HOME APPLIANCES SALES FORECASTING AND RECOMMENDATION

Achmad Luckyta Fasyni

COMPANY PROFILE:
PT DENKI KOBO

PT. Denki Kobo is one of the oldest and largest distributors of Japanese home appliances in the country. They sell a myriad of home appliances from air conditioners, refrigerators, rice cookers, washing machines, vacuum cleaners, etc.

ABOUT THE PROJECT

BUSINESS ANALYSIS

The client have yet to employ proper data analysis techniques in their business. They need to predict the demand of the goods and ultimately
the future sales number.

DATA PRACTICE

The data was still manually recorded and as such, owing to human errors. The client need to improve their data collection methods,

PROJECT TASKS

01 DATA OVERVIEW
Write a data description report.

02 DATA CLEANING
Take out the data belonging to PT. D.K. Works.

03 MODELLING
Demonstrate a method for predicting the client's future sales numbers.

04 FORECASTING
Predict when, if it happens, a total of 30 million goods is sold by the client in 2021.

05 RECOMMENDATION 1
Suggest to the client which product(s) you think they should cut.

06 RECOMMENDATION 2
Propose a necessary data collection method.

01.

DATA DESCRIPTION REPORT

Dataset and Worksheet:
https://github.com/LuckyFasyni/Home-Appliances-Sales-Forecasting-Recommendation

Amount of Files:
46 csv files (sales dataset) and 1 xlsx file (margin dataset)

SALES DATASET

A group of dataset which explains sales for each product every day.

- Data Collection Practice
  Collected into separate files divided by quarters every year. From Q1 2010 (in '2010Q1.csv') to Q2 2021 (in '2021Q2.csv').

- Date Range
  '2010-02-15' (February 15th, 2010) to '2021-04-29' (April 29th, 2021)

- Data Quality
  There are 105630 rows (78.27%) with missing values in at least one of the columns. No duplicated data.

DATA STRUCTURE

FEATURE	DESCRIPTION	DATA TYPE	NON-NULL
Date	The date when the sales happen	object	134964
Category1	The first word of product category	object	134964
Category2	The second word of product category	object	93196
Category3	The third word of product category	object	30603
Maker	The maker or the brand of the product	object	134964
Sales	The total amount of the product sold on that day	float64	129588
Identifier	The identifier code of the product	object	134964

MARGIN DATASET

A dataset containing the average margin of profit for every brand they have sold in the past 3 years.

- Data Collection Practice
  Collected into single xlsx file.

- Data Quality
  There are no missing values and duplicated data

DATA STRUCTURE

FEATURE	DESCRIPTION	DATA TYPE	NON-NULL
Category	The product category	object	52
Maker	The maker or the brand of the product	object	52
Margin	The average margin of profit for the past 3 years	int64	52

02.
DATA CLEANING

Take out the data belonging to PT. D.K. Works.

WORKFLOW

1
Drop missing values from sales dataset (only 3.9% of total data)

2
Merge 3 product category columns into 1 on sales dataset

3
Crosscheck if product between sales and margin dataset is matched

4
Merge sales and margin dataset into 1 fact dataset

TAKE OUT D.K. WORKS DATA

A product with an identifier of “CTCNSTZEDP” belongs to PT. Denki Kobo (the client), while a product with an identifier of “PRN3TK7HWK” belongs to PT. D.K. Works (the sister company).

Notice that the data belonging to PT D.K. Works has the number 3 and 7 in the identifier code. Since the client and PT. D.K. Works had already parted ways, they requested that the latter’s data be taken out.

Cleaned dataset
129588 records

```
factsales_df[~factsales_df['Identifier'].str.contains('\d')].reset_index(drop=True)
```

Only client’s dataset
89498 records

03.

MODELLING

Demonstrate a method for predicting the client's future sales numbers.

ARIMA ALGORITHM

ARIMA, or Autoregressive Integrated Moving Average, is a popular time series forecasting model used to make predictions based on historical data.

ARIMA models incorporate three key components:

- Autoregression (p), which involves predicting future values based on past values of the same variable;
- Differencing (d), which involves transforming the data to achieve stationarity; and
- moving average (q), which involves modeling the error term as a linear combination of past error terms.

By combining these components, ARIMA models can capture the underlying patterns and relationships in time series data, making them useful for predicting future values and identifying trends and patterns in the data.

WORKFLOW

Feature Selection
Only Date and Sales columns will be included

Resampling
Resample the dataset to weekly frequency to handle inconsistent date interval

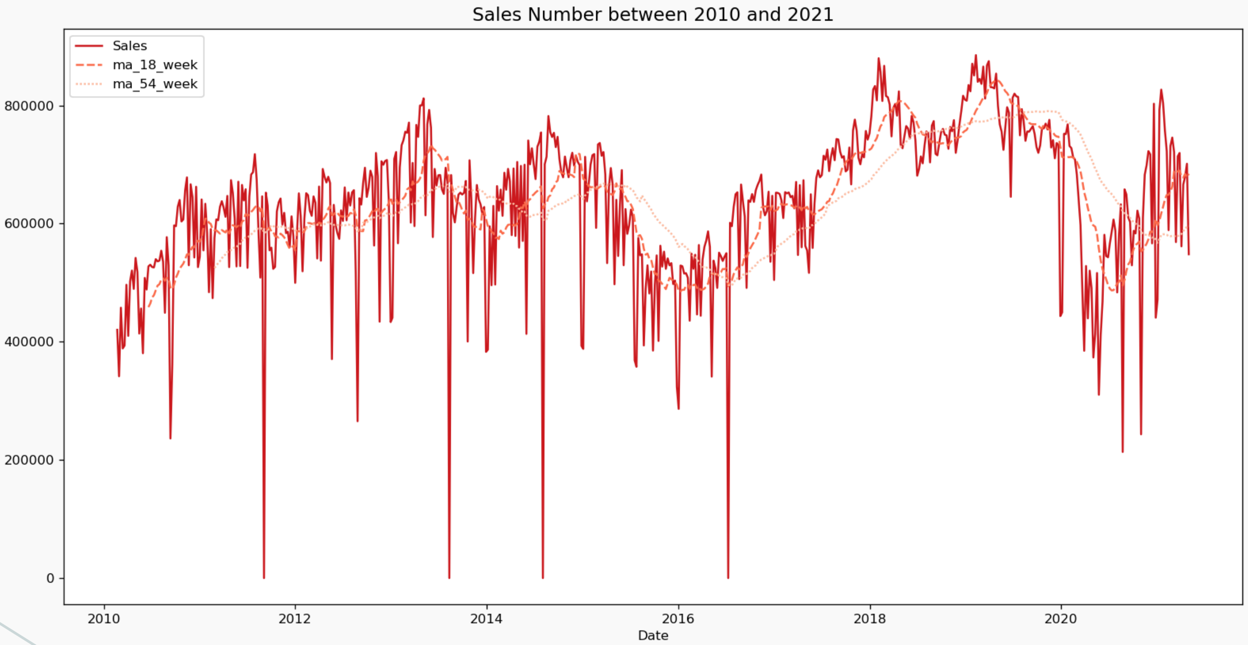
Check for Stationarity
Turns out the data is non-stationary as p-value = 0.052

Hyperparameter Tuning
Determine the order of the model using the Akaike Information Criterion (AIC) through various d, p, and q parameter values

Model Training
Results:
MSE = 9458556739.2 - RMSE = 97255.11
Stiil need to be improved!

Forecasted Sales Visualization

STATIONARITY CHECK



AdFuller Test Results:

ADF Statistic:
-3.6449034811734395

p-value:
0.05214714534222171

because p_value > 0.05, hence 'the data is **not stationary**

HYPERPARAMETER TUNING

Best Model:

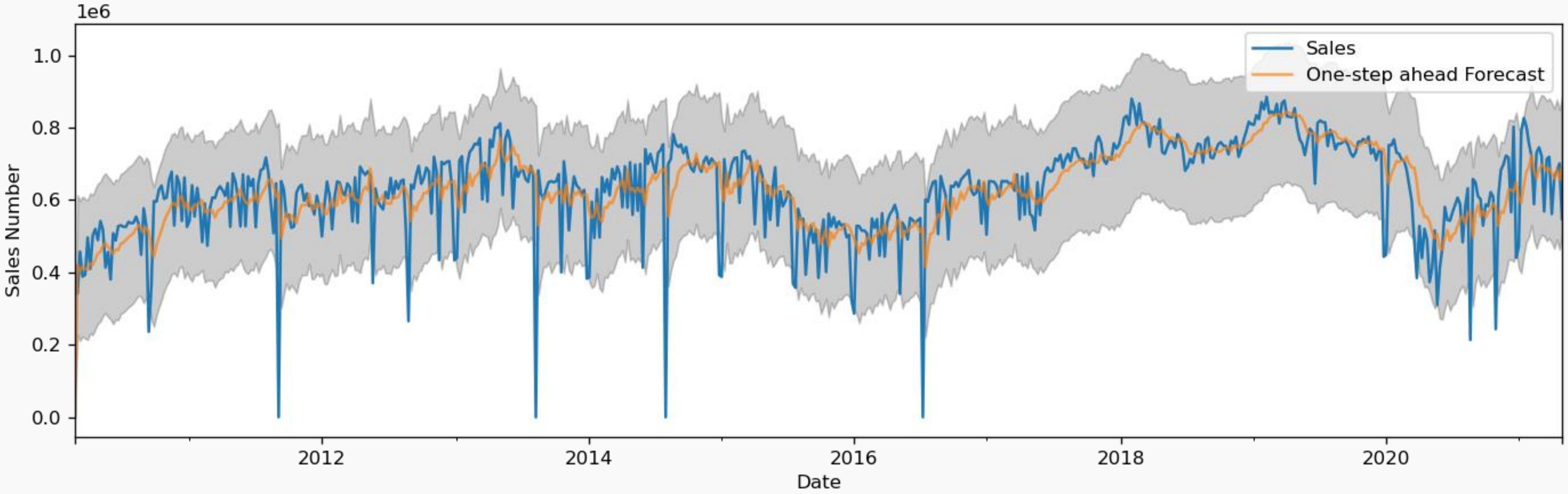
SARIMAX(0, 1, 2)x(0, 0, [1, 2, 3], 12)

With AIC = 15075.251

```
Best model:  ARIMA(0,1,2)(0,0,3)[12]
Total fit time: 101.677 seconds
                                      SARIMAX Results
=============================================================================================
Dep. Variable:                                     y   No. Observations:                  585
Model:             SARIMAX(0, 1, 2)x(0, 0, [1, 2, 3], 12)   Log Likelihood           -7531.626
Date:                               Thu, 06 Apr 2023   AIC                          15075.251
Time:                                       15:14:24   BIC                          15101.471
Sample:                                   02-21-2010   HQIC                         15085.470
                                        - 05-02-2021
Covariance Type:                                 opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.7717      0.031    -25.000      0.000      -0.832      -0.711
ma.L2         -0.0373      0.037     -1.000      0.317      -0.110       0.036
ma.S.L12      -0.0582      0.049     -1.179      0.238      -0.155       0.039
ma.S.L24      -0.0762      0.064     -1.193      0.233      -0.201       0.049
ma.S.L36      -0.0591      0.056     -1.063      0.288      -0.168       0.050
sigma2      9.726e+09   6.03e-12   1.61e+21      0.000    9.73e+09    9.73e+09
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              4924.17
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               0.55   Skew:                            -2.64
Prob(H) (two-sided):                  0.00   Kurtosis:                        16.21
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 3.36e+35. Standard errors may be unstable.
```

MODEL TRAINING

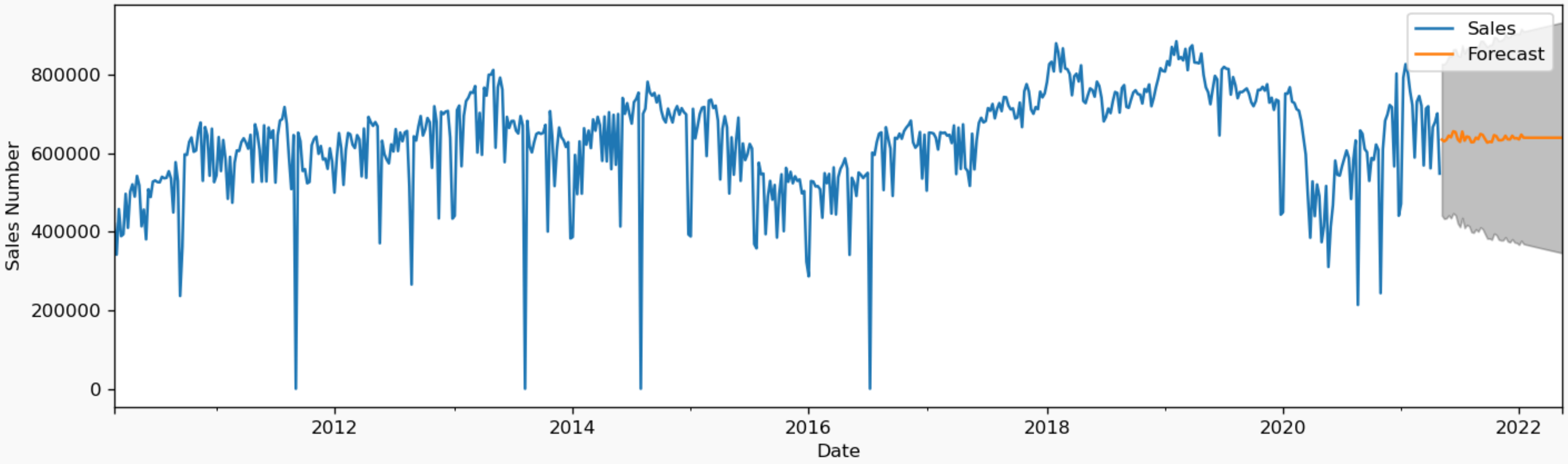


Model Performance Evaluation

MSE = 9458556739.2 & RMSE = 97255.11 → compared to Mean Benchmark = 634365.11

The model is performing well! The RMSE is lower than the mean benchmark, but still need to be improved!

FORECASTED FUTURE SALES



04.

FORECASTING

Predict when, if it happens, a total of 30 million goods is sold by the client in 2021.

WORKFLOW

1
Capture predicted sales number in a data frame

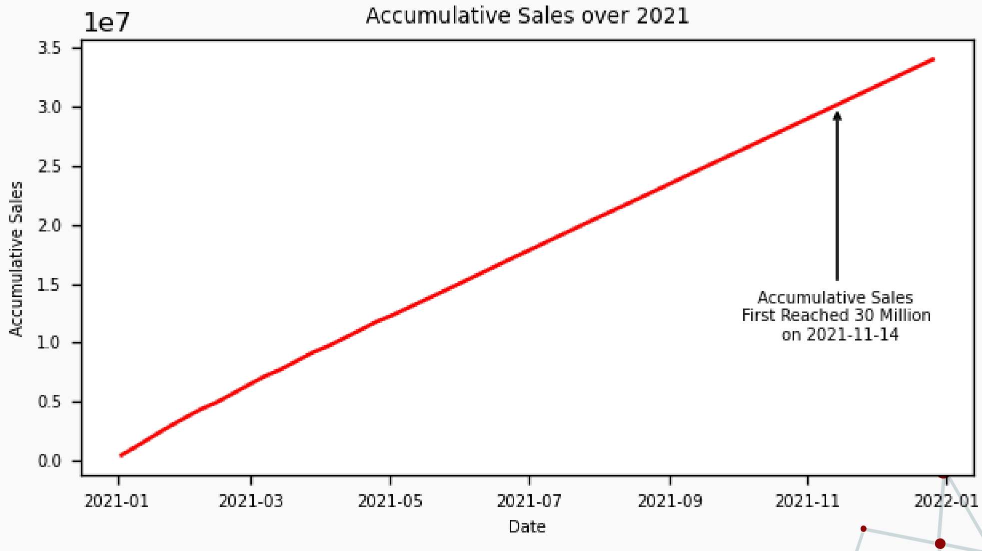
2
Concatenate both actual and predicted sales number data in 2021

3
Calculate the accumulative sales number through 2021 year

4
Accumulative sales data visualization

FORECASTED ACCUMULATIVE SALES

The client had set a target to sell a total of 30 million goods before the year 2021 ends. At their current sales rate, this goal can be accomplished on **November 14th, 2021** or **the second week of November.**



05.

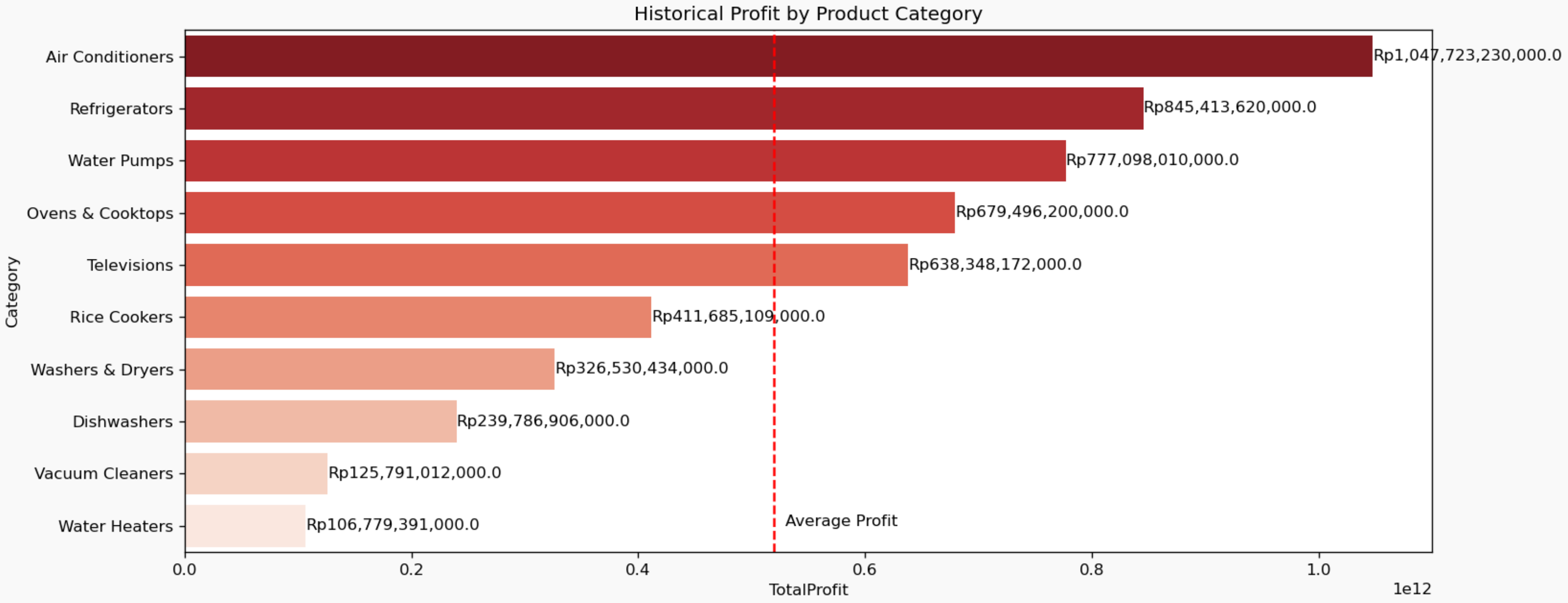
PRODUCT EFFICIENCY RECOMMENDATION

Suggest to the client which product(s) you think they should cut.

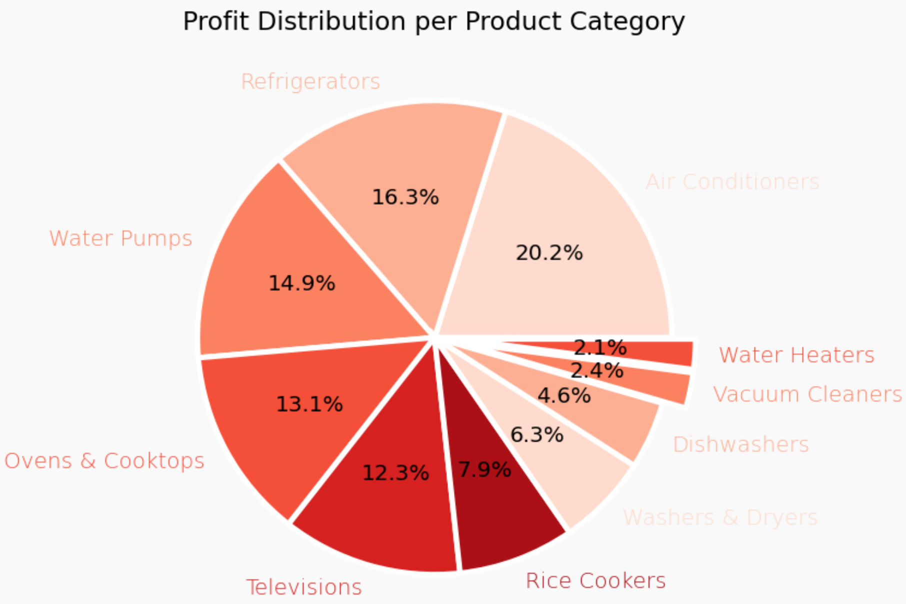
PANDEMIC HITS HARD!

Due to the unexpected COVID-19 pandemic, the client is heavily considering shutting down several of their warehouses, and as they will have limited space to store their goods, they will be forced to cut one or a few of their product lines.

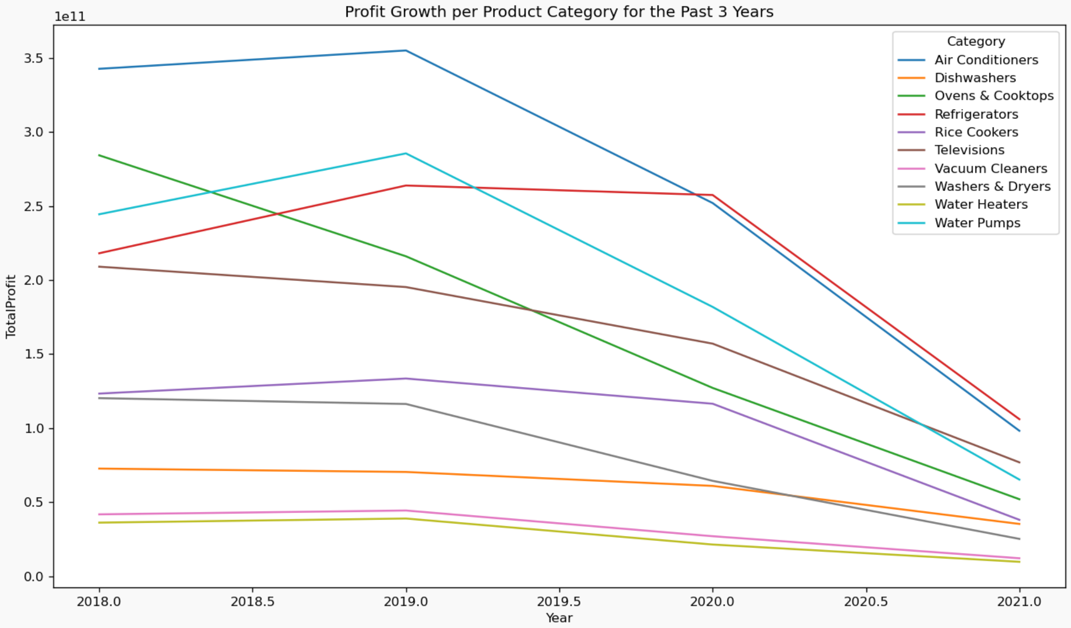
WATER HEATERS AND VACUUM CLEANERS ARE THE LEAST PROFITABLE PRODUCTS



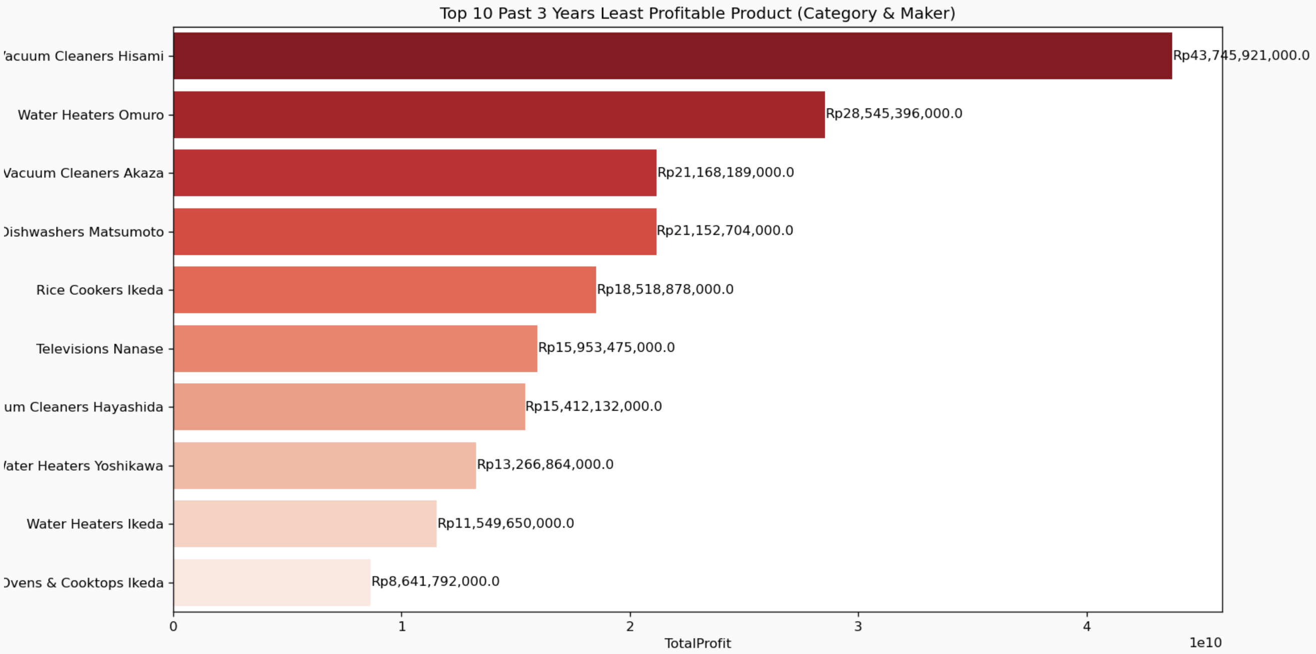
PAST 3 YEARS PROFIT DISTRIBUTION



UNDERWHELMING PERFORMANCE BY WATER HEATERS AND VACUUM CLEANERS



TOP 10 LEAST PROFITABLE PRODUCTS



ANALYSIS

1. Based on product category, `Water Heaters`, `Vacuum Cleaners`, and `Dishwashers` are the least profitable among other categories. Contribute total 9% from the Total Profit in the past 3 years.

2. But shutting down one product category without further analysis on the competition landscape seems a big deal since it will decrease the diversification that the client has had in the last years. Moreover it could also because of the limited product choice in that category.

If we see the profit growth through past 3 years, there are several product who have been consistantly suffered an underwhelming performance such as:

1. `Ikeda Ovens & Cooktops`
2. `Ikeda Water Heaters`
3. `Yoshikawa Water Heaters`
4. `Hayashida Vacuum Cleaners`
5. `Nanase Televisions`,
6. `Ikeda Rice Cookers`,
7. `Matsumoto Dishwashers`,
8. `Akaza Vacuum Cleaners`,
9. `Omuro Water Heaters`, and
10. `Hisami Vacuum Cleaners`.

Combine those product sales, they only contribute 4.75% of all product sales number and 3.8% of all profit.

06.

DATA COLLECTION BEST PRACTICES

Propose a necessary data collection method.

CONTEXT

Based on the information provided, the client's data collection practice involves manually recording daily sales data by the company's sales managers since early 2010. This manual process may have resulted in some days where the sales data is not available due to human error.

The daily sales data is consolidated by the set product category and maker. However, the product category data is stored in three different columns, where a column contains only a single word and empty entries are marked as "null".

Sales data is not available for weekends and public holidays as the warehouses only operate on weekdays. For audit purposes, the sales data is broken down into separate files divided by quarters.

RECOMMENDATION

1. **Automate Data Collection**: Implement a system to automatically record daily sales data that can reduce the human error and make sure all days are recorded. The client can use any POS (point-of-sales) service system. Additionally, POS systems can integrate with other software systems, such as accounting or customer relationship management (CRM) software, to provide a comprehensive view of business operations.

2. **Train Sales Officers and Managers**: Provide training to the sales managers on data keeping best practices and how the sales officers to input data accurately.

3. **Standardize Data**: Standardize the way data is stored, for example, store product category data in a single column with no empty entries.

4. **Include Weekend and Public Holiday Data**: Consider adding data for weekends and public holidays even if the warehouses are closed during these periods, as this could help in analyzing sales patterns.

5. **Consolidate Quarterly Sales Data**: Consolidate quarterly sales data into a single file instead of separate files to simplify data analysis.

RECOMMENDATION

As for the data the client could collect to accommodate future data analysis, the client could consider the following:

Geographical Data

Information on the geographic location of the customers to better understand the demand for certain products in different regions.

Customer Demographics

Age, gender, and income could provide valuable insights into the target market.

Inventory Data

Inventory levels and turnover could help in optimizing supply chain and inventory management.

THANKS

Do you have any questions?

luckyfasyni@gmail.com
+62 857 7496 4599