PRICE INTELLIGENCE

15th JUNE 2022 (WEDNESDAY)

1430 - 1530 (MYT)

ONLINE
The Department of Statistics Malaysia (DOSM) has initiated the implementation of BDA under the project of Statistics Big Data Analytics (STATSBDA).
STATSBDA OBJECTIVE

To modernise data collection

To allow high accuracy

To reduce respondents burden

To use as supplements for existing data in production of certain statistics

To produce new statistical indicators

Objective
Price Intelligence

- leveraging the capability of Big Data in collecting large data from various sources and transform them into better structure

- different prices of the same goods can be obtained through various online retailer websites, providing a modernized price data collection

- Transform from unstructured data into structured data to perform analysis
Objective

01

to give better insight in consumer price analysis and monitoring

02

to establish new set of data source, either to complement or as a reference for the calculation of Consumer Price Index

New Data Collection Methodology
New data collection process introduced to cover price of goods selling online

Value Creation

Transform of Business Process
New data collected enable to create holistic landscape of CPI monitoring process including engagement with other government agencies

Holistic View in Online and Offline Prices
Allow monitoring and forecast future price trend and as valuable input for price control decisions by government
## PRICE INTELLIGENCE ARCHITECTURE

<table>
<thead>
<tr>
<th>DATA ACQUISITION</th>
<th>ETL</th>
<th>PRICE LAKE DEVELOPMENT</th>
<th>ANALYSIS</th>
<th>VISUALIZATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>External Data Sources</strong>&lt;br&gt;Online Retailer</td>
<td><strong>Online Retail Database</strong></td>
<td><strong>Raw Data Management</strong></td>
<td>Trending &amp; Forecast&lt;br&gt;Descriptive&lt;br&gt;Dependency &amp; Correlation&lt;br&gt;Price Basket Enrichment</td>
<td><strong>Dashboard</strong>&lt;br&gt;Tableau</td>
</tr>
<tr>
<td><strong>Website Government/Agency</strong>&lt;br&gt;FAMA&lt;br&gt;SAM&lt;br&gt;NDZ&lt;br&gt;SAH&lt;br&gt;DOFM</td>
<td><strong>Intelligent Product Classification (COICOP)</strong></td>
<td><strong>Enrichment Data Management</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DOSM Price Database (CPI)</strong>&lt;br&gt;</td>
<td><strong>FTP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Data Acquisition
   - Collect & Extract data from various sources, both internal & external

2. Data Management - Price Lake Development
   - Store structured massive data

3. Data Analysis
   - Descriptive
   - Exploration
   - Comparison

4. Data Visualization

Note: COICOP → Classification of Individual Consumption According to Purpose
Example: Groceries/Fresh Fish

Information extracted:
- Page URL
- Product title
- Product unit
- Product category
- Price
- Original price
- Additional details
Example: Electronic Items


Information extracted:
- Page URL
- Product title
- Product category: Electronic Items
- Product ID: 2TC50AD1X
- Price: RM1,597
- Original price: RM2,200
- Additional details:
  - Brand: Sharp
  - Model: 2TC50AD1X
  - Specifications:
    - General:
      - Brand: Sharp
      - Model: 2TC50AD1X
      - Product Type: LED TVs
    - Additional details:
      - Brand: Sharp
      - Model: 2TC50AD1X
      - Product Type: LED TVs
PI Dictionary – Classification at 4 Digit level of COICOP

Pending/Accepted refers to the matching result at the 4 Digit COICOP level

Partially/exact refers to the matching result with MCOICOP

11.3% exact match
88.7% partially match

The word most often appear on the item description

*Information based on 231 item specifications updates in PI Dictionary
## PI Dictionary – Text matching scoring

Specification that meet requirement of items in CPI basket of goods with high scoring value will return exact match, mapped to 7 digits item specification MCOICOP code

<table>
<thead>
<tr>
<th>Item Specification</th>
<th>6D Code</th>
<th>Item Code</th>
<th>Score</th>
<th>Match Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huggies Dry Diapers M 6-11 kg 72 pieces each</td>
<td>121321</td>
<td>1213212</td>
<td>75.14</td>
<td><strong>EXACT</strong></td>
</tr>
<tr>
<td>Drypers Wee Wee Dry Disposable Diapers M 6-11kg 74pcs each</td>
<td>121321</td>
<td>1213211</td>
<td>76.59</td>
<td><strong>EXACT</strong></td>
</tr>
<tr>
<td>Huggies Dry Diapers M 6-11 kg 72 pieces each</td>
<td>121321</td>
<td>1213212</td>
<td>75.14</td>
<td><strong>EXACT</strong></td>
</tr>
<tr>
<td>Huggies Dry Diapers M 6-11 kg 72 pieces each</td>
<td>121321</td>
<td>1213212</td>
<td>75.14</td>
<td><strong>EXACT</strong></td>
</tr>
<tr>
<td>Drypers Wee Wee Dry Disposable Diapers M 6-11kg 74pcs each</td>
<td>121321</td>
<td>1213211</td>
<td>76.59</td>
<td><strong>EXACT</strong></td>
</tr>
<tr>
<td>Drypers Wee Wee Dry Disposable Diapers M 6-11kg 74pcs each</td>
<td>121321</td>
<td>1213211</td>
<td>76.59</td>
<td><strong>EXACT</strong></td>
</tr>
</tbody>
</table>

Meanwhile, broad specification items has only 6 digits item code to be matched with online item specification. Items is considered matching as long as the item is the same regardless of brand name, units etc

<table>
<thead>
<tr>
<th>Item Specification</th>
<th>6D Code</th>
<th>Item Code</th>
<th>Score</th>
<th>Match Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drypers Wee Wee Dry Disposable Diapers Sxl 15+kg 40pcs each</td>
<td>121321</td>
<td></td>
<td>67.12</td>
<td><strong>PARTIAL</strong></td>
</tr>
<tr>
<td>Huggies Dry Pants Diapers 19-14kg 50pcs each</td>
<td>121321</td>
<td></td>
<td>64.44</td>
<td><strong>PARTIAL</strong></td>
</tr>
<tr>
<td>Tena Value Diapers, medium 8 pack x 12s</td>
<td>121321</td>
<td></td>
<td>50.86</td>
<td><strong>PARTIAL</strong></td>
</tr>
</tbody>
</table>
PI Dictionary – Rule based classification

PI Dictionary has been built in order to get the best match to MCOICOP 6/7 digits based on description of the items using text matching scoring.

Dictionary will be updated regularly based on basket of goods in CPI.

COICOP : Classification of Individual Consumption According to Purpose
1. Click Data Monitoring on the top left of the screen
2. Click Folder Name and Click file name
3. Status of the generated data will be displayed.
   Click Download to download to local PC
**PRICE INTELLIGENCE ANALYTICS AND VISUALIZATION**

a) Trend

Trend analysis is used to evaluate data patterns based on linear approaches. There are various time series analysis techniques that can be used, such as ARIMA, Exponential Smoothing, Holt Winter, Linear Trend, Exponential and Level Aggregation.

b) Descriptive

Descriptive analysis, more emphasis on data exploration summaries such as: Mean, Median, Standard Deviation, Variants, Histogram / Skewness, etc.
During the pandemic Covid-19, online flight ticket prices has been used as a reference-proxy in the compilation of the CPI.

**Item specification (Route)**
- Kedah to Kuala Lumpur
- Johor Bahru to Kuala Lumpur
- Kuantan to Kuala Lumpur
- Penang to Kuala Lumpur
- Kuala Terengganu to Kuala Lumpur

Implications of the Movement Control Order (MCO), all the price data collection at the outlets has been suspended. As part of mitigation plan, data crawling has been done during the MCO for 20 main product CPI as below:

- CAP RAMBUTAN HIJAU SST RICE 5% 10KG (RICE)
- JATI BERAS SUPER SPESIAL 10KG (RICE)
- JASMINE RICE SUPER SPECIAL 5% 10KG (RICE)
- JASMINE RICE SUPER SPECIAL TEMPATAN 5KG (RICE)
- AYAM BERSIH (PELBAGAI BAHAGIAN) (CHICKEN)
- IKAN BAWAL HITAM (FISH)
- IKAN CENCARU (FISH)
- IKAN KEMBUNG (FISH)
- KUETIAU BASAH (FLAT RICE NOODLES)
- MEE KUNING BASAH (NOODLES)
- UBI KENTANG (POTATO)
- BAWANG BESAR (ONION)
- CILI KERING KERINTING (DRIED CHILLI)
- KACANG BUNCIS (FRENCH BEAN)
- KUBIS BULAT (TEMPATAN) (CABBAGE)
- LOBAK MERAH (CARROT)
- TOMATO
- SANTAN KELAPA (FRESH COCONUT MILK)
- TELUR AYAM GRED B (HEN'S EGGS GRADE B)
- *MINYAK MASAK (PELBAGAI JENAMA) (COOKING OIL)
APPLICATIONS OF PRICE INTELLIGENCE

HOUSE RENTAL PRICES FOR INTERNATIONAL COMPARISON PROGRAM (ICP)

45.6% housing rental data for International Comparison Program 2017 (ICP2017) submission were using online price data as an additional data source to fulfill the requirements of ICP.

19 types of housing specification with different sizes consisting of:
- Single-detached house
- Attached house (row house)
- Studio apartment
- One-bedroom apartment
- Two-bedroom apartment
- Three-bedroom apartment

RENTAL HOUSING DATA SOURCED USED FOR ICP 2017
A study on administrative data by other government agency versus DOSM prices data collection
- Price collection for monitoring and enforcement
- 133 similar items between DOSM and administrative data – mostly on perishable items

### OBJECTIVE

<table>
<thead>
<tr>
<th>1</th>
<th>COMPARISON</th>
<th>2</th>
<th>DISPLAY</th>
<th>3</th>
<th>RECOMMENDATIONS</th>
<th>4</th>
<th>OPTIMIZING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>To study the difference in price and price movement between DOSM data and administrative data</td>
<td></td>
<td>Displays statistics on the distribution of prices of goods collected by DOSM and administrative data.</td>
<td></td>
<td>Provide recommendations for the replacement of outlets and reduction of price collection in the field by using the price collection channeled from administrative data.</td>
<td></td>
<td>Optimizing usage of data sources in other government agencies as a complement in generating CPI</td>
</tr>
</tbody>
</table>
RECENT WORKS – ADMINISTRATIVE DATA

Trend Harga Purata Bulanan dan Peratus Perbezaan Harga antara DOSM

Products with highest number of observation and minimum difference in price between sources
CHALLENGES/LIMITATION

1. Crawlers
2. Legal issues
3. Data Storage
4. Data continuity for the compilation of CPI
1. Forecasting prices of fish and vegetable using web scraped price micro data
   Mazliana Mustapa, Raja Rajeswari Ponnusamy, Ho Ming Kang

2. Online and Offline Prices: Measuring selected home appliance’s products
   Mohd Saiful Husain and Norsyela Muhammad Noor Mathivanan

3. Analysis of the Mobile Phones Prices Malaysia using Web Scraped Data
   Nur Hurriyatul Huda Abdullah Sani

Table 3: ARIMA best model

<table>
<thead>
<tr>
<th>Item</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Bream</td>
<td>$Y_t = 2.4205 - 0.333Y_{t-1} + 0.6382Y_{t-2} + 0.0594 + 0.9688e_{t-1}$</td>
</tr>
<tr>
<td>Solar Kuning</td>
<td>$Y_t = 1.4794 - 1.6648Y_{t-1} - 0.7431Y_{t-2} + 0.7924e_{t-1}$</td>
</tr>
<tr>
<td>Green Spinach</td>
<td>$Y_t = 1.1813 + 1.3246Y_{t-1} - 0.7774Y_{t-2} + 0.1008 - 0.8323e_{t-2}$</td>
</tr>
<tr>
<td>Kangkung</td>
<td>$Y_t = 0.9995 + 1.7293Y_{t-1} - 0.7776Y_{t-2} + 0.1023 - 0.8315e_{t-2}$</td>
</tr>
<tr>
<td>Long Beans</td>
<td>$Y_t = 0.5460 + 3.667Y_{t-1} + 0.0730 + 0.4322e_{t-1} + 0.2638e_{t-2}$</td>
</tr>
<tr>
<td>Basah, Cencur, Kempong, Round Cabbage and Savii Japun</td>
<td>$Y_t = \mu + Y_{t-1}$ where: $\mu =$ mean of the changes of period to period</td>
</tr>
</tbody>
</table>
Wan Ahmad Ridhuan Wan Jaafar  
Core Team Big Data Analytics  
Department of Statistics Malaysia  

**Team Price Intelligence**  
mazliana@dosm.gov.my  
ridhuan@dosm.gov.my  
noradilah.adnan@dosm.gov.my  

https://statsbda.dosm.gov.my/
THANK YOU

StatsMalaysia
www.DOSM.gov.my

#KeluargaMalaysia

2022  2012-2022  20 OKT  2016-2030