DOTDOTDOT

THEME 3:

ECONOMY PRICE CATCHER

PRICE PREDICTION MODEL BASED ON HOUSEHOLD ITEMS' PRICE TRENDS

TEAM MEMBERS:

AQILAH SALIHAH
ANIS MUNIRAH
ZAHRA AZALEA

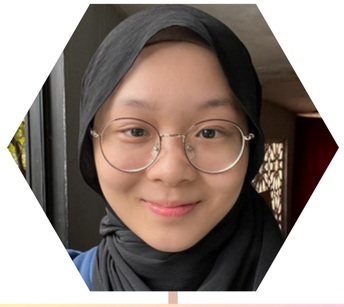
INTRODUCTION



AQILAH SALIHAH



ANIS MUNIRAH



ZAHRA AZALEA

OUR BACKGROUND

We are all 2nd-year Data Science students.

KEY METHODOLOGY

- Data Integration
- Data Preprocessing
- Hypothesis Testing
- Data Visualization
- Prediction Modeling

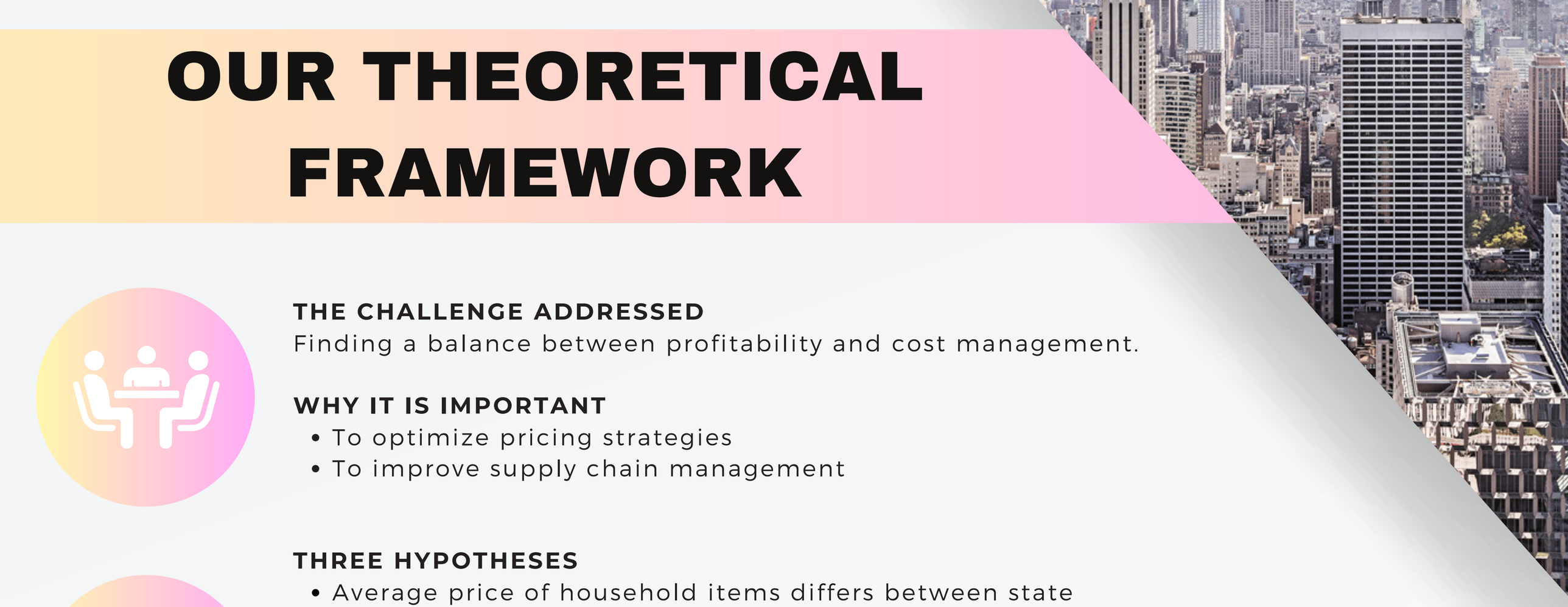
OUR MAIN FOCUS

To incorporate key findings into a business strategy to maximize profit and improve cost management.

KEY FINDINGS

- The average price of item changes by state due to living costs.
- Different types of premises offer different average prices of items.

OUR THEORETICAL FRAMEWORK



THE CHALLENGE ADDRESSED

Finding a balance between profitability and cost management.

WHY IT IS IMPORTANT

- To optimize pricing strategies
- To improve supply chain management

THREE HYPOTHESES

- Average price of household items differs between state
- Different types of premises have varying impacts on item prices
- Prices of household items increases over time

WHY IT IS RELEVANT

- To understand market trends and make informed decisions to avoid loss.
- To improve business strategy for better future performance.

DATA DESCRIPTION

Data Preprocessing

1. Data Loading: loaded from external sources using the ‘pd.read_parquet’ function.
2. Data Concatenation:concatenated using ‘pd.concat’ to create one combined DataFrame named priceCatcher, which contains data for multiple months.
3. Data Merging: prices catcher data was merged with lookup datasets on the item and premise code.
4. Column Dropping: Several columns are dropped from the DataFrame using PC = PC.drop(columns= ['address', 'premise_code', 'item_code', 'premise', 'item', 'unit'])

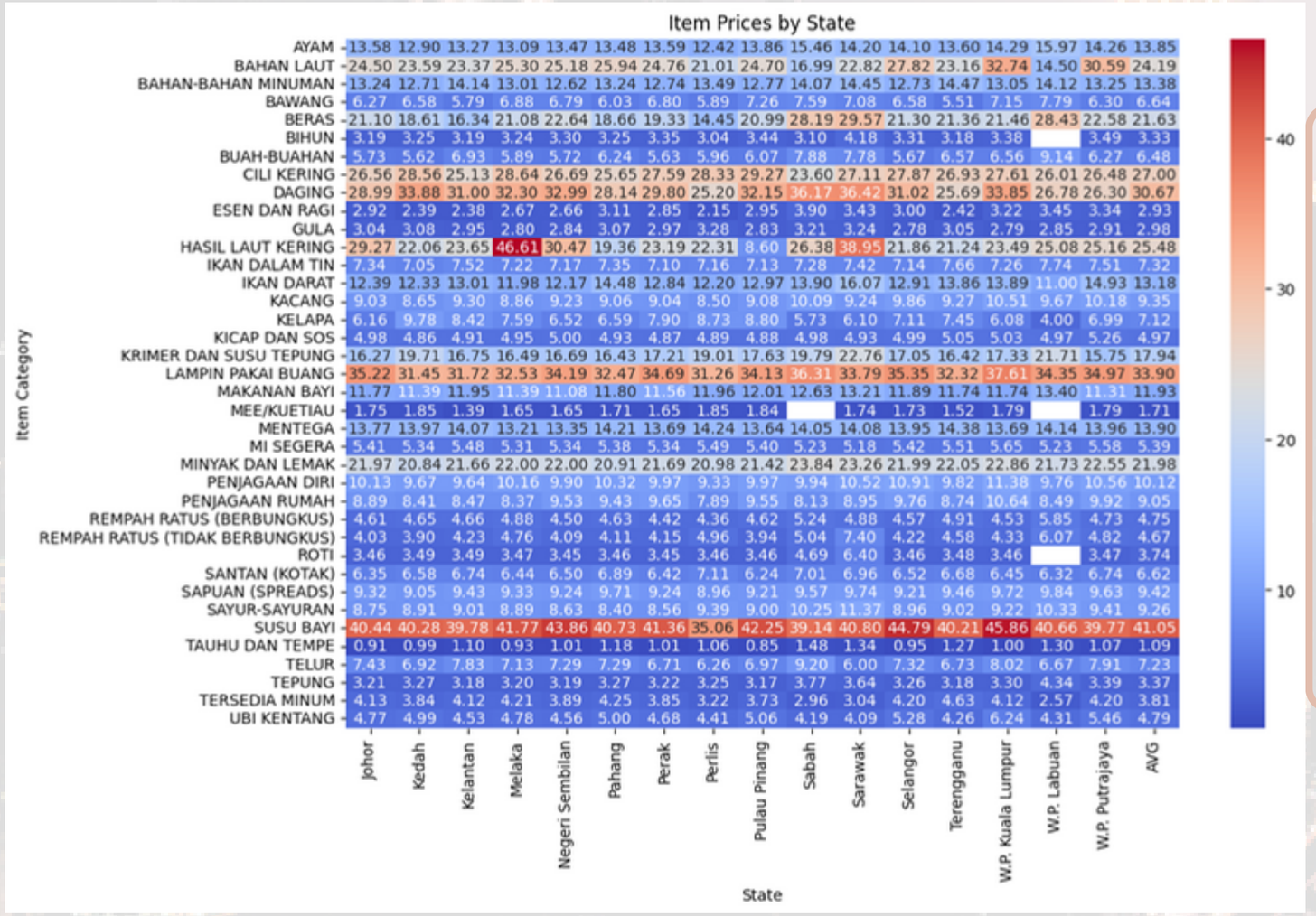
Datasets used

- pricecatcher_2023-09
- pricecatcher_2023-08
- pricecatcher_2023-07
- pricecatcher_2023-06
- lookup_premise
- lookup_item

Missing data handling

a. Checking for Missing Data: ‘print(PC.isnull().sum())’ is used to display the count of missing values for each column.
b. Dropping Rows with Missing Data: Missing data is handled by dropping rows with missing values using ‘PC = PC.dropna()’. After this step, the DataFrame no longer contains rows with missing values.

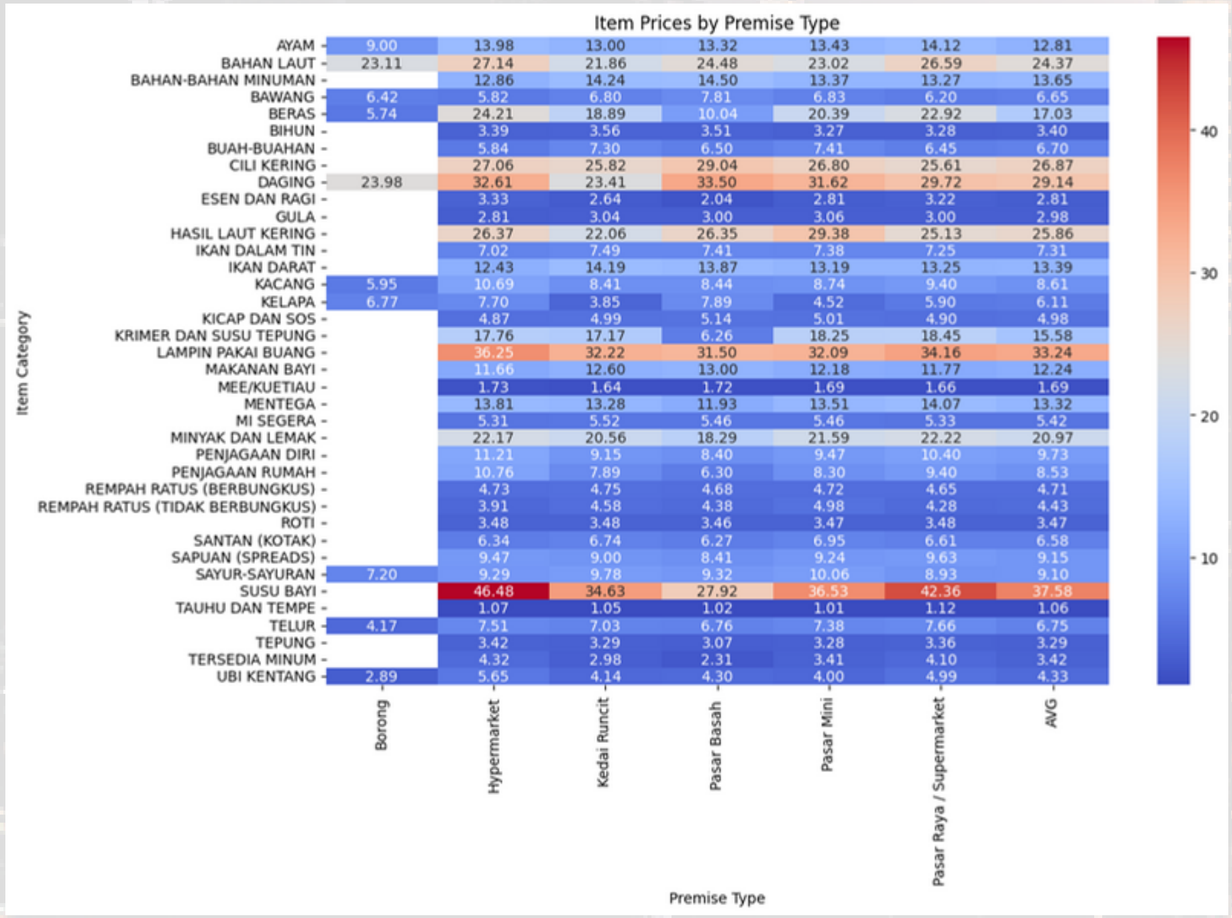
DATA DESCRIPTION



EDA

- The item price heatmap, which visualizes how item prices varies across states. Warmer colors (e.g., red) indicate higher average prices, while cooler colors (e.g., blue) indicate lower average prices.
- The availability of item categories and states in the dataset, emphasizing the diversity of items and regions included in the analysis.

DATA DESCRIPTION

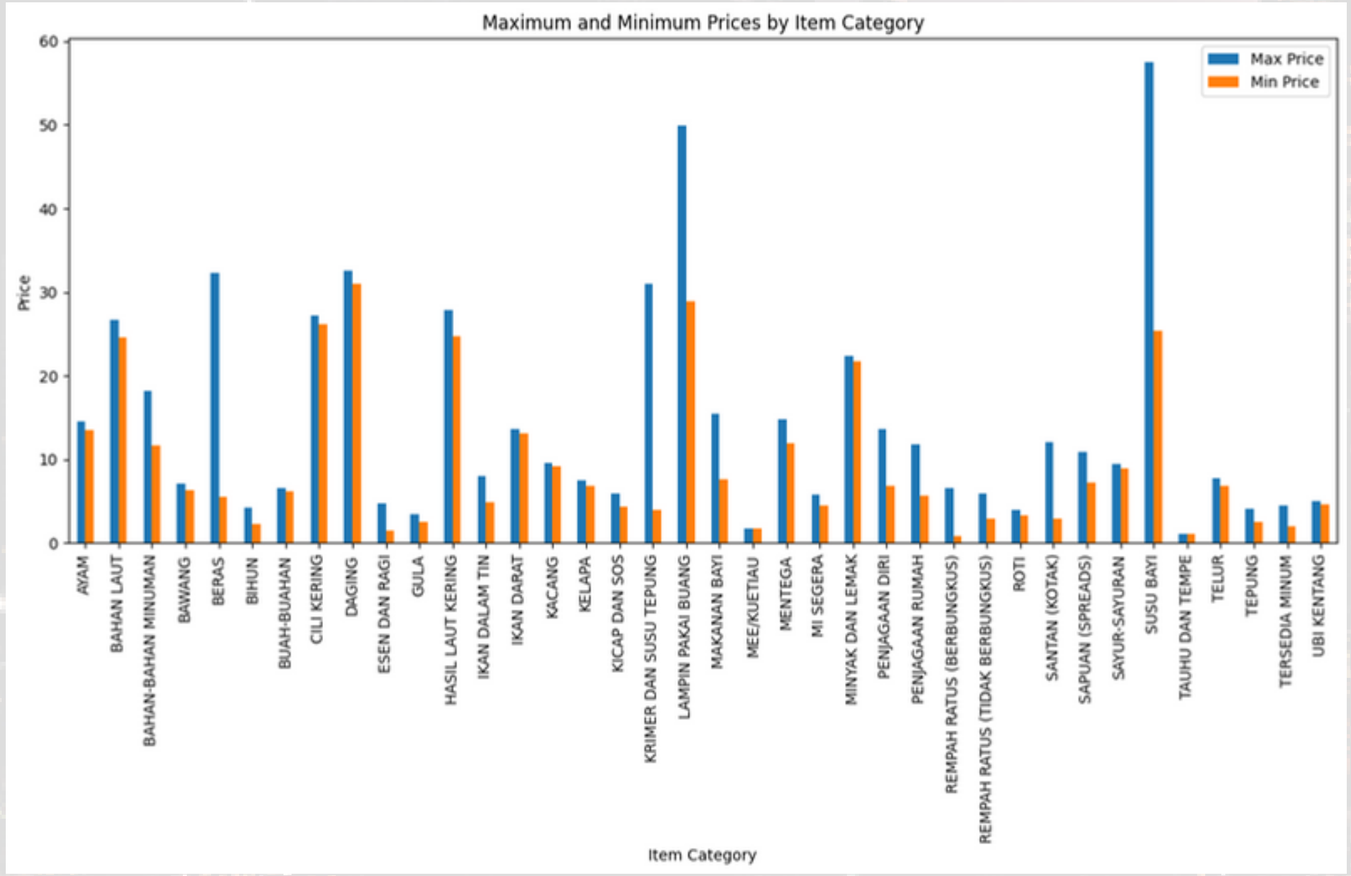


EDA

- The heatmap, which visually illustrates how average prices for item categories vary across different types of premises. Warmer colors (e.g., red) indicate higher average prices, while cooler colors (e.g., blue) represent lower average prices.

- The ability to understand how the choice of premise type may impact the prices of specific item categories, which can be valuable for decision-making or market analysis.

DATA DESCRIPTION



EDA

- The bar plot illustrating the price range (from minimum to maximum) for “item_category”, which gives an overview of price variability of each item.
- This visual representation allows for a quick identification of items with high or low price fluctuations within the time frame.

METHODOLOGY

01 **Data Integration**

- Four months of Price Charter data are combined integrated with the Lookup item dataset and Lookup premise dataset

02 **Data Preprocessing**

- the combined datasets were cleaned by dropping the unwanted columns and removing the missing values

03 **Hypothesis testing**

- Data analysis techniques were use to find the relationships of the data to test the hypothesis.

04 **Data Visualization**

- the dataset then is exported to power bi, python libraries (matplotlib.pyplot & seaborn) to visualize the results

05 **Prediction Modeling**

- using Prophet time series forecasting tool and python a prediction model was made to predict future prices of the house hold items

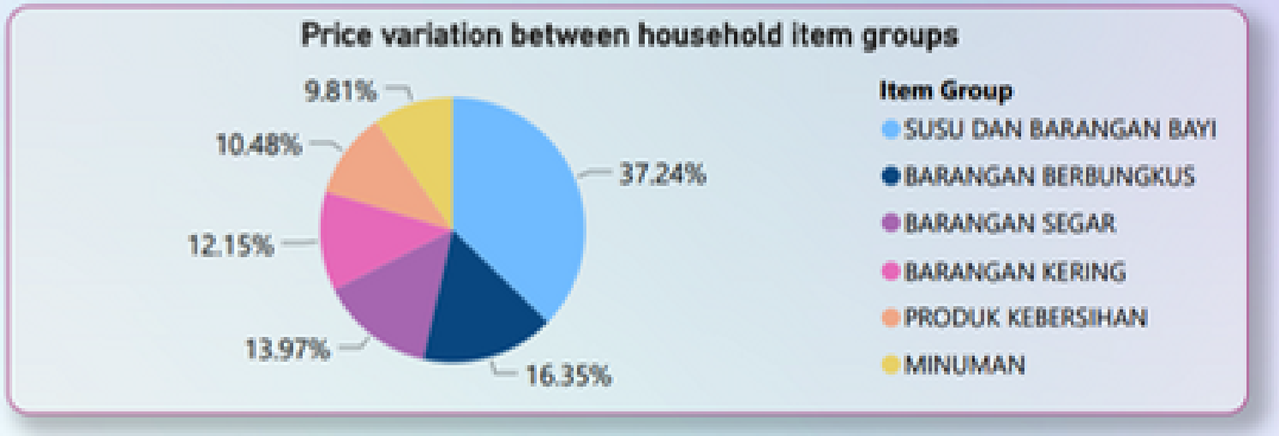
RESULTS

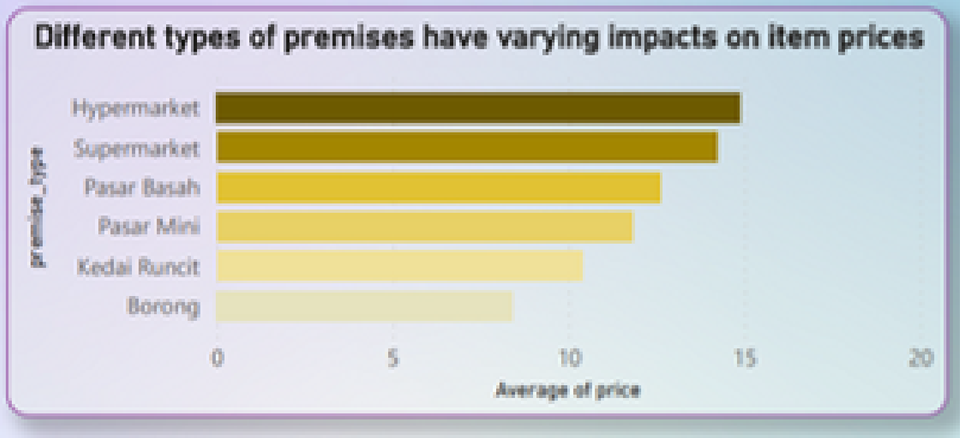
Hypothesis 1: Average price of household items differs between state

- States with a higher cost of living such as Kuala Lumpur tend to have higher average prices for household items, as these expenses are often closely tied to the overall cost of living rather than Perlis.
- The variations in average household item prices between states are not solely random, but rather a result of a consumer preferences and market dynamics.

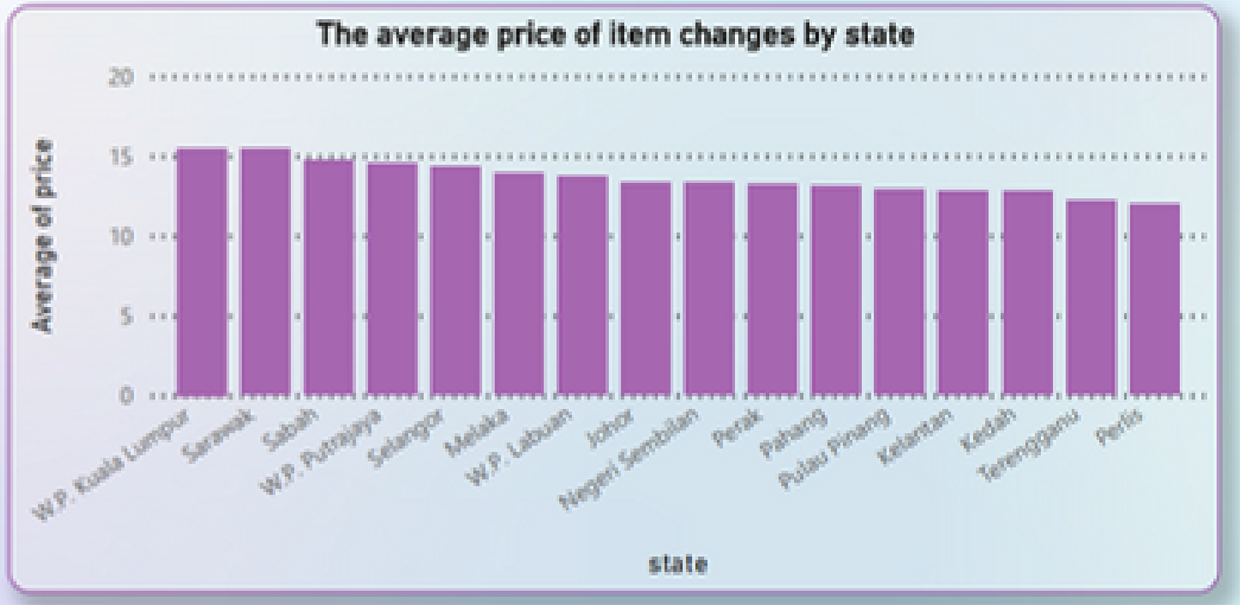
Hypothesis 2: Different types of premises have varying impacts on item prices

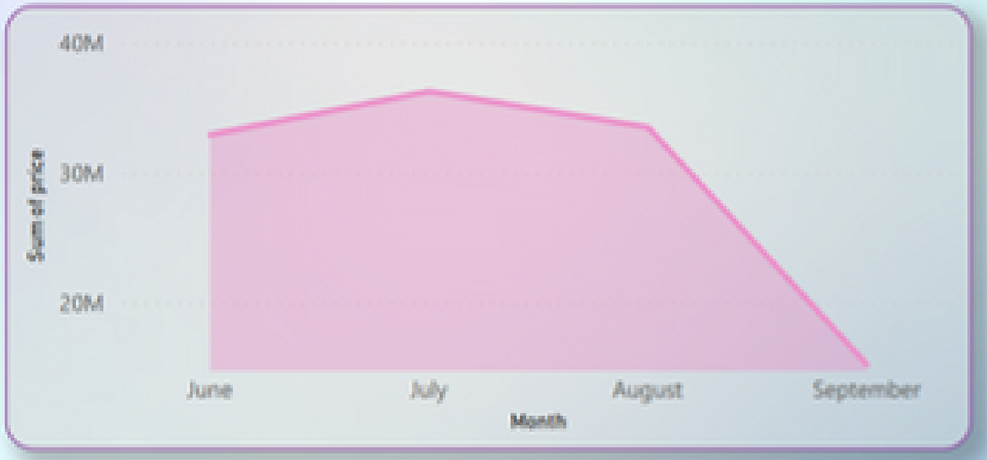
- Hypermarket has the highest average price than Borong due to premium brands and higher-quality items are typically associated with higher prices.
- Pasar basah who focus on fresh produce is higher than pasar mini
- Larger stores such as supermarket may have to manage higher overhead expenses, while smaller stores like kedai runcit may have smaller proportional overhead expenses



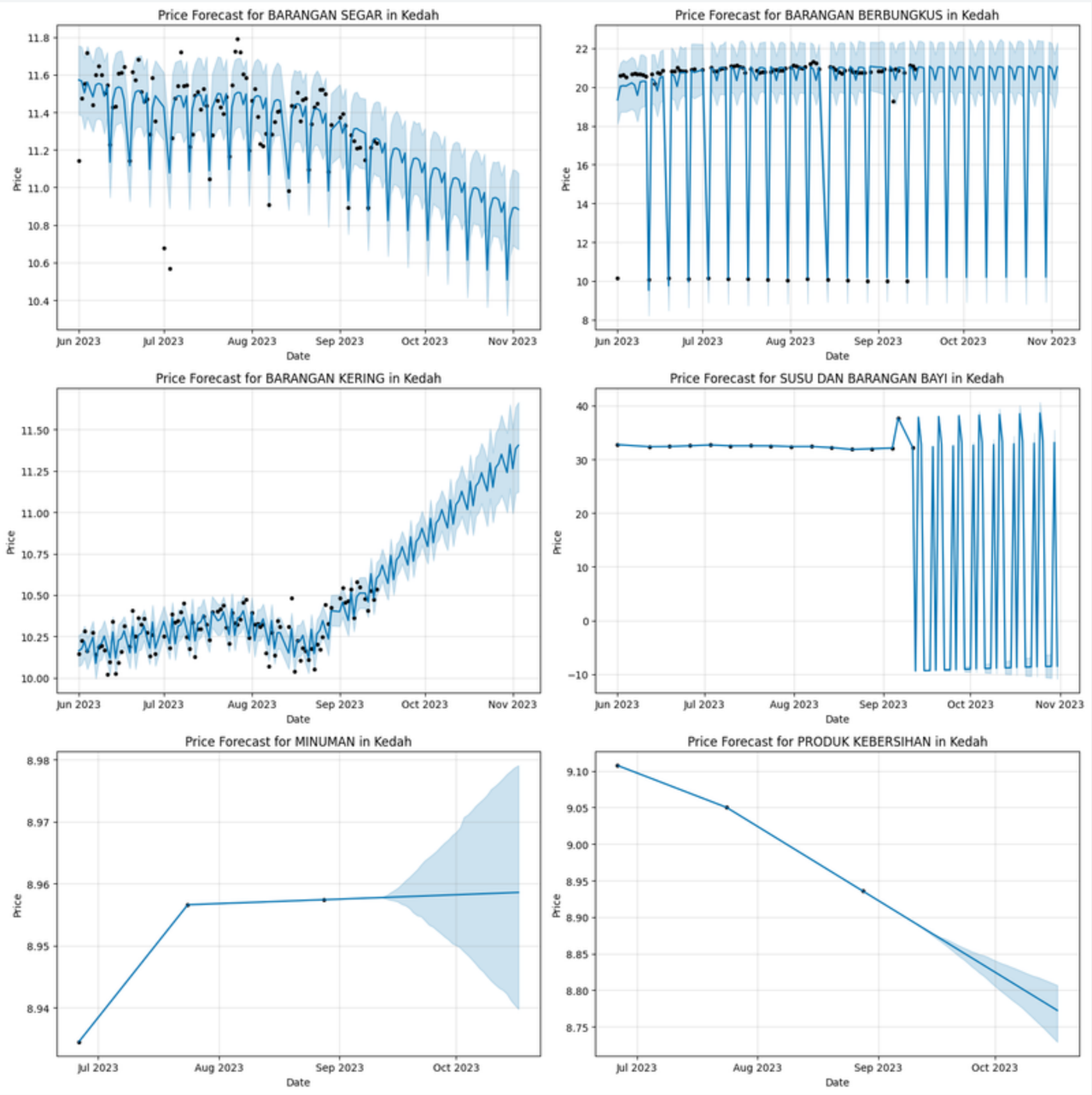


Different demand dynamics influence price variation between household item groups.





At 15.48, W. P. Kuala Lumpur had the highest average of price than Perlis due to cost of living



Hypothesis 3: Prices of household items increases over time

- A price forecasting model was built to predict the future price of the items in chosen state for the next 50 days.
- Based and the price forecast for KEDAH only items grouped in “BARANG KERING” shows an increase in price.
- Other item group such as “Barang SEGAR” and “PRODUK KEBERSIHAN” shows a downward trend in prices.
- “BARANG BERBUNGKUS”, “SUSU DAN BARANGAN BAYI” and “MINUMAN shows a constant price

RECOMMENDATION



Analyze competitors and adjust pricing strategies.

Optimize supply chain strategies to align with product demand and market trends.

Collect more data by engaging with customers and obtaining their feedback.

CONCLUSION & DISCUSSION

Lessons Learnt

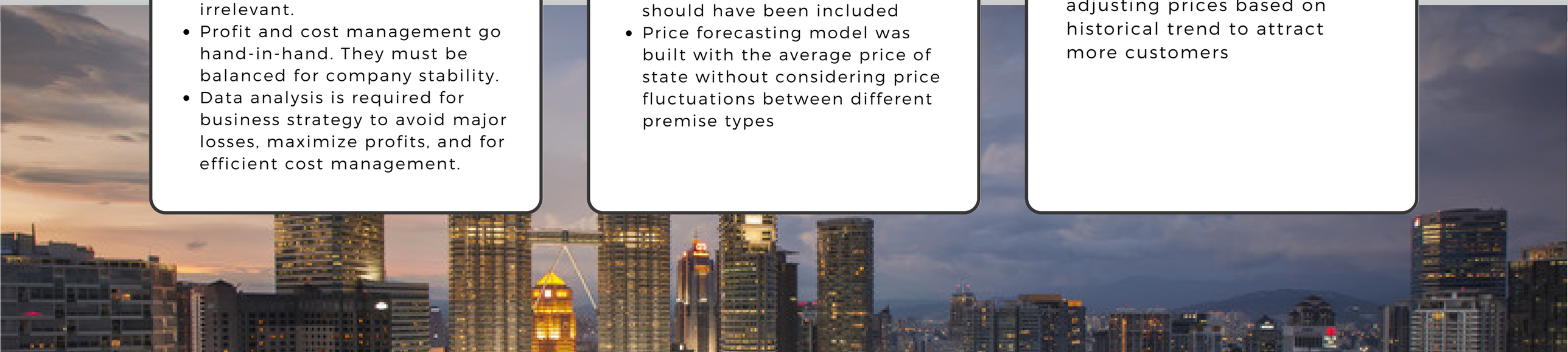
- A complete and prepared data is required for an analysis. Without proper data cleaning, the findings would become irrelevant.
- Profit and cost management go hand-in-hand. They must be balanced for company stability.
- Data analysis is required for business strategy to avoid major losses, maximize profits, and for efficient cost management.

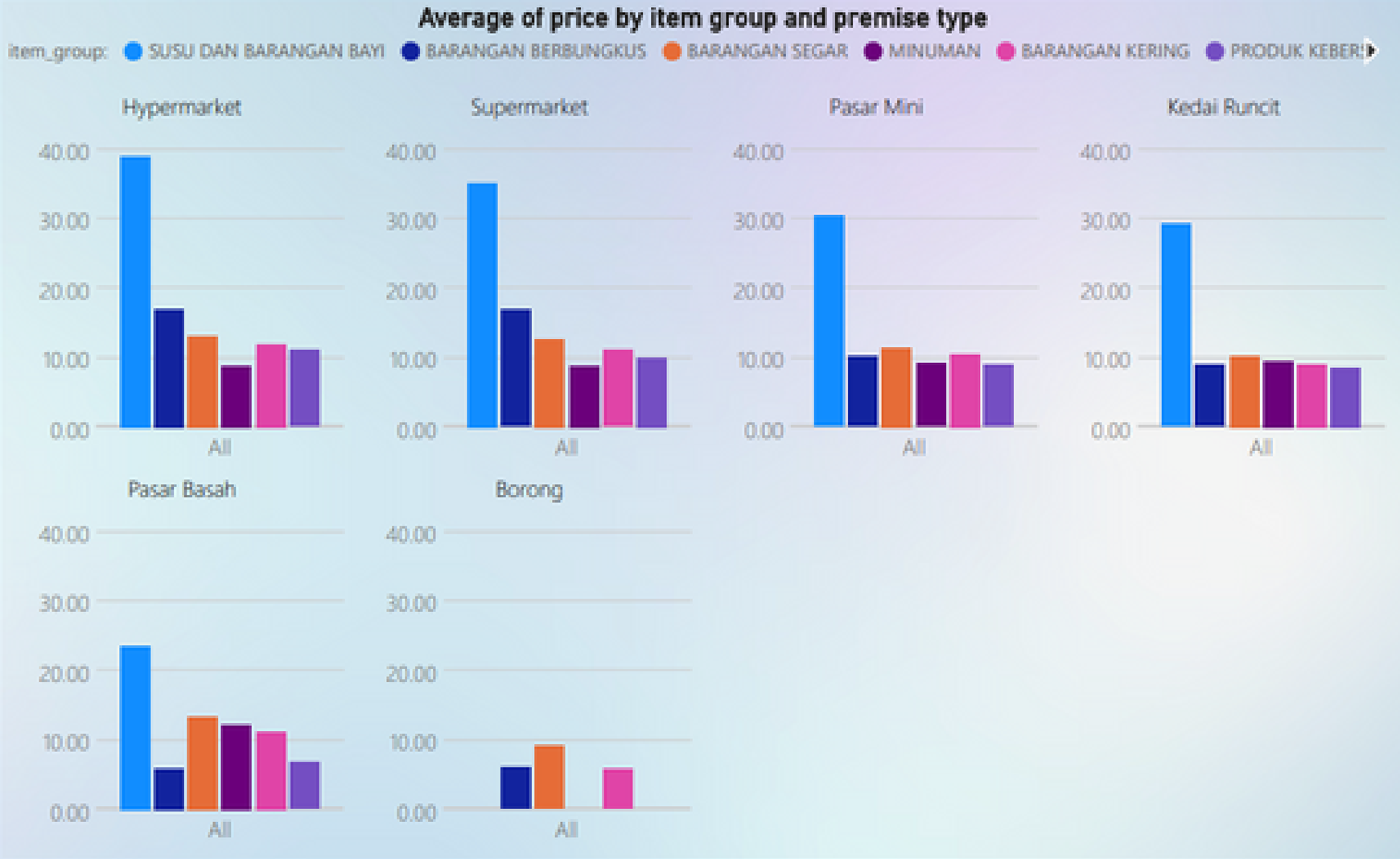
Limitations

- The dataset's timeframe is limited to four months, monthly and yearly trends are not fully capture. Earlier historical data should have been included
- Price forecasting model was built with the average price of state without considering price fluctuations between different premise types

Implication of the study

Business and retailers can use the insights to optimize pricing strategies by adjusting prices based on historical trend to attract more customers




Average of price by item group and premise type
item_group:
SUSU DAN BARANGAN BAYI
BARANGAN BERBUNGKUS
BARANGAN SEGAR
MINUMAN
BARANGAN KERING
PRODUK KEBER
Hypermarket
Supermarket
Pasar Mini
Kedai Runcit
Pasar Basah
Borong
40.00
30.00
20.00
10.00
0.00
All

APPENDIX

APPENDIX

code repository link:
https://github.com/aqilahsalihah/PriceCatcher_Forecaster/blob/main/dotdotdot_DATATHON.ipynb



DOTDOTDOT