



PRIME MINISTER'S DEPARTMENT
DEPARTMENT OF STATISTICS MALAYSIA

STATISTICAL, ECONOMIC AND SOCIAL
RESEARCH AND TRAINING CENTRE
FOR ISLAMIC COUNTRIES (SESRIC)
**TRAINING COURSE ON BIG DATA
APPLICATIONS ON PRICE STATISTICS**

PRICE INTELLIGENCE



15th JUNE 2022 (WEDNESDAY)



1430 - 1530 (MYT)



ONLINE



20 OCT



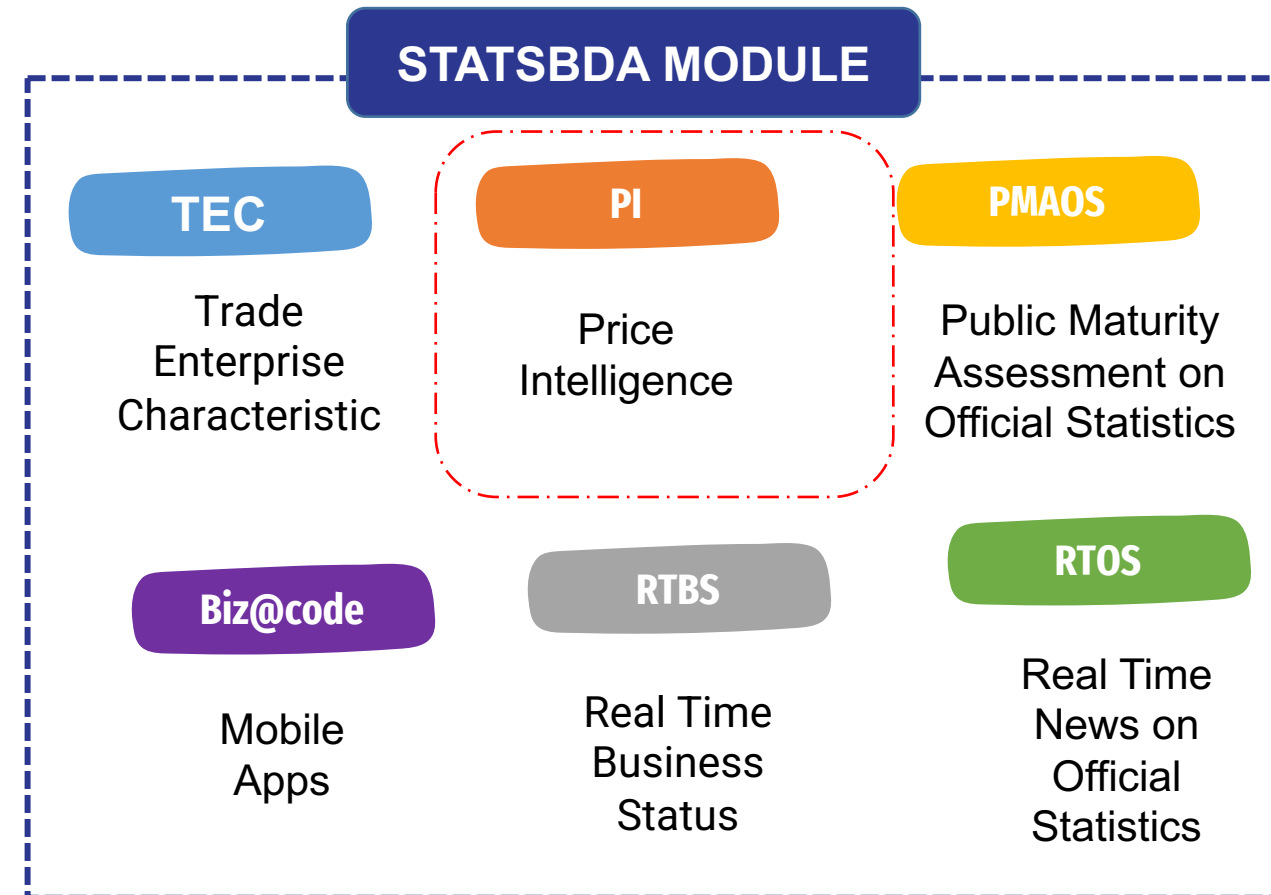
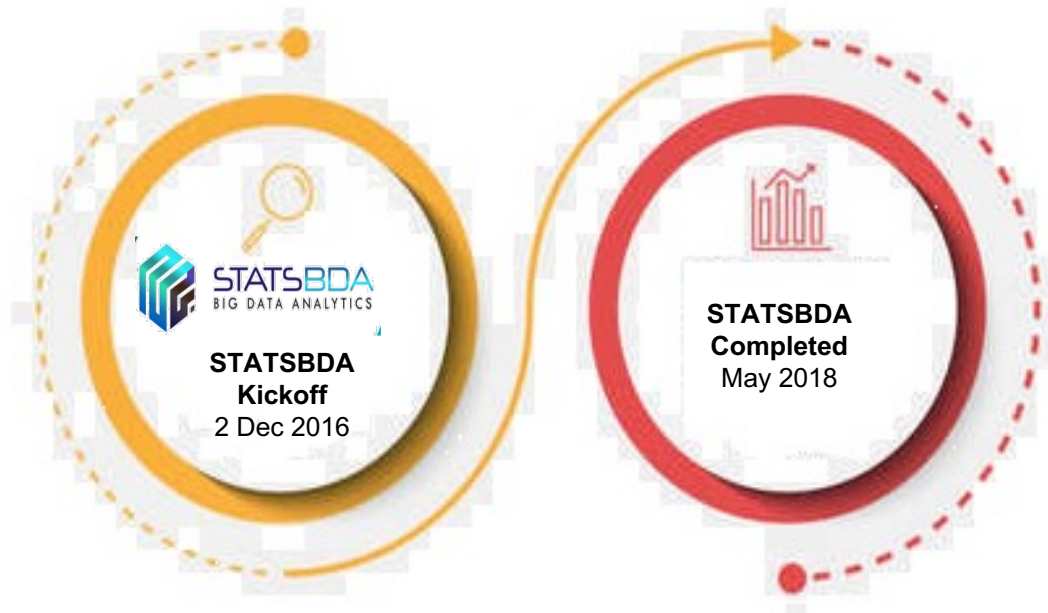
2016 - 2030



Wan Ahmad Ridhuan Wan Jaafar
Department of Statistics Malaysia



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







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



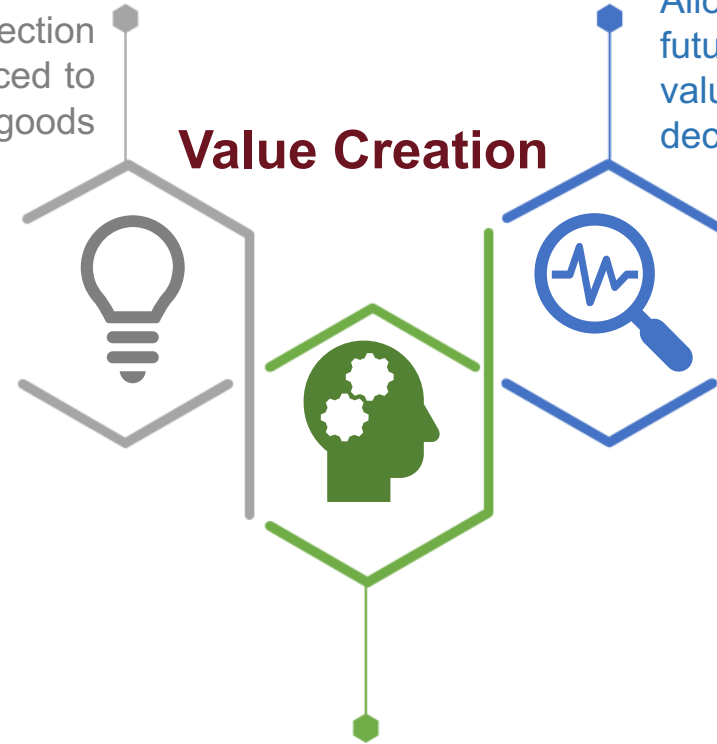
PI Visualization

Data Management

New Data Collection Methodology

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

Value Creation



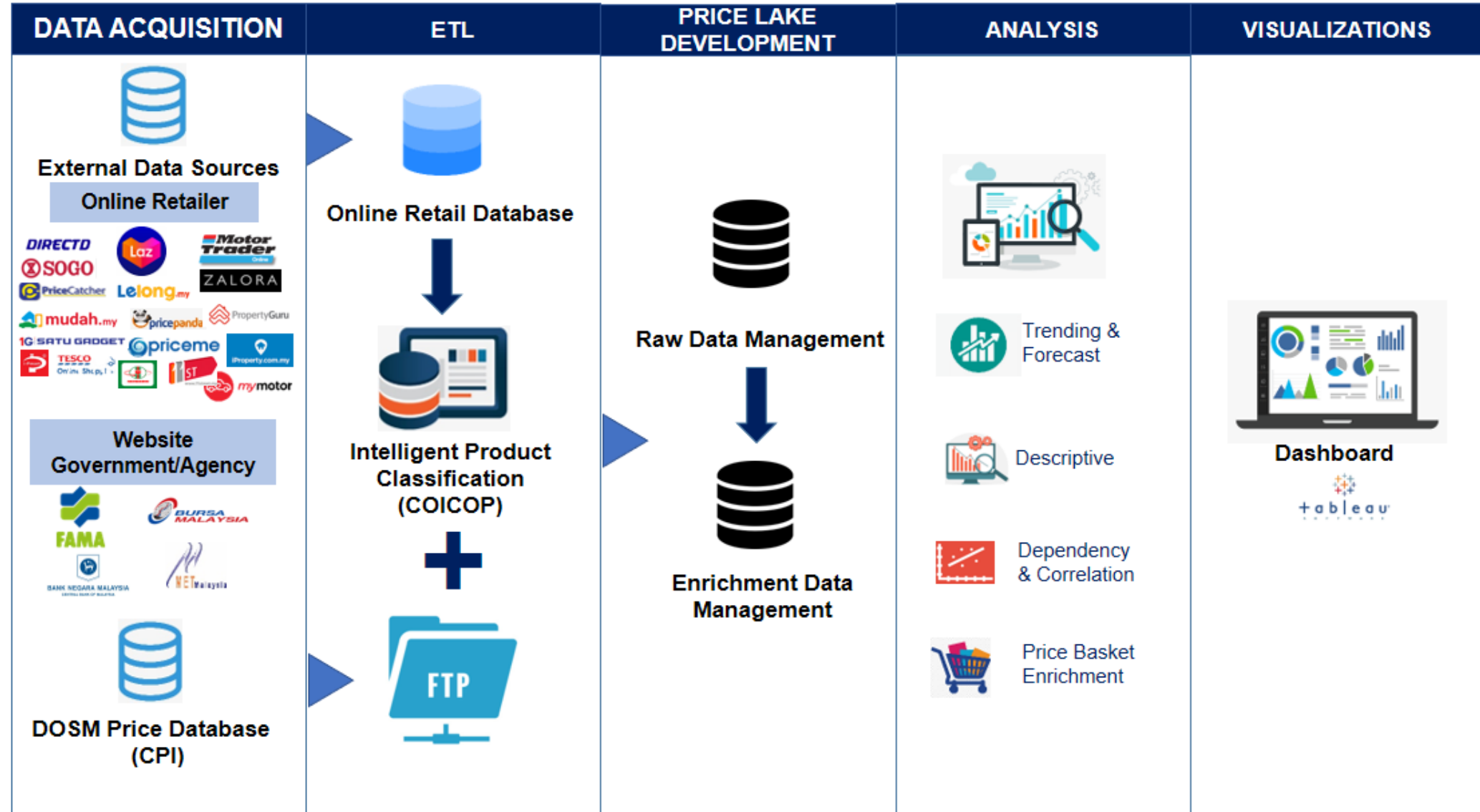
Holistic View in Online and Offline Prices

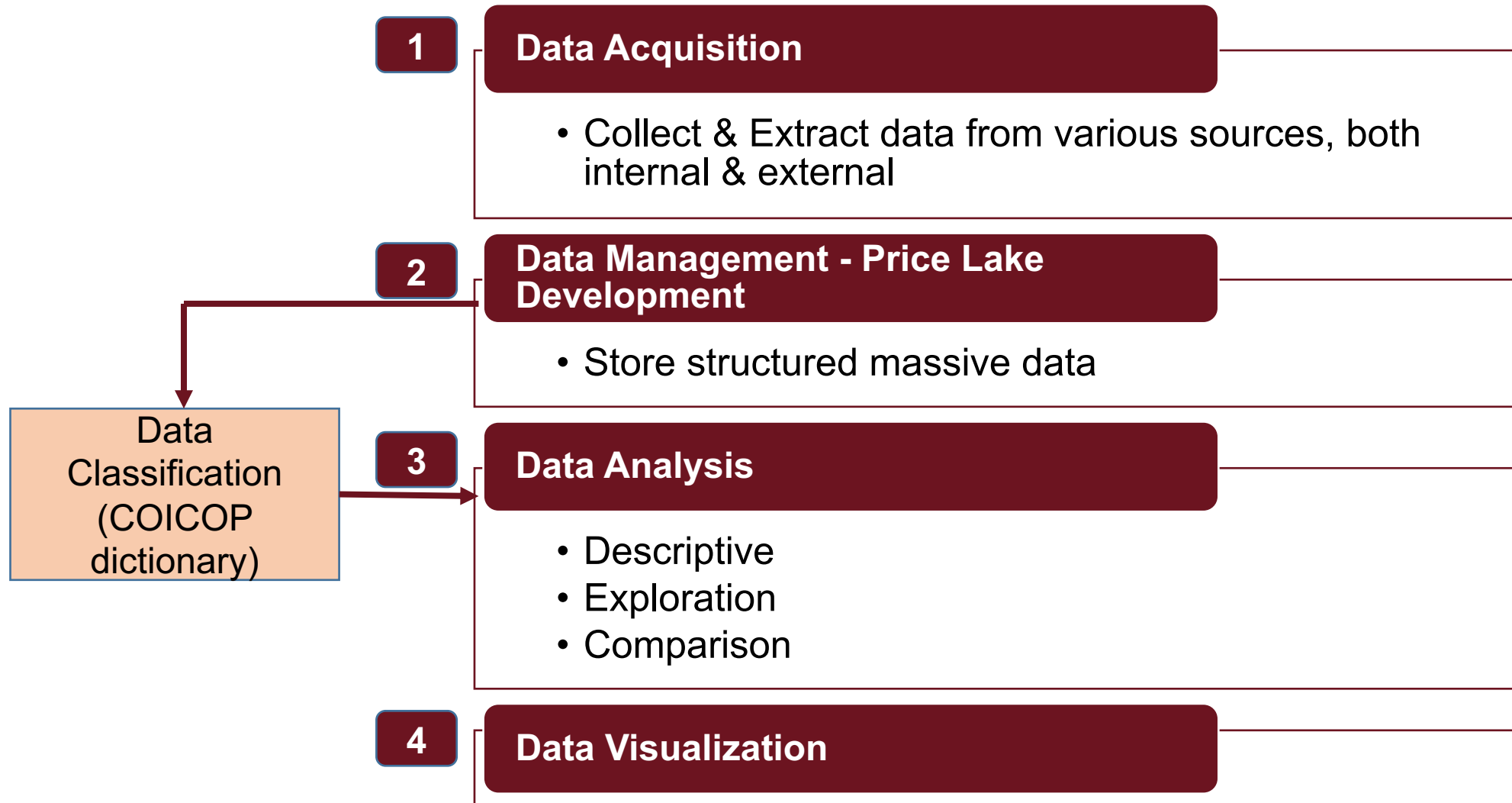
Allow monitoring and forecast future price trend and as valuable input for price control decisions by government

Transform of Business Process

New data collected enable to create holistic landscape of CPI monitoring process including engagement with other government agency

PRICE INTELLIGENCE ARCHITECTURE





Note : COICOP → Classification of Individual Consumption According to Purpose



Example : Groceries/Fresh Fish

Offer



**Gula Prai Caster Sugar
500g**

**Now RM2.25, Before RM2.39,
Save RM0.14**
Promotion valid until
19/05/2021

RM 2.25 RM2.25/each

1 **Add**



Mabong (Indian mackerel)
[Rest of category >](#)

☐ Quantity ☒ Weight

RM 3.81 RM17.99/kg

1.06kg **Add**

Information extracted

- Page URL
- Product title
- Product unit
- Product category
- Price
- Original price
- Additional details



Example : Electronic Items

harveynorman.com.my/tv-and-audio/tvs-en-2/tvs-en-3/sharp-2tc50ad1x-50-full-hd-led-tv.html

Sign In APU webmail NEWS Portal Individual - isi-web... Regression in R | Da... National Quality As... Malaysia << SEA-E...

Enjoy Free delivery* when you spend a minimum of RM100! Shop Now!

Catalogues Store finder Career Wishlist Compare Account Contact us

BROWSE PRODUCTS > Harvey Norman > What are you looking for? 0

TV & AUDIO > TV > TV > SHARP 2TC50AD1X 50" FULL HD BASIC TV

Sharp 2TC50AD1X 50" Full HD Basic TV

RM1,597
WAS RM2,499

ELECTRICAL & IT JUMBO SALE!
★★★★★
Be the first to review this product

Write a review

1

ADD TO CART

FIND IN STORE

Product Availability

Type your address

✓ Nationwide Delivery
✓ Shop With Confidence
✓ Over 20 Stores Nationwide

Available for 24 Months Interest Free Payments
RM66.54 per month
Min. Total Purchase Values Apply. Find Out More Here

Specifications

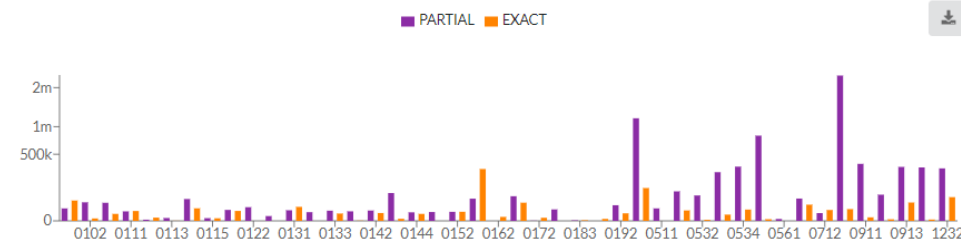
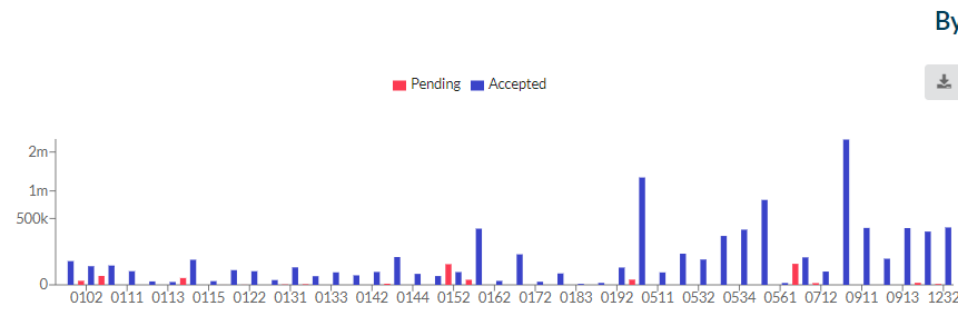
General	
Brand	Sharp
Model	2TC50AD1X
Product Type	LED TVs

Information extracted

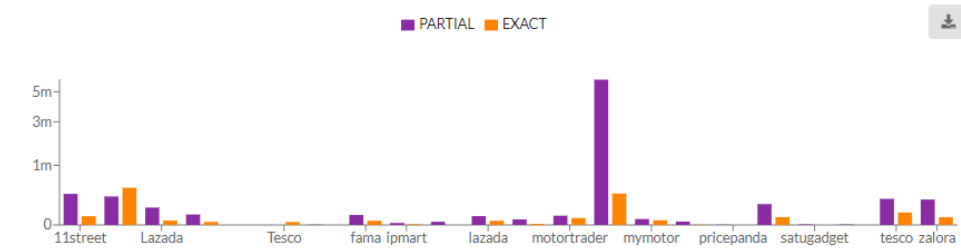
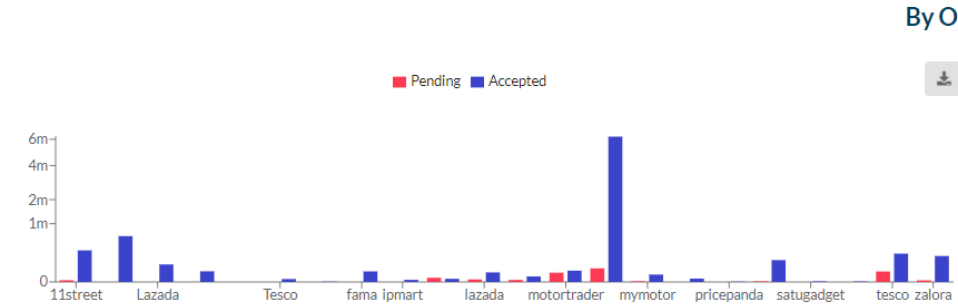
- Page URL
- Product title
- Product category
- Product ID
- Price
- Original price
- Additional details
 - Brand
 - Model



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

highly mesin singapore japan peti blouse basuh galaxy myset apple korea oppo latest incredible newest price paradise products sony refrigerator sleeve size dispenser machine 5s 6s merah over
plus xiaomi note **iphone** swap cuckoo white purifier baju shirt items pro air quartz penapis watch panasonic water huawei redmi toshiba trends

*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

Item Specification	6D Code	Item Code	Score	Match Type
 huggies dry diapers m 6-11 kg 72 pieces each	121321	1213212	75.14	EXACT
 drypers wee wee dry disposable diapers m 6-11kg 74pcs each	121321	1213211	76.59	EXACT
 huggies dry diapers m 6-11 kg 72 pieces each	121321	1213212	75.14	EXACT
 huggies dry diapers m 6-11 kg 72 pieces each	121321	1213212	75.14	EXACT
 drypers wee wee dry disposable diapers m 6-11kg 74pcs each	121321	1213211	76.59	EXACT
 drypers wee wee dry disposable diapers m 6-11kg 74pcs each	121321	1213211	76.59	EXACT

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

Item Specification	6D Code	Item Code	Score	Match Type
 drypers wee wee dry disposable diapers xxi 15+kg 40pcs each	121321		67.12	PARTIAL
 huggies dry pants diapers l 9-14kg 50pcs each	121321		64.44	PARTIAL
 tena value diapers, medium 8 pack x 12s	121321		50.86	PARTIAL



PI Dictionary – Rule based classification



PI Dictionary

Search...

☐ Only show Inactive

Page Size

+

Classifier	4D Code ▲	6D Code	Description	Sample Keyword	Sample Variation	Item Code	Status
wearables	1213	121321	Lampin pakai buang untuk bayi/ Diapers for baby	Baby Diapers Diapers Disposable Diapers Drypers Genki Goon Huggies Mamypoko Pampers Whoopee	Classic Diapers Drynights Night Pants Petpet Super Tape Ultra Wee Wee Dry	1213211	✓
wearables	1213	121321	Lampin pakai buang untuk bayi/ Diapers for baby	Baby Diapers Disposable Diapers Genki Goon Huggies Lampin Pakai Buang Bayi Mamypoko Pampers Tesco Loves Whoopee	Baby Diapers Dry Drynights Drypant Extra Pants Tape Wee Wee Dry Wwd	1213212	✓
wearables	1213	121322	Syampu/perapi/ Shampoo/conditioner	CONDI Clairol PANTENE Perapi SHM SHP SSK Sunsilk Chinese Herbal Repairing Syampu		1213223	✓

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

DATA SET GENERATOR



1

STATSBDA
BIG DATA ANALYTICS

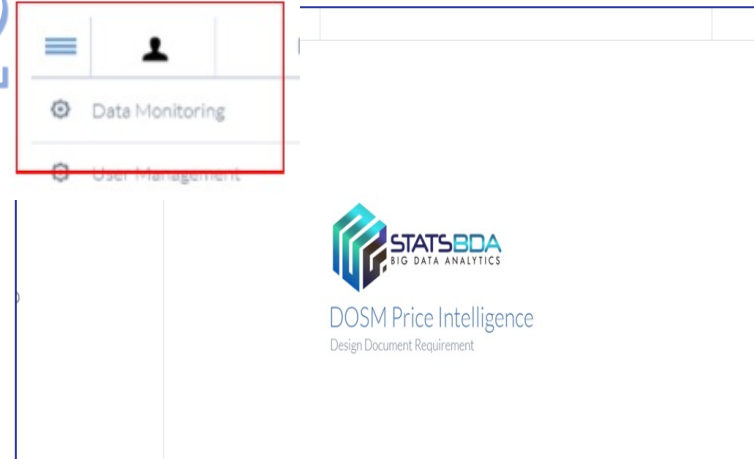
DOSM Price Intelligence

username

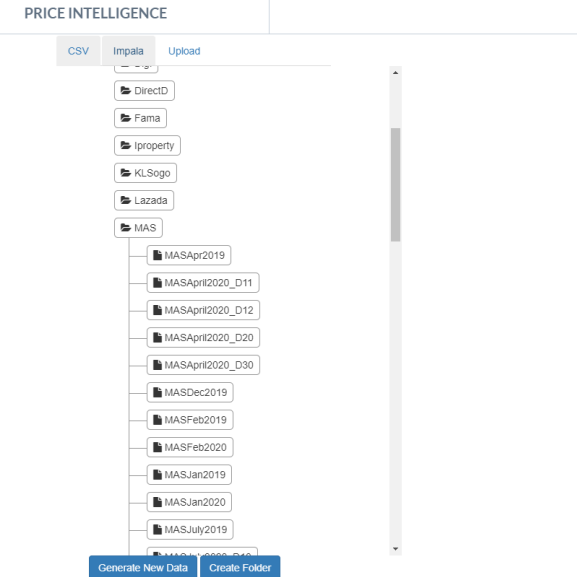
password

Login

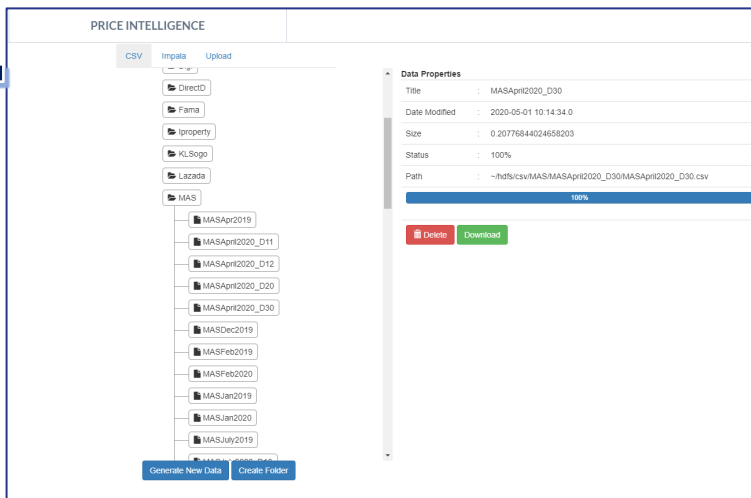
2



3



4



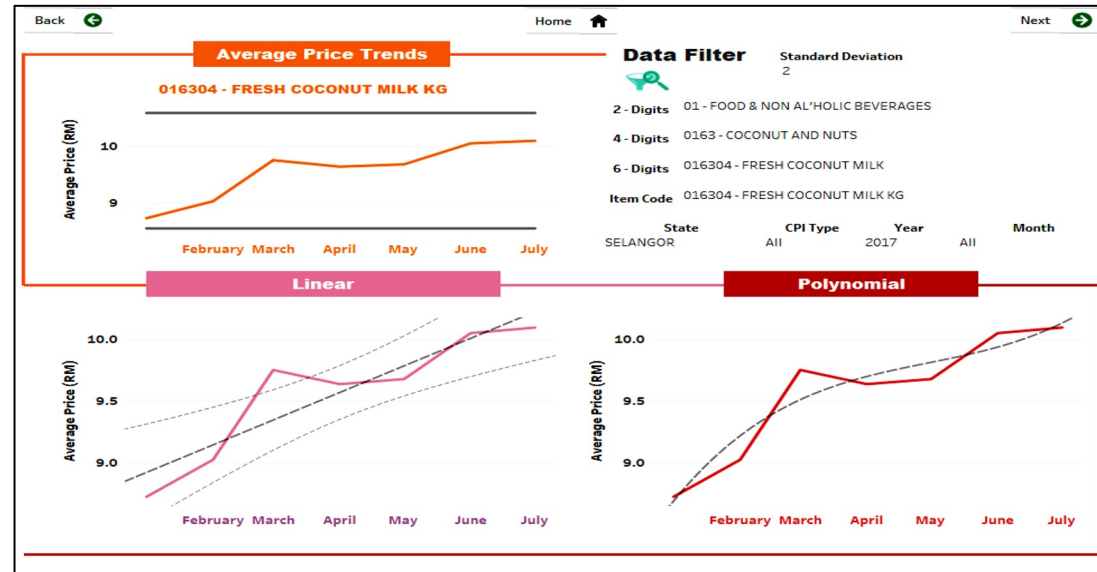
5

1	dates	site	flight_class	date_flight	flight_code	arrive	price	from	depart	dest
2	20200413	malaysiaairlines	business_smart	20200420	mh2	5:25	16,747	kuala lum	23:00	london lhr
3	20200413	malaysiaairlines	economy_flex	20200420	mh2	5:25	6,196	kuala lum	23:00	london lhr
4	20200413	malaysiaairlines	economy_promo	20200420	mh2	5:25	0	kuala lum	23:00	london lhr
5	20200413	malaysiaairlines	economy_smart	20200420	mh2	5:25	0	kuala lum	23:00	london lhr
6	20200413	malaysiaairlines	business_smart	20200420	mh3010	10:00	0	kota kinat	6:10	tawau twu
7	20200413	malaysiaairlines	economy_flex	20200420	mh3010	10:00	348	kota kinat	6:10	tawau twu
8	20200413	malaysiaairlines	economy_promo	20200420	mh3010	10:00	0	kota kinat	6:10	tawau twu
9	20200413	malaysiaairlines	economy_smart	20200420	mh3010	10:00	0	kota kinat	6:10	tawau twu
10	20200413	malaysiaairlines	business_smart	20200420	mh3012	14:50	0	kota kinat	12:00	tawau twu
11	20200413	malaysiaairlines	economy_flex	20200420	mh3012	14:50	503	kota kinat	12:00	tawau twu
12	20200413	malaysiaairlines	economy_promo	20200420	mh3012	14:50	0	kota kinat	12:00	tawau twu
13	20200413	malaysiaairlines	economy_smart	20200420	mh3012	14:50	0	kota kinat	12:00	tawau twu
14	20200413	malaysiaairlines	business_smart	20200420	mh3014	15:05	0	kota kinat	9:30	tawau twu
15	20200413	malaysiaairlines	economy_flex	20200420	mh3014	15:05	348	kota kinat	9:30	tawau twu
16	20200413	malaysiaairlines	economy_promo	20200420	mh3014	15:05	0	kota kinat	9:30	tawau twu
17	20200413	malaysiaairlines	economy_smart	20200420	mh3014	15:05	0	kota kinat	9:30	tawau twu
18	20200413	malaysiaairlines	business_smart	20200420	mh3016	15:05	0	kota kinat	12:15	tawau twu
19	20200413	malaysiaairlines	economy_flex	20200420	mh3016	15:05	348	kota kinat	12:15	tawau twu
20	20200413	malaysiaairlines	economy_promo	20200420	mh3016	15:05	0	kota kinat	12:15	tawau twu
21	20200413	malaysiaairlines	economy_smart	20200420	mh3016	15:05	0	kota kinat	12:15	tawau twu
22	20200413	malaysiaairlines	business_smart	20200420	mh3041	7:05	0	kota kinat	6:30	labuan ibu
23	20200413	malaysiaairlines	economy_flex	20200420	mh3041	7:05	190	kota kinat	6:30	labuan ibu
24	20200413	malaysiaairlines	economy_promo	20200420	mh3041	7:05	0	kota kinat	6:30	labuan ibu
25	20200413	malaysiaairlines	economy_smart	20200420	mh3041	7:05	0	kota kinat	6:30	labuan ibu
26	20200413	malaysiaairlines	business_smart	20200421	mh3042	7:40	0	labuan ibu	7:05	kota kinabalu bki
27	20200413	malaysiaairlines	economy_flex	20200421	mh3042	7:40	190	labuan ibu	7:05	kota kinabalu bki
28	20200413	malaysiaairlines	economy_promo	20200421	mh3042	7:40	0	labuan ibu	7:05	kota kinabalu bki
29	20200413	malaysiaairlines	economy_smart	20200421	mh3042	7:40	0	labuan ibu	7:05	kota kinabalu bki
30	20200413	malaysiaairlines	business_smart	20200420	mh3043	9:00	0	kota kinat	8:25	labuan ibu
31	20200413	malaysiaairlines	economy_flex	20200420	mh3043	9:00	190	kota kinat	8:25	labuan ibu
32	20200413	malaysiaairlines	economy_promo	20200420	mh3043	9:00	0	kota kinat	8:25	labuan ibu
33	20200413	malaysiaairlines	economy_smart	20200420	mh3043	9:00	0	kota kinat	8:25	labuan ibu
34	20200413	malaysiaairlines	business_smart	20200421	mh3044	16:10	0	labuan ibu	15:35	kota kinabalu bki
35	20200413	malaysiaairlines	economy_flex	20200421	mh3044	16:10	190	labuan ibu	15:35	kota kinabalu bki
36	20200413	malaysiaairlines	economy_promo	20200421	mh3044	16:10	0	labuan ibu	15:35	kota kinabalu bki
37	20200413	malaysiaairlines	economy_smart	20200421	mh3044	16:10	0	labuan ibu	15:35	kota kinabalu bki

Download Data

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

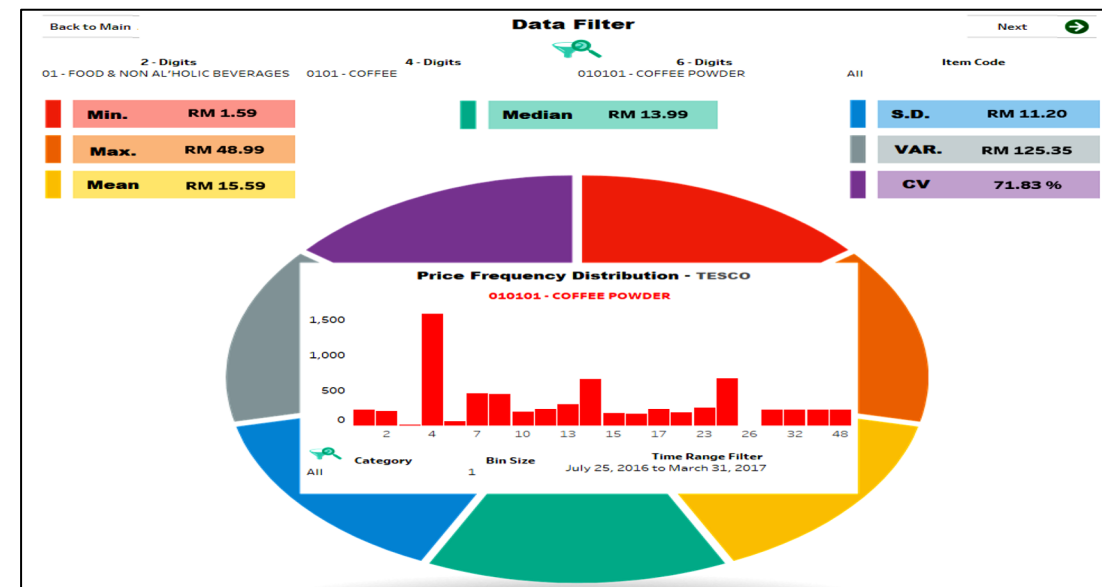


a) Trend

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



MARCH –
MAY 2020

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)

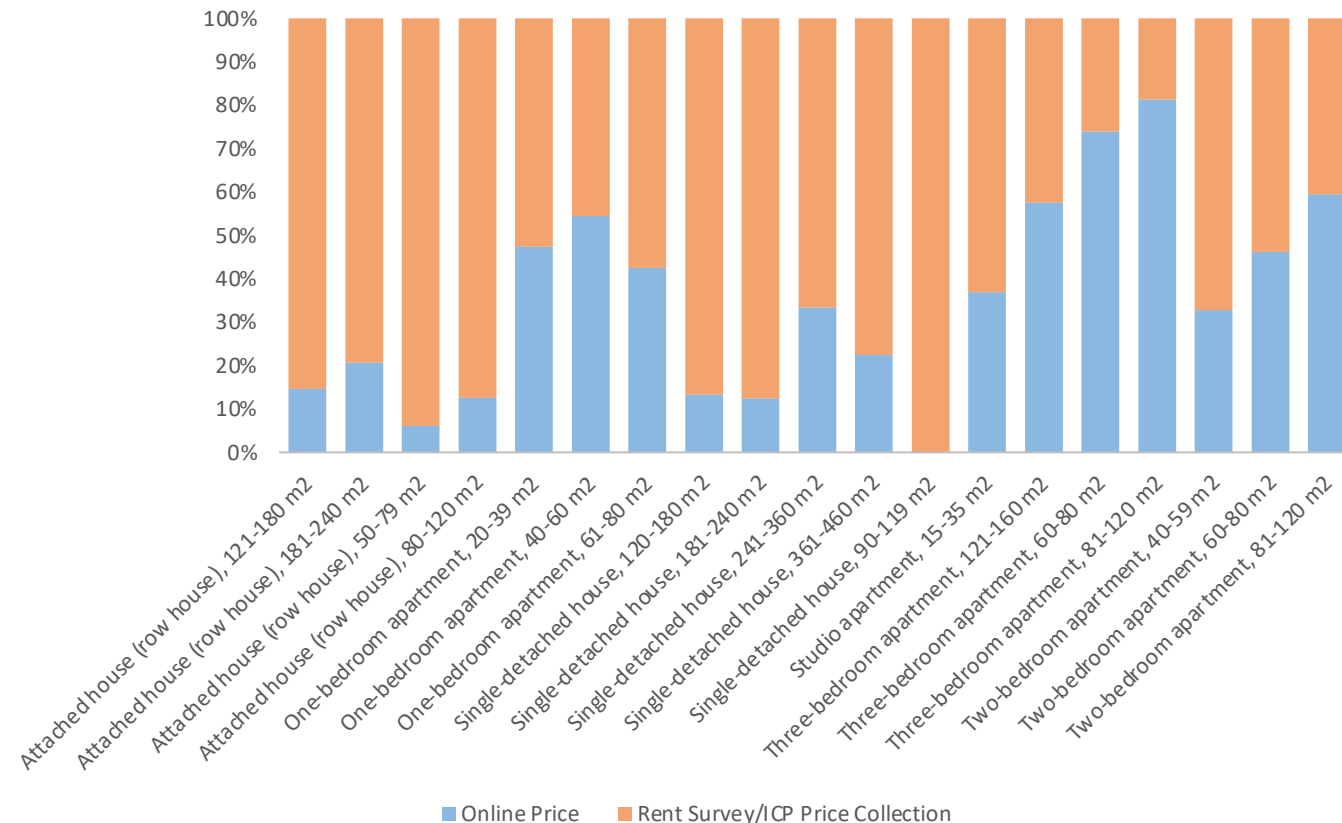




HOUSE RENTAL PRICES FOR INTERNATIONAL COMPARISON PROGRAM (ICP)

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

RENTAL HOUSING DATA SOURCED USED FOR ICP 2017



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



- 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

1

COMPARISON

To study the difference in price and price movement between DOSM data and administrative data

2

DISPLAY

Displays statistics on the distribution of prices of goods collected by DOSM and administrative data.

3

RECOMMENDATIONS

Provide recommendations for the replacement of outlets and reduction of price collection in the field by using the price collection channeled from administrative data

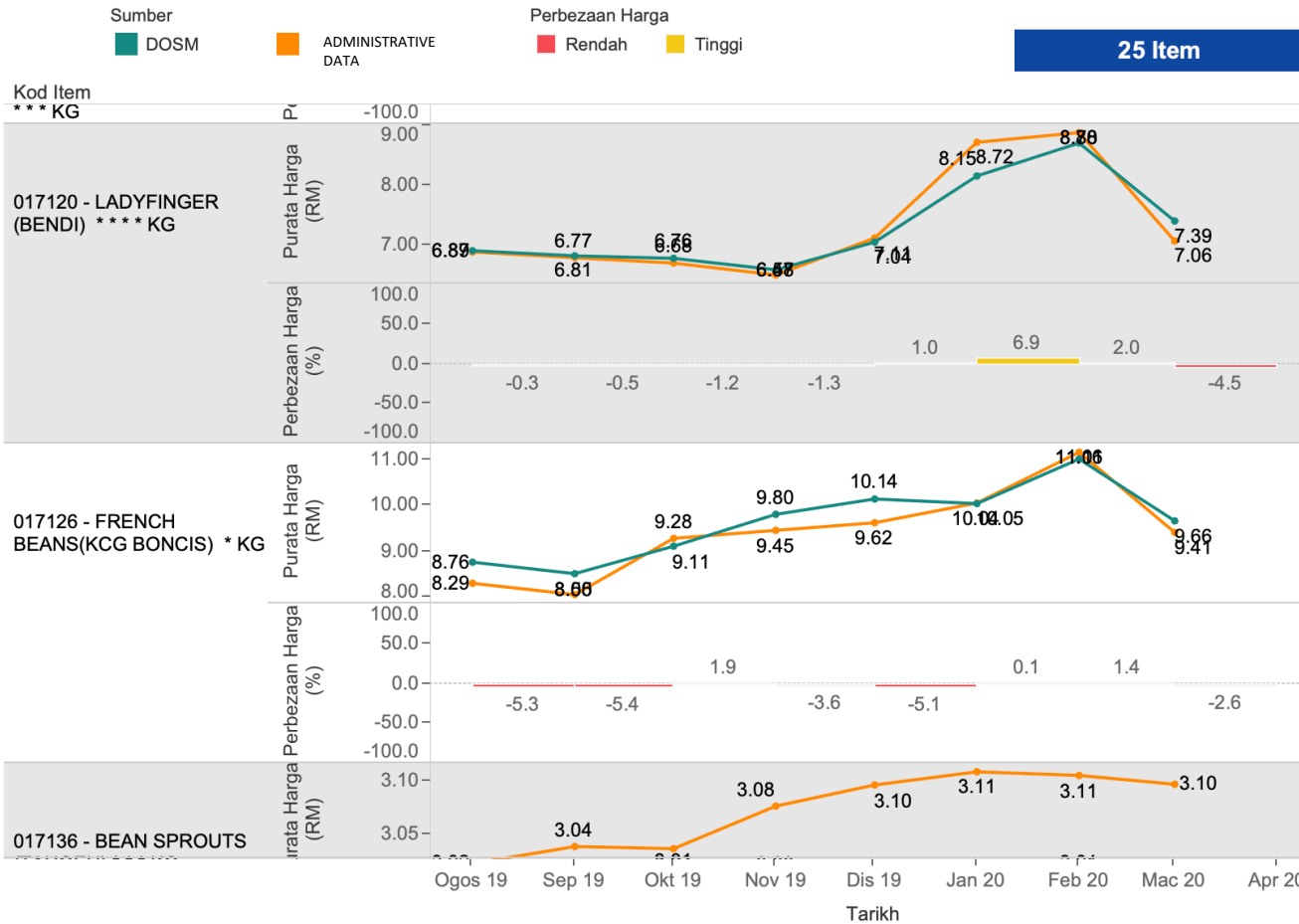
4

OPTIMIZING

Optimizing usage of data sources in other government agencies as a complement in generating CPI



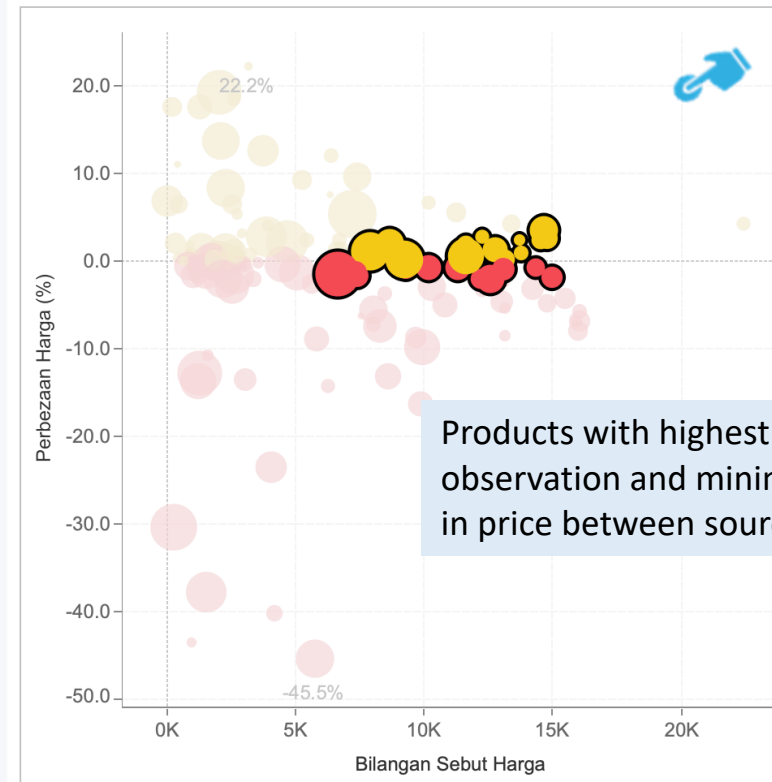
Trend Harga Purata Bulanan dan Peratus Perbezaan Harga antara DOSM



Taburan Perbezaan Harga bagi Setiap Item

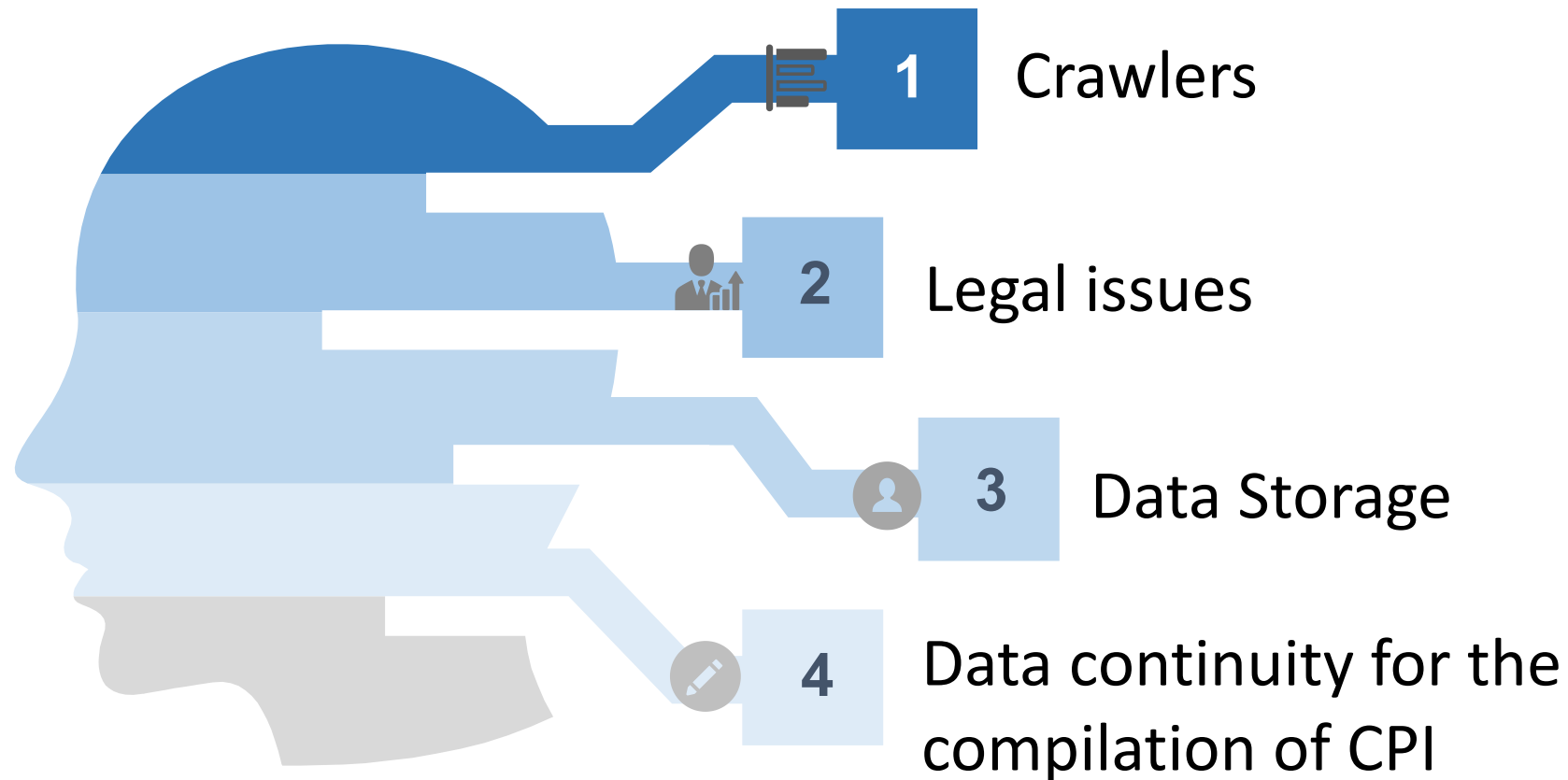
vs DOSM dan Bilangan Sebut Harga

Perbezaan Harga: Rendah (Red), Tinggi (Yellow)



Nota:
Kebesaran saiz titik di carta merujuk kepada purata harga item tersebut

CHALLENGES/LIMITATION





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

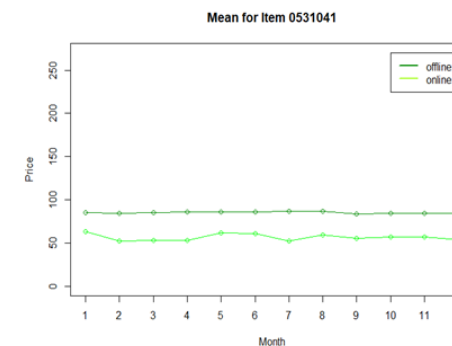
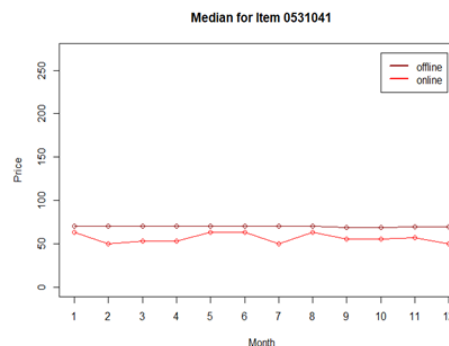
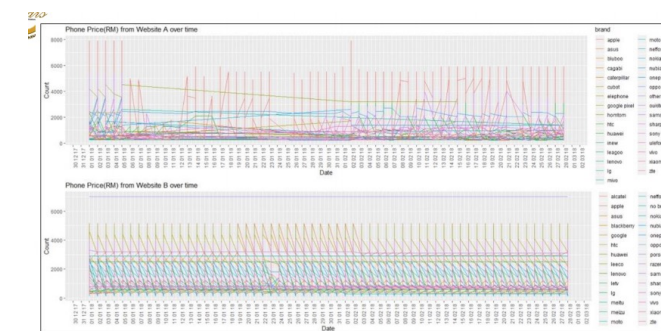
- Mohd Saiful Husain and Norsyela Muhammad Noor Mathivanan

3 Analysis of the Mobile Phones Prices Malaysia using Web Scraped Data

- Nur Hurriyatul Huda Abdullah Sani

Table 3: ARIMA best model

Item	Model
Red Bream	$Y_t = 2.4205 - 0.0333Y_{t-1} + 0.6382Y_{t-2} + 0.0594 + 0.9688e_{t-1}$
Selar Kuning	$Y_t = 1.4794 - 1.6648Y_{t-1} - 0.7431Y_{t-2} + e_t - 0.7924e_{t-1}$
Green Spinach	$Y_t = 1.1813 + 1.7246Y_{t-1} - 0.7774Y_{t-2} + 0.1008 - 0.8126e_{t-1}$
Kangkung	$Y_t = 1.0995 + 1.7293Y_{t-1} - 0.7776Y_{t-2} + 0.1023 - 0.8315e_{t-1}$
Long Beans	$Y_t = 0.5460 + 0.3667Y_{t-1} + 0.0780 + 0.4322e_{t-1} + 0.2638e_{t-2}$
Bawal, Cencaru, Kembong, Round Cabbage and Sawi Jepun	$Y_t = \mu + Y_{t-1}$ where μ : mean of the changes of period to period





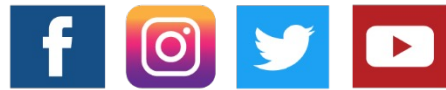
<https://statsbda.dosm.gov.my/>

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

THANK YOU



StatsMalaysia

www.DOSM.gov.my



2022



2012-2022



20 OKT



2016-2030

