



# STATSBDA: Accelerates Statistics through Big Data

Speaker:

Nur Aziha Mansor@Noordin

Principal Assistant Director

Methodology & Research Division

Department of Statistics Malaysia

# Outlines

**About DOSM**

**BDA Journey**

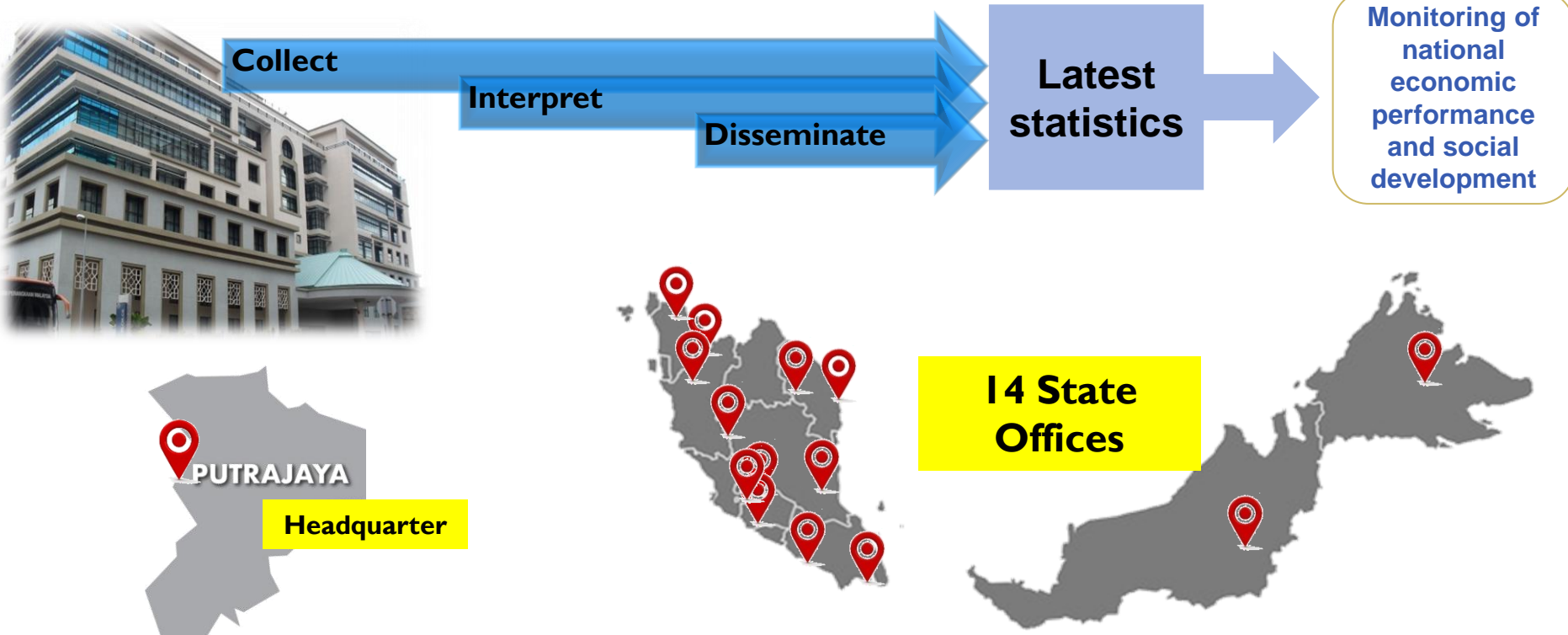
**DOSM BDA Environment**

**STATSBDA Architecture**

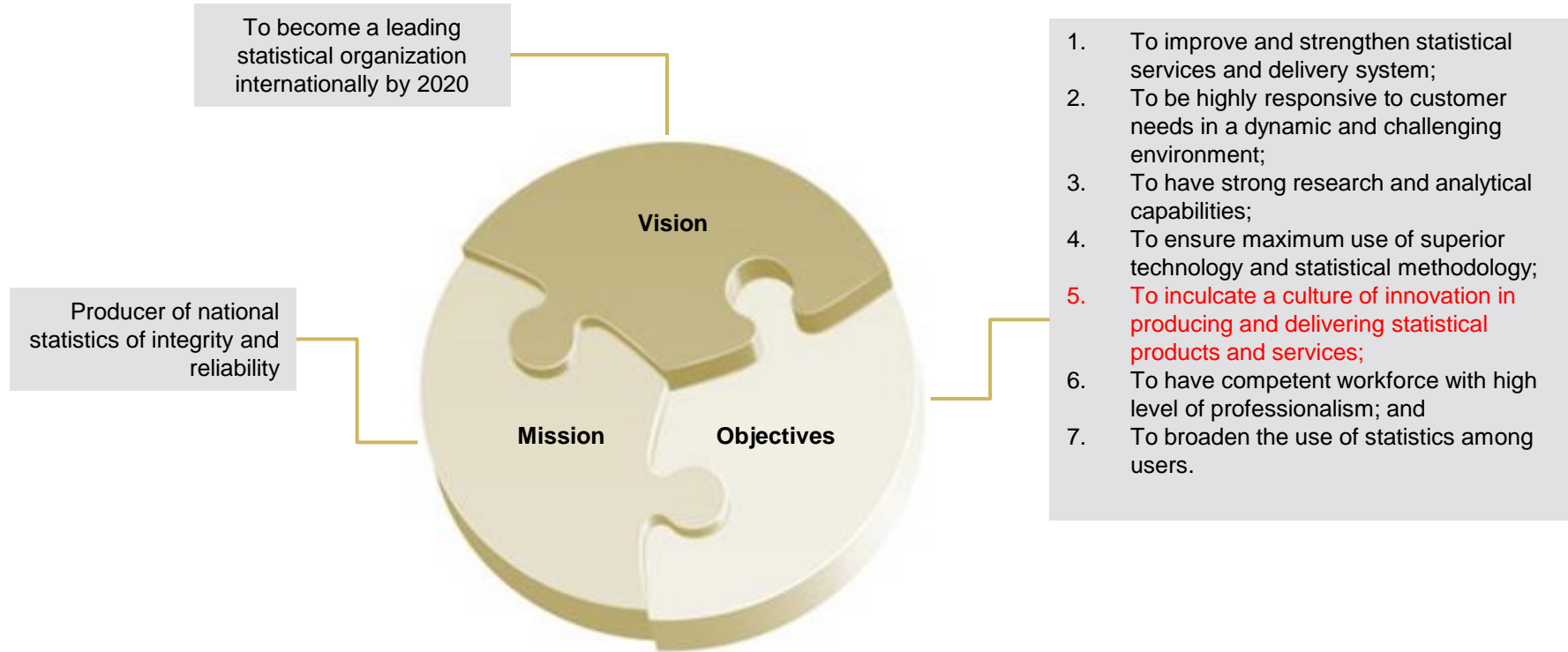
**STATSBDA Module**

**Conclusion**

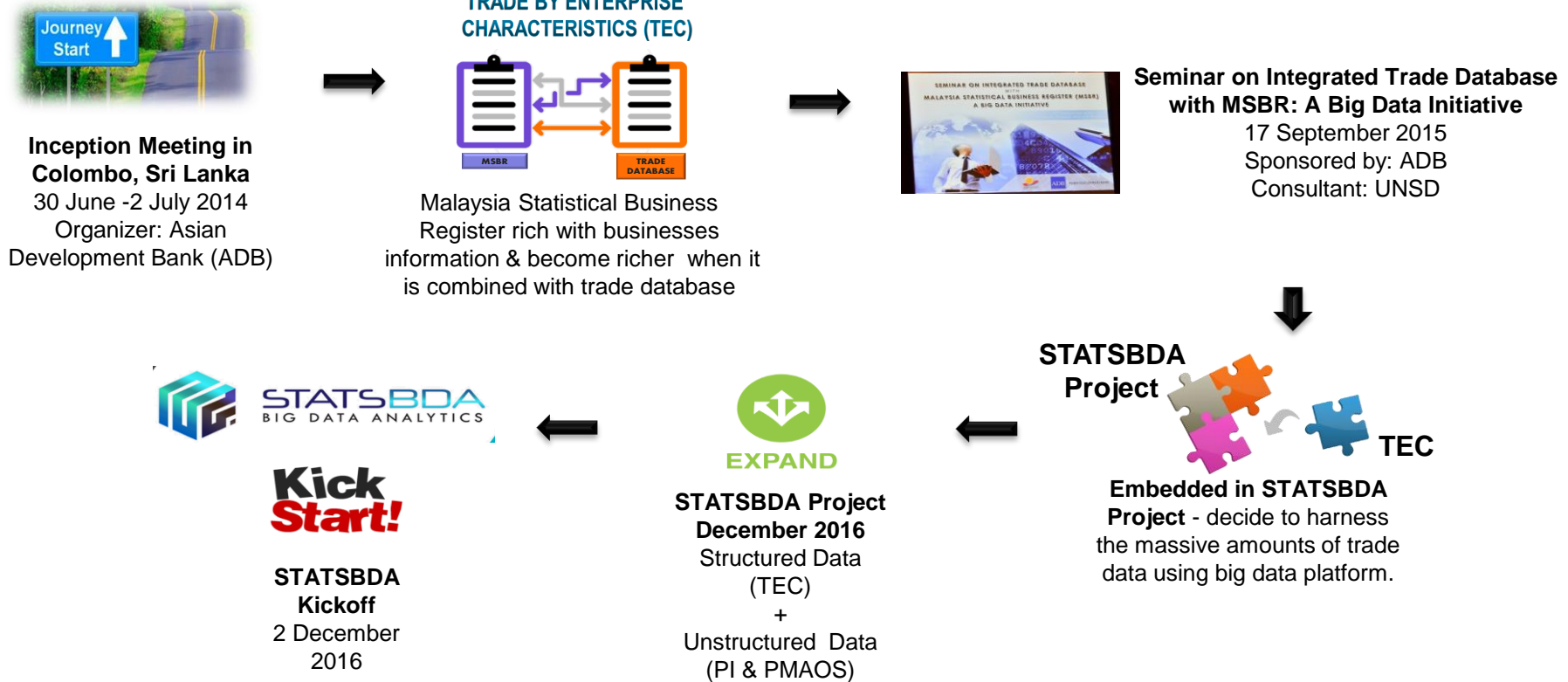
# About DOSM



# About DOSM

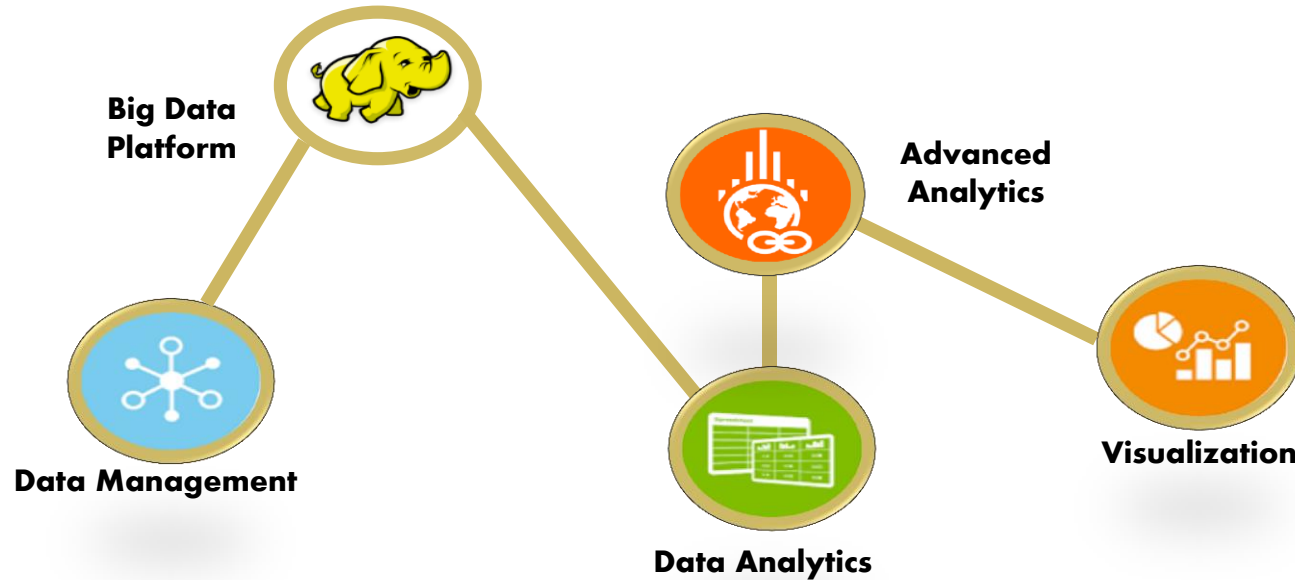


# Big Data Analytics Journey

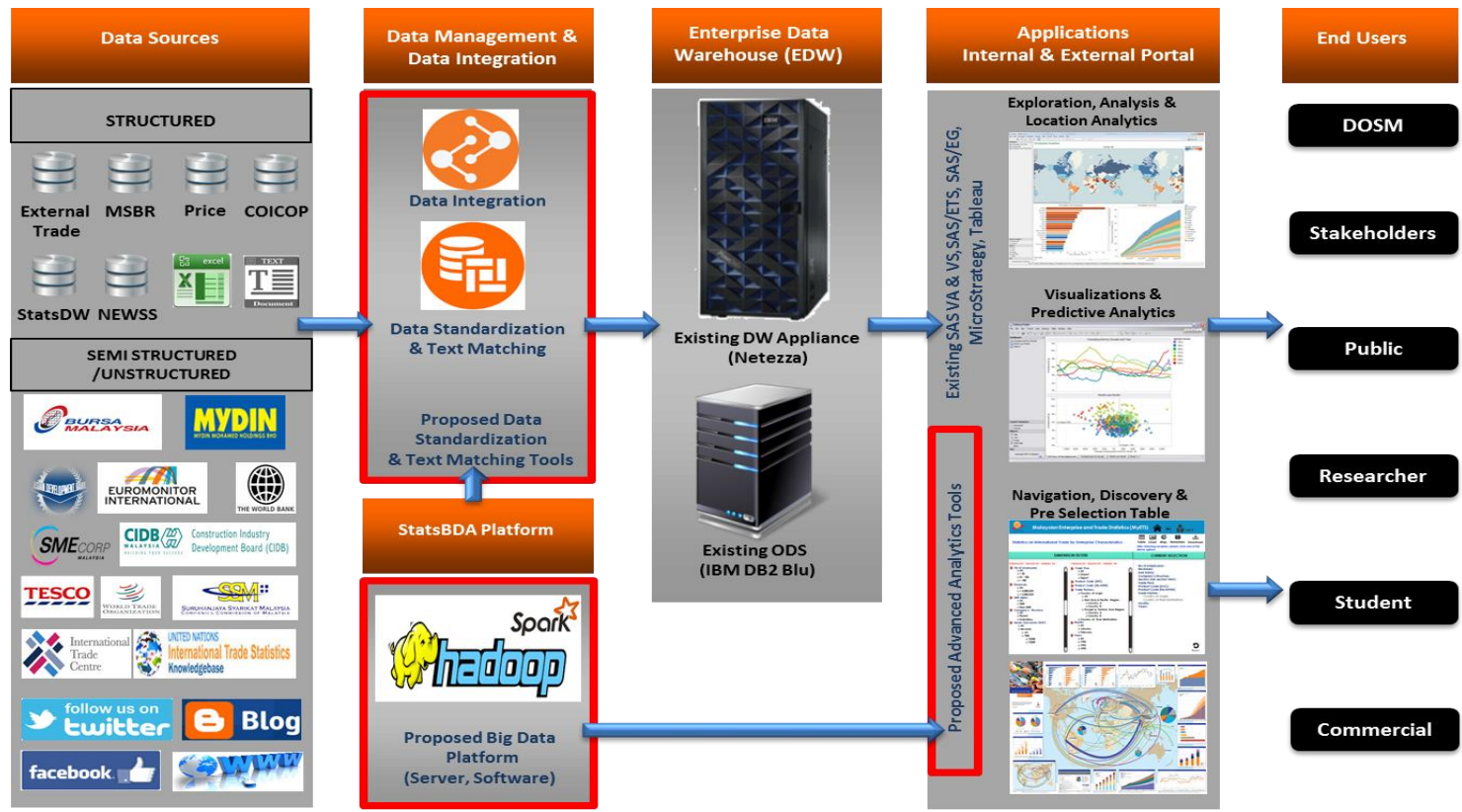


# DOSM Big Data Analytics Environment

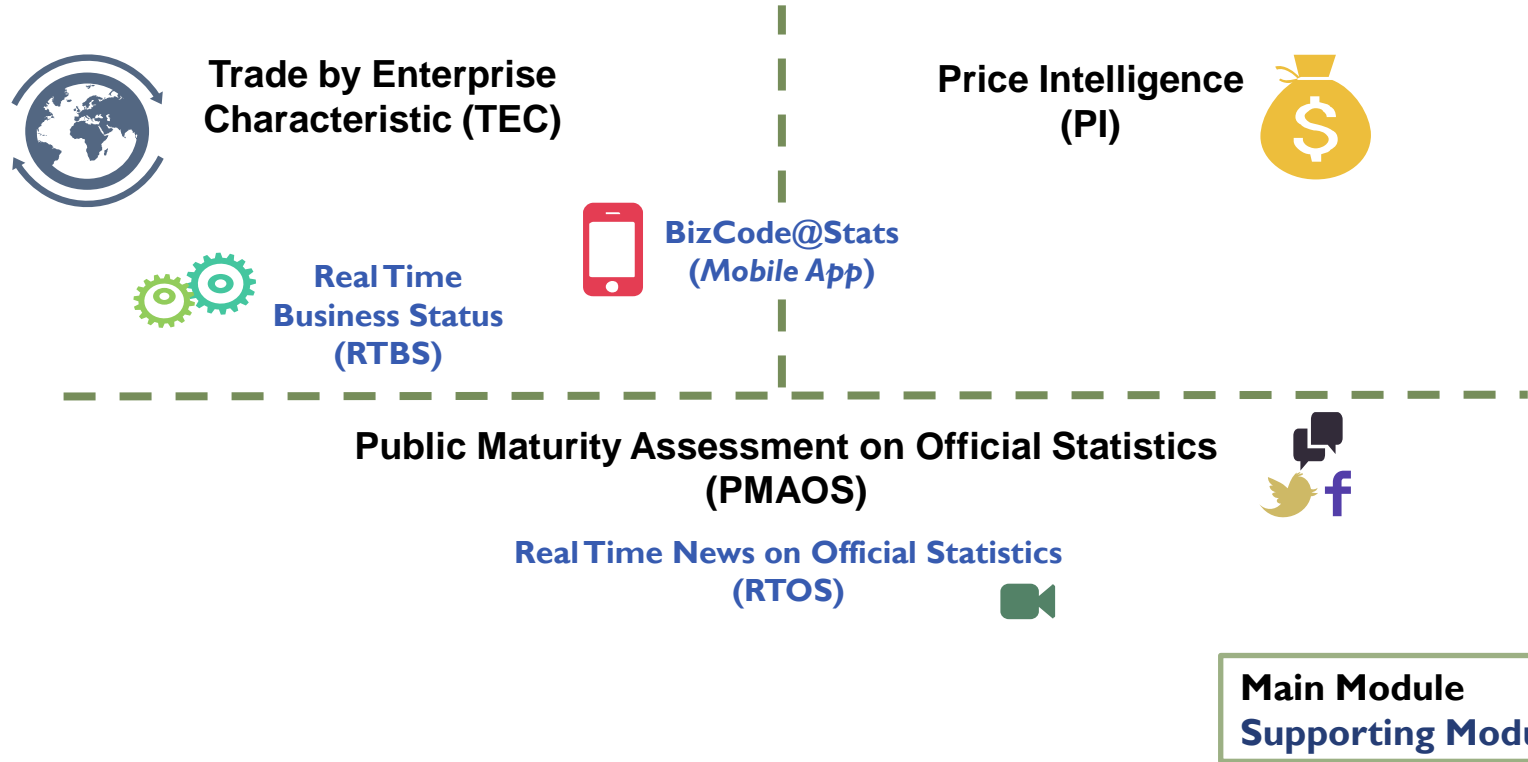
## Integrated Big Data Analytics Environment



# STATSBDA Architecture



# STATSBDA Modules





## STATSBDA

**Structured  
Data**

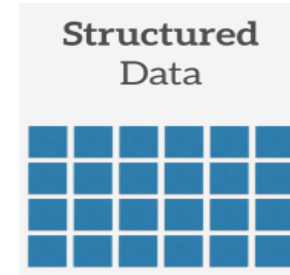
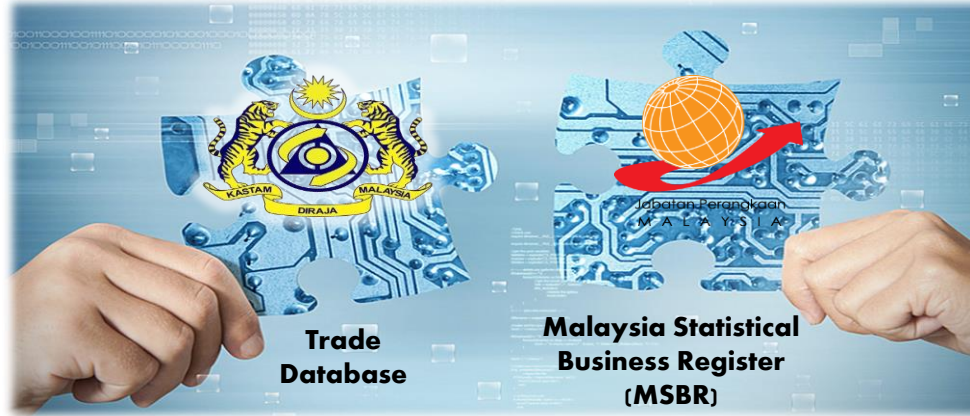
**Unstructured Data**

**Trade by Enterprise  
Characteristic (TEC)**

**Price Intelligence (PI)**

**Public Maturity  
Assessment on Official  
Statistics (PMAOS)**

# Trade by Enterprise Characteristic (TEC)

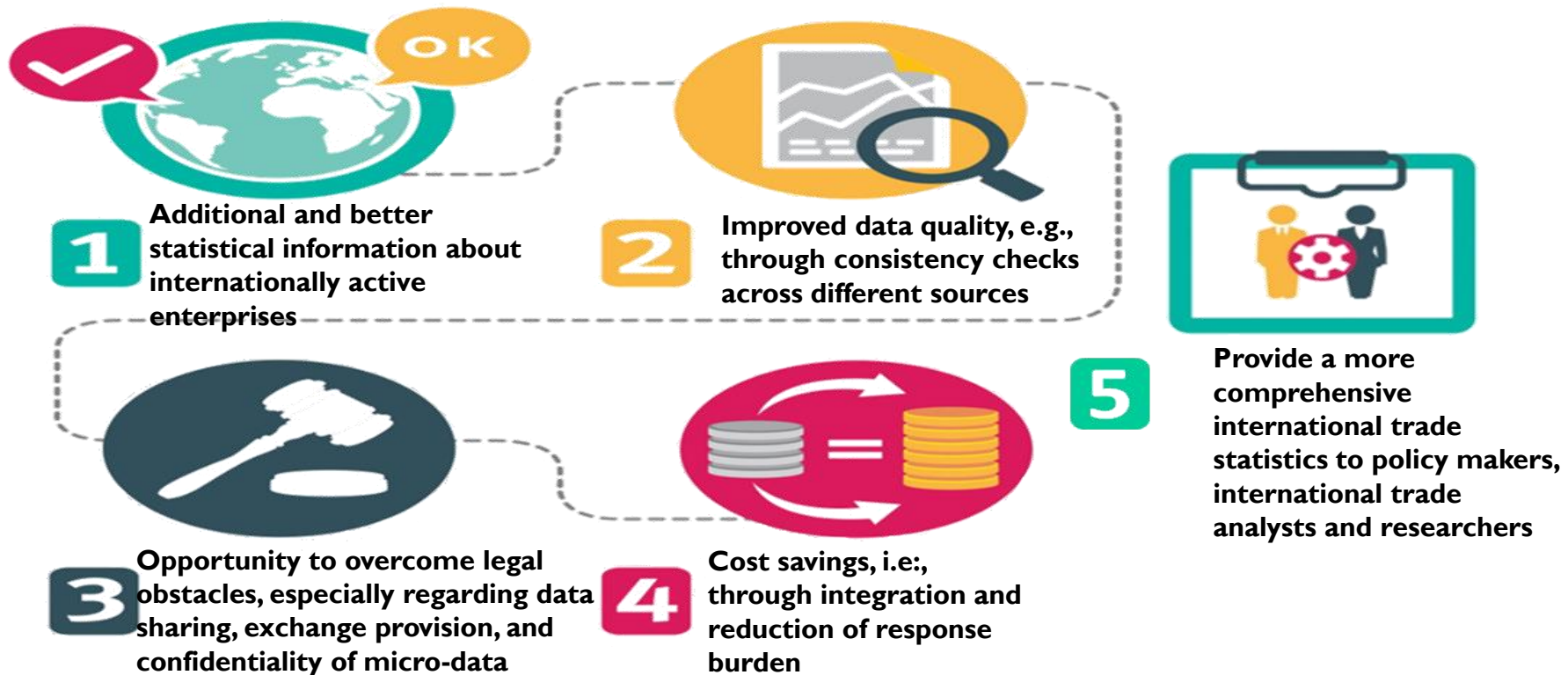


High volume of administrative data from JKDM is integrated with MSBR to **provide value added** to an existing international trade statistics **without initiating new survey**.

These statistics are able to identify the characteristics of companies or businesses involved global market.



# Trade by Enterprise Characteristic (TEC)



# Trade by Enterprise Characteristic (TEC)

- **Descriptive**

- Combination of Trade and Company Info (Trade Enterprise Characteristic)
- Able to deep dive to more granular level
- Benefit for DOSM to enhance productivity: Stakeholders, Public

- **Forecasting**

- Gain ability to forecast
- Maximize the potential of the products or services produced or offered by Malaysia
- Have an expectation for next period of trade value
- Alert users when trend is out of expectation and DOSM enable to find the reasons at early stage

# Trade by Enterprise Characteristic (TEC)

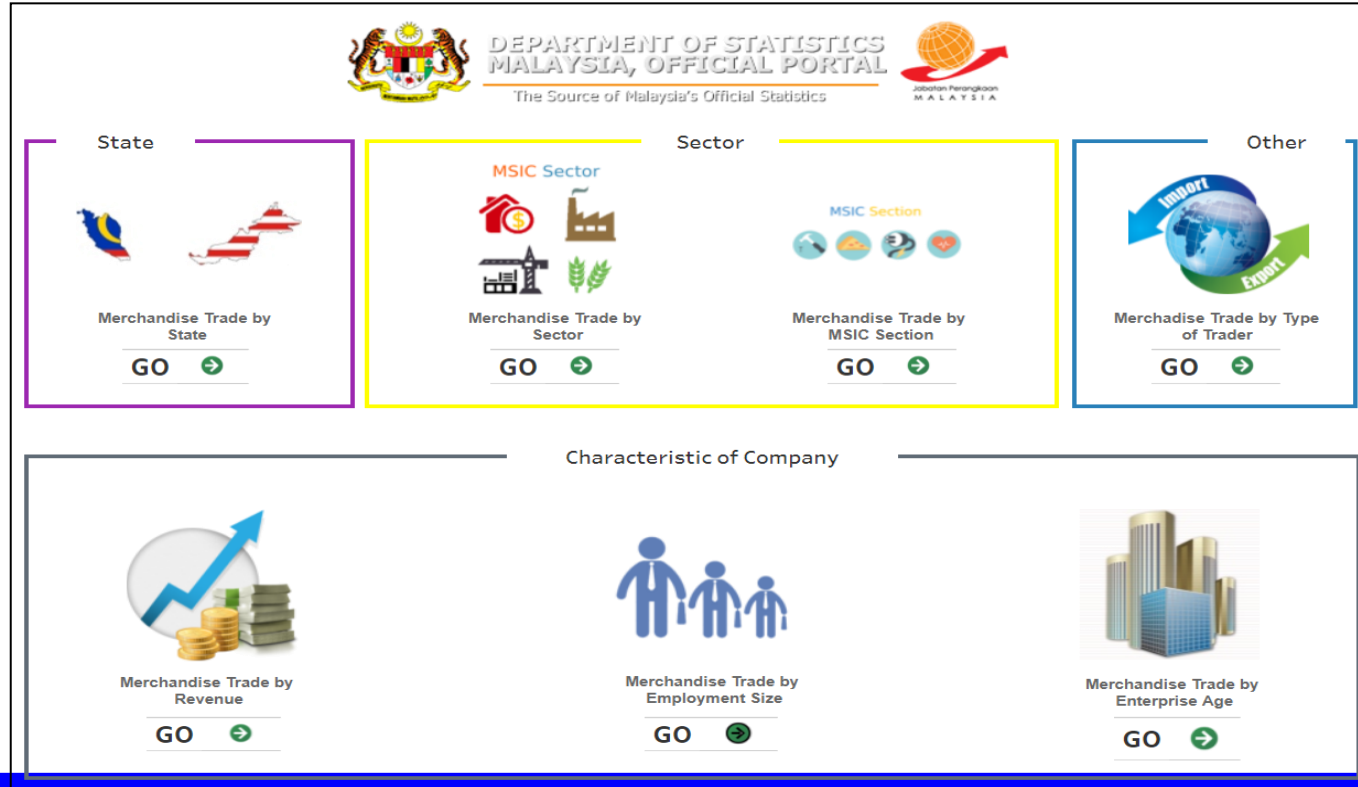
## **Descriptive**

- 1) Overall
- 2) Product Code
- 3) Major Trading Partners
- 4) State
- 5) FTA & Non-FTA Trading Partners Tag
- 6) Two-Ways Trader Statistics
- 7) Border State Statistics
- 8) High-Tech Product Statistics
- 9) Broad Economic Category Statistics
- 10) Halal Products Export
- 11) Top Exporter / Importer by Product
- 12) Product by Importer Exporter
- 13) Business Status
- 14) Trade Statistics by Sector
- 15) Trade Statistics by Section (MSIC)
- 16) Economic Region Statistics
- 17) Revenue & Employment Size Band Statistics
- 18) Manufacturing Statistics
- 19) Enterprise Age Statistics
- 20) High Demand Product Statistics

## **Forecasting**

- 1) Forecast Total Import/Export Value by Sector for Overall and Among SME
- 2) Forecast Import/Export Value for Each State
- 3) Forecast Number of SME Importers/Exporters by Sector
- 4) Forecast Total Import/Export Value by Top 10 Trading Partners
- 5) Forecast Retained Import Value for Each Broad Economic End Use Category
- 6) Forecast Import/Export Value for Each State
- 7) Forecast Total Import/Export Value of FTA Partner Country
- 8) Sustainability of Top 5
- 9) Forecast the Percentage Contribution in Export Value of Overall SME and Import & Export Value of Overall MNC Respectively at Monthly Level
- 10) Forecast the Overall Import & Export Trade Value at Monthly Level

# Trade by Enterprise Characteristic (TEC) - Visualization



## Merchandise Trade by State

Year: 2016 Trade Type: (All)

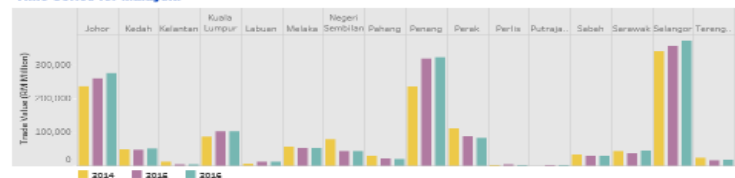
Import Value: RM 69,133.57 Million Export Value: RM 745,300.59 Million Total Value: RM 1,434,522.57 Million Balance of Trade: RM 66,166.92 Million

### Exports & Imports by State



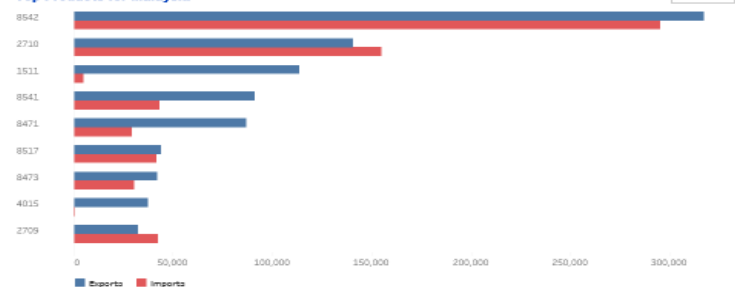
States	Share
Selangor	25.03%
Penang	22.51%
Johor	19.32%
Kuala Lumpur	7.00%
Kedah	5.52%
Malaka	4.60%
Perak	3.74%
Negeri Sembilan	2.71%
Sarawak	2.09%
Sabah	2.01%
Pahang	1.37%
Terengganu	1.29%
Perlis	0.17%
Labuan	0.08%
Kelantan	0.35%
Putrajaya	0.07%

### Time Series for Malaysia



\*Only Choose 1 State from Time Series

### Top Products for Malaysia



## Merchandise Trade by State

Year: 2016 Trade Type: (All)

Import Value: RM 117,219.00 Million Export Value: RM 159,000.15 Million Total Value: RM 277,189.01 Million Balance of Trade: RM 42,000.24 Million

### Exports & Imports by State



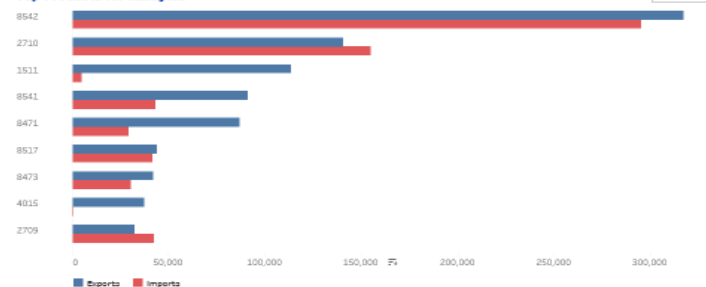
States	Share
Selangor	25.03%
Penang	22.51%
Johor	19.32%
Kuala Lumpur	7.00%
Kedah	5.52%
Malaka	4.60%
Perak	3.74%
Negeri Sembilan	2.71%
Sarawak	2.09%
Sabah	2.01%
Pahang	1.37%
Terengganu	1.29%
Perlis	0.17%
Labuan	0.08%
Kelantan	0.35%
Putrajaya	0.07%

### Time Series for Johor



\*Only Choose 1 State from Time Series

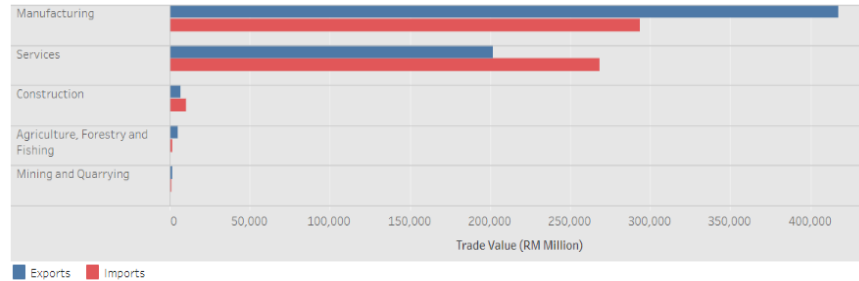
### Top Products for Malaysia



## Merchandise Trade by Sector

Year: 2016 Trade Type: (All)

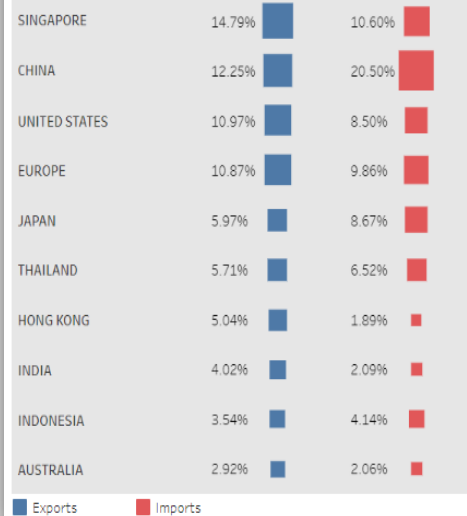
### Exports & Imports Value by Sector



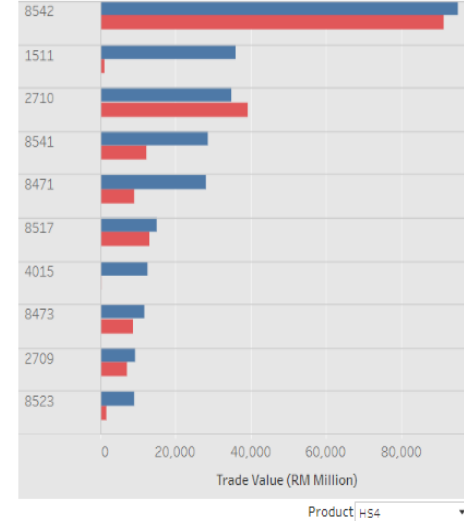
### Exports & Imports Share by Year



### Top Partner Country For All Sector



### Top Product For All Sector





# Price Intelligence (PI)

*National Statistical Institutes are increasingly using data from the internet as a source for their statistics*

## Modernization of data collection tools

### Adoption of web scraping techniques

- Automated processes implemented using a web crawler

## Data is gathered and copied from the web

- A collection of automated and semi automated techniques

## Data retrieval and analysis

Discovering;

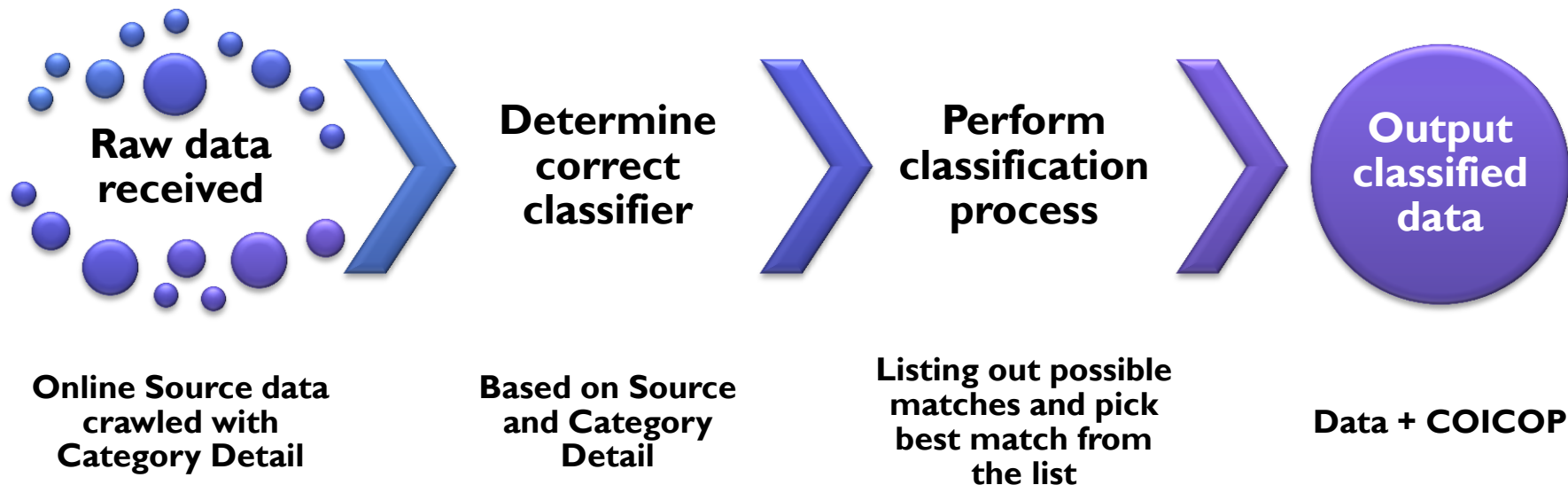
- Previously unknown patterns in data; and
- Relationship that can be used for prediction of user-relevant quantiles

*Statistical Journal of the IAOS 28(2012)*

Unstructured  
Data



# Price Intelligence (PI)



# Price Intelligence (PI)

1  
Improve on price  
data quality

2

Enables statisticians to  
react better to the  
increasing amount of data  
sources on the internet

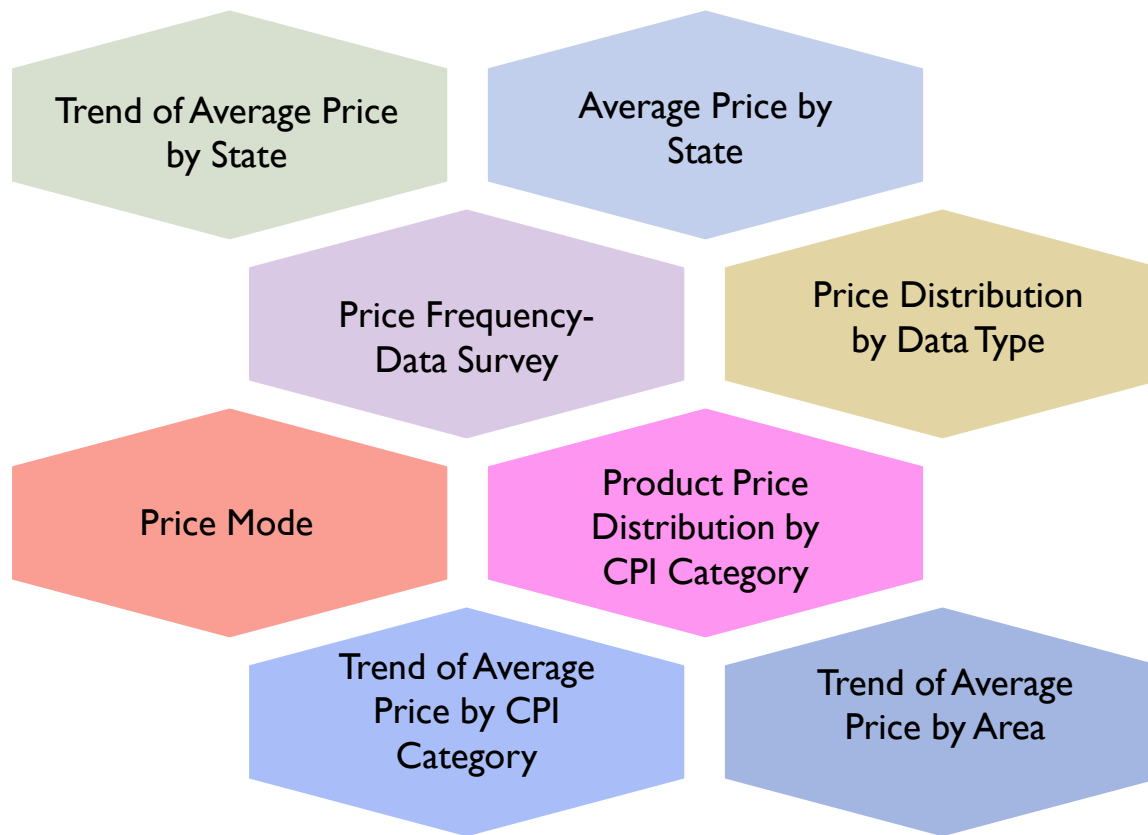
4  
Offer an access to  
crawl large  
amounts of data

3

Reduce the overall  
workload for data  
collection



# Price Intelligence (PI) - Visualization

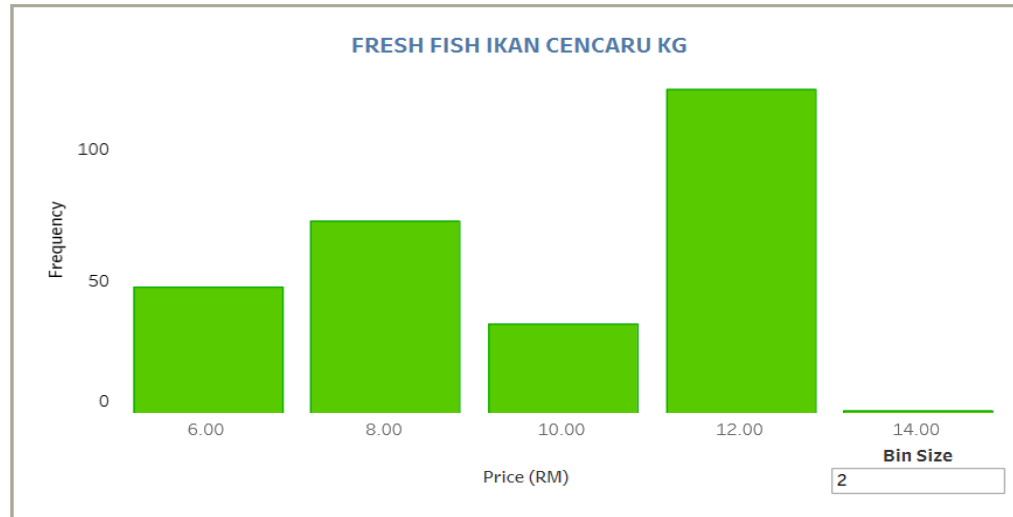


## Price Frequency - Online Source 1

<b>Freq.</b>	279
<b>Min.</b>	RM 6.99
<b>Max.</b>	RM 14.00

<b>Mean</b>	RM 10.38
<b>Median</b>	RM 10.00

<b>S.D.</b>	RM 2.17
<b>VAR.</b>	RM 4.72
<b>CV</b>	20.93 %



State: 
 Area: 
 Year: 
 Month: 
 Day:

2 - Digits:

4 - Digits:

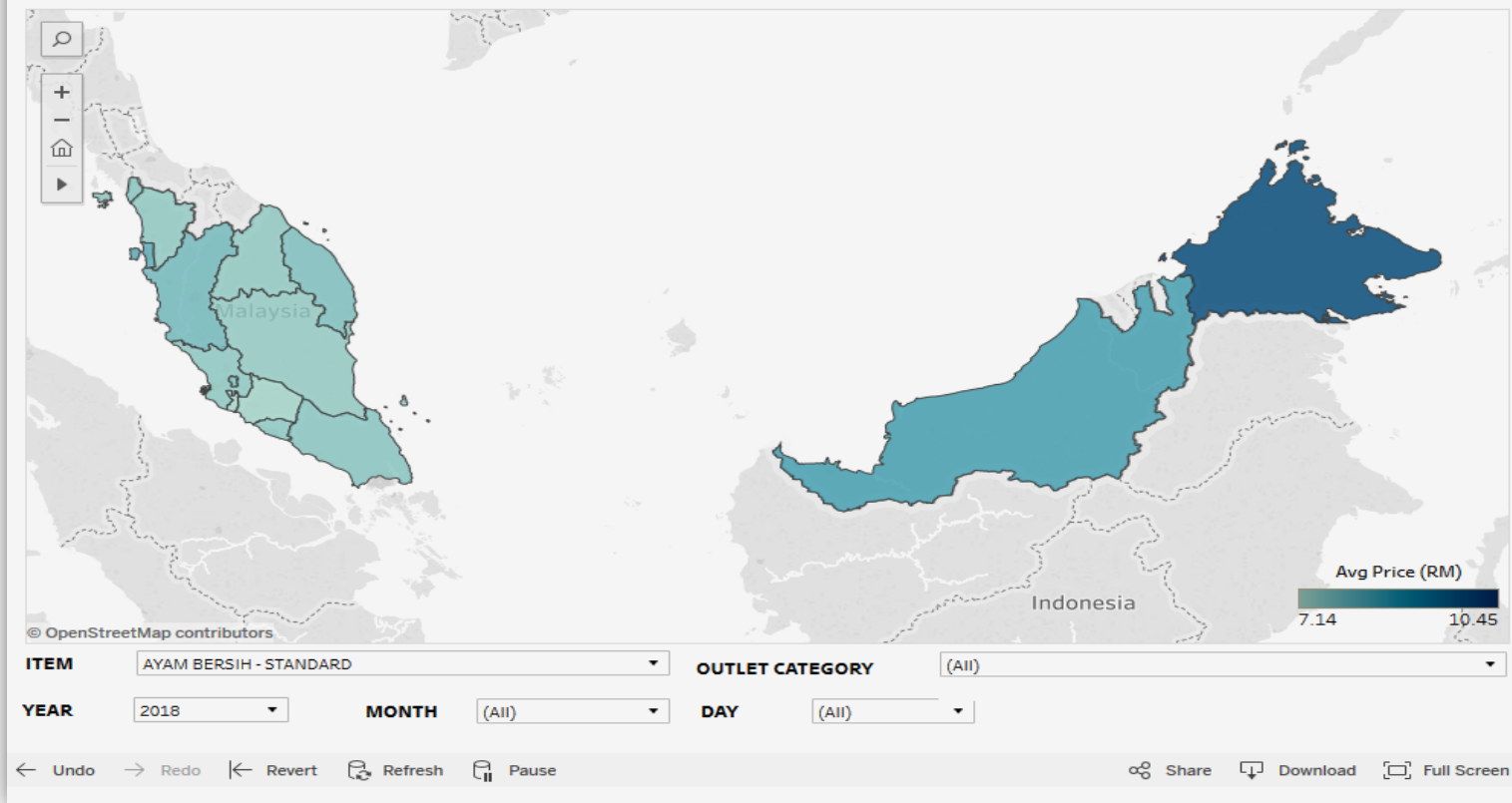
6 - Digits:

Item Code:

Outlet Category:

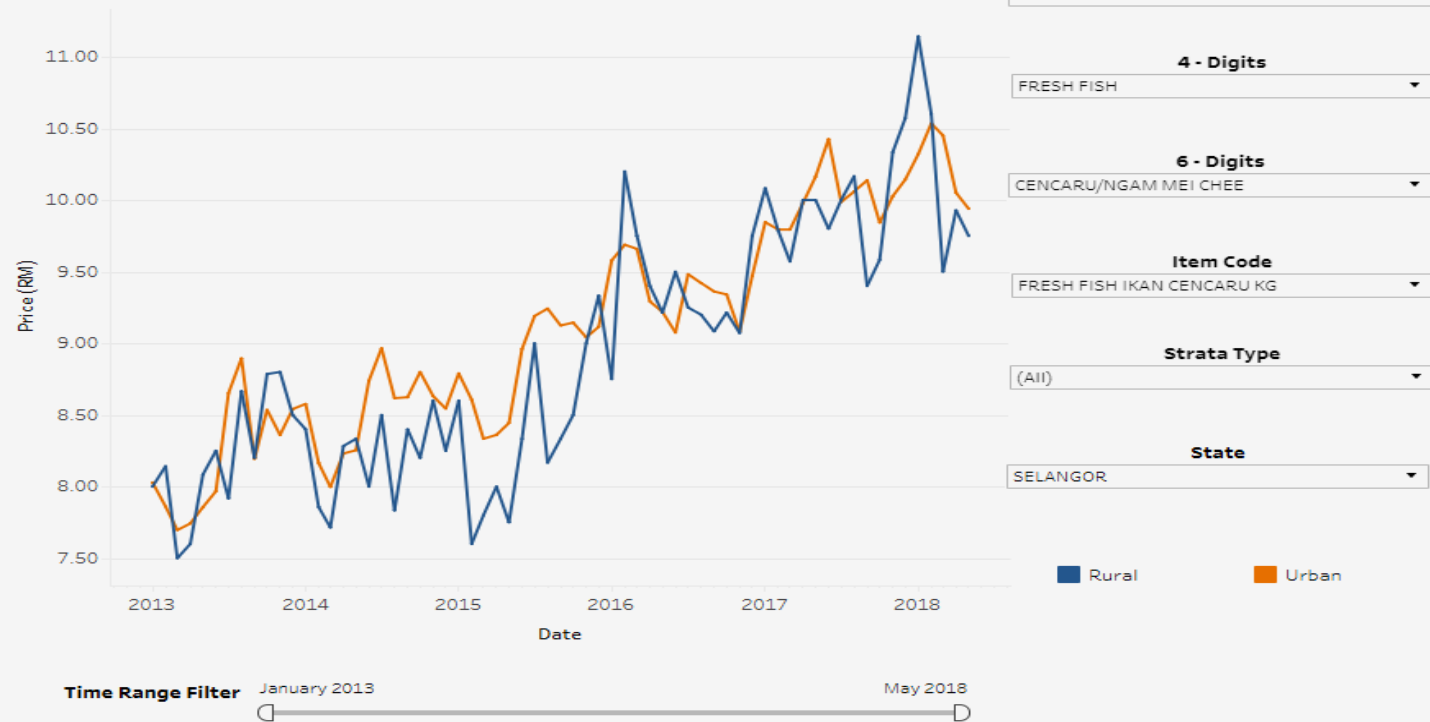
## Average Price by State

AYAM BERSIH - STANDARD



## Trend of Average Price by Strata Category

### FRESH FISH IKAN CENCARU KG



# Public Maturity Assessment on Official Statistics (PMAOS)

Opinion or sentiment data is generated through social channels in the form of reviews, chats, shares, likes tweets, etc.



*Topic Analysis*



*Social Media Account Monitoring*



*Report*



*Article Upload*



*Analysis*



*Visualisation*



Unstructured  
Data

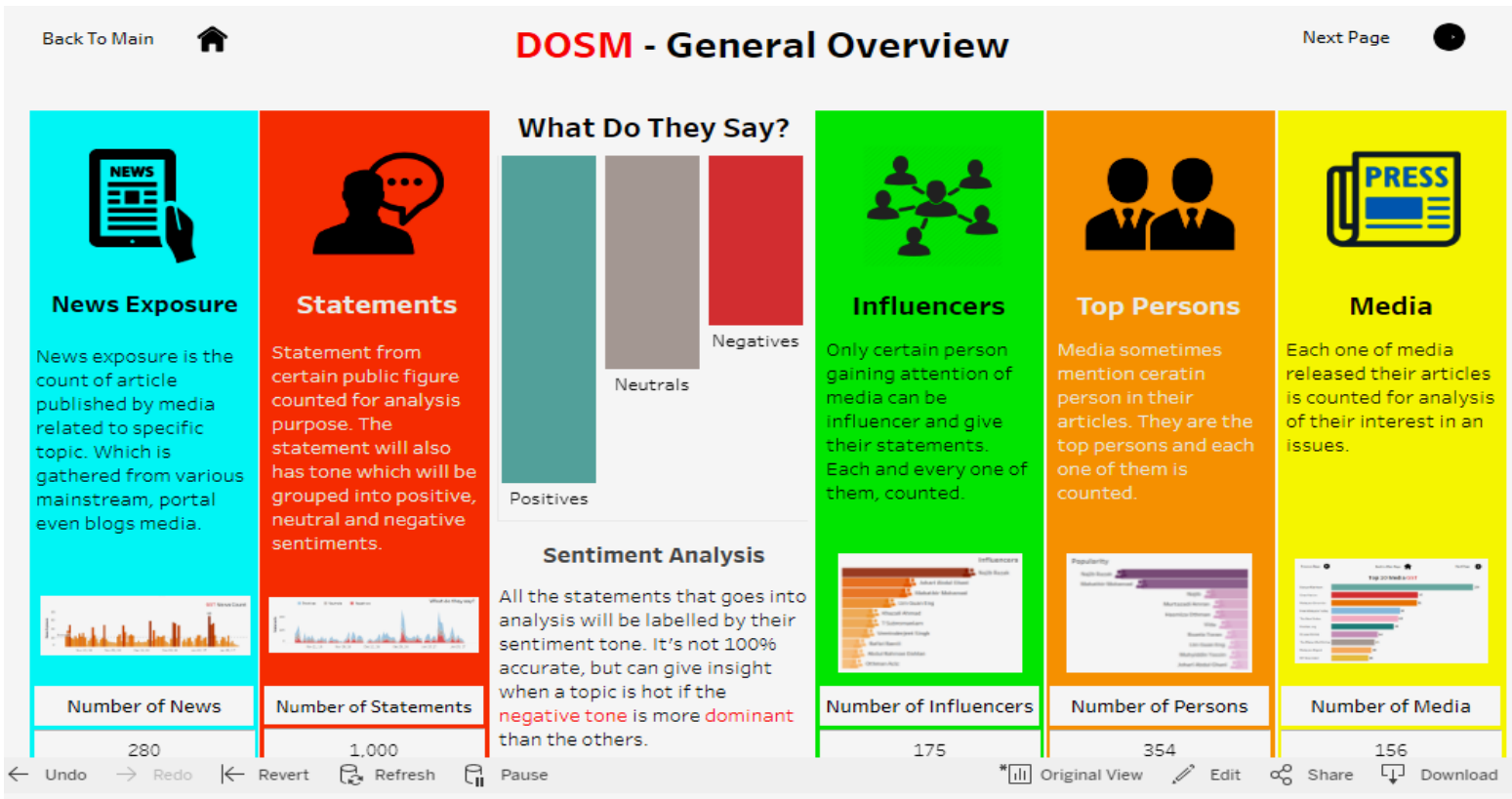


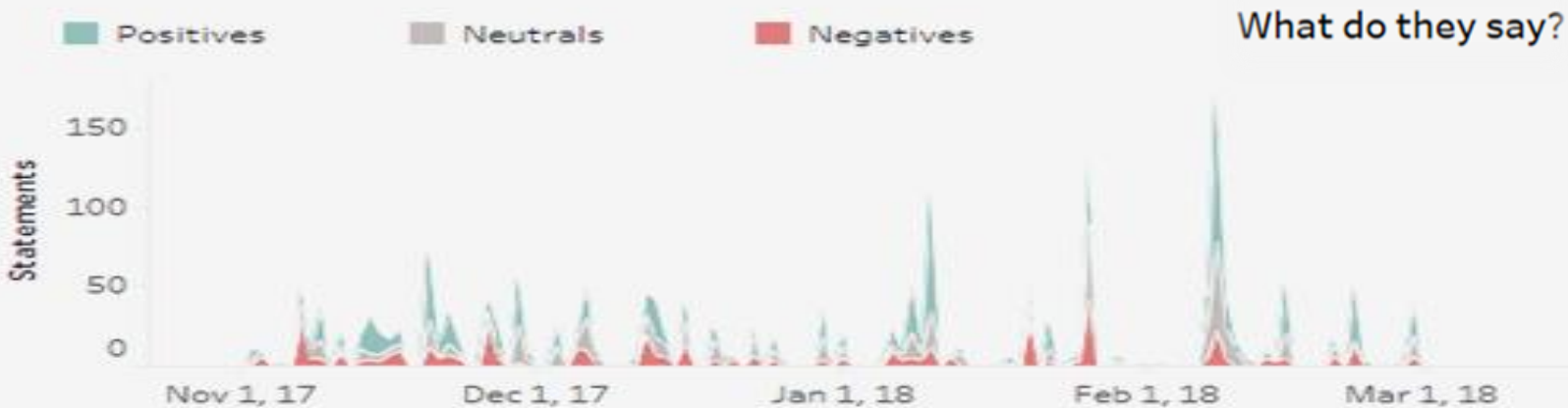
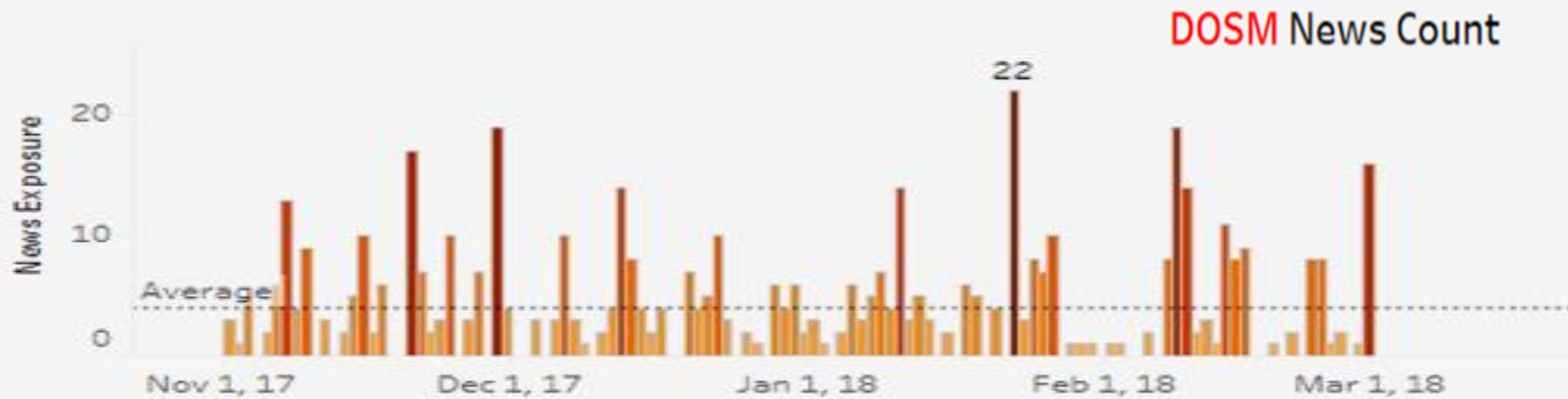
To measure degree of  
“happiness” of Malaysia  
community towards official  
statistics



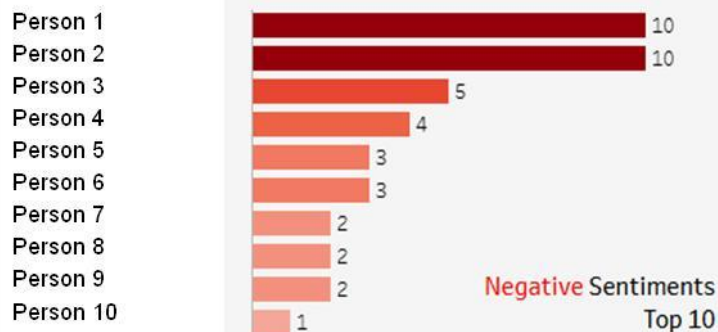
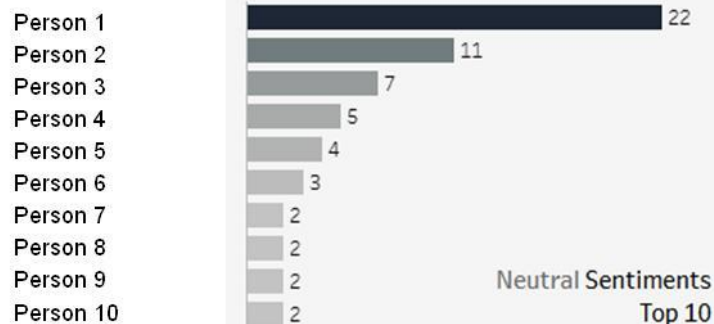
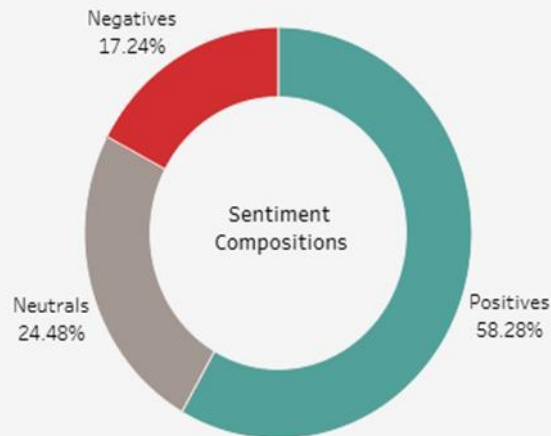
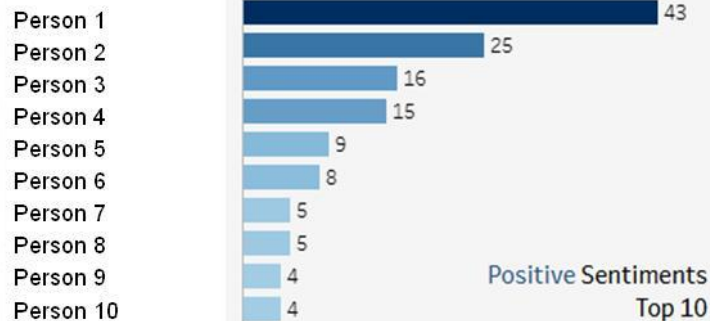
# Public Maturity Assessment on Official Statistics (PMAOS)

## - Visualization

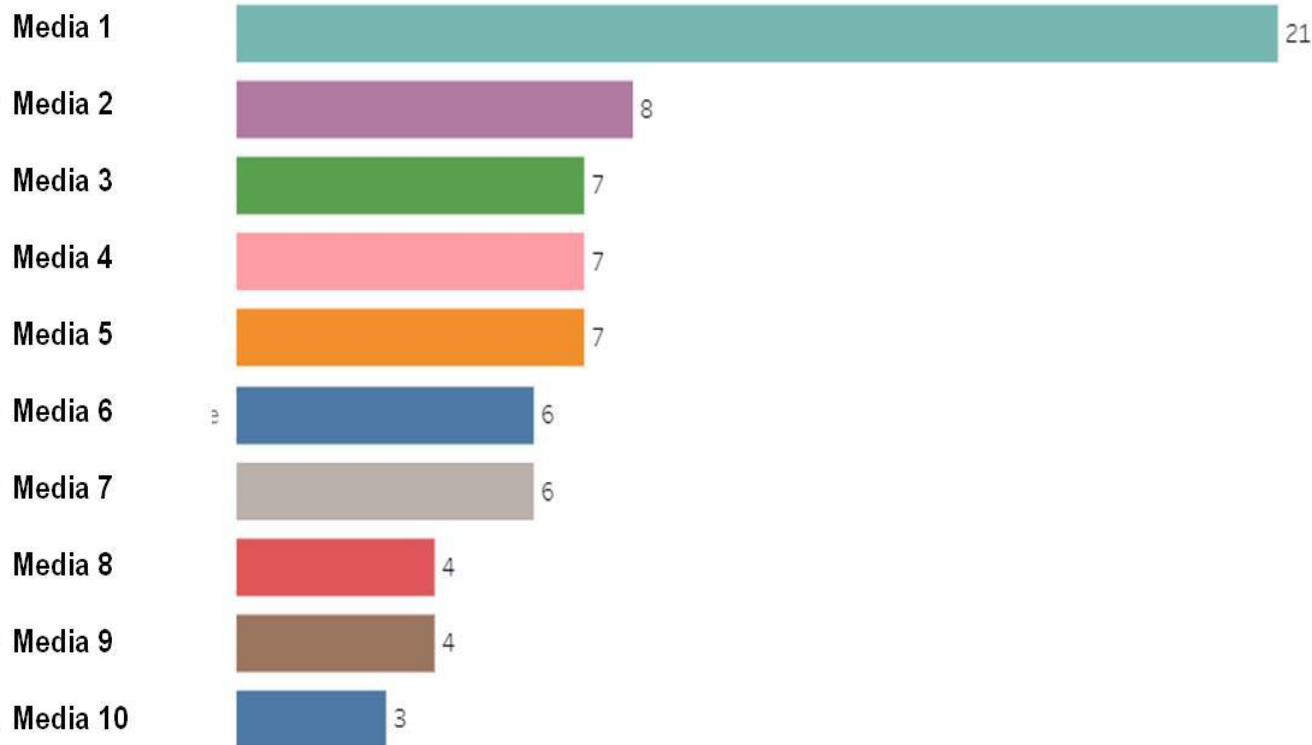




## Top 10 Influencers By Sentiment Tone



## Top 10 Media DOSM



# Conclusion



# Thank You