**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.

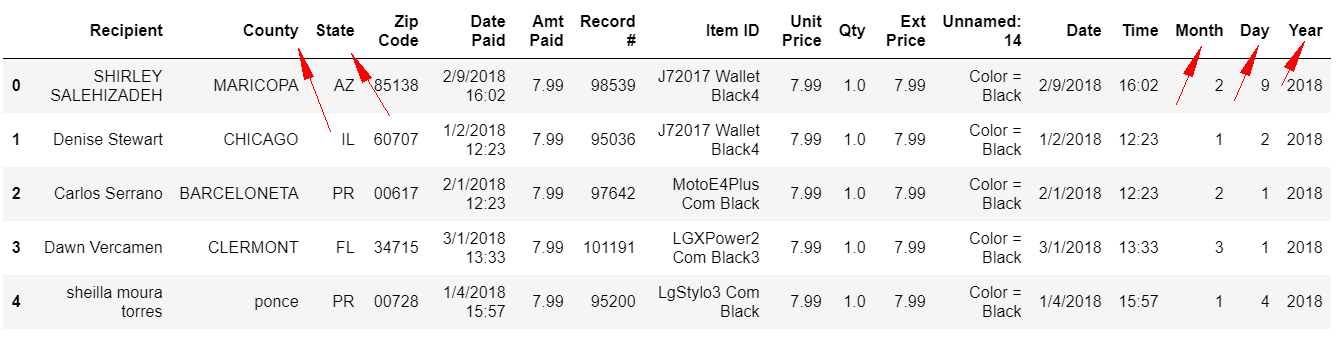
These tables include various info about various phone case models. Objective to aggregate and visualize purchase, inventory, and shipping info about a product at a glance.

|  |  |
| --- | --- |
| 1) | Store\_Order\_Detail.csv |
| 2) | Shipping\_Cost\_Store.csv |
| 3) | Item\_Inventory\_Price.csv |

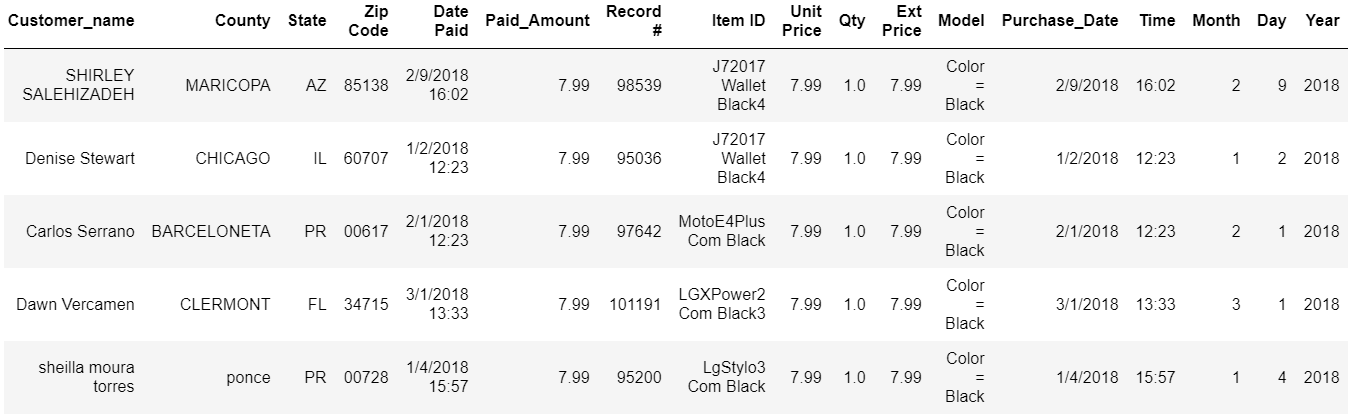
**Transformation**

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

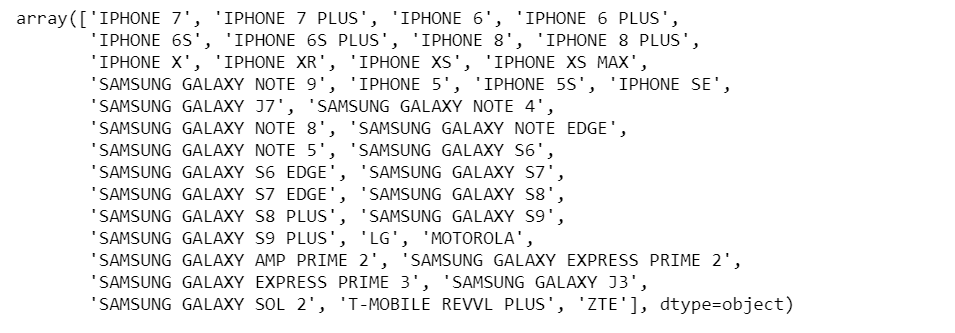
* Separated the Date and time from the “Date Paid” column in order to use filter by date only, “Date” column was split in 3 columns to have Month, Day and Year. Then we separated Address into multiple columns: City and State in order to filter independently.



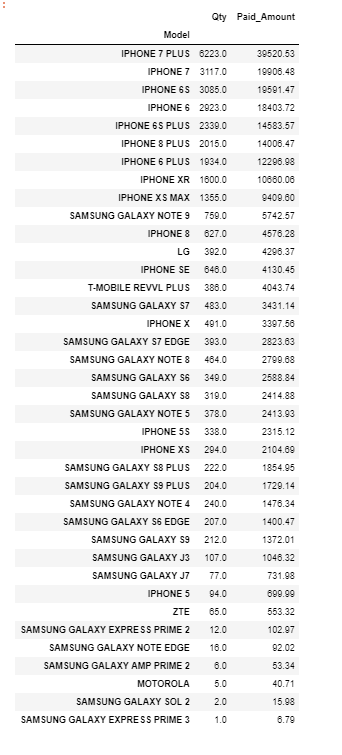
* Rename columns was used to create logical titles for the columns we needed. Unnecessary columns were dropped.



* 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.

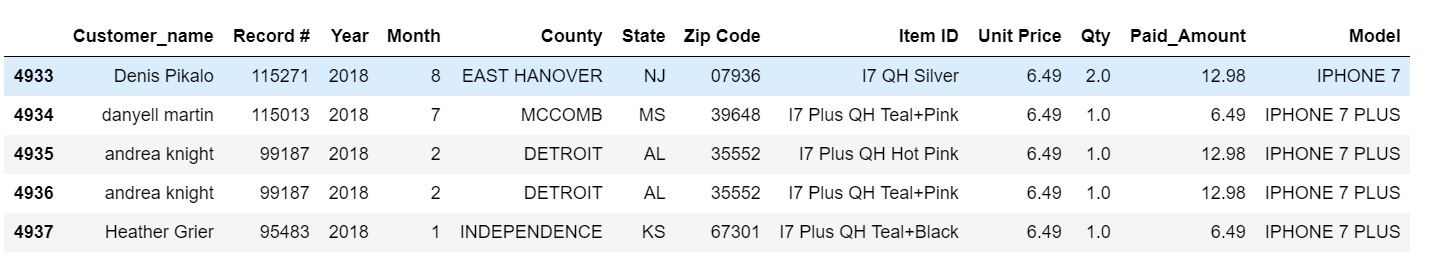


* Group by all items we need by model name

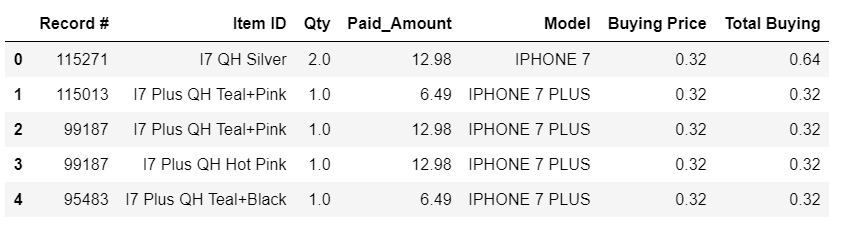


* Merge three CSV files and clean. Merge by item id and model. Get three final tables:

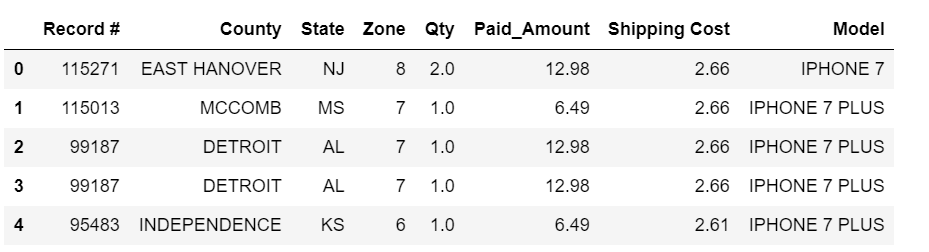
*Sales Table:*



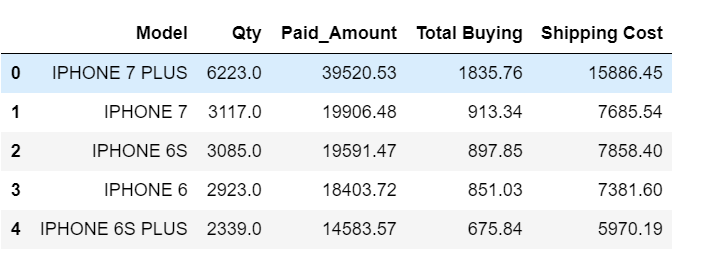
*Sales with buying table:*



*Sales with Shipping Tables:*

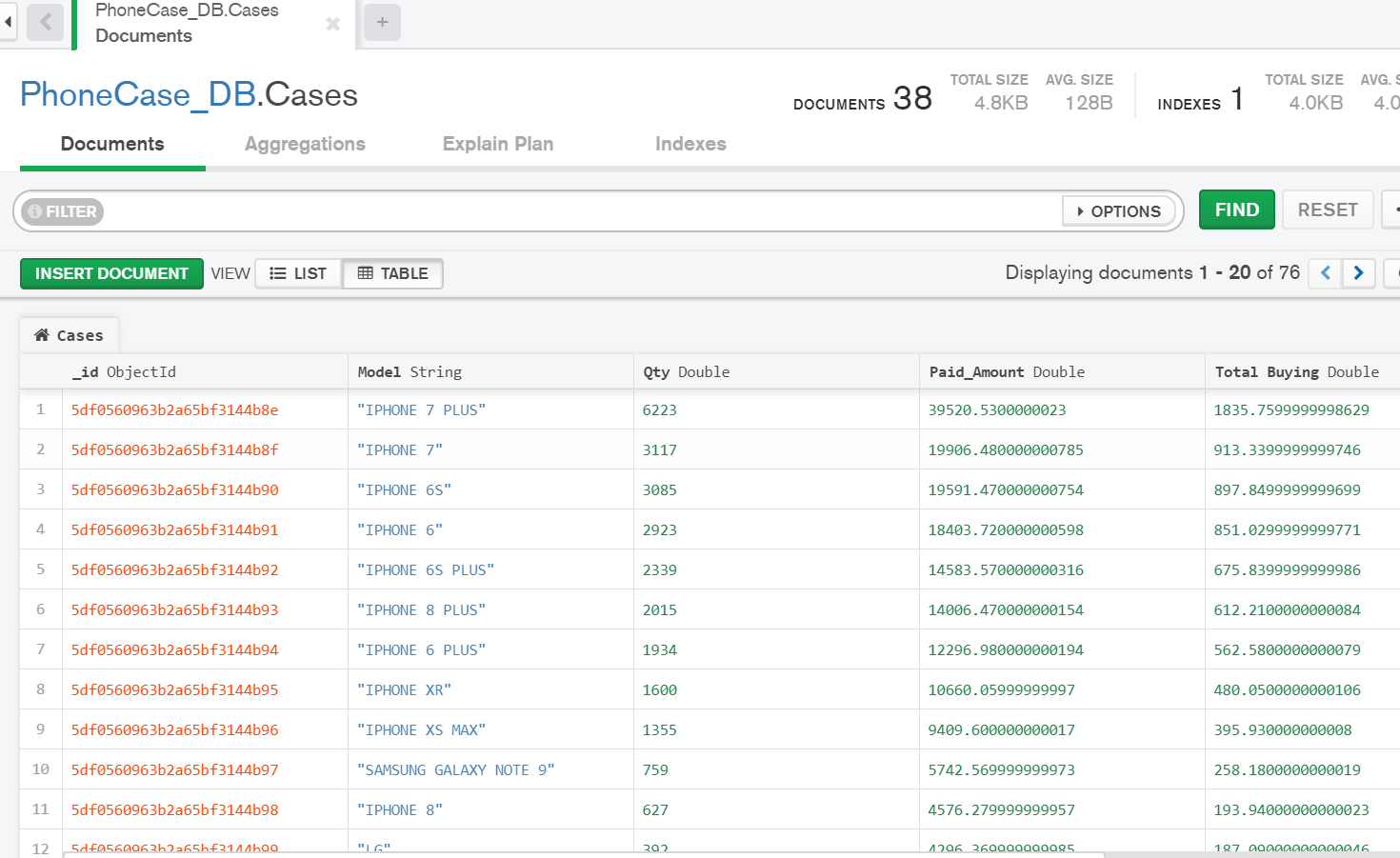


After all cleaning we created new tables using the Group by function. We were able to obtain detailed information. For example, “2018 Model Sales Summary”, “2018 Inventory purchase Cost”, “2018 Inventory purchase Cost”.



**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