Big Data Analytics
Department of Statistics Malaysia
Experience

The International Symposium on the Use of Big Data for Official Statistics
16-18 October 2019
Hangzhou, China

Big Data and Department of Statistics Malaysia, DOSM

Outlines

01 Introduction

02 Big Data Platform

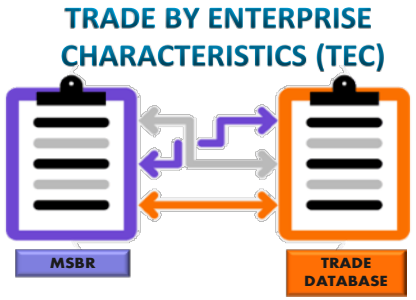
03 Methodology & Results

04 Way Forward

STATSBDA JOURNEY

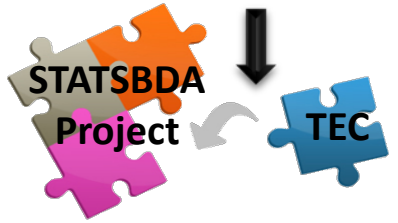
Inception Meeting in Colombo, Sri Lanka
30 June -2 July 2014
Organizer: ADB

CDTA: 8594 - Statistical Business Registers for Improved Information on Small Medium-Sized and Large Enterprises.



MSBR rich with businesses information & become richer when it is combined with trade database

Seminar on Integrated Trade Database with MSBR: A Big Data Initiative
17 September 2015
Sponsored by: ADB
Consultant: Mr. Markie Muryawan
(Chief International Merchandise, Trade Statistics Section, UNSD)



Embedded in STATSBDA Project - decide to harness the massive amounts of trade data using big data platform.

EXPAND

STATSBDA Project December 2016
Structured Data (TEC)
+
Unstructured Data (PI & PMAOS)

STATSBDA BIG DATA ANALYTICS

Kick Start!

STATSBDA Kickoff
2 December 2016

Big Data Platform
Department of Statistics Malaysia, DOSM

TRADE BY ENTERPRISE CHARACTERISTIC (TEC)

Integration of MSBR with trade database - to identify the enterprises that are engaged in international markets and what are their characteristics.

PRICE INTELLIGENCE (PI)

The modernization of data collection mainly consists of the adoption of web scraping techniques to scrape price data from related website for CPI compilation and improving the quality of the data.

PUBLIC MATURITY ASSESSMENT ON OFFICIAL STATISTICS (PMAOS) & REAL TIME NEWS ON OFFICIAL STATISTICS (RTOS)

The analysis and assessment of the degree of "happiness" of Malaysia community towards to official statistics published by DOSM through online social media.

REAL TIME BUSINESS STATUS (RTBS)

Integrate data from companies Commission of Malaysia(CCM) to DOSM environment to enable the organization to have direct access of information of the business entities.

BizCode@Stats

Department of Statistics Malaysia (DOSM)

BizCode@Stats is a platform for sharing activity code and statistics classification from DOSM which developed based on International Classification Standards. BizCode@Stats was developed to allow users to get code and description for interactive and mobile.

Trade by Enterprise Characteristic (TEC)

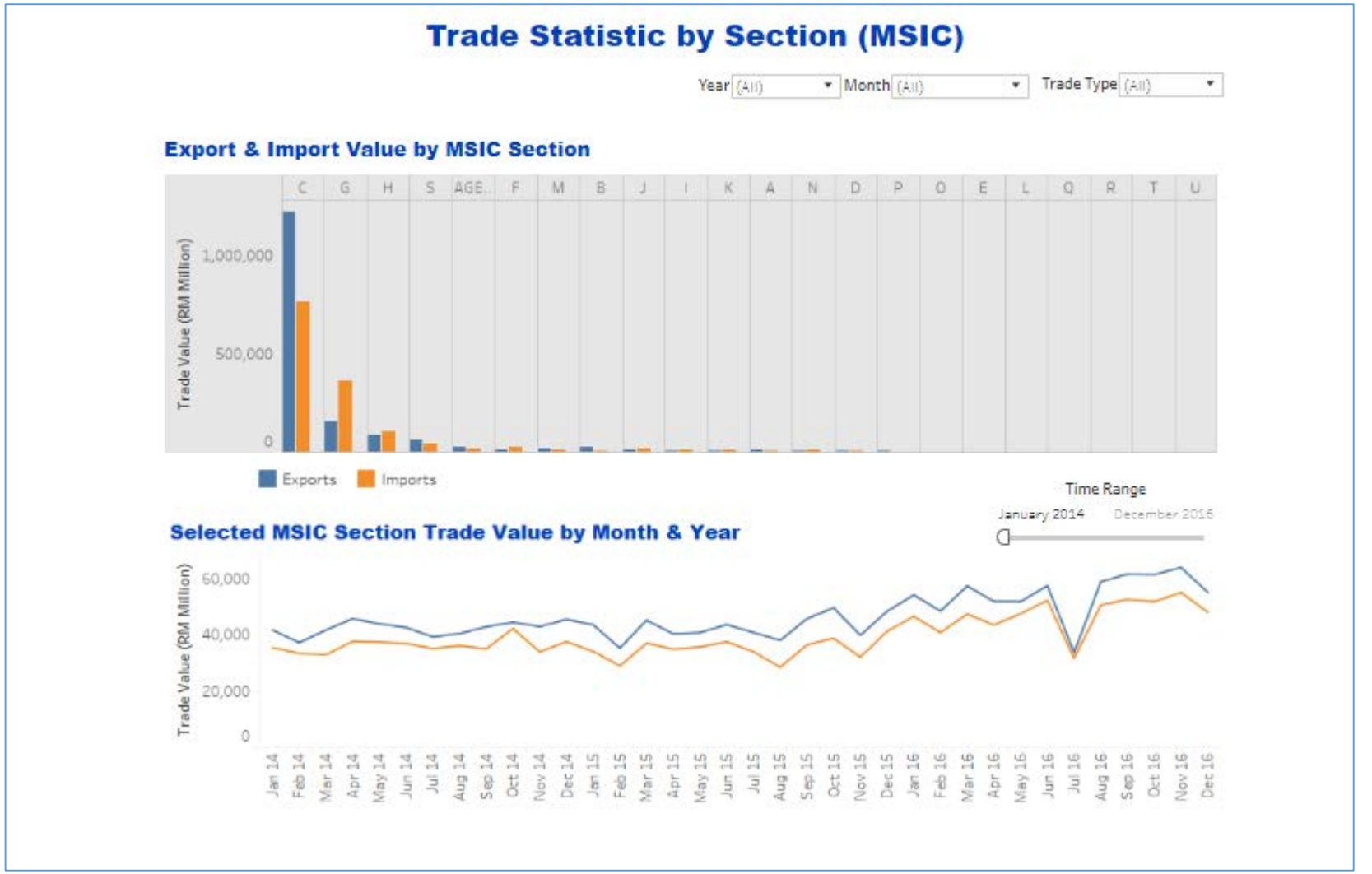
Integration of MSBR with trade database - to identify the enterprises that are engaged in international markets and what are their characteristics.

- **Objectives:**
  - the linkage between these two databases can describe the structure and characteristics of trading enterprises involved in the global market
  - benefit for organization; to enhance productivity: Internal & Explicit stakeholders, Public
  - to provide statistical indicators on trade by enterprise characteristics

- **Example of results:**
  - forecast import/export value for each state
  - forecast number of SME importers/exporters by sector
  - forecast retained import value for each broad economic end use category

Trade by Enterprise Characteristic (TEC)-Internal Dashboard

Trade by Enterprise Characteristic (TEC)-Interactive Dashboard



Price Intelligence (PI)

The modernization of data collection mainly consists of the adoption of web scraping techniques to scrape price data from related website for CPI compilation and improving the quality of the data.

- **Objectives:**
  - to create a price list of different goods and provide the solution for consumers on the best prices available
  - Improve on price data quality
  - Offer an access to crawl large amounts of data
  - Reduce the overall workload for data collection
- **Methodology:**
  - via Web Scrapping
- **Example of results/analysis:**
  - trend of Average Price by State
  - average Price by State
  - trend of Average Price by Area
  - product Price Distribution by CPI Category

Price Intelligence (PI)- Internal Dashboard

DEPARTMENT OF STATISTICS MALAYSIA, OFFICIAL PORTAL
The Source of Malaysia's Official Statistics

PRICE INTELLIGENCE

Manual Survey

GO

1Pengguna

GO

TESCO

GO

Mudah

GO

Online Sources

GO

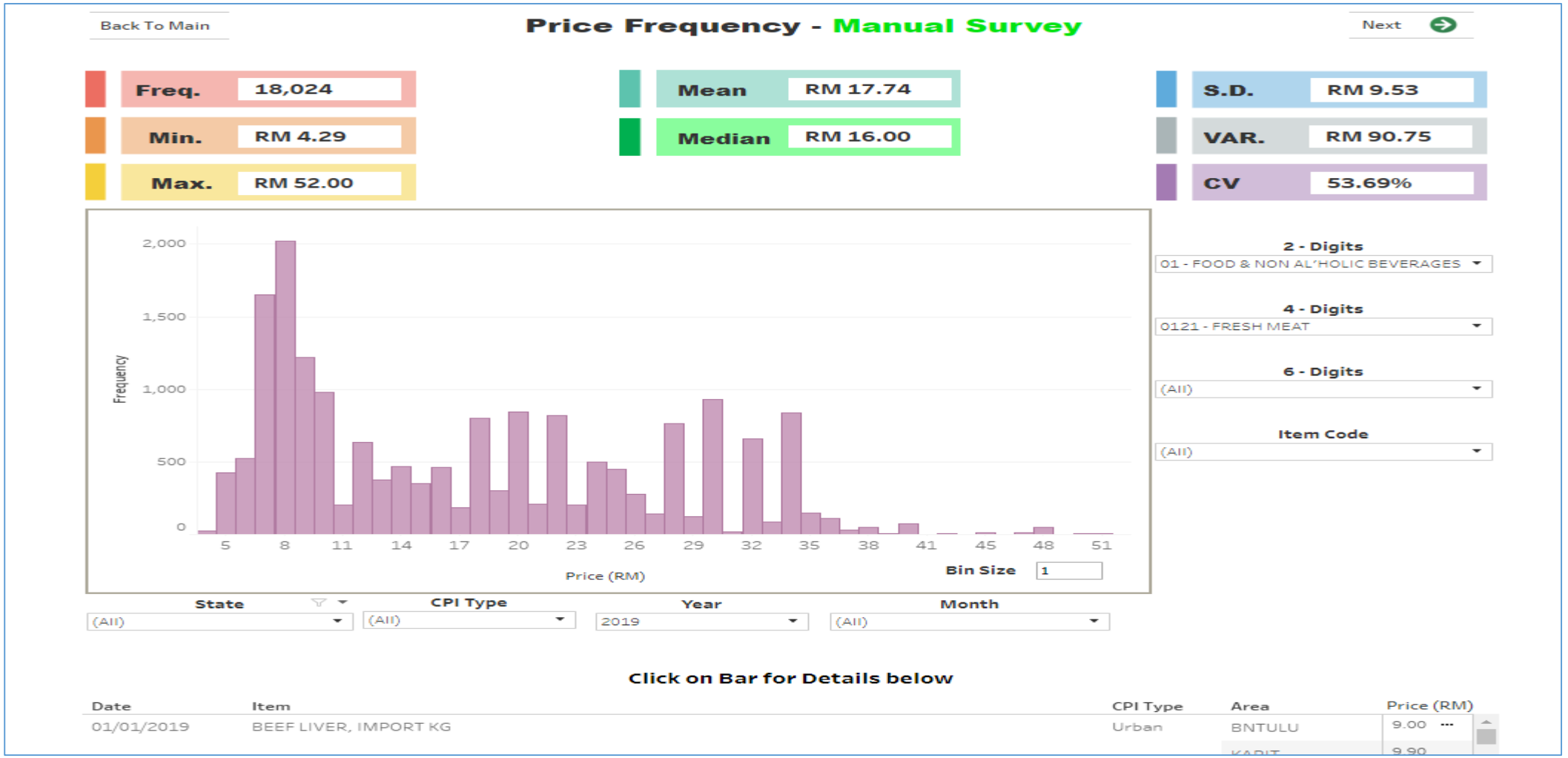
Combined Sources

GO

Index Data Correlation

GO

Price Intelligence (PI)- Internal Dashboard



REAL TIME BUSINESS STATUS (RTBS)

Integrate data from Companies Commission of Malaysia(CCM) to DOSM environment; to enable the organization to have direct access of information of the business entities.

- **Objectives:**
  - to build a strategic partnership between respective agencies
  - ways to get a direct access of business entities information
  - to provide web service
- **Methodology:**
  - strategic partnership between respective agencies

REAL TIME BUSINESS STATUS (RTBS)

REAL TIME BUSINESS STATUS (RTBS)

Integrate data from companies Commission of Malaysia(CCM) to DOSM environment to enable the organization to have direct access of information of the business entities.

The information received has the characteristics of integrity that is timely, accurate and reliable.

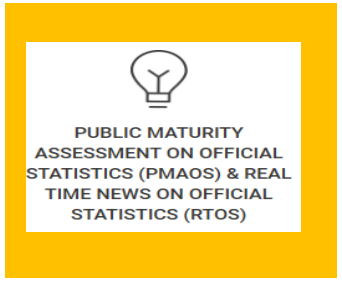
Checking business status centrally

Allow DOSM to get latest business status registered with SSM through SSM Middleware

SSM Middleware provides a platform for accessing information DOSM business entities registered with SSM

Reduce manual work processes where business information requests are made in batch form

Public Maturity Assessment on Official Statistics (PMAOS) & Real Time News on Official Statistics (RTOS)



The analysis and assessment of the degree of "happiness" of Malaysia community towards to official statistics published by DOSM through online social media.

- **Objectives:**
  - PMAOS and RTOS created to take all voices of community on statistics from the social media (Facebook comments and tweets).
  - Determine the degree of "happiness" of the community towards the official statistics.
- **Example of results:**
  - Opinion or sentiment data is generated through social channels in the form of reviews, chats, shares, likes tweets, etc.

Public Maturity Assessment on Official Statistics (PMAOS) & Real Time News on Official Statistics (RTOS)- Internal Dashboard

Back To Main

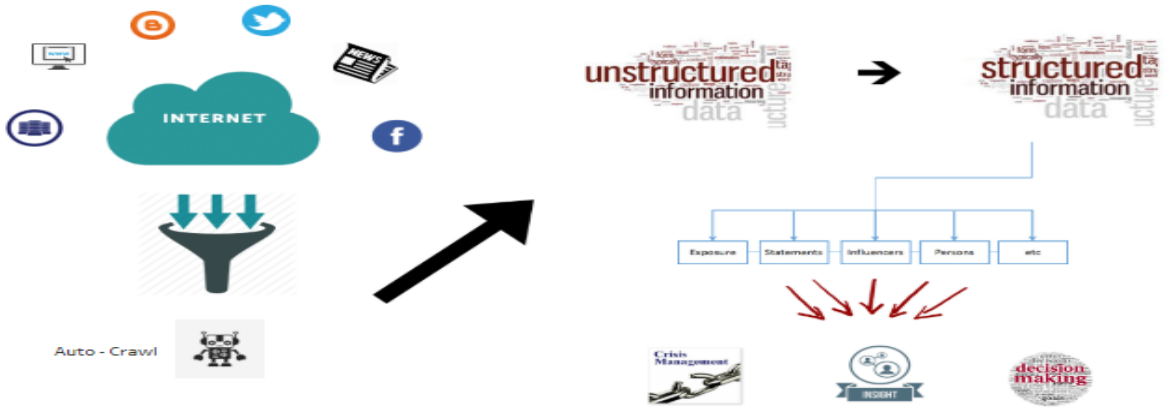
DEPARTMENT OF STATISTICS MALAYSIA
The Source of Malaysia's Official Statistics

DOSM MEDIA INTELLIGENCE

ISSUE MONITORING	TWITTER MONITORING	FACEBOOK MONITORING	TOPIC DEPENDENCY	ARTICLE BANK
GO	GO	GO	GO	GO

What is DOSM Media Intelligence?

DOSM Media Intelligence is a media source to spread information to the public in fast mode. Furthermore, DOSM is observed daily issues that happened around us. In specific, there are several components that DOSM observed such as: topic issue, the most influencing person, the related organization and etc. The reason why DOSM need the media intelligence is to simplify the analysis process of some topic. DOSM will be able to prevent, manage and solved the issue based on the analysis result using media intelligence.



Public Maturity Assessment on Official Statistics (PMAOS) & Real Time News on Official Statistics (RTOS)- Internal Dashboard

Back To Main

DOSM - General Overview

Next Page

News Exposure

News exposure is the count of article published by media related to specific topic. Which is gathered from various mainstream, portal

Number of News

632

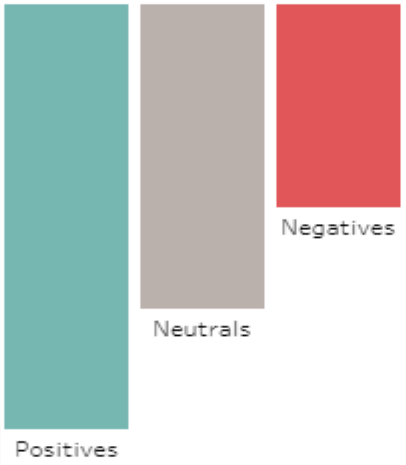
Statements

Statement from certain public figure counted for analysis purpose. The statement will also has tone which will be grouped into positive, neutral and negative sentiments

Number of Statements

2,748

What Do They Say?



Sentiment Analysis

All the statements that goes into analysis will be labelled by their sentiment tone. It's not 100% accurate, but can give insight when a topic is hot if the negative tone is more dominant than the others.

Influencers

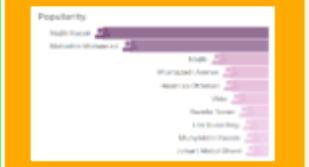
Only certain person gaining attention of media can be influencer and give their statements. Each and every one of them, counted.

Number of Influencers

293

Top Persons

Media sometimes mention ceratin person in their articles. They are the top persons and each one of them is counted

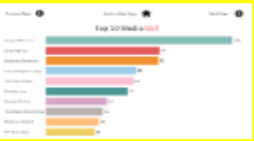


Number of Persons

613

Media

Each one of media released their articles is counted for analysis of their interest in an issues.



Number of Media

239

STATISTICAL CLASSIFICATION(BizCode@Stats)

BizCode@Stats
Department of Statistics Malaysia
(DOSM)

BizCode@Stats is a platform for sharing activity code and statistics classification from DOSM which developed based on International Classification Standards. Allow users to access the application via handheld devices.

- **Objectives:**
  - is built as a platform for sharing codes and the classification of economic statistics in Malaysia
- **Methodology:**
  - mobile Application
  - the application via handheld devices.

DOSM Journey on BIG DATA

Way Forward

- DOSM is in preliminary phase; collaboration with Malaysian Space Agency (MYSA)
- Satellite Image
- To strengthen Malaysia Statistical Address Register (MSAR)

70 years 1949 - 2019 Department of Statistics Malaysia

TERIMA KASIH & THANK YOU

HARI STATISTIK NEGARA	62nd ISI World Statistics Congress ISI 2019 Kuala Lumpur 18 - 23 August 2019	POPULATION & HOUSING CENSUS MALAYSIA 2020 YOUR DATA IS OUR FUTURE	HIES/BA 2019 Data Shaping People's Wellbeing	WRC 2019 Kaji Perdagangan Borong & Runcit	SUSTAINABLE DEVELOPMENT GOALS MALAYSIA
20 OCT	18 - 23 AUG 2019	JULY 2020 (ACTUAL MYCENSUS)	JAN – DEC 2019	MAR – SEPT 2019	2015 - 2030

#StatsMalaysia | #MyStatsDay | #ISIWSC2019 | #MyCensus2020 | #HIES2019 | #MyRetailCensus2019 | #LeaveNoOneBehind