Practical Excel automation with pandas and xlwings

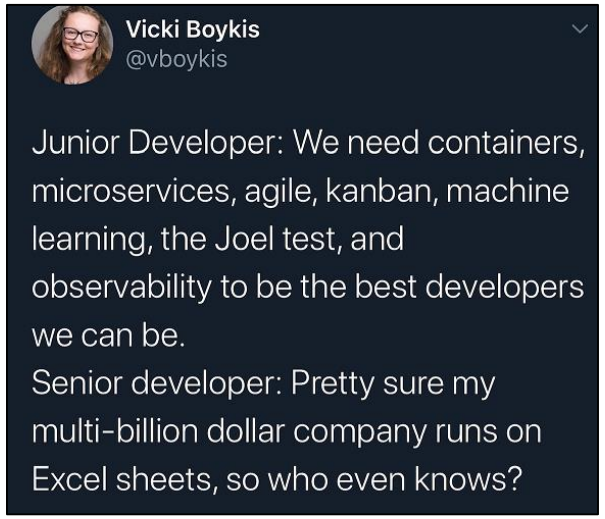
Chicago Python Data SIG presents Python Powered Healthcare, 02/19/2020

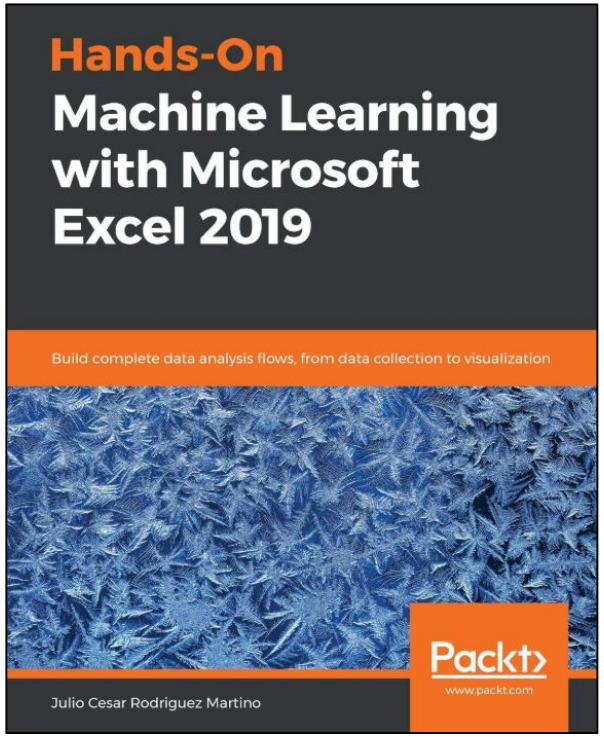
by Paul Zuradzki

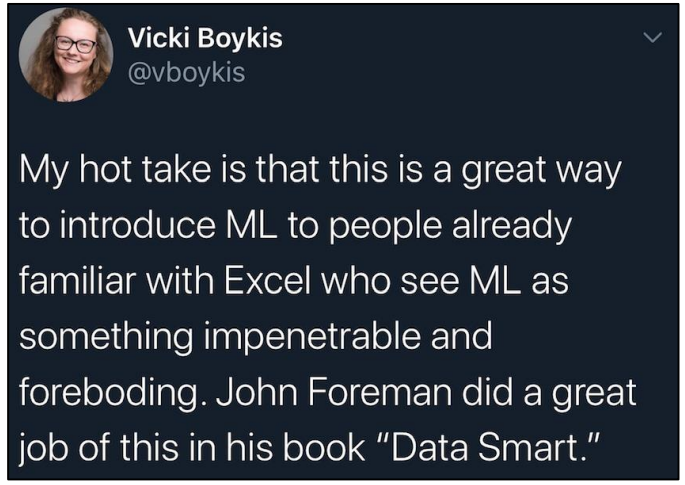
@paulzuradzki
paulzuradzki.com

Excel makes the world go round

Accessible delivery method. Sometimes simple is good. Rapid prototyping.







Meme Citation
https://github.com/chris1610/pbpython/blob/master/presentations/Escaping-Excel-Hell-with-Python-and-Pandas.pdf

General use cases

- Bulk merging from many sheets or workbooks
- Bulk splitting to many sheets or workbooks
- Bulk editing and styling to replace VBA tasks
- Static dashboards
- And more…

Health data examples

- Medical record review audits – sample member encounters by provider and evaluate chart elements with pass/fail metric
- Static dashboards and prototypes for quality measures/HEDIS
- Splitting up static provider reports
- Processing Excel source files from CMS or other third parties
  - E.g., for MSSP/NextGen, CMS sends out formatted enrollment data with columns that must be carefully parsed

Example: Medical Record Review Audit

Reviewers/nurses inspect electronic medical record to check for consistency with claims and confirm quality measures that may not be reflected in claims (ex: BMI assessments)

Group ID XXXXXXXXXXXXX
Group Name Doctor Strange LLC

	Review Date	Member ID	Member Last Name, First Name	D.O.B.	Age	Chart Type	DOS	DOS	DOS	1. Significant illness	2. Physical exam	3. Allergies	4. Past medical History	5. Diagnosis consistent with findings	6. Treatment consistent with diagnosis	7. Notation of follow up care
1	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	7/31/2019	xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx	Yes	Yes	Yes	Yes	Yes	Yes	Yes
9		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
10		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
11		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
12		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
13		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
14		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
15		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
16		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
17		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
18		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
19		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
20		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
21		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
22		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
23		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
24		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
25		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
26		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
27		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
28		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
29		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
30		xxxxxxx	xxxxxxx	xxxxxxx	xxxxxxx	ADULT	xxxxxxx	xxxxxxx	xxxxxxx							
						Total Number Of Elements Evaluated				8	8	8	8	8	8	8
						Patient Files (Yes)				8	8	8	8	8	8	8
						Patient Files (No)				0	0	0	0	0	0	0
						Patient Sample Size (Yes+No)				8	8	8	8	8	8	8
						% of Patients (Yes/Sample)				100%	100%	100%	100%	100%	100%	100%

Example: Medical Record Review Audit, continued

Providers are scored by chart element. Low-scorers get re-audited.

Dr Strange LLC
Practitioner Performance Evaluation

CY2020/MY2019

	Group ID:	**XXXXXXXXXXXXX**
	Group Name:	**Doctor Strange LLC**
	Address:	**400 S State St, Chicago, IL 60605**
	Type of medical record:	**Paper** **Electronic (name of system)**
	Chart Elements	
100.00%	1. Significant illness and medical conditions are indicated on the problem list.	
100.00%	2. History and Physical exam for presenting complaints containing relevant psychological and social conditions affecting the patient's medical/behavioral health, including mental health, and substance abuse status are documented.	
100.00%	3. Medication allergies and adverse reactions are prominently noted in the medical record or notation of NKA or NKDA.	
100.00%	4. Past medical history is easily identified and includes serious accidents, operations, and illnesses. For children and adolescents (18 years and younger), past medical history relates to prenatal care, birth, operations, and childhood illnesses.	
100.00%	5. Working diagnoses are consistent with clinical findings.	
100.00%	6. Treatment plans are consistent with diagnoses.	
100.00%	7. Encounter forms or notes have a notation, when indicated, regarding follow-up care calls or visits. Time of return is noted in weeks, months, or as needed.	
100.00%	8. Unresolved problems from previous office visits, referrals, diagnostic testing, and status of preventive screenings are addressed in subsequent visits.	
	9. Consent forms are maintained in the medical record.	
100.00%	10. If a consultation is requested, there is a note from the consultant in the record.	
100.00%	11. Consults, labs, and imaging reports are reviewed and initialed by the ordering physician. Consultation and abnormal laboratory and imaging study results have an explicit notation in the record of follow up plans.	
100.00%	12. Name and telephone numbers of emergency contact are documented.	
37.50%	13. An immunization record for children AGES 18 AND YOUNGER is up to date or an appropriate history has been made in the medical record (for adults).	
100.00%	14. Documentation of discharge summaries.	
100.00%	15. Evidence of member/patient identification information, on each page of the medical record.	
100.00%	16. One-time identification of personal/biographical data, including date of birth, age, gender, marital status, home and work telephone number, home and work address, mailing address (if different from home), and school/employer.	

Should this be in a "BI" tool or web app?

Ain't nobody got time for that

Split data pseudocode

from national aviation[1] stats to state-specific tabs

1) Read data set into dataframe
2) For each GroupBy variable (state)
    3) Subset by group variable (state)
    4) Create new sheet for group variable
    5) Transfer subset to associate sheet

[1] Unfortunately, it's hard to give a public talk on health data with actual health data due to privacy regulations
[2] To simulate de-identified health data, see CMS Medicare Claims Synthetic Public Use Files (SynPUFs)

Other tools

- xlwings
- openpyxl
- xlsxwriter
- pandas (pd.to_excel, pd.from_excel)
- gspread for Google Sheets
- VBA
  - I think it’s encouraging for folks to know that if they invest in learning Python, they can replace or enhance much of their VBA workflows

Parting thoughts

- G Sheets are awesome but Excel prevalence > G Sheets
  - Don’t put any PHI on un-protected cloud anyway
- xlwings requires Excel to be installed
- openpyxl works directly with the file and the XML data in the Excel file, so openpyxl may be preferred if deploying on a light server (ex: processing data files)
- xlwings opens an application instance. Manage resources carefully (setup/teardown). Not super fast – but better than humans copy/pasting.
- xlwings vs openpyxl: with xlwings, I like that you can avoid the “for row in rows, for col in cols” double nested loop pattern when all you want to do is copy a data set. It’s a more natural copy/paste-like syntax with dataframes.

```
for row in sheet['A1':'C3']:
        for cell in row:
                print(cell.coordinate, cell.value)
```

```
sht.range('A1').value = df
```