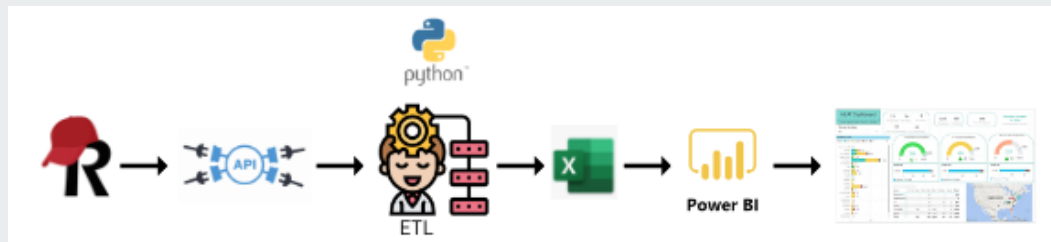


# REDCap Interesting Use Case

**Efficient Data Workflow Automation:  
Integrating REDCap API, Python, and PowerBI for  
Seamless & Real-time Multicenter Clinical Trials  
Reporting**



Ally Qi, MPH  
Clinical Trials Operations Manager  
Vanderbilt Coordinating Center (VCC)  
REDCap Fest  
2/21/2024



**Question:**



**How much time do you spend on  
creating or updating PPT slides per  
day/week/month?**

# Framing Eighteen coils in cerebral Aneurysms Trial (*FEAT*)

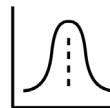
stryker



Multi-center, prospective, randomized trial comparing the utilization of 0.014-0.0155" coils (**18 coils**) vs. smaller diameter coils (**standard 10 coils**) in mid-sized aneurysm treatment



**651** participants enrolled across **25** sites in U.S.

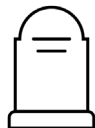


Random **1:1** assignment



**Primary Outcome:**  
Angiographic Occlusion rate on follow-up angiography at 12-18 months

## Secondary Outcomes



Treatment **related morbidity and mortality**, as measured by the NIH Stroke Scale



**Packing density** as measured by volumetric filling of the aneurysm



**Modified Rankin Scale** at 3-6- and 12-18-months post-coiling



**Re-hemorrhage and re-treatment rates**

# Challenges - What if?



Meetings rescheduled?  
1x, 2x? Per week/month?



Cross-database data  
analysis



Special metrics of  
interest by PI



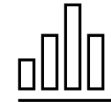
Site PI contact  
info



Accurate/up-to-date  
count of participant status



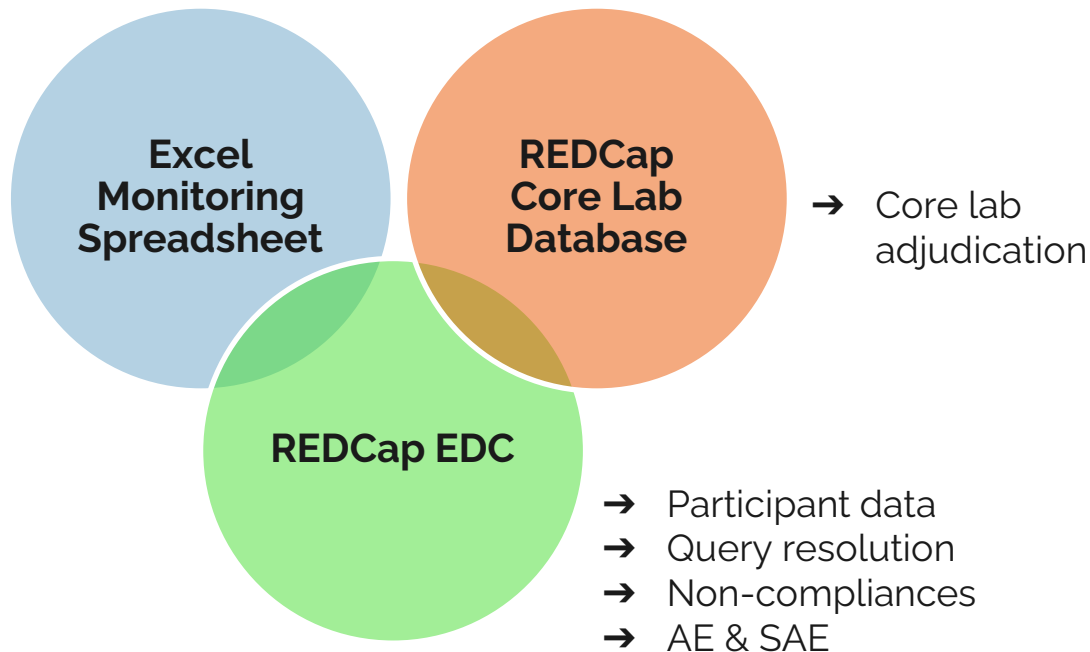
Unexpected questions to  
drill down into different  
categories or levels of data



Multi-site trial  
performance (PD & SAE &  
Query Resolution &  
Inventory)

# Cross-database Data Wrangling

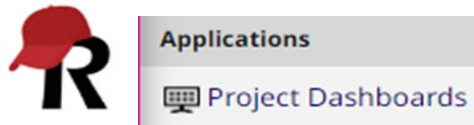
- Participant status
- Follow-up window
- Enrollment dates
- Monitoring status per instrument level
- Site contact info



# What're my options of data visual tools?



PowerPoint slides



REDCap Dashboard



Google Data Studio



**PowerBI** or Tableau

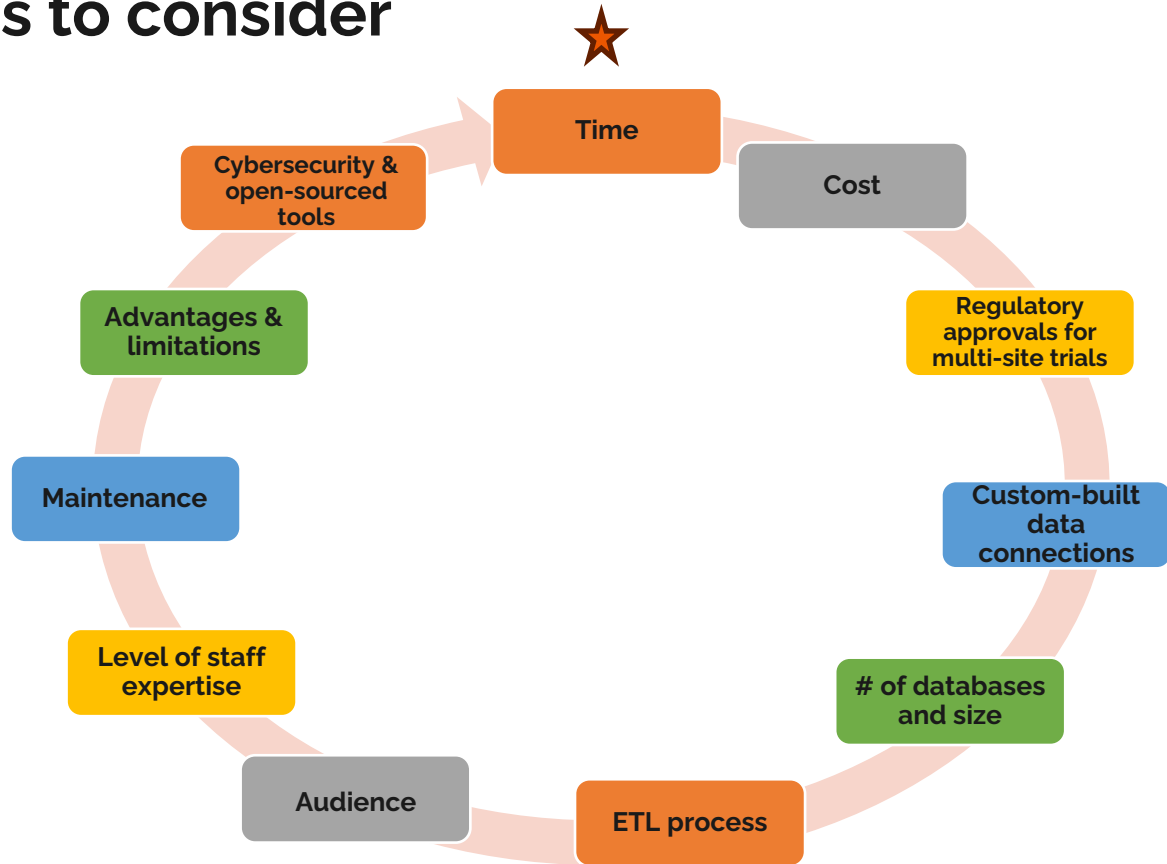


R & R Shiny

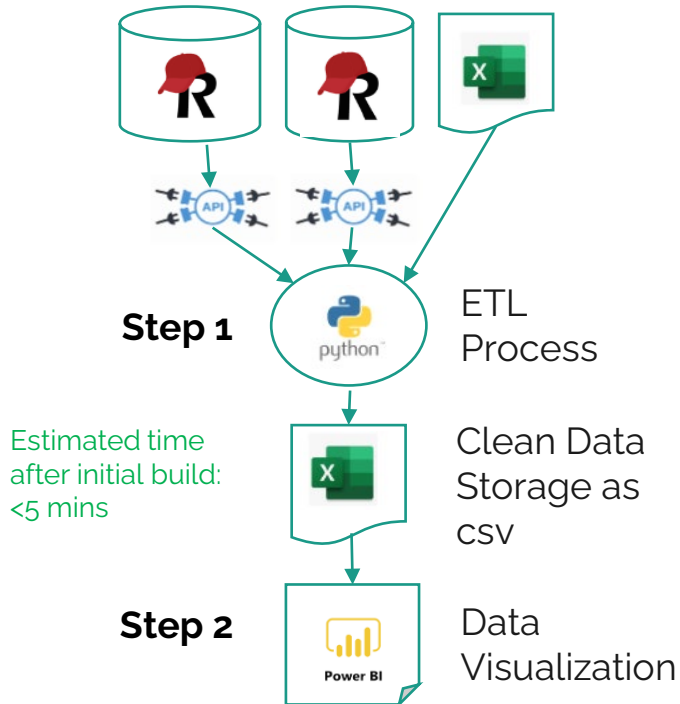


JavaScript & HTML & CSS

# Factors to consider



# Solution! A Seamless Data Pipeline



## Initial build:

1. In Python Jupyter Notebook, import data from 2 REDCap databases using API and 1 SharePoint Excel sheet to complete the ETL process. Then export clean data to flat csv files in a dedicated folder.
2. In PowerBI, build a dashboard to display key study metrics.

## Benefits:

1. Free (limited version)\*
2. Fast building time
3. Fastest loading time, closest to real time
4. Full automation
5. Reproducible & replicable
6. Secure
7. No regulatory approval needed
8. Low maintenance (debugging)



# Key Metrics Identification



- **Data Safety Monitoring Plan (DSMP) and Study Protocol**
- Defines what can and can't be shared and protects blinding



- **Study Type**
- Investigational product(s): drugs or medical devices or interventions or others



- **Number of Time Points**
- Longitudinal vs single time point



- **Database structure**
- IP Arms (single vs multiple) and phases (start up, maintenance, and site closure)



- **Number of Sites**
- Single site vs Multi-sites

# Key Metrics Identification



Enrollment:  
# at each time point,  
change over time

Participant Status  
Incomplete: lost to  
follow-up, withdrawn,  
death, etc.

# of participants in  
window under each  
time point

Inclusion & Exclusion  
Criteria

Primary & Secondary  
Outcomes  
Blinded vs open-  
labeled?

Safety Events:  
AE & SAE

Non-compliances

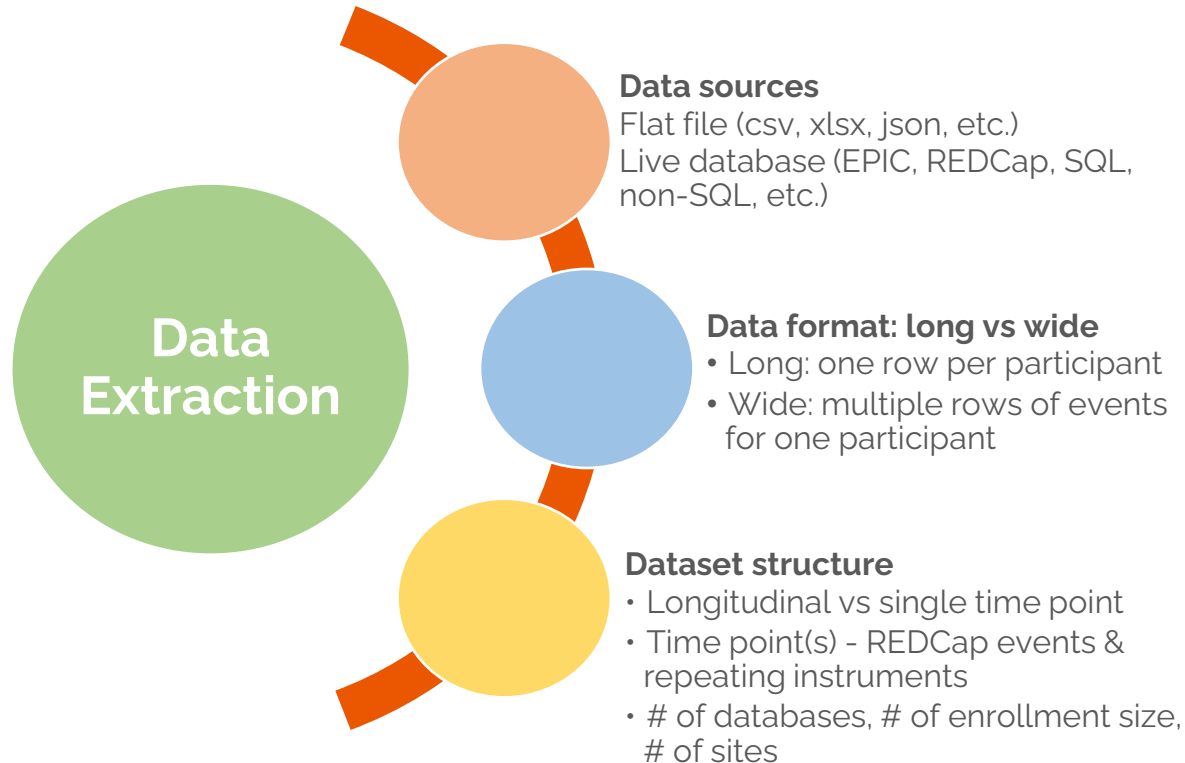
REDCap Queries (open  
vs closed)  
Median days to full  
resolution

Inventory Tracking:  
drugs, medical  
devices, imaging

Site Activation Dates &  
Enrollment Speed


Special metrics of  
interest - PI


# Data Extraction (ETL)




# Basic REDCap Report Export


## Choose export format


☒  CSV / Microsoft Excel (raw data)

☐  CSV / Microsoft Excel (labels)

☐  SPSS Statistical Software

☐  SAS Statistical Software

☐  R Statistical Software

☐  Stata Statistical Software

☐  CDISC ODM (XML)

## De-identification options (optional)

The options below allow you to limit the amount of sensitive information that you are exporting out of the project. Check all that apply.

### Known Identifiers:

- ☐ Remove All Identifier Fields (tagged in Data Dictionary)
- ☐ Hash the Record ID field (converts record name to an unrecognizable value)

### Free-form text:

- ☐ Remove unvalidated Text fields (i.e. Text fields other than dates, numbers, etc.)
- ☐ Remove Notes/Essay box fields

### Date and datetime fields:

- ☐ Remove all date and datetime fields
- OR —
- ☐ Shift all dates by value between 0 and 364 days (shifted amount determined by algorithm for each record)  
[What is date shifting?](#)

[Deselect all options](#)

## Advanced data formatting options

### Export blank values for gray Form Status?

All Form Status fields with a gray status icon can be exported either as a blank value or as "0" (Incomplete). Hint: Blank values are recommended if the data will be imported back into REDCap, in which this preserves the gray status icons for all the imported records.

Export gray Form Status fields with value of "0" ▼

### Set CSV delimiter character

Set the delimiter used to separate values in the CSV data file (only valid for CSV Raw Data and CSV Labels export formats):

, (comma) - default ▼

### Force all numbers into a specified decimal format?

You may choose to force all data values containing a decimal to have a specified decimal character (comma or period/full stop). This will be applied to all calculations and number-validated text values in the export file.

Use fields' native decimal format (default) ▼

**NOTE:** Your data formatting selections above will be remembered in the future and will be pre-selected upon your next export.

# Intro to API (Application Programming Interface)



Your API token for project [redacted]

The API token below is **ONLY** for you and will work **ONLY** with this project. This token allows special access to REDCap data and **should NOT be shared with others**. If you think your token has been compromised, then please contact your REDCap administrator immediately **AND** either delete or regenerate your token by using the buttons below.

API Token: [redacted] 

Finished using the API for this project? If so, please delete your token for security reasons.

Think someone else knows your token? If so, please regenerate your token for security reasons.

APIs (Application Programming Interfaces) are used to:

- provide a standardized way for data communication
- act as a bridge between different systems and allow them to exchange data seamlessly



# REDCap API

**Applications**

- Project Dashboards
- Alerts & Notifications
- Multi-Language Management
- Data Exports, Reports, and Stats
- Data Import Tool
- Data Comparison Tool
- Logging and Email Logging
- File Repository
- User Rights and DAGs
- Customize & Manage Locking/E-signatures
- Randomization
- Data Quality and Resolve Issues
- API and API Playground**
- REDCap Mobile App

## API Playground

The API Playground is an interface that allows experimentation with the REDCap API without actually writing any code. You can explore all the different API methods and their various options to customize a given API request. You may even execute a real API request and see the exact response that REDCap returns from the request. If you are interested in creating an API script, the section at the bottom will provide code samples for various programming languages to give you a head start. For details on the capabilities of the REDCap API and how to use it, please see the [REDCap API documentation](#).

Select an API method from the drop-down list below, after which it will load any other options that are specific to that method.

API Method:

Production - Some APIs have been disabled for the playground because they affect data

Format:

Report ID:

Raw Labels:

Raw Headers:

Checkbox Labels:

Set CSV delimiter character:

Errors:

### Response

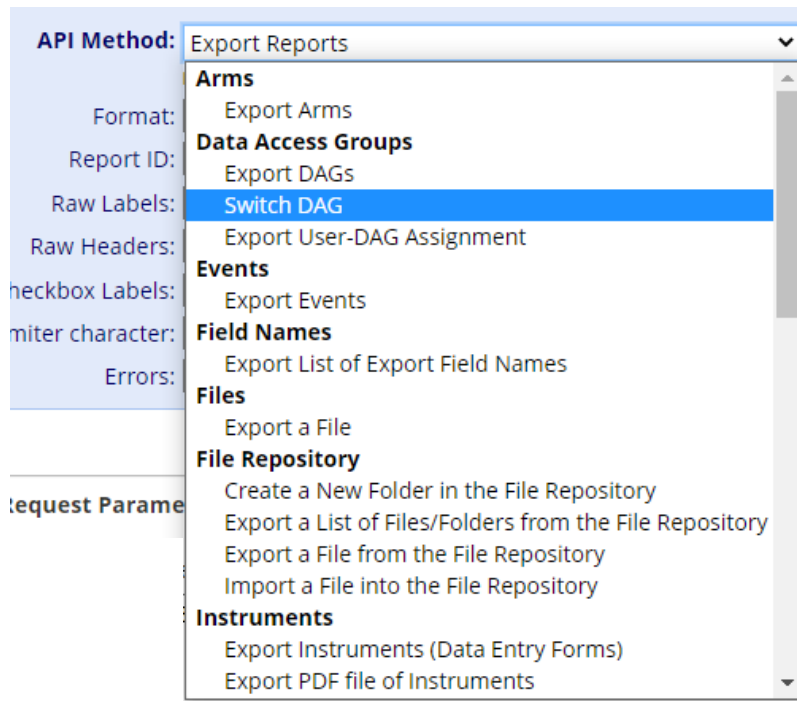
Click the Execute Request button to execute a real API request, and it will display the API response in a text box below.

Displayed in the box below is the code you would use to execute this API request in the selected programming language.

```
#!/usr/bin/env python
import requests
data = {
    'token': ' ',
    'content': 'report',
    'format': 'csv',
    'report_id': '330456',
    'csvDelimiter': '',
    'rawOrLabel': 'label',
    'rawOrLabelHeaders': 'label',
    'exportCheckboxLabel': 'true',
```



# What can I export from REDCap using API?



The screenshot shows the 'API Method:' dropdown menu in the REDCap interface. The menu is open, displaying a list of export options. The 'Switch DAG' option is highlighted in blue. The options are organized into categories: Arms, Data Access Groups, Events, Field Names, Files, File Repository, and Instruments. The 'Format:' field is set to 'Format:'. The 'Report ID:' field is empty. The 'Raw Labels:' field is empty. The 'Raw Headers:' field is empty. The 'checkbox Labels:' field is empty. The 'delimiter character:' field is empty. The 'Errors:' field is empty. The 'Request Parameters' field is empty.

**API Method:** Export Reports ▼

- Arms**
  - Export Arms
- Data Access Groups**
  - Export DAGs
  - Switch DAG**
  - Export User-DAG Assignment
- Events**
  - Export Events
- Field Names**
  - Export List of Export Field Names
- Files**
  - Export a File
- File Repository**
  - Create a New Folder in the File Repository
  - Export a List of Files/Folders from the File Repository
  - Export a File from the File Repository
  - Import a File into the File Repository
- Instruments**
  - Export Instruments (Data Entry Forms)
  - Export PDF file of Instruments

- Arms
- Data Access Groups
- Events & Field Names
- Files & File Repository
- Instruments
- Logging
- Metadata(data dictionary)
- Projects
- Records
- Repeating Instruments and Events
- Reports
- REDCap Version
- Survey
- Users & User Privileges & Roles



# Intro to Python Pandas

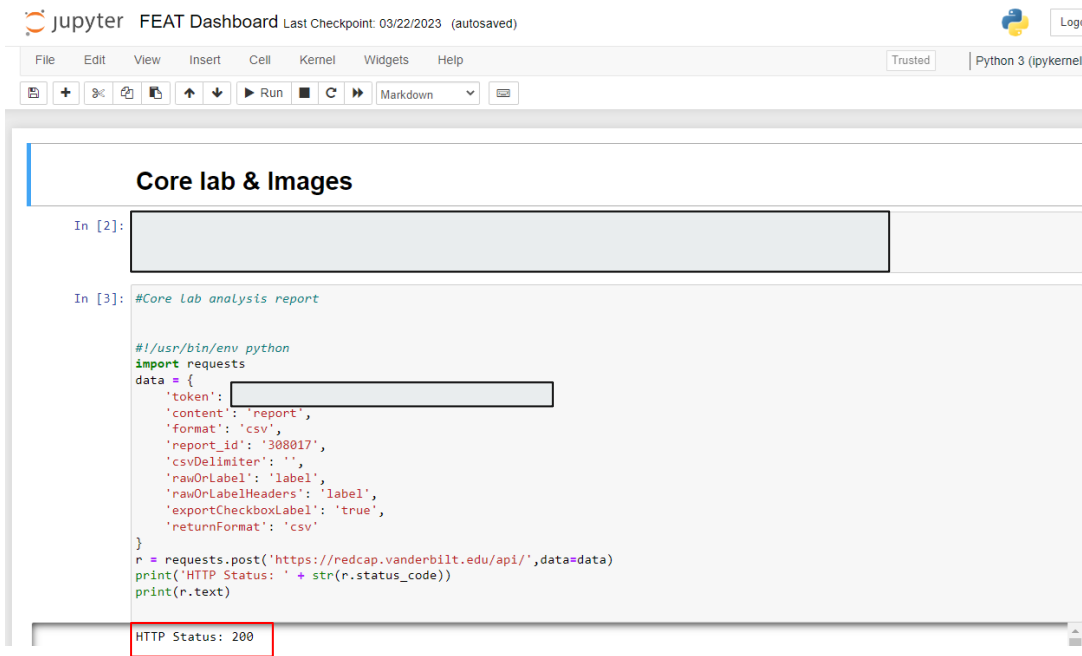


Python Pandas:

- a powerful open-source data analysis and manipulation library
- provides easy-to-use data structures and data analysis tools
- provides a fast and efficient way to work with structured data
- has its intuitive and flexible API



# Data Transformation (ETL)



The screenshot shows a Jupyter Notebook titled "FEAT Dashboard" with a last checkpoint of "03/22/2023 (autosaved)". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, running, and markdown. The notebook content is divided into two cells. The first cell, labeled "In [2]:", contains a redacted area. The second cell, labeled "In [3]:", contains a Python script for a REST API call. The script imports the 'requests' library and defines a data dictionary with fields like 'token', 'content', 'format', 'report\_id', 'csvDelimiter', 'rawOrLabel', 'rawOrLabelHeaders', 'exportCheckboxLabel', and 'returnFormat'. It then makes a POST request to 'https://redcap.vanderbilt.edu/api/' and prints the HTTP status and response text. The output of the second cell is "HTTP Status: 200", which is highlighted with a red box.

```
Core lab & Images

In [2]: [REDACTED]

In [3]: #Core Lab analysis report

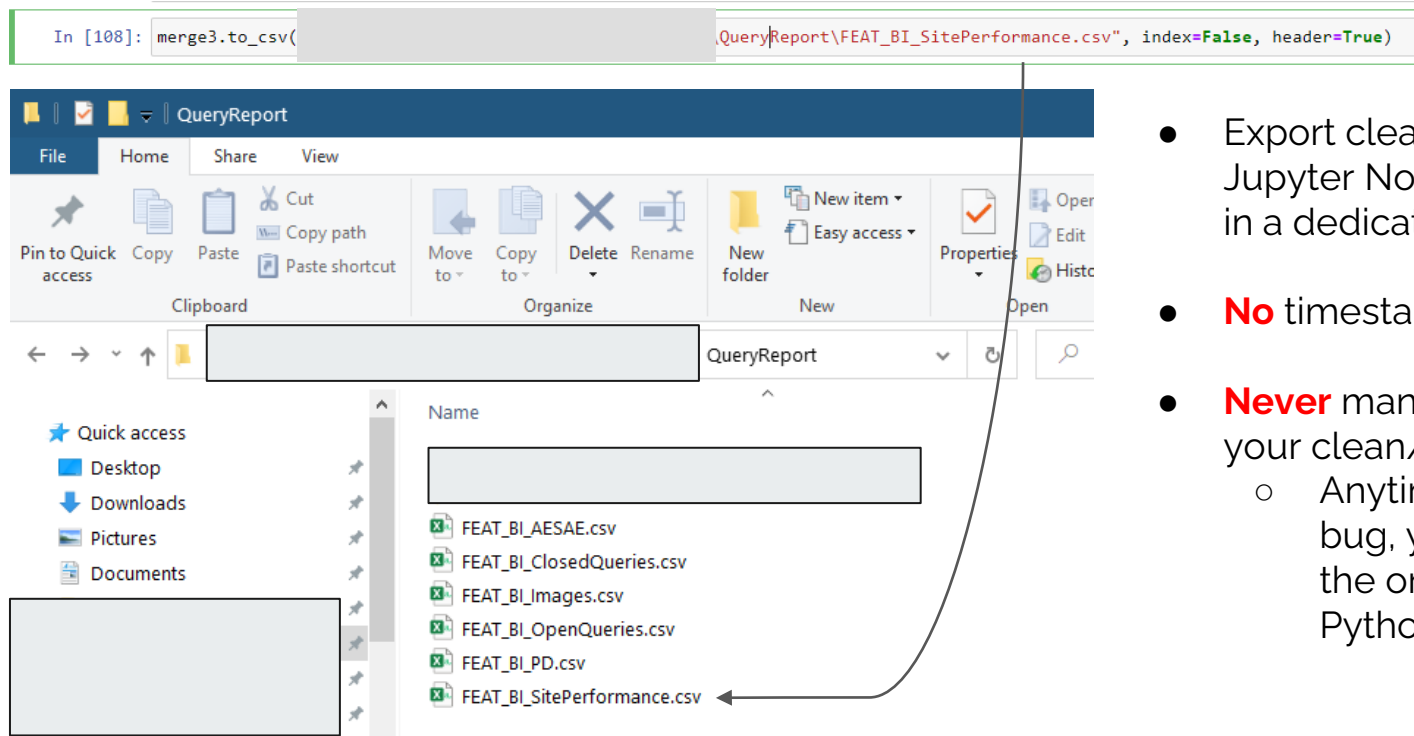
#!/usr/bin/env python
import requests
data = {
    'token': [REDACTED],
    'content': 'report',
    'format': 'csv',
    'report_id': '308017',
    'csvDelimiter': '',
    'rawOrLabel': 'label',
    'rawOrLabelHeaders': 'label',
    'exportCheckboxLabel': 'true',
    'returnFormat': 'csv'
}
r = requests.post('https://redcap.vanderbilt.edu/api/', data=data)
print('HTTP Status: ' + str(r.status_code))
print(r.text)

HTTP Status: 200
```

The API request was successful.

- 6 API Calls from 2 REDCap databases
- Each API call takes ~ 1 minute, depending on internet connection
- Import raw data and convert it to **Pandas dataframes** for data transformation
- Regular debugging is needed for maintenance

# Data Load (ETL)



- Export clean data from Jupyter Notebook to csv files in a dedicated folder
- **No** timestamp in file names
- **Never** manually open/edit your clean/flat csv data files
  - Anytime if you catch a bug, you can correct the original data or Python code

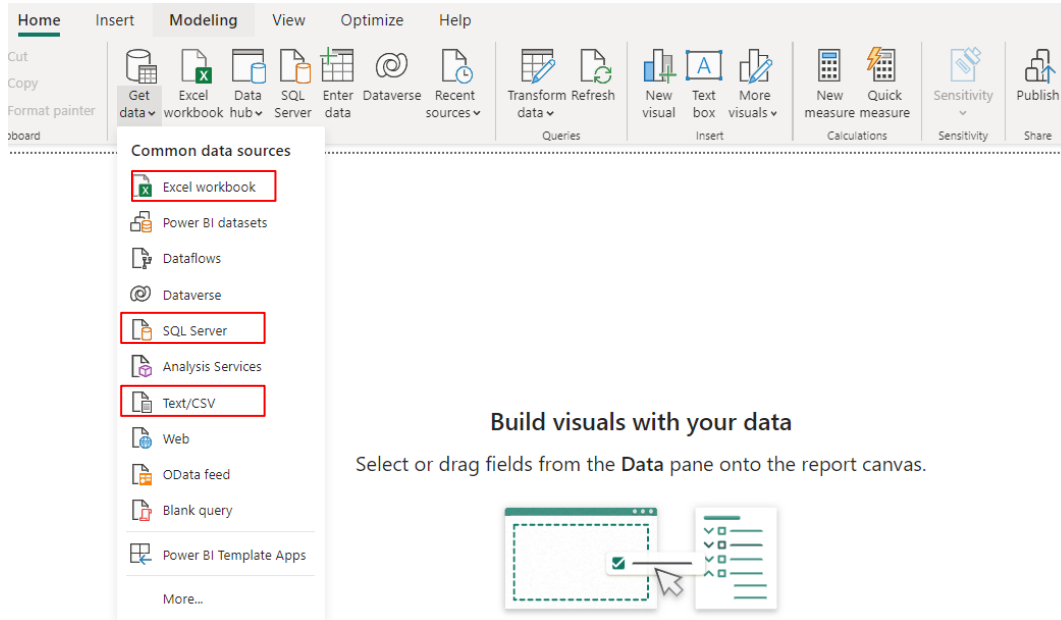
# Intro to PowerBI Dashboard



PowerBI dashboard:

- a data visualization tool that allows users to create interactive and dynamic dashboards to analyze and present data
- connect to various data sources, such as Excel spreadsheets, databases, and cloud-based services, to collect and transform data into meaningful insights

# PowerBI Dashboard Setup



The screenshot shows the Power BI ribbon with the 'Modeling' tab selected. The 'Get data' dropdown menu is open, displaying a list of common data sources. The following items are highlighted with red boxes:

- Excel workbook
- SQL Server
- Text/CSV

Other data sources listed include Power BI datasets, Dataflows, Dataverse, Analysis Services, Web, OData feed, Blank query, and Power BI Template Apps. A 'More...' link is at the bottom of the list.

## Build visuals with your data

Select or drag fields from the **Data** pane onto the report canvas.



## Get Data

Search

All

File

Database

Power Platform

Azure

Online Services

Other

Other

Web

SharePoint list

OData Feed

Active Directory

Microsoft Exchange

Hadoop File (HDFS)

Spark

Hive LLAP

R script

Python script

ODBC

OLE DB

REDCap Connector

# Demo!



This demo dashboard used simulated datasets. All site info is publicly available on [Clinicaltrials.gov](https://clinicaltrials.gov).

# Conclusion

## How to connect all metrics together to tell a story?

- Updates: good news vs bad news
- Action items & decisions for the team
- Discussion topics

## An iterative process until all needs and wants are met

- Feedback from audience

## Less is more!

## Time is of the essence!

## What's the true value of this?

- Using **free\*** and **existing** technologies to build a **seamless, fast and automatic** data pipeline that provides **accurate and real-time** data for team members (esp. Pls) to make **timely** decisions.

# Questions? Feedback? More?

Contact Me

Email: [ally.qi@vumc.org](mailto:ally.qi@vumc.org)

Office: 615-875-0619



# Thank you!

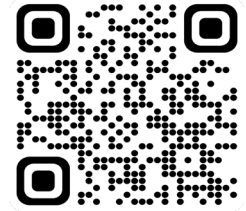
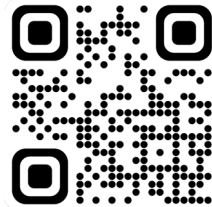
VCC family:

**Stephanie Smith, Jessica Marlin, David McKeel, Kelly Walsh, John Graves, Courtney Jordan, and LSS.**

Leadership:

**Amanda Bistran-Hall, Krista Vermillion, Jess Collins, Bree Burks, Dr. Michael Froehler**

For more info about VCC:  
<https://vcc.vumc.org>



**National PI: Dr. J Mocco, MD, MS**  
Kalmon D. Post Professor and Senior Vice Chair  
Cerebrovascular Center Director  
Department of Neurological Surgery  
Mount Sinai Health System