OVERVIEW ON BIG DATA
-STATSBDA-

14 JUNE 2022 (TUESDAY)
14:30 – 15:30 (MYT)
ONLINE

Ms. Mazliana Mustapa
Department of Statistics Malaysia
• How can we extract useful information from the data?
• How can we integrate our knowledge and experience in the industries with those data?
  • How certain we are the data are powerful enough to give us insight?
  • How can the data help us minimize operational cost?

Statistics can help us build new knowledge from the data where we can get from variety of sources. They have enormous effects on industrial innovations.
Living in the era of big data, our daily life is surrounded by statistics. Statistics provide useful information on what is happening around us in the past, current and future.
MAXIMIZE BENEFIT USING BIG DATA

Data, computational power and connectivity

Analytics and intelligence

Human machine interaction

Digital to physical conversion
Inception Meeting in Colombo, Sri Lanka
30 June - 2 July 2014
Organizer: Asian Development Bank (ADB)

Malaysia Statistical Business Register 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: UNSD

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

STATSBDA Project December 2016
Structured Data (TEC) + Unstructured Data (PI & PMAOS)
DOSM BIG DATA ANALYTICS ENVIRONMENT

Integrated big data analytics environment

Big data platform

Advanced analytics

Visualization

Structured & unstructured data analytics

Data management
High volume of administrative data from JKDM is integrated with MSBR to provide value added to an existing international trade statistics without having to run a new survey.

These statistics are able to identify the characteristics of companies or businesses involved global market.
• Descriptive
  ➢ Combination of Trade and Company Info (Trade Enterprise Characteristic)
  ➢ Able to deep dive to more granular level
  ➢ Benefit for organization to enhance productivity: Internal & Explicit 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
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

WHAT IS PRICE INTELLIGENCE (PI)?
**b) Descriptive**
Descriptive analysis, more emphasis on data exploration summaries such as: Mean, Median, Standard Deviation, Variants, Histogram / Skewness, etc.

**a) Trend**
Trend analysis is used to evaluate data patterns based on linear approaches. There are various techniques used in price analysis, such as ARIMA, Exponential Smoothing, Holt Winter, Linear Trend, Exponential and Level Aggregation.
Opinion or sentiment data is generated through social channels in the form of reviews, chats, shares, likes, tweets, etc.
• **Keyword** is the main important thing in developed topic

• Using Operator such as “( )” “AND” “OR” as additional rules.
DOSM: Manufacturing sales up 6.8pc to RM126.5b in August 2021

Uzir Mahidin said the growth in sales value for August 2021 was driven by the petroleum, chemical, rubber and plastic products subsector (31.4 per cent), especially in the manufacture of coke and refined petroleum products industries... Mohd Uzir noted that sales value for export-oriented industries, which accounted for 75.3 per cent of total
WHAT IS BizCode@STATS?

BizCode @ Stats is a DOSM robust platform for statistical codes and classifications developed on international classification basis to be used by public in delivering information simplicity.

This mobile applications should serves users with similar services to those accessed on PCs and it could be accessed even without internet connection.

This app can be downloaded on both Android and iOS mobile devices (Apple) through Gallery of Malaysian Governemnt Mobile Aplication (GAMMA).
Strategic partnership with SSM through online system integration between DOSM and SSM which is known as SSM Middleware. It is a web service where DOSM needs to send requests based on specified format to get latest businesses’ / companies’ information.
WELCOME TO STATSBDA PORTAL
Powerful Data, Valuable Result, Informative Decision

https://statsbda.dosm.gov.my/
DATA SCIENCE LAB

Equipped with 4 PCs with statistical software packages like SAS, SPSS, R, Stata
DOSM INITIATIVES

DOSM has taken the initiative by implementing innovations in data collection through non-official and non-conventional sources.

In line with the Big Data phenomenon and the Government's intention to drive ICT services, the Malaysian Government has announced implementation of Public Sector Big Data Analytics (DRSA)

DOSM has taken proactive action by developing the Big Data Project which is known as STATSBDA (from December 2016 until May 2018)

To complement and strengthen the development of DOSM Big Data initiative, the Analytics and Predictive Section in Methodology & Research Division has been established in Nov 2017. Human capital strengthening programs also have been implemented.

9 DOSM officers have been recognized as SAS Certified Data Scientist in 2018 and DOSM also encourages its officers to further studies in Data Science field.

This section consists of five (5) officers:
- One (1) E48;
- Two (2) E41;
- One (1) E32 and
- One (1) E29.

In May 2020, Big Data Analytics Core Team has been established by gathering officers who have certification in Data Science and as well as those who possess the related skills.
**Formed with a membership of 5 officers**

**ANALYTICS UNIT**

1 Assistant Director
1 Asst. Stats Officer

**PREDICTIVE UNIT**

1 Assistant Director
1 Asst. Stats Officer

**Formed with a membership of 27 officers**

**LEAD BY STATISTICIAN (E54)**

**DOSM : ECONOMIC SERVICE SCHEMES (E)**

7 SAS Certified Data Science
4 Masters in Data Science
10 Statistical Competencies Development Programme
1 Asst. Officer (STATSBDA)

**DOSM : INFORMATION SYSTEM SERVICE SCHEMES (F)**

2 SAS Certified Data Science
3 Asst. Officer (STATSBDA)

**NOV 2017 – MAY 2020**

Chief Assistant Director (E48)
Methodology and Research Division

**FROM MAY 2020**

ADVISOR
Chief Information Officer (CIO)
Business Analyst

Artificial Intelligence / Machine Learning Developer

Advanced Analytics Developer / Data Scientist

Java/JavaScript/NodeJs/PHP Developer

Data Engineer

Mobile Apps Developer

Python Developer

Visualization Developer

Big Data Engineer

System Engineer

10 TALENT
SOCIAL PROTECTION DATABASE (PDPS)
Ensure a more equitable distribution of social assistance

DATABASE OF INFORMAL SECTOR AND MICRO INDUSTRY (PDSIIM)
Identify affected groups of individuals in the Informal Sector and Micro Industries

LABOUR MARKET INFORMATION ANALYTIC PALTAFORM (LMIAP)
Integrate labor market information in Malaysia

MOBILE POSITIONING DATA (MPD)
Exploring the use of location data and user movement through telecommunication networks for the production of tourism, migration and population statistics in Malaysia

EARTH OBSERVATION (EO)
Estimating Malaysia's rubber plantation area and production productivity using Satellite Imagery and Machine Learning enganggar kawasan penanaman
Establishment of **National Big Data Analytics Center (NBDAC)**

- **Aims** to bring together professionals from multi-disciplinary groups from Ministries / Departments / Agencies. Those professionals specialize in data science area such as data engineering, statistical science, mathematics, visualization, machine learning, data mining, data modeling and subject matter experts.

- **Various services schemes and grades will be appointed** to cover the scope of function, complexity and workload.
THANK YOU