Griffith UNIVERSITY

AdvAna *and the Guru*

The future's so bright I gotta wear shades

Presented by CLANCY BIRRELL
Advanced Analytics Manager, Office of Planning Services

Path

GoT

Focus

Projects

Shades

Glossary of terms

Griffith University
Office of Planning Services

AdvAna – ADVanced ANAlytics

Descriptive:

Just the facts

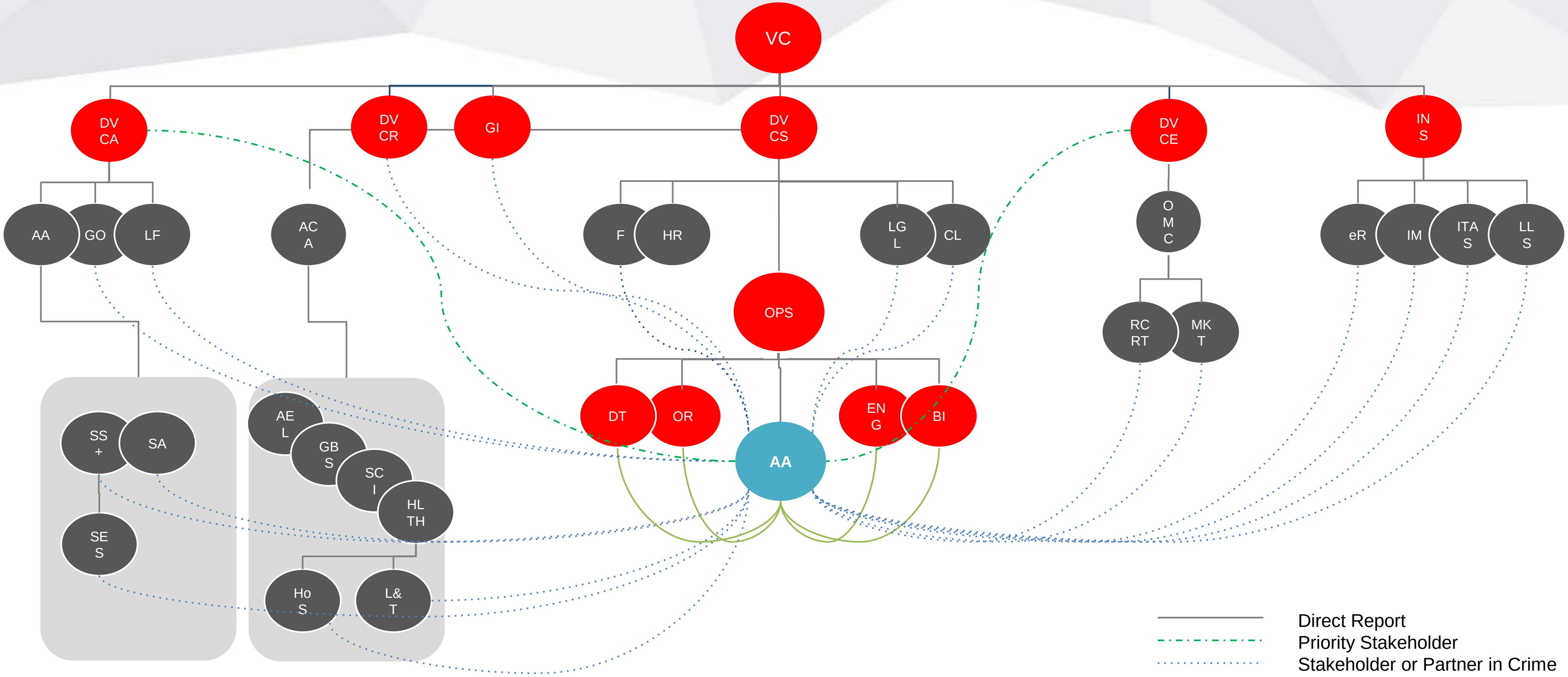
Predictive:

Predict the future based on the past

Prescriptive:

What happens if I push this shiny red button?

The world according to...

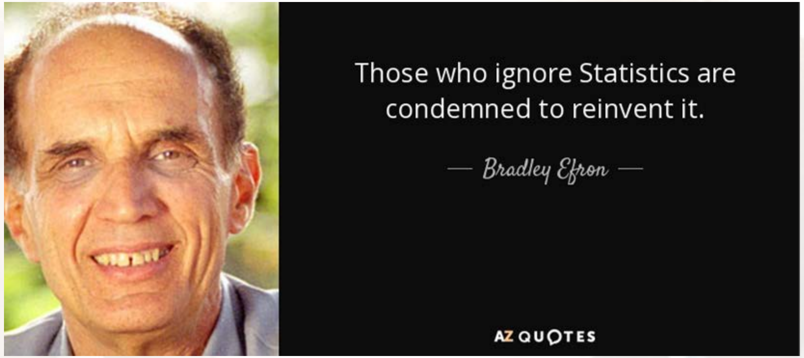


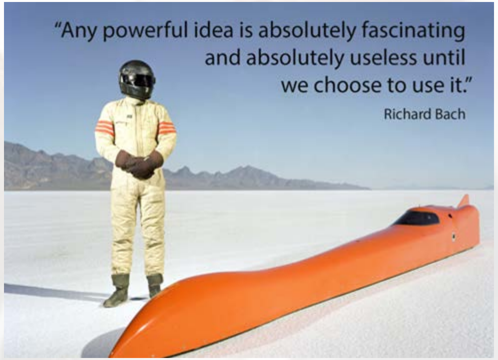
A word on Guru's

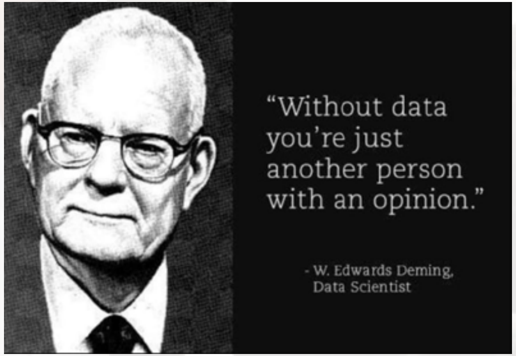
The gurus I have met personally, as well as those whose careers and teachings I have studied at a distance, range from crooks who could be quickly dismissed to teachers who were brilliant but flawed, to those who, while still human, seemed to possess so much compassion and clarity of mind that they were nearly flawless ...

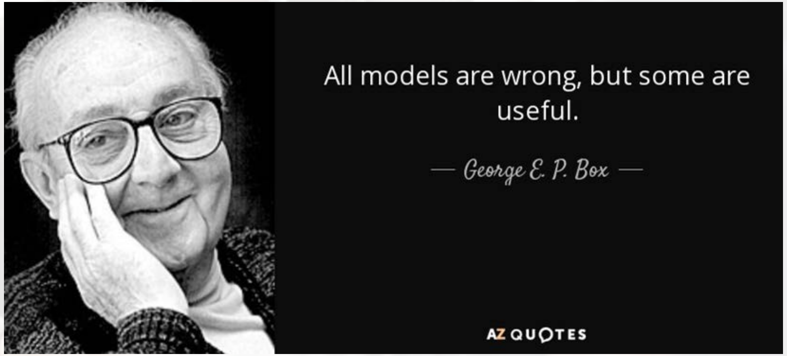
Sam Harris: Waking Up

Therefore ~









www.analyticshero.com/2012/10/25/31-essential-quotes-on-analytics-and-data/

Griffith UNIVERSITY

Applied statistics vs. data science

Jeremy Jarvis
@jeremyjarvis
Follow

"A data scientist is a statistician who lives in San Fransisco"
#monkigras pic.twitter.com/HypLL3Cnye

5:13 AM - 30 Jan 2014

1,469 836

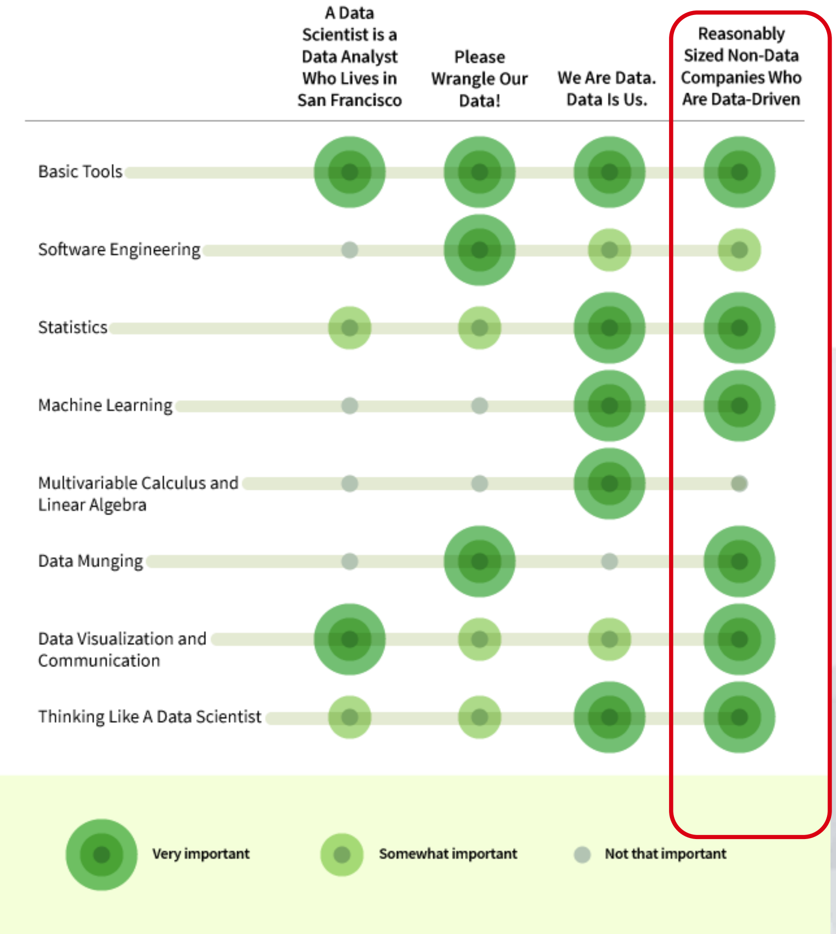
Big Data Borat
@BigDataBorat
Follow

Data Science is statistics on a Mac.

8:32 AM - 27 Aug 2013

610 269

Accenture LLP [US] | https://www.accenture.com/us-en/blogs/blogs-the-difference-between-data-science-and-statistics



blog.udacity.com/2014/11/data-science-job-skills.html

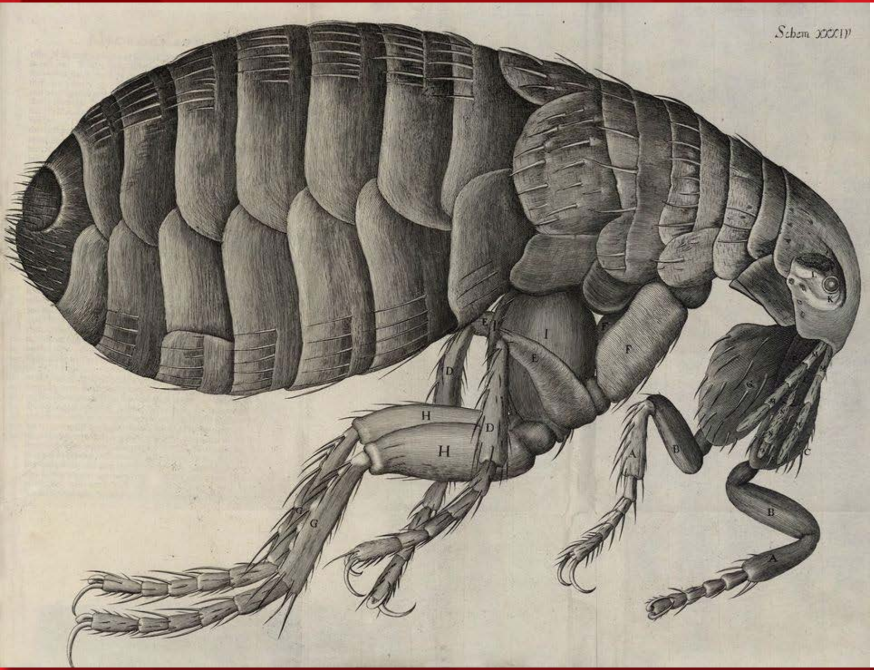
Griffith UNIVERSITY

Applied statistics vs. data science

- Standard practice
- Accepted methods
- Pre built data
- Groomed data
- Simple data sources
- Pre built point and click tools
- Running fixed and well designed experiments
- Statistical inference
- German Engineering Style Rigour
- Long form reporting of results
- Usually single domain application
- Develop statistical analyses

- No standards
- Anything that works
- Data munging
- Messy all types of data
- Many sources
- Code slinging & data carpentry
- Experimenting any which way you can
- Some statistical inference but more focus on ROI and business results
- As required rigour
- Short form results, business style strategic summaries and agile results
- Cross domain
- Develop data products

Aaargh parasites!



- We don't exist without data!
- Low quality data = low quality analytics
- Limited access to data = limited analytics
- If the data don't exist we don't analyse it

But above all ~

It all starts with a well formulated question.


FOCUS
Griffith
UNIVERSITY
Office of Planning Services

Student lifecycle


Recruitment


Application

Enrolment


1st Year


2nd Year


Argh Year


Completion


Continuation


Employment

Griffith UNIVERSITY
Office of Planning Services

Student Lifetime Value Optimisation

Promotability

Staff Retention Prediction

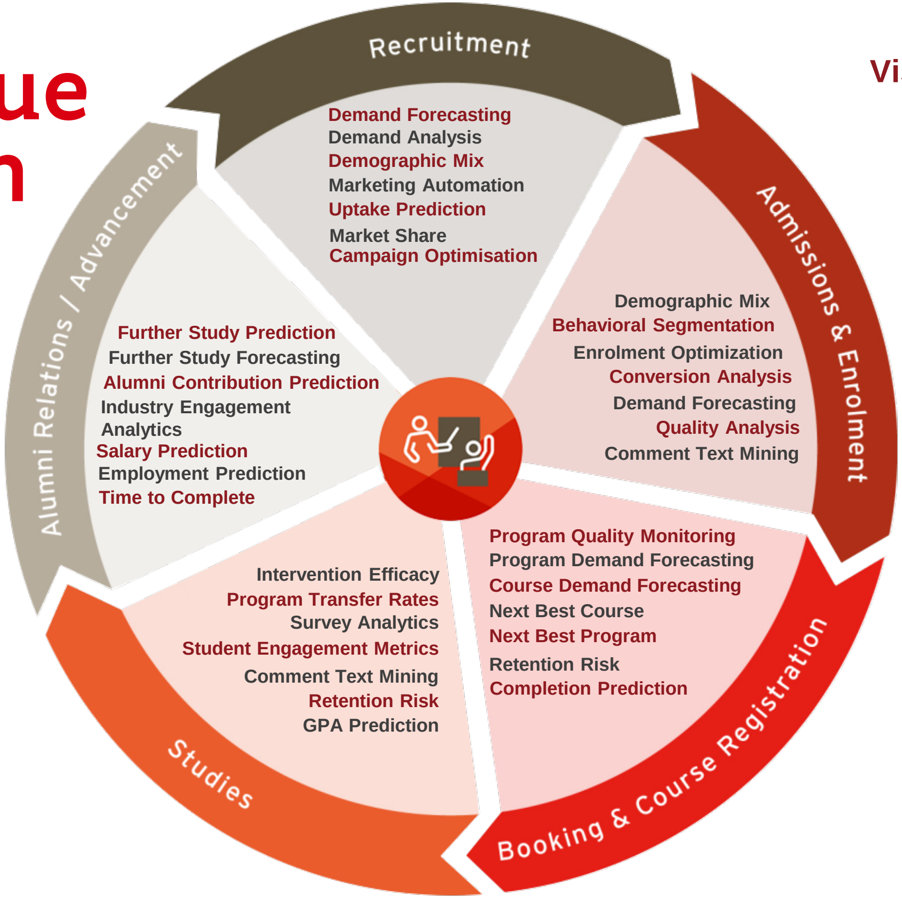
HR Analytics

STRATEGY ANALYSIS

Funding Sensitivity Analysis

Enterprise Retention Risk

Research Project Expected Cost



Analytics Server

Enterprise Visual Analysis Tool

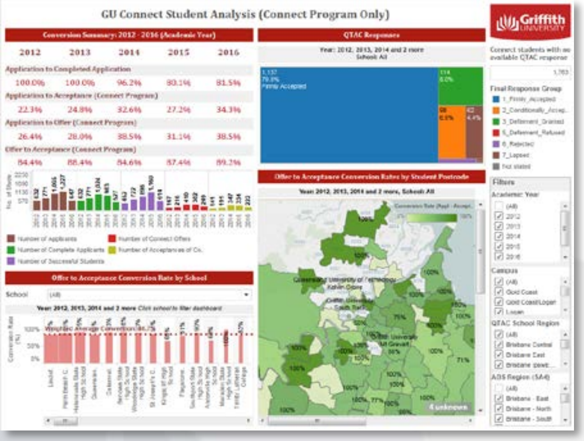
Student 360º

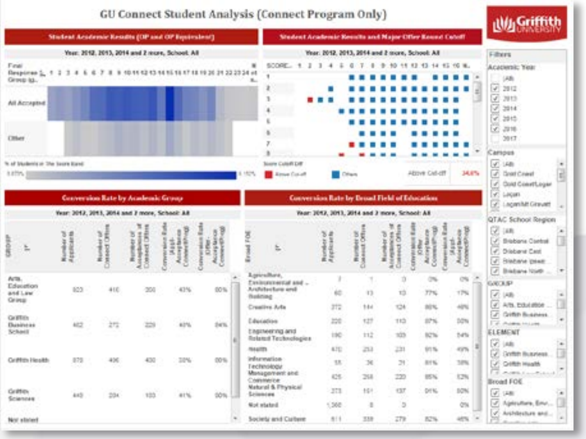
Data Lake

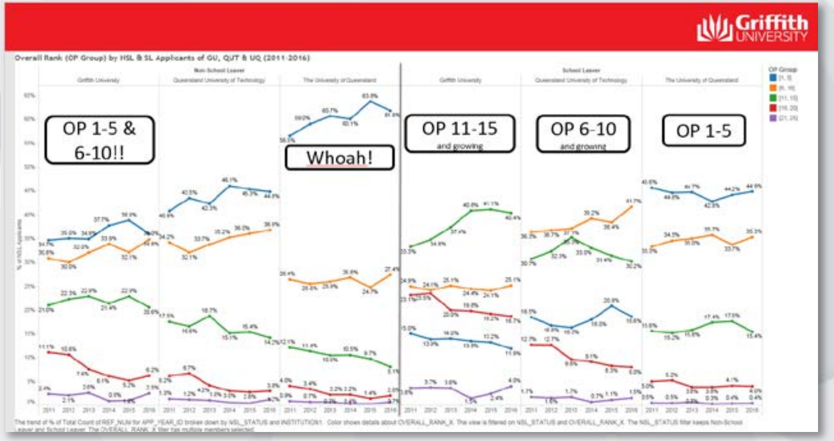
DECISION SUPPORT

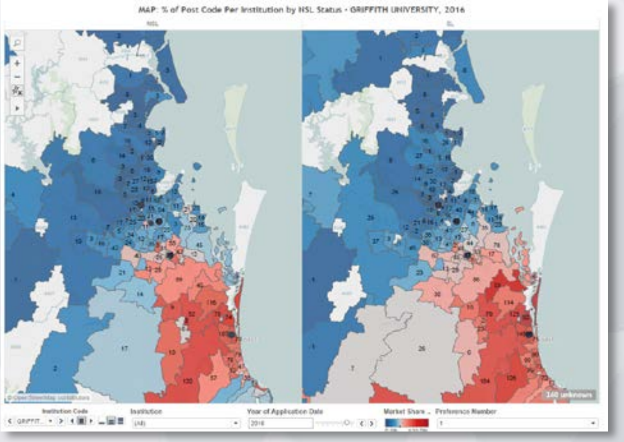
Griffith UNIVERSITY

Recruitment



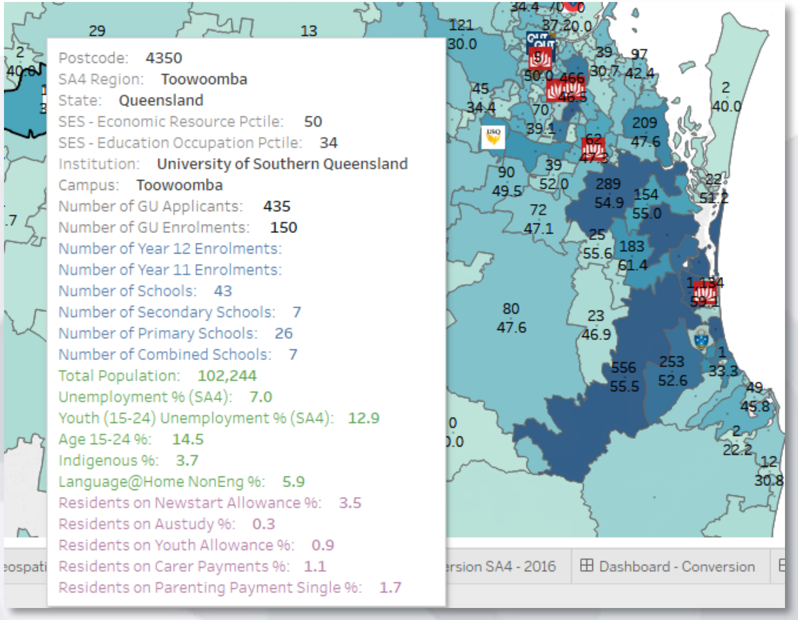




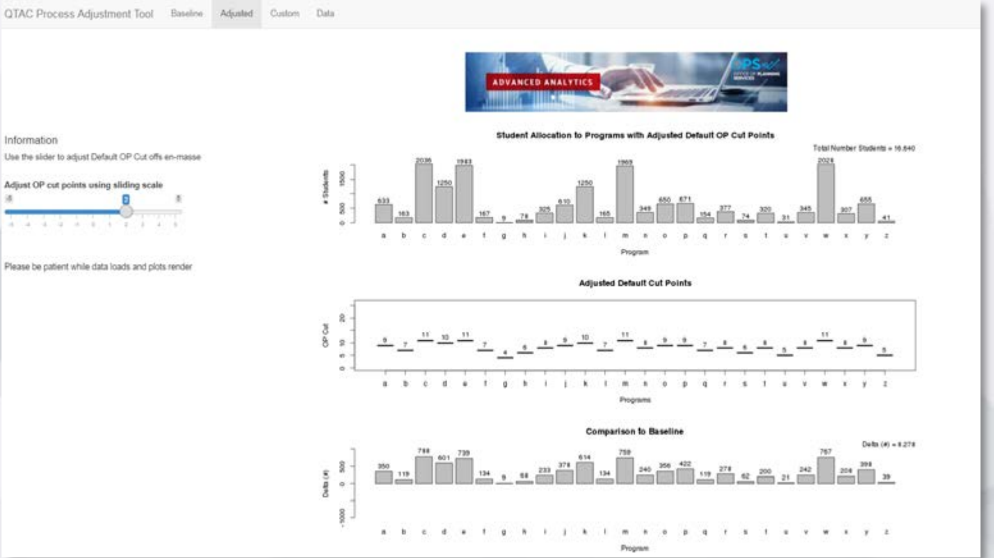


Griffith UNIVERSITY

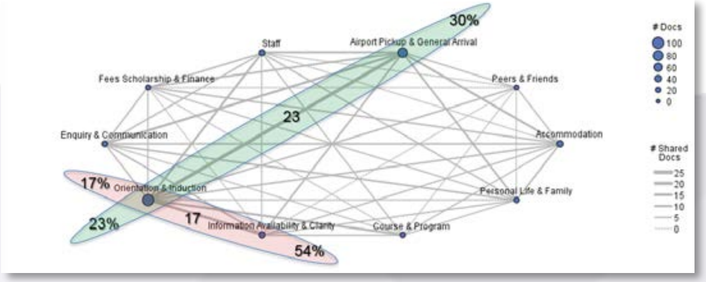
Application + Enrolment



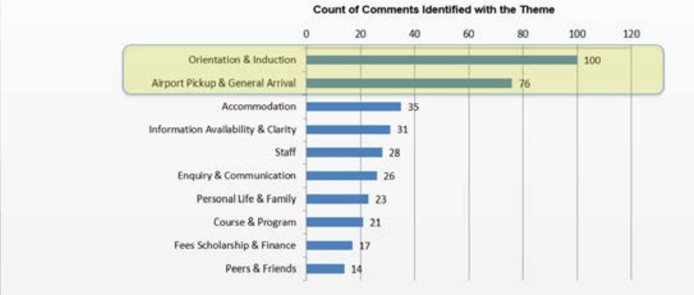
Application + Enrolment



Category Web (Co-occurrence of Themes)



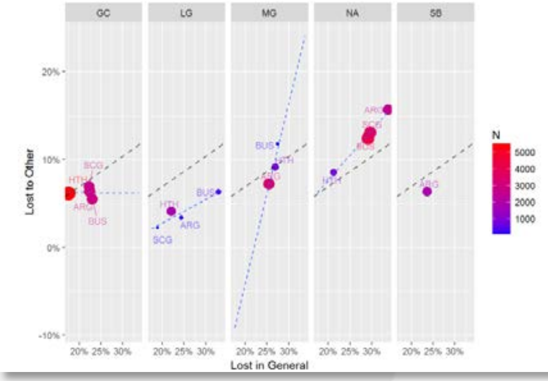
Top 10 Key Themes



Griffith UNIVERSITY

During studies

Q: What is the relationship between general loss of students vs lost to another institution?



Prefer QUT

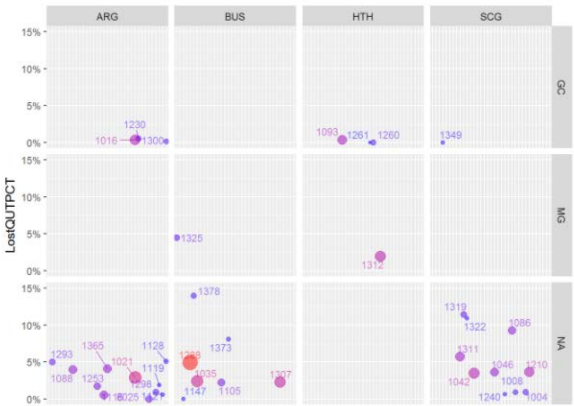
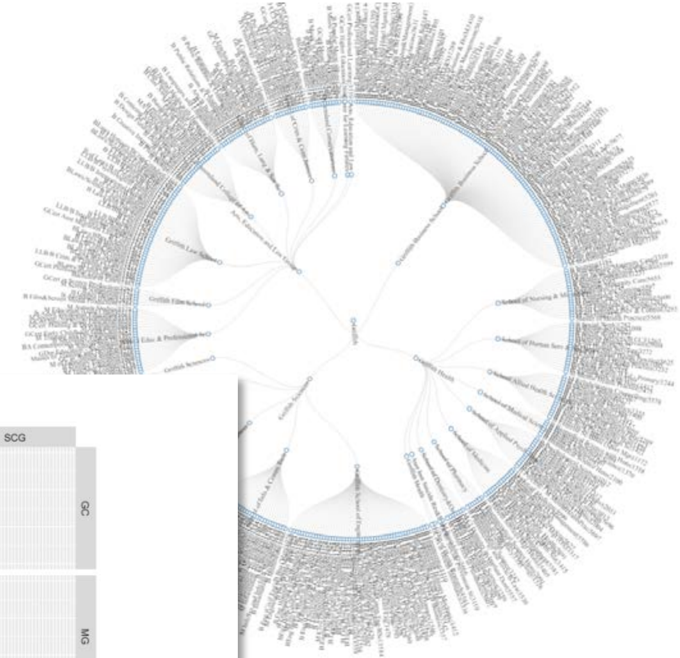
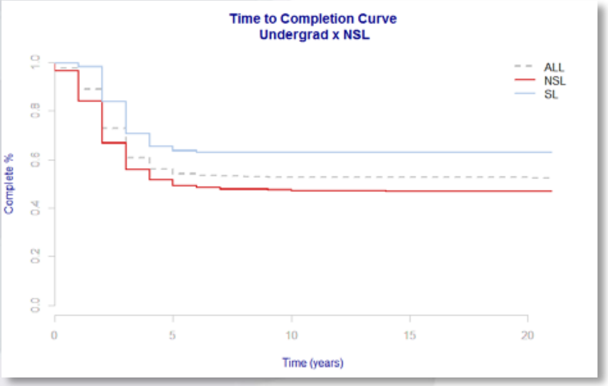


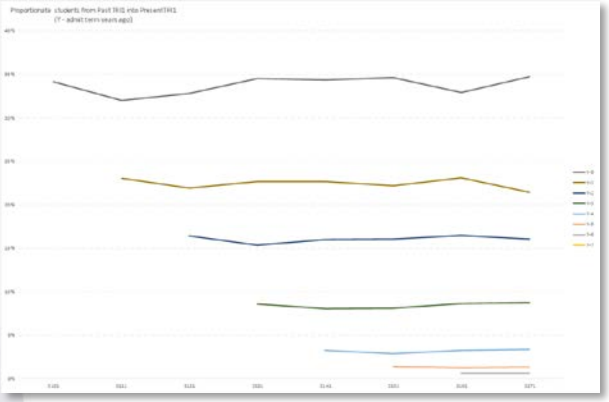
Figure 10.

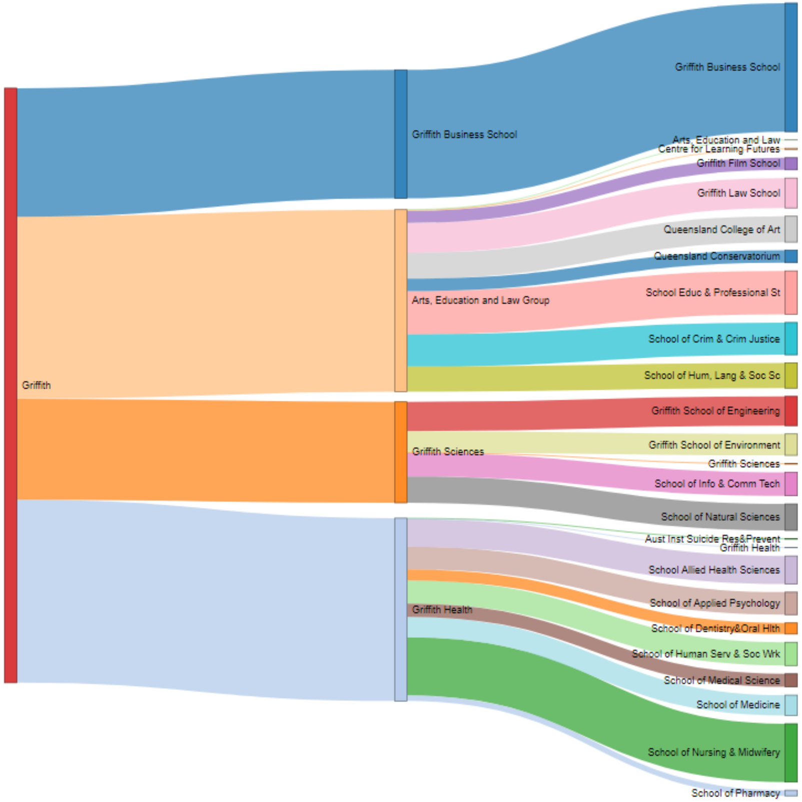
- BUS program 1378 - B Business/B Communication at Nathan campus has the highest rate of lost to QUT but is small in size
- SCG at Nathan campus seem to reflect a higher rate of lost with QUT preferences than other groups or campuses



During studies



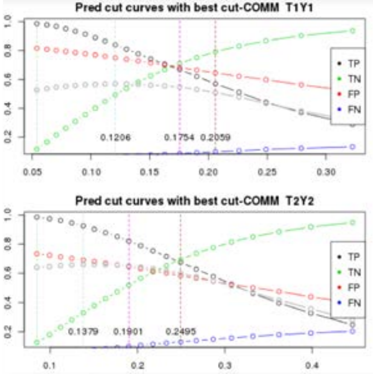




Retention

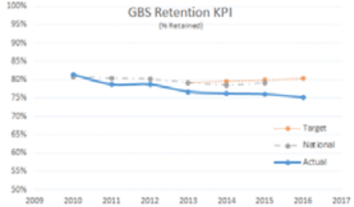
Home › r2017t2 › Code

Name
- step0.2-oracon.R
- step1.1.0-retn enrl.R
- step1.1.1-retn enrl.R
- step1.2.0-retn smpl.R
- step1.3.0-retn lst rsn.R
- step1.4.0-retn cmpl.R
- step2.1.0-retn stndng.R
- step2.10.0-retn mtrgrm.R
- step2.11.0-retn admappl.R
- step2.12.0-retn admapplprg.R
- step2.13.0-retn stdnt.R
- step2.14.0-retn crsgrd.R
- step2.15.0-retn sclshp.R
- step2.16.0-retn pthwy.R
- step2.17.0-RETN_GEO_TRANSIT.jar
- step2.17.1-retn sa1.R
- step2.17.2-retn dist.R
- step2.17.3-retn census.R
- step2.18.0-retn satg.R
- step2.2.0-retn bb.R
- step2.3.1-retn secset.R
- step2.3.2-retn secset.R
- step2.3.3-retn secset.R
- step2.3.4-retn secset.R



Bachelor Degree Retention

The proportion of all non-graduating bachelor degree students enrolled at census date in trimester one of a year (base year) who are enrolled in any Griffith program at the census date in trimester one of the next year.



Target:

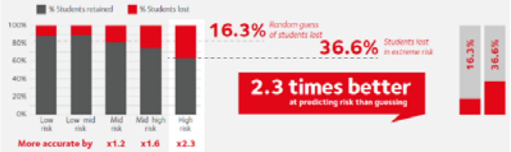
To exceed the national average for student retention by 2017

Predictive Analytics Models

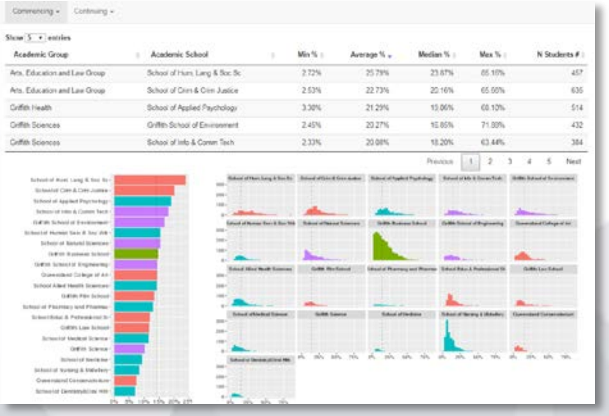
Predict Retention Risk

Machine Learning

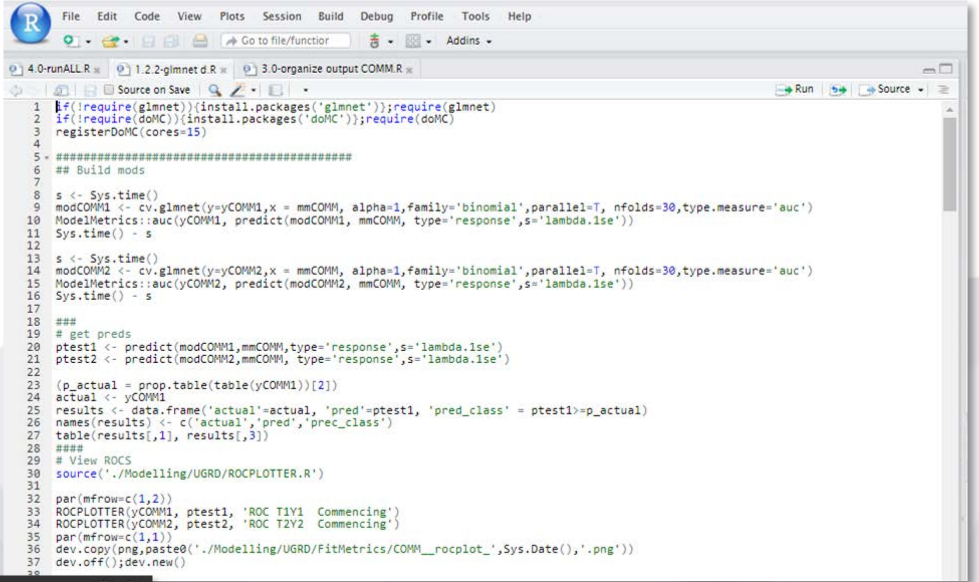
- ✓ 1000 variables considered
- ✓ Balances True Positive & False Positives
- ✓ Identifies patterns in noisy data
- ✓ Sensitive to low volume signals
- ✓ **Available NOW**



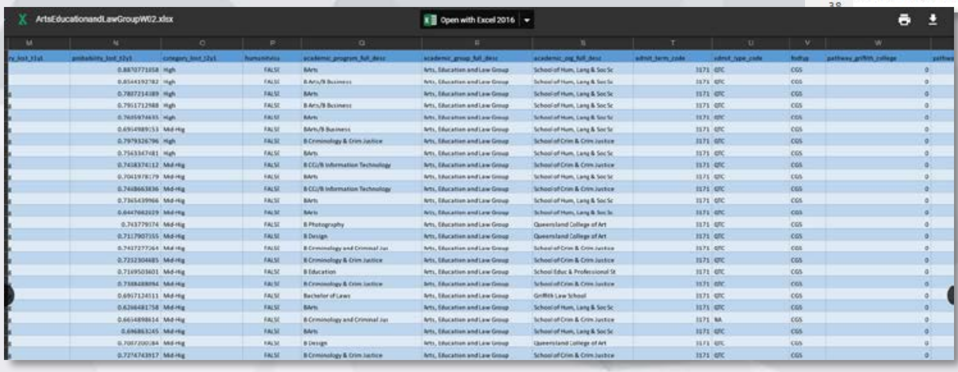
Retention



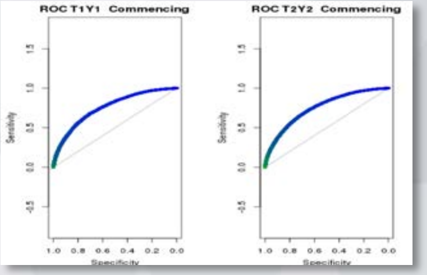
Commencing
Continuing
Show 5 entries
Academic Group | Academic School | Min % | Average % | Median % | Max % | N Students #
Arts, Education and Law Group | School of Hum, Lang & Soc Sc | 2.72% | 25.79% | 23.87% | 85.10% | 457
Arts, Education and Law Group | School of Crim & Crim Justice | 2.53% | 22.73% | 20.16% | 65.60% | 635
Griffith Health | School of Applied Psychology | 3.30% | 21.29% | 15.06% | 68.10% | 514
Griffith Sciences | Griffith School of Environment | 2.45% | 20.27% | 15.85% | 71.80% | 432
Griffith Sciences | School of Info & Comm Tech | 2.33% | 20.08% | 18.20% | 63.44% | 384
Previous 1 2 3 4 5 Next



File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
4.0-runALL.R
1.2.2-glmnet d.R
3.0-organize output COMM.R
Source on Save
Run
Source
if(!require(glmnet)){install.packages('glmnet')};require(glmnet)
if(!require(doMC)){install.packages('doMC')};require(doMC)
registerDoMC(cores=15)
##
Build mods
s <- Sys.time()
modCOMM1 <- cv.glmnet(y=yCOMM1,x = mmCOMM, alpha=1,family='binomial',parallel=T, nfolds=30,type.measure='auc')
ModelMetrics::auc(yCOMM1, predict(modCOMM1, mmCOMM, type='response',s='lambda.1se'))
Sys.time() - s
s <- Sys.time()
modCOMM2 <- cv.glmnet(y=yCOMM2,x = mmCOMM, alpha=1,family='binomial',parallel=T, nfolds=30,type.measure='auc')
ModelMetrics::auc(yCOMM2, predict(modCOMM2, mmCOMM, type='response',s='lambda.1se'))
Sys.time() - s
###
# get preds
ptest1 <- predict(modCOMM1,mmCOMM,type='response',s='lambda.1se')
ptest2 <- predict(modCOMM2,mmCOMM, type='response',s='lambda.1se')
(p_actual = prop.table(table(yCOMM1))[2])
actual <- yCOMM1
results <- data.frame('actual'=actual, 'pred'=ptest1, 'pred_class' = ptest1>=p_actual)
names(results) <- c('actual','pred','prec_class')
table(results[,1], results[,3])
####
# View ROCS
source('./Modelling/UGRD/ROCPLOTTER.R')
par(mfrow=c(1,2))
ROCPLOTTER(yCOMM1, ptest1, 'ROC T1Y1 Commencing')
ROCPLOTTER(yCOMM2, ptest2, 'ROC T2Y2 Commencing')
par(mfrow=c(1,1))
dev.copy(png,paste0('./Modelling/UGRD/FitMetrics/COMM__rocplot_',Sys.Date(),'.png'))
dev.off();dev.new()



ArtsEducationandLawGroupW02.xlsx
Open with Excel 2016



ROC T1Y1 Commencing
ROC T2Y2 Commencing
Sensitivity
Specificity

Analytics server


...about those shades
Griffith UNIVERSITY
Office of Planning Services

Employability

Possibilities:

- Statistical Learning
- Machine Learning Predictions
- Cluster Segmentation
- Support operational intervention testing

GOS-L
graduate outcomes survey – longitudinal

What is it?

Measured by Employability Rate

Graduate Outcome Survey

$$\frac{\textit{Number of graduates } \textbf{\textit{in}} \textit{ Full Time Employment}}{\textit{Number of graduates in } \textbf{\textit{Available}} \textit{ for Full Time Employment}}$$

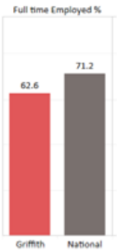
4 Months from Graduation Date

Factors

1) Graduates who are full time employed
2) Graduates who are available to be full time employed

Under the assumption:

comparative measure that represents graduate *population* Employability Rate



Levers

Results are influenced by a few things, some of which are

Population Employability Rates differ per FoE, and

Griffith FoE mix is different for both

1) Responders
2) Full Time Employed

These two things are not equal and have different impacts

Employment rate in QLD, Brisbane & Gold Coast Vs. other states and capital cities

Work readiness of our graduates

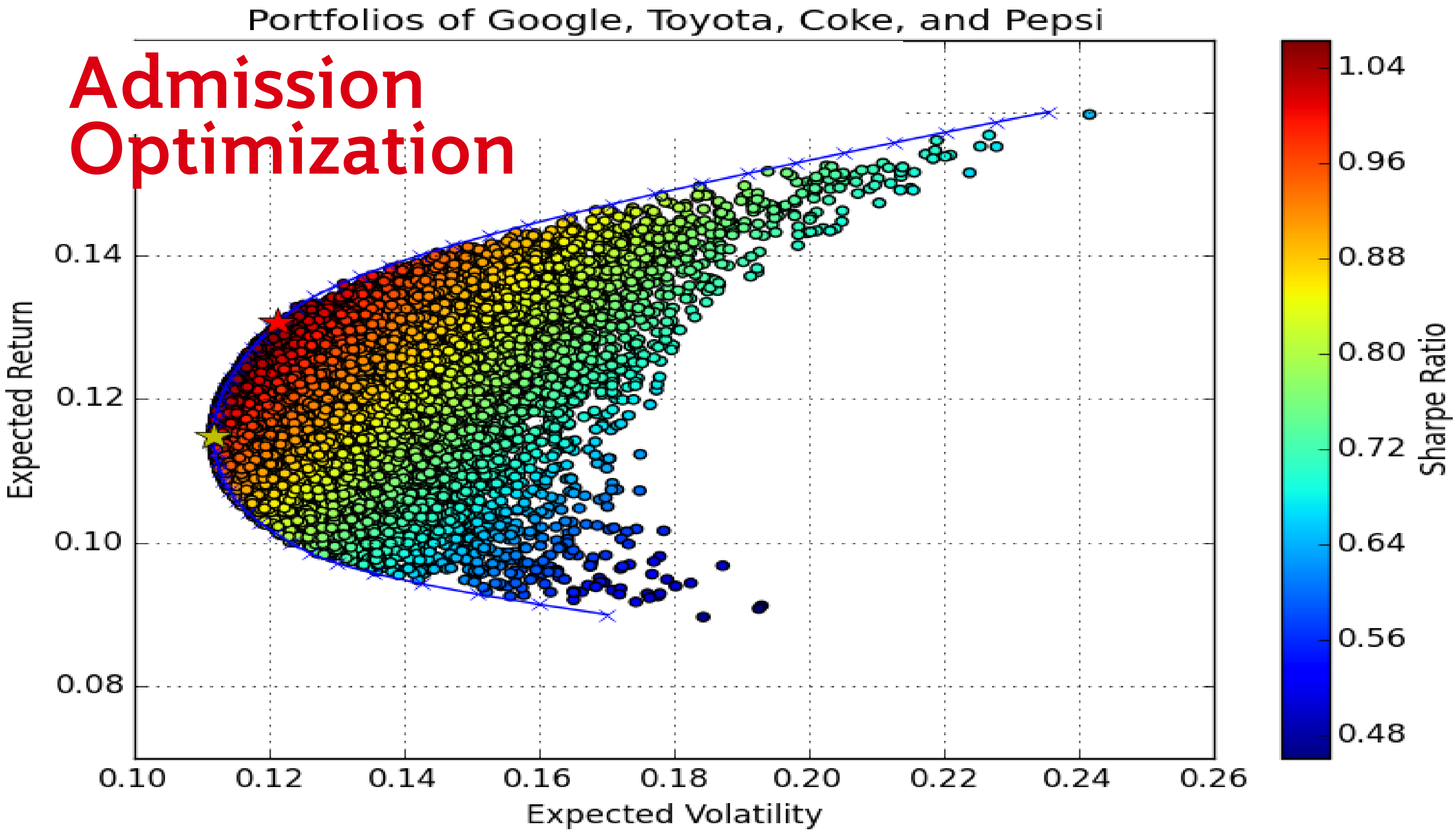
Griffith UNIVERSITY

Enterprise data visualisation

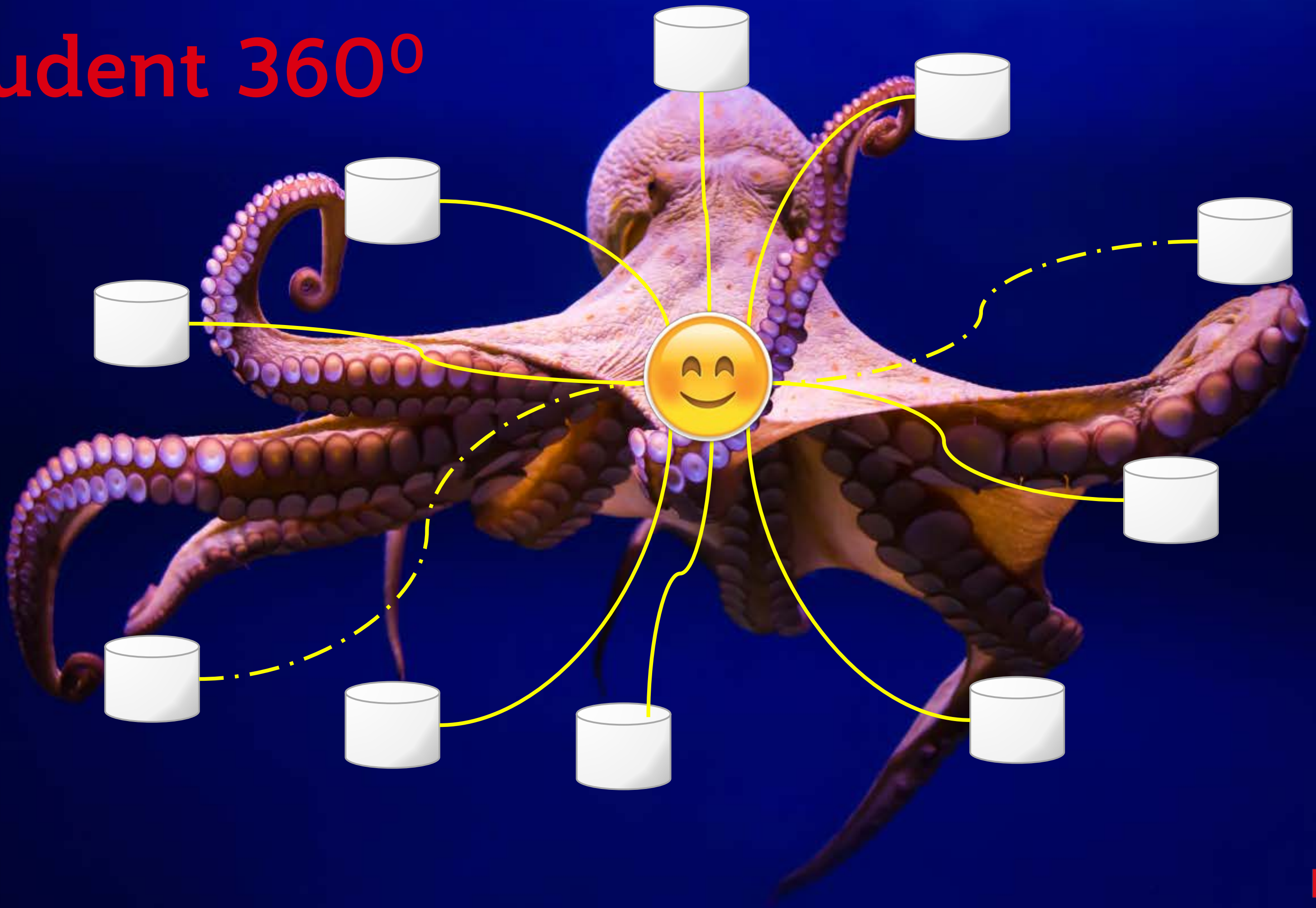
Griffith UNIVERSITY
Office of Planning Services

Data Lake

Admission Optimization



Student 360^0



Success measures

GPA prediction

PGRD Conversion Probability

RSCH Conversion Probability

Research Success Predictions:

- Publications,
- Time to Complete,
- Grant $’s

Finance

Student Lifetime Value

Research Project Expected Cost

HR Analytics

Promotability

Retention: Pre and Post Hire

Future Salary

Fraud

So much to do…

What is the path, past, present,future of analytics nirvana at griffith

Where Clancy takes us through the hopes, tears, fears and findings from the OPS Advanced Analytics team with examples of real life products, projects and future possibilities.

Clancy is an Applied Statistician/Data Scientist. He has experience in all data analytics paradigms and extensive data mining experience. He has qualifications in mathematics, statistics, machine learning and data science.

You'll get a taste of the practical every day application of advanced analytics @ Griffith as well as some dreamy stuff that we'd like to be doing in the future.

What is Data Science, What is Descriptive, Predictive & Prescriptive Analytics?

Is deep learning or big data on our dance card?

What about AI?

What are the questions at Griffith that Analytics can help answer, the big and the not so big?


QUESTIONS

THANK YOU

Griffith UNIVERSITY

Office of Planning Services

CLANCY BIRRELL

Advanced Analytics Manager, Office of Planning Services

+61 7 3735 8073 | clancy.birrell@griffith.edu.au