Economic and Social Commission for Western Asia

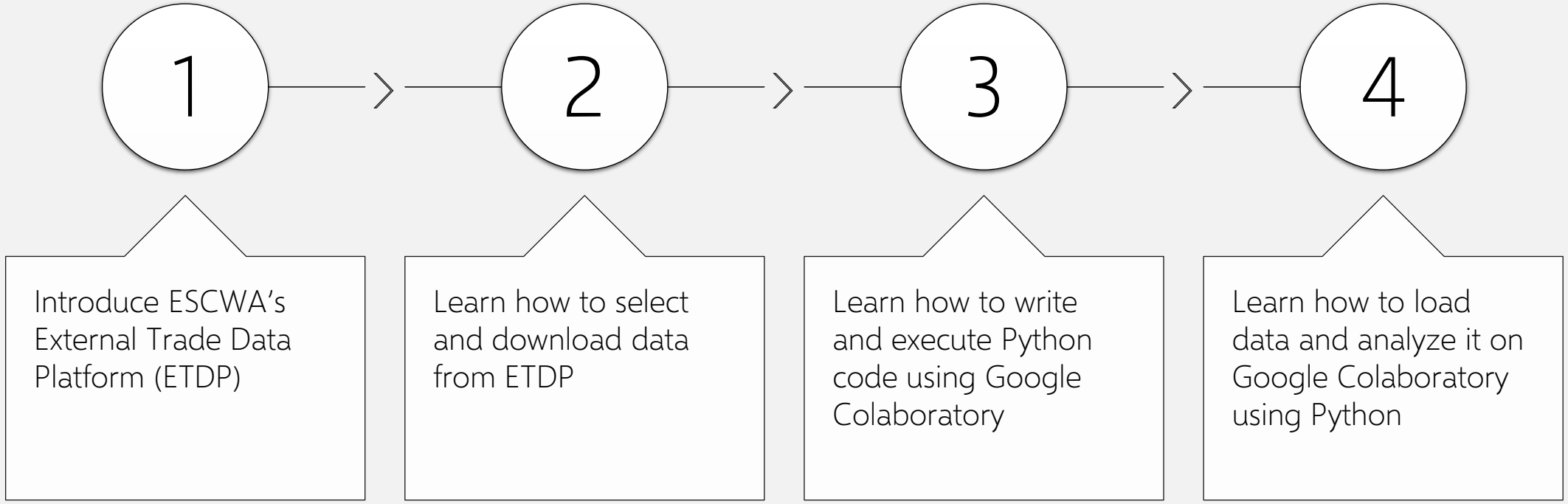
Python for Data Analytics: Trade Data Case Study

Mohamad Hossary
ESCWA

UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity Dignified Life

Goals



Agenda

- **Day 1:** Introduction to Python and the fundamentals of trade data.
- **Day 2:** Hands-on trade data exploration and visualization.
- **Day 3:** Calculating trade indicators and applying unsupervised machine learning.

UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity Dignified Life

Day 1: Python Fundamentals & Introduction to Trade Data

Introduction to Python

- A versatile, high-level programming language known for its readability and simplicity.
- It has a wide range of uses from web and application development to data analysis, machine learning, and AI

Why Use Python for Data Analysis?

- **Efficiency:** Write complex analysis with fewer lines of code compared to other languages.
- **Scalability:** Python's libraries are optimized to handle large datasets that are too big for traditional software.
- **Rich Ecosystem:** A vast collection of open-source libraries means you don't have to build everything from scratch.
- **Integration:** Easily connects with databases, web sources, and other statistical tools.

How Do We Work with Python?

Python is installed on your computer like any other application.

Several options are available for writing and running Python code:

- **Simple Text Editors:** Tools like Notepad can be used, but they lack helpful features.
- **IDEs (Integrated Development Environments):** Dedicated applications like VSCode or PyCharm provide everything needed for coding.
- **Web Applications:** Interactive tools like Jupyter Notebooks allow for a mix of code, text, and visualizations.

The Power of Jupyter Notebooks

An open-source web application for creating and sharing documents that contain live code, equations, visualizations, and narrative text.

Interactive "Cells": Run small chunks of code one at a time, see the output immediately, and make changes easily. This is ideal for exploratory data analysis.

Data Storytelling: Combine your analysis (the code) with your interpretation (the text and charts) in a single, easy-to-read document.

Supports many programming languages but is the standard for data science in Python.

Our Tool for this Workshop: Google Colab

A free, cloud-based version of Jupyter Notebook provided by Google.

Zero-Setup Required: No installation needed. You can start writing Python code instantly using just a web browser and a Google account.

Collaboration & Sharing: Easily share your notebooks with colleagues, just like a Google Doc.

Free Resources: Provides access to powerful computing resources (like GPUs) for free, which is essential for advanced machine learning tasks.

https://colab.research.google.com/

The Rise of Big Data in Official Statistics

NSOs are increasingly looking beyond traditional surveys to new, non-traditional data sources. "Big Data" refers to data characterized by high volume, velocity, and variety. In general, Big Data:

- Provides more timely and granular data.
- Can supplement or replace costly surveys.
- Offers new insights into economic and social phenomena.

Python is the key tool for processing and analyzing these new data sources.

Leveraging New Data Sources with Python

The UN Global Working Group on Big Data has identified several key sources NSOs can use:

Earth Observation	Satellite imagery for agriculture, environment, and land use statistics
AIS Data	Maritime traffic data for tracking trade flows and economic activity
Mobile Phone Data	Anonymized data for mobility, population density, and tourism statistics
Scanner Data	Retail scanner data for calculating price indices and consumption patterns

UN Task Teams on Big Data

The UN Committee of Experts on Big Data (UN-CEBD) coordinates global collaboration through several task teams. Their goal is to develop and share best practices for using Big Data for official statistics. For example:

- **Earth Observation Team:** Investigates using satellite data to improve statistics on agriculture, environment, and transport.
- **AIS Data Team:** Develops methods for using maritime data to measure trade, emissions, and fishery activity.
- **Training & Capacity Building Team:** Provides guidance and develops networks to enhance skills at NSOs worldwide.

Python Basics: Variables & Data Types

- Variables store data. Python automatically detects the data type.
- Datatypes include:

`int`: Integers (e.g., `age = 30`)
`float`: Decimals (e.g., `price = 25.99`)
`str`: Text (e.g., `name = "Ahmad"`)
`bool`: True/False (e.g., `is_adult = True`)

Python Basics: Operators

Operators perform actions on variables and values.

Arithmetic:
`+`, `-`, `*`, `/`, `**` (exponent)

Comparison:
`==`, `!=`, `>`, `<`, `>=`, `<=`

Logical:
`and`, `or`, `not`

Python Basics: Control Flow

Control the order in which code is executed

If/elif/else	tests conditions in order, running a block of code for the first condition that is true, and executing the final “else” block only if all previous conditions are false.
for	A “for” loop executes a block of code a specific number of times, once for each item in a sequence such as a list, a string, or a range of numbers.
while	A “while” loop repeatedly executes a block of code as long as a specified condition remains true.

Python Basics: Functions

- A function is a named, **reusable block of code** that performs a specific task, which helps you organize your program and avoid repeating yourself.
- Functions can accept inputs (as parameters) to customize their actions and can **optionally** send a result back (using the *return* keyword)

Python Basics: Data Structures

- Lists [ ]: Ordered and changeable collections.

```
my_list = [1, "apple", 3.5]
```

- Tuples (): Ordered and *unchangeable* collections.

```
my_tuple = (1, "apple", 3.5)
```

- Dictionaries {}: Unordered collections of key-value pairs.

```
my_dict = {“name”:”Ahmad”, “age”:32}
```

Introduction to Merchandise Trade Data

- Data on the exchange of goods between countries.
- **Key Components:**

Reporter:
The country reporting the data

Trade Flow:
Exports, Imports, Re-Exports

Partner:
The country with whom the trade is conducted.

Value:
The monetary value of the trade (typically in USD)

Product:
Products classified using HS codes

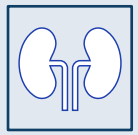
Year:
The year the trade occurred

The ESCWA Trade Data Portal (ETDP)

https://etdp.unescwa.org

- A comprehensive resource for accessing and analyzing trade data for the Arab region.
- Provides detailed data on imports and exports by product and partner country.
- **Functionality:** Allows users to query, filter, and download trade data for analysis.

Characteristics of the Data on ETDP



Data is reported using the Harmonized Commodity Description and Coding System (HS)

Data is reported at the HS6 level

Trade data is standardized to the HS17 version of the nomenclature to enable time series analysis

Day 1 - Exercises

- Practice defining variables and using operators.
- Write simple functions and control flow statements.
- Create and manipulate lists, dictionaries, and tuples.

UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity Dignified Life

Day 2: Exploring Trade Data

Reminder (Merchandise Trade Data)

- Data on the exchange of goods between countries.
- **Key Components:**

Reporter:
The country reporting the data

Trade Flow:
Exports, Imports, Re-Exports

Partner:
The country with whom the trade is conducted.

Value:
The monetary value of the trade (typically in USD)

Product:
Products classified using HS codes

Year:
The year the trade occurred

Reminder (ETDP)

https://etdp.unescwa.org

- A comprehensive resource for accessing and analyzing trade data for the Arab region.
- Provides detailed data on imports and exports by product and partner country.
- **Functionality:** Allows users to query, filter, and download trade data for analysis.

The 6-Digit HS Code Structure

- HS codes have a hierarchical structure where each level provides more detail. The first 6 digits are universal.

Example: Dates (`080410`)

Chapter `08` **:** Edible fruit and nuts; peel of citrus fruit or melons.

Heading `0804` **:** Dates, figs, pineapples, avocados, guavas, mangoes...

Subheading `080410` **:** -- Dates

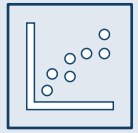
Countries can add further digits (e.g., HS-8 or HS-10) for their own tariff or statistical needs, but the first 6 digits remain consistent globally.

The Workflow

Today, we will follow a standard data analysis workflow:

- **Import Libraries:** Load the necessary tools (Pandas and Plotly).
- **Load Data:** Read the trade data from CSV files into Pandas DataFrames.
- **Ask Questions:** Formulate questions about the data.
- **Manipulate & Analyze:** Use Pandas to aggregate and calculate results to answer our questions.
- **Visualize:** Use Plotly Express to create charts that communicate our findings.

What Insights Are We Looking to Extract from Our Data?



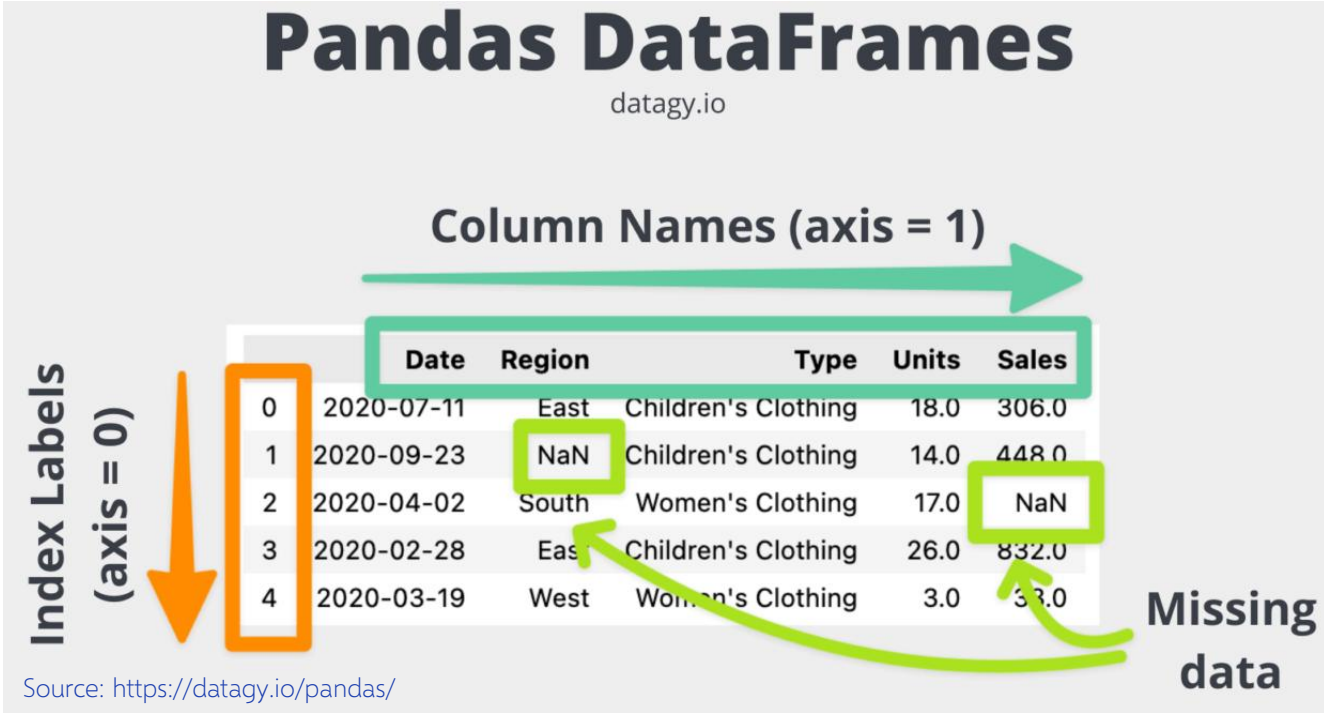
Understand the overall trade patterns in the Arab region

Understand the Arab region’s trade activity with the United States

Understand the Arab region’s trade vulnerability (Day 3)

Loading Data with Pandas

- We use pd.read_csv() or pd.read_excel() to load our data (depending if the data is csv or excel).
- The result is a **DataFrame**, which is like spreadsheet or a table.



Source: https://datagy.io/pandas/

Aggregating with .groupby()

To find the total export value for each product, we use the "split-apply-combine" strategy with `.groupby()`.

1. **Split:** Group the DataFrame by product code (HS4Id).
2. **Apply:** For each product group, calculate the sum of its `TradeValue`.
3. **Combine:** Collect the results into a new DataFrame.

Ploting with Plotly Express

- Plotly is a powerful and simple tool to plot data

```
import plotly.express as px

# Create an interactive bar chart
fig = px.bar(top_product_exports,
             y='TradeValue',
             x='description_wrapped', # The product description
             title="Top Products Exported By The Arab Region")

# Display the figure
fig.show()
```

Ploting with Plotly Express

```
px.bar(data_frame=my_df, x="Column_A", y="Column_B", color="Column_C", title="My Chart Title")
```

px.bar(): The function for the chart type (e.g., .scatter(), .line(), .pie()).

data_frame: The Pandas DataFrame that holds your data.

x and y: The names of the columns to plot on the axes.

color (Optional): The name of a column to use for coloring the marks.

title (Optional): The text to display as the chart's title.

Day 2 - Exercises

- Practice data exploration:
  - Analyze the Arab region's trade relationship with China.
  - Explore the non-oil/gas economy by excluding relevant HS codes.
  - Conduct an intra-Arab trade analysis to see which countries trade most with each other

UNITED NATIONS
الاسكوا
ESCWA

Shared Prosperity Dignified Life

Day 3: Calculating Indicators and Applying Machine Learning

Why Calculate Trade Indicators?

- They provide a standardized way to measure and compare economic performance.
- They help us understand complex phenomena like economic openness, trade balance, and regional integration.
- They are essential for evidence-based policymaking.

Indicator 1: Trade Openness Index

Measures how reliant a country's economy is on international trade.

$$\text{Trade Openness} = \frac{\text{Exports} + \text{Imports}}{\text{GDP}} \times 100$$

Indicator 2: Foreign Trade Coverage Ratio

Shows if a country's export earnings can cover its import costs.
A value > 100 indicates a trade surplus.

$$\text{Coverage Ratio} = \frac{\text{Total Exports}}{\text{Total Imports}} \times 100$$

Indicator 3: Intra-regional Exports Share

Measures how much of a country's exports stay within the Arab region.

$$\text{Intra-regional Exports Share} = \frac{\text{Exports to Arab Region}}{\text{Total Exports}} \times 100$$

Indicator 4: Regional Market Share

Shows a country's dominance in the intra-regional export market.

$$\text{Regional Market Share} = \frac{\text{Intra-Regional Exports of Country}}{\text{Total Intra-Regional Exports of All Arab Countries}} \times 100$$

Introduction to Unsupervised Machine Learning

What is it? A type of machine learning where the algorithm learns patterns from unlabeled data.

Clustering: The most common unsupervised learning task. The goal is to group similar data points together.

Our Goal: We will use clustering to automatically group Arab countries based on their trade indicator profiles, without any preconceived notions.

K-Means Clustering & The Elbow Method

K-Means is an algorithm that groups data into a pre-defined number (*K*) of clusters.

how do we choose *K*? We use the **Elbow Method**.

We plot the algorithm's performance for different values of *K*. The "elbow" of the curve—where performance gains start to diminish—suggests the optimal number of clusters.

Day 3 - Exercises

- Practice indicator calculation and clustering:
  - Calculate the Herfindahl-Hirschman Index (HHI) for both product and market concentration.
  - Use these new concentration indices to perform another round of clustering.
  - Analyze and interpret your new cluster results.

Requests to ESCWA

- If you found this training helpful and you would like to request trainings or technical assistance, kindly send an official letter from your country to the following:
- Mr. Karim Khalil, Secretary Of Committee, Political Affairs: khalil31@un.org
- Include in the cc the following:
  - Mr. Tarik Alami, Acting Director of Cluster 4: alamit@un.org
  - Mr. Mohamad Hossary, Data Scientist, Trade Statistics: mohamad.hossary@un.org