nielsen

Big Data and Econometrics: 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"

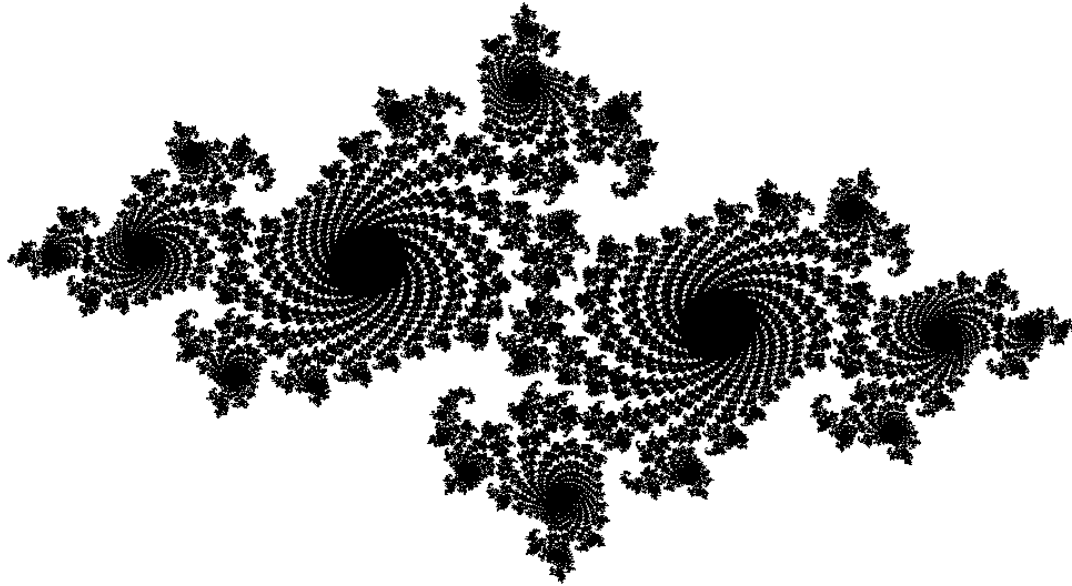Disparate data sets allow for better validation and creation of new insights

X 8 Billion

VS

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?

BUZZ
Volume of conversations

SENTIMENT
NSS™ (Net Sentiment Score)

TOPICS
Identification of conversation topics

EMOTIONS
7 pairs of opposite emotions (proprietary AI model)

KPIs
Purchase intent, recommendation, price, availability

ENGAGEMENT
Consumer activity generated by corporate activity

INFLUENCERS
Based on buzz or posts, likes and shares

IMAGES
Logo and image theme detection

Twitter   Video   Facebook*   Instagram*   Reddit**   Weibo**   VK**   Blogs   Forums   News   Reviews
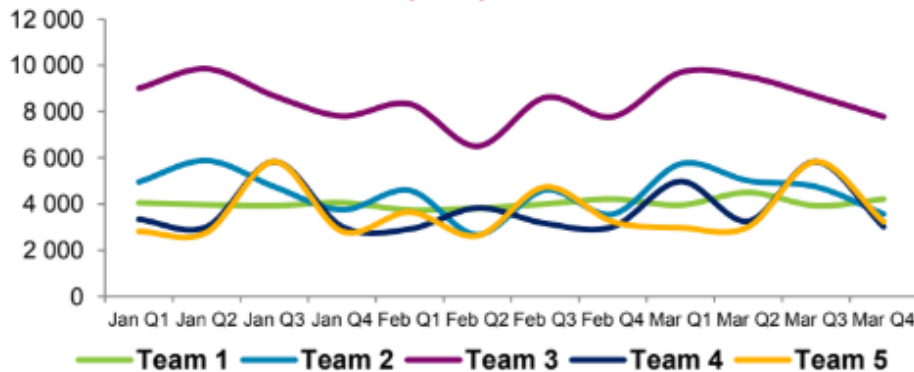
* For Facebook we harvest data based on page search (not keyword search), and are only able to harvest from public pages For Instagram we are only able to harvest from up to two brand or group pages.
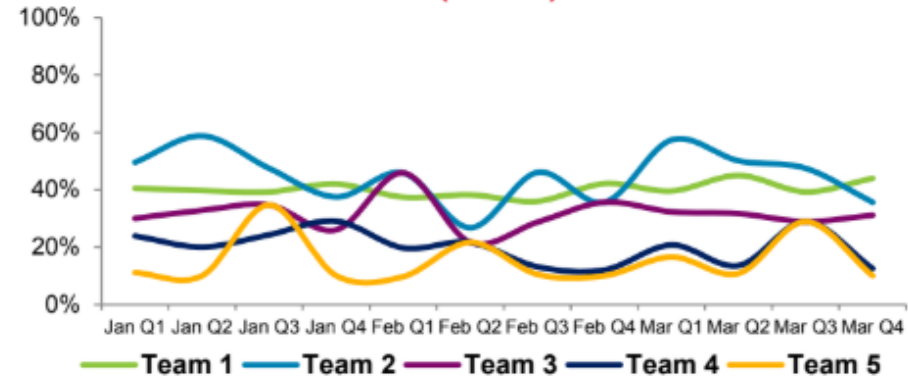** An additional data cost may be incurred for these these sources.
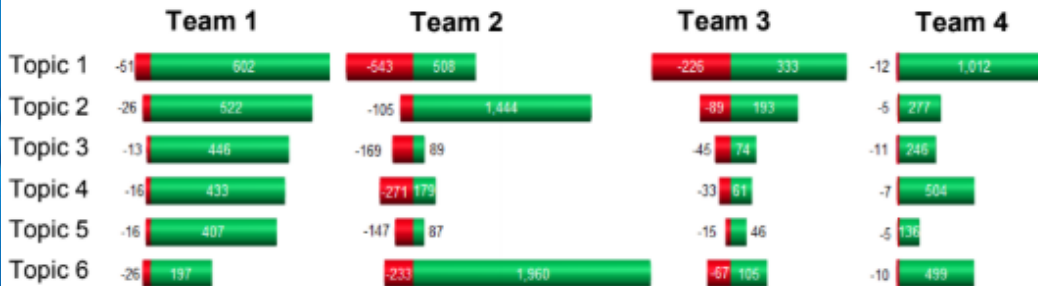
Example of premiership football teams

SHARE OF VOICE (SoV) – QUARTERLY TREND

Legend: Team 1, Team 2, Team 3, Team 4, Team 5

NET SENTIMENT SCORE (NSS™) – QUARTERLY TREND

Legend: Team 1, Team 2, Team 3, Team 4, Team 5

TOPICS – VOLUME & SENTIMENT

	Team 1	Team 2	Team 3	Team 4
Topic 1	-51 / 602	-543 / 508	-226 / 333	-12 / 1,012
Topic 2	-26 / 522	-105 / 1,444	-89 / 193	-5 / 277
Topic 3	-13 / 446	-169 / 89	-46 / 74	-11 / 246
Topic 4	-16 / 433	-271 / 179	-33 / 61	-7 / 504
Topic 5	-16 / 407	-147 / 87	-15 / 46	-5 / 136
Topic 6	-26 / 197	-233 / 1,960	-67 / 105	-10 / 499

KEY INFLUENCERS
mentioned in the context of Premier League

Football Influencer	Buzz	NSS™
Name 1	14,905	32%
Name 2	10,934	45%
Name 3	4,042	14%
Name 4	3,043	55%

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
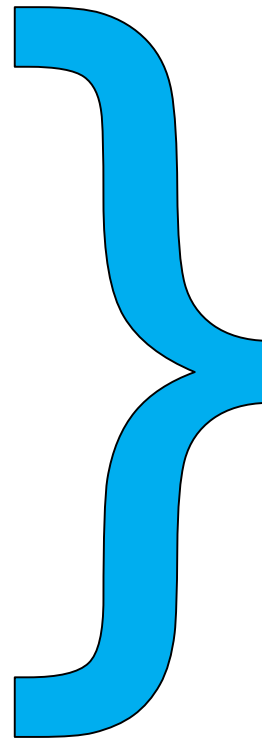
The Food Consumption Score (FCS)

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.
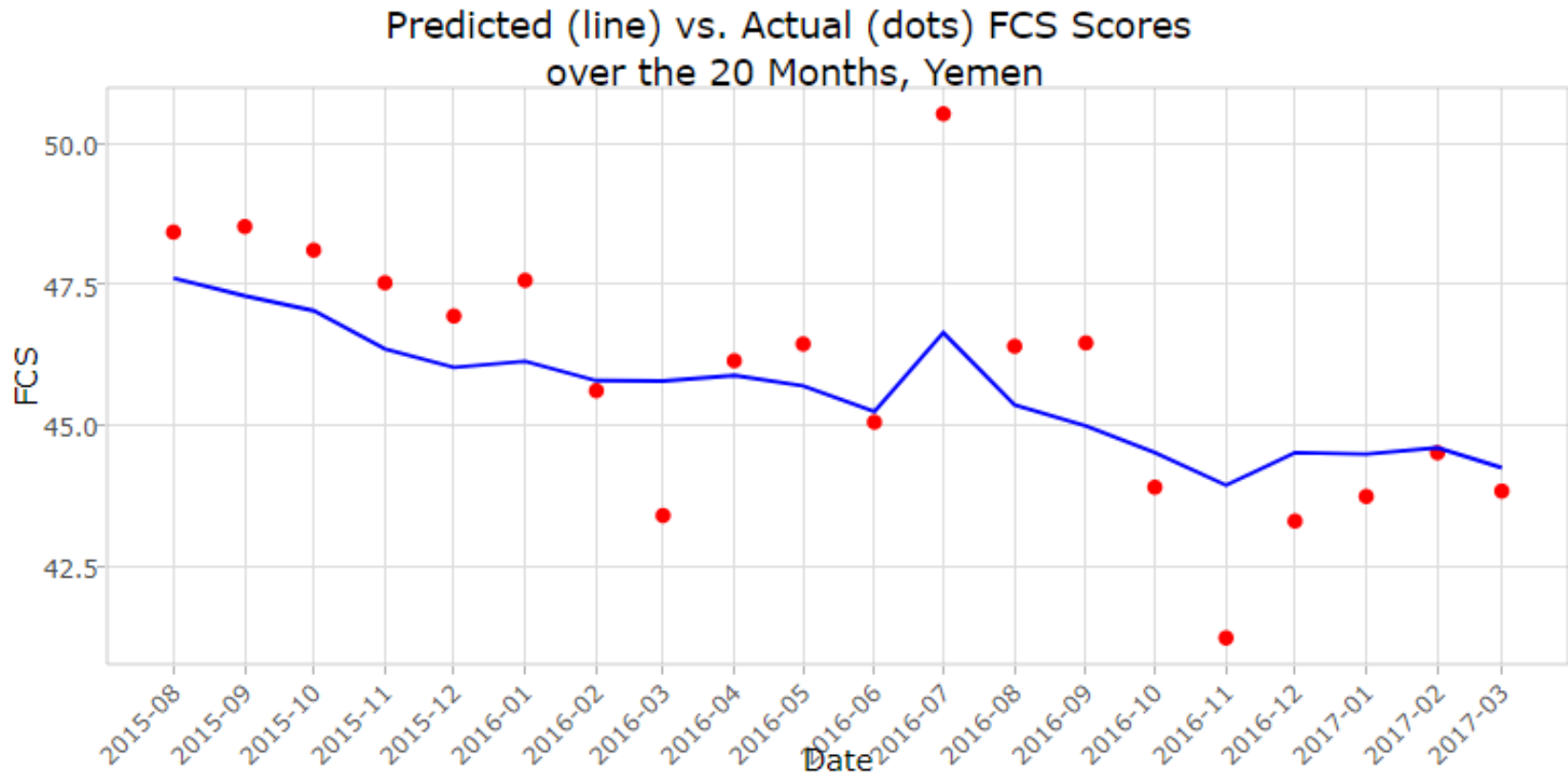
Measure of success: A model that can predict the FCS with 80% accuracy

Commodity prices showed little predictive ability but the previous months score by food type showed some strong correlations to the next months FCS
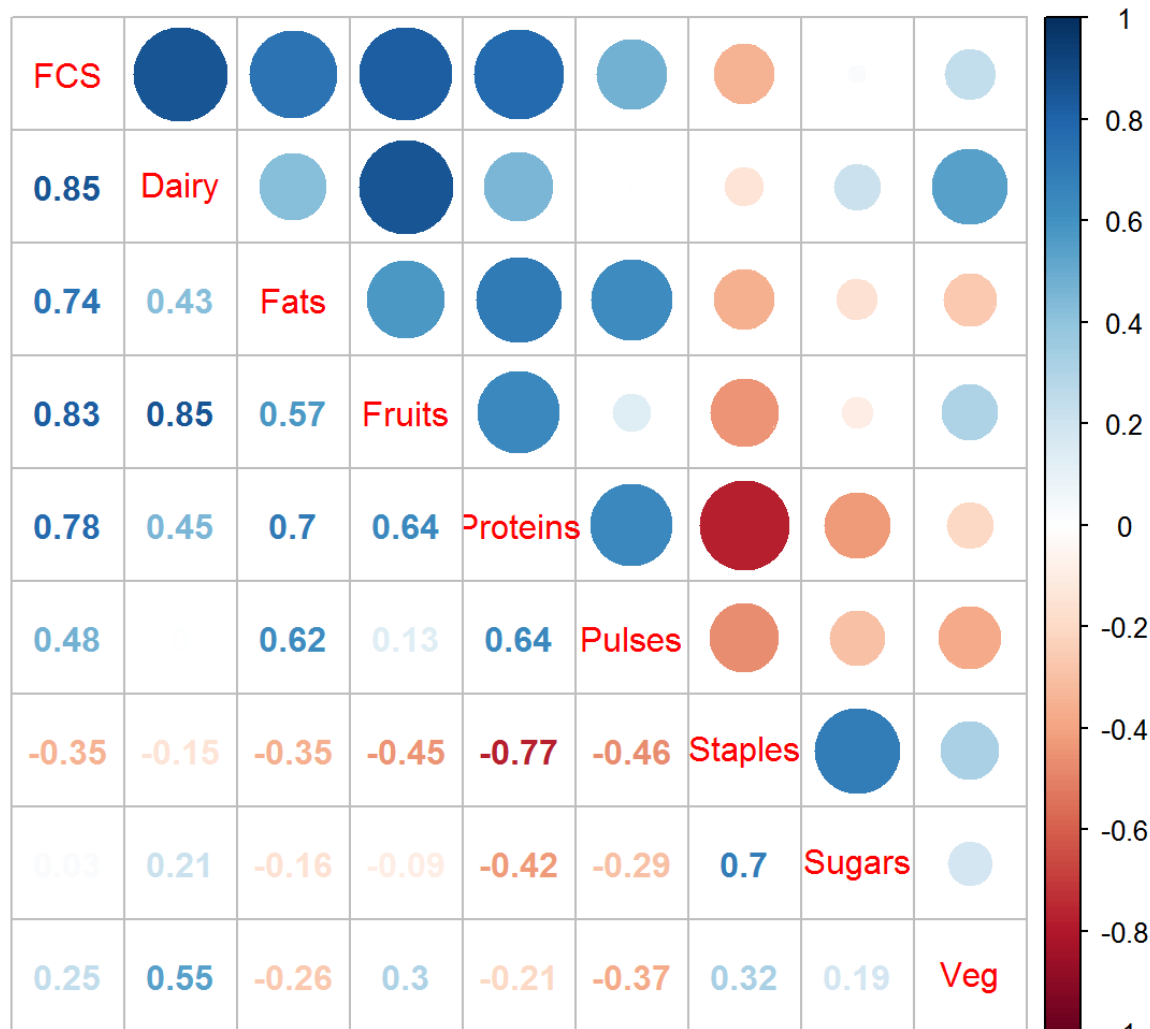
The final model showed an average accuracy of 9[...]

Addressing goal 2 of the UN Sustainable Development Goals

Predicted (line) vs. Actual (dots) FCS Scores over the 20 Months, Yemen

Home stores vs Unemployment

THE RETAIL AUDIT PROCESS

5. Reported Outputs

4. Statistical expansion
of sample to universe

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

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

1. Establish Retail Trade Universe
Define store types and gain information on shop numbers & turnover

Establishing the Retail Trade Universe

Rolling Establishment survey is conducted every six months
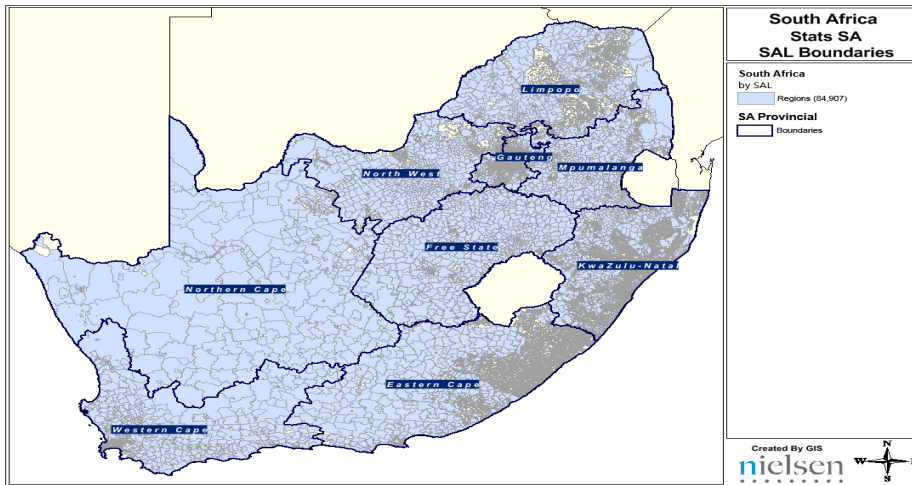
Starting with the smallest statistical representation of the country

(South Africa = Small Area Layers)

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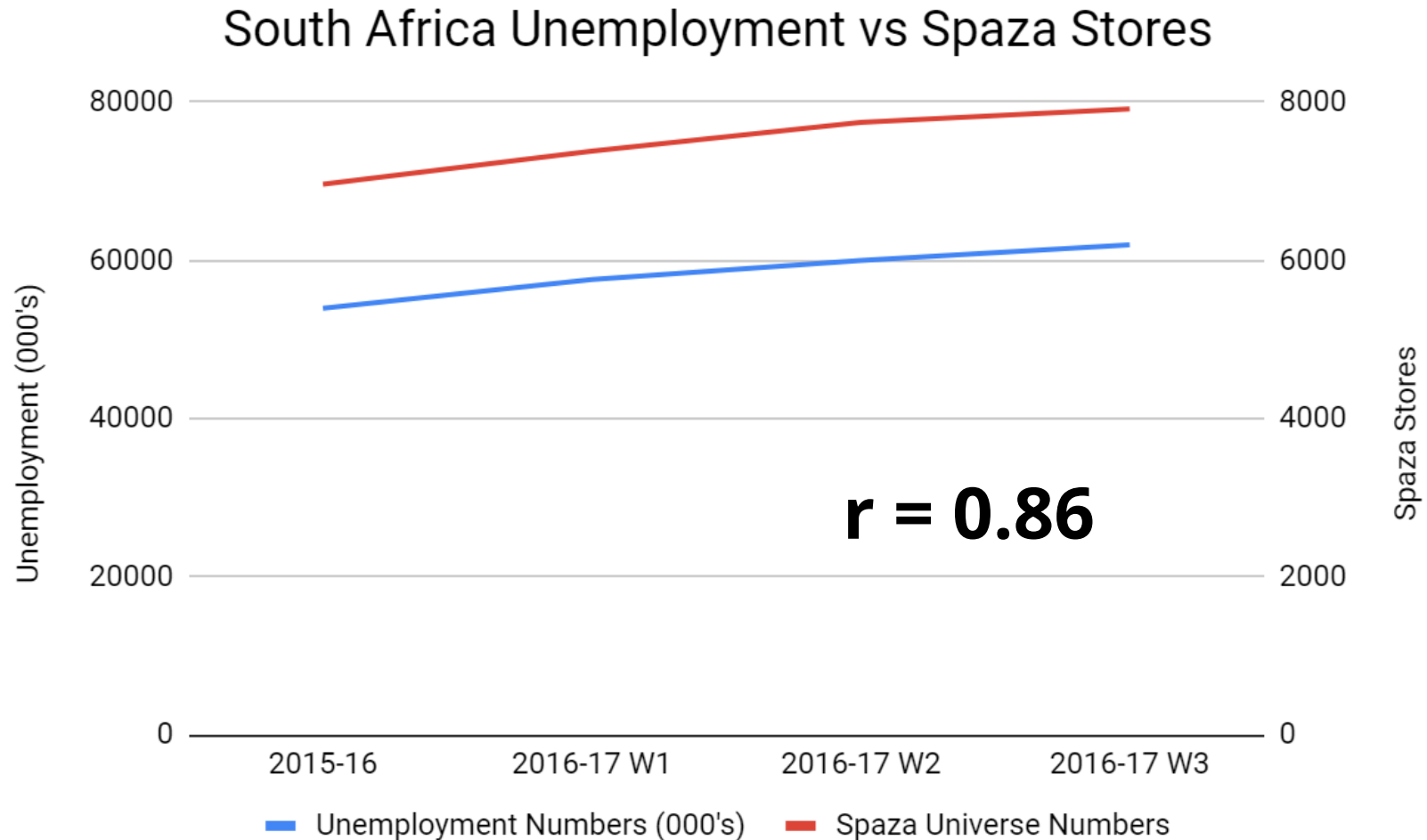

Statistics South Africa

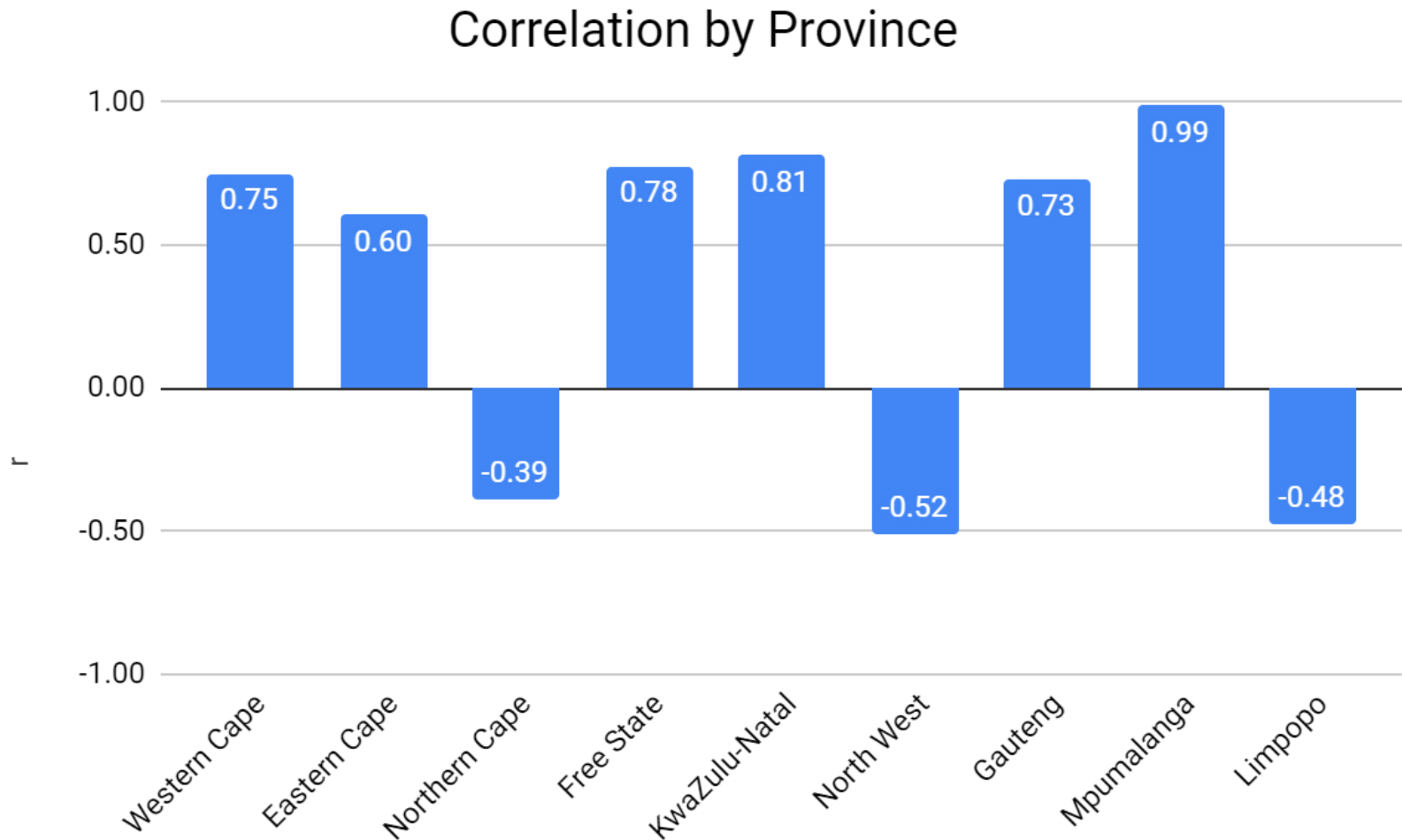# A spaza shop is an informal convenience shop business in South Africa, usually run from home.

"People who lose their job, or can't find work in the major urban areas are returning home and opening spazas"

Nielsen Data Acquisition

The trend of Spaza Stores follows that of unemployment with a correlation of 86%

South Africa Unemployment vs Spaza Stores

r = 0.86

Unemployment (000's) — left axis: 0, 20000, 40000, 60000, 80000

Spaza Stores — right axis: 0, 2000, 4000, 6000, 8000

x-axis: 2015-16, 2016-17 W1, 2016-17 W2, 2016-17 W3

— Unemployment Numbers (000's) — Spaza Universe Numbers

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 > abs(0.6))

Correlation by Province

Statistics South Africa (Unemployment Numbers)
Nielsen Data Universe Estimation (Spaza Universe Numbers)

What are the implications?

Addressing goal 8 of the UN Sustainable development goals

	Change YoY in Unemployment	Change YoY in Spaza Numbers	% of Unemployment change accounted for by spaza openings
2016-17 W1	363,000	4,179	1.15%
2016-17 W2	243,000	3,628	1.49%
2016-17 W3	197,000	1,723	0.87%

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

nielsen