Zyx Global Corp Technical Report

By: Yan Zhang, Ray Speakmoore, Christian Fernandez

Extraction

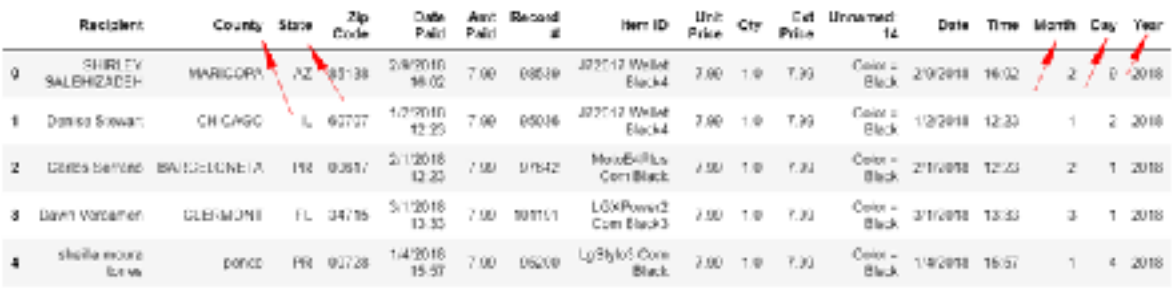
We retrieved the data set from Zyx Global Corp, which is Vicky's online store. The three different data set from various departments handling the product during its consumer lifecycle. These tables include various info about various phone case models. Objective to aggregate and visualize purchase, inventory, and shipping info about a product at a glance.

1)	Store_Order_Detail.csv
2)	Shipping_Cost_Store.csv
3)	Item_Inventory_Price.csv

Transformation

We performed the following cleaning actions to the raw data (CSV files);

- Separated the Date and time from the "Date Paid" column in order to use filter by date only, "Date" column was split in 3 columns to have Month, Day and Year. Then we separated Address into multiple columns: City and State in order to filter independently.



	Recipient	County	State	Zip Code	Date Paid	Amt Paid	Record #	Item ID	Unit Price	Qty	Ext Price	Unnamed: 14	Date	Time	Month	Day	Year
0	SHIRLEY SALEHIZADEH	MARICOPA	AZ	85138	2/8/2018 16:02	7.00	08539	J27217 Wallet Black4	7.00	1.0	7.00	Color = Black	2/8/2018	16:02	2	8	2018
1	Donna Stewart	CHICAGO	IL	60727	1/2/2018 12:23	7.00	05036	J27217 Wallet Black4	7.00	1.0	7.00	Color = Black	1/2/2018	12:23	1	2	2018
2	Gilles Bertaux	BARCELONETA	PR	00617	2/1/2018 12:23	7.00	07642	MotoBillLa Cam Black	7.00	1.0	7.00	Color = Black	2/1/2018	12:23	2	1	2018
3	Dawn Vanderlan	GLENMONT	FL	34736	3/1/2018 13:33	7.00	08111	LGXPower2 Cam Black3	7.00	1.0	7.00	Color = Black	3/1/2018	13:33	3	1	2018
4	sheila moore torres	ponce	PR	00728	1/4/2018 15:57	7.00	05209	LgStylo3 Cam Black	7.00	1.0	7.00	Color = Black	1/4/2018	15:57	1	4	2018

- Rename columns was used to create logical titles for the columns we needed. Unnecessary columns were dropped.

Customer_name	County	State	Zip Code	Date Paid	Paid_Amount	Record #	Item ID	Unit Price	Qty	Ext Price	Model	Purchase_Date	time	Month	Day	Year
[illegible]	MARICOPA	AZ	85138	2/6/2018 [illegible]	7.99	98539	J72017 Wallet Hearts	7.99	1.0	7.99	Color - Rainb	2/6/2018	18:02	2	6	2018
[illegible]	CHICAGO	IL	60707	1/2/2018 [illegible]	7.99	97876	J72017 Wallet Black2	7.99	1.0	7.99	Color - Black	1/2/2018	12:23	1	2	2018
[illegible]	BARCELONETA	PR	00617	2/1/2018 [illegible]	7.99	97647	MotoE4Plus Com Black	7.99	1.0	7.99	Color - Black	2/1/2018	12:23	2	1	2018
[illegible]	CLERMONT	FL	34715	3/1/2018 [illegible]	7.99	101191	LGXPower2 Com Black2	7.99	1.0	7.99	Color - Black	3/1/2018	13:53	3	1	2018
[illegible]	ponce	PR	00730	1/4/2018 [illegible]	7.99	95229	LgStylo2 Com Black	7.99	1.0	7.99	Color - Black	1/4/2018	15:57	1	4	2018

- Filter was used for all item names with pandas.replace function to get all the capable models we use. All item names in the kept the same format and renamed by normalizing different names for same name items.

```
array(['IPHONE 7', 'IPHONE 7 PLUS', 'IPHONE 6', 'IPHONE 6 PLUS',
       'IPHONE 6S', 'IPHONE 6S PLUS', 'IPHONE 8', 'IPHONE 8 PLUS',
       'IPHONE X', 'IPHONE XR', 'IPHONE XS', 'IPHONE XS MAX',
       'SAMSUNG GALAXY NOTE 9', 'IPHONE 5', 'IPHONE 5S', 'IPHONE SE',
       'SAMSUNG GALAXY J7', 'SAMSUNG GALAXY NOTE 4',
       'SAMSUNG GALAXY NOTE 8', 'SAMSUNG GALAXY NOTE EDGE',
       'SAMSUNG GALAXY NOTE 5', 'SAMSUNG GALAXY S6',
       'SAMSUNG GALAXY S6 EDGE', 'SAMSUNG GALAXY S7',
       'SAMSUNG GALAXY S7 EDGE', 'SAMSUNG GALAXY S8',
       'SAMSUNG GALAXY S8 PLUS', 'SAMSUNG GALAXY S9',
       'SAMSUNG GALAXY S9 PLUS', 'LG', 'MOTOROLA',
       'SAMSUNG GALAXY AMP PRIME 2', 'SAMSUNG GALAXY EXPRESS PRIME 2',
       'SAMSUNG GALAXY EXPRESS PRIME 3', 'SAMSUNG GALAXY J3',
       'SAMSUNG GALAXY SOL 2', 'T-MOBILE REVVL PLUS', 'ZTE'], dtype=object)
```

- Group by all items we need by model name

Model	Qty	Paid Amount
IPHONE 7 PLUS	5221.0	38520.53
IPHONE 7	3117.0	19905.49
IPHONE 6S	3035.0	18561.47
IPHONE 6	2823.0	18403.72
IPHONE 6S PLUS	2236.0	14583.57
IPHONE 6 PLUS	2215.0	14005.47
IPHONE 8 PLUS	1824.0	12295.68
IPHONE XR	1506.0	10660.08
IPHONE XS MAX	1256.0	9459.63
SAMSUNG GALAXY NOTE 8	756.0	5742.57
IPHONE 8	627.0	4575.23
LG	592.0	4286.27
IPHONE SE	548.0	4130.45
T-MOBILE REVVL PLUS	536.0	4040.74
SAMSUNG GALAXY S7	483.0	3431.14
IPHONE X	481.0	3387.55
SAMSUNG GALAXY S7 EDGE	393.0	3023.63
SAMSUNG GALAXY NOTE 9	384.0	2789.68
SAMSUNG GALAXY S9	346.0	2583.04
SAMSUNG GALAXY S8	316.0	2414.08
SAMSUNG GALAXY NOTE 5	276.0	2013.93
IPHONE 5S	226.0	2015.12
IPHONE XS	284.0	2104.69
SAMSUNG GALAXY S8 PLUS	232.0	1654.68
SAMSUNG GALAXY S9 PLUS	204.0	1529.14
SAMSUNG GALAXY NOTE 4	246.0	1476.04
SAMSUNG GALAXY S6 EDGE	207.0	1403.47
SAMSUNG GALAXY S6	212.0	1272.01
SAMSUNG GALAXY J3	107.0	1045.32
SAMSUNG GALAXY J7	77.0	731.98
IPHONE 5	94.0	569.69
ZTE	66.0	553.32
SAMSUNG GALAXY EXPRESS PRIME 2	12.0	102.97
SAMSUNG GALAXY NOTE EDGE	15.0	92.02
SAMSUNG GALAXY AMP PRIME 2	8.0	83.24
MOTOROLA	5.0	40.71
SAMSUNG GALAXY GOLD 2	2.0	15.98
SAMSUNG GALAXY EXPRESS PRIME 3	1.0	5.79

- Merge three CSV files and clean. Merge by item id and model. Get three final tables:

Sales Table:

	Customer_name	Record #	Year	Month	County	State	Zip Code	Item ID	Unit Price	Qty	Paid_Amount	Model
4933	Doris Poole	118271	2018	8	EAST HANOVER	NJ	07936	I7 GH Silver	6.49	2.0	12.98	IPHONE 7
4934	darryl martin	118013	2018	7	MCCOMB	MS	39648	I7 Plus GH Teal+ Plus	6.49	1.0	6.49	IPHONE 7 PLUS
4935	andrea knight	98987	2018	2	DETROIT	AL	36552	I7 Plus GH Matt Plus	6.49	1.0	12.98	IPHONE 7 PLUS
4936	andrea knight	98987	2018	2	DETROIT	AL	36552	I7 Plus GH Teal+ Plus	6.49	1.0	12.98	IPHONE 7 PLUS
4937	Heather Giller	90403	2018	1	INDEPENDENCE	KS	67301	I7 Plus GH Teal+Black	6.49	1.0	6.49	IPHONE 7 PLUS

Sales with buying table:

	Record #	Item ID	Qty	Paid_Amount	Model	Buying Price	Total Buying
0	115271	I7 QH Silver	2.0	12.98	IPHONE 7	0.32	0.64
1	115013	I7 Plus QH Teal+Pink	1.0	6.49	IPHONE 7 PLUS	0.32	0.32
2	99187	I7 Plus QH Teal+Pink	1.0	12.98	IPHONE 7 PLUS	0.32	0.32
3	99187	I7 Plus QH Hot Pink	1.0	12.98	IPHONE 7 PLUS	0.32	0.32
4	95483	I7 Plus QH Teal+Black	1.0	6.49	IPHONE 7 PLUS	0.32	0.32

Sales with Shipping Tables:

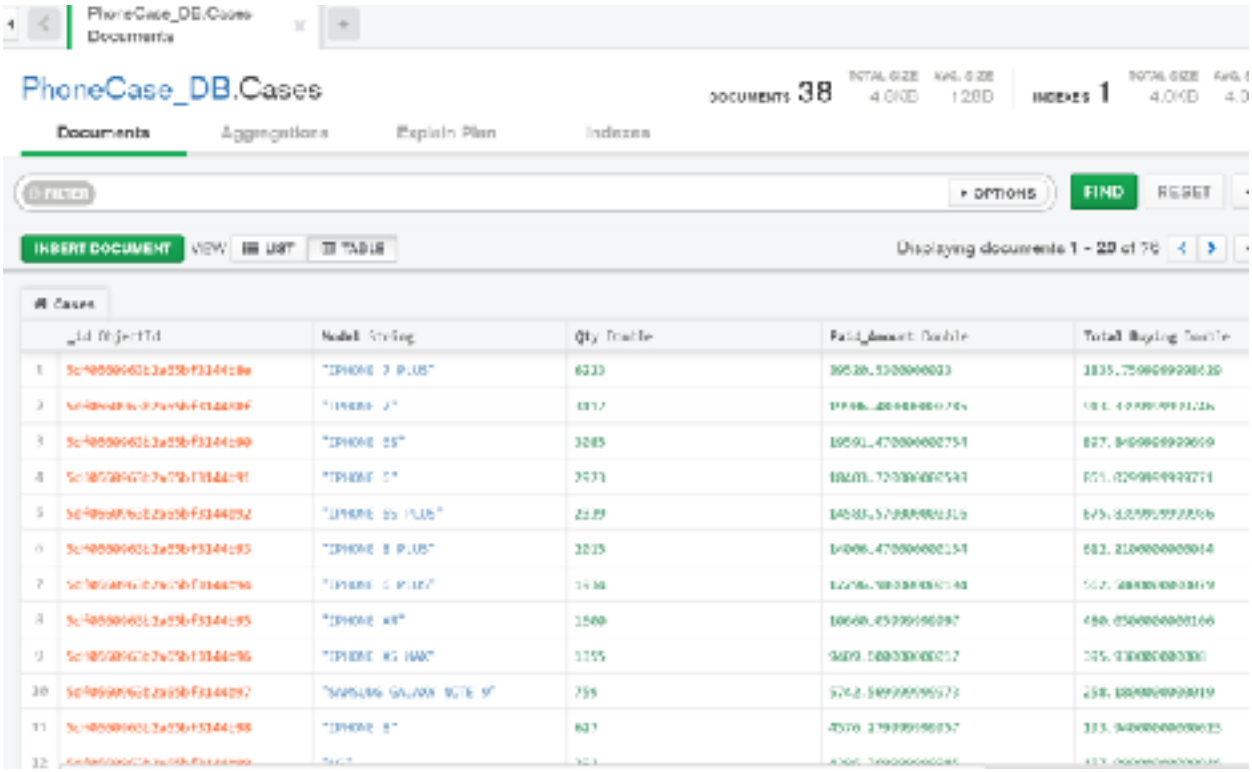
	Record #	County	State	Zone	Qty	Paid_Amount	Shipping Cost	Model
0	115271	EAST HANOVER	NJ	8	2.0	12.98	2.66	IPHONE 7
1	115013	MCCOMB	MS	7	1.0	6.49	2.66	IPHONE 7 PLUS
2	99187	DETROIT	AL	7	1.0	12.98	2.66	IPHONE 7 PLUS
3	99187	DETROIT	AL	7	1.0	12.98	2.66	IPHONE 7 PLUS
4	95483	INDEPENDENCE	KS	6	1.0	6.49	2.61	IPHONE 7 PLUS

After all cleaning we created new tables using the Group by function. We were able to obtain detailed information. For example, “2018 Model Sales Summary”, “2018 Inventory purchase Cost”, “2018 Inventory purchase Cost”.

	Model	Qty	Paid_Amount	Total Buying	Shipping Cost
0	IPHONE 7 PLUS	6223.0	39520.53	1835.76	15886.45
1	IPHONE 7	3117.0	19806.48	913.34	7685.54
2	IPHONE 6S	3065.0	19591.47	897.65	7858.40
3	IPHONE 6	2923.0	18403.72	851.03	7381.60
4	IPHONE 6S PLUS	2339.0	14583.57	675.64	5970.19

Load

After extracting and transforming the data, we then loaded the data into a MongoDB database.



Summary

We used these datasets so we could identify the iPhone 7 Plus and iPhone 7 were the two of the most popular sellers in the store. Apple cases sell better than Samsung cases in 2018. Also iPhone XR was expected to be an important item to our inventory data, but we did not see significant sales on this item. Since the newer item is priced higher than the older iPhone 7/7P, iPhone /8/8P, iPhone X/XR. Then we can assume that Iphone 8/8P may be next year's best seller based on our analysis of the iPhone case popularity for lower priced iPhone cases.

Best Sales: iPhone 7Plus
Best Brand: Apple
Shipping cost most: Zone 8
Most purchasing cost: iPhone 7plus

ETL Project
Group: Yan Zhang, Ray Speakmoore, Christian Fernandez.
2019 USC Viterbi Data Analytics Boot Camp