Before we get started...

Turn on Video
Seeing each other's faces will make the experience immersive.

Let's Discuss
Contribute to the class! You're driving the content with me.

Ask Questions
Feel free to ask questions anytime via chat or verbally.

DAY 1 AGENDA

INTRODUCTIONS	01
SPRINT OVERVIEW AND EXPECTATIONS	02
SURVEY BASICS	03
RECAP OF BASIC STATS AND PYTHON	04
BASIC PANDAS	05

01
INTRODUCTIONS

aaronstaclara

Rick Aaron Sta.Clara

Data Scientist, GrowSari
Data Science Instructor, Eskwelabs

Data scientist with a degree in chemical engineering. Background includes engineering, energy and fintech.

🚀 Currently working in the fintech arm of GrowSari, a tech startup aiming to empower more than 1M sari-sari stores in the Philippines.

🤖 Previously assisted in Eskwelabs as a Data Science Mentor.

📊 Core data science expertise: (1) forecasting, (2) credit scoring and (3) analytics

UNIVERSITY OF THE PHILIPPINES 1908

growsari
All-in-One Negosyo Power!

ESKWELABS

GROUP 1

Your teammates for the duration of Sprint 1.

Basty Vergara
Chief Data Officer
LinkedIn
@Went#9326

1. **Mark Pintac**
2. **Angelica Divine Alan**
3. **Michael Arsolon**
4. **Isabelle Gilay**
5. **Jamil Alawi**
6. **Ericson Tulda**

GROUP 2

Your teammates for the duration of Sprint 1.

John Patrick Laurel
Data Analyst
LinkedIn
@pats#6678

1. **Karla Concepcion**
2. **Karen Bioy**
3. **Sofia Calvo**
4. **Rex Fuentes**
5. **Enrico Asuncion**
6. **Chiara Perez**

GROUP 3

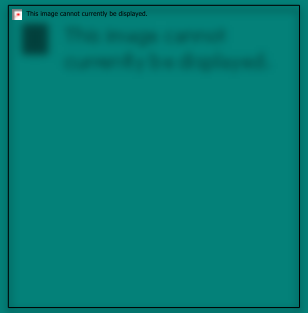
Your teammates for the duration of Sprint 1.

Fred Baldo
Audit Senior Manager
LinkedIn
@Fred0824#3263

1. **Lee Mark Chavez**
2. **Lynmar Didal**
3. **Andre Dometita**
4. **John Andrew Chua**
5. **Christian Pantoja**
6. **Tristan Tiu**

GROUP 4

Your teammates for the duration of Sprint 1.



Alyssa Ty
Senior Data Engineer at Rappler
LinkedIn
@sasaaty#2731

1. **Ephraim Cabanlit**
2. **Lesmon Saluta**
3. **Andre Pardillo**
4. **Bien Aculan**
5. **John Emerson Nuqui**
6. **Keith Yadao**

INTRODUCE YOURSELF

To jumpstart our lecture series, let's do a QUICK round of introductions:

1. What's your name?
2. What's your background?
3. What do you hope to learn in this sprint?

02
SPRINT OVERVIEW AND EXPECTATIONS

ESKWELABS

Towards Financial Inclusion (FI)

Using survey data to assess FI status of the Philippines

Sprint 1, Sept 2022

Rick Aaron Sta.Clara

MOTIVATION OF SPRINT

In line with the National Strategy for Financial Inclusion (NSFI) 2022-2028 by Bangko Sentral ng Pilipinas (BSP), this sprint aims to:

(1) Profile financial inclusion (FI) metrics in the Philippines using survey data from World Bank.
(2) Formulate policy recommendations to further improve access to financial services particularly to vulnerable sectors.

Institutions

The **World Bank** is an international financial institution that provides loans and grants to low and middle-income economies with aims (1) to end extreme poverty and (2) to promote shared prosperity.

The **Bangko Sentral ng Pilipinas** is the central bank and monetary authority of the republic of the Philippines. Primarily, they are in charge of regulating the financial system of the Philippines.

What is financial inclusion?

According to **BSP,** "financial inclusion is a state in which everyone, especially the vulnerable sectors, has effective access to a wide range of financial services".

This can be achieved thru access to financial products and services that must be able to serve needs for "saving, payments, financing, investing, or getting insured".

Focus for this sprint:

1. Who are these vulnerable groups?
2. What is their profile?
3. How can we bridge the gap?

The Data

The **Global Financial Inclusion (Global Findex) Database 2017** is a database containing survey responses from 150,000 adults age 15 and above in over 140 economies. This includes updated indicators on access to and use of formal and informal financial services. Focus is on the Philippines but we might touch upon other countries for benchmarking.

The most recent survey database, the Global Financial Inclusion Database 2021, will be released Fall 2022.

Data points can be broadly categorized into two categories:

Demographic

Data related to age, income group, education and employment.

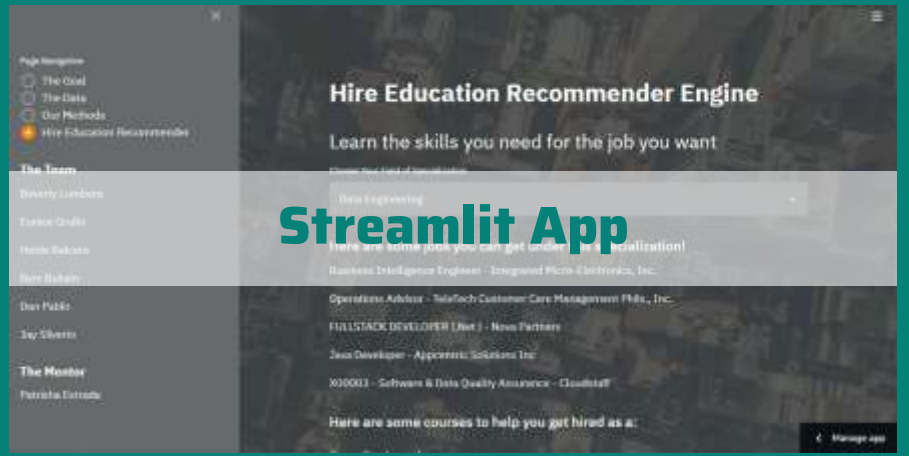
Access to Financial Services

Data related to access to debit card, credit card, savings, borrowings and payments.

SPRINT TIMELINE

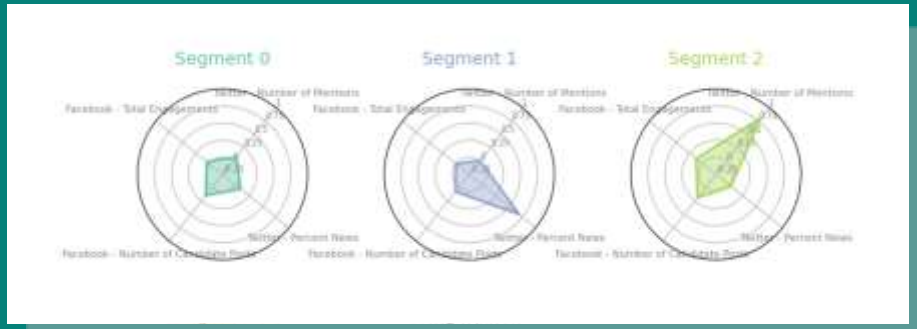
	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
Week 1	Sprint Overview **Pandas:** Basic data wrangling techniques	**Pandas:** Advanced data wrangling techniques Intro to data visualization	Data visualization Introduction to Streamlit	**Lab activity:** Exploratory Data Analysis	Data storytelling Creating your own Streamlit web page Mentor-free day	Review of Week 1
Week 2	Introduction to Machine Learning Clustering	Introduction to Linear Regression and Logistic Regression Mentor-free day	Introduction to deployments Deployment using Streamlit Cloud and Hugging Face Spaces	Sprint Project Prep Day 1	Sprint Project Prep Day 2	Sprint Presentation

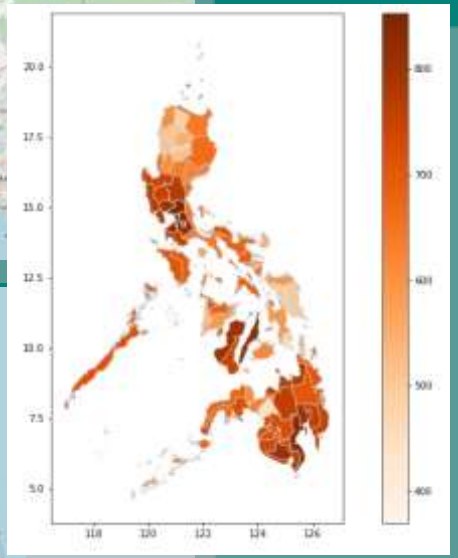
KEY DELIVERABLES

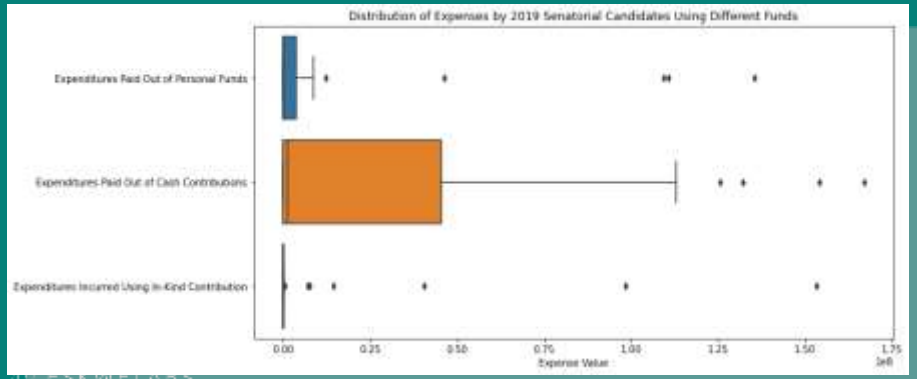


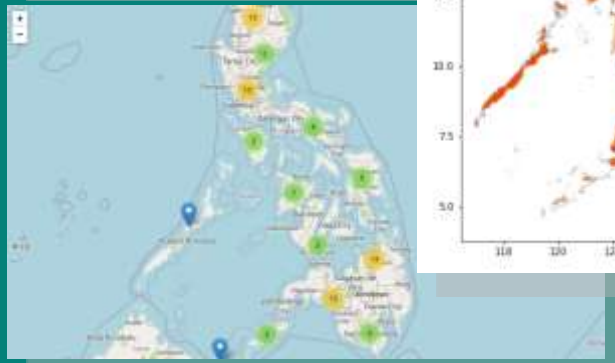
*Sample output for a different use case

SNIPPETS OF OUTPUT









ACTIVITY. In breakout rooms, access the Global Findex 2017 using this link https://microdata.worldbank.org/index.php/catalog/3324#metadata-identification. Explore the metadata (aka the data about the data) and discuss the following with your groupmates. Note that you will use this dataset in the next 2 weeks so it is important to have a good grasp of what it is about.

1. What are the FI variables in the dataset? How are they encoded?
2. Based on the metadata, what factors do you think can we correlate with access to financial services?
3. What analysis will be interesting to tackle for this? Example: discrepancy of access to financial services by gender.

Allotted time: 15 minutes

BREAK TIME ☕

Mamma Mia
Soundtrack
by TT Club

Be back by 07:10 PM

Rest your eyes
Look away from your screen for a while

ESKWELABS

03
SURVEY BASICS

Survey 101

Survey basics & data interpretation

Survey

Method of collecting data from a fraction of a population

"Surveys provide a speedy and economical means of determining facts about our economy and about people's knowledges, attitudes, beliefs, expectations, and behaviors."

— *American Statistical Association (ASA)*

Survey data collection

In-person interviews
Pen-and-paper or computer-assisted interviews

Self-administered pen-and-paper interviews

Telephone surveys

Self-administered computer surveys
Typically web interviewing

Basics of Survey Sampling

How do you know if the sample you are using captures the population you want to study?

- **Select respondents at random to avoid selection bias**
  Random sampling is essential to getting a good and representative dataset.

- **Sampling should be feasible and affordable**
  It is too expensive and time-consuming to interview everyone in a population. However, sampling reduces accuracy of interviewing everyone in a population. There should be a balance of feasibility and affordability against accuracy.

- **Sample should be large enough to meet required precision of results**
  There are minimum sample sizes that can provide acceptable levels of precision. A sample that is bigger than the minimum will have an increased precision. Any size less than the minimum, however, will likely result in an imprecise and inaccurate outcome.

“Biased naman ‘yang surveys.”

How accurate are survey results in estimating actual votes in the Philippines?

Pre-election Pulse Asia survey was conducted on April 16-21 with only 2,400 samples. Meanwhile, actual votes reached 56M. Despite the significant difference in sample size, we can clearly see how close the survey results approximated the actual results.

Carefully-designed surveys are powerful!

	% of Votes from Pulse Asia (April 16-21)	% of Actual Votes
Marcos	56	58.77
Robredo	23	27.94
Pacquiao	7	6.81
Domagoso	4	3.59
Lacson	2	1.66

04
RECAP OF BASIC PYTHON AND STATS

RECAP OF BASIC PYTHON AND STATS

Quiz Time

Go to https://www.menti.com/1mhftxnh7m and enter the game pin **1618 6977**.

What do I need to know to get thru this sprint?

1. **Practice your basic Python programming skills.** Remember your variables, data types, string and number manipulation, lists, tuples, functions, dictionaries and logic (if-else, for, while).
2. **Recall your stats.** Remember your central tendency measures, standard deviation, variance, distributions, correlations and hypothesis tests.
3. **Be comfortable with errors.** A lot of fellows get frustrated with this especially during their first few coding sessions. During the entirety of the cohort (and during the rest of your data science work), you WILL encounter errors. To debug:
    a. Always think about what you want to accomplish.
    b. Break it down into small steps.
    c. Search about it online (usually, StackOverflow). Be specific about what you want your code to do.
    d. If it still does not make sense, ask help from your mentors and instructor.
4. **Be inquisitive.** Always think beyond the data and the techniques.
5. **Have fun!**

04
PANDAS - BASICS

PANDAS

pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

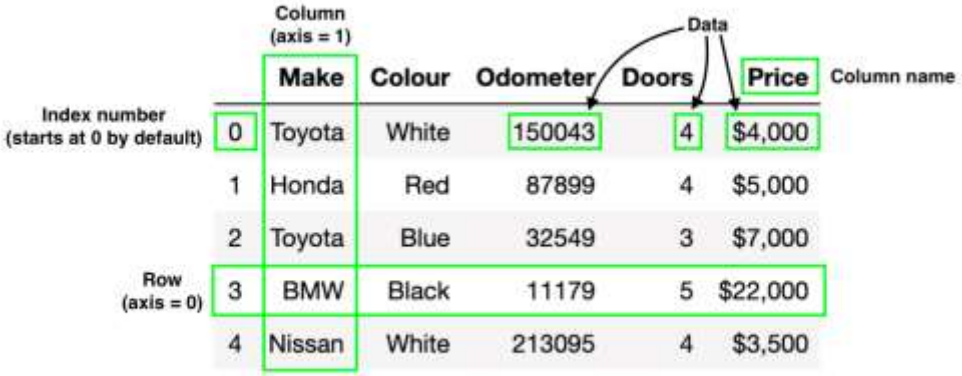
*Description from https://pandas.pydata.org/

ANATOMY OF PANDAS

A pandas DataFrame is similar to your SQL table or Excel sheet!

In the industry, we mainly use pandas to wrangle data which be used to prepare the data for downstream data science tasks like analytics (reporting, dashboarding) or machine learning.

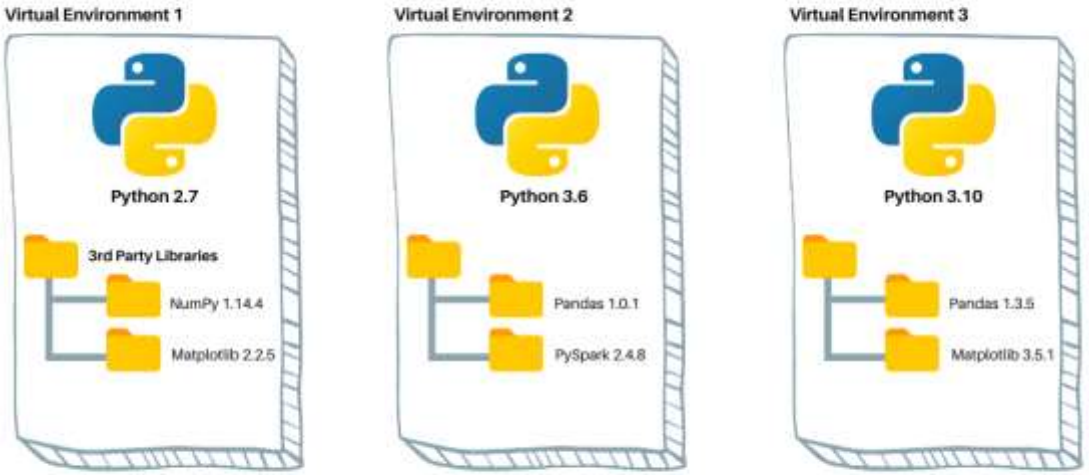
In most cases, you will spend MAJORITY of your time cleaning the data.



Before we start coding...

1. Make sure you have downloaded the materials for this sprint.
2. Make sure you followed the instructions in README.txt for a smooth coding experience. Prerequisites: Python, terminal navigation.
3. For any technical help, please reach out to me or your mentors.

Virtual environments



dataquest.io

LET'S CODE!

This is going to be a **code-along session**. If you can't follow, feel free to open your mic and interrupt. You may also ask your mentors for help. Again, encountering errors is natural. 😀

ACTIVITY. In breakout rooms, load & explore the FI survey dataset (micro_world.csv). Try to answer the following questions for each. Discuss findings with the rest of your group.

1. EASY. How many female respondents are there in the Philippines?
2. AVERAGE. What percentage of the Filipino respondents have BOTH credit card and debit card? Hint: browse through the data description to find the needed columns.
3. DIFFICULT. Compare the richest 20% to the poorest 20% of Filipino respondents in terms access to debit cards. Is there a discrepancy?

Allotted time: 20 minutes

Sign up as early as NOW

Even days before our deployment lecture next week, please create an account for each of the following:

1. **GitHub**. Access https://github.com/.
2. **Streamlit Cloud**. Access https://streamlit.io/cloud.
3. **Hugging Face**. Access https://huggingface.co/.

Hugging Face

TAKE-HOME READING:

National Strategy for Financial Inclusion 2022-2028 of Bangko Sentral ng Pilipinas

Ask:

1. Who are the vulnerable groups?
2. What are their profiles?
3. How does the government plan to bridge the gap?

End of Sprint 1 Day 1
Any questions?

Q & A