The Internet of Things

Aaron Hart
Dr. Rosaria Silipo
Phil Winters
At the 2012 KNIME UGM……

Dr. Killian Thiel
Dr. Tobias Kötter
White Papers and complete workflows available on the KNIME Public Server!

Text Mining for Sentiment

Network Mining for Relevance

Drill Down on special cases

Analytics for Prediction
Last year....

Time Series + Machine Learning + Big Data

Dr. Rosaria Silipo

Manufacturing
Chemical
Life Science
Transportation
Utilities
Automotive
Cyber Security

The Irish Energy Trials
White Paper and Complete Workflows Available!

- Telemetry Data
- Time Series Analysis with clustering
- Measurable / Applied to the Business
- SENSIBLE usages of Big Data
The Internet of Things

Illustration by CRISTINA BYVIK

Use Public Data Please....
Washington DC
Sensors!
What Is Capital Bikeshare?

Capital Bikeshare puts over 2500 bicycles at your fingertips. You can choose any of the over 300 stations across Washington, D.C., Arlington and Alexandria, VA and Montgomery County, MD and return it to any station near your destination. Check out a bike for your trip to work, Metro, run errands, go shopping, or visit friends and family. Join Capital Bikeshare for a day, 3 days, a month, a year or try our new daily key option, and have access to our fleet of bikes 24 hours a day, 365 days a year. The first 30 minutes of each trip are free. Each additional 30 minutes incurs an additional fee.
Street Maps

Topology / Elevations

Weather

Holiday Schedules

<table>
<thead>
<tr>
<th>HOLIDAY</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Year's Day</td>
<td>Tuesday, January 1</td>
<td>Wednesday, January 1</td>
<td>Thursday, January 1</td>
</tr>
<tr>
<td>Martin Luther King Jr.</td>
<td>Monday, January 20</td>
<td>Monday, January 20</td>
<td>Monday, January 19</td>
</tr>
<tr>
<td>Memorial Day</td>
<td>Monday, May 27</td>
<td>Monday, May 26</td>
<td>Monday, May 25</td>
</tr>
<tr>
<td>Independence Day</td>
<td>Thursday, July 4</td>
<td>Friday, July 4</td>
<td>Friday, July 3</td>
</tr>
<tr>
<td>Labor Day</td>
<td>Monday, September 2</td>
<td>Monday, September 1</td>
<td>Monday, September 7</td>
</tr>
<tr>
<td>Thanksgiving</td>
<td>Thursday, November 28 Friday, November 29</td>
<td>Thursday, November 27 Friday, November 28</td>
<td>Thursday, November 26 Friday, November 27</td>
</tr>
<tr>
<td>Christmas</td>
<td>Tuesday, December 24 Wednesday, December 25</td>
<td>Wednesday, December 24 Thursday, December 25</td>
<td>Thursday, December 24 Friday, December 28</td>
</tr>
<tr>
<td>New Year's Eve</td>
<td>Tuesday, December 31</td>
<td>Wednesday, December 31</td>
<td>Thursday, December 31</td>
</tr>
</tbody>
</table>
KNIME and the Internet of Things

Read Bike
Read all the Raw Bike Transaction Records

Enrich
Enrich with new data

Expand
Expand from bikes to stations and paths

Investigate
various Investigations

Predict
Create Prediction Process

Drill In
Learn from Segments and Routes
Reading the Sensor Data

Output data - 3:224 - Column Resorter

<table>
<thead>
<tr>
<th></th>
<th>Bike#</th>
<th>Start date/time</th>
<th>Start station</th>
<th>End station</th>
<th>Date_1</th>
<th>Date_2</th>
<th>End date</th>
<th>End date_2</th>
<th>End date_3</th>
<th>Member type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>W00749</td>
<td>21 Mar. 2011  03:58:00</td>
<td>14th &amp; Harvard St NW</td>
<td>16th &amp; Harvard St NW</td>
<td>3110</td>
<td>01 Apr. 2011 00:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>W01048</td>
<td>21 Mar. 2011  03:20:00</td>
<td>39th &amp; L St NW</td>
<td>7th &amp; Water St SW / SW Waterfront</td>
<td>3122</td>
<td>01 Apr. 2011 00:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>W00340</td>
<td>21 Mar. 2011  03:42:00</td>
<td>Lincoln Park / 15th &amp; East Capitol St NE</td>
<td>13th &amp; H St NE</td>
<td>31619</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>W00862</td>
<td>21 Mar. 2011  03:45:00</td>
<td>Pennsylvania Ave &amp; Dupont Circle NW</td>
<td>15th &amp; P St NW</td>
<td>31102</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>W00449</td>
<td>21 Mar. 2011  03:22:00</td>
<td>Park Rd &amp; Holmead Place NW</td>
<td>15th &amp; M St NW</td>
<td>31619</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>W00583</td>
<td>21 Mar. 2011  03:34:00</td>
<td>4th St &amp; Massachusetts Ave NW</td>
<td>3104</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>W00318</td>
<td>21 Mar. 2011  03:31:00</td>
<td>37th &amp; O St NW / Georgetown University</td>
<td>21st &amp; M St NW</td>
<td>3122</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>W00577</td>
<td>21 Mar. 2011  03:30:00</td>
<td>14th &amp; Harvard St NW</td>
<td>31105</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>W00416</td>
<td>21 Mar. 2011  03:29:00</td>
<td>Logan &amp; Mt Pleasant NW</td>
<td>7th &amp; T St NW</td>
<td>31107</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>W00903</td>
<td>21 Mar. 2011  03:26:00</td>
<td>13th St &amp; New York Ave NW</td>
<td>Georgia Ave and Fairmont St NW</td>
<td>3122</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>W00682</td>
<td>21 Mar. 2011  03:25:00</td>
<td>New York Ave &amp; 15th St NW</td>
<td>21st &amp; M St NW</td>
<td>3122</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>W00389</td>
<td>21 Mar. 2011  03:21:00</td>
<td>Calvert St &amp; Woodley Place NW</td>
<td>14th St &amp; Spring Rd NW</td>
<td>3101</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>W00191</td>
<td>21 Mar. 2011  03:20:00</td>
<td>14th &amp; V St NW</td>
<td>3101</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>W00418</td>
<td>21 Mar. 2011  03:18:00</td>
<td>17th &amp; K St NW [formerly 17th &amp; L St NW]</td>
<td>15th &amp; P St NW</td>
<td>3121</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>W00627</td>
<td>21 Mar. 2011  03:17:00</td>
<td>21st &amp; I St NW</td>
<td>3105</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>W00492</td>
<td>21 Mar. 2011  03:16:00</td>
<td>14th &amp; M St NW</td>
<td>3122</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>W00343</td>
<td>21 Mar. 2011  03:15:00</td>
<td>11th &amp; K Street NW</td>
<td>14th &amp; Spring Rd NW</td>
<td>3103</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>W00798</td>
<td>21 Mar. 2011  03:13:00</td>
<td>14th &amp; Rhode Island Ave NW</td>
<td>3104</td>
<td>31 Mar. 2011 23:00:00</td>
<td>Registered</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reading Sensor data is always messy...... but KNIME is good at it!
Enrich

XLS Reader → Total Station Capacity → Elevation → Add Elevation → Math Formula → distance → Time Difference String Manipulation → Add Weather

stabon names, addresses, locations, lat/long

Enrich

Use REST GET Resources to retrieve elevation

[Diagram of KNIME workflow with nodes and connections indicating string manipulation, joiner, and column filter, and a table showing appended data with columns for Row ID, new column, location coordinates, and elevation.]
Station and Bike Facts

Over 3 years

307 Stations
2963 Bikes
19.4% Casual Bikers
5.9m Bike Moves
The 30 minute rule...

What Is Capital Bikeshare?
Capital Bikeshare puts over 2500 bicycles at your fingertips. You can choose any of the over 300 stations across Washington, D.C., Arlington and Alexandria, VA and Montgomery County, MD and return it to any station near your destination. Check out a bike for your trip to work, Metro, run errands, go shopping, or visit friends and family. Join Capital Bikeshare for a day, 3 days, a month, a year or try our new daily key option, and have access to our fleet of bikes 24 hours a day, 365 days a year. The first 30 minutes of each trip are free. Each additional 30 minutes incurs an additional fee.

<table>
<thead>
<tr>
<th></th>
<th>Under 30</th>
<th>30 and Over</th>
</tr>
</thead>
<tbody>
<tr>
<td>Casual</td>
<td>63%</td>
<td>37%</td>
</tr>
<tr>
<td>Subscriber</td>
<td>98%</td>
<td>2%</td>
</tr>
<tr>
<td>Overall</td>
<td>91%</td>
<td>9%</td>
</tr>
</tbody>
</table>
Most Popular Segment:

Jefferson Memorial to the Lincoln Memorial

89% Casual Users!
The Business Challenge:
Even MORE of a Business Challenge

Any Station without bikes for 1 hour:
$XXXX Per Violation

Any Station with no free slots for 1 hour:
$XXXX Per Violation
The Business Challenge:  12\textsuperscript{th} and Bell St.   -587 Bikes
The Business Challenge: 12th Street and Bell  -587 Bikes
“Helpful” Bike Enthusiasts

Capital Bikeshare Tracker

Current Status & Outages  |  Status History  |  Outage History

Bikes in use (approx.): 90

Presently Empty and Full: 3

Empty and Full Stations: 3

<table>
<thead>
<tr>
<th>Terminal</th>
<th>Station Name</th>
<th>Status</th>
<th>Start</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>31107</td>
<td>Lamont &amp; Mt Pleasant NW</td>
<td>empty</td>
<td>16:16:52</td>
<td>50</td>
</tr>
<tr>
<td>31229</td>
<td>New Hampshire Ave &amp; T St NW</td>
<td>empty</td>
<td>16:21:56</td>
<td>45</td>
</tr>
<tr>
<td>31505</td>
<td>Eckington Pl &amp; Q St NE</td>
<td>empty</td>
<td>16:55:24</td>
<td>11</td>
</tr>
</tbody>
</table>

The data above is valid as of 2/9 17:04:23

http://www.cabitracker.com/status.php
Capital Bikeshare Response

http://bikeportland.org/2013/03/10/behind-the-scenes-of-capital-bikeshare-84006
Station Totals per hour
Creating the Video.....

OSM
Image Processing
Group Loop
Station Totals per hour
Lean Restocking Alert System

Goal: Implement an alert signal for when restocking (both adding and removing bikes) might be needed in a station.

Target Variable. Flag and Shifted contain the current human operated restocking information.

We want 1 hour warning! Lag(flag-1)
Lean Restocking Alert System

Input Features.

- Weather related features
- Number of registered and casual people showing up
- Station infos (name and max. number of docks)
- Calendar infos (working day, holiday, date)
- Past infos
- Count as the number of bikes added and removed at each hour
- Adjusted cumulative sum as the number of bikes available at the station at a given hour
- Bike ratio = adjusted cumulative sum/total docks available.
- Move to Percentage
Feature Elimination Loop

Two options:
1. Use all the input features (no thinking required, just a powerful machine)
2. Select the most useful input features via the “Feature Elimination” loop

At each step one input feature is removed - i.e. the input feature whose removal produces the smallest error increase.
"Backward Feature Elimination Filter" node
The input feature subset with the smallest error (81% accuracy) is forwarded to the final model training block:

- Hour of the day
- working day (Y/N)
- Current Bike Ratio
- Terminal (station code)

Bike Ratio past and weather infos do not seem to be relevant!
Total Bikers Number Prediction by the Hour

Registered (blue) vs. Casual (red) Bikers
Casual vs. Registered: Auto-Correlation Matrix (24h)
Time Series Prediction

(from last year’s KNIME UGM):

RMS error from a “Numeric Scorer” node
Time Series Prediction

(from last year’s KNIME UGM):
- Past lags and seasonality index had to be defined by hand.
- We tried a few values and we chose the best.
- Hypotheses for the bikers data: 24h seasonality index, 10-20 past hour values needed

This year’s UGM:

We select the best values programmatically using the Optimization Parameter Loop!
Parameter Optimization Loop

Looping over different values for lagging and seasonality index (Brute Force or Hillclimbing).
Collecting the RMS error at the end of the loop.
Selecting the parameter set with minimum RMS error.
### “Bikers Prediction” workflow

#### Parameter Optimization Loop

- **Read Data**
  - File Reader: `hours.csv` (registered casual date weather)

- **Pre-processing**
  - Dataframe sorting

- **Optimization**
  - Find best lag and seasonality index

- **Remove Seasonality**

- **TableRow To Variable**
  - Optimize lag and seasonality index in auto-regressive model

- **Time Series Prediction Metanodes**

### Seasonality, Past lags, RMS Error

<table>
<thead>
<tr>
<th>Seasonality index</th>
<th>Past lags</th>
<th>RMS Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>casual</td>
<td>1 hour</td>
<td>20</td>
</tr>
<tr>
<td>registered</td>
<td>24 hours</td>
<td>10</td>
</tr>
</tbody>
</table>
Conclusions....

- Best Input Feature Subset using the “Feature Elimination” Metanode
- Best Parameter Set for Time Series Prediction using the “Parameter Optimization” node
- Weather influence on bikers is overrated!
- Casual Bikers 24 hours Seasonality is also not so relevant!
Top Net Bike Change Stations

<table>
<thead>
<tr>
<th>Row ID</th>
<th>Name</th>
<th>Terminal</th>
<th>Count(stations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row98_Row49</td>
<td>Massachusetts Ave &amp; Dupont Circle NW</td>
<td>31200</td>
<td>366953</td>
</tr>
<tr>
<td>Row230_Row20</td>
<td>Columbus Circle / Union Station</td>
<td>31623</td>
<td>278418</td>
</tr>
<tr>
<td>Row99_Row22</td>
<td>15th &amp; P St NW</td>
<td>31201</td>
<td>255757</td>
</tr>
<tr>
<td>Row111_Row57</td>
<td>17th &amp; Corcoran St NW</td>
<td>31214</td>
<td>206565</td>
</tr>
<tr>
<td>Row77_Row14</td>
<td>14th &amp; V St NW</td>
<td>31101</td>
<td>187431</td>
</tr>
<tr>
<td>Row126_Row...</td>
<td>New Hampshire Ave &amp; T St NW</td>
<td>31229</td>
<td>187031</td>
</tr>
<tr>
<td>Row220_Row76</td>
<td>Eastern Market Metro / Pennsylvania Ave &amp; 7th St SE</td>
<td>31613</td>
<td>181194</td>
</tr>
<tr>
<td>Row80_Row17</td>
<td>Adams Mill &amp; Columbia Rd NW</td>
<td>31104</td>
<td>169684</td>
</tr>
<tr>
<td>Row101_Row24</td>
<td>14th &amp; Rhode Island Ave NW</td>
<td>31203</td>
<td>160891</td>
</tr>
<tr>
<td>Row125_Row90</td>
<td>8th &amp; H St NW</td>
<td>31228</td>
<td>159584</td>
</tr>
</tbody>
</table>
An aside on visualization
Finding route data
(Just google it)
REST Webservices with KREST

http://maps.googleapis.com/maps/api/directions/json?
&origin=lat,long&destination=lat,long&mode=bicycling

https://developers.google.com/maps/documentation/directions/
Plotting with ggplot2

require(ggplot2)

clean_theme = theme(text = element_blank(),
  panel.background = element_blank(),
  panel.border = element_blank(),
  axis.title.x = element_blank(),
  axis.title.y = element_text(size=14, face="italic", colour = "black"),
  axis.text = element_text(size=12, face="italic", colour = "black"),
  legend.text = element_blank(),
  legend.key = element_blank(),
  legend.position = "none",
  panel.grid = element_blank()
)

plot = ggplot() + geom_path(aes(x=knime.in$"Lat",
  y=knime.in$"Long",
  group = knime.in$"route",
  alpha = knime.in$"Count(a)",
  color = factor(knime.in$"route")),
  knime.in)

plot + clean_theme
And now some analysis...

Net Changes

Total Traffic Volume

16th and Harvard

Bike sources

Bike Sinks

Union Station
Stations with deficits and surpluses

16th and Harvard

Bike sources

Bike Sinks

Union Station
Network Analysis
Top 250 Routes

Urban/Students

Dupont Circle

Tourists

Suburban

Union Station
Union Station Subnet

- D St & Maryland Ave NE
- 13th & D St NE
- Lincoln Park / 13th & East Capitol St NE
- Columbus Circle / Union Station
- 8th & F St NE
- L'Enfant Plaza / 7th & C St SW
- 3rd & H St NE
- 11th & H St NE
- 6th & H St NE
- King St Metro
- 15th St & Massachusetts Ave SE
Dupont Circle Subnet

- Massachusetts Ave & Dupont Circle NW
- 14th & V St NW
- 16th & Harvard St NW
- Adams Mill & Columbia Rd NW
- Park Rd & Holmead Pl NW
- Columbia Rd & Belmont St NW
- Connecticut Ave & Newark St NW / Cleveland Park
- 36th & Calvert St NW / Glover Park
- Georgetown Harbor / 30th St NW
- C & O Canal & Wisconsin Ave NW
- 25th St & Pennsylvania Ave NW
1) 16th and Harvard

2) Dupont Circle

3) Georgetown Harbor
Possible Next Steps

Combine Prediction with Path !!!!

Enrich through analyzing images
Lessons Learned

Sensor Data is simple... but can still be messy
- get it in, get it clean!

Connect it!
- Common keys, Lat/long, time, etc.
- No common key? Network analysis + Imaging

Enrich / Expand
- KNIME Transformation
- External Sources; REST calls; Palladian, etc.

Explore
- Visualization (over time!) OSM, Imaging
- Network Analysis
- Correlation matrix!

Prediction
- Feature Elimination + “Classic Techniques”
- “Brute Force” Feature Elimination + “Classic Techniques”
The Internet of Things

Illustration by CRISTINA BYVIK

Use Public Data Please....
The Internet of Things: White paper, Workflows, Data will be available