Business Intelligence and Analytics applied to Public Housing

Doctoral Consortium @ ADBIS 2019
September 8th, 2019 in Bled, Slovenia

E. Scholly\textsuperscript{1,2}, C. Favre\textsuperscript{1}, E. Ferey\textsuperscript{2}, S. Loudcher\textsuperscript{1}

\textsuperscript{1}University of Lyon, Lyon 2, ERIC EA 3083
\textsuperscript{2}BIAL-X
Introduction
Context

A business issue
A business issue

- Public Housing: dwellings, occupants, overdue, patrimony, ...
A business issue

- **Public Housing**: dwellings, occupants, overdue, patrimony, ...
A business issue

- **Public Housing**: dwellings, occupants, overdue, patrimony, ...

Three main thematics
A business issue

- **Public Housing**: dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence (BI)**: ETLs, data warehouses, OLAP, ...
A business issue

- **Public Housing**: dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence (BI)**: ETLs, data warehouses, OLAP, ...
- **Data Science (DS)**: knowledge extraction, Machine Learning, ...
A business issue

- **Public Housing** : dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence** (BI) : ETLs, data warehouses, OLAP, ...
- **Data Science** (DS) : knowledge extraction, Machine Learning, ...
- **Big Data** : Volume, Variety, Velocity, ...
A business issue

- **Public Housing**: dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence (BI)**: ETLs, data warehouses, OLAP, ...
- **Data Science (DS)**: knowledge extraction, Machine Learning, ...
- **Big Data**: Volume, Variety, Velocity, ...

How does all this blend?
A business issue

- **Public Housing**: dwellings, occupants, overdue, patrimony, ...

Three main thematics

- **Business Intelligence (BI)**: ETLs, data warehouses, OLAP, ...
- **Data Science (DS)**: knowledge extraction, Machine Learning, ...
- **Big Data**: Volume, Variety, Velocity, ...

→ How does all this blend?
What data?

Several data sources

1. Internal data
   - Landlord's data
   - Dwellings, occupants, overdue, ...
   - Mostly relational data
   - BI analyses, simple DS analyses

2. External data
   - Open data (+ social networks)
   - Environment
   - (possibly) Big Data
   - Advanced DS analyses
What data?

Several data sources

1. Internal data

• Landlord's data
• Dwellings, occupants, overdue, ...
• Mostly relational data
• BI analyses, simple DS analyses

2. External data
• Open data (+ social networks)
• Environment
• (possibly) Big Data
• Advanced DS analyses
What data?

Several data sources

1. **Internal data**
   - Landlord’s data
   - Dwellings, occupants, overdue, ...
   - Mostly relational data
   - BI analyses, simple DS analyses
What data?

Several data sources

1. **Internal data**
   - Landlord’s data
   - Dwellings, occupants, overdue, ...
   - Mostly relational data
   - BI analyses, simple DS analyses

2. **External data**
What data?

Several data sources

1. **Internal data**
   - Landlord’s data
   - Dwellings, occupants, overdue, ...
   - Mostly relational data
   - BI analyses, simple DS analyses

2. **External data**
   - Open data (+ social networks)
   - Environment
   - (possibly) Big Data
   - Advanced DS analyses
1. Introduction

2. Data storage and management

3. Attractiveness

4. First results and future outcomes
Data storage and management
Business Intelligence (BI)

Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]
### Business Intelligence (BI)
Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

### Business Analytics (BA)
The use of Data Science methods on a company’s data

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]
**Business Intelligence (BI)**
Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

**Business Analytics (BA)**
The use of Data Science methods on a company’s data

What about BI?

- BI&A

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]
Business Intelligence and Analytics

**Business Intelligence (BI)**

Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

**Business Analytics (BA)**

The use of Data Science methods on a company’s data

What about BI?

- BI&A
- BI & BA

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]
Business Intelligence and Analytics

Business Intelligence (BI)
Methods and tools for collecting, storing, organizing and analyzing data to support decision-making

Business Analytics (BA)
The use of Data Science methods on a company’s data

What about BI ?

- BI&A
- BI & BA
- BI → BA

[Chen et al., 2012, Larson and Chang, 2016, Mortenson et al., 2015, Baars and Ereth, 2016, Gröger, 2018]
Run BI and BA analyses...
Run BI and BA analyses...

- Separately
Run BI and BA analyses...

• Separately
• Together
Run BI and BA analyses...

- Separately
- Together
- (possibly) on Big Data
Run BI and BA analyses...

- Separately
- Together
- (possibly) on Big Data
Run BI and BA analyses...

- Separately
- Together
- (possibly) on Big Data

**Data Intelligence**

Perform analyses, simple or advanced, on all types of data
Run BI and BA analyses...

• Separately
• Together
• (possibly) on Big Data

Data Intelligence
Perform analyses, simple or advanced, on all types of data

→ How?
Data Lakes

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

[Miloslavskaya and Tolstoy, 2016]
Data Lakes

<table>
<thead>
<tr>
<th>Data Lake [Dixon, 2010]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.</td>
</tr>
</tbody>
</table>

Two main characteristics

- *Schema-on-read*
A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*
- *Data variety*

---

**Data Lake** [Dixon, 2010]

[Miloslavskaya and Tolstoy, 2016]
Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*
- Data variety
Data Lakes

Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

• Schema-on-read
• Data variety

→ Need for a metadata system

[Miloslavskaya and Tolstoy, 2016]
Data Lake [Dixon, 2010]

A data lake is a large repository of heterogeneous raw data, supplied by external data sources and from which various analyses can be performed.

Two main characteristics

- *Schema-on-read*
- Data variety

→ Need for a metadata system

Big research field

[Miloslavskaya and Tolstoy, 2016]
Attractiveness
Data Intelligence in practice

Data sources
Extract

Data Lake
Load

Storing area

Metadata

DWH

(Re)Load

BI analyses
OLAP, reporting, ...

BA analyses
Statistics, clustering, forecasting, indicators, ...

(Meta)data exploration

Analysis
Transform
Defining attractiveness

Attractiveness of what?

1. Dwelling
2. Residency
3. Neighborhood
Defining attractiveness

Attractiveness of what?

1. Dwelling
Defining attractiveness

Attractiveness of what?

1. Dwelling
2. Residency
Defining attractiveness

Attractiveness of what?

1. Dwelling
2. Residency
3. Neighborhood
Attractiveness of what?

1. Dwelling (internal)
2. Residency
3. Neighborhood
Defining attractiveness

Attractiveness of what?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood
Defining attractiveness

Attractiveness of what?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)
Defining attractiveness

Attractiveness of what?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

→ Strategic Patrimony Plan
Defining attractiveness

Attractiveness of what?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

→ Strategic Patrimony Plan

Advanced indicators
Defining attractiveness

Attractiveness of what?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

→ Strategic Patrimony Plan

Advanced indicators

• Machine Learning algorithms
Defining attractiveness

Attractiveness of what?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

→ Strategic Patrimony Plan

Advanced indicators

- Machine Learning algorithms
- Back-feeding the lake
Defining attractiveness

Attractiveness of what?

1. Dwelling (internal)
2. Residency (internal - external)
3. Neighborhood (external)

→ Strategic Patrimony Plan

Advanced indicators

• Machine Learning algorithms
• Back-feeding the lake
• Enrich BI analyses
First results and future outcomes
Work done with P. N. Sawadogo [Sawadogo et al., 2019, Scholly et al., 2019]
First contribution

Work done with P. N. Sawadogo [Sawadogo et al., 2019, Scholly et al., 2019]

• Our definition of a Data Lake
• Key features for metadata systems
• Metadata typology in three categories
• MEtadata model for DAta Lakes (MEDAL)

Presented at 4 PM in this room!
What’s next?

Work in progress

Implementation(s) of MEDAL
Retrieve all data
Development of a complete data lake
Tests and comparisons
What’s next?

Work in progress

• Implementation(s) of MEDAL
What’s next?

Work in progress

- Implementation(s) of MEDAL
- Retrieve all data
What’s next?

Work in progress

- Implementation(s) of MEDAL
- Retrieve all data
- Development of a complete data lake
What’s next?

Work in progress

- Implementation(s) of MEDAL
- Retrieve all data
- Development of a complete data lake
- Tests and comparisons
Thank you for your attention!

Questions?


