Python-Powered Analytics in Excel: Leveling Up Skills for Management Scholars and Educators

If we teach today as we taught yesterday, we rob our children of tomorrow.

– John Dewey

The purpose of this workshop is fourfold: introduce, demonstrate, encourage participation, and review the newly released Python add-in via Excel feature. This update occurred toward the end of August 2023 and is accessible for individuals willing to use a beta version of Excel if they have Office 365. The relevance and significance of this professional development workshop (PDW) is to bring awareness of new analytic capabilities within the Excel spreadsheet software, which are likely to come on the market in the next year. In designing this workshop, SWAM participants do not need to be proficient in Python. Individuals with no experience and experts alike can benefit from attending. variables. The presenter's GitHub repo will provide hypothetical data (with Python syntax) for attendees to practice along with additional files (PowerPoint slides, instructional documents). A screenshot of the Python interface is viewable in the Appendix of this proposal.

Description of the Intended Audience

The intended audience for this workshop is management scholars (students and faculty) who desire to learn about new software tools that can aid and enhance their analytic workflow. Instructors and teaching professors should find the applied example analyses useful as Python will be used to perform analytics on salary data, performance, and team dynamic visualizations, all as part of integration within Excel. Instructors are expected to gain fresh insights to evaluate and model their student’s performance after attending this workshop. Furthermore, considering utility in research and/or teaching capacities, attendees will gain insights on how to clean/merge data and build models currently not possible in Excel (logistic regression, KNN, etc).

Proposed Session Format and Structure

A 1.5-hour workshop is requested to cover the material in this PDW. However, there is flexibility to adjust the session length if a shorter time is necessary for the SWAM program. Ideally, the room should be set up for attendees to use their computers, much like a classroom or boardroom. This workshop will use a combination of PowerPoint slides and the presenter sharing his screen during demonstrations. Hypothetical data for training purposes will be used so attendees can perform the analyses. Because this PDW is designed to be engaging by using demonstrations, a projector will be necessary for the presenter. The format and structure are detailed below:

Timetable & Topic

Part 1: Introduction and Getting Started	
5 minutes	Introduction
8– 10 minutes	Python Language & IDEs, a Refresher
8 – 10 minutes	Opting into the Excel Beta Version and Setup
Part 2: Python vs. Excel in Advanced Analytics	
10 – 15 minutes	Advantages of Python over Excel
5 – 8 minutes	Python Integration in Excel – Benefits & Downsides
8 – 10 minutes	Python Libraries
8 – 10 minutes	Break
Part 3: Hands-on Python Practice in Excel for Management	
5 – 10 minutes	Example 1: Employee Salary Analysis (pandas, matplotlib)
10 – 12 minutes	Example 2: Analyzing Student Performance (pandas, seaborn)
10 – 12 minutes	Example 3: Team Dynamics Visualization (networkx)
Part 4: Libraries, Applications, and Closing	
4 – 7 minutes	Applications for Management
8 – 10 minutes	Should you try it?

Technology and software are required. Participants need to bring a laptop computer (Windows or Mac operating system) with Wi-Fi capabilities. This PDW will use Excel which has Python integrated from a beta version. Instructions will be demonstrated on how to get the beta version. Also, CSV files for "hands-on" analysis will be provided. The presenter will provide details to his GitHub account to access the files. At a minimum, attendees should have a Google Colab account, even if they have never used the IDE. Being this is cloud-based, it will minimize installing code locally on computers. I will use this IDE to demonstrate and draw the connection to how the code is used within Excel. Below, I created YouTube videos on establishing a Colab account. Ideally, I would like to make these three videos I made to be available before the workshop for individuals who are not familiar with Python and Colab. If this workshop is approved, I will make an additional video on how to get the beta version of Excel and use syntax in the PowerShell prompt to initiate the integration manually.

Setting Up & and Using Google Colab Videos

Video 1: https://bit.ly/drkeeler-colab1
Video 2: https://bit.ly/drkeeler-colab2
Video 3: https://bit.ly/drkeeler-colab3

Attendee Value

This workshop is anticipated to benefit a wide range of Federation of Business Discipline attendees, especially individuals in management, decision science, and analytics. Given the prevalent usage of Excel as a spreadsheet, educating attendees on the integrated analytic capabilities of Python will be novel.

The first part of this workshop is to set the stage for the integration of Python into the Excel concept. Beyond giving an overview of the workshop, a handout and demonstration on how to opt into the beta version of Excel will be provided. Furthermore, a guided setup via PowerShell

to be shown to manually initiate the beta version if the primary method is problematic. The first part will also have a brief discussion of the idea behind object orientation and the concept of data frames, data types, loops, conditions, and functions. This will be done in a way that allows attendees to have a frame of reference with Excel (e.g. conditions, functions, etc.).

The second part of the workshop will highlight some accessible analytic techniques that are currently unavailable in the Data Analysis module of Excel. This part's theme could be considered management research analyses in Python. Logistic regression, training/testing validation, and ICC grouping analytics will be covered. There will be a demonstration of practical applications, especially in HR and OB research. This component will conclude with a rundown of the benefits and downsides of using the Python add-in. Importantly, attendees need to understand the software is being run on Microsoft servers in the cloud; there exist potential issues to consider (data transmission breaches, the necessity of Internet service, etc.).

The third part of the workshop will allow attendees to practice with three different data samples, assuming they have the beta version installed. The presenter will provide access to sample data from his GitHub account at https://github.com/justinbkeeler/swam2024_pdw. Exercise #1 will involve an employee salary scenario where the pandas and matplotlib libraries will be used to visualize and analyze employee salary data in relation to role, department, experience, and performance. The objective is to identify potential inequities in pay that need to be addressed by HR. Exercise #2 will encompass fictitious student data (created by the presenter) to evaluate performance and generate descriptive summaries, correlation, and data visualizations using both heatmaps and scatter plots. Lastly, Exercise #3 will entail using the “network” library to perform network analysis. Centrality analysis and network visualization will also be performed to help attendees understand how potential departments and individuals within a

company are interconnected. Each of these demonstrations are expected to take 5-8 minutes using the Python code. Importantly, the purpose is to show the capabilities of Python from within Excel that would not usually be possible.

The final part of the workshop would involve a discussion of preloaded libraries (matplotlib, numpy, pandas, seaborn, and statsmodels) as well as currently importable libraries (beautifulsoup4, ipthon, genism, scikit-learn, etc.). High-level real-world applications, samples provided by Microsoft, and time for Q&A will conclude this part of the workshop.

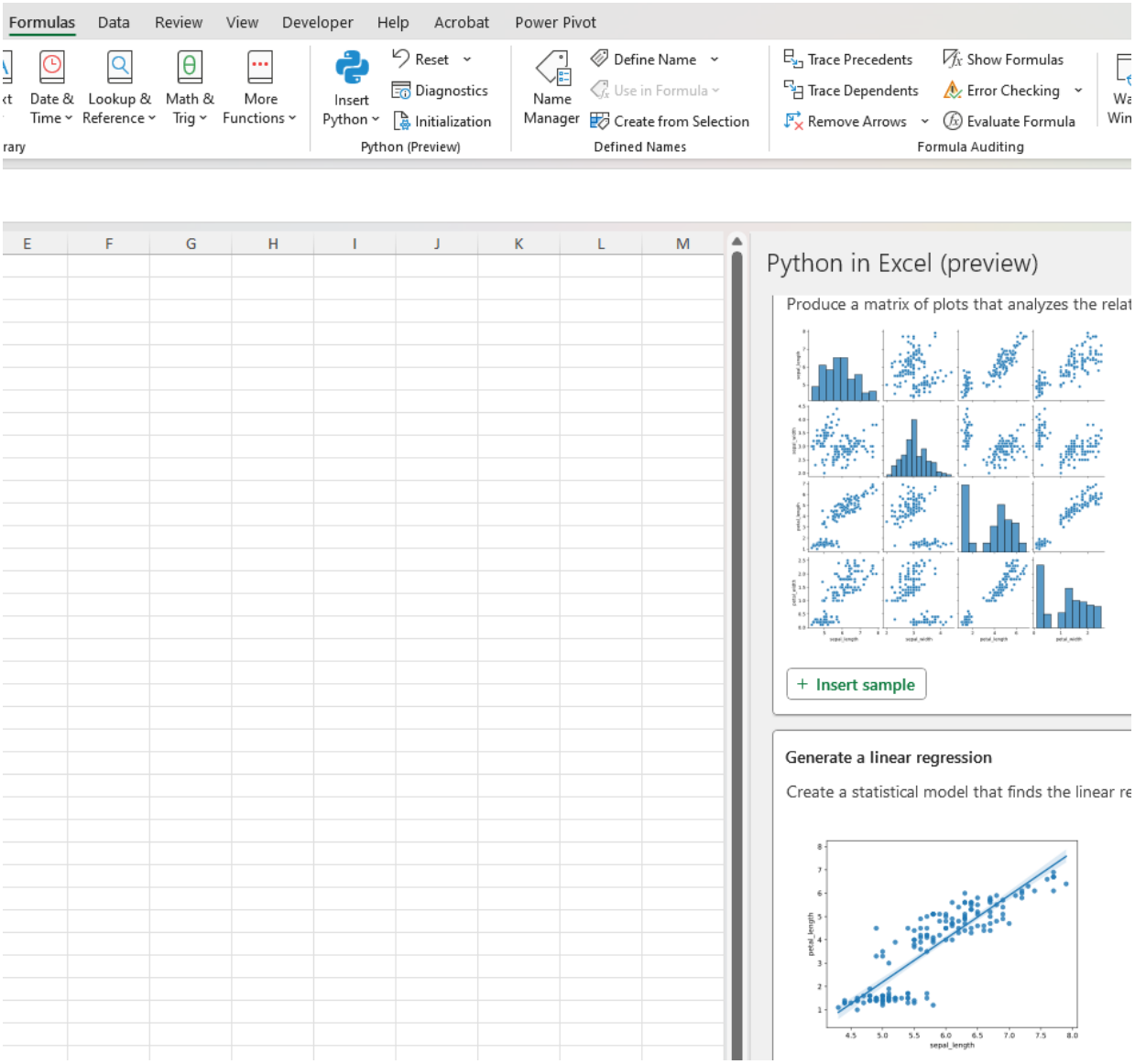
Benefit Highlights

- Raise awareness of multifunctionality in analytics that is currently possible with the new Python integration into Excel.
- Give a high-level overview of the Python foundation and what is meant by object-oriented programming.
- Discuss the limitations of using a cloud-based version of Python.
- Provide an opportunity for attendees to practice and try three different analytic examples to represent research and personal use.

About the Presenter

The presenter has taught statistics and analytics via different programming languages for over eight years. Over the last two years, he has taught approximately 300 graduate students in an Introduction to Python for Business course at a D1/R2 university, which he personally developed. He has also taught and developed an Introduction to Analytics using R. Recently, an institution on the East Coast contracted with him as a subject matter expert to advise and develop a course called, "Decision Making and Analytic Techniques for Non-Profits." This course uses Python programming for graduate students without any prior programming experience. The presenter is well-versed in data visualization and data storytelling, too. He currently has research under review in the *Journal of Organizational Effectiveness: People and Performance*. Beyond R and Python, the presenter is skilled in Tableau, Alteryx, SPSS, AMOS, and Mplus for an array of analytics as well as techniques related to machine learning and neural networks.

Appendix


Formulas
Data
Review
View
Developer
Help
Acrobat
Power Pivot
Date & Time
Lookup & Reference
Math & Trig
More Functions
Insert Python
Reset
Diagnostics
Initialization
Python (Preview)
Name Manager
Define Name
Use in Formula
Create from Selection
Defined Names
Trace Precedents
Trace Dependents
Remove Arrows
Show Formulas
Error Checking
Evaluate Formula
Formula Auditing
E
F
G
H
I
J
K
L
M
Python in Excel (preview)
Produce a matrix of plots that analyzes the relat
+ Insert sample
Generate a linear regression
Create a statistical model that finds the linear re
petal_length
sepal_length