# Business Analytics With XLKitLearn: An Excel Frontend For scikit-learn

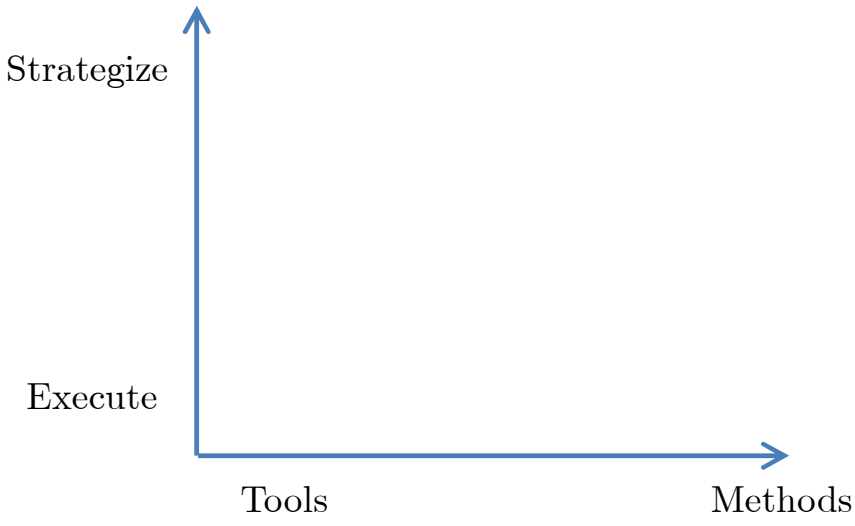
*Daniel Guetta*
Columbia Business School
guetta@gsb.columbia.edu

May 28th 2021



1

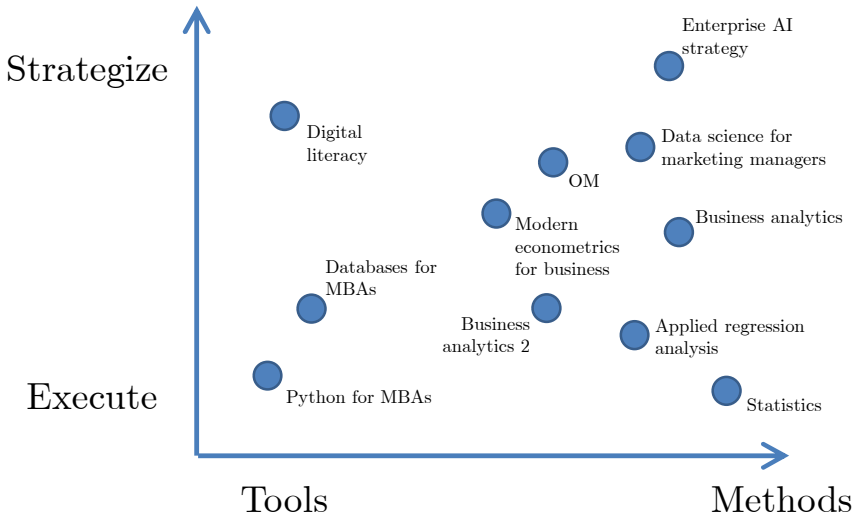
## Analytics for MBAs






2

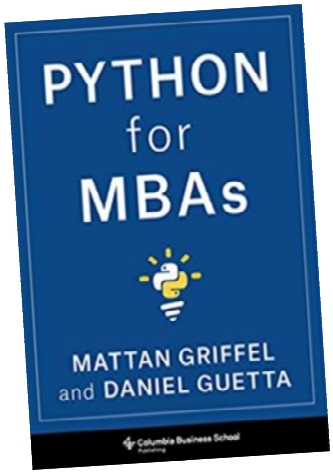
# Analytics for MBAs






3

# Analytics for MBAs

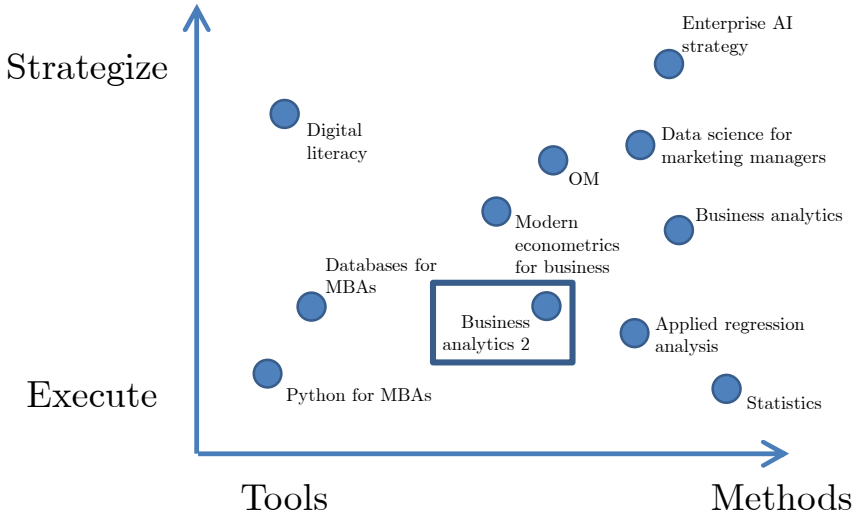



pythonformbas.com



4

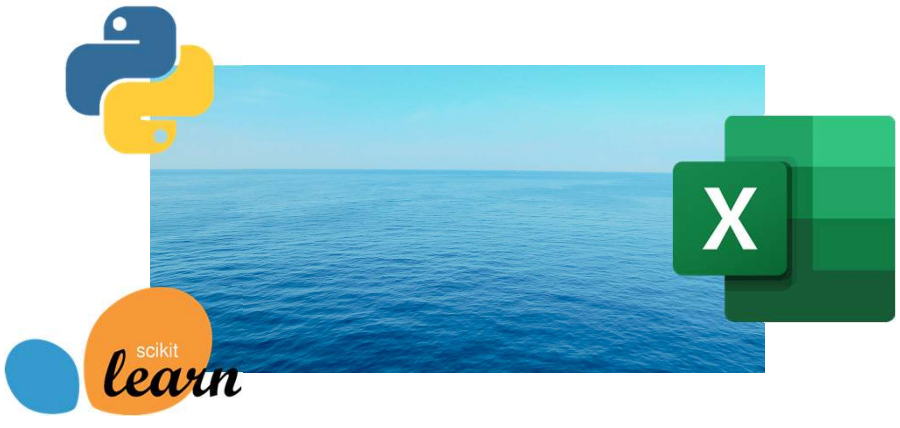
# Analytics for MBAs






5

# Business Analytics 2






6

## Business Analytics 2

Potential options

- Use one of many off-the-shelf software products?
  - Black boxes
- Use R/Python and give students code to copy/paste?
  - Poor experience
- Cover everything theoretically without hands-on work?
  - Can be difficult to absorb; dry



7

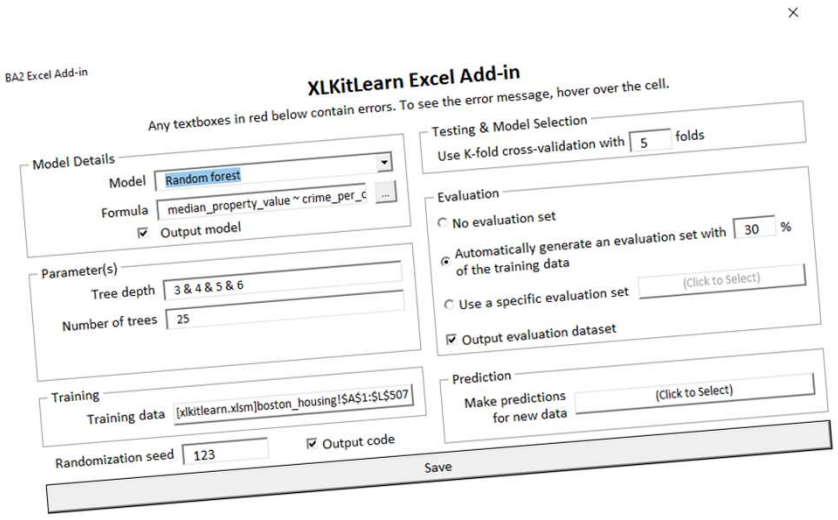
## Our Solution: XLKitLearn

danguetta.github.io/xlkitlearn



8

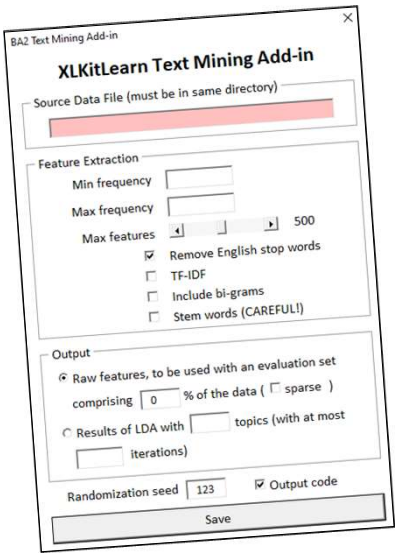
# Predictive analytics






9

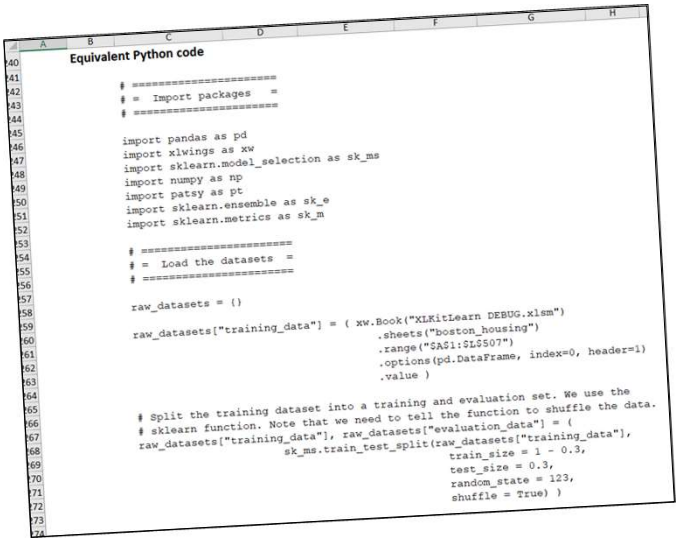
# Text analytics






10

# Each run generates Python



Equivalent Python code

```
# ======================
# =  Import packages   =
# ======================

import pandas as pd
import xlwings as xw
import sklearn.model_selection as sk_ms
import numpy as np
import patsy as pt
import sklearn.ensemble as sk_e
import sklearn.metrics as sk_m

# ======================
# =  Load the datasets  =
# ======================

raw_datasets = {}

raw_datasets["training_data"] = ( xw.Book("XLKitLearn DEBUG.xlsm")
                                    .sheets("boston_housing")
                                    .range("$A$1:$L$507")
                                    .options(pd.DataFrame, index=0, header=1)
                                    .value )


# Split the training dataset into a training and evaluation set. We use the
# sklearn function. Note that we need to tell the function to shuffle the data.
raw_datasets["training_data"], raw_datasets["evaluation_data"] = (
                        sk_ms.train_test_split(raw_datasets["training_data"],
                                               train_size = 1 - 0.3,
                                               test_size = 0.3,
                                               random_state = 123,
                                               shuffle = True) )
```