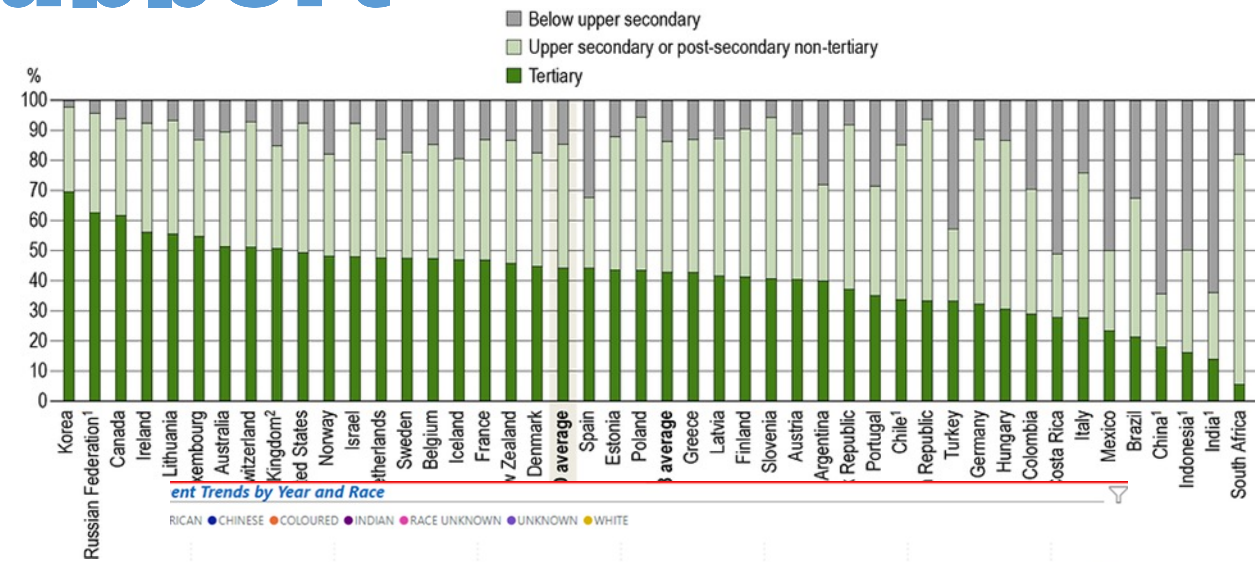
University of the Witwatersrand

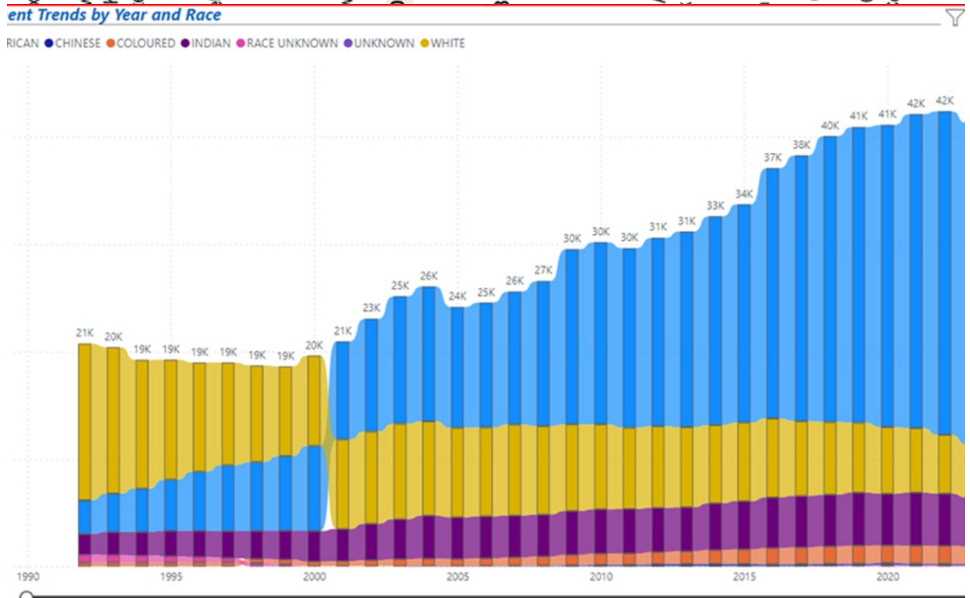
Reflections on Siyaphumelela 2.0

Siyaphumelela Conference June 2024

Context: Increased Enrolment and Diversification Underpins Need for Support



OECD 2019 Study of Tertiary Attainment – SA has lowest level of tertiary attainment of 38 members



Wits increase and diversification of enrolment over 30 years – schooling background and academic preparedness

Interacting Contextual Elements: Predictive Modelling for Student Success

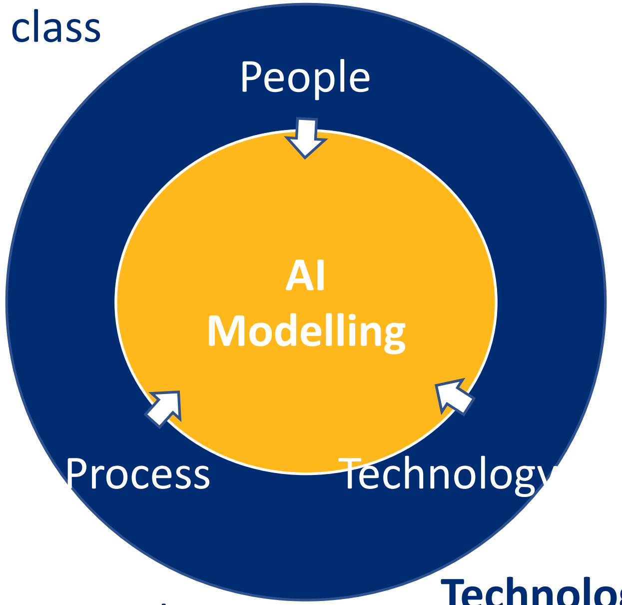
UG throughput rate in minimum time for 2020 cohort: 35.4%

Governance (4)
- Student Success Framework
- Data Governance Framework
- Monitoring & Evaluation Framework
- Student Success Committee

Governance

People (5) ***
- Teaching and learning interaction in class
- Tutoring and mentoring
- Advisors arranging interventions
- Divisions offering student support

Process (3)
Capture of course structures in AMS
Capture of Marks
Capture of Interventions



AI Models (1)
Enables faculties to proactively identify the students that will need support before entering campus. Refined (BQ, LMS, marks etc.)

Technology (2)
- Data Warehouse
- Data Lake

Store/QA the data that feeds the models

Leading and Lagging Indicators
Course Pass Percentage Reports

Monitor progress and make Adjustments to max impact

Stages of Analytic Maturity (Davenport, 2013)

Analytics 3.0 (Impact)

Organisations realise measurable business impact from the combination of traditional analytics and big data. They begin to understand what type of interventions work for which students and at which point in the learning cycle, and which policies and practices can avoid roadblocks to contribute to student success

M&E of all student success projects

SIYA 3.0

Analytics 2.0 (Predictive)

Capitalises on the emergence of large, fast-moving, external, and unstructured data from sources such as learning management systems (LMS), biographical questionnaires, food bank access records, lab access, classroom attendance to understand the factors that lead to student success

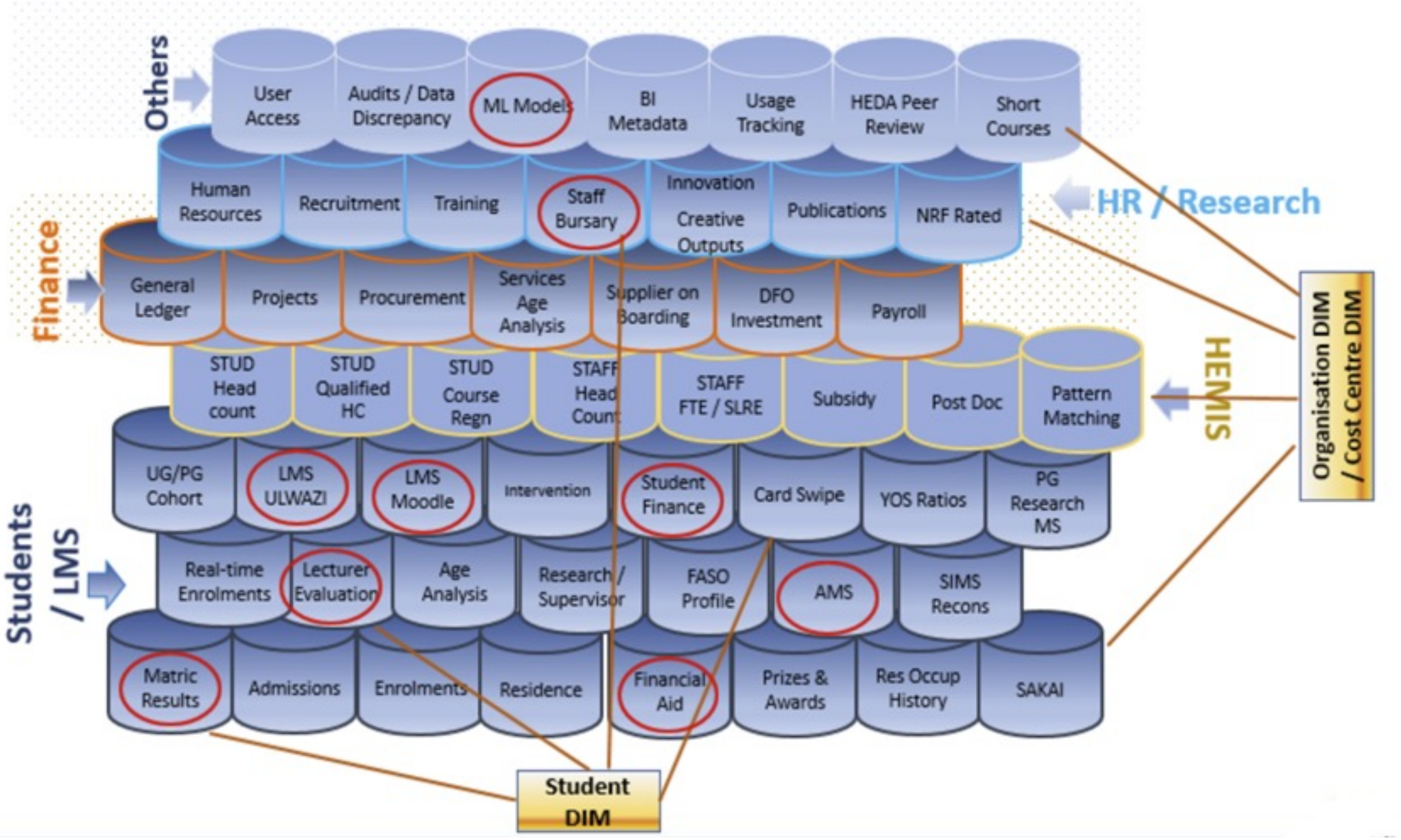
SIYA 2.0

Analytics 1.0 (Descriptive)

- Assembling business intelligence systems and expertise to drive reporting and descriptive analytics
- Standard measures of student success
- Persistence, completion and placement rates

SIYA 1.0

Wits University Data Warehouse (EDW)



50+ MD Data Marts
Data Prepared,
Structured, and
Quality Assured
For Analysis

Lessons Learnt from Siyaphumelela 2.0

- Student success is not a project, it’s a programme
- People come and go but the programme needs to remain intact
- Sustained support from the highest level is crucial
- Members of the SSC need to be decision makers and represent a constituency
- You can’t get to Cape Town without going through the Karoo

Lessons Learnt from Siyaphumelela 2.0

- Engage the academics – student success begins in class
- Foreground students in the student success efforts and provide them with information
- Be clear on which students to assist
- Set a vision (targets) for improvement
- Regularly engage the advisors

Plans for Siyaphumelela 3.0

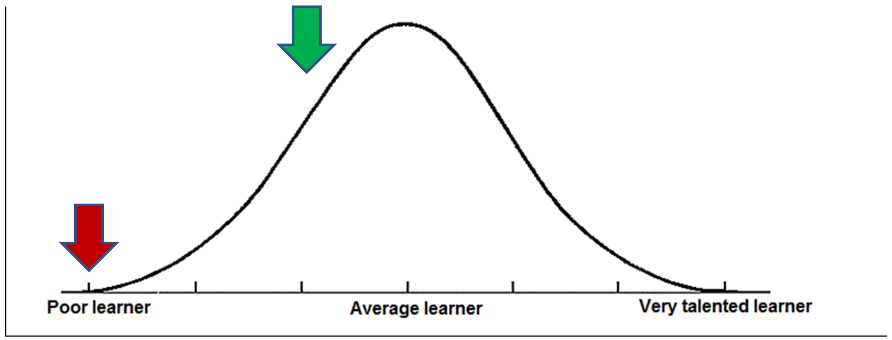
Need to drive impact to move the needle

Establish a monitoring and evaluation framework

Be clear on where you want to go before beginning the journey (KPIs)

Every student success project will have a monitoring and evaluation component

Understand who our advisors have been helping and which interventions lead to success



STUDENT INVOLVEMENT

Research conducted by **Thakatau Hlakotsa**, one of our student interns from EBE who became a Data Scientist in BIS (presented by intern James Zungunde), into determining which type of interventions assists which type of students

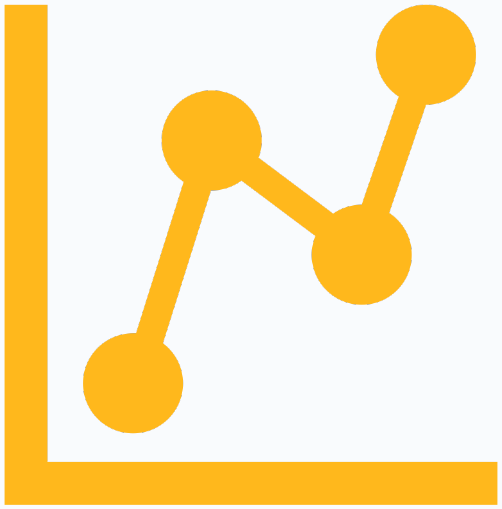
Research conducted by intern **Francis Phiri** on using AI to improve data quality

Statistical/Machine Learning Models:

1 Identify students who **require help** ahead of time.

2 Assess the **effect of Interventions on student performance**.

3 Discover **features that impact student success** via the predictive power of the feature in the model.



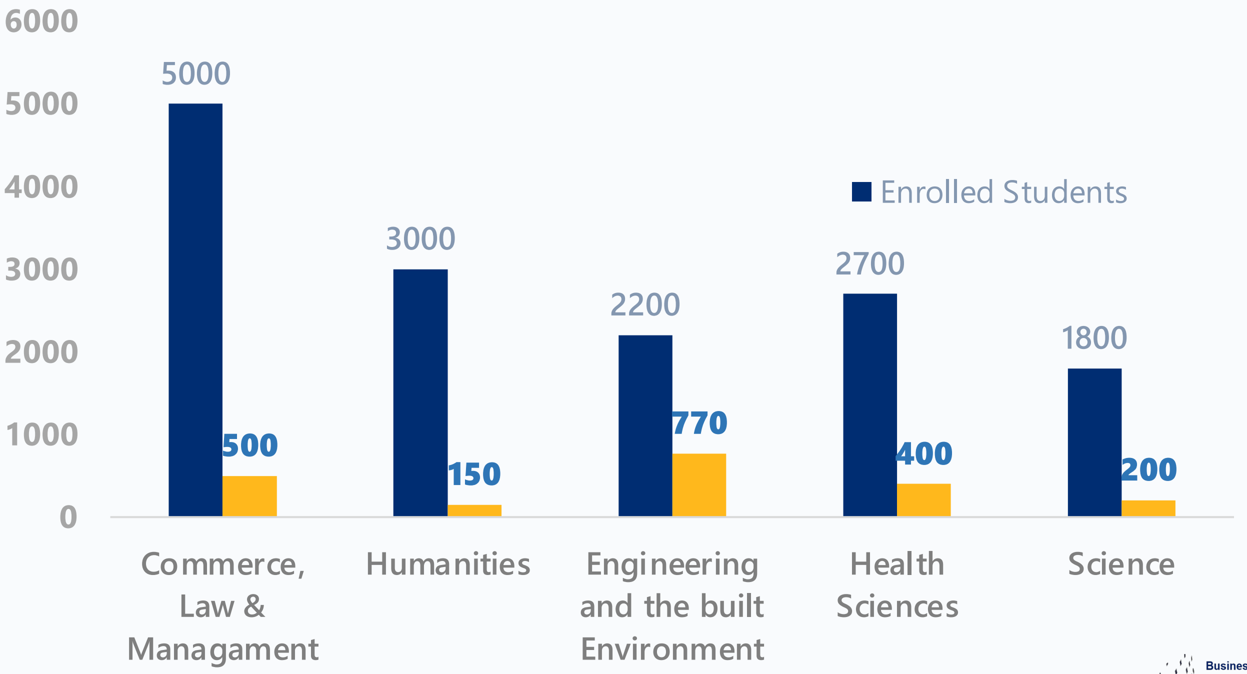
Statistical/Machine Learning Models:

- Step 1: Identifying at-risk student.
- Data shared via Power BI dashboards.
- ****Numbers are made up****

Example of how **AT-RISK students are identified per Faculty**

2020

AT-RISK Students



Business Intelligence Services

Statistical/Machine Learning Models:

- Linking AT-RISK information and Intervention Capturing Information to determine who has received an intervention

700!

Have an **intervention**

2020 AT-RISK students

Key Performance Indicator

Statistical/Machine Learning Models:

Analysing Interventions together with AT-RISK data to determine where we can obtain the greatest impact.

Assessment of previous data points towards a positive relationship between student success and interventions.

Specifically, we see that the more interventions for AT-RISK students coincide with better pass rates for AT-RISK students.

Looking into Issue Types data to see if students with specific issues are more influenced by Interventions.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

BUSINESS INTELLIGENCE SERVICES

Machine-Learning Aided Data Quality Assurance

Francis Phiri – BIS Intern, MSc Student

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

100 1922 2022

Presentation Outline

1. Impacts of Poor Data Quality
2. Data Quality Dimensions
3. AI/ML in Data Quality Management
4. AI/ML in Data Quality Framework

The Cost of Poor Data Quality

Data Quality → Information Quality → Decision Quality → Business Outcomes → Organization Performance

Incorrect Decisions

- Incorrect student rejections or admissions
- Incorrect student course registrations
- Inaccurate decisions on financial aid eligibility

Reduced Productivity

- Time wasted in accuracy validation
- Time wasted in data recollection
- Time wasted in correcting erroneous data

Missed Opportunities

- Poor data leads to inaccurate forecasting and misleading predictions
- Inaccurate forecasting may lead to bad decisions which may drive away student applications

Reputation Damage

- Poor data quality leads to bad decisions which may lead to poor student experiences, and this can impact the university's reputation

Accreditation Risks

- Poor data may lead to inaccurate assessments of an institution's accreditation status

Data Quality Dimensions

AI/ML in Data Quality Management

	Non-Learning DQ Approaches	Learning DQ approaches
Scope	• Precise, well defined business rules • Limited coverage, focusing on few DQ dimensions (completeness, duplication, validity)	• More variation and “fuzziness” of business rules • Higher coverage, including additional DQ dimensions (provenance, accuracy, consistency)
Creation	• Business rules learned from business experts, validated with data • Hard-coded in DQ platform by DQM manager	• Business rules are learned from data, validated by business experts • Custom development by data scientist built on big data infrastructure
Run	• Automatic checks performed against constraints on regular basis	• Automatic checks performed against the model on regular basis
Update	• Punctual update of checks to add/remove/change procedure in place • Manual by DQM manager	• Retraining of model based on updated data and monitoring of models and its results • Little to no intervention
Follow-up process	• Same channel • Similar follow-up process • Accepted as such by users	• Same channel • Similar follow-up process • Training to raise awareness of users for uncertainty of results/false positives

Data matching

Anomaly Detection

Data Lineage

Data Augmentation

Missing Data Prediction

Fix Recommendations

Large Language Models

NLP

Data Quality Dashboard

80% Provenance
Threshold: 90%

82% Duplication
Threshold: 85%

92% Completeness
Threshold: 90%

73% Accuracy
Threshold: 90%

95% Validity
Threshold: 90%

AI/ML DQ Framework