



Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC-Bank Indonesia Satellite Seminar on “*Big Data*” at the ISI Regional Statistics Conference 2017

Bali, Indonesia, 21 March 2017

Using online property advertisements data as a proxy for property market indicators¹

Kumala Kristiawardani and Irfan Sampe,
Bank Indonesia

¹ This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Using Online Property Advertisements Data as a Proxy for Property Market Indicators



Bank Indonesia:
Kumala Kristiawardani
Irfan Sampe

IFC – Bank Indonesia Satellite Seminar on Big Data
Bali, 21 March 2017

Topics Covered

- Background
- Data Sources
- Methodology
 - Data Acquisition
 - Data Issues
 - Data Preparation
 - Data Processing
- Results
- Conclusions

Background

- A boom and bust in residential property prices is perhaps the most widely discussed topic in recent financial crises
 - ❑ Residential property prices were fell in the 1990s, following the US recession in 1990-1991;
 - ❑ In Japan, residential property prices fell continuously as the economy collapsed in Japan around 1990;
 - ❑ In 2007, the housing market crash was the cause of the financial crisis in US.
- Bank Indonesia has an important task to not only to safeguard monetary stability, but also financial system stability
- Hence, monitoring residential property prices (with other asset prices) is crucial for Bank Indonesia to achieve its main task.

Background

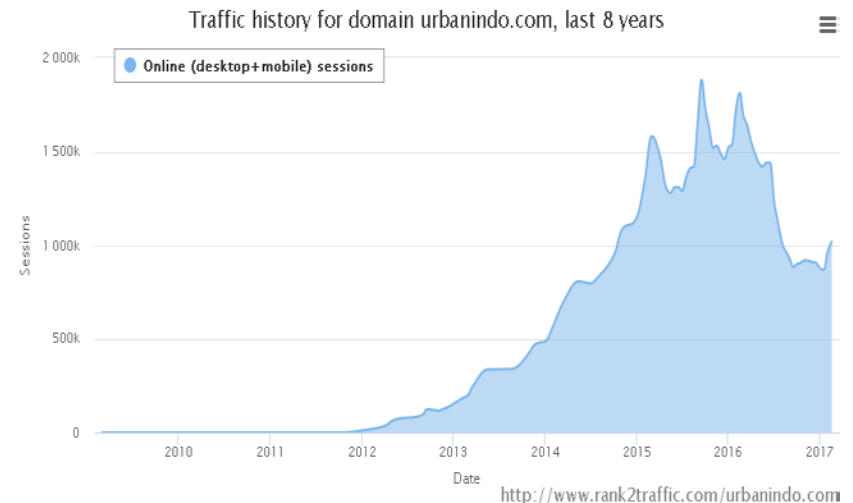
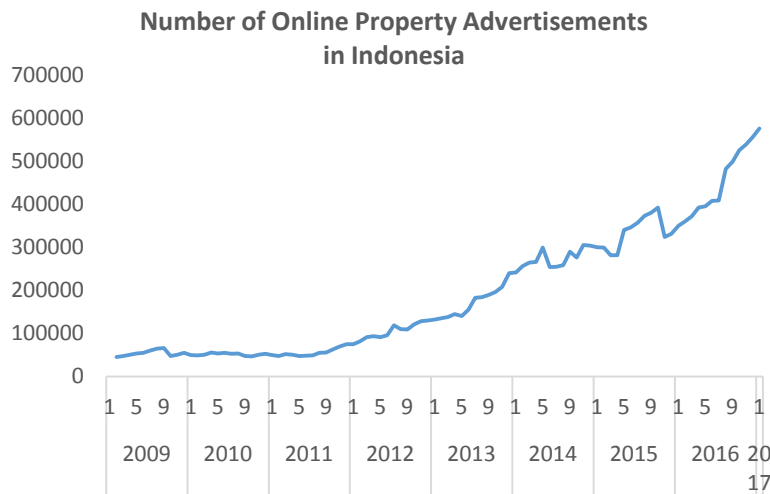
- Currently, the Bank Indonesia's primary data sources for monitoring Residential Property price are:
 - ❑ Residential Property Price Survey for primary house, conducted quarterly in **16 big cities**.
 - ❑ Residential Property Price Survey for secondary market, conducted quarterly only in **9 big cities**.

The data published at **six weeks** after the end of the survey period

- How do “big data” give the added value for Bank Indonesia in monitoring residential property market?

Background

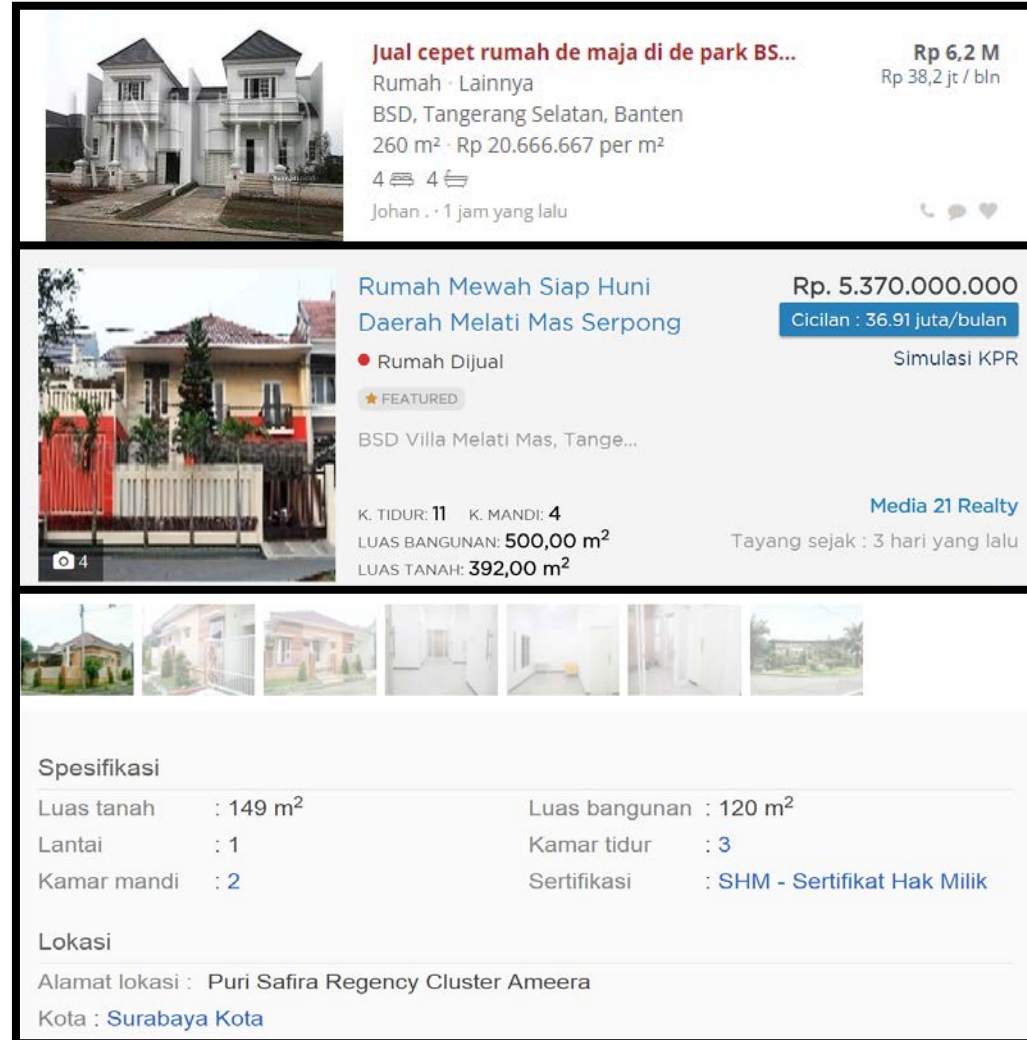
- The people's behaviour change in finding and selling the house (especially for secondary market)
 - ❑ Traditional: property agent, advertisement in newspaper
 - ❑ Now: search through internet (google, property online website, mobile apps)



Data Sources

3 biggest property online website in Indonesia (share 56 %)

- Title
- Status of property : sell/rent
- Type of property (house/apartment/villa/condotel/condominium)
- Advertising time : Starting & end date
- Property price
- Land & building size
- Number of bedroom & bathroom
- Address



Jual cepat rumah de maja di de park BS... **Rp 6,2 M**
Rp 38,2 jt / bln

Rumah · Lainnya
BSD, Tangerang Selatan, Banten
260 m² · Rp 20.666.667 per m²
4 🛏 4 🚿
Johan · 1 jam yang lalu

Rumah Mewah Siap Huni **Rp. 5.370.000.000**
Daerah Melati Mas Serpong **Cicilan : 36.91 juta/bulan**

• Rumah Dijual Simulasi KPR

★ FEATURED

BSD Villa Melati Mas, Tange...

K. TIDUR: 11 K. MANDI: 4
LUAS BANGUNAN: 500,00 m²
LUAS TANAH: 392,00 m²

Media 21 Realty
Tayang sejak : 3 hari yang lalu

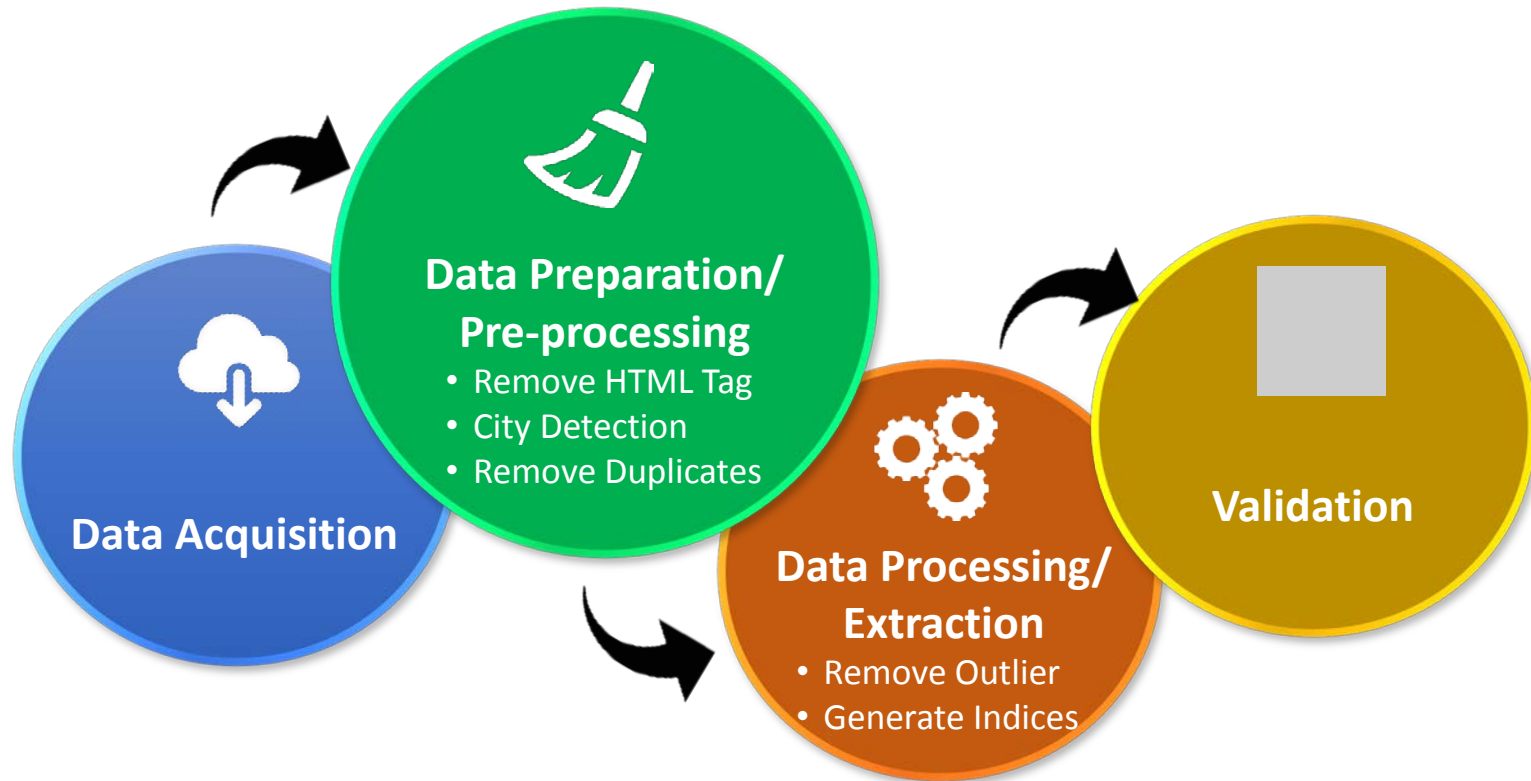
Spesifikasi

Luas tanah	: 149 m ²	Luas bangunan	: 120 m ²
Lantai	: 1	Kamar tidur	: 3
Kamar mandi	: 2	Sertifikasi	: SHM - Sertifikat Hak Milik

Lokasi

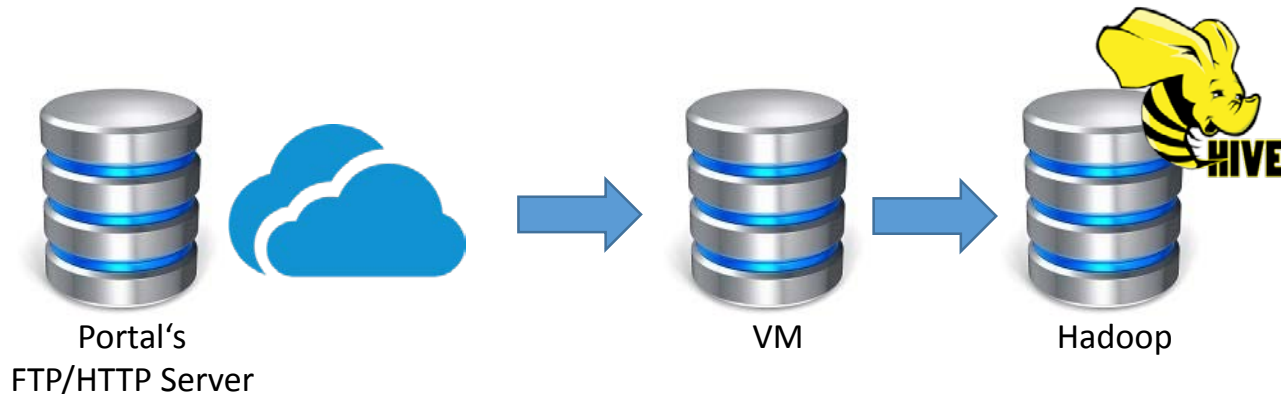
Alamat lokasi : Puri Safira Regency Cluster Ameera
Kota : Surabaya Kota

Methodology



Data Acquisition

- Property portal shared the data using FTPS/HTTPS. The files are password protected
- Available in the 1st week every month
- Loaded into Hadoop
- \approx 2.2 million ads/month



Data Issues

- Human error in data entry, i.e:
 - Price = Rp. 0, Price = Rp. 16 trillion (\$ 1.2 billion) on small size property
 - Land Size = 0 sqm, Land Size = 1 sqm
 - Typo on city/regency name
- Not standardized address data (freetext field)
 - District/sub district, e.g: Bogor, Bgr
 - Street name without district name, e.g: Jl. Kesadaran Sukmajaya
- Duplicate ads that are caused by:
 - One property can be advertised by more than one seller in a single portal
 - One property can be advertised by one seller across portals
 - Ads re-post after expiration date

Data Preparation/Pre-Processing

City Detection

- Map district/sub-district into city/regency using BPS's* Master Kabupaten,
- Map address into city/regency using Google Maps Geocoding API

Kampung Rambutan → Jakarta Selatan
Jl. Kesadaran Sukmajaya → Depok

*Indonesian Central Bureau of Statistics (BPS)

Remove Duplicates

Advertisements are identical if:

- The same attributes values on city/regency, land size, building size, number of bathrooms, and number of bedroom
- Price difference $\leq 5\%$
- String similarity score for address and ads title ≥ 0.8 (scale of 1) → using Levenshtein Distance

Data Processing/Extraction

Remove Outlier

- Removing properties with:
 - Land size and building size is empty (NULL)
 - Land size < 21 sqm and > 10.000 sqm
 - building size < 21 sqm > 10.000 sqm
- Applying price/sqm threshold
- Applying Median Absolute Deviation (MAD)

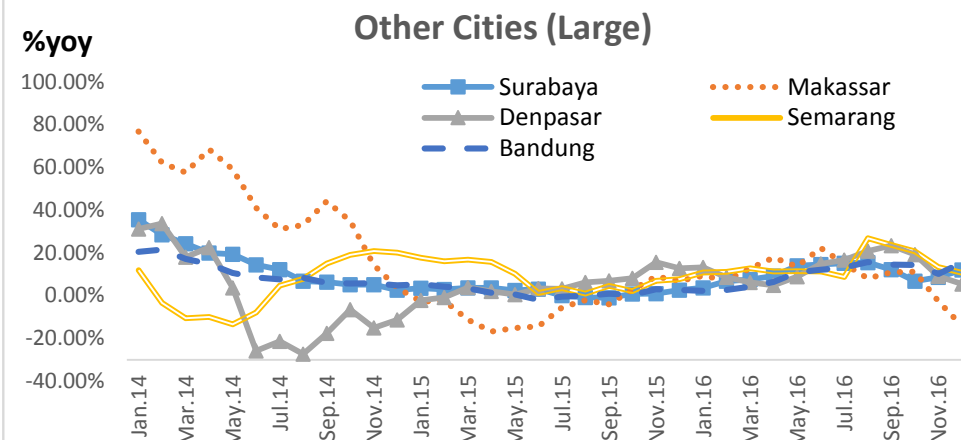
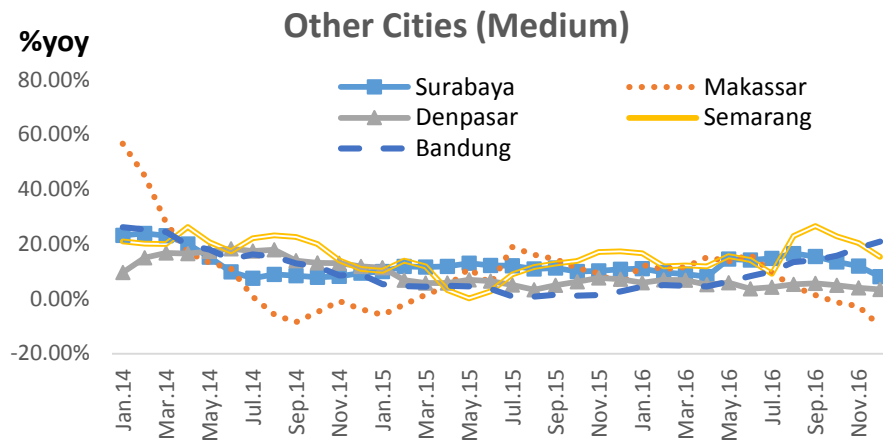
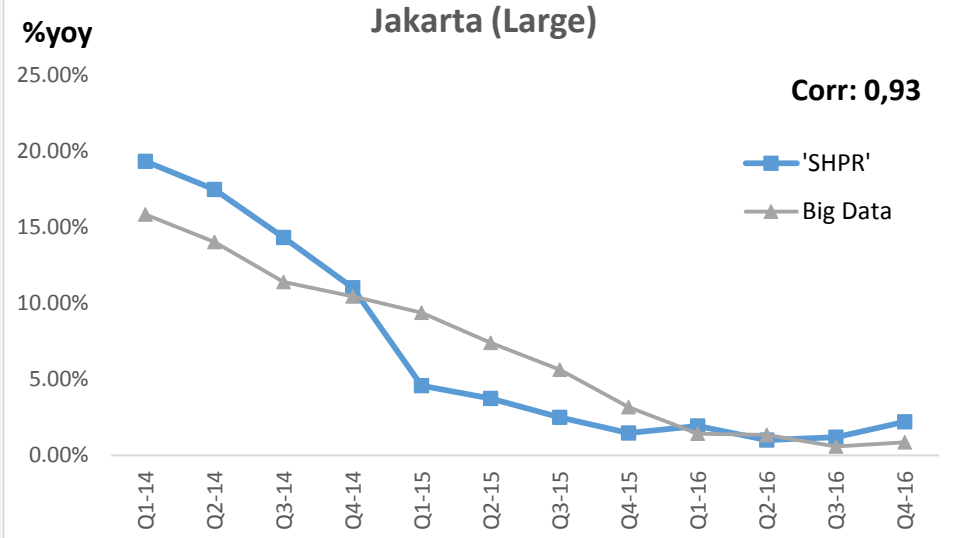
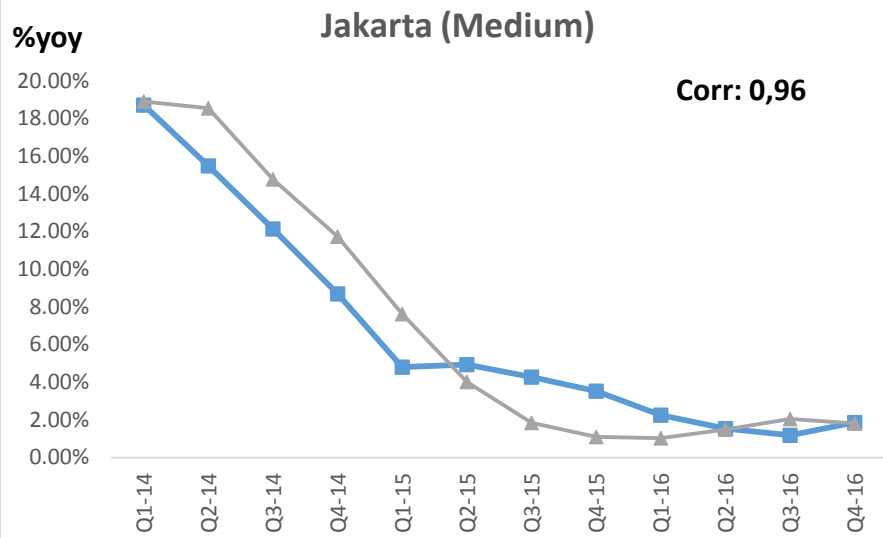
Generate Indices

- Landed house only
- Properties are divided into 3 types (based on building size):
 - Small: < 80 sqm
 - Medium: 80 – 150 sqm
 - Large: > 150 sqm
- Indices are generated per city/regency
 - Price (AVG: average of property price)
 - Supply (COUNT: number of active property ads)

Results Obtained

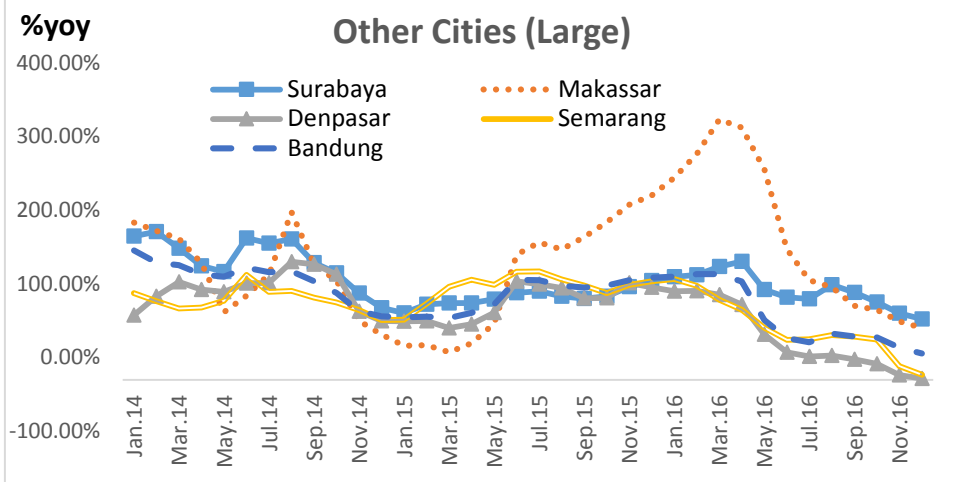
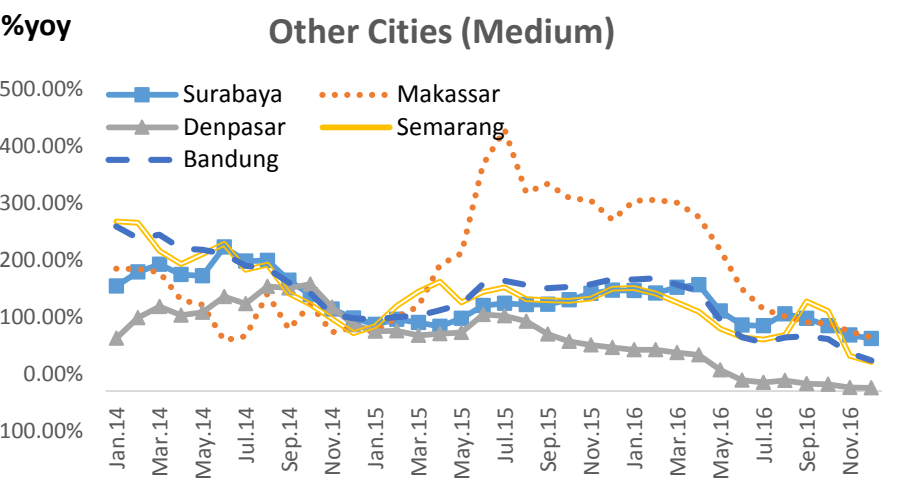
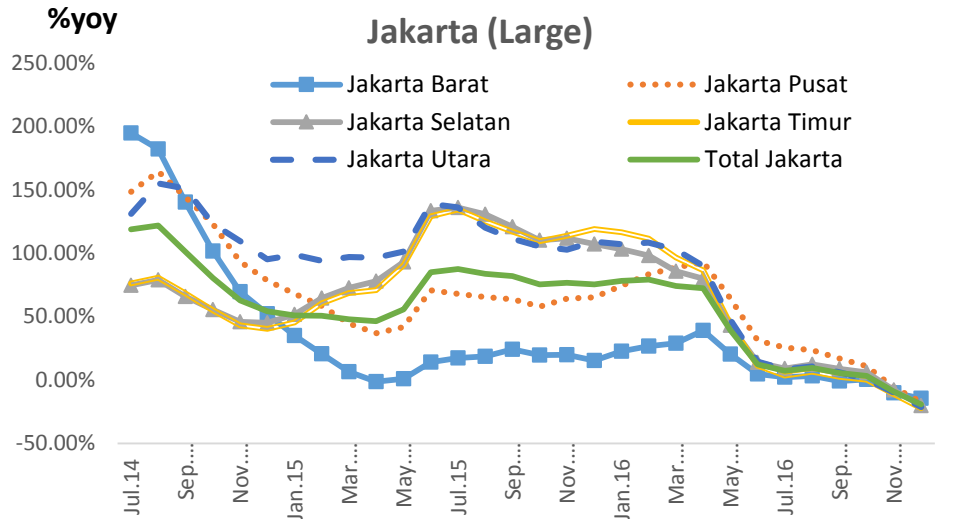
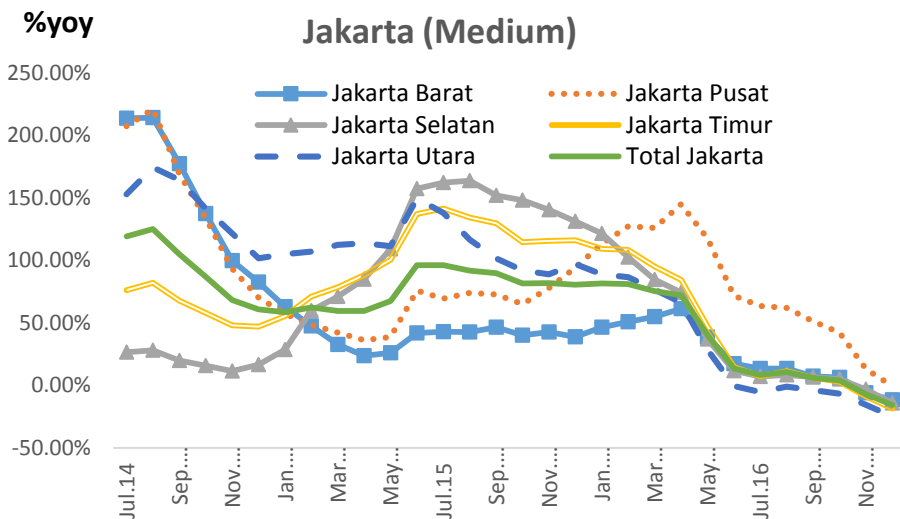
Price Index

Base period: Q2 2015



Results Obtained

Supply Index



Conclusions

- Online property ads data are potentially used as a proxy of price and supply indicators in Indonesia's residential property market.
- However, there are some limitations in conducting the research due to data availability and quality, i.e:
 - Short periode of data (only available since 2013)
 - The sold status is rarely updated by the seller

Terima Kasih

(Thank you)