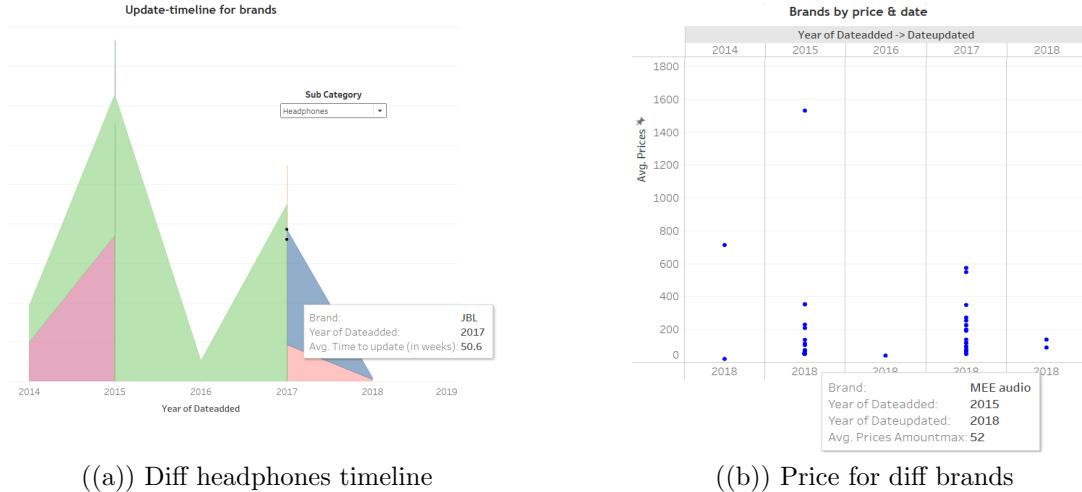


Figure 32: Headphone categories with different brands

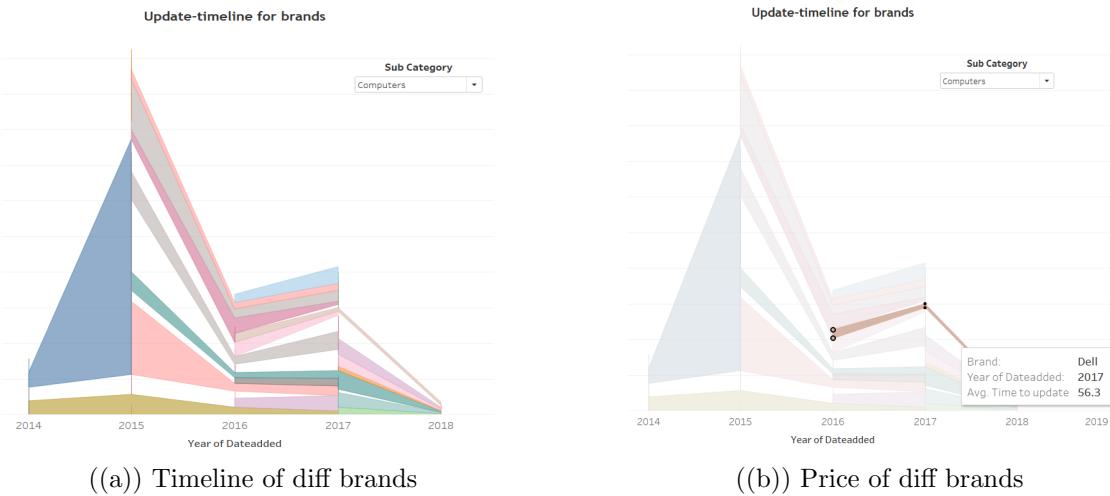


Figure 33: Computer brands

Mobiles were put on sale quickly. Connecting it to our next analysis on sales & discount, the highest avg.discount was given on TVs, which agrees with it being on shorter shelf-time (quickly sold on sale).

Also, the % of products being on sale was greater in widely used items such as Mobiles, computers, TVs etc than products of Furniture, Office, Musical etc. Products such as Memory, Radio, cases didn't see much price difference even if given on sale, probably could be because less users interested in buying them. Almost all of the on-sale products were of New condition. This was contrary to the belief that seller could put a refurbished/used product on sale more likely than new one. We couldn't identify any definite pattern between sale products and their shipping terms.

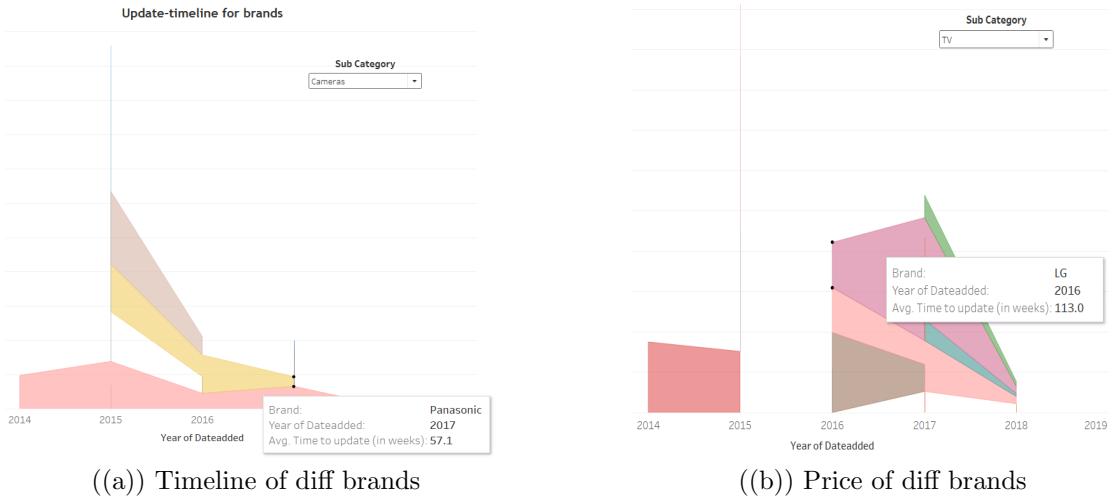


Figure 34: Camera and TV brand visualizations

Avg price of popular brands (Apple, Samsung etc) was higher in sale than normal, which doesn't make sense (could be induced bias while pre-processing or the dataset had prior errors). Highest avg.discount was given on Saturdays' and lowest on sunday. New conditioned items had the highest avg.price which agrees with common intuition. Manufacturer refurbished items seems to sell for higher price than seller refurbished ones, this could be reasoned as manufacturer being rebuilding them with genuine products and the former being more capable of repairing the product.

8 Competition Analysis By Bhashkar

I have analysed how stiff is the competition among various websites and how their selling of products differ due to this.

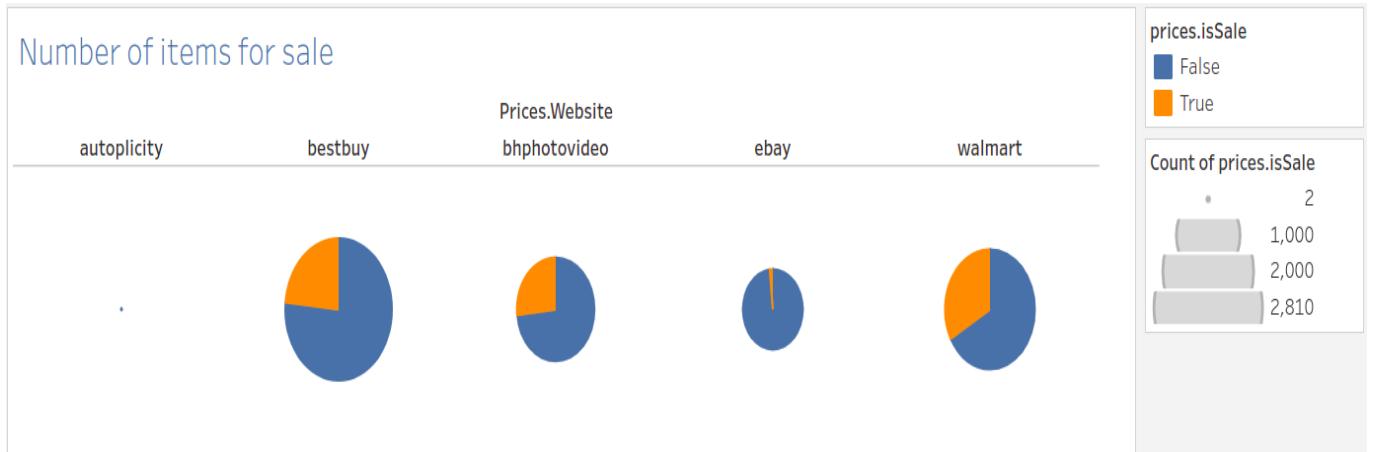


Figure 35: Number of items for sale of each website

From the above figure it can be seen that walmart has put most items for sale among other competitors website then comes bestbuy and bhphotovideo and ebay has the least number of its items for sale. Since autoplicity has only two items it is already out of competition. Also, walmart is the only website giving discount on its products as can be seen in visualisations done by Akshay and Urja. Putting items for sale is a common strategy to get ahead of competitors in this case the only tough competitors which walmart is trying to surpass is bestbuy.

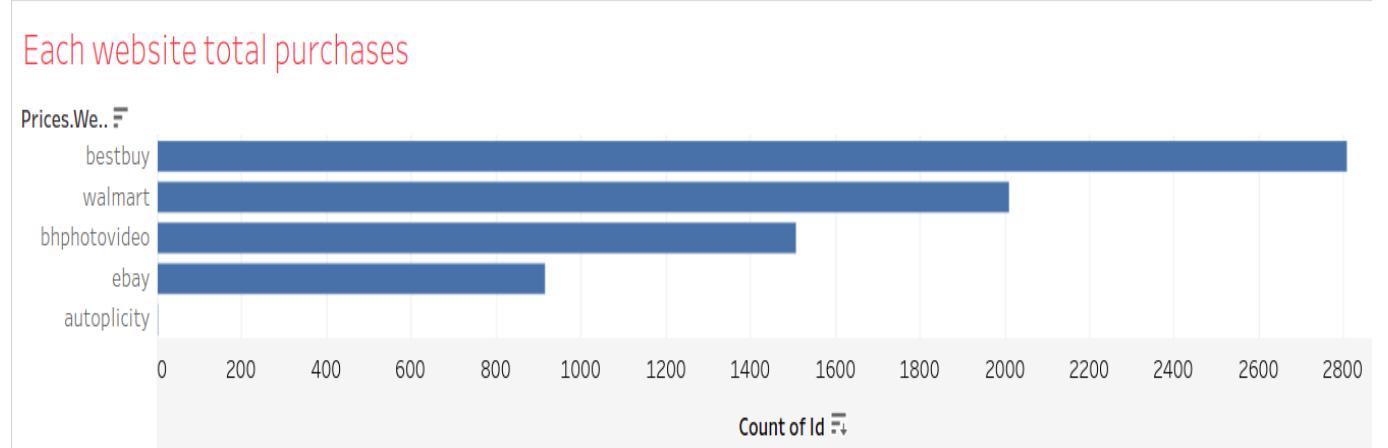


Figure 36: Each Website total products

But bestbuy is far ahead of walmart in terms of product count as can be seen in above visualisation.

Above two visualisations hint that walmart is not as big product selling website as bestbuy but is trying its best to become top website.



Figure 37: Condition of products sold by each website

Above visualisation shows how various websites compete in selling more and more new products. Though ebay and bestbuy is selling all three types of products namely, new, pre-owned and refurbished with bestbuy selling least number of pre-owned, refurbished products but most products of all websites are new. All best buy has almost all of its products as new which is bad news for its competitors.

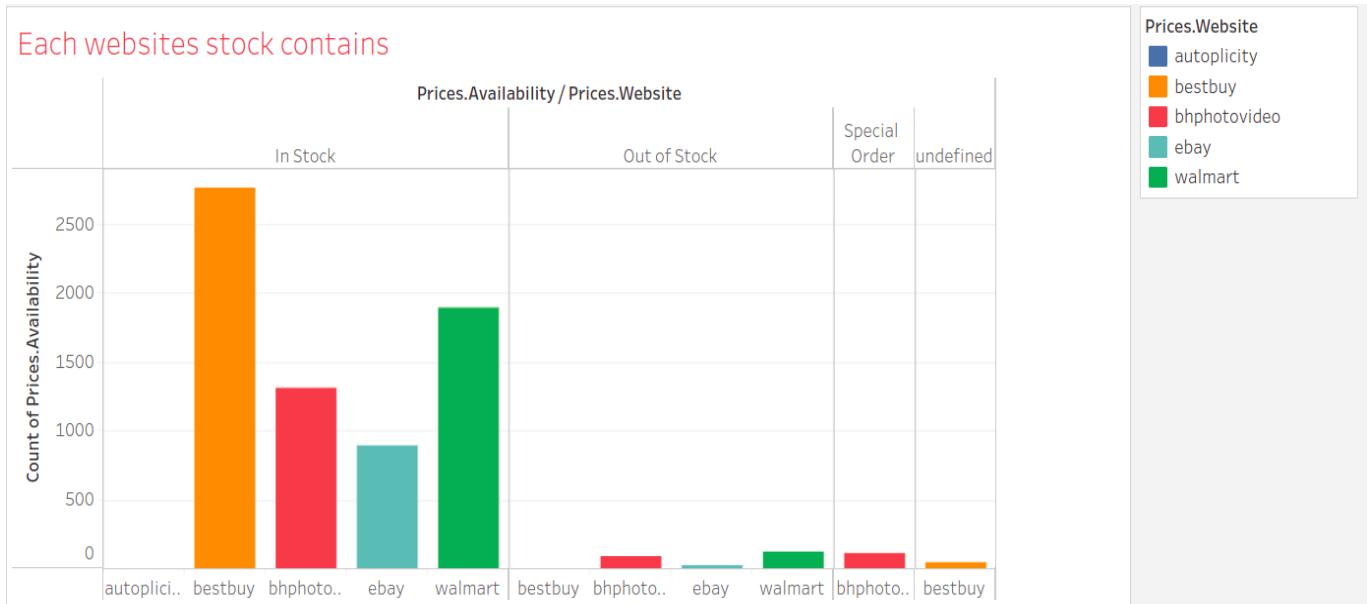


Figure 38: Each Website stock contains

Above Visualisation shows how many products of each website are in stock. Here also, bestbuy is far ahead of its competitors even though it has more number of products but most of its products are in Stock which is again bad news for its competitors. Also, Walmart has most number of out of stock products which is because it was the website who has put most number of its products for sale and when more products are put for sale than chances of it getting out of stock is higher which is what happened in case of walmart. Walmart tried hard to go ahead of its top competitor bestbuy and put more items for sale but did not handled the case of out of stock products properly. Also, bhphotovideo is most number of special order products which means its kind of selling special and rare products. Which is good as it is also its best by trying unique strategy to become successful not by selling more common products but by selling more unique products.

Count of free shipping each website provides



Figure 39: Each Website Free shippings

Above Visualisation shows how all website competes in providing free shippings to its customers. It can be seen that most products of bestbuy shipping charges are undefined it means that it can be free or not. Here, the top website providing free shippings is bhphotovideo which also makes shipping expedited which is a unique feature of it. bestbuy is below par in competition in terms of shipping prices compared to other websites.

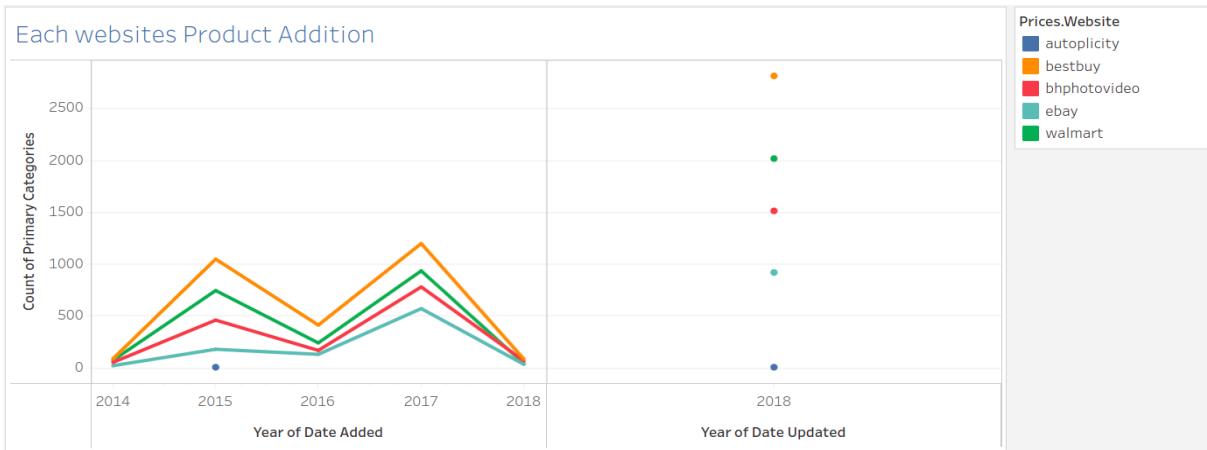


Figure 40: Each Website Product addition/update trends

Above Visualisation shows the product addition/Update trends across websites. Here also bestbuy has added/Updated most number of products each year that's why now it has most number of product for selling as compared to other websites.

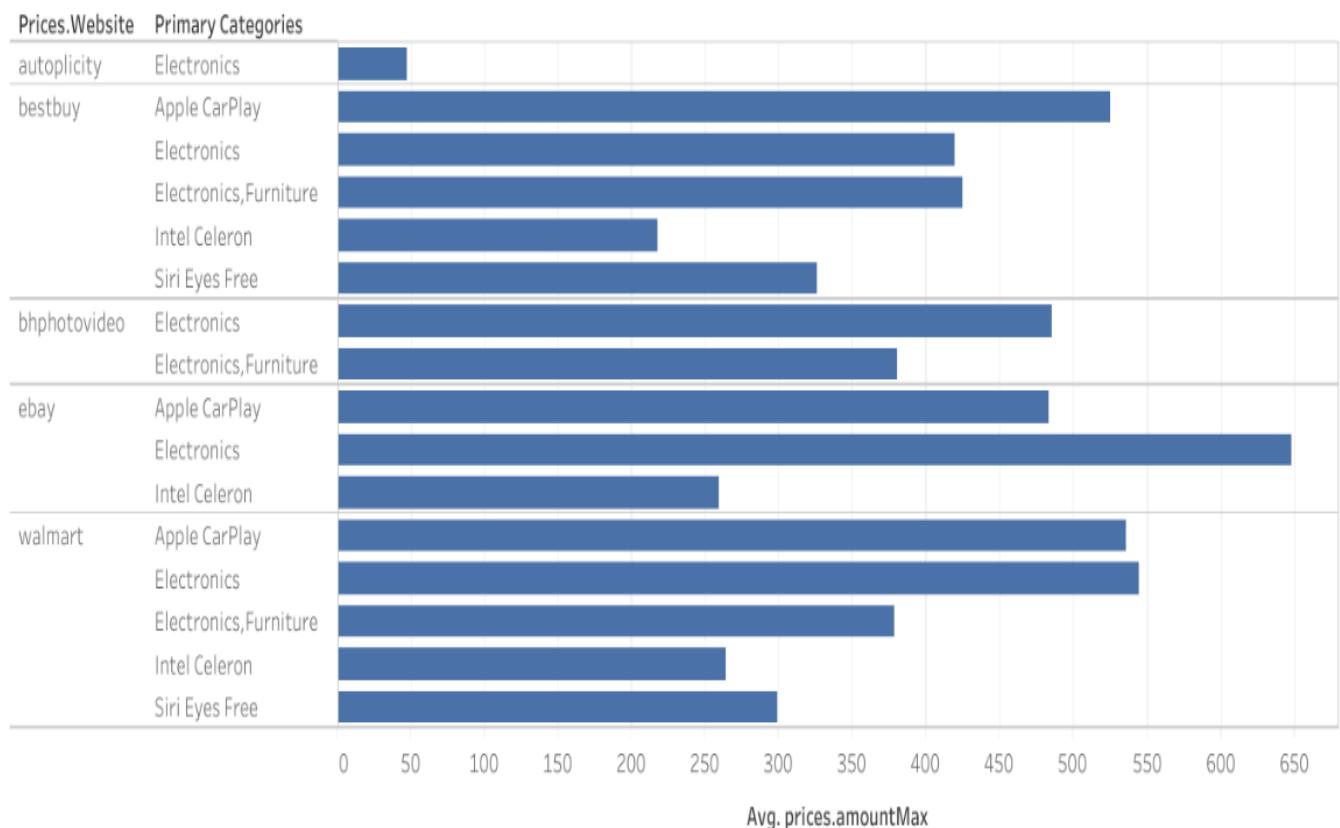


Figure 41: Each Website Product's maximum price trends

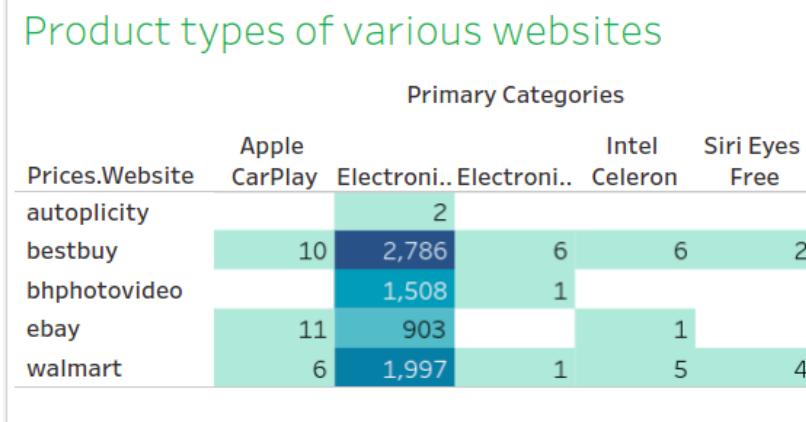


Figure 42: Each Website Product type count

Above two Visualisations show each website competes in selling primary category of products.Like bestbuy and walmart sells all categories of products but here one thing is in common that all websites have abundance of electronics products.And coming to prices the most

costliest product of bestbuy if Apple CarPlay related while other websites costliest product is electronics related. Also, the most costliest electronics products are sold by ebay.

8.1 Summary of Competitive Analysis By Bhashkar

After such thorough visualisation and analysis I have come to conclusion that bestbuy website is the best website when it comes to quantity and quality as bestbuy has most number of products compared to other websites also most of the products it has is in Stock and new.

Walmart comes second in terms of products quantity. While Walmart has most of its products for sale but it is not able to handle the sale load and more of its products are Out of Stock. But its good that all the products of Walmart are new hence in terms of quality and sales walmart is the top website.

bestphotovideo comes third. It has devised a new strategy to overcome its competitors by selling less common products though getting high number of special orders.

Ebay comes fourth with still a progressing website but there a lot which needs to be improved like number of products, decreasing avg. price of electronics items etc.

autoplicity comes last as it has just started selling products with only two products being put up for selling.

References

1. <https://www.tableau.com/>