



2024 TET HOLIDAY FORECASTING

Exploratory Data Analysis & Model Engineering

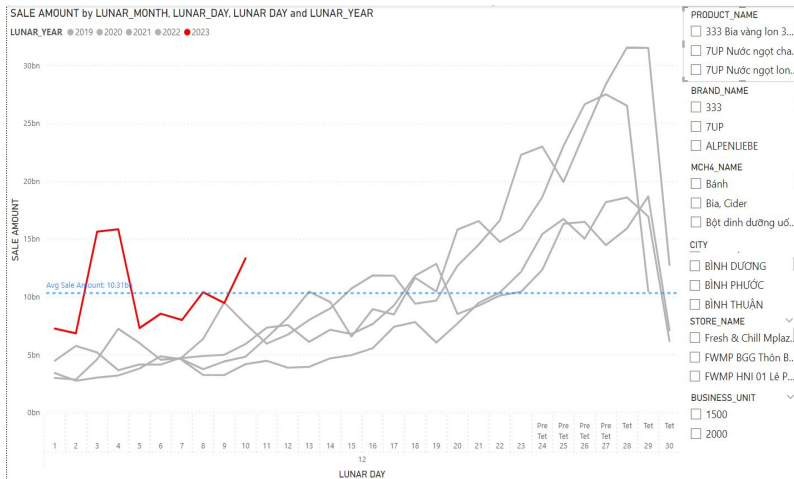


Data collection

I collected historical sales data for Tet holiday-specific SKUs from WCM stores for December (lunar month) for the timestamp of 4 year 2019-2023.

The dataset encompassed information on sales, transactions, product categories, store locations, and dates.

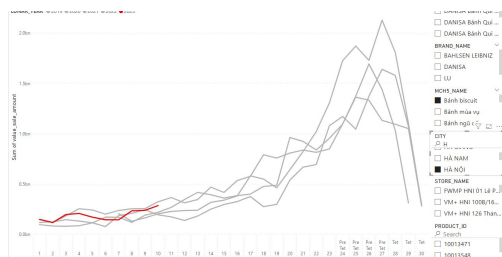
TREND ANALYSIS



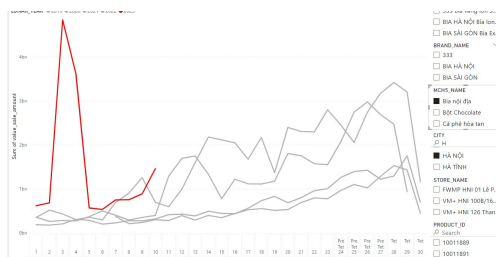
Overall for the whole nationwide, demand for Tet specific SKUs (Beer, Cookies, Chocolate, Wine...) observed a sharp increase from 20th Dec (lunar calendar).

Different products may encounter different purchasing patterns due to customer preference and regional difference.

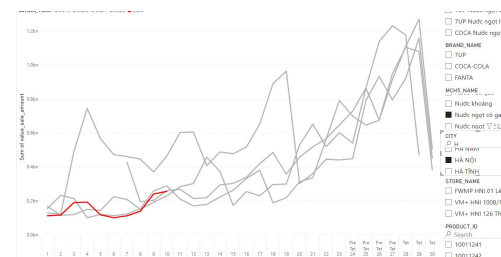
TREND ANALYSIS - by MCH5 Category



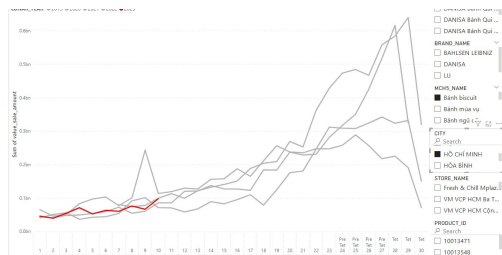
1.1. Bánh Biscuit - Ha Noi



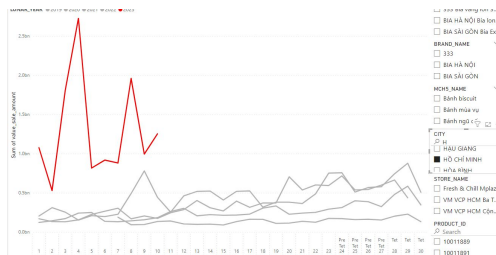
2.1. Bia nội địa- Ha Noi



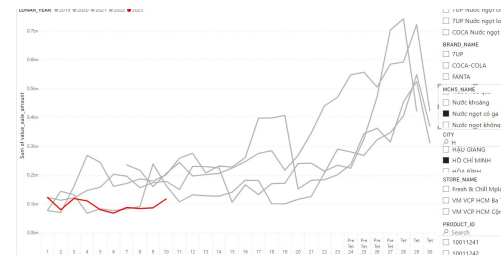
3.1. Nước ngọt có ga - Ha Noi



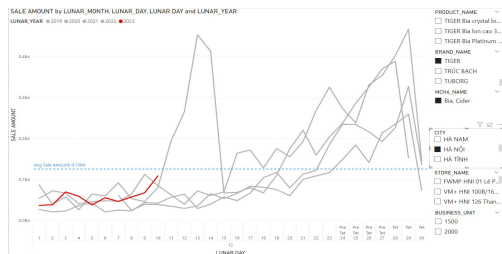
1.2. Bánh Biscuit - HCM



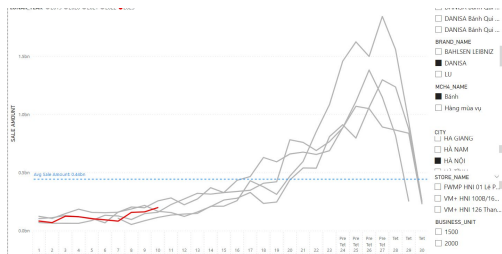
2.2. Bia nội địa - HCM



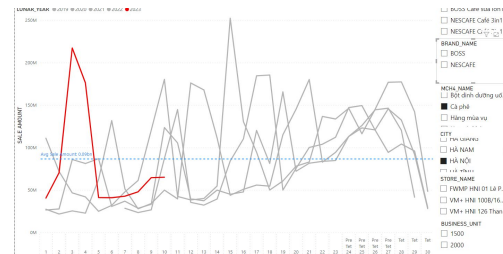
3.2. Nước ngọt có ga - HCM



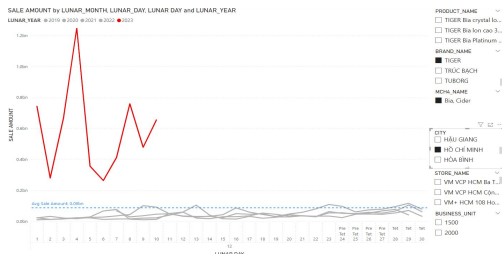
1.1. Tiger - Ha Noi



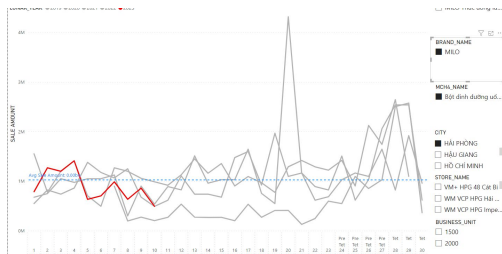
2.1. Danisa - Ha Noi



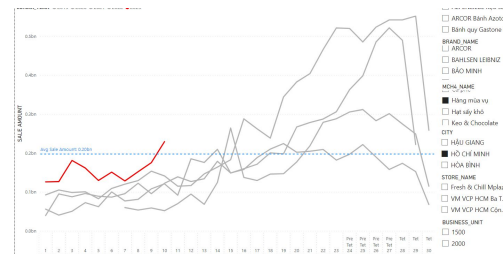
3.1. Ca Phe - Ha Noi



1.1. Tiger - HCM

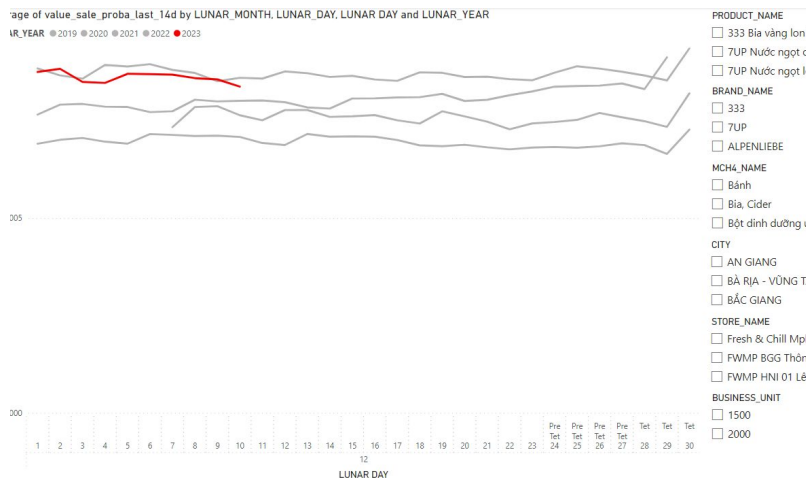


2.2. Milo - Hai Phong



3.2. Seasonal products - HCM

PROBABILITY OF PURCHASING



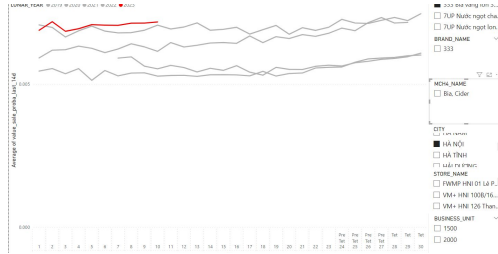
I also check if the lunar new year impact the probability of purchasing.

To determine the probability of purchasing a specific SKU, I counted the total number of visits (transactions) within the past 14 days and calculated the ratio between transactions of the SKU and total store transactions.

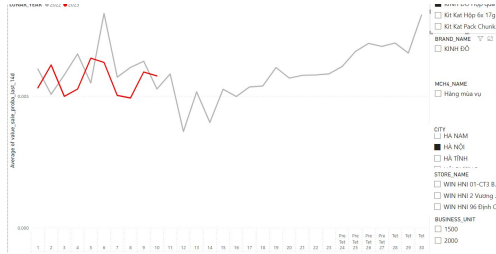
$$\text{proba_of_sale}(\text{sku}_i) = \frac{\text{transactions_last_14d}(\text{sku}_i)}{\text{total_store_transactions}}$$

Overall for all products, the probability of purchasing is quite stable.

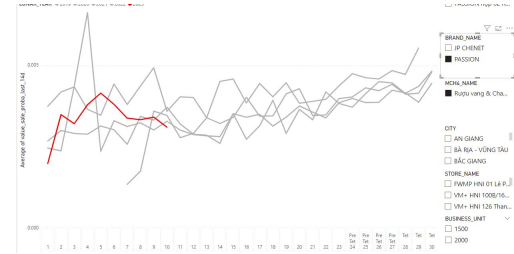
PROBABILITY OF PURCHASING - By Product



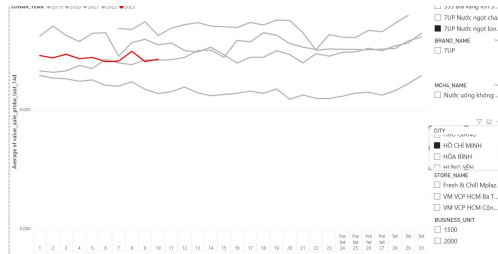
1.1. 333 Beer - Ha Noi



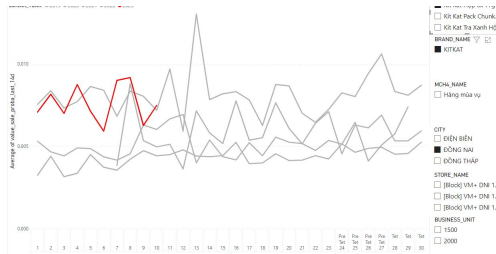
2.1. Kinh do Hop qua tet - Ha Noi



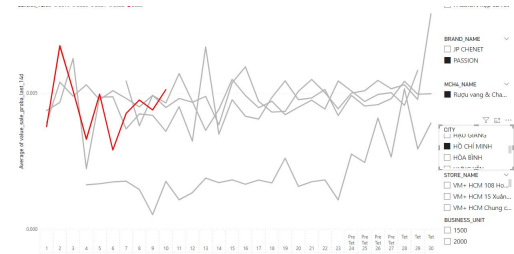
3.1. PASSION Red Wine - Ha Noi



1.1. 7UP - HCM

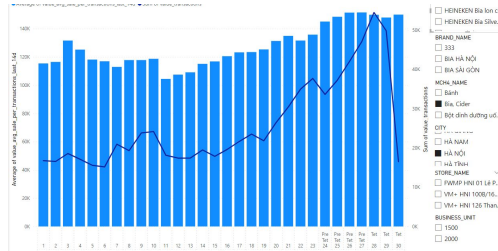


2.2. KITKAT - Dong Nai

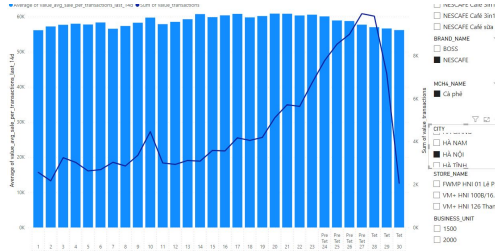


3.2. PASSION Red Wine - HCM

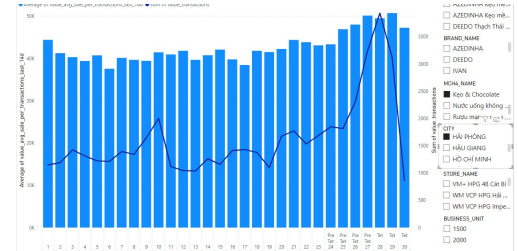
Average Order Value & Visits



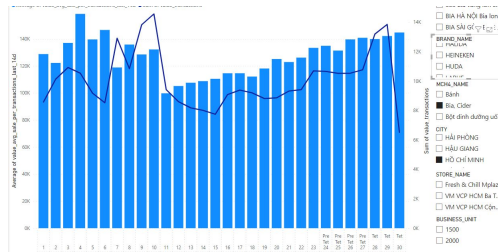
1.1. Bia - Ha Noi



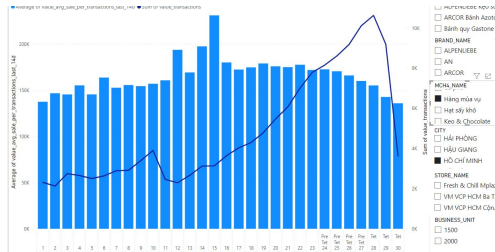
2.1. Coffee - Ha Noi



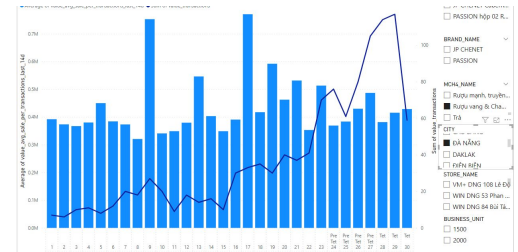
3.1. Keo & Chocolate - Hai Phong



1.1. Bia - HCM



2.2. Seasonal products - HCM



3.2. Red Wine - Da Nang



EXPLORATORY DATA ANALYSIS

Key takeaways

Impact of Lunar New Year Holidays

The analysis revealed that the Lunar New Year holidays have a dramatic impact on demand, particularly in terms of the number of visits to WCM stores for purchasing items in preparation for Tet. This surge in visits indicates the significance of the Tet holiday season in driving consumer demand.

Probability of Purchasing

Contrary to expectations, the probability of purchasing remained relatively stable across different products. This finding implies that the increase in visits during the Tet holiday season does not significantly affect the likelihood of customers making a purchase.

Average Order Value

For Tet-related products such as beer, wine, and cookies, there was a noticeable increase in average order value. This finding suggests that customers tend to purchase larger quantities of these products during the Tet holiday season compared to other periods.



FORECAST APPROACH

From historical data, we know how much of each category/SKU we will sale at each day prior to Tet (30th December - Lunar Calendar)

product / MCH5	city	lunar month	lunar day	% rev
red ruby / Bia nội địa	HN	12	1	1%
			2	2%
			3	1%
		
			30	20%

Suppose that we are at day 20th December lunar calendar, we know how much we already sold of each product at each store in the past 14d

Forecast(red ruby / HN / store 1565 / day n) = Actual (day 6th-19th) / % Rev (day 6th - 19th) * %Rev (day n)



FORECAST APPROACH

To forecast sales for each SKU at the store and daily level during the Tet sale period, I developed a forecasting model. The model incorporates the average sale per transaction, probability of sale, and forecasted store transactions.

Formular:

Sale_forecast(sku_i, store_i, d) = Avg_sale_per_transaction_last14d(sku_i, store_i) * Proba_sale_last_14d(sku_i, store_i) * **Forecast_store_transactions(store_i, d)**

Forecast_store_transactions(store_i, d) = Avg_store_trnx_last_14d (store_id) * avg_uplift (d) * avg_uplift(dow) * **Risk_control(store_id, city)**

Risk_control(store_id, city) = uplift_sale(sku_id, d, city) * contribution_sale_of_store_by_city_last_14d(store_i, city)

In which: d is lunar calendar date



SIMULATION FORECAST

Suppose that we are at day 20th Dec 2022 by Lunar Calendar (11th Jan 2023), I used all past data for the Tet season (lunar month is december only) and forecast for the next 10 days from 20th - 30th Dec 2022 (lunar calendar).

Results:

Aggregate at store level for all Tet products: Accuracy is > 51% averagely for all stores (> 3000)

Aggregate at product level: Accuracy is 49% averagely for all products (200)

For more details, please check the following link:

Agg results: https://docs.google.com/spreadsheets/d/1dx9_Ml-vnvHJgyUXcy8fPRZFzkXBMvuYvO7ln9yg4/edit?usp=sharing

Raw results: <https://drive.google.com/file/d/1TI60x2hw9ipC7healJ1hi9CFBxeilxda/view?usp=sharing>