Understanding Migration Flows with Mobile Data.
Challenges & Opportunities

Pedro A. de Alarcon, PhD
Head of Big Data for Social Good

AI for Social Good, Doha (Qatar). Feb 2019
"When I think about social good, I think about the commitments we have all made with the UN when it comes to the 17 Sustainable Development Goals for 2030. Forging a relationship between our big data work for social good is fundamental, especially as 80% of the 6 billion mobile phones in the world are in developing countries, which is where we can have the greatest impact."

—Jose Maria Álvarez Pallete, Chairman of Telefonica
Big Data for Social Good: Our commitment with the SDGs

Poverty & Development Metrics
- Improving SDG metrics with telco data (Spain, Central America)

Air Quality in Urban Areas
- Pollution forecasting (Spain, Brazil)
- Quantifying emissions with mobility data (Germany)

Natural Disaster Response
- Magic box initiative (Colombia)
- Measuring earthquake’s impact (Mexico)

Climate change in rural areas
- Internal Forced Displacements (Colombia)
- Smart Livestock Farming (Ecuador)

Epidemics & spread forecasting
- Analyzing the spread of Zika (Colombia, Brazil)
- Analyzing the spread of Measles (Brazil)
- Mobility impact analysis of swine flu (Mexico)
### SDGs and mobile phone metadata

#### The opportunities of mobile phone metadata

<table>
<thead>
<tr>
<th>Indicator</th>
<th>SDGs for which the indicator is relevant</th>
<th>Country in which the pilot project was carried out</th>
<th>The administrative level on which the groundtruthing was performed with 0 indicating the national level and higher numbers indicating smaller areas</th>
<th>Categorized adjusted $R^2$ indicating how much of the variance is captured by the spatial model</th>
<th>Categorized Root Mean Squared Error measures the standard deviation of the unexplained residuals</th>
<th>Possible disaggregation into sub groups like age and gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxy indicator predictable with mobile phone metadata</td>
<td>SDGs for which the indicator is relevant</td>
<td>Country in which the pilot project was carried out</td>
<td>The administrative level on which the groundtruthing was performed with 0 indicating the national level and higher numbers indicating smaller areas</td>
<td>Categorized adjusted $R^2$ indicating how much of the variance is captured by the spatial model</td>
<td>Categorized Root Mean Squared Error measures the standard deviation of the unexplained residuals</td>
<td>Possible disaggregation into sub groups like age and gender</td>
</tr>
<tr>
<td>Multidimensional Poverty Index</td>
<td>1.1, 1.2</td>
<td>Sudan, Philippines</td>
<td>2 &amp; 3</td>
<td>&gt;0.75</td>
<td>&lt;0.05</td>
<td>gender</td>
</tr>
<tr>
<td>Population Density</td>
<td>various</td>
<td>Senegal</td>
<td>pending</td>
<td>pending</td>
<td>aging groups &amp; gender</td>
<td></td>
</tr>
<tr>
<td>Literacy Rate</td>
<td>4.4, 4.6</td>
<td>Senegal</td>
<td>2, 3 &amp; 4</td>
<td>&gt;0.9</td>
<td>&lt;0.05</td>
<td>gender</td>
</tr>
<tr>
<td>Share of Women</td>
<td>10.2 &amp; various others</td>
<td>Senegal</td>
<td>3</td>
<td>&gt;0.9</td>
<td>&lt;0.05</td>
<td>n/a</td>
</tr>
<tr>
<td>Electricity Access Rate</td>
<td>7.1, 7.b, 9.1, 11.1</td>
<td>Senegal</td>
<td>3</td>
<td>&gt;0.85</td>
<td>&lt;0.15</td>
<td>n/a</td>
</tr>
<tr>
<td>Share of Minority Groups</td>
<td>10.2 &amp; various others</td>
<td>Senegal</td>
<td>3</td>
<td>&gt;0.9</td>
<td>&lt;0.05</td>
<td>n/a</td>
</tr>
<tr>
<td>Primary Completion Rate</td>
<td>4.1, 4.5, 4.6, 5.1</td>
<td>Senegal</td>
<td>3</td>
<td>&gt;0.9</td>
<td>&lt;0.05</td>
<td>n/a</td>
</tr>
<tr>
<td>Migration Patterns</td>
<td>10.7, 8.8</td>
<td>Turkey</td>
<td>Manually defined</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Commuting Patterns</td>
<td>10.7, 8.5, 8.8</td>
<td>Turkey</td>
<td>Manually defined</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Mobile phone “metadata” typically means Call Details Records.

- Per antenna KPIs (activity, pop. density...)
- Social Graph
- Mobility
When moving, call logs are like breadcrumbs

But...
XDRs mean more breadcrumbs (up to 5x)
Mobility Insights from Telco Data

- **Type**
  - Macro
  - Local
  - Behavioural

- **Source**
  - Active
  - Passive
  - WiFi
  - Small Cell
  - Apps
  - Websites

- **Events**
  - Billions of network events
  - 24 / 7 / 365
Mobile metadata + CRM + Spatial Semantics

- **Settles**
  - Demographic
  - Mode
- **Journeys**
  - Purpose
  - Frequency
  - Route
  - Segmentation
The Individual Trajectories Matrix

Livelihoods (approximation)

Mobility profile (monthly)

Livelihood Zone Target

Binarization

Profile 4
Profile 3
Profile 2
Profile 1

ALL RUTURED PEOPLE

MONTH
The Individual Trajectories Matrix

Adjust temporal resolution to days, weeks, months...

<table>
<thead>
<tr>
<th>PhoneID</th>
<th>Week1</th>
<th>Week2</th>
<th>…</th>
<th>WeekN</th>
</tr>
</thead>
<tbody>
<tr>
<td>sdqadw</td>
<td>Bogotá</td>
<td>Bogotá</td>
<td>..</td>
<td>Bogotá</td>
</tr>
<tr>
<td>fw4efef</td>
<td>Cali</td>
<td>Cali</td>
<td>…</td>
<td>Bogotá</td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td>…</td>
<td></td>
</tr>
</tbody>
</table>

Most common location: Adjust spatial resolution to district, municipality, Department...

Full customer base (>11M users)
Climate change is exacerbating extreme climate phenomena. Rural populations are specially impacted and, in many cases, forced to migrate to urban areas.
La Guajira (Colombia). Severe drought ... for years.

Analysing forced displacements due to climate variability in Colombia

- Telefónica leveraged open datasets from SIAC (Environmental Information System of Colombia) to identify regions with high vulnerability to drought conditions and limited ability to recover, focusing on departments La Guajira, Tolima and Huila.

- Mobility insights from anonymized mobile network data helped to identify clusters of users with specific long-term movement patterns, indicative of internal displacement.

- We combined these movement patterns with census data from Colombia’s National Administrative Department of Statistics to estimate the number of people displaced from the focal region during the period of the drought.
Visualization tool to navigate & discover insights
Challenges

- Limitations to create group profiling:
  - The new personal data privacy regulation (GDPR).
  - Most users are pre-paid subscribers (scarce information from CRM)
  - Addressing different market share rates within the same country
  - Filtering out people < 18yo
- Very few official and reliable **ground truth sources** to compare with.
Opportunities
Building robust IDP indicators from multidimensional data integration (from public and private sources)

Satellite Imagery & Computer Vision
Digital Globe

Social Media
Facebook and others

Telco Data
Telefonica
Thank you!

www.luca-d3.com

Follow us on Twitter: @pdealarcon @LUCA_D3