**FOOD RETAIL ANALYSIS**

We investigate the relationships between four products: pasta, pasta sauce, syrup, and pancake mix. Two years' worth of purchase data are included in the data set.

The data includes two years' worth of purchase information that has been matched to consumer and brand information. To enhance sales, we want to identify a trend in consumer behavior.

The analysis is focused on maximizing sales using coupons. In retail, coupons are now a crucial tool. By grouping them with already popular things, they are frequently utilized for promotional activities, for fostering consumer loyalty, and for boosting sales of a product.

To find Specific period to concentrate on implement promotion (hour, day of week, and month)

This data has limitation on anonymous datetime then it not clearly analyst. We assume start date and year to simulation and build new column of datetime data (in real situation we should ask from data owner to clarify actual datetime).

Focus points:

* Units sales
* Dollar sales
* Coupon usage
* Store Geography
* Store location
* Product category (commodity)

**PYTHON DATA PREPARATION AND DATA OVERVIEW**

File provided as .csv

Here the relationship between table creates form csv files.

Diagram

Description automatically generated



**Data Overview**

Text

Description automatically generated

Check for missing value using msno.matrix(pandas.dataframe)

A picture containing histogram

Description automatically generated Shape, rectangle

Description automatically generated

A picture containing background pattern

Description automatically generated Graphical user interface

Description automatically generated with medium confidence

**Data Cleaning and manipulation**

* product dataframe

Checked no duplicated.

Text

Description automatically generated

Graphical user interface, text

Description automatically generated

product size is not right deal with it.

Text

Description automatically generated

Brand lists : 131 brands - product have 4 categories

Text

Description automatically generated

* store dataframe

Graphical user interface, text, application

Description automatically generated

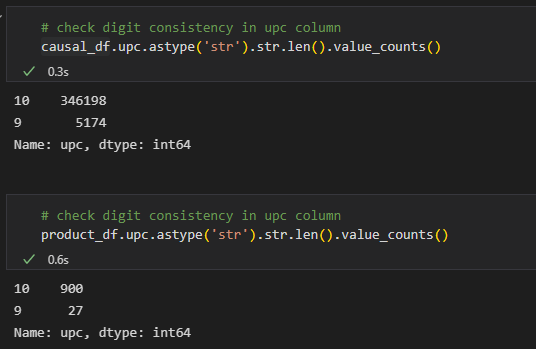
- unique zip code is 299 locations some store is in same zip code total 387 stores.

- zip code tell the location : try cross check with us state zip code

- I assume this is USA zip code (can compare with the list from us zip code: <https://github.com/zauberware/postal-codes-json-xml-csv/blob/master/data/US.zip>)

* causal dataframe

upc number have 10 and 9 digit all the table reference seems not to be any issue here.



Display description.

Text

Description automatically generated

Require more detail on feature description.

Text

Description automatically generated

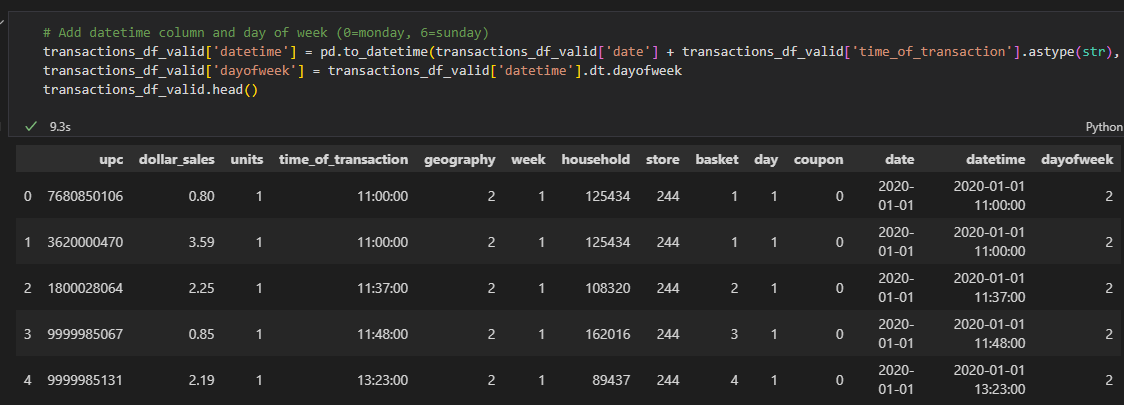
* transaction dataframe

- notice that time digit is not always 4 digits

Text

Description automatically generated

Day and week are anonymous use day number and week number and time to create final datetime column. Assume it to be in year 2020 to 2021



clearly see 2 group of geography store (north and south region)

Graphical user interface, text

Description automatically generated

After cleaning and transform save to new csv to load in Tableau

Text

Description automatically generated with medium confidence

EDA

Coupon usage overall

Text

Description automatically generated

Data summary from each feature or variable or dimension (column)

| **Variable** | **Definition** | **Values** |
| --- | --- | --- |
| basket | unique identifier of a trip made by a Household to a store | numeric key |
| brand | unique name of a Brand to which multiple commodities can belong | character values |
| commodity | unique name of a commodity | pancake mixes, syrups, pasta sauce & pasta |
| coupon | indicates a coupon was used | 1 = Yes, 0 = No |
| day | Chronologically ordered identifier for Day of transaction | numeric, 1 to 728 |
| display\_desc | Product location in temporary in-store display | character values |
| feature\_desc | Product location on weekly mailer | character values |
| geography | identifier for a geography | only two groups, 1 or 2 |
| household | unique identifier for a household(customer) | numeric key |
| store | unique identifier for a store | numeric, 1 to 387 |
| store\_zip\_code | Zip code of the store | standard zip codes |
| time\_of\_transaction | military time of transaction | 4-digit military time |
| upc | Uniquely identifies a commodity of a brand | numeric key |
| week | Chronologically ordered identifier for a week | numeric, 1 to 104 |

Data visualization and get insight.

ML

Sales and Unit forecast with timeseries. Such as:

* use fbprophet model
* ARIMA