FOCiS 2024 Workshop: Extracting and Analyzing Data Using AI Tools in ImmPort

Yusuf Ashktorab
Reuben Sarwal
Sammi Smith
Alicia Williamson
Sanchita Bhattacharya

FOCiS®
Federation of Clinical Immunology Societies

ImmPort
Bioinformatics For The Future Of Immunology

Overview of tutorial

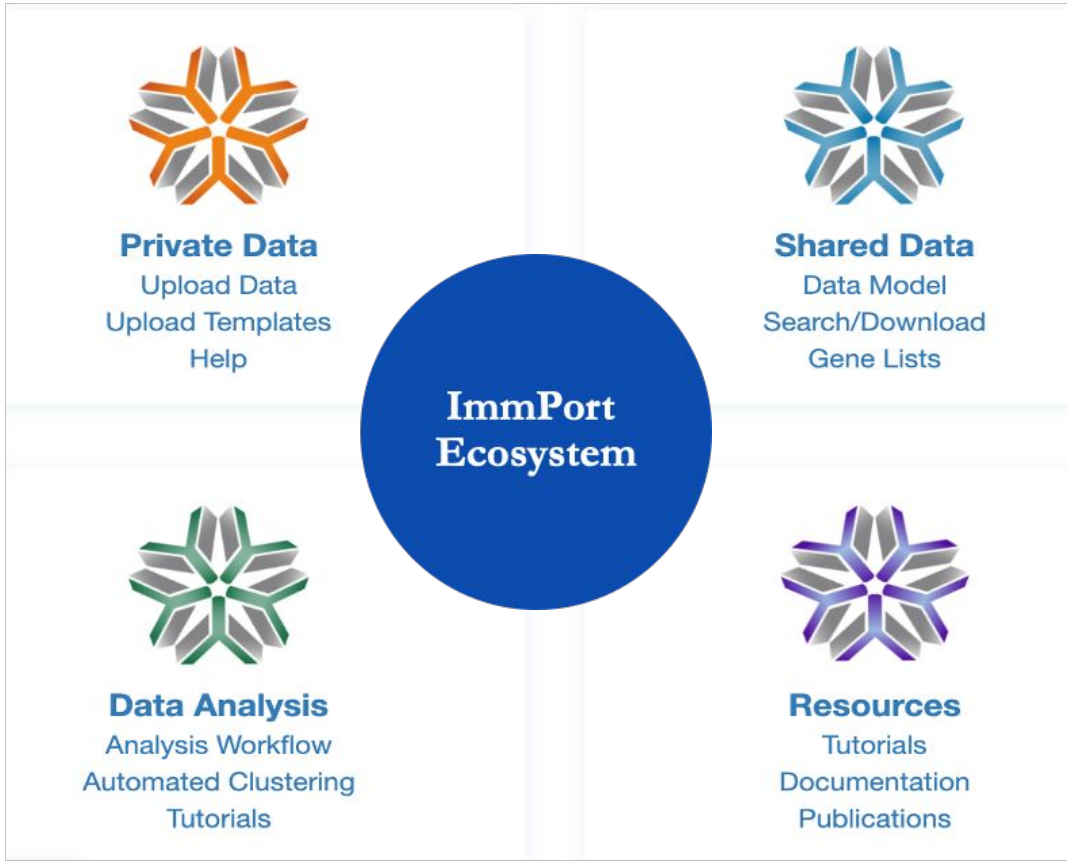
Scientific Question:
Do vaccine responses vary across different vaccines when considering factors such as age, race, and gender?

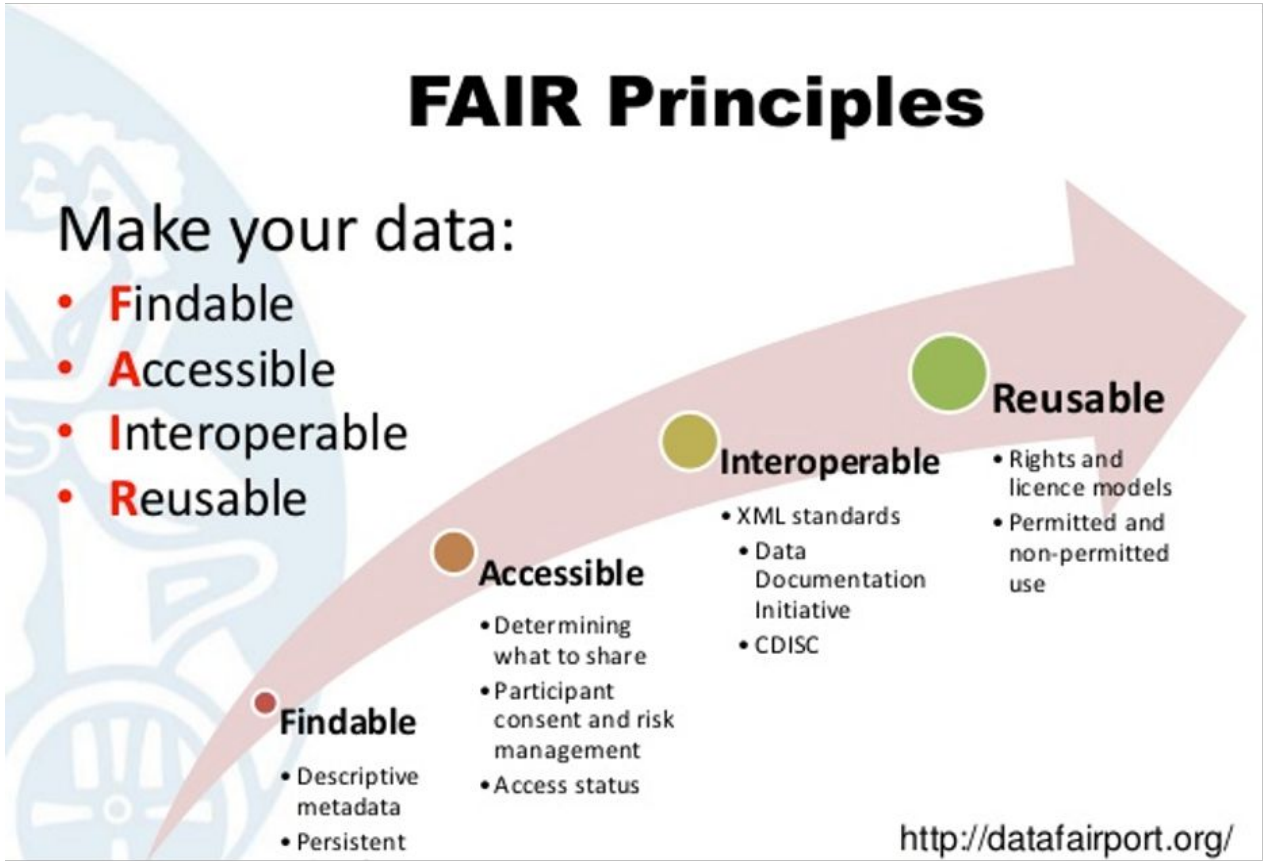
In this tutorial we will explore how to use the ImmPort database, ChatGPT, and Jupyter notebooks to extract, explore, and analyze vaccine response data, specifically using ELISA data as an example.

ImmPort

Bioinformatics For The Future Of Immunology

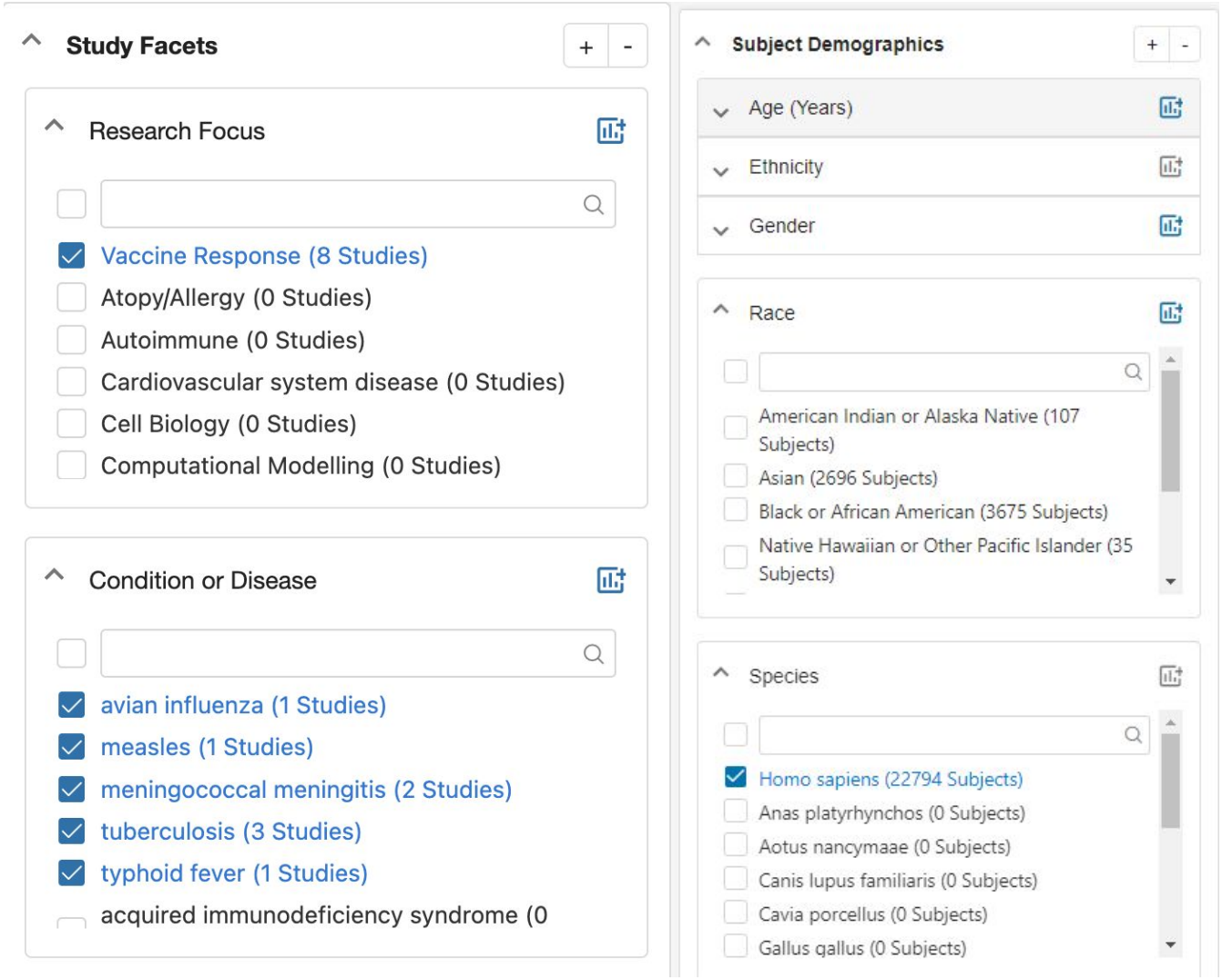
ImmPort data portal was developed to collect and share research and clinical trials data from NIAID/DAIT funded researchers





Let's explore some data!

Step 1: Identify areas of interest



Ex: Vaccine response in Humans, select these Conditions/Diseases: Avian influenza, measles, meningococcal meningitis, tuberculosis, and typhoid fever

[Link to search]

Parts of this tutorial

This tutorial is spread out into 4 main sections:

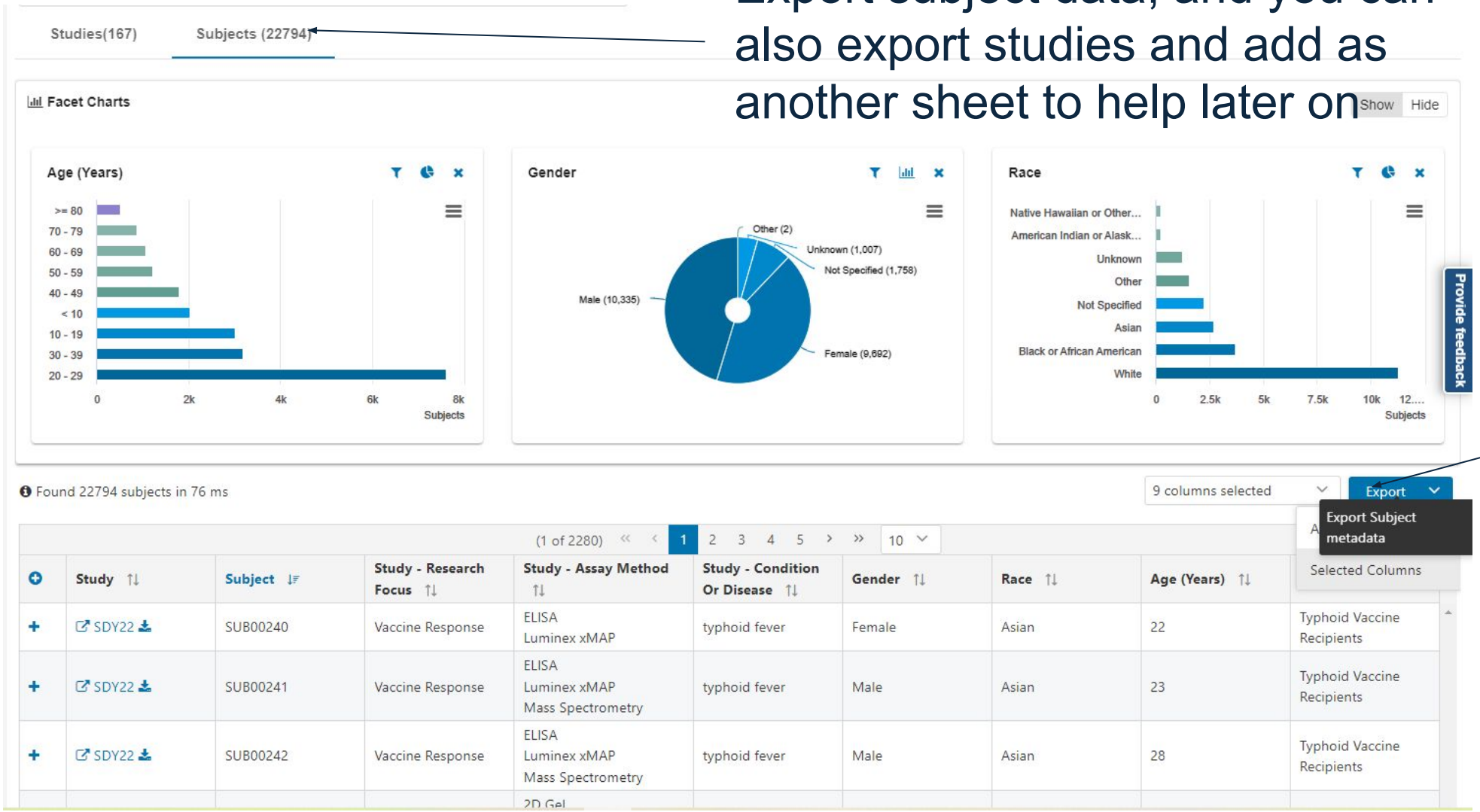
1. Downloading and exploring data in ImmPort
2. Generating useful figures in GPT4 for exploratory data analysis
3. Using the ImmPort API to pull assay and other relevant data for downstream analysis
4. Using GPT4 to help and run downstream analysis of assay data, specifically ELISA

Part 1

Downloading and exploring data in ImmPort

Step 2: Export relevant data

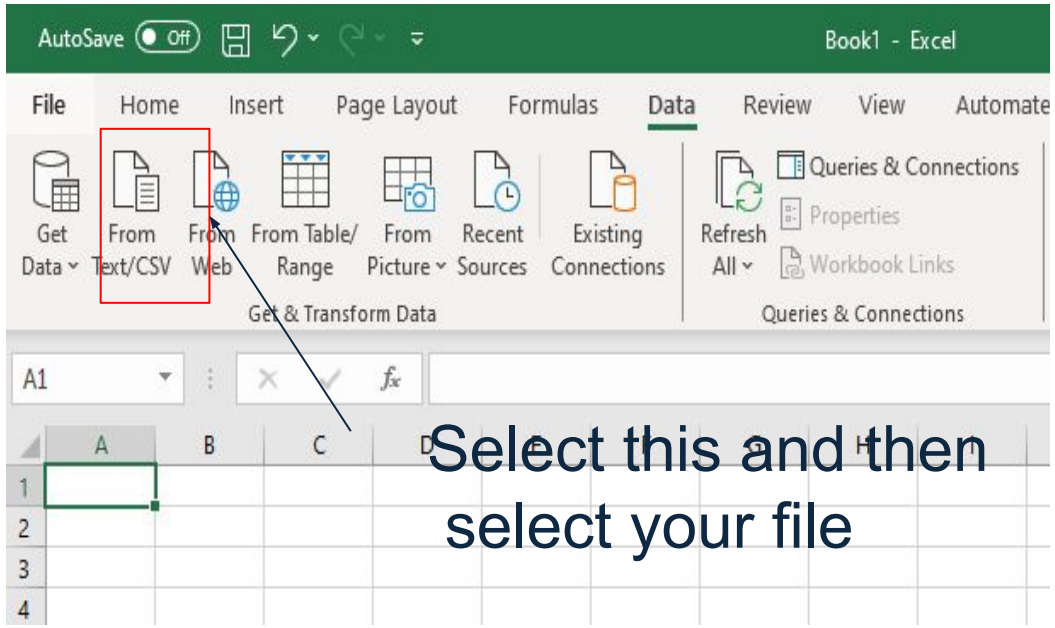
Export subject data, and you can also export studies and add as another sheet to help later on



Preferred to export all unless you know you wont need it later

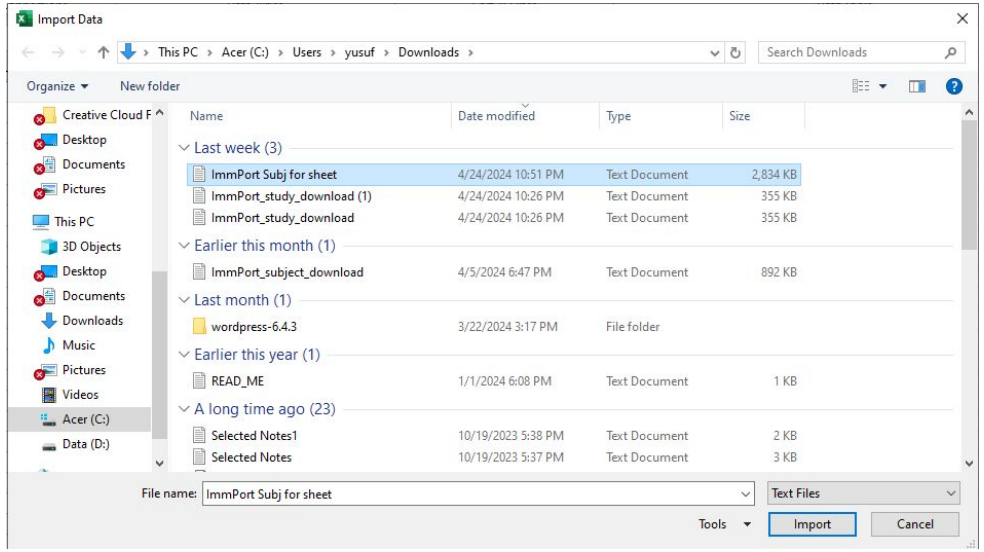
Convert .txt to CSV

1)



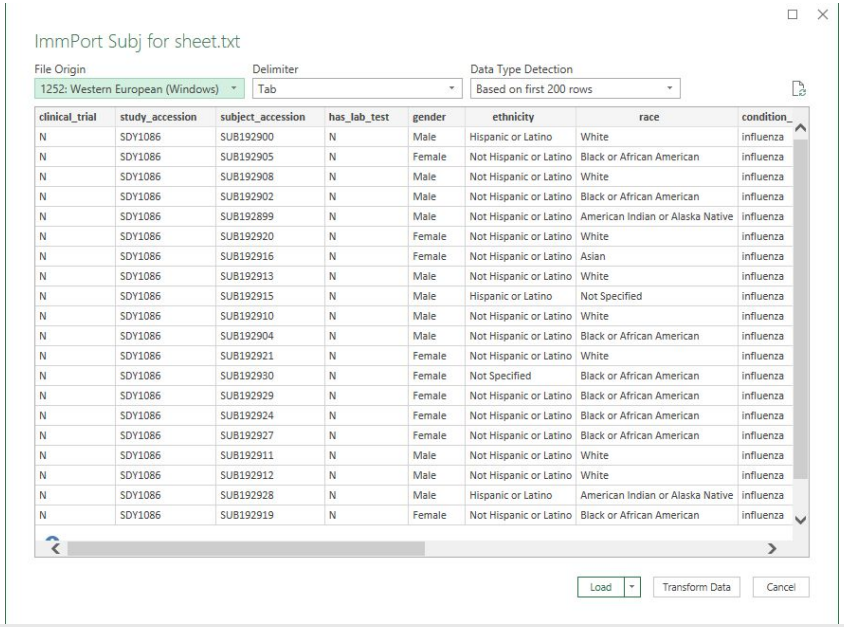
Select this and then select your file

2)



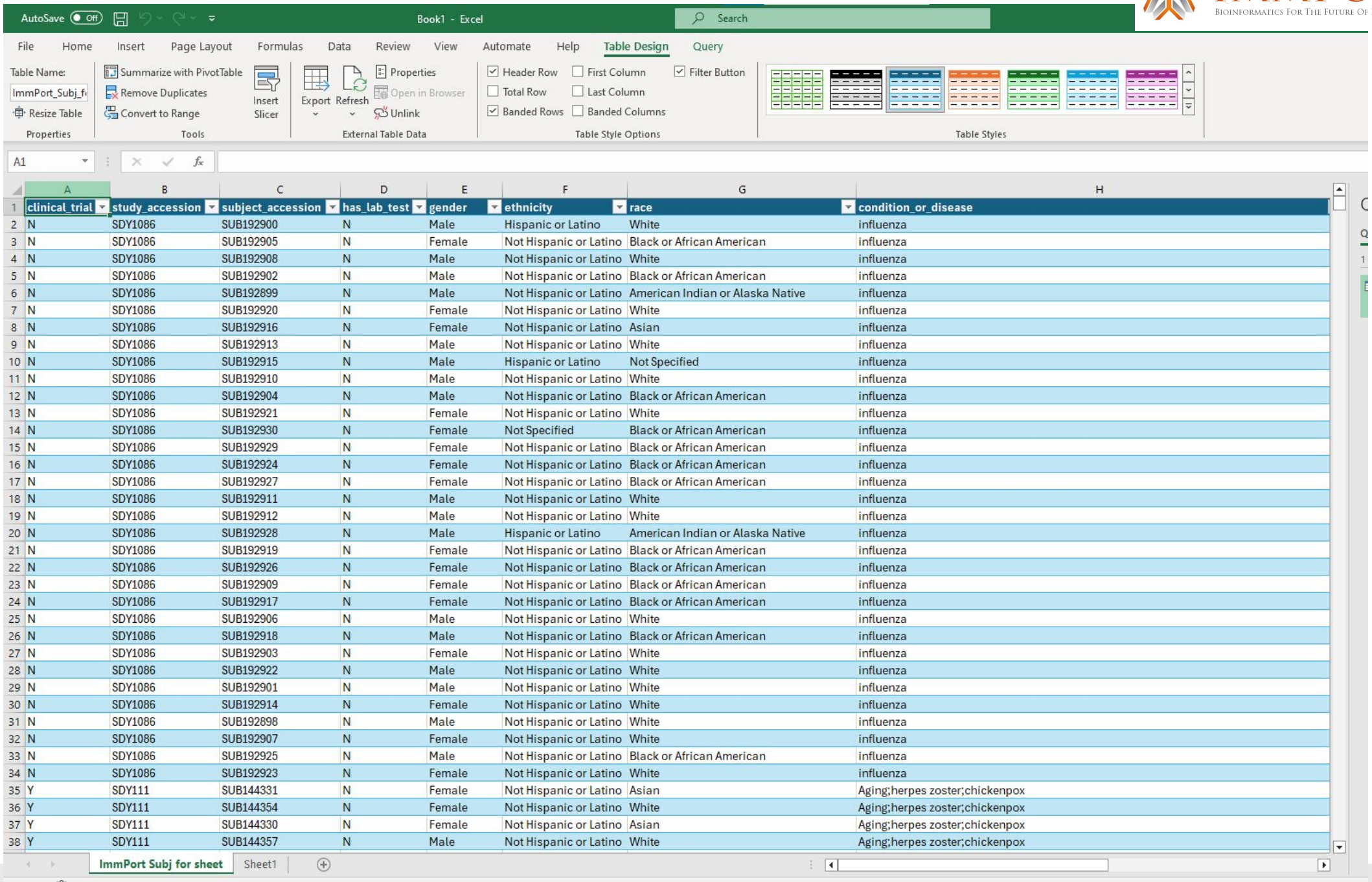
Select the downloaded file

3)



It will give you a preview, simply load What it gives.

It should look something like this, simply save your file as a .csv or .xlsx, this will be what we are working with.



Book1 - Excel

clinical_trial	study_accession	subject_accession	has_lab_test	gender	ethnicity	race	condition_or_disease
N	SDY1086	SUB192900	N	Male	Hispanic or Latino	White	influenza
N	SDY1086	SUB192905	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192908	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192902	N	Male	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192899	N	Male	Not Hispanic or Latino	American Indian or Alaska Native	influenza
N	SDY1086	SUB192920	N	Female	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192916	N	Female	Not Hispanic or Latino	Asian	influenza
N	SDY1086	SUB192913	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192915	N	Male	Hispanic or Latino	Not Specified	influenza
N	SDY1086	SUB192910	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192904	N	Male	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192921	N	Female	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192930	N	Female	Not Specified	Black or African American	influenza
N	SDY1086	SUB192929	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192924	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192927	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192911	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192912	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192928	N	Male	Hispanic or Latino	American Indian or Alaska Native	influenza
N	SDY1086	SUB192919	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192926	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192909	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192917	N	Female	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192906	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192918	N	Male	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192903	N	Female	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192922	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192901	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192914	N	Female	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192898	N	Male	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192907	N	Female	Not Hispanic or Latino	White	influenza
N	SDY1086	SUB192925	N	Male	Not Hispanic or Latino	Black or African American	influenza
N	SDY1086	SUB192923	N	Female	Not Hispanic or Latino	White	influenza
Y	SDY111	SUB144331	N	Female	Not Hispanic or Latino	Asian	Aging;herpes zoster;chickenpox
Y	SDY111	SUB144354	N	Female	Not Hispanic or Latino	White	Aging;herpes zoster;chickenpox
Y	SDY111	SUB144330	N	Female	Not Hispanic or Latino	Asian	Aging;herpes zoster;chickenpox
Y	SDY111	SUB144357	N	Male	Not Hispanic or Latino	White	Aging;herpes zoster;chickenpox

Why did we choose these studies?

- These studies were chosen specifically due to their simplicity and the lack of data cleaning required for the purpose of this tutorial, but that should not be a criteria for your own work.
- GPT can also be used to help you clean data, an example is the fact that studies can have multiple conditions or diseases, or inclusion criteria which you don’t want in your study, you have to be careful with these things, using the API can help you get more context about the studies and help with data cleaning.

Part 2

Generating useful figures in GPT4 for exploratory data analysis, data driven hypothesis testing

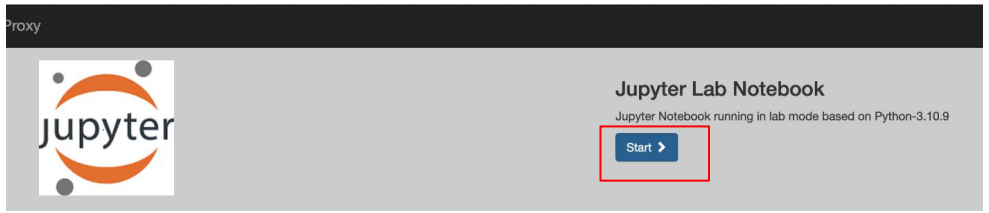
Logging in to Butte Lab Server

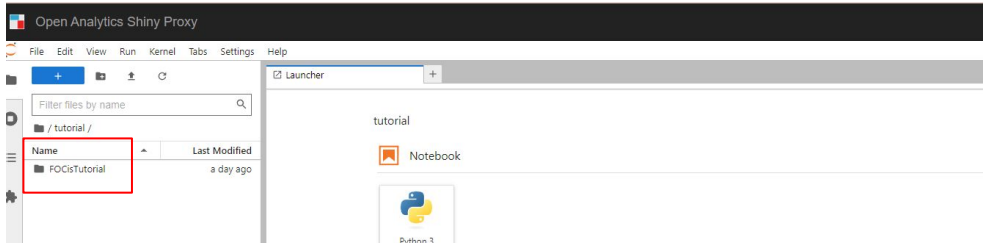
focis2024.net

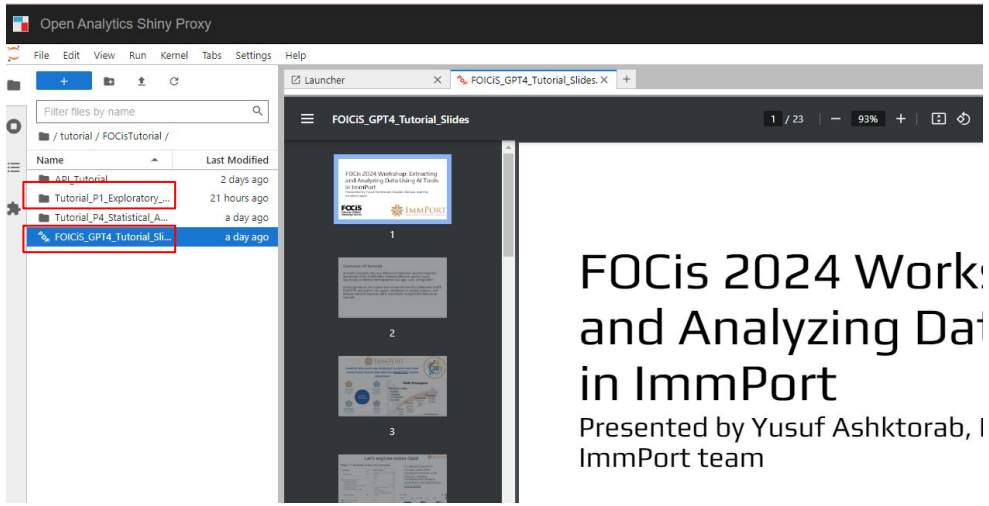
Email:
Email you used to register

Password:
ImmportFocis2024!

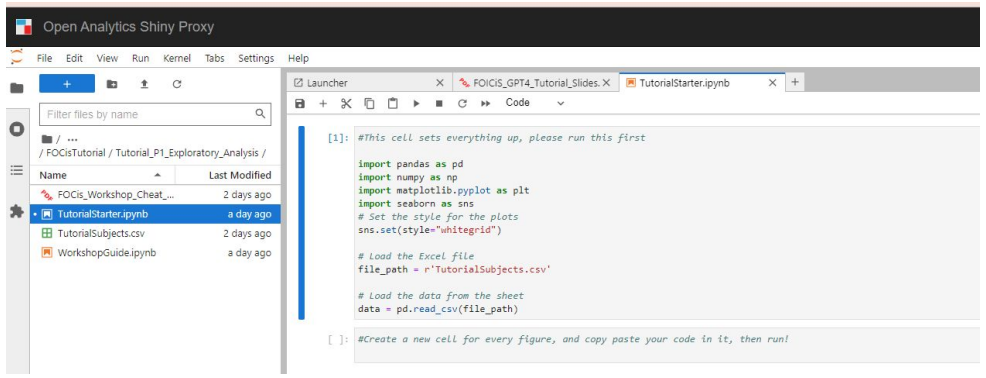
Accessing Tutorial and Data







1) First press start
2) Next open the first folder by double clicking
3) Open the PDF FOCiS Slides by double clicking, this will allow you to follow along
4) Next double click on Tutorial P2 folder, and then open the Tutorial Starter notebook by double clicking agin



Disclaimer!

- If you have the free version of GPT4, you have a limit on number of prompts, so please be diligent in what prompts you use and try to be as prompt efficient as possible to get the most out of the tutorial.

- The code and interpretations presented in this tutorial have been generated with the assistance of OpenAI's GPT-4. While every effort has been made to ensure the accuracy and reliability of the information, it is important to recognize that AI-generated content may not be perfect and should be used with discretion. For more information about OpenAI and GPT-4, please visit openai.com.

Video Walkthrough

Now Take some time to Explore!

- Can't think of figure ideas? Worksheet contains ideas of figures to generate (page 1)
- Trouble prompting? (refer to pages 2 and 3 of worksheet)
- Still stuck? Raise your hand and ask for help, or check the WorkshopGuide notebook
- Too easy? Raise your hand and we will give you new challenges!

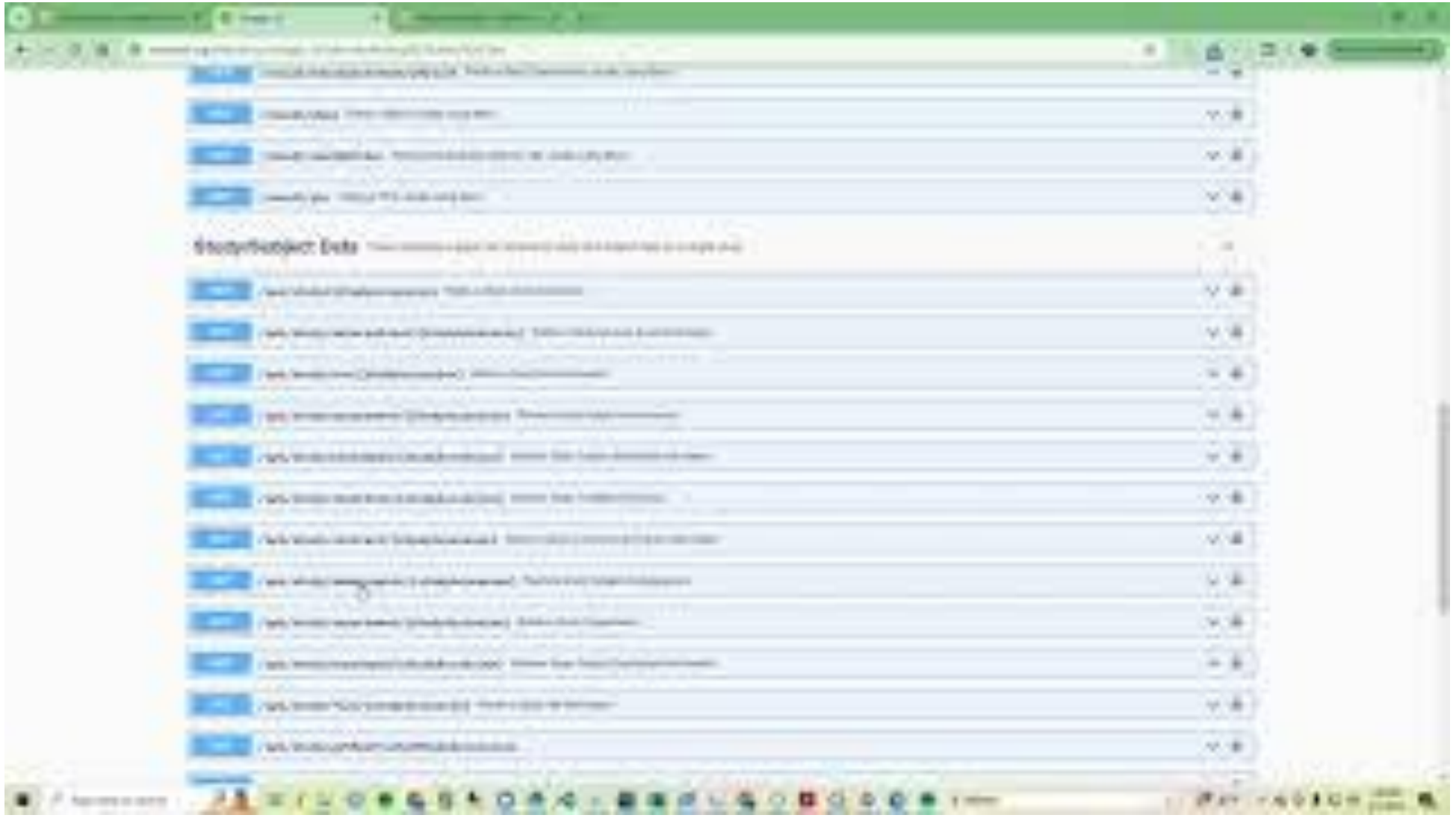
Bringing People Back

- Was anyone able to think of and make any interesting figures? Please share!
- What issues did you run into? What improvements can we make?
- What did you learn, what went well?
- Did this want to make you use AI tools like ChatGPT more in your own work?

Part 3

Using the ImmPort API to pull assay and other relevant data for downstream analysis

Using the API to extract the assay data



Part 4

Using GPT4 to help and run downstream analysis of assay data, specifically ELISA

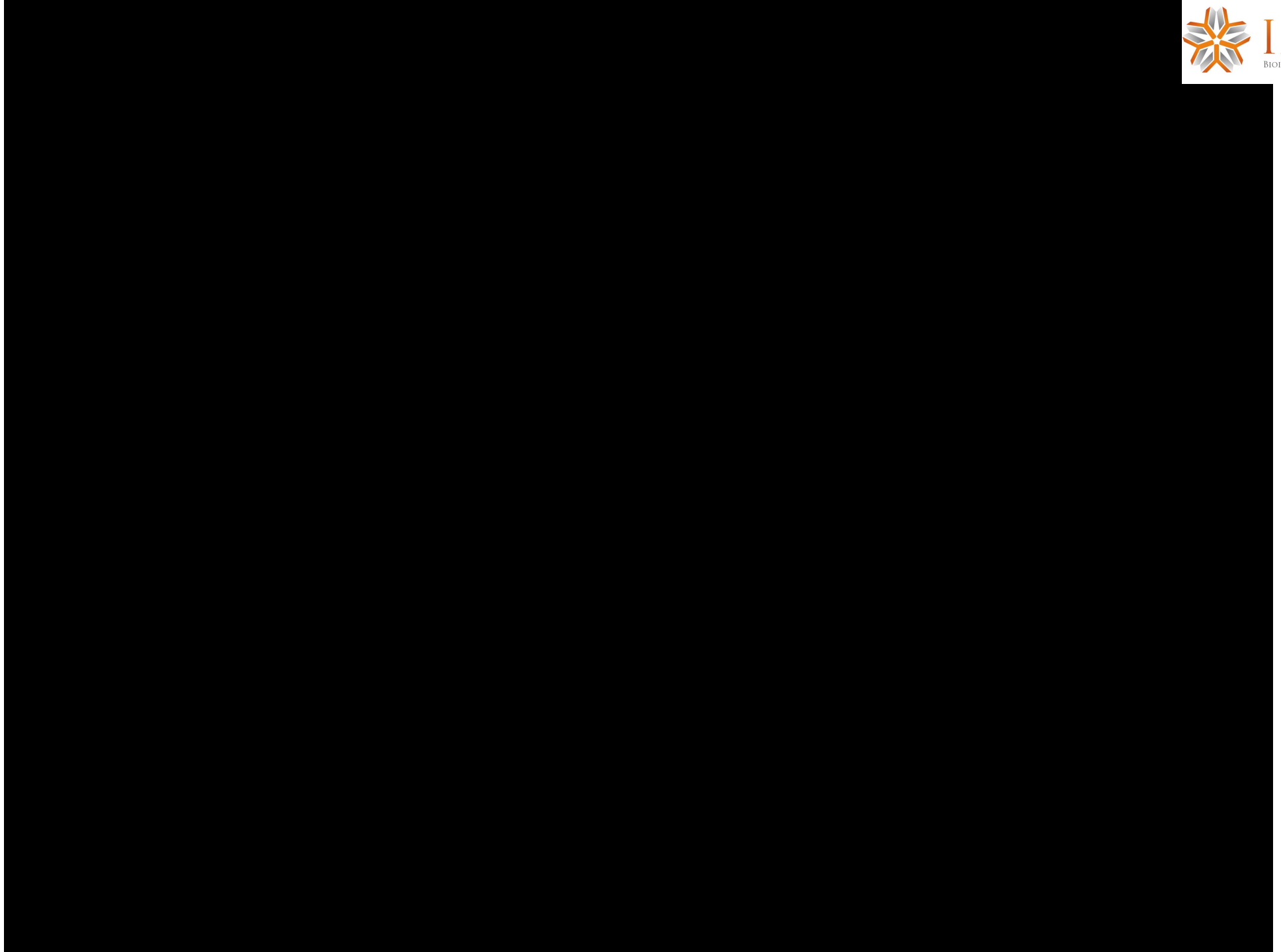
Let's Familiarize ourselves!

Open the ELISATUTPROJ excel sheet. This is the ELISA data that we pulled with the API earlier. Take some time to look through it, specifically take note of the fact that we have TIV and LAIV vaccine groups, IgG, IgM, and IgA analyte data, and measurements across several different days.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
	ageEve	stralPo	lyteAcc	ytePre	analyteReporte	mAcces	armNa	mpleA	sample	clinic	ethnici	imentA	mpleA	gende	ement	Age	edVisitA	race	studyTi	imeColle	jectAcc	ectPhe	nitRepo	lueRep	mentAc	yTimeT	dy_acce	tion_or	sea
25	Age at enr	North America			IgA	ARM2343	LAIV 2011	BS722134	Serum	N	Not Hispa	EXP13739	ES792123	Female	ELISA	63	PV3182	White	Immune R	0	SUB13618	LAIV Vacc	ng/ml	6256.4824	TRT1453	Time of in	SDY396	influenza	
26	Age at enr	North America			IgG	ARM2343	LAIV 2011	BS722134	Serum	N	Not Hispa	EXP13739	ES792124	Female	ELISA	63	PV3182	White	Immune R	0	SUB13618	LAIV Vacc	ng/ml	246923.04	TRT1453	Time of in	SDY396	influenza	
27	Age at enr	North Am	ANA817	IgM	IgM	ARM2343	LAIV 2011	BS722134	Serum	N	Not Hispa	EXP13739	ES792125	Female	ELISA	63	PV3182	White	Immune R	0	SUB13618	LAIV Vacc	ng/ml	7038.5522	TRT1453	Time of in	SDY396	influenza	
28	Age at enr	North America			IgA	ARM2343	LAIV 2011	BS722145	Serum	N	Not Hispa	EXP13739	ES792126	Male	ELISA	44	PV3182	White	Immune R	0	SUB13618	LAIV Vacc	ng/ml	7115.8110	TRT1453	Time of in	SDY396	influenza	
29	Age at enr	North America			IgG	ARM2343	LAIV 2011	BS722145	Serum	N	Not Hispa	EXP13739	ES792127	Male	ELISA	44	PV3182	White	Immune R	0	SUB13618	LAIV Vacc	ng/ml	189780.17	TRT1453	Time of in	SDY396	influenza	
30	Age at enr	North Am	ANA817	IgM	IgM	ARM2343	LAIV 2011	BS722145	Serum	N	Not Hispa	EXP13739	ES792128	Male	ELISA	44	PV3182	White	Immune R	0	SUB13618	LAIV Vacc	ng/ml	9167.2080	TRT1453	Time of in	SDY396	influenza	
31	Age at enr	North America			IgA	ARM2343	LAIV 2011	BS722157	Serum	N	Not Hispa	EXP13739	ES792129	Female	ELISA	53	PV3182	Black or A	Immune R	0	SUB13618	LAIV Vacc	ng/ml	15561.838	TRT1453	Time of in	SDY396	influenza	
32	Age at enr	North America			IgG	ARM2343	LAIV 2011	BS722157	Serum	N	Not Hispa	EXP13739	ES792130	Female	ELISA	53	PV3182	Black or A	Immune R	0	SUB13618	LAIV Vacc	ng/ml	288687.43	TRT1453	Time of in	SDY396	influenza	
33	Age at enr	North Am	ANA817	IgM	IgM	ARM2343	LAIV 2011	BS722157	Serum	N	Not Hispa	EXP13739	ES792131	Female	ELISA	53	PV3182	Black or A	Immune R	0	SUB13618	LAIV Vacc	ng/ml	12516.099	TRT1453	Time of in	SDY396	influenza	
34	Age at enr	North America			IgA	ARM2343	LAIV 2011	BS722169	Serum	N	Not Hispa	EXP13739	ES792132	Female	ELISA	44	PV3182	White	Immune R	0	SUB13618	LAIV Vacc	ng/ml	18499.509	TRT1453	Time of in	SDY396	influenza	

More about our data

- The TIV (Trivalent Influenza Vaccine) targets three strains of the influenza virus - typically two A strains (H1N1 and H3N2) and one B strain—and is a traditional inactivated vaccine administered via injection.
- The LAIV (Live Attenuated Influenza Vaccine) contains live but weakened influenza viruses, designed to stimulate a stronger immune response, and is typically administered as a nasal spray.
- The dataset includes IgG (Immunoglobulin G), the most common antibody in blood circulation; IgM (Immunoglobulin M), the first antibody produced in response to an infection; and IgA (Immunoglobulin A), which plays a critical role in mucosal immunity.
- Measurements were taken at multiple time points across different days, allowing for analysis of immune response over time.



Final Exploration

Now take some time to play with the notebook, and use GPT 4 to see if you can recreate some of the plots, or make new ones!

Thank You For Attending!

Please take a moment to fill out this quick survey about your experience here!

ImmPort

Bioinformatics For The Future Of Immunology

FOCiS®

Federation of Clinical Immunology Societies