Big Data and Economic Metrics: Some corporate case studies

Jason Naicker
As data gets bigger
We are generating data faster than ever

6.7 billion store transactions per month

30+ million retail items measured

1.6 trillion online impressions per year

40,000 Google searches per second
It’s important to differentiate between Big Data, an “a lot of data”
Disparate data sets allow for better validation and creation of new insights

X 8 Billion
Chaos - Order - Chaos

The more assumptions we make, the less predictable the result.
The growth of data science as discipline leads to a need for more data.
Sentiment Analysis
What can we measure using sentiment analysis?

<table>
<thead>
<tr>
<th>BUZZ</th>
<th>SENTIMENT</th>
<th>TOPICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume of conversations</td>
<td>NSS™ (Net Sentiment Score)</td>
<td>Identification of conversation topics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EMOTIONS</th>
<th>KPIs</th>
<th>ENGAGEMENT</th>
<th>INFLUENCERS</th>
<th>IMAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 pairs of opposite emotions (proprietary AI model)</td>
<td>Purchase intent, recommendation, price, availability</td>
<td>Consumer activity generated by corporate activity</td>
<td>Based on buzz or posts, likes and shares</td>
<td>Logo and image theme detection</td>
</tr>
</tbody>
</table>

Twitter  Video  Facebook*  Instagram*  Reddit**  Weibo**  VK**  Blogs  Forums  News  Reviews

* For Facebook we harvest data based on page search (not keyword search), and are only able to harvest from public pages. For Instagram we are only able to harvest from up to two brand or group pages.
** An additional data cost may be incurred for these sources.
Example of premiership football teams
Possible applications:

Sentiment analysis could be used to provide a streaming understanding of consumer sentiment.

- Overall sentiment
- Sentiment specific to specific socio-political-economic issues or policies
- Future state: early crises warning
WFP: Predicting Food Consumption Score (FCS) using external data
The Food Consumption Score (FCS)

via Questionnaire

Dietary Diversity

Food Frequency

FCS
(Caloric Sufficiency)
#PROJECT8HACKFORHUNGER

Nielsen virtual hackathon to solve a key issue raised by the WFP:

Question

How can changes in food prices (or food security) be predicted on the basis of publicly available data?

Output

A predictive model taking in public information such as commodity prices, rainfall, or any other publicly available information.

Measure of success: A model that can predict the FCS with 80% accuracy
Commodity prices showed little predictive ability but the previous months score by food type showed some strong correlations to the next months FCS.
The final model showed an average accuracy of 97%

Addressing goal 2 of the UN Sustainable Development Goals
Home stores vs Unemployment
1. Establish Retail Trade Universe
   Define store types and gain information on shop numbers & turnover

2. Design representative sample
   How many of each store type do we need to represent the measured Retail Trade Universe?

3. Data collection
   Collect data both Scan/POS and Manual Audit data

4. Statistical expansion
   of sample to universe

5. Reported Outputs

THE RETAIL AUDIT PROCESS
Establishing the Retail Trade Universe

Rolling Establishment survey is conducted every six months

Starting with the smallest statistical representation of the country
(South Africa = Small Area Layers)

Field auditors walk every road and footpath to count the stores in the SAL as well as their characteristics

Statistics South Africa
A spaza shop is an informal convenience shop business in South Africa, usually run from home.
“People who lose their job, or can’t find work in the major urban areas are returning home and opening spazas”
The trend of Spaza Stores follows that of unemployment with a correlation of 86%.

\[ r = 0.86 \]

Statistics South Africa (Unemployment Numbers)
Nielsen Data Universe Estimation (Spaza Universe Numbers)
At a regional level six of nine provinces show a strong correlation \((r > \text{abs}(0.6))\).
What are the implications?

Addressing goal 8 of the UN Sustainable development goals

<table>
<thead>
<tr>
<th></th>
<th>Change YoY in Unemployment</th>
<th>Change YoY in Spaza Numbers</th>
<th>% of Unemployment change accounted for by spaza openings</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-17 W1</td>
<td>363,000</td>
<td>4,179</td>
<td>1.15%</td>
</tr>
<tr>
<td>2016-17 W2</td>
<td>243,000</td>
<td>3,628</td>
<td>1.49%</td>
</tr>
<tr>
<td>2016-17 W3</td>
<td>197,000</td>
<td>1,723</td>
<td>0.87%</td>
</tr>
</tbody>
</table>

Difficulties in opening stores:
- Access to credit
- Access to goods
- Entrepreneurial Skills

Levers that can be pulled:
- Governmental loans to grassroots entrepreneurs
- Creation of networks to enable buying partnerships and easier distribution
- Small scale educational programs via training colleges